

Robust Moving Horizon Estimation with Huber Loss for GNSS-Based Vehicle Localization under Outliers

H. S. U. Hernández , A. Capovilla , L. L. Giovanini , and G. M. Sánchez 

Abstract—This work presents a robust *MHE* formulation using the Huber loss (*MHE-H*), which replaces the quadratic residual penalty with a Huber-weighted term, preserving least-squares efficiency for nominal errors while bounding the influence of large deviations. Robustness is characterized via an influence-function and gross-error sensitivity analysis, and a practical robust global asymptotic stability (*p-RGAS*) result is established under incremental input/output-to-state stability (*i-IOSS*) detectability, linking robustness and stability within a unified nonlinear framework. Guidelines relating the Huber threshold to noise scale and horizon length are also derived. Tested on seven vehicle trajectories combining public and independently collected GNSS/IMU data under real and synthetic outliers, *MHE-H* achieves the lowest worst-case position and heading errors versus *EKF*, robust *EKF*, *MHE*, and redescending-loss *MHE*, at moderate added computational cost.

Link to graphical and video abstracts, and to code:
<https://latam.ti.eer9.org/index.php/transactions/article/view/10917>

Index Terms—moving horizon estimation, state-estimation, outliers, Huber-loss

I. INTRODUCTION

STATE estimation is fundamental to feedback control, monitoring, and optimization of engineering systems, where model-based estimators — Kalman filtering (*KF*) [1], particle filters [2], and Moving Horizon Estimation (*MHE*) [3] — combine system dynamics with measurements to reconstruct the system state. Real measurements, however, are corrupted by sensor faults, unmodeled dynamics, external disturbances, and large deviations (*outliers*) that violate the Gaussian assumptions underlying classical quadratic-cost estimators, degrading accuracy or causing divergence [4]. *MHE* is particularly well suited to address this challenge because it naturally incorporates constraints and nonlinear dynamics while estimating the current state via a constrained optimization over a moving window of recent measurements [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Guillermo Valencia-Palomo (*Corresponding author: Hugo Sacha Uriel Hernandez*).

The authors wish to thank the *Universidad Nacional de Litoral*, the *Agencia Nacional de Promoción Científica y Tecnológica* and the *Consejo Nacional de Investigaciones Científicas y Técnicas* (CONICET) from Argentina, for their support.

Hugo Sacha Uriel Hernandez, A. Capovilla, L. L. Giovanini, and G. M. Sánchez are with Research Institute for Signals, Systems and Computational Intelligence, UNL, Santa Fe, Argentina (e-mails: hsuhernandez@sinc.unl.edu.ar, acapovilla@sinc.unl.edu.ar, lgiovanini@sinc.unl.edu.ar, and gsanchez@sinc.unl.edu.ar).

Existing robust *MHE* strategies fall into two main categories. Combinatorial approaches, such as leave-one-out *MHE* [5]–[7], improve robustness by solving a sequence of least-squares problems that omit potentially corrupted measurements, at the expense of increased computational cost. Divergence-based approaches, such as the Bayesian reformulation of *MHE*, down-weight outliers via a robust divergence metric rather than discarding measurements outright [8]. A third line of work incorporates the Huber loss directly into the *MHE* cost [9]–[13], which has been shown to handle sporadic outliers in linear systems and to robustly estimate tethered-aircraft pose against ℓ_2 -norm *MHE* and an *EKF*.

Despite this progress, existing robust *MHE* formulations largely lack a quantitative characterization of how robust the resulting estimator is, and under what conditions its stability is preserved. Statistical robustness theory offers two well-established tools for this purpose: the *influence function*, which measures the local sensitivity of an estimator to infinitesimal contamination, and the *gross-error sensitivity*, its supremum over all contamination points, which certifies whether an estimator’s worst-case bias remains bounded.

Recent work has begun applying these tools to *MHE*, but a joint treatment that couples influence-function-based robustness analysis with formal stability guarantees—particularly via incremental input/output-to-state stability (*i-IOSS*)—remains limited for Huber-based *MHE* applied to real-world localization problems [14]–[16].

Building on these results, this paper proposes a robust Huber-loss *MHE* method (*MHE-H*) for nonlinear systems. This paper makes three contributions. First, it formulates a Huber-loss *MHE* for nonlinear systems with measurement outliers. Second, it derives a local *KKT*-based contamination-sensitivity bound and conditions for finite gross-error sensitivity. Third, under exponential *i-IOSS* and specified arrival-cost conditions, it establishes a practical estimation-error bound and validates the method on GNSS/IMU localization data.

The rest of the paper is structured as follows. Section II introduces the notation, the problem setting, and the Huber loss. Section III formulates the outlier-robust *MHE* problem. Section IV examines robustness through the influence function and gross-error sensitivity, and then studies stability. Section V presents the experimental results on real GNSS data, combining naturally occurring outliers from field recordings with synthetic contamination under controlled conditions. Section VI closes the paper and highlights future directions.

II. PRELIMINARIES AND SETUP

A. Notations

Let $\mathbb{Z}$, $\mathbb{Z}_{[a,b]}$, and $\mathbb{Z}_{\geq a}$ denote the integers, the integer interval $[a, b]$ with $b > a$, and the integers larger or equal than a , respectively. Boldface symbols denote finite ($\mathbf{w}_n := \{w_1, \dots, w_n\}$) or infinite ($\mathbf{w} := \{w_1, \dots, w_n, \dots\}$) sequences. For $x \in \mathbb{R}^{n_x}$, $\|x\|_A^2 := x^\top A x$ and $|x|_A := \sum_{j=1}^{n_x} a_j |x_j|$ denote the A -weighted ℓ_2 and ℓ_1 norms, respectively; $\|x\|$ and $\|x\|_{[a,b]}$ denote the supremum norm over $\mathbb{Z}_{\geq 0}$ and $\mathbb{Z}_{[a,b]}$; $x^\top$ is the transpose; $L_2([0, \infty), \mathbb{R}^n)$ is the space of summable n -dimensional sequences; and I_n is the $n \times n$ identity matrix. A sequence $\mathbf{w}$ is bounded if $\|\mathbf{w}\| < \infty$, and $\mathcal{W}(w_{\max}) := \{\mathbf{w} : \|\mathbf{w}\| \leq w_{\max}\}$; it converges to zero if $|w_k| \rightarrow 0$ as $k \rightarrow \infty$. Functions $\gamma \in \mathcal{K}$, $\mathcal{K}_\infty$, $\zeta \in \mathcal{L}$, and $\beta \in \mathcal{KL}$ follow the standard comparison-function definitions.

B. Problem Statement

Consider the discrete-time system

$$\begin{aligned} x_{k+1} &= f(x_k, w_k) & x_0 &= \bar{x}_0, \\ y_k &= h(x_k) + v_k + o_k, \end{aligned} \quad (1)$$

with state $x_k \in \mathcal{X} \subseteq \mathbb{R}^{n_x}$, output $y_k \in \mathcal{Y} \subseteq \mathbb{R}^{n_y}$, process disturbance $w_k \in \mathcal{W} \subseteq \mathbb{R}^{n_w}$, measurement noise $v_k \in \mathcal{V} \subseteq \mathbb{R}^{n_y}$, and outliers $o_k \in \mathcal{O} \subseteq \mathbb{R}^{n_y}$, all in $L_2([0, \infty))$ and the sets $\mathcal{X}, \mathcal{Y}, \mathcal{V}, \mathcal{W}, \mathcal{O}$ compact, convex, and containing the origin. The outlier sequence o_k is not modelled probabilistically; its effect is absorbed into the measurement residual $\hat{v}_k$. Consequently, robustness is achieved through a loss function rather than an explicit stochastic outlier model.

In the following, we assume f is locally Lipschitz and h is continuous. The initial state x_0 is unknown but bounded around a prior $\bar{x}_0$ with error $e_{\max} \geq 0$

$$\bar{x}_0 \in \mathcal{X}_0 := \{\bar{x}_0 : \|x_0 - \bar{x}_0\| \leq e_{\max}\}, \quad \mathcal{X}_0 \subseteq \mathcal{X}. \quad (2)$$

Remark 1: The framework extends readily to systems with known input u_k , i.e., $x_{k+1} = f(x_k, u_k, w_k)$, since u_k is measured exactly and identical across all candidate trajectories, introducing no decision variables. We therefore omit it here, retaining the notation of (1), with f time-varying through u_k and all regularity and detectability constants holding uniformly over $\mathcal{U} \subseteq \mathbb{R}^{n_u}$.

Given measurements $\mathbf{y}_{[k-N,k]}$, MHE estimates $\hat{x}_{k|k}$ by solving

$$\begin{aligned} \min_{\hat{x}_{k-N|k}, \hat{\mathbf{w}}_{N|k}} \quad & \Gamma_{k-N|k}(\chi) + \sum_{j=k-N}^{k-1} \ell(\hat{w}_{j|k}, \hat{v}_{j|k}) \\ & + \|\hat{v}_{k|k}\|_{R^{-1}}^2 \\ \text{s.t.} \quad & \begin{cases} \chi = \hat{x}_{k-N|k} - \bar{x}_{k-N}, \\ \hat{x}_{j+1|k} = f(\hat{x}_{j|k}, \hat{w}_{j|k}), \quad j \in \mathbb{Z}_{[k-N,k-1]}, \\ y_j = h(\hat{x}_{j|k}) + \hat{v}_{j|k}, \quad j \in \mathbb{Z}_{[k-N,k]}, \\ \hat{x}_{j|k} \in \mathcal{X}, \hat{w}_{j|k} \in \mathcal{W}, \hat{v}_{j|k} \in \mathcal{V}, \end{cases} \end{aligned} \quad (3)$$

where $\hat{x}_{j|k}$ and $\hat{w}_{j|k}$ are the estimated state and process noise at time j using the measurements $\mathbf{y}_{[k-N,k]}$ available at time k , and $\hat{v}_{j|k} = y_j - h(\hat{x}_{j|k})$ is the residual. The arrival cost

$$\Gamma_{k-N|k}(\chi) = \|\hat{x}_{k-N|k} - \bar{x}_{k-N}\|_{P_{k-N|k}^{-1}}^2 \quad (4)$$

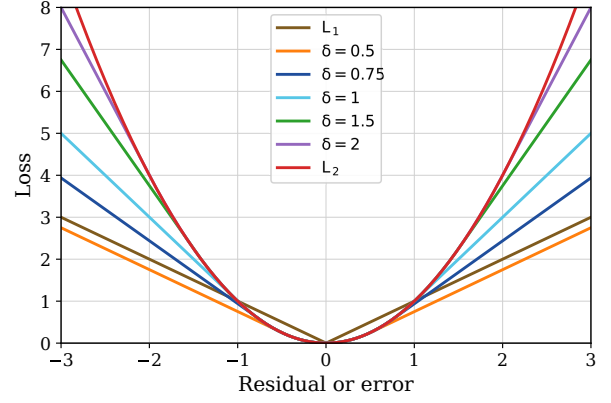


Fig. 1. Comparison of $L_\delta(\hat{v}_{i,j|k}, \delta_i)$ for different values of δ_i .

summarizes past data via the recursively updated prior $\bar{x}_{k-N}$ and information matrix $P_{k-N|k}^{-1}$, while the stage cost

$$\ell(\hat{w}_{j|k}, \hat{v}_{j|k}) = \|\hat{w}_{j|k}\|_{Q^{-1}}^2 + \|\hat{v}_{j|k}\|_{R^{-1}}^2 \quad (5)$$

penalizes the process disturbance (regularization) and residual, with $Q, R \succ 0$ being the disturbance and noise covariance matrices.

Estimator (3) is robustified by replacing the residual penalty in (5) with a robust loss function that downweights the $\hat{v}_{j|k}$ induced by outliers $o_{j|k}$

$$\ell_\eta(\hat{w}_{j|k}, \hat{v}_{j|k}) = \|\hat{w}_{j|k}\|_{Q^{-1}}^2 + |\boldsymbol{\eta}(\hat{v}_{j|k})|_D, \quad (6)$$

where $\boldsymbol{\eta}(\hat{v}_{j|k}) = [\eta_i(\hat{v}_{i,j|k})]_{i=1}^{n_y}$ collects the downweighted residuals at time j and $\eta_i \in \mathcal{K}_\infty$.

C. Robust Loss Function

The Huber function combines quadratic and linear penalties, retaining least-squares efficiency and differentiability for small residuals ($|\hat{v}_{i,j|k}| \leq \delta_i$), while limiting the influence of large ones ($|\hat{v}_{i,j|k}| > \delta_i$) [17]

$$L_{\delta_i}(\hat{v}_{i,j|k}, \delta_i) = \begin{cases} \frac{1}{2} |\hat{v}_{i,j|k}|^2, & |\hat{v}_{i,j|k}| \leq \delta_i, \\ \delta_i |\hat{v}_{i,j|k}| - \frac{1}{2} \delta_i^2, & |\hat{v}_{i,j|k}| > \delta_i, \end{cases} \quad (7)$$

where δ_i sets the quadratic-to-linear transition threshold per measurement channel, balancing sensitivity and robustness [18]. Fig. 1 illustrates $L_\delta(\hat{v}_{i,j|k}, \delta_i)$ for representative values of δ_i , showing how larger thresholds recover the quadratic (least-squares) regime while smaller ones approach the ℓ_1 loss [19].

III. OUTLIER-ROBUST MHE

We now formulate a robust MHE that accurately estimates the states of system (1) from measurements corrupted by outliers. Replacing the quadratic residual penalty in (5) by the Huber-weighted term (7) yields the stage cost

$$\ell_\delta(\hat{w}_{j|k}, \hat{v}_{j|k}, \delta) = \|\hat{w}_{j|k}\|_{Q^{-1}}^2 + |\mathbf{L}_\delta(\hat{v}_{j|k}, \delta)|_D \quad (8)$$

where $D = \text{diag}(d_1, \dots, d_{n_y}) \succ 0$ is used for residual-whitening and cost weighting, and $\mathbf{L}_\delta(\hat{v}_{j|k}, \delta) = [L_{\delta_i}(\hat{v}_{i,j|k}, \delta_i)]_{i=1}^{n_y}$ collects the downweighted residuals at time

$j \in \mathbb{Z}_{[k-N,k]}$. The first term regularizes $\hat{x}_{[k-N+1,k]}$ via the disturbance energy, while the second downweights outlier-corrupted residuals rather than discarding them outright.

Substituting $\ell_\delta(\cdot)$ for $\ell(\cdot)$ in (3) gives the proposed optimization problem

$$\begin{aligned} \min_{\hat{x}_{k-N|k}, \hat{w}_{N|k}} \quad & \Gamma_{k-N|k}(\chi) + \sum_{j=k-N}^{k-1} \ell_\delta(\hat{w}_{j|k}, \hat{v}_{j|k}, \delta) \\ & + |\mathbf{L}_\delta(\hat{v}_{k|k}, \delta)|_D \\ \text{s.t.} \quad & \begin{cases} \chi = \hat{x}_{k-N|k} - \bar{x}_{k-N}, \\ \hat{x}_{j+1|k} = f(\hat{x}_{j|k}, \hat{w}_{j|k}), \quad j \in \mathbb{Z}_{[k-N,k-1]}, \\ y_j = h(\hat{x}_{j|k}) + \hat{v}_{j|k}, \quad j \in \mathbb{Z}_{[k-N,k]}, \\ \hat{x}_{j|k} \in \mathcal{X}, \hat{w}_{j|k} \in \mathcal{W}, \hat{v}_{j|k} \in \mathcal{V}. \end{cases} \end{aligned} \quad (9)$$

Robustness against outliers is achieved by construction: For large residuals, the Huber penalty grows only linearly rather than quadratically, so outlying measurements no longer dominate the optimization cost as they do under a pure ℓ_2 penalty. The next section formalizes this property via the influence function and gross-error sensitivity, and establishes stability of (9) under the *i-IOSS* detectability condition.

Assumption 1: In the following we assume that the optimization problem (9) is well posed and a solution exists $\forall k \in \mathbb{Z}_{\geq 0}$.

IV. THEORETICAL PROPERTIES

This section establishes the main theoretical properties of estimator (9). Section IV-A characterizes its statistical robustness through the influence function and gross-error sensitivity. Section IV-B establishes *p-RGAS* of the estimation error. Additional assumptions on the problem data are introduced as needed.

Assumption 2: Let $\hat{x}_{N|k}^*$ denote a local minimizer of $J(\mathbf{y}_{N|k}, \hat{\mathbf{w}}_{N|k}, \hat{x}_{k-N|k}, \delta)$ in (9), with associated optimal residuals $\hat{\mathbf{v}}_{N|k}^*$. Assume the non-degeneracy condition

$$|\hat{v}_{i,j|k}^*| \neq \delta_i \quad \forall i \in \mathbb{Z}_{[1,n_y]}, j \in \mathbb{Z}_{[k-N+1,k]}, \quad (10)$$

then there exists a neighbourhood $\mathcal{N}(\hat{x}_{N|k}^*)$ on which each residual component $\hat{v}_{i,j|k}$ keeps a constant sign relative to its threshold ($\forall \hat{v}_{i,j|k} \in \mathcal{N} : |\hat{v}_{i,j|k}| < \delta_i$ or $|\hat{v}_{i,j|k}| > \delta_i$). Consequently $J(\mathbf{y}_{N|k}, \hat{\mathbf{w}}_{N|k}, \hat{x}_{k-N|k}, \delta)$ is twice continuously differentiable on $\mathcal{N}$ with respect to $\hat{x}_{N|k}^*$, and its Hessian is nonsingular at $\hat{x}_{N|k}^*$.

Condition (10) holds generically since the set $\{\hat{v}_{i,j|k} \in \mathcal{V} : |\hat{v}_{i,j|k}| = \delta_i\}$ has Lebesgue measure zero in residual space, and the underlying measurement noise has a continuous distribution, so violations of (10) occur with probability zero at any isolated local minimiser.

A. Robustness analysis

Robustness characterizes an estimator's resilience to outliers and deviations from the assumed distributional model. We use Hampel's influence function [20] as a local robustness diagnostic, quantifying how a small contamination in $\mathbf{y}_{[k-N,k]}$ perturbs the estimator output.

Remark 2: Assumption 1 ensures the existence of a solution to (9) but not uniqueness. The influence function

should be understood as evaluated at the locally unique, non-degenerate minimizer satisfying (10), restricting Hampel's classical framework to a local neighbourhood.

The robustness analysis will be carried out using a linear-time-varying (*LTV*) approximation of the nonlinear system (1) obtained by linearizing $f(\cdot)$ and $h(\cdot)$ around the estimated-state trajectory $\hat{x}_{N|k}$ to characterize the local robustness properties. Since the influence function is inherently a local (first-order) sensitivity measure, the *LTV* linearization around $\hat{x}_{N|k}$ is not an additional source of approximation error but the natural representation of the problem in the neighborhood $\mathcal{N}(\hat{x}_{N|k})$ where Assumption 2 holds; hence it suffices to certify the robustness properties.

Influence function: The influence function of an estimator at a contamination point y_s^+ is given by

$$\mathbb{IF}(y_s^+, T, F) = \lim_{\epsilon \rightarrow 0^+} \frac{T((1-\epsilon)F + \epsilon y_s^+) - T(F)}{\epsilon}, \quad (11)$$

where $\hat{x}_{k|k} = T(F)$ denotes the solution of (9) viewed as a functional T that maps the joint measurement distribution F (and the state prior) to the current state estimate $\hat{x}_{k|k}$. It is characterized locally by the first-order stationarity conditions of (9)

$$\nabla_x J(\mathbf{y}_{N|k}, \hat{\mathbf{w}}_{N|k}, \hat{x}_{k-N|k}) + \lambda^T \nabla c(\hat{\mathbf{w}}_{N|k}, \hat{x}_{k-N|k}) = 0 \quad (12)$$

where $c(\hat{\mathbf{w}}_{N|k}, \hat{x}_{k-N|k})$ denotes the dynamic constraints. For estimation problem (9), this condition becomes

$$P_{k-N|k}^{-1} \chi - C_{k-N|k}^T D \Psi(\hat{v}_{k-N|k}) - A_{k-N|k}^T \lambda_{k-N} = 0, \quad (13)$$

for $k-N < j < k$

$$-C_{j|k}^T D \Psi(\hat{v}_{j|k}) + \lambda_{j-1} - A_{j|k}^T \lambda_j = 0, \quad (14)$$

$$-C_{k|k}^T D \Psi(\hat{v}_{k|k}) + \lambda_{k-1} = 0, \quad (15)$$

$$Q^{-1} \hat{w}_{j|k} - W_{j|k}^T \lambda_j = 0, \quad (16)$$

$$\hat{x}_{j+1|k} - A_{j|k} \hat{x}_{j|k} - W_{j|k} \hat{w}_{j|k} = 0, \quad (17)$$

where $A_{j|k}$, $W_{j|k}$ and $C_{j|k}$ are the linearized matrices around the state estimated trajectory $\hat{x}_{N|k}$ and

$$\Psi(\hat{v}_{j|k}, \delta) = [\psi(\hat{v}_{i,j|k}, \delta_i)]_{i=1}^{n_y}, \quad (18)$$

with

$$\psi_{i,j|k} = \frac{\partial L_{\delta_i}(\hat{v}_{i,j|k}, \delta_i)}{\partial \hat{v}_{i,j|k}} = \begin{cases} \hat{v}_{i,j|k}, & |\hat{v}_{i,j|k}| \leq \delta_i, \\ \delta_i \text{sign}(\hat{v}_{i,j|k}), & |\hat{v}_{i,j|k}| > \delta_i. \end{cases} \quad (19)$$

To eliminate the process-noise and multiplier variables from the stationarity system (13)-(17), we first solve (16) for $\hat{w}_{j|k}$ and substitute it into the dynamics (17), giving the multiplier recursion

$$\lambda_j = S_j (\hat{x}_{j+1|k} - A_{j|k} \hat{x}_{j|k}) \quad (20)$$

where $S_j = (W_{j|k} Q W_{j|k}^T)^{-1}$. Substituting this expression for λ_j back into the state-stationarity equations (13)-(15) eliminates both $\hat{w}_{j|k}$ and λ_j , leaving a single reduced system expressed purely in terms of the state trajectory $\hat{x}_{N|k}$ and the residuals $\hat{\mathbf{v}}_{N|k}$. Collecting these equations compactly yields (12) in the form

$$H_Q \hat{x}_{N|k} + H_R (\hat{\mathbf{v}}_{N|k}, \delta) = 0, \quad (21)$$

where $H_R(\hat{v}_{N|k}, \delta)$ collects the Huber-weighted residual terms

$$H_R(\cdot) = [C_{k-N|k}^T D \Psi(\hat{v}_{k-N|k}) \cdots C_{k|k}^T D \Psi(\hat{v}_{k|k})] \quad (22)$$

and $H_Q \in \mathbb{R}^{(N+1)n_x \times (N+1)n_x}$ is a block-tridiagonal matrix whose blocks h_{Qji} , with $j, i \in \mathbb{Z}_{[k-N, k]}$ indexing the states $\hat{x}_{j|k}$ of the stacked trajectory, are given by

$$\begin{aligned} h_{Qk-N, k-N} &= P_{k-N|k}^{-1} + A_{k-N|k}^T S_{k-N} A_{k-N|k}, \\ h_{Qk-N, k-N+1} &= -A_{k-N|k}^T S_{k-N}, \end{aligned} \quad (23)$$

for $k-N < j < k$

$$h_{Qji} = \begin{cases} -S_i A_{i|k} & i = j-1, \\ S_{i-1} + A_{i|k}^T S_i A_{i|k} & i = j, \\ -A_{i-1|k}^T S_{i-1} & i = j+1, \\ 0 & \text{otherwise,} \end{cases} \quad (24)$$

and

$$\begin{aligned} h_{Qk, k-1} &= -S_{k-1} A_{k-1|k}, \\ h_{Qk, k} &= S_{k-1}, \end{aligned} \quad (25)$$

where $S_j = (W_{j|k} Q W_{j|k}^T)^{-1}$ for $j \in \mathbb{Z}_{[k-N, k-1]}$.

By Assumption 2, the reduced stationarity map in (21) is continuously differentiable in a neighborhood of $\hat{x}_{N|k}$, and its Jacobian $\nabla_x^2 J(\delta)$ is nonsingular there; these two conditions are precisely what allow the implicit-function theorem to be applied.

The estimator (9) solves (12) under the nominal data-generating distribution F . Now, consider the contaminated distribution $F_\epsilon = (1 - \epsilon)F + \epsilon y_s^+$ obtained from replacing $y_s^+ \in \hat{y}_{N|k}$ with an outlying $y_s = y_s^+$. Differentiating the reduced stationarity system with respect to the contamination parameter and applying the implicit-function theorem yields the influence function of (9) given by

$$\mathbb{IF}(y_s^+, T, F) = (\nabla_x^2 J(\delta))^{-1} C_{s|k}^T D \Psi(\hat{v}_{s|k}, \delta) \quad (26)$$

where

$$\begin{aligned} \nabla_x^2 J(\delta) &= H_Q + \frac{\partial}{\partial \delta} H_R(\hat{v}_{N|k}, \delta), \\ \frac{\partial}{\partial \delta} H_R(\hat{v}_{N|k}, \delta) &= [C_{k-j|k}^T D \Theta(\hat{v}_{k-j|k}, \delta) C_{k-j|k}]_{j=0}^N, \\ \Theta(\hat{v}_{j|k}, \delta) &= [\theta(\hat{v}_{i,j|k}, \delta_i)]_{i=1}^{n_y} \end{aligned} \quad (27)$$

and

$$\theta_{i,j|k} = \frac{\partial \psi(\hat{v}_{i,j|k}, \delta_i)}{\partial \hat{v}_{i,j|k}} = \begin{cases} 1, & |\hat{v}_{i,j|k}| \leq \delta_i, \\ 0. & \end{cases} \quad (28)$$

Equations (26) and (27) show that the influence function is governed by the clipped derivative $\psi_{i,j|k}$ and its switching derivative $\theta_{i,j|k}$, both controlled by the Huber threshold δ (Fig. 2). Isolating the contribution of a single residual, the stage-wise influence scales with $\psi(\hat{v}_{i,k-j|k})$, modulated by $(\nabla_x^2 J(\delta))^{-1}$, the system dynamics $A_{j|k}$, and the measurement sensitivity through $C_{j|k}$. Three regimes follow directly:

- For small residuals ($|\hat{v}_{i,j|k}| \leq \delta_i$), the influence is approximately linear, as in least-squares, modulated by the system dynamics,
- For large residuals ($|\hat{v}_{i,j|k}| > \delta_i$), the influence saturates at δ , up to the linear transformation by $(\nabla_x^2 J(\delta))^{-1}$, bounding the effect of any single outlier,

- As $\delta \rightarrow \infty$, $\Theta_{j|k} \rightarrow I$ and $\nabla_x^2 J(\delta)$ recovers the Kalman-filter information matrix [21].

This limiting behavior of *MHE-H* is consistent with classical full-information estimation theory, which showed that for unconstrained linear systems with Gaussian disturbances, the moving-horizon/full-information cost Hessian coincides with the inverse of the Kalman filter's error covariance [21]. In our setting, the arrival-cost weight $P_{k-N|k}^{-1}$ and stage-cost curvature $A_{j|k}^T S_j A_{j|k}$ reduce to the same Riccati-recursion terms as the unconstrained linear *MHE* studied by Rao [22] when the Huber threshold $\delta \rightarrow \infty$, i.e., the Huber loss recovers the pure ℓ_2 penalty.

Our result therefore generalizes this classical equivalence to a nonlinear robust loss: in equation (26), $(\nabla_x^2 J(\delta))^{-1}$ plays the same role as the steady-state Kalman covariance in the unconstrained linear case. The Kalman information matrix is recovered as $\delta \rightarrow \infty$, while finite δ values yield bounded influence without departing from this full-information interpretation. It reduces the allowable influence of large residuals—improving robustness at the cost of efficiency under purely Gaussian noise—while recovering standard least-squares behavior with unbounded influence as δ increases.

Gross-error sensitivity: In the previous subsection, we derived the influence function of *MHE-H* under generic point contamination. To assess the estimator's robustness, we analyze worst-case conditions over all admissible contamination points, output channels, and residual magnitudes. Specifically, we consider the supremum of the influence function over all possible contamination locations. This leads to the concept of *gross-error sensitivity* [20], defined by

$$\begin{aligned} \gamma(\delta) &= \sup_{y_s^+ \in \mathcal{Y}} \|\mathbb{IF}(y_s^+, T, F)\|, \\ &= \sup_{y_s^+ \in \mathcal{Y}} \|(\nabla_x^2 J(\delta))^{-1} C_{s|k}^T D \Psi(\hat{v}_{s|k}, \delta)\|. \end{aligned} \quad (29)$$

To find a bound to $\gamma(\delta)$, we split the norm in Eq. (29) into three factors: the Hessian-inverse sensitivity, the weighting matrix D , and the Huber score Ψ , each bounded separately below

$$\|\mathbb{IF}(y_s^+, T, F)\| \leq \|(\nabla_x^2 J(\delta))^{-1} C_{s|k}^T\| \|D\| \|\Psi(\hat{v}_{s|k}, \delta)\|. \quad (30)$$

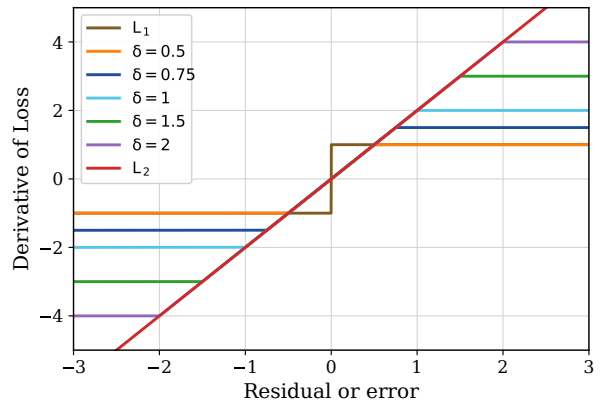


Fig. 2. Comparison of the influence function $\psi_{i,j|k}$ for the L_1 , Huber and L_2 losses for different values of the Huber threshold δ .

Since D is a bounded diagonal matrix, and each component of $\Psi(\hat{v}_{s|k}, \delta)$ is clipped at δ_i ,

$$\|D\| \|\Psi(\hat{v}_{s|k}, \delta)\| \leq \bar{d} \bar{\delta} \sqrt{n_y}. \quad (31)$$

where

$$\bar{d} := \max_{i \in [1, n_y]} d_i, \quad \bar{\delta} := \max_{i \in [1, n_y]} \delta_i. \quad (32)$$

Substituting

$$\gamma(\delta) \leq \bar{d} \bar{\delta} \sqrt{n_y} \sigma_{\max}((\nabla_x^2 J(\delta))^{-1} C_{s|k}^T) \quad (33)$$

$$\leq \bar{d} \bar{\delta} \sqrt{n_y} (\lambda_{\min}^{-1}(\nabla_x^2 J(\delta))) \sigma_{\max}(C_{s|k}^T). \quad (34)$$

As stated in Section IV-A, as $\delta \rightarrow \infty$, $\nabla_x^2 J(\delta) \rightarrow H_\infty$, a fixed δ -independent matrix which is the Kalman-filter information matrix.

Taking the limit in (33) as $\delta \rightarrow \infty$, using $\nabla_x^2 J(\delta) \rightarrow H_\infty$ from Section IV-A,

$$\lim_{\delta \rightarrow \infty} \gamma(\delta) \leq \frac{\bar{d} \sqrt{n_y} \sigma_{\max}(C_{s|k}^T)}{\lambda_{\min}(H_\infty)} \lim_{\delta \rightarrow \infty} \bar{\delta} = \infty. \quad (35)$$

That is, $\gamma(\delta)$ diverges linearly in δ , correctly recovering the unbounded gross-error sensitivity of the quadratic-cost (least-squares) MHE in this limit, and consistent with the least-squares result of Cao et al. [8] discussed in the Conclusion. For any fixed, finite δ , by contrast, $\gamma(\delta)$ remains bounded by (33), which is the intended robustness certificate: no single contaminated measurement, however extreme, can produce an unbounded shift in the MHE-H estimate.

B. Robust Stability

We establish robust global asymptotic stability *RGAS* of the proposed estimator under exponential *i-IOSS*. The proof follows the standard *MHE* structure: (i) optimality bounds the state and disturbance estimation errors; (ii) exponential *i-IOSS* propagates these bounds to the current state error; (iii) the sliding-window update causes the arrival-cost error to decay geometrically; and (iv) combining these yields the *RGAS* bound of Theorem 1. Relative to classical proofs, the key additional ingredient is a uniform $\mathcal{K}_\infty$ lower bound on the Huber stage cost.

Definition 1: System (1) is *i-IOSS* if there exist $\beta \in \mathcal{KL}$ and $\gamma_w, \gamma_v \in \mathcal{K}$ such that, for any pair of initial conditions $z, \tilde{z} \in \mathbb{R}^{n_x}$ and any feasible disturbance and output sequences $(\mathbf{w}, \mathbf{y}), (\tilde{\mathbf{w}}, \tilde{\mathbf{y}})$,

$$|x_k - \tilde{x}_k| \leq \beta(|x_0 - \tilde{x}_0|, k) + \gamma_w(\|\mathbf{w}_{k-1} - \tilde{\mathbf{w}}_{k-1}\|) + \gamma_v(\|h(\mathbf{x}_{k-1}) - h(\tilde{\mathbf{x}}_{k-1})\|) \quad (36)$$

holds $\forall k \in \mathbb{Z}_{\geq 0}$, where $x_k = x(k, z, \mathbf{w})$, $\tilde{x}_k = x(k, \tilde{z}, \tilde{\mathbf{w}})$.

This notion of nonlinear detectability compares two trajectories and their sensitivity to external perturbations. Its exponential form is

$$|x_k - \tilde{x}_k| \leq c_0 \rho^k |x_0 - \tilde{x}_0| + c_w \|\mathbf{w}_{k-1} - \tilde{\mathbf{w}}_{k-1}\| + c_v \|\mathbf{y}_k - \tilde{\mathbf{y}}_k\|, \quad (37)$$

with $c_0, c_w, c_v > 0$ and $\rho \in (0, 1)$.

Assumption 3: System (1) is exponentially *i-IOSS* with constants c_0, ρ, c_w, c_v as in (37), and admits a Lyapunov function V satisfying

$$\alpha_1 \|x - \tilde{x}\|^2 \leq V(x, \tilde{x}) \leq \alpha_2 \|x - \tilde{x}\|^2, \quad (38)$$

$$V(f(x, w), f(\tilde{x}, \tilde{w})) \leq \rho^2 V(x, \tilde{x}) + \sigma_w \|w - \tilde{w}\|^2 + \sigma_v \|h(x) - h(\tilde{x})\|^2, \quad (39)$$

for some $\alpha_1, \alpha_2, \sigma_w, \sigma_v > 0$.

Conditions (38)–(39) provide a constructive certificate of Assumption 3, verifiable via linear matrix inequalities [23].

Definition 2: Estimator (9) is practical *RGAS* if there exist $\beta \in \mathcal{KL}$, $\gamma_w, \gamma_v \in \mathcal{K}$ and a $c \geq 0$ - the practical bound - such that, for any pair of initial conditions $z, \tilde{z} \in \mathbb{R}^{n_x}$ and any feasible disturbance and output sequences $(\mathbf{w}, \mathbf{y}), (\tilde{\mathbf{w}}, \tilde{\mathbf{y}})$,

$$|x_k - \tilde{x}_k| \leq \beta(|x_0 - \tilde{x}_0|, k) + \gamma_w(\|\mathbf{w}_{k-1} - \tilde{\mathbf{w}}_{k-1}\|) + \gamma_v(\|h(\mathbf{x}_{k-1}) - h(\tilde{\mathbf{x}}_{k-1})\|) + c \quad (40)$$

holds $\forall k \in \mathbb{Z}_{\geq 0}$, where $x_k = x(k, z, \mathbf{w})$, $\tilde{x}_k = x(k, \tilde{z}, \tilde{\mathbf{w}})$.

For estimator (9) to be robustly stable, the cost J must be compatible with the *i-IOSS* detectability bound.

Assumption 4: The arrival cost (4) satisfies, for all χ

$$\underline{p} \|\chi\|^2 \leq \Gamma_{k-N|k}(\chi) \leq \bar{p} \|\chi\|^2 \quad (41)$$

with $0 < \underline{p} \leq \bar{p} < \infty$ uniform in k , where $\underline{p} = \lambda_{\min}(P_{k-N|k}^{-1})$ and $\bar{p} = \lambda_{\max}(P_{k-N|k}^{-1})$.

The following lemma establishes analytic bounds on the Huber loss required for the stability proof.

Lemma 1: For any $\hat{v}_{i,j|k}$ and $\delta_i > 0$, the Huber loss (7) satisfies

$$\frac{1}{2} \frac{\hat{v}_{i,j|k}^2}{1 + \frac{1}{\delta_i} |\hat{v}_{i,j|k}|} \leq L_\delta(\hat{v}_{i,j|k}, \delta_i) \leq \frac{1}{2} \hat{v}_{i,j|k}^2. \quad (42)$$

Proof. See Appendix A. $\square$

Defining $\sigma_L(s) : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ by

$$\sigma_L(s) = \frac{1}{2} \frac{s^2}{1 + \frac{s}{\underline{\delta}}} \quad \text{with} \quad \underline{\delta} = \min_{i \in [1, n_y]} \delta_i, \quad (43)$$

Lemma 1 shows $\sigma_L \in \mathcal{K}_\infty$ and $L_\delta(\hat{v}_{i,j|k}, \delta_i) \geq \sigma_L(|\hat{v}_{i,j|k}|)$ — the required uniform $\mathcal{K}_\infty$ lower bound. The Huber cost is further bounded above by the purely quadratic cost

$$J \leq \Gamma_{k-N|k}(\chi) + \sum_{j=k-N}^k \|\hat{w}_{j|k}\|_{Q^{-1}}^2 + \|\hat{v}_{j|k}\|_D^2, \quad (44)$$

so the true trajectory remains feasible for the (9) optimization and the quadratic bound used in the *i-IOSS* stability proof remains valid.

Assumption 5: There exist σ_L and $\underline{d} > 0$ such that

$$|L_\delta(\hat{v}_{j|k}, \delta)|_D \geq \underline{d} \sum_{i=1}^{n_y} \sigma_L(|\hat{v}_{i,j|k}|), \quad (45)$$

where $\sigma_L \in \mathcal{K}_\infty$ and $\underline{d} = \min_{i \in [1, n_y]} d_{ii}$.

With these ingredients, we obtain sufficient conditions for *MHE* (9) to be practically robustly globally asymptotically stable (*p-RGAS*)

Theorem 1: Suppose Assumption 3 holds with $\rho \in (0, 1)$ and $c_0, c_w, c_v \in \mathbb{R}_{>0}$, Assumptions 4 and 5 holds, and the estimation horizon N satisfies

$$N \geq N_{\min} := \left\lceil -\frac{1}{2 \log \rho} \log \left(1 + \frac{c_w^2}{\underline{q} \underline{p}} + \frac{c_v^2 n_y}{\underline{d} \underline{p} \bar{\delta}} \right) \right\rceil, \quad (46)$$

then the *MHE* estimate $\hat{x}_{k|k}$ is *p-RGAS* — i.e., it satisfies the *RGAS*-type bound (47) up to a bounded, outlier-dependent offset term rather than converging exactly — such that

$$\|\hat{x}_k - x_k\| \leq \beta(\|\hat{x}_0 - x_0\|, k) + \gamma_w(\|w_{N|k}\|_\infty) + \gamma_v(\|v_{N|k}\|_\infty) + \gamma_o(\|o_{N|k}\|_\infty) + c, \quad (47)$$

where

$$\begin{aligned} \beta(s, k) &= \sqrt{6\kappa \bar{p}} c_0 \rho^{kN} s, \quad \gamma_w(s) = C_{\gamma_w} s, \\ \gamma_v(s) &= C_{\gamma_v} s, \quad \gamma_o(s) = C_{\gamma_o} s, \quad c = \sqrt{\kappa \bar{C}_o \bar{\delta}}. \end{aligned} \quad (48)$$

Proof See Appendix B. $\square$

The constants ρ, c_0, c_w, c_v can be computed via Linear Matrix Inequality (*LMI*)

$$\begin{aligned} &V(f(x, w), f(\tilde{x}, \tilde{w})) - \rho^2 V(x, \tilde{x}) \\ &- \sigma_w \|w - \tilde{w}\|^2 - \sigma_v \|h(x) - h(\tilde{x})\|^2 \preceq \mathbf{0} \end{aligned} \quad (49)$$

for $V(x, \tilde{x}) = (x - \tilde{x})^\top P_\delta (x - \tilde{x})$, solvable with standard semidefinite programming tools [15].

The horizon requirement $N_{\min}$ captures an intuitive trade-off: for highly detectable systems ($\rho \ll 1$), few measurements suffice, so $N_{\min}$ is small; for weakly detectable systems ($\rho \rightarrow 1$), a longer horizon is required. A large arrival-cost weight $P_{k-N|k}^{-1}$ (and thus $\underline{p}$) reduces $N_{\min}$, reflecting that accurate prior information lessens the need for a long horizon, whereas large disturbance-to-cost ratios $c_w^2/(\underline{q} \underline{p})$ and $c_v^2/(\underline{d} \underline{p})$ increase it.

The *p-RGAS* proof of Theorem 1 assumes access to the exact local minimizer of the *MHE* cost (Assumption 2); in practice, the experiments rely on warm-started FATROP solves that converge reliably across all tested scenarios, but formally extending the guarantee to account for finite-iteration or early-terminated solutions remains an open direction.

C. Tuning δ

There is no universally optimal choice of δ for *MHE-H*; instead, δ should be selected based on the noise scale and the required level of robustness. It governs the trade-off between **efficiency** — the estimator's variance relative to the Cramer–Rao lower bound, maximized by least-squares — and **robustness** — stability and well-behaved performance under outliers. A large δ ($\delta \in (2\sigma_v, 3\sigma_v)$) yields high efficiency, close to least-squares, down-weighting only very large residuals; a moderate δ ($\delta \in (\sigma_v, 1.5\sigma_v)$) trades some efficiency for improved rejection of frequent medium-size outliers under mildly heavy-tailed noise. A small δ ($\delta < \sigma_v$) saturates the influence function early, offering strong robustness at the cost of efficiency, treating even moderate deviations as potential outliers.

This trade-off also determines how strongly residuals propagate through the estimated trajectory. Each residual $\hat{v}_{j|k}$ influences not only $\hat{x}_{j|k}$ but also nearby states through the stage and arrival costs: a large δ allows outlier effects to

propagate over multiple states. In contrast, a large δ saturates the innovation early, anchoring the estimate more to the model and prior data. Practically, δ should also be adjusted for noise character and horizon length: if the noise is near-Gaussian and the model is accurate, a larger δ avoids unnecessarily clipping informative residuals. Likewise, with a large horizon N , many consistent residuals carry redundant information, so a smaller δ can safely clip occasional outliers; with a small N , each residual is critical for observability, so an overly small δ risks frequent clipping and a sluggish, noisy estimate.

Table I summarizes practical operating regimes, assuming residuals are normalized by a robust estimate $\hat{\sigma}_v$ of the innovation standard deviation.

V. EXPERIMENTAL VALIDATION

The presented *MHE-H* method is evaluated using real-world multisensor data from the i2Nav-Robot dataset [24], together with an additional dataset collected specifically for this work in an experimental campaign conducted by the authors using the group's own instrumented vehicle. This experimental design allows the estimator to be assessed under both naturally occurring GNSS degradations and controlled measurement contamination.

We compare against a plain EKF; a quadratic-cost *MHE* (*MHE*); a χ^2 -gated robust EKF that inflates the measurement covariance according to the residual-rejection rule of Zhou et al. [25] (EKF-Z); and an *MHE* using the redescending loss of Deniz et al. [26] (*MHE-D*).

A. Dataset

All seven sequences used in this section are outdoor trajectories. Six of them — building00, building02, street00, street01, street02, and playground00 — come from the i2Nav-Robot dataset, collected with an AgileX Ranger Mini3 ground robot instrumented with an ADIS16465 IMU (200 Hz), wheel odometry (50 Hz), and a NovAtel OEM719 GNSS receiver operated in RTK mode (1 Hz fix rate). building00 and building02 were

TABLE I
TUNING REGIMES FOR THE HUBER THRESHOLD

Regime	δ range	Efficiency	Robustness	When to use
Efficiency-oriented	$1.1\hat{\sigma}_v - 1.6\hat{\sigma}_v$	High, near least-squares	Moderate; protects against occasional anomalies	Noise nearly Gaussian, model well specified, outliers rare
Robustness-oriented	$0.7\hat{\sigma}_v - \hat{\sigma}_v$	Reduced relative to least-squares	High; limits repeated influence of large innovations	Heavy-tailed noise or frequent outliers; estimator becomes more model-dominated
Diagnostic	$\ll \hat{\sigma}_v$	Low; inefficient for state-estimation	Very high; approximates an ℓ_1 -like cost	Measurements contain frequent faults/anomalies; primarily useful for outlier detection

recorded next to tall buildings, producing a classic *urban-canyon* effect: signal blockage and multipath reflections off the surrounding structures degrade the RTK fix and are the source of the real outlier bursts used throughout this section.

Crucially, the i2Nav-Robot dataset provides its own reference (ground-truth) trajectory for every sequence, generated by a dedicated, navigation-grade GNSS/INS system that is physically independent of the NovAtel OEM719 receiver used here as the estimator's measurement, reporting centimeter-level position accuracy at high rate.

In addition to the publicly available i2Nav-Robot dataset, the authors conducted an independent field experiment specifically for this study. The resulting seventh sequence was collected on a skid-steer ground vehicle platform equipped with wheel odometry (20 Hz), an Xsens MTi-630 IMU (20 Hz), and two u-blox ZED-F9P GNSS receivers: a standalone receiver (10 Hz) used as the system's measurement, and an RTK receiver with a moving-base heading solution used to build the ground truth (position and yaw).

B. System and Measurement Models

The estimators share the 3-state kinematic model $x_k = [p_{x,k}, p_{y,k}, \psi_k]^\top$, where (p_x, p_y) is the position in the local NED frame and ψ the heading. Between consecutive GNSS epochs t_k, t_{k+1} , the wheel-odometry and the yaw increments given by the IMU are pre-integrated into a body-frame input $u_k = [\Delta s_{x,k}, \Delta s_{y,k}, \Delta \psi_k]^\top$ where $\Delta s_{x,k}$ and $\Delta s_{y,k}$ denote the incremental vehicle displacement expressed in the body frame, and $\Delta \psi_k$ is the accumulated yaw rotation over the interval. The resulting discrete-time process model is

$$p_{x,k+1} = p_{x,k} + \Delta s_{x,k} \cos \psi_k - \Delta s_{y,k} \sin \psi_k + w_{x,k}, \quad (50)$$

$$p_{y,k+1} = p_{y,k} + \Delta s_{x,k} \sin \psi_k + \Delta s_{y,k} \cos \psi_k + w_{y,k}, \quad (51)$$

$$\psi_{k+1} = \psi_k + \Delta \psi_k + w_{\psi,k}, \quad (52)$$

with w_k the process noise.

For the i2Nav sequences, GNSS provides position-only measurements according to

$$y_k = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} x_k + v_k, \quad (53)$$

For the separately recorded trajectory used in the experimental evaluation, the heading provided directly by the IMU is additionally included as a measurement. In this case, the measurement model is defined as

$$y_k = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} x_k + v_k, \quad (54)$$

The gyroscope measures the angular velocity about the vertical axis, ω_z , which is integrated over each sampling interval to obtain the corresponding yaw increment, $\Delta \psi_k$. These yaw increments are used as inputs to the process model, while the absolute heading provided by the IMU constitutes an additional measurement available to the estimator.

Because the GNSS receiver reports geodetic coordinates (latitude, longitude, altitude), all measurements and the ground truth are converted to a common local NED frame, origin at the first fix of each sequence, via the exact WGS-84 ellipsoidal transformation [27].

C. Performance Metrics

Let $e_{p,k} = \|\hat{p}_k - p_k\|_2$ denote the 2D Euclidean position error and $e_{\psi,k}$ the yaw error (wrapped to $(-180^\circ, 180^\circ]$) at the k -th valid GNSS epoch of a run. Following the translational and rotational error definitions of Zhi et al. [28], each run is summarized by four metrics:

- **ATE** (Absolute Translational Error): $\sqrt{\text{mean}_k e_{p,k}^2}$ [m];
- **MTE** (Maximum Translational Error): $\max_k e_{p,k}$ [m];
- **AYE** (Absolute Yaw Error): $\sqrt{\text{mean}_k e_{\psi,k}^2}$ [deg];
- **MYE** (Maximum Yaw Error): $\max_k e_{\psi,k}$ [deg].

Despite the name, ATE and AYE are root-mean-square, not arithmetic-mean, errors; MTE and MYE are the worst-case (peak) error observed within a run.

D. Implementation and Parameters

The EKF and EKF-Z baselines use a closed-form per-epoch update and require no nonlinear solver. The MHE, MHE-D and MHE-H estimators are implemented in Python 3.10.12 using the open-source symbolic optimization framework CasADi [29] (v. 3.7.2); the resulting nonlinear programs are solved with the FATROP solver [30]. All MHE-based runs are warm-started. All simulations were performed on desktop computer equipped with a 2.8 GHz Intel i5 with 16 GB RAM, running the 64-bit version of Ubuntu Linux 22.04.

Unless stated otherwise, all five estimators share the fixed tuning $Q = \text{diag}(10^{-3}, 10^{-3}, 1.2 \times 10^{-5})$ [$\text{m}^2, \text{m}^2, \text{rad}^2$] and $R = 0.1I_2$ [m^2] for the position-only sequences, with $P_0 = \text{diag}(1, 1, (10^\circ)^2)$. Q reflects unmodeled process error (wheel slip), since sensor-only noise is negligible by comparison. R is set conservatively above the receiver's nominal RTK-fixed precision to account for degraded tracking conditions, yielding $Q \ll R$: the estimator favors the kinematic model over individual GNSS corrections in the absence of a robust weighting mechanism.

For the independently collected sequence, where the IMU heading is additionally available as a direct measurement, R is augmented to $\text{diag}(0.1, 0.1, \sigma_\psi^2)$, where $\sigma_\psi = 1^\circ \approx 0.017[\text{rad}]$ is set according to the dynamic yaw accuracy specified in the IMU datasheet.

The initial state x_0 is obtained directly from the data at the start of each sequence (initial position and orientation). The arrival-cost covariance $P_{k-N|k}$ is recursively updated through a standard EKF.

Remark 3 (Limitations of manual weight selection): The weights Q , R and D are fixed offline from the physical error budget of the platform, as above. This is a modelling commitment, and it is the main practical limitation of the present formulation: the noise covariances are rarely known exactly, and in the field they are time-varying — R in particular should arguably inflate whenever the receiver drops from an RTK-fixed to a float or standalone solution. The Huber penalty mitigates but does not remove this fragility: it bounds the influence of any individual residual, yet it cannot fully compensate for systematically mis-scaled R , since δ is itself interpreted relative to the residual scale (Section IV-C). Learning-based *MHE* schemes address precisely this

gap by tuning the weighting matrices online from data through differentiable optimisation — for example the neural *MHE* of [31] and its trust-region variant [32]. Those approaches are complementary rather than competing: they adapt the weights but do not provide a bounded-influence or *p*-RGAS certificate, whereas *MHE-H* provides both for a fixed, analytically characterised loss. Combining a learned *Q/R* with a Huber residual penalty — and, in the same spirit, adapting δ from a running robust estimate of the innovation scale — is a natural direction, which we leave for future work.

E. Synthetic Outlier Generation

The synthetic contamination used in Sections V-F and V-G injects *bootstrap-resampled real outlier bursts*. Bursts are mined once from *building00* and *building02*: consecutive-epoch runs whose GNSS-to-ground-truth position error exceeds 3 m, giving two libraries with different character and natural outlier rate — *building00* contributes 21 isolated spikes (median length 1 epoch, 23.4% of its epochs) and *building02* contributes 3 sustained bursts (median length 24 epochs, consistent with prolonged multipath, 4.6% of its epochs). For a target contamination rate ε and severity level s , bursts are drawn (with replacement) from a mixture of the two libraries — interpolated between their natural outlier rates as a function of ε — and placed at random, non-overlapping positions (minimum 3-epoch gap) on a clean sequence until the target fraction of contaminated epochs is reached; each drawn burst's real (e_x, e_y) error vector is inserted verbatim, preserving its temporal shape and *x-y* correlation rather than perturbing each axis independently. All inserted bursts are then rescaled by a single common factor so that their mean magnitude matches a value linearly interpolated between the two libraries' mean burst magnitudes (≈ 5.3 m at $s = 0$, ≈ 19.6 m at $s = 1$). This decouples *how large* outliers are (s) from *how often* they occur (ε).

F. Sensitivity to the Huber Threshold and the Horizon Length

MHE-H was swept over $N \in \{5, 10, 15\}$ and $\delta \in \{0.01, 0.1, 0.25, 0.5, 1.0, 1.345\}$, alongside the quadratic *MHE* ($\delta \rightarrow \infty$) as a reference, over the four clean base sequences (*playground00*, *street00*, *street01*, *street02*). Each cell combines three contamination rates $\varepsilon \in \{0.05, 0.15, 0.25\}$, three severity levels $s \in \{0, 0.5, 1\}$, and five random seeds, for 180 runs per cell, obtained by injecting bootstrap-resampled real outlier bursts into each sequence. The choice of (δ^*, N) is based on the four accuracy metrics defined in Section V-C.

Fig. 3 shows the resulting ATE, MTE, AYE and MYE, averaged per cell; the outlined cell in each panel marks that metric's own minimum. The configuration $\delta^* = 0.1$, $N = 5$, is the best or joint-best configuration in three of the four metrics: it attains the lowest MTE, AYE, and MYE, while ATE is lower at $N = 10$. We therefore retain $N = 5$ as the single horizon shared by *MHE*, *MHE-D* and *MHE-H* in the remainder of the experimental evaluation.

G. Robustness Under Synthetic Contamination

The five estimators were compared under controlled synthetic contamination on the four clean sequences (*playground00*, *street00*, *street01*, *street02*). Outlier bursts were injected independently along two axes: contamination rate $\varepsilon \in \{0.05, 0.15, 0.25\}$ and severity level $s \in \{0, 0.5, 1\}$. For each combination, ten independent random seeds were used, resulting in 360 runs per estimator.

Table II summarizes the estimation accuracy, computational cost, and solver reliability over all runs. Error metrics are reported as mean $\pm$ standard deviation across runs. For all estimators, a run is flagged as a failure if its ATE exceeds 25 m. For *MHE*-based estimators, a run is additionally considered a failure if the solver successfully solves fewer than 99% of the sliding windows or raises an exception. Per-run timing statistics (mean, median, p95, p99) are averaged over the 360 runs; the reported maximum is the worst-case epoch over all runs (maximum of the per-run maxima).

MHE-H attains the lowest mean MTE, AYE, and MYE among the five estimators. Its mean ATE (2.79 m) is close to that of *EKF-Z* (2.63 m), while its dispersion remains relatively high under the considered contamination scenarios. *MHE-D* presents a higher failure rate than *MHE-H*, suggesting that the Huber loss's bounded tail weighting is more numerically stable than the fully redescending loss used by Deniz under this contamination model. The results also highlight the computational cost of the different approaches. The *EKF*-based methods require less than 1 ms per estimation step on average, whereas the *MHE*-based methods require approximately 1.8 ms. The robust *MHE* formulations introduce only a small additional computational cost relative to the quadratic *MHE*.

These results are consistent with the theoretical analysis in Section IV-A, so it achieves competitive worst-case metrics even though its mean ATE retains relatively high variance across runs with differing outlier realizations.

Fig. 4 further illustrates the behavior of the estimators as the contamination severity increases. The results are averaged over contamination rates and random seeds, with the shaded regions representing ± 1 standard deviation. *EKF-Z*, *MHE-D*, and *MHE-H* remain comparatively stable as the severity increases, whereas the non-robust *EKF* and quadratic *MHE* exhibit a substantially larger degradation. This indicates that the benefit of robust loss functions becomes increasingly pronounced as the magnitude of the measurement contamination grows.

H. Validation on Field Sequences with Real GNSS Outliers

The five estimators were then run on the two i2Nav-Robot sequences that contain naturally occurring GNSS outliers: *building00* ($\approx 23\%$ of epochs flagged as outliers) and *building02* ($\approx 5\%$). Table III reports the resulting error metrics for a single realization per sequence and estimator.

On *building00*, which exhibits the highest proportion of naturally occurring GNSS outliers, the robust *MHE* formulations provide a substantial improvement over the non-robust estimators. *MHE-D* and *MHE-H* provide comparable performance.

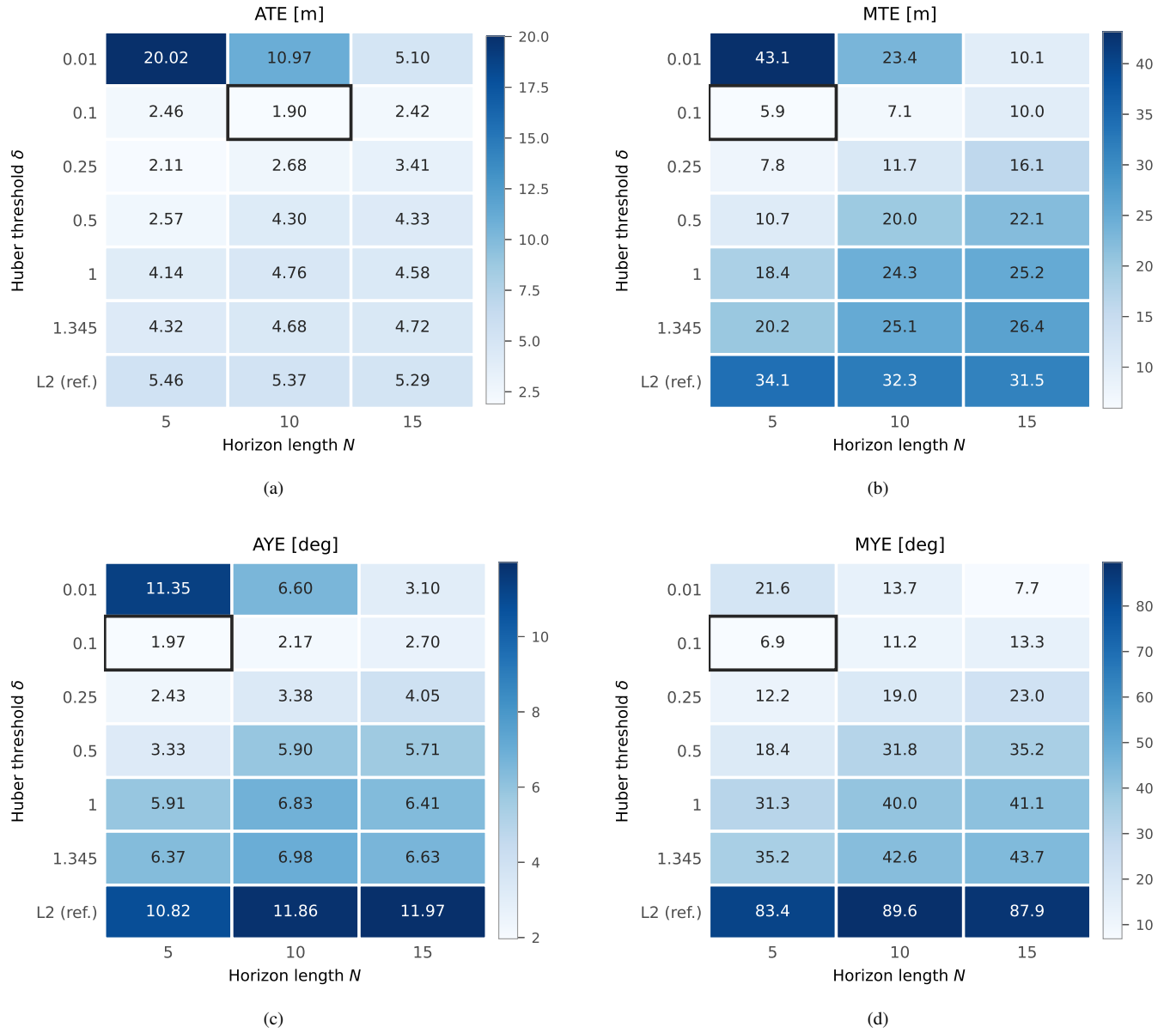


Fig. 3. MHE-H estimation accuracy over the (δ, N) grid, averaged over 180 runs per cell across the four base sequences: (a) ATE, (b) MTE, (c) AYE, (d) MYE.

TABLE II
ERROR METRICS AND COMPUTATIONAL COST UNDER SYNTHETIC CONTAMINATION

Estimator	Position Error [m]		Yaw Error [deg]		Computation time [ms]					Mean iter.	Failure [%]
	ATE	MTE	AYE	MYE	Mean	Median	95th	99th	Max		
EKF	5.39±3.27	29.08±18.68	8.68±5.92	56.52±38.69	0.31	0.31	0.36	0.43	0.61	—	0.00
MHE	5.87±3.56	34.04±22.24	11.24±7.91	84.19±55.55	1.76	1.66	2.36	2.95	4.42	2.92 ± 0.06	0.00
EKF-Z	2.63±2.89	9.35±7.79	2.81±2.48	12.17±8.09	0.33	0.33	0.38	0.46	0.66	—	0.28
MHE-D	6.45±6.82	15.68±16.08	3.55±3.76	9.15±7.70	1.80	1.70	2.42	2.99	4.72	4.09 ± 0.45	2.22
MHE-H	2.79±5.97	7.04±10.48	2.35±4.41	6.94±6.94	1.79	1.68	2.40	2.96	4.69	3.47 ± 0.21	1.67

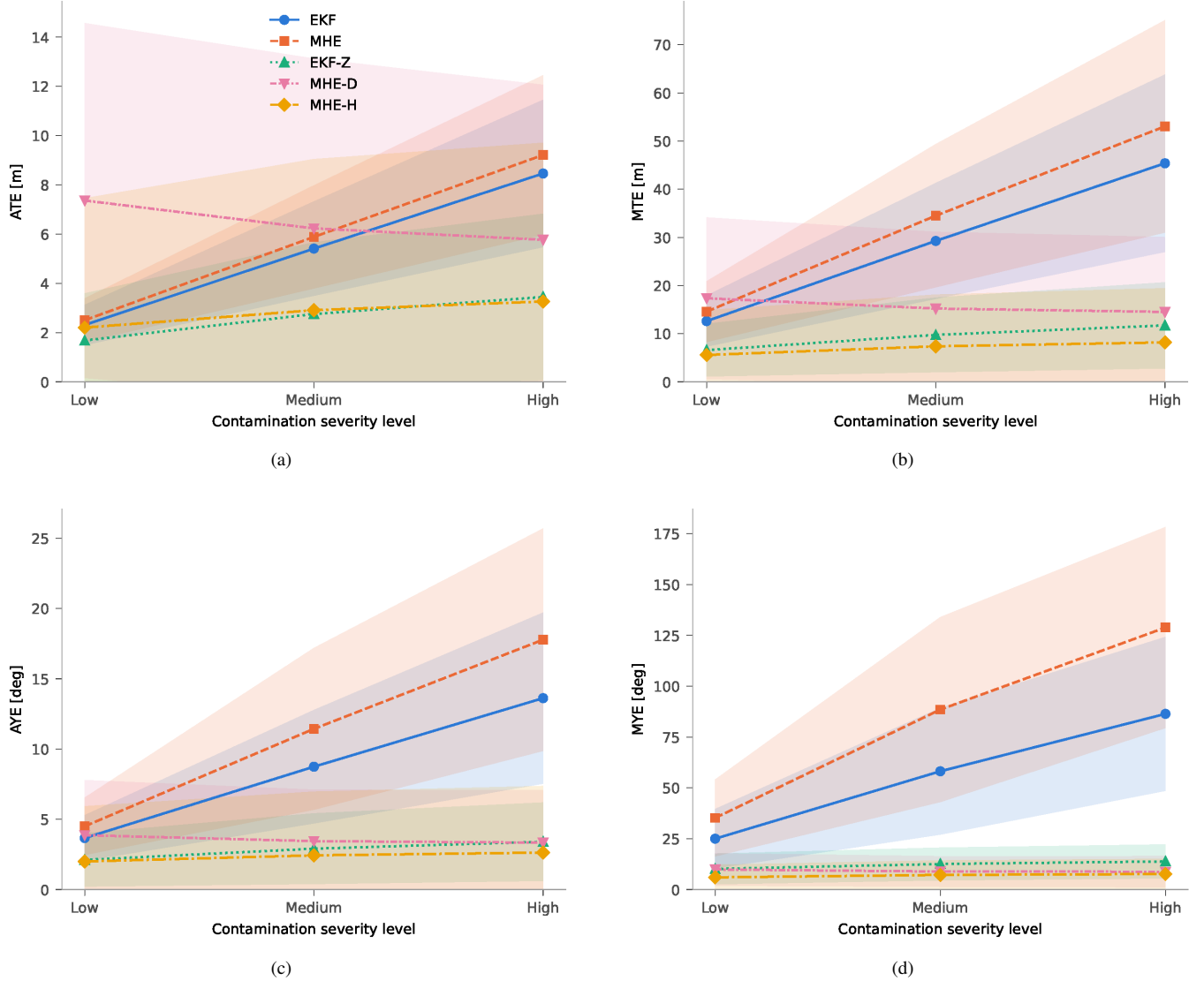


Fig. 4. Estimation error vs. contamination severity level. (a) ATE, (b) MTE, (c) AYE, (d) MYE.

TABLE III
ERROR METRICS ON FIELD SEQUENCES WITH REAL GNSS
OUTLIERS

Seq.	Est.	Position Error [m]		Yaw Error [deg]	
		ATE	MTE	AYE	MYE
building00	EKF	11.10	50.31	13.92	103.06
	MHE	11.74	50.90	16.96	136.30
	EKF-Z	4.92	16.00	3.82	14.51
	MHE-D	2.47	6.00	1.19	3.84
	MHE-H	2.56	6.24	1.99	6.64
building02	EKF	1.27	11.43	2.11	21.38
	MHE	1.27	12.27	2.23	21.82
	EKF-Z	1.49	7.00	1.30	8.29
	MHE-D	3.51	9.05	1.63	6.89
	MHE-H	3.09	8.04	2.17	7.15

In *building02*, where naturally occurring GNSS outliers are less frequent, the differences between the estimators are less pronounced. The standard EKF and MHE achieve the

lowest mean position error, whereas the robust MHE variants yield slightly higher ATE values. Nevertheless, the robust formulations substantially reduce the worst-case position error: MHE-H achieves an MTE of 8.04 m, compared with 11.43 m for the EKF, while MHE-D achieves the lowest MYE of 6.89°. These results indicate that, when outliers are relatively sparse, robustification may provide greater benefits in terms of worst-case errors than in terms of mean accuracy. This trade-off occurs because *building02*'s outliers are sparse and sustained rather than frequent, so quadratic estimators' mean error stays low while MHE-H's down-weighting slightly discounts informative nominal residuals, yet still suppresses the large transient deviations that dominate worst-case metrics.

I. Validation on an Independently Collected Sequence

The seventh sequence was evaluated both in its original, uncontaminated form and under the same synthetic contamination methodology as Section V-G: three contamination rates,

TABLE IV
ERROR METRICS AND COMPUTATIONAL COST ON THE SEVENTH SEQUENCE: ORIGINAL UNCONTAMINATED RUN (SINGLE REALIZATION, NO STD) AND UNDER SYNTHETIC CONTAMINATION (MEAN $\pm$ STD OVER 90 RUNS/ESTIMATOR)

Estimator	Cond.	Position Error [m]		Yaw Error [deg]		Computation time [ms]					Mean iter.
		ATE	MTE	AYE	MYE	Mean	Median	95th	99th	Max	
EKF	Clean	1.15	1.70	0.94	3.49	0.35	0.33	0.46	0.64	1.09	–
	Contam.	4.73 ± 2.62	31.63 ± 21.78	0.94 ± 0.00	3.50 ± 0.01	0.34	0.33	0.42	0.54	3.88	–
MHE	Clean	1.15	1.71	0.93	3.60	1.75	1.63	2.53	3.77	5.58	1.93
	Contam.	5.51 ± 3.14	37.14 ± 25.04	0.93 ± 0.00	3.59 ± 0.01	1.74	1.67	2.10	3.21	8.79	1.94 ± 0.00
EKF-Z	Clean	1.15	1.70	0.94	3.49	0.35	0.34	0.38	0.49	0.94	–
	Contam.	1.54 ± 0.23	5.20 ± 1.33	0.95 ± 0.02	3.57 ± 0.16	0.36	0.35	0.45	0.57	2.89	–
MHE-D	Clean	1.15	1.71	0.93	3.78	1.90	1.73	3.00	4.45	5.90	3.93
	Contam.	1.20 ± 0.09	1.99 ± 0.38	0.93 ± 0.00	3.78 ± 0.00	1.83	1.74	2.28	3.33	9.08	3.93 ± 0.00
MHE-H	Clean	1.15	1.71	0.97	3.74	1.82	1.73	2.28	4.18	6.02	3.83
	Contam.	1.21 ± 0.04	2.39 ± 0.22	1.01 ± 0.04	3.87 ± 0.23	1.80	1.74	2.17	3.30	8.45	3.48 ± 0.16

three severity levels and ten random seeds, for $3 \times 3 \times 10 = 90$ runs per estimator. Table IV reports the resulting error metrics and computational cost for both conditions.

On the original, uncontaminated run, all five estimators produce essentially equivalent performance. This shared floor matches the receiver's own data sheet rather than reflecting any estimator limitation: the u-blox ZED-F9P reports a standalone horizontal accuracy on the order of 1.5 m, against centimeter-level accuracy for the RTK receiver used as ground truth. The observed ATE of ≈ 1.15 m is thus consistent with the nominal precision of the standalone GNSS measurement itself.

Under synthetic contamination, the same pattern as Section V-G reproduces on this independently collected platform: EKF-Z, MHE-D and MHE-H stay within a narrow band of their clean-run baseline, while the non-robust EKF and MHE degrade sharply (roughly $4\text{--}5\times$ their clean baseline). Because the injected bursts corrupt only the GNSS position channel, AYE and MYE are barely affected for any estimator. Fig. 5 illustrates this on a representative contaminated realization: EKF and MHE briefly diverge at each injected burst while the three robust estimators track the ground truth closely throughout.

VI. CONCLUSION

This paper introduced *MHE-H*, a robust Moving Horizon Estimator that replaces the quadratic measurement penalty with the Huber loss to mitigate outlier-corrupted GNSS measurements in vehicle localization. Beyond the estimator itself, the paper's main contribution is a theoretical framework that connects statistical robustness with dynamic stability. The influence-function and gross-error sensitivity analysis quantify how a contaminated measurement propagates through the estimated trajectory. In parallel, the practical robust global asymptotic stability (*p-RGAS*) result shows that the resulting estimation error remains bounded under generic *i-IOSS* detectability, without relying on restrictive linear or Gaussian assumptions.

Recent generalized *MHE* work based on β -divergence losses proves standard least-squares *MHE* has infinite gross-error sensitivity [8], while their divergence-weighted estimator remains bounded. *MHE-H* parallels this conclusion via the Huber threshold δ but additionally couples the bound explicitly

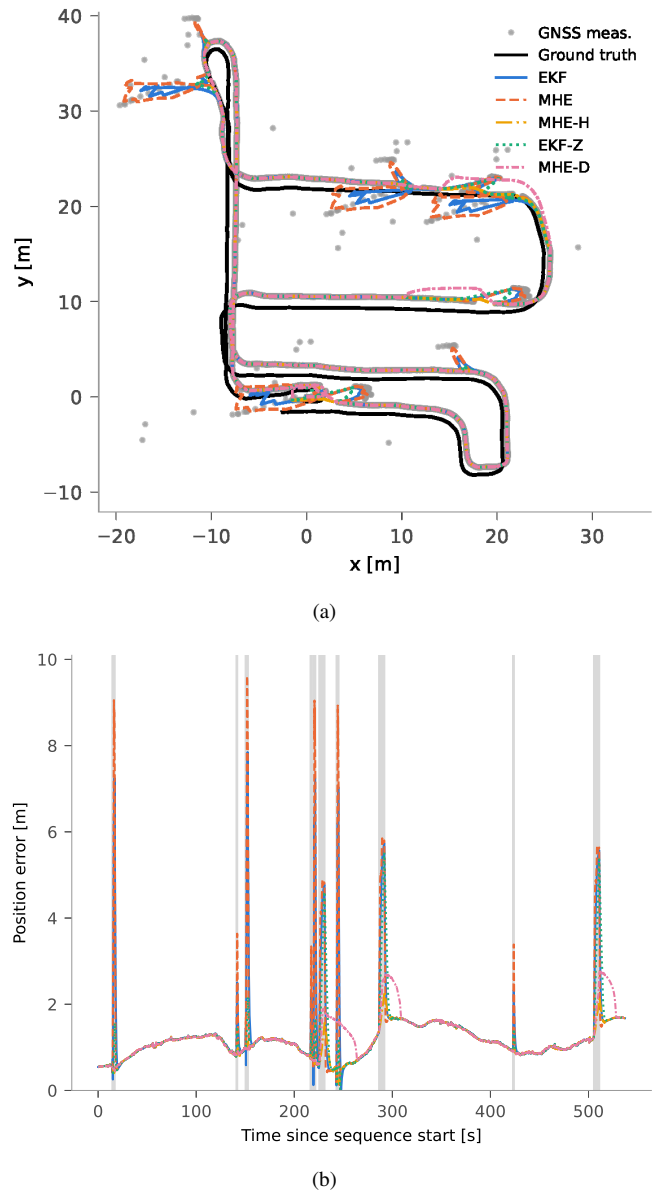


Fig. 5. A representative contaminated realization of the seventh sequence. (a) Estimated trajectories and ground truth; (b) position error over time. Shaded regions indicate the injected outlier bursts.

to *i*-IOSS detectability constants, yielding a stability-linked robustness certificate absent from the divergence-based analysis.

Across seven test sequences, two field-recorded runs with naturally occurring GNSS outliers, and an independently collected trajectory under synthetic contamination — *MHE-H* consistently matched or improved upon the accuracy of *EKF*, standard *MHE*, a χ^2 -gated robust *EKF*, and a redescending-loss *MHE* variant, particularly in worst-case metrics such as maximum position error. The Huber threshold δ was shown to govern an explicit efficiency-robustness trade-off, with $\delta \approx 1.1 \hat{\sigma}_v$ to $1.6 \hat{\sigma}_v$ favoring near-least-squares efficiency under near-Gaussian noise and δ below $\hat{\sigma}_v$ favoring aggressive outlier rejection under heavy-tailed contamination, giving practitioners a principled tuning rule rather than an ad hoc choice.

The moderate computational overhead relative to standard *MHE* (roughly 1.8 ms per estimation step versus sub-millisecond *EKF* updates) indicates *MHE-H* remains viable for real-time localization at typical GNSS/IMU sampling rates.

Future work will focus on three directions: embedding *MHE-H* in an online, resource-constrained implementation suitable for automotive-grade hardware; extending the influence-function framework to adaptively tune δ from a running estimate of the innovation scale rather than a fixed offline value; and investigating multirate sensor-fusion strategies that combine asynchronous GNSS, IMU, and odometry streams within the same robust-stability framework established here.

APPENDIX A PROOF OF LEMMA 1

The upper bound follows directly from definition of Huber loss: for small residuals ($|\hat{v}_{i,j|k}| \leq \delta_i$) equality holds by definition (7). In the large residuals regime ($|\hat{v}_{i,j|k}| > \delta_i$), the inequality $2\delta_i |\hat{v}_{i,j|k}| \leq \hat{v}_{i,j|k}^2 + \delta_i^2$, hence $L_\delta(\hat{v}_{i,j|k}, \delta_i) \leq 1/2 \hat{v}_{i,j|k}^2$. To analyse the lower bound, we define

$$g(\hat{v}_{i,j|k}, \delta_i) := L_\delta(\hat{v}_{i,j|k}, \delta_i) - \frac{1}{2} \frac{\hat{v}_{i,j|k}^2}{1 + \frac{1}{\delta_i} |\hat{v}_{i,j|k}|}. \quad (55)$$

For small residuals $g(\cdot)$ verifies

$$g(\hat{v}_{i,j|k}, \delta_i) = \frac{1}{2} \hat{v}_{i,j|k}^2 \frac{\frac{1}{\delta_i} |\hat{v}_{i,j|k}|}{1 + \frac{1}{\delta_i} |\hat{v}_{i,j|k}|} \geq 0. \quad (56)$$

In the large residuals regime, $t = |\hat{v}_{i,j|k}|/\delta_i > 1$ and

$$\frac{g}{\delta_i^2} = t - \frac{1}{2} - \frac{1}{2} \frac{t^2}{1+t} = \frac{t^2 + t - 1}{2(1+t)} > 0. \quad (57)$$

APPENDIX B PROOF OF THEOREM 1

The core mechanism in any *MHE* stability proof is that the optimizer cannot produce a cost higher than the cost evaluated at the true disturbance sequence. We formalize this observation into a bound on the state and disturbance estimation errors.

Lemma 2: Let $(\hat{x}_{k-N|k}^*, \hat{w}_{N|k}^*)$ be an optimizer of the *MHE* problem (9). Under Assumptions 3 and 4, the following inequality holds

$$\begin{aligned} & \underline{p} \|\hat{x}_{k-N|k}^* - x_{k-N}\|^2 + \underline{q} \|\hat{w} - w^*\|^2 \\ & + \underline{d} \sum_{j=k-N}^k \sum_{i=1}^{n_y} \sigma_L (|\hat{v}_{i,j|k}|) \leq \mathcal{J}_k \end{aligned} \quad (58)$$

where

$$\begin{aligned} \mathcal{J}_k &= \bar{p} \|x_{k-N} - \bar{x}_{k-N}\|^2 + \bar{C}_w \|w^*\|^2 \\ & + \bar{C}_v \|v^*\|^2 + \bar{C}_o \bar{\delta}, \\ \bar{C}_w &= N \max_{i \in [1, n_x]} q_i, \bar{C}_v = (N+1) n_y \bar{d}, \bar{C}_o = \bar{C}_v \bar{\delta}. \end{aligned} \quad (59)$$

Proof. Since $(\hat{x}_{k-N|k}^*, \hat{w}_{k-N|k}^*)$ is an optimizer of the *MHE* problem (9), optimality implies

$$J(\mathbf{y}_{N|k}, \hat{w}_{N|k}^*, \hat{x}_{k-N|k}^*) \leq J(\mathbf{y}_{N|k}, \mathbf{w}_{N|k}, x_{k-N|k}), \quad (60)$$

where the right-hand side is the cost evaluated at the true state and disturbance trajectory $(x_{k-N|k}, \mathbf{w}_{N|k})$ that generates the measurements, which is feasible for problem (9) since it satisfies the system dynamics 1 exactly.

We first lower-bound the optimal cost by isolating the contributions of the prior mismatch, disturbance-estimation error, and residual penalty in terms of quantities that can later be related to the state-estimation error. Using Assumption 4 and Lemma 1 we obtain

$$\begin{aligned} J &= P_{k-N|k}^{-1} \|\hat{x}_{k-N|k} - \bar{x}_{k-N}\|^2 + \sum_{j=k-N+1}^k q_j \hat{w}_j^2 \\ & + \sum_{i=1}^{n_y} d_i L_\delta(\hat{v}_{i,j|k}, \delta_i), \\ & \geq \underline{p} \|\hat{x}_{k-N|k} - x_{k-N}\|^2 - \underline{p} \|x_{k-N} - \bar{x}_{k-N}\|^2 \\ & + \underline{q} \|\hat{w}_{N|k} - \mathbf{w}_{N|k}\|^2 - \underline{q} \|\mathbf{w}_{N|k}\|^2 \\ & + \underline{d} \sum_{j=k-N+1}^k \sum_{i=1}^{n_y} \sigma_L (|\hat{v}_{i,j|k}|). \end{aligned} \quad (61)$$

To compare the disturbance term with the true disturbance, write $\hat{w}_{k-j} = \frac{1}{2}(\hat{w}_{k-j} - w_{k-j})^2 - w_{k-j}$, and use the elementary inequality to obtain $\hat{w}_{k-j}^2 \geq \frac{1}{2}(\hat{w}_{k-j} - w_{k-j})^2 - w_{k-j}^2$. Applying this bound component wise yields the stated disturbance-error contribution.

Next, we upper-bound the right-hand side of (60) by evaluating the Huber term along the true residuals $\hat{v}_{j|k} = v_j + o_j$. For any scalar arguments, the Huber loss satisfies $L_\delta(s, \delta_i) \leq \delta_i |s|$ when $|s| > \delta_i$, and $L_\delta(s, \delta_i) \leq 1/2 s^2$ otherwise. Hence it is globally upper-bounded by the piecewise-linear envelope $L_\delta(s, \delta_i) \leq \delta_i |s|$. Using the triangle inequality $\|v_j + o_j\| \leq \|v_j\| + \|o_j\|$ and the fact that $L_\delta(s, \delta_i)$ is convex with slope bounded by δ_i , we have

$$L_\delta(v_j + o_j, \delta_i) \leq L_\delta(v_j, \delta_i) + \delta_i \|o_j\| \quad (62)$$

Summing over the horizon and output channels gives

$$\begin{aligned} \sum_{j=k-N+1}^k \sum_{i=1}^{n_y} d_i L_\delta(\hat{v}_{i,j|k}, \delta_i) &\leq \sum_{j=k-N+1}^k \sum_{i=1}^{n_y} d_i L_\delta(v_{i,j|k}, \delta_i) \\ &+ \sum_{j=k-N+1}^k \sum_{i=1}^{n_y} d_i \delta_i \|o_{i,j}\| \end{aligned}$$

Combining this with Assumption 4 for the arrival cost and the upper quadratic bound in Lemma 1, we obtain

$$\begin{aligned} J^* &= P_{k-N|k}^{-1} \|x_{k-N|k} - \bar{x}_{k-N}\|^2 + \sum_{j=k-N+1}^k q_j w_{j|k}^{*2} \\ &\quad + \sum_{i=1}^{n_y} d_i L_\delta(v_{i,j|k}^*, \delta_i^*) \\ &\leq \bar{p} \|x_{k-N|k} - \bar{x}_{k-N}\|^2 + \bar{C}_w \|w\|^2 + \bar{C}_v \|v\|^2 \\ &\quad + \bar{C}_o \bar{\delta}, \end{aligned} \quad (63)$$

where $\bar{\delta} = \max_{i \in [1, n_y]} \delta_i$ and $\bar{d} = \max_{i \in [1, n_y]} d_i$. Here $\bar{C}_w = N \max_{i \in [1, n_x]} q_i$ absorbs the horizon-wise accumulation of the process-disturbance weights q_i , $\bar{C}_v = (N+1) n_y \bar{d}$ collects the upper bound on the weighted nominal measurement-loss terms over all N stages and n_y output channels, and $\bar{C}_o = (N+1) n_y \bar{d} \bar{\delta}$ captures the additional linear growth induced by the outlier magnitudes through the Huber envelope $L_\delta(s, \delta_i) \leq \delta_i |s|$.

Finally, substituting the lower bound (61) and the upper bound (63) into the optimality relation (60), and rearranging terms yields (58)-(59). $\square$

Lemma 2 bounds the disturbance and residual estimation errors $\hat{w} - w$ and $\hat{v} - v$. We now propagate these bounds through the system dynamics to obtain a bound on the state-estimation error $\hat{x}_k - x_k$.

The next lemma propagates the optimization-based bound of Lemma 2 through the exponential i -IOSS dynamics to convert the cost residual $\mathcal{J}_k$ into a direct bound on the current state-estimation error $\|\hat{x}_{k|k} - x_k\|^2$.

Lemma 3: Under Assumptions 3 and 4, $\hat{x}_k - x_k$ satisfies

$$\|\hat{x}_k - x_k\|^2 \leq \kappa \mathcal{J}_k, \quad (64)$$

where

$$\kappa := 3 \left[\frac{c_0^2}{\underline{p}} \rho^{2N} + \frac{c_w^2}{\underline{q}} + 2 \frac{c_v^2 \bar{\delta}}{\underline{d}} \left(1 + \frac{\bar{v}}{\underline{d}} \right) \right]. \quad (65)$$

Proof. Under Assumption 3, the system is exponentially i -IOSS, so applying (37) over the horizon $[k-N, k]$ to the true and estimated trajectories gives

$$\|\hat{x}_k - x_k\| \leq c_0 \rho^N \|\hat{x}_{k-N|k} - x_{k-N}\| + c_w \|\hat{w}_{N|k} - w_{N|k}\|_\infty + c_v \|\hat{v}_{N|k}\|_\infty.$$

Applying the inequality $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$, which yields

$$\begin{aligned} \|\hat{x}_k - x_k\|^2 &\leq 3c_0^2 \rho^{2N} \|\hat{x}_{k-N|k} - x_{k-N}\|^2 \\ &\quad + 3c_w^2 \|\hat{w}_{N|k} - w_{N|k}\|_\infty^2 + 3c_v^2 \|\hat{v}_{N|k}\|_\infty^2. \end{aligned}$$

Since the supremum norm is bounded by the Euclidean norm for finite-dimensional vectors, we have

$$\begin{aligned} \|\hat{x}_{k|k} - x_k\|^2 &\leq 3c_0^2 \rho^{2N} \|\hat{x}_{k-N|k} - x_{k-N}\|^2 \\ &\quad + 3c_w^2 \|\hat{w}_{N|k} - w_{N|k}^*\|^2 \\ &\quad + 3c_v^2 \sum_{j=k-N+1}^k \sum_{i=1}^{n_y} \hat{v}_{i,j|k}^2. \end{aligned} \quad (66)$$

Bounding the first two terms directly via Lemma 2 we obtain

$$\|\hat{x}_{k-N|k} - x_{k-N}\|^2 \leq \frac{1}{\underline{p}} \mathcal{J}_k, \quad (67)$$

$$\|\hat{w}_{N|k} - w_{N|k}\|^2 \leq \frac{1}{\underline{q}} \mathcal{J}_k, \quad (68)$$

To convert the residual term into a quadratic bound, use the Huber lower-envelope relation introduced in Lemma 1

$$s^2 \leq 2\bar{\delta} \left(1 + \frac{\bar{v}}{\underline{d}} \right) \sigma_L(s) := C_\delta \sigma_L(s) \quad \forall 0 \leq s \leq \bar{v}, \quad (69)$$

where $\bar{v}$ is a uniform bound on the residual magnitude over the estimation window and $C_\delta = 2\bar{\delta}(1 + \bar{v}/\underline{d})$. We now convert the Huber residual term into a quadratic bound, so that the residual contribution over the horizon can be combined with the Euclidean error terms in the i -IOSS estimate

$$\sum_{j=k-N+1}^k \sum_{i=1}^{n_y} \hat{v}_{i,j|k}^2 \leq C_\delta \sum_{j=k-N+1}^k \sum_{i=1}^{n_y} \sigma_L(|\hat{v}_{i,j|k}|) \leq \frac{C_\delta}{\underline{d}} \mathcal{J}_k, \quad (70)$$

Substituting these three bounds into (66), gives

$$\|\hat{x}_k - x_k\|^2 \leq 3 \frac{c_0^2}{\underline{p}} \rho^{2N} \mathcal{J}_k + 3 \frac{c_w^2}{\underline{q}} \mathcal{J}_k + 3 \frac{c_v^2}{\underline{d}} C_\delta \mathcal{J}_k, \quad (71)$$

so that $\|\hat{x}_k - x_k\|^2 \leq \kappa \mathcal{J}_k$ holds with

$$\kappa = 3 \left[\frac{c_0^2}{\underline{p}} \rho^{2N} + \frac{c_w^2}{\underline{q}} + \frac{c_v^2}{\underline{d}} C_\delta \right]. \quad (72)$$

Substituting the definition $C_\delta = 2\bar{\delta}(1 + \bar{v}/\underline{d})$ then gives the expanded expression used in Theorem 1. $\square$

The term $\mathcal{J}_k$ in (64) contains $P_{k-N|k}^{-1} \|x_{k-N} - \bar{x}_{k-N}\|^2$, which involves the prior $\bar{x}_{k-N}$ produced by the previous MHE step. The following lemma shows that the prior mismatch term embedded in $\mathcal{J}_k$ decays geometrically under the sliding-window update. This allows Lemma 3 to be expressed in terms of the initial uncertainty and bounded disturbances, which is the key step needed to prove Theorem 1.

Lemma 4: Define the arrival cost error at step k as $e_k := \|x_{k-N} - \bar{x}_{k-N}\|$, the sliding-window update $\bar{x}_{k-N} = \hat{x}_{k-N+1|k}$ yields

$$\begin{aligned} e_k^2 &\leq 3c_0^2 \rho^{2kN} e_0^2 + \frac{3c_w^2}{(1-\rho^N)^2} \|w\|^2 \\ &\quad + \frac{3c_v^2}{(1-\rho^N)^2} \|v\|^2 + \frac{3c_o^2}{(1-\rho^N)^2} \|o\|^2. \end{aligned} \quad (73)$$

Proof. Applying the exponential i -IOSS property (37) to the pair $(x_{k-N+1}, \hat{x}_{k-N+1|k})$, given the sliding-window update $\bar{x}_{k-N} = \hat{x}_{k-N+1|k}$, both propagated from time $k-N$ over N steps with respective disturbances $(w_{N|k}, v_{N|k})$ and $(\hat{w}_{N|k}, \hat{v}_{N|k})$

$$\begin{aligned} e_{k+1} &\leq c_0 \rho^N \|x_{k-N} - \bar{x}_{k-N}\| + c_w \|w_{N|k}\|_\infty \\ &\quad + c_v \|v_{N|k}\|_\infty + c_o \|o_{N|k}\|_\infty. \end{aligned} \quad (74)$$

Next, we bound w, v and o by their supremum over the full horizon. Applying (74) recursively from step 0 to step k

$$\begin{aligned} e_k &\leq \rho^{Nk} e_0 + \sum_{j=0}^{k-1} c_w \|w\|_\infty \rho^{Nj} + \|v\|_\infty c_v \rho^{Nj} \\ &\quad + \|o\|_\infty c_o \rho^{Nj} \\ &\leq c_0 \rho^{kN} e_0 + \frac{c_w}{1-\rho^N} \|w\|_\infty + \frac{c_v}{1-\rho^N} \|v\|_\infty \\ &\quad + \frac{c_o}{1-\rho^N} \|o\|_\infty, \end{aligned} \quad (75)$$

and squaring gives (73). $\square$

Remark 4: Substituting (73) into $\mathcal{J}_k$, the coefficient of e_0^2 in the resulting state error bound (64) is $3\kappa P_{k-N|k}^{-1} c_0^2 \rho^{2kN}$. The estimate $\hat{x}_{k|k}$ overcomes the initial state uncertainty if

$$3\kappa P_{k-N|k}^{-1} c_0^2 \rho^{2N} < P_{k-N|k}^{-1}, \quad (76)$$

which is verified if condition (46) is satisfied.

Finally, we combine the instantaneous state-error bound of Lemma 3 with the geometric decay of the arrival-cost mismatch from Lemma 4 to obtain a single estimate suitable for proving Theorem 1. From Lemma 3, equation (64) and the definition (59) of $\mathcal{J}_k$

$$\|\hat{x}_k - x_k\|^2 \leq \kappa P_{k-N|k}^{-1} e_k^2 + \kappa \bar{C}_w \|w\|^2 + \kappa \bar{C}_v \|v\|^2 + \kappa \bar{C}_o \bar{\delta}. \quad (77)$$

Substituting (73) we have

$$\begin{aligned} \|\hat{x}_k - x_k\|^2 &\leq 3\kappa \bar{p} c_0^2 \rho^{2kN} e_0^2 + \left(\kappa \bar{C}_w + \frac{3\kappa \bar{p} c_w^2}{(1-\rho^N)^2} \right) \|w\|^2 \\ &\quad + \left(\kappa \bar{C}_v + \frac{3\kappa \bar{p} c_v^2}{(1-\rho^N)^2} \right) \|v\|^2 + \kappa \bar{C}_o \bar{\delta} + \frac{3\kappa \bar{p} c_o^2}{(1-\rho^N)^2} \|o\|^2. \end{aligned}$$

The initial error $e_0 = x_0 - \hat{x}_0$ can be rewritten in term of the estimation error as follows

$$e_0 \leq \|x_0 - \hat{x}_0\| + \|\hat{x}_0 - \bar{x}_0\| \leq \|x_0 - \hat{x}_0\| + e_{\max}. \quad (78)$$

Substituting into previous equation and using the Cauchy–Schwartz inequality, we have

$$\begin{aligned} \|\hat{x}_k - x_k\|^2 &\leq 6\kappa \bar{p} c_0^2 \rho^{2kN} \|x_0 - \hat{x}_0\|^2 \\ &\quad + 6\kappa \bar{p} c_0^2 \rho^{2kN} e_{\max}^2 + C_{\gamma_w}^2 \|w\|^2 \\ &\quad + C_{\gamma_v}^2 \|v\|^2 + \kappa \bar{C}_o \bar{\delta} + \frac{3\kappa \bar{p} c_o^2}{(1-\rho^N)^2} \|o\|^2, \end{aligned} \quad (79)$$

where

$$C_{\gamma_w}^2 := \kappa \bar{C}_w + 3 \frac{\kappa \bar{p} c_w^2}{(1-\rho^N)^2}, \quad (80)$$

$$C_{\gamma_v}^2 := \kappa \bar{C}_v + 3 \frac{\kappa \bar{p} c_v^2}{(1-\rho^N)^2}, \quad (81)$$

$$C_{\gamma_o}^2 := \kappa \bar{C}_o + 3 \frac{\kappa \bar{p} c_o^2}{(1-\rho^N)^2}, \quad (82)$$

$$c := \sqrt{\kappa \bar{C}_o \bar{\delta}}. \quad (83)$$

Taking square roots of (79) and using the triangular inequality $\sqrt{a+b+c+d} \leq \sqrt{a} + \sqrt{b} + \sqrt{c} + \sqrt{d}$ we obtain

$$\begin{aligned} \|\hat{x}_k - x_k\| &\leq \sqrt{6\kappa \bar{p} c_0 \rho^{kN}} \|x_0 - \hat{x}_0\| \\ &\quad + \sqrt{6\kappa \bar{p} c_0 \rho^{kN}} e_{\max} \\ &\quad + C_{\gamma_w} \|w\| + C_{\gamma_v} \|v\| + C_{\gamma_o} \|o\| + c. \end{aligned} \quad (84)$$

Defining

$$\beta_0(s, k) := \sqrt{6\kappa \bar{p} c_0 \rho^{kN}} (s + e_{\max}), \quad (85)$$

$$\gamma_w(s) := C_{\gamma_w} s, \quad (86)$$

$$\gamma_v(s) := C_{\gamma_v} s \quad (87)$$

$$\gamma_o(s) := C_{\gamma_o} s \quad (88)$$

then, (84) gives the *RGAS* bound (47)

$$\begin{aligned} \|\hat{x}_k - x_k\| &\leq \beta(\|\hat{x}_0 - x_0\|, k) + \gamma_w(\sup_{j \leq k} \|w_j\|) \\ &\quad + \gamma_v(\sup_{j \leq k} \|v_j\|) + \gamma_o(\sup_{j \leq k} \|o_j\|) + c \end{aligned} \quad (89)$$

where

- $\beta(\cdot, k) \in \mathcal{K}$ for each fixed $k \geq 0$ because it is continuous and strictly increasing. $\beta_0(0, k) = \sqrt{6\kappa \bar{p} c_0 \rho^{kN}} e_{\max} \geq 0$, the standard $\mathcal{KL}$ form shift by $e_{\max}$. $\beta_0(s, k) = \sqrt{6\kappa \bar{p} c_0 \rho^{kN}} s$ and the constant absorbed into the initial condition;
- $\beta(s, \cdot) \in \mathcal{L}$ since $\rho^{kN} \rightarrow 0$ as $k \rightarrow \infty$, using $\rho \in (0, 1)$ and $N \geq 1$; hence $\beta_0 \in \mathcal{KL}$;

- $\gamma_w, \gamma_v \in \mathcal{K}$.

Condition $N \geq N_{\min}$ ensures $3\kappa c_0^2 \rho^{2N} < 1$. This implies that $\sqrt{6\kappa \bar{p} c_0} \rho^{kN}$ decays strictly to zero, which is precisely the $\mathcal{L}$ requirement for $\beta \in \mathcal{KL}$. Without this condition, the ρ^{kN} factor might not be small enough at $k = 1$ to offset the amplification by $\sqrt{6\kappa \bar{p} c_0}$, and the recursion could fail to contract.

Defining the tail supremum $\|w\|_{[k_0, \infty]} := \sup_{k \geq k_0} \|w_k\|$. If $w_k \rightarrow 0$ and $v_k \rightarrow 0$ as $k \rightarrow \infty$, then for every $\varepsilon > 0$ there exists k_0 such that $\|w\|_{[k_0, \infty]} \leq \varepsilon$ and $\|v\|_{[k_0, \infty]} \leq \varepsilon$. Applying the bound of Theorem 1 over the window $[k_0, k]$ instead of $[0, k]$, and using $\beta(\cdot, k - k_0) \rightarrow 0$ as $k \rightarrow \infty$, gives

$$\begin{aligned} \limsup_{k \rightarrow \infty} \|\hat{x}_{k|k} - x_k\| &\leq \gamma_w(\varepsilon) + \gamma_v(\varepsilon) \\ &\quad + C_{\gamma_o} \|o\|_{[k_0, \infty]} + c. \end{aligned} \quad (90)$$

Since ε is arbitrary, letting $\varepsilon \rightarrow 0$ shows

$$\limsup_{k \rightarrow \infty} \|\hat{x}_{k|k} - x_k\| \leq C_{\gamma_o} \limsup_{k \rightarrow \infty} \|o\|_{[k, \infty]} + c. \quad (91)$$

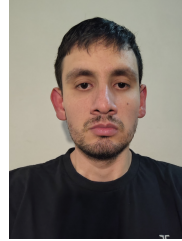
If outliers vanish asymptotically as well ($o_k \rightarrow 0$), the estimation error remains bounded by the practical offset $c = \sqrt{\kappa \bar{C}_o \bar{\delta}}$, which vanishes as $\bar{\delta} \rightarrow 0$ and grows with the Huber threshold; if outliers persist with bounded amplitude, the error is additionally inflated by $C_{\gamma_o} \|o\|$. This is precisely the sense in which the estimator is *practically* *RGAS*: exact convergence is traded for a bounded, $\bar{\delta}$ -governed offset.

REFERENCES

- [1] M. Khodarahmi and V. Maihami, “A review on kalman filter models,” *Archives of Computational Methods in Engineering*, vol. 30, no. 1, pp. 727–747, 2023, doi: 10.1007/s11831-022-09815-7.
- [2] M. Speekenbrink, “A tutorial on particle filters,” *Journal of Mathematical Psychology*, vol. 73, pp. 140–152, 2016, doi: 10.1016/j.jmp.2016.05.006.
- [3] J. B. Rawlings and D. A. Allan, “Moving horizon estimation,” in *Encyclopedia of Systems and Control*. Springer, 2021, pp. 1352–1358, doi: 10.1007/978-3-030-44184-5_4.
- [4] D. Simon, *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, 2006.
- [5] A. Alessandri and M. Awawdeh, “Moving-horizon estimation for discrete-time linear systems with measurements subject to outliers,” in *53rd IEEE conference on decision and control*. IEEE, 2014, pp. 2591–2596, doi: 10.1109/CDC.2014.7039785.
- [6] A. Alessandri and M. Awawdeh, “Moving-horizon estimation with guaranteed robustness for discrete-time linear systems and measurements subject to outliers,” *Automatica*, vol. 67, pp. 85–93, 2016, doi: 10.1016/j.automatica.2016.01.015.
- [7] M. J. Awawdeh, T. Faisal, A. Bashir, A. I. N. Alshbatat, and R. T. Momani, “Moving-horizon estimation approach for nonlinear systems with measurement contaminated by outliers,” *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 22, no. 1, pp. 219–231, 2024, doi: 10.12928/telkomnika.v22i1.25230.
- [8] W. Cao, C. Liu, Z. Lan, S. E. Li, W. Pan, and A. Alessandri, “Robust bayesian inference for moving horizon estimation,” *Automatica*, vol. 173, p. 112108, 2025, doi: 10.1016/j.automatica.2024.112108.
- [9] N. Haverbeke, “Efficient numerical methods for moving horizon estimation,” *Diss., Katholieke Universiteit Leuven, Heverlee, Belgium*, 2011.
- [10] K. Geebelen, A. Wagner, S. Gros, J. Swevers, and M. Diehl, “Moving horizon estimation with a huber penalty function for robust pose estimation of tethered airplanes,” in *2013 American Control Conference*. IEEE, 2013, pp. 6169–6174, doi: 10.1109/ACC.2013.6580805.
- [11] T. Sandy, L. Stadelmann, S. Kerscher, and J. Buchli, “Confusion: Sensor fusion for complex robotic systems using nonlinear optimization,” *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1093–1100, 2019, doi: 10.1109/LRA.2019.2894168.

- [12] J. Dichgans, J. Kallwies, and H.-J. Wuensche, "Robust vehicle tracking with monocular vision using convolutional neuronal networks," in *2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE, 2020, pp. 297–302, doi: 10.1109/MFI49285.2020.9235213.
- [13] Y. Guo, P. Ji, Y. Zheng, X. Sun, and F. Ma, "Rmhe-enhanced mpc for eye-in-hand visual servoing under measurement outliers and dynamic disturbances," in *Journal of Physics: Conference Series*, vol. 3176, no. 1. IOP Publishing, 2026, p. 012018, doi: 10.1088/1742-6596/3176/1/012018.
- [14] N. Deniz, M. Murillo, G. Sanchez, and L. Giovanini, "Robust stability of moving horizon estimation for non-linear systems with bounded disturbances using adaptive arrival cost," *IET Control Theory & Applications*, vol. 14, no. 18, pp. 2879–2888, 2020, doi: 10.1049/iet-cta.2019.0523.
- [15] J. D. Schiller, S. Muntwiler, J. Köhler, M. N. Zeilinger, and M. A. Müller, "A lyapunov function for robust stability of moving horizon estimation," *IEEE Transactions on Automatic Control*, vol. 68, no. 12, pp. 7466–7481, 2023, doi: 10.1109/TAC.2023.3280344.
- [16] I. Krauss, V. G. Lopez, and M. A. Müller, "Sample-based moving horizon estimation," in *2026 European Control Conference (ECC)*. IEEE, 2026, pp. 870–875, doi: 10.48550/arXiv.2510.24191.
- [17] P. J. Huber, "Robust estimation of a location parameter," in *Breakthroughs in statistics: Methodology and distribution*. Springer, 1992, pp. 492–518, doi: 10.1007/978-1-4612-4380-9_35.
- [18] Q. Wang, Y. Ma, K. Zhao, and Y. Tian, "A comprehensive survey of loss functions in machine learning," *Annals of Data Science*, pp. 1–26, 2020, doi: 10.1007/s40745-020-00253-5.
- [19] W. Hu, "Generic stability implication from full information estimation to moving-horizon estimation," *IEEE Transactions on Automatic Control*, vol. 69, no. 2, pp. 1164–1170, 2023, doi: 10.1109/TAC.2023.3277315.
- [20] F. R. Hampel, "The influence curve and its role in robust estimation," *Journal of the American Statistical Association*, vol. 69, no. 346, pp. 383–393, 1974, doi: 10.1080/01621459.1974.10482962.
- [21] C. V. Rao, J. B. Rawlings, and D. Q. Mayne, "Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations," *IEEE Transactions on Automatic Control*, vol. 48, no. 2, pp. 246–258, 2003, doi: 10.1109/TAC.2002.808470.
- [22] C. V. Rao, J. B. Rawlings, and J. H. Lee, "Constrained linear state estimation—a moving horizon approach," *Automatica*, vol. 37, no. 10, pp. 1619–1628, 2001, doi: 10.1016/S0005-1098(01)00115-7.
- [23] D. A. Allan and J. B. Rawlings, "Robust stability of full information estimation," *SIAM Journal on Control and Optimization*, vol. 59, no. 5, pp. 3472–3497, 2021, doi: 10.1137/20M1329135.
- [24] H. Tang, T. Zhang, L. Wang, X. Ding, M. Yuan, Z. Xiang, J. Chen, Y. Bian, S. Liu, Y. Wang *et al.*, "i2nav-robot: A large-scale indoor-outdoor robot dataset for multi-sensor fusion navigation and mapping," *arXiv preprint arXiv:2508.11485*, 2025, doi: 10.48550/arXiv.2508.11485.
- [25] X. Zhou, M. Zhang, J. Hu, L. Li, and X. Guan, "A robust in-motion coarse alignment method for low-accuracy sins and gps integrated system," *IEEE/ASME Transactions on Mechatronics*, vol. 30, no. 1, pp. 611–622, 2024, doi: 10.1109/TMECH.2024.3397244.
- [26] N. N. Deniz, G. M. Sanchez, F. A. A. Cheein, and L. L. Giovanini, "Robust moving horizon estimation for autonomous agricultural vehicles with gnss outliers using a robust loss function," *IEEE Robotics and Automation Letters*, 2025, doi: 10.1109/LRA.2025.3606377.
- [27] U. S. D. M. Agency, *Department of Defense World Geodetic System 1984: its definition and relationships with local geodetic systems*. Defense Mapping Agency, 1987, vol. 8350.
- [28] M. Zhi, C. Deng, H. Zhang, H. Tang, J. Wu, and B. Li, "Rngc-viwo: Robust neural gyroscope calibration aided visual-inertial-wheel odometry for autonomous vehicle," *Remote Sensing*, vol. 15, no. 17, p. 4292, 2023, doi: 10.3390/rs15174292.
- [29] J. A. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl, "Casadi: a software framework for nonlinear optimization and optimal control," *Mathematical Programming Computation*, vol. 11, pp. 1–36, 2019, doi: 10.1007/s12532-018-0139-4.
- [30] L. Vanroye, A. Sathya, J. De Schutter, and W. Decré, "Fatrop: A fast constrained optimal control problem solver for robot trajectory optimization and control," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 10036–10043, doi: 10.1109/IROS55552.2023.10342336.
- [31] B. Wang, Z. Ma, S. Lai, and L. Zhao, "Neural moving horizon estimation for robust flight control," *IEEE Transactions on Robotics*, vol. 40, pp. 639–659, 2023, doi: 10.1109/TRO.2023.3331064.
- [32] B. Wang, X. Chen, and L. Zhao, "Trust-region neural moving horizon estimation for robots," in *2024 IEEE International Conference on*

Robotics and Automation (ICRA). IEEE, 2024, pp. 2059–2065, doi: 10.1109/ICRA57147.2024.10611703.



Hugo S. U. Hernández is a Bioengineer and a PhD candidate at the Universidad Nacional del Litoral (UNL), Argentina, and a member of the Research Institute for Signals, Systems and Computational Intelligence (sinc(i) - CONICET). His research focuses on advanced state-estimation methods, particularly Moving Horizon Estimation, with applications to autonomous ground vehicle localization and multi-sensor fusion.



software for collaborative research initiatives.

Agustín Capovilla is a Support Staff (CPA) member at the Research Institute for Signals, Systems and Computational Intelligence (sinc(i) - CONICET) and a Professor at the Universidad Nacional del Litoral (UNL). He is in the final year of a Computer Engineering degree at the Facultad de Ingeniería y Ciencias Hídricas, UNL. He is actively involved in robotics projects, specializing in the design and construction of autonomous vehicles, electronic circuit design, and embedded systems programming, as well as the implementation and integration of



Leonardo L. Giovanini is a Principal Researcher at the Research Institute for Signals, Systems and Computational Intelligence (sinc(i) - CONICET) and Professor at the Universidad Nacional del Litoral (UNL). He received his PhD in Engineering Science from UNL. His research interests focus on advanced modeling, optimal control, and estimation techniques for complex systems, with applications to networked systems, precision farming, and robotics.



Guido M. Sánchez is a Researcher at the Research Institute for Signals, Systems and Computational Intelligence (sinc(i) - CONICET) and an Associate Professor at the Universidad Nacional del Litoral (UNL). He received his PhD in Engineering from UNL. His research focuses on advanced automatic control systems and estimation, particularly for autonomous ground vehicles, including path-tracking, sensor fusion, nonlinear model predictive control and nonlinear moving horizon estimation.