On Robust Statistics for GNSS Single Point Positioning

Daniel Medina[1], Haoqing Li[2], Jordi Vilà-Valls[3] and Pau Closas[2]

Abstract— **Navigation problems are generally solved applying least-squares (LS) adjustments. Techniques based on LS can be shown to perform optimally when the system noise is Gaussian distributed and the parametric model is accurately known. Unfortunately, real world problems usually contain unexpectedly large errors, so-called outliers, that violate the noise model assumption, leading to a spoiled solution estimation. In this work, the framework of robust statistics is explored in order to provide a robust solution to the Global Navigation Satellite Systems (GNSS) single point positioning (SPP) problem. Considering that GNSS observables may be contaminated by erroneous measurements, we survey most popular approaches for robust regression and how they can be adapted into a general methodology for robust SPP.**

I. INTRODUCTION

Global Navigation Satellite Systems (GNSS) play a fundamental role on prospective applications of Intelligent Transportation Systems (ITS), as the main source of positioning information. Besides, GNSS provides timing synchronization to critical applications such as power grid or the stock market. However, GNSS performance can be easily degraded by natural phenomena and signal reflection. Navigation in urban scenarios results particularly challenging due to the presence of severe multipath effects, inducing large errors in the observed pseudorange measurements. Most positioning techniques are based on maximum likelihood (ML) estimation, since the later provides optimal solutions under the assumption of Gaussian distributed observation noise. Although this assumption is generally fulfilled for nominal GNSS open-sky conditions, positioning on signal-degraded scenarios constitutes a challenge for ML estimators such as the least-squares (LS) [1].

Thus, the GNSS community has devoted great efforts towards the development of resilient navigation solutions. One of the most popular approaches is based on solution separation –also known as consistency-checking–, where a statistical test is applied to the estimated residuals to verify whether the Gaussian assumption is fulfilled. Otherwise, combinations of subsets excluding one observation are computed and the statistical test applied again. This procedure is repeated until a fault-free subset is found. Advanced Receiver Autonomous Integrity Monitoring (ARAIM) is possibly the most well-known representative of the solution separation approach, becoming the *de facto* navigation method for vertical guidance in the aviation domain [2], [3]. Numerous other works have adapted consistency-checking navigation algorithms for single point positioning (SPP) in signal-degraded scenarios [4]–[7]. Unfortunately, with the deployment of growing availability GNSS constellations, solution separation methods present challenging scalability issues since their computation complexity constitutes a combinatorial problem with the number of observations and outliers, eventually becoming an intractable problem.

Robust statistics provides an alternative framework for the definition of navigation methods resilient against multiple erroneous observations. Originally suggested for general data analysis in the early 1970s [8]–[10], robust estimators has experienced substantial research growth and its use has extended to manifold fields: signal processing [11]–[13], biomedical [14], [15], power systems [16], etc. The application of robust estimators to compute position, velocity and time (PVT) solutions in satellite-based navigation has appealed numerous authors, both for memory-less single point positioning [17]–[20] and for recursive estimation [21]–[23]. In that PVT context, the performance of robust techniques have been demonstrated on both simulated and real data, and this paper attempts at characterizing those estimators in terms of quantities relevant to the robust statistics literature. It is out of the scope of this work the consideration of robust statistics in other components of a GNSS receiver, however it is remarkable its use in proposing the so-called Robust Interference Mitigation (RIM) paradigm [24], which presents a robust cross-ambiguity function (CAF) that provides the receiver with enhanced anti-jamming resiliency [25]–[29].

This work introduces the principles of robust statistics for regression problems and presents three of the most popular robust methods: M-, S- and MM-estimators. Besides, a comprehensive guide on the implementation of such techniques for solving the GNSS single point positioning problem is detailed. Moreover, the specific challenges on the application of robust estimators for GNSS positioning are discussed. In this article, simulation experiments are carried out to evaluate the positioning capabilities of the M-, S- and MM-estimators against classical LS. In those experiments, the pseudorange observations are contaminated with a percentage of outliers, ranging from 10 to 40 %, of different magnitude. Finally, the Gaussian efficiency and the capability of mitigating the effects of outliers is addressed over different data sizes, to verify the importance of data redundancy for the performance of robust estimators.

The rest of the paper is organized as follows. In Section

This work has been partially supported by the DGA/MRIS (2018.60.0072.00.470.75.01) and NSF under Awards CNS-1815349 and ECCS-1845833.

[1]D. Medina is with the Institute of Communications and Navigation, German Aerospace Center (DLR), Germany. [2]H. Li and P. Closas are with the Dept. of Electrical & Computer Engineering, Northeastern University, Boston MA, USA. [3]J. Vilà-Valls is with Institut Supérieur de l'Aéronautique et de l'Espace (ISAE-SUPAERO), University of Toulouse, France.

2, the basics of robust estimation are introduced. Section 3 relates the specific implementation details of using robust techniques in the GNSS single point positioning problem. Section 4 present the simulation results and discusses the performance of robust estimators. Finally, in Section 5 the outlook along with the next steps on the future work are presented.

II. Background on Robust Estimation

A traditional way to represent 'well-behaved' data is to assume that the underlying noise is normal distributed, with known parametrization,

$$\varepsilon \sim \mathcal{N}\left(\mu, \sigma^2\right). \tag{1}$$

Thus, classical regression methods assume that ε belongs to an *exactly* known parametric distribution [30]. If this assumption holds, the LS estimate is optimal. However, in many practical situations Gaussianity does not hold and we may speak of *approximately normal* measurements, for instance due to larger probability of observations occurring at the tails of the distribution. In those circumstances, the aforementioned optimality is lost for LS, and robust estimators become relevant to provide *close-to-optimal* results in non-nominal conditions. The concept of approximate normality can be formalized by considering that a proportion $1-\epsilon$ of the observations are affected by Gaussian noise, while with complementary probability $0 \leq \epsilon \leq 1$ the data is contaminated by an unknown (potentially) non-Gaussian distribution,

$$\varepsilon \sim (1-\epsilon)\, G + \epsilon\; H \tag{2}$$

where $G = \mathcal{N}\left(\mu, \sigma^2\right)$ is the nominal Gaussian distribution and H is an arbitrary contaminating distribution. Notice that another approach for modeling outliers involves the use of *heavy-tailed* distributions, whose tails tend to zero at a slower rate than the Gaussian distribution. Cauchy, Laplace, Student-t or α-stable distributions are examples of such heavy-tailed densities. This section introduces a basic notion on robust statistics and on some of the most well-known robust estimators for regression problems. For a detailed theoretical analysis of robust statistics, the reader is referred to classical textbooks [10], [31], [32], or the recent works [11], [33] for its application to a variety of signal processing problems.

A. Robust Statistics Dictionary

Some basic concepts from robust statistics are introduced in this section. First we define *qualitative robustness* adopting Hampel's definition [8]. In plain words, if a bounded change in the distribution of the observations is seen as a bounded change in the distribution of the estimates, then the claim is that the estimator is robust. More precisely, let $\mathcal{X} = \{x_1, \dots, x_n\}$ be a set of i.i.d. observations from a distribution F, and let $T_n = T_n(\mathcal{X})$ be a sequence of estimates. Then T_n is called *robust* at $F = F_0$ if the sequence of maps of distributions[1], $\mathfrak{L}_F(T_n)$ is equicontinuous at F_0, that is, if we take a suitable distance d_*, in the space of probability measures, and assume that for all $\delta_2 > 0$ there exists a $\delta_1 > 0$ such that,

$$d_*(F_0, F) \leq \delta_1 \Rightarrow d_*\left(\mathfrak{L}_{F_0}(T_n), \mathfrak{L}_F(T_n)\right) \leq \delta_2\,. \tag{3}$$

Another important metric is the *breakdown point* ϵ^*, which was first defined as the smallest percentage of contamination that can cause the estimator to take on arbitrarily large aberrant value [8]. Later, the concept of breakdown point on finite sets was introduced in [34]. Thus, taken any sample $\mathcal{X}$ of n observations and any estimator $T \triangleq T_n$. Let $\beta\left(m, T, \mathcal{X}\right)$ be the supremum of $\|T\left(\mathcal{X}'\right) - T\left(\mathcal{X}\right)\|$ for all corrupted samples $\mathcal{X}'$ where m of the original n observations are replaced by arbitrary values. Then, the breakdown point of an estimator T is defined as

$$\epsilon^*_m\left(T, \mathcal{X}\right) = \min\left\{\frac{m}{n}, \beta\left(m, T, \mathcal{X}\right)\right\} \tag{4}$$

If a set of observation is to follow a mixture model as in Eq. 2, those healthy observations following a known distribution are referred to as *inliers*. On the other hand, observations that are well separated from the majority of the data are generally referred to as *outliers* within the framework of robust statistics.

Robust estimators provide resiliency to outliers, but they do it at the price of some performance degradation under the nominal model, that is when all observations are inliers. The way to quantify such degradation is through the so-called *loss-of-efficiency* (LoE), that is defined as the ratio of performances between the optimal method (e.g., the LS) and the robust estimator, both using measurements from the nominal model, $\epsilon = 0$.

B. Robust Estimates for Regression Problems

Consider a linear regression problem $y_t = \mathbf{z}_t^\top \mathbf{x} + n_t$, with $t = 1, \dots, N$, and $\mathbf{x}$ a parameters vector to be estimated, or in vector form, $\mathbf{y} = \mathbf{Z}\mathbf{x} + \mathbf{n}$. We can define a vector $\mathbf{r} = \mathbf{y} - \mathbf{Z}\mathbf{x}$ of observation residuals. The regression is generally solved applying a LS estimator (minimization of the ℓ_2-norm of the residuals),

$$\hat{\mathbf{x}}_{LS} = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{Z}\mathbf{x}||_2^2 \Rightarrow \arg\min_{\mathbf{x}} \sum_{i=1}^{N} \left(\frac{r_i(\mathbf{x})}{\sigma_i}\right)^2, \tag{5}$$

which is optimal when the Gaussian noise assumption for $\mathbf{n}$ holds. However, it lacks robustness as even a single outlier could completely spoil the estimation. A first approach towards protecting against outlying measurements is the *least-absolute value* or ℓ_1, consisting on the substitution of the squared residuals as

$$\hat{\mathbf{x}}_{\ell_1} = \arg\min_{\mathbf{x}} \sum_{i=1}^{N} \left|\frac{r_i(\mathbf{x})}{\sigma_i}\right|. \tag{6}$$

Nonetheless, the ℓ_1 method retains a sum of residuals and thus the influence of outliers is still unbounded. This problem can be generalized by considering a general loss function $\rho(x)$ (a.k.a. ρ-function). For instance, $\rho_{\text{LS}}\left(x\right) = x^2$ and $\rho_{\ell_1}\left(x\right) = |x|$ correspond to the aforementioned estimation

[1] $\mathfrak{L}_F(T_n)$ stands for the distribution of an estimator (or test statistic) T_n under F.

approaches. The framework of robust statistics proposes loss functions $\rho(\cdot)$ such that the estimates are nearly optimal when the noise is exactly normal and nearly optimal when the noise is approximately normal (e.g., contaminated normal). We define $\psi(x) = \frac{\partial \rho(x)}{\partial x}$, called the influence function (a.k.a. ψ-function). Several robust estimators of regression have been proposed in the literature, the most popular being: *i*) M-estimate, *ii*) S-estimate, and *iii*) MM-estimate. In the sequel, the loss functions for robust statistics are introduced, alonside some details on the robust estimators, for which Fig. 1 provides some pictorial support.

1) Huber and Tukey Families of Loss Functions: The key idea behind robust estimation is to use loss functions which appropriately penalize measurements with outliers. Several loss functions exist in the literature, the most common being Huber and Tukey's bisquare families of functions. The family of Huber functions is defined as

$$\rho_a^H(x) = \begin{cases} x^2 & \text{if} \quad |x| \leq a \\ 2a|x| - a^2 & \text{if} \quad |x| > a \end{cases}, \tag{7}$$

$$\psi_a^H(x) = \begin{cases} x & \text{if} \quad |x| \leq a \\ a \operatorname{sgn}(x) & \text{if} \quad |x| > a \end{cases}, \tag{8}$$

$$W_a^H(x) = \min\left\{1, \frac{a}{|x|}\right\}, \tag{9}$$

then $\rho_a^H(x)$ is quadratic around 0 and increases linearly with x. In the case of location estimation, the limit cases, $a \to \infty$ and $a \to 0$ correspond to the mean and median estimates, respectively. A desirable property of ρ-functions is boundedness, which implies redescending ψ-functions that tend to 0 at infinity. A popular choice is the Tukey's bisquare or biweight family of functions,

$$\rho_c^B(x) = \begin{cases} 1 - \left(1 - \left(\frac{x}{c}\right)^2\right)^3 & \text{if} \quad |x| \leq c \\ 1 & \text{if} \quad |x| > c \end{cases} \tag{10}$$

$$\psi_c^B(x) = x\left(1 - \left(\frac{x}{c}\right)^2\right)^2 I(|x| \leq c), \tag{11}$$

$$W_c^B(x) = \left(1 - \left(\frac{x}{c}\right)^2\right)^2 I(|x| \leq c), \tag{12}$$

with $c > 0$ a constant parameter and $I(|x| \leq c)$ the indicator function, i.e., $I(|x| \leq c) = 1$ if $|x| \leq c$, and 0 if $|x| > c$.

Typically, the constant parameter in both functions are fixed to achieve a given efficiency to the normal distribution. For a 95% of efficiency, $a = 1.345$ for the Huber function, and $c = 4.685$ for the Tukey function.

2) M-estimator: the M-estimate of regression is defined as

$$\hat{\mathbf{x}}_M = \arg\min_{\mathbf{x}} \sum_{i=1}^{N} \rho\left(\frac{r_i(\mathbf{x})}{\sigma_i}\right), \tag{13}$$

with $\hat{\sigma}_i$ an estimate of the scale of errors n_i, or equivalently, as the solution to

$$\sum_{i=1}^{N} \psi\left(\frac{r_i(\mathbf{x})}{\sigma_i}\right) \frac{\partial\left(r_i(\mathbf{x})/\sigma_i\right)}{\partial \mathbf{x}} = \sum_{t=1}^{N} \psi\left(\frac{y_t - \mathbf{z}_t^\top \hat{\mathbf{x}}_M}{\sigma_i}\right) \mathbf{z} = \mathbf{0}, \tag{14}$$

which is commonly solved by an Iteratively Reweighted LS (IRLS), with an instrumental weight function defined as

$$W(x) = \begin{cases} \psi(x)/x, & \text{if } x \neq 0 \\ \psi'(0), & \text{if } x = 0 \end{cases}, \tag{15}$$

to provide the convenient alternative formulation,

$$\sum_{i=1}^{N} W\left(r_i/\sigma_i\right) \partial\left(\left(\frac{r_i}{\sigma_i}\right)^2\right) / \partial \mathbf{x} = 0. \tag{16}$$

Solving such system requires finding the state estimate *per se* as well as the weights for each of the observations according to the corresponding weighting function. Notice that a normalization using the dispersion of the residuals σ_i is included in the formulation, because these estimates are not scale equivariant. An estimate of the dispersion must be used, $\hat{\sigma}_i$, for instance, the normalized median absolute deviation MAD, defined as

$$\mathrm{MAD}(\mathbf{x}) = c_m \ \mathrm{Med}(|\mathbf{x} - \mathrm{Med}(\mathbf{x})|) \tag{17}$$

being $\mathrm{Med}(\mathbf{x})$ the median of $\mathbf{x}$, and c_m a normalizing constant (≈ 1.4815 for the normal case).

3) S-estimator: the S-estimate of regression is defined as the estimator that minimizes the robust scale M-estimate,

$$\hat{\mathbf{x}}_S = \arg\min_{\mathbf{x}} s_M(\mathbf{r}(\mathbf{x})), \tag{18}$$

with $s_M(\mathbf{r}(\mathbf{x}))$ the M-estimate of scale, which satisfies

$$\frac{1}{N} \sum_{i=1}^{N} \rho\left(\frac{r_i(\mathbf{x})}{s_M(\mathbf{r}(\mathbf{x}))}\right) = b, \tag{19}$$

and thus,

$$\hat{\mathbf{x}}_S = \arg\min_{\mathbf{x}} \sum_{i=1}^{N} \rho\left(\frac{r_i(\mathbf{x})}{\hat{s}}\right), \hat{s} = s_M(\mathbf{r}(\hat{\mathbf{x}}_S)) \tag{20}$$

Again this is solved by an IRLS. A typical choice is the bisquare scale with $\rho(x) = \min\{1 - (1 - x^2)^3, 1\}$ and $b = 0.5$. In this case, $W(x) = \min\{3 - 3x^2 + x^4, 1/x^2\}$, where it's clear that larger values of x have smaller weights. The main problem of S-estimators is that they cannot achieve simultaneously a high BP and high efficiency at the normal distribution.

4) MM-estimator: this robust estimator is designed to achieve both high efficiency and high BP simultaneously. If we consider two bounded loss functions, ρ_0 and ρ_1, which satisfy $\rho_1 < \rho_0$, then the MM estimator is defined as

$$\hat{\mathbf{x}}_{MM} = \arg\min_{\mathbf{x}} \sum_{i=1}^{N} \rho_1\left(\frac{r_i(\mathbf{x})}{s_N(\mathbf{r}(\hat{\mathbf{x}}_1))}\right), \tag{21}$$

where $\hat{\mathbf{x}}_1$ is consistent and high BP estimate of $\mathbf{x}$, and $s_N(\mathbf{r}(\hat{\mathbf{x}}_1))$ is the M-estimate of scale of the residuals of $\hat{\mathbf{x}}_1$, computed using ρ_0 and b.

The MM-estimate is build up with three steps:

1) Compute an initial consistent S-estimate of $\mathbf{x}$, $\hat{\mathbf{x}}_1$, with a high BP but possibly low normal efficiency.
2) Compute a M-estimate of the scale of the residuals $s_N(\mathbf{r}(\hat{\mathbf{x}}_1))$ using the high BP estimate $\hat{\mathbf{x}}_1$.

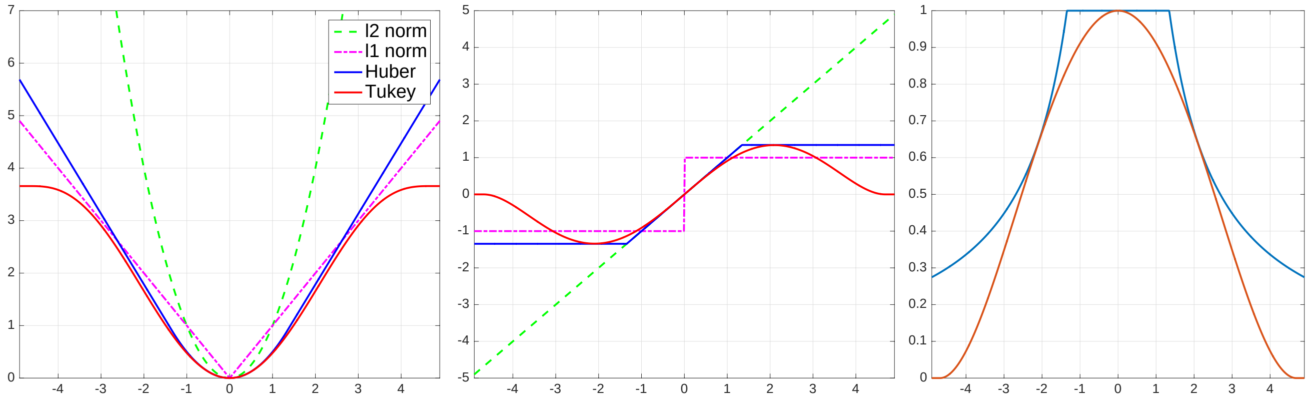


Fig. 1: $\rho(x)$ (left), $\psi(x)$ (middle) and $W(x)$ (right) for: i) ℓ_2-norm (LS), ii) ℓ_1-norm, iii) Huber fct with $a = 1.345$, and iv) Tukey fct with $c = 4.685$.

3) Compute the regression M-estimate starting at $\hat{\mathbf{x}}_1$, considering the robust scale estimate $s_N(\mathbf{r}(\hat{\mathbf{x}}_1))$ and using the recursive IRLS solution.

III. Robust Statistics for GNSS Positioning

The GNSS-based positioning principle consists in solving a geometric problem from the measured ranges to the visible satellites, whose positions are known. Assuming that $n \geq 4$ satellites are tracked, then the observation model to relate the code pseudoranges to the unknown receiver coordinates is as follows:

$$R_i = \|\mathbf{p}_i - \mathbf{p}\|^2 + c\left(\delta t - \delta t_i\right) + I_i + Tr_i + \varepsilon_i \tag{22}$$

where the subscript $i = \{1, \ldots, n\}$ refers to the i-th satellite, R is the observed pseudorange, $\mathbf{p}_i$ and $\mathbf{p}$ denote the satellite and receiver positions respectively, c is the speed of light and δt is the receiver clock offset. Besides, I and Tr denote the ionospheric and tropospheric corrections and ε agglomerates the remaining unmodeled errors (e.g., multipath effects, instrumental delays, phase biases, etc.). The GNSS positioning problem is generally formulated as a regression problem as:

$$\mathbf{y} = h\left(\mathbf{x}\right) + \boldsymbol{\varepsilon} \tag{23}$$

where $\mathbf{y}$ is the n-dimensional observation vector, $h\left(\cdot\right)$ is the observation model from (22) and $\mathbf{x} = \left[\mathbf{p}^\top, c\delta t\right]^\top$ is the state to be estimated. The LS adjustment is the most commonly used method for the estimation of the regression problem of Eq. 23. Since GNSS SPP involves a nonlinear observation model, the problem is typically linearized and solved applying an iterative Gauss-Newton method as follows

$$\Delta\mathbf{x} = \left(\mathbf{H}^\top\mathbf{W}\mathbf{H}\right)^{-1}\mathbf{H}^\top\mathbf{W}\mathbf{y} \tag{24}$$

$$\hat{\mathbf{x}}^k = \hat{\mathbf{x}}^{k-1} + \Delta\mathbf{x} \tag{25}$$

where $\mathbf{H}$ is the Jacobian matrix for the observation model, also known as geometry matrix. That linearization is performed around some guess point $\hat{\mathbf{x}}^{k-1}$ for the k-th iteration of the method, and $\Delta\mathbf{x}$ provides the update for that iteration which will be used to linearize at iteration $k+1$ as in (25). $\mathbf{W}$ is the weighting matrix for the observations. Classical SPP solutions take $\mathbf{W}$ as the inverse of the observations covariance matrix $\mathbf{R}$. Stochastic modelling of pseudorange observations has been a recurrent topic within the GNSS community. A simplification commonly used is to assume that the observations noise is uncorrelated, zero-mean normal distributed $\varepsilon_i \sim \mathcal{N}\left(0, \sigma_i^2\right)$ [35]. Thus, the covariance is given by

$$\mathbf{R} = \mathbf{W}^{-1} = \text{diag}\left(\sigma_i^2, \ldots, \sigma_n^2\right) \tag{26}$$

where σ_i^2 is derived from combining the uncertainty of the different error sources (satellite ephemeris and clock, ionosphere, troposphere, multipath and receiver noise), as in [36], [37] or from error models dependent on the satellite elevation and/or the signal carrier-to-noise density ratio [38]–[40].

Algorithm 1 describes the IRLS process for the robust estimation of the GNSS SPP. Notice that WLS (a short for weighted least squares) refers to the iterative Gauss-Newton described in (24)-(25), and MAD is defined in (17). N and δ denote the maximum number of iterations of the iterative Gauss-Newton method and the convergence criteria, respectively. The choice of the influence function and the scale estimate is subject on the robust estimator applied – e.g., for the M-estimator, one might use the Huber function in (8) and the MAD as scale estimate.

Remarkably, there are certain specific challenges associated to the GNSS-based positioning problem that we point out in this paper. On the one hand, the observation model $h\left(\cdot\right)$ is nonlinear. Thus, the IRLS procedure for finding the observations weights based on the M-estimator concatenates with the iterative LS used for dealing with the model nonlinearity. On the other hand, the GNSS problem is characterized by presenting *fat* data samples, namely, there is a low redundancy of observations. Since generally only around a dozen satellites are tracked and at least four parameters are to be estimated, GNSS SPP constitutes a severe case of low redundancy regression problem [41]. Lastly, the general assumption on robust statistics of independent and identically

Algorithm 1 IRLS procedure for robust SPP

Initial WLS $\rightarrow \{\mathbf{x}^0, \mathbf{r}^0\} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2_{W^{-1}}$
Initial scale $\rightarrow \hat{\sigma}^0 = \text{MAD}\left(\mathbf{r}^0\right)$
Normalized residuals $\mathbf{d}^0 = \mathbf{r}^0 / \hat{\sigma}^0$
for $k := 1, 2, \dots N$ **do**
  $\mathbf{w}^k = \psi(\mathbf{d}^{k-1}) / \mathbf{d}^{k-1}$
  $\mathbf{W} = \text{diag}\left(w_1^k, \dots, w_n^k\right)$
  WLS $\rightarrow \mathbf{x}^k, \mathbf{r}^k$
  Estimate scale $\rightarrow \hat{\sigma}\left(\mathbf{r}^k\right)$
  $\mathbf{d}^k = \mathbf{r}^k / \hat{\sigma}^k$
  if $\|\mathbf{x}^k - \mathbf{x}^{k-1}\| < \delta$ **then**
    Stop
  end if
end for

distributed noise is not met for the GNSS case. Not only are GNSS observations noise uniquely described using stochastic models, but the assumption of independent noise can be violated for satellites of similar direction-of-arrival (e.g., for multipath and none line of sight effects), or for all satellites (e.g., under the influence of a jamming attack or an ionospheric storm).

IV. Test and Results

The performance of robust estimators, as well as classical LS for GNSS positioning, is compared using a simulation environment. Different % of outlying observations ϵ and outlier magnitudes are considered. The magnitude of the outliers α is defined as the ratio between inlier, or *healthy* observations, and outliers. The sky plot of the tracked satellite is as shown in Fig. 1 and results are obtained after averaging 10^4 Monte Carlo runs. For the simulation of

TABLE I: Parameters for the Monte Carlo simulation.

UTC time	15/05/2017 09 : 30 : 00
Location	Koblenz, Germany (50°21'56"N, 7°35'55"E)
Number of satellites n	10
Observation variance noise [m^2]	4
Outlier percentage ϵ	0 - 10 - 20 - 30 - 40
Outlier magnitude α	1 - 3 - 6 - 10 - 30 - 60 -100

the observations, a simplified model was considered, where the pseudorange measurements correspond to the geometric distance to the satellites and a simulated receiver clock offset, thus omitting the effects due to atmospheric propagation and the errors in the satellites ephemeris and clock offsets. The inlier observation noise follows a zero-mean normal distribution, whose variance noise is indicated in Table I. Similarly, the noise present in outlying observations follows a zero-mean normal distribution with a variance α times larger than that of healthy observations. First, a simulation is realized based on the actual positions of GPS satellites, as shown in the sky plot of Fig. 2, assuming that the receiver is located in Koblenz (Germany) in May 2017. The evaluated estimators are the regular WLS, the M estimator on the Huber function ($a = 1.345$), the S estimator on the Tukey function ($a = 4.685$ and $b = 0.5$) and the MM estimator (applying on a first stage a S estimator for the scale estimate and later a M estimator, using the same tuning parameters as stated previously). The first row of Fig. 3 shows the

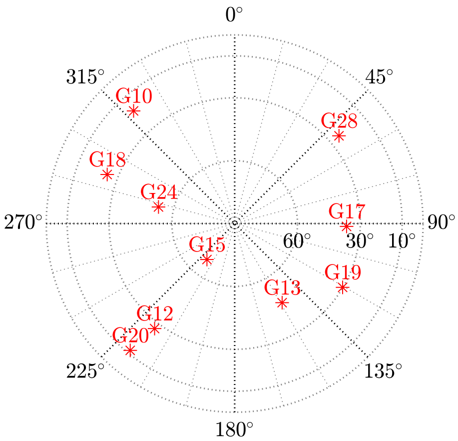
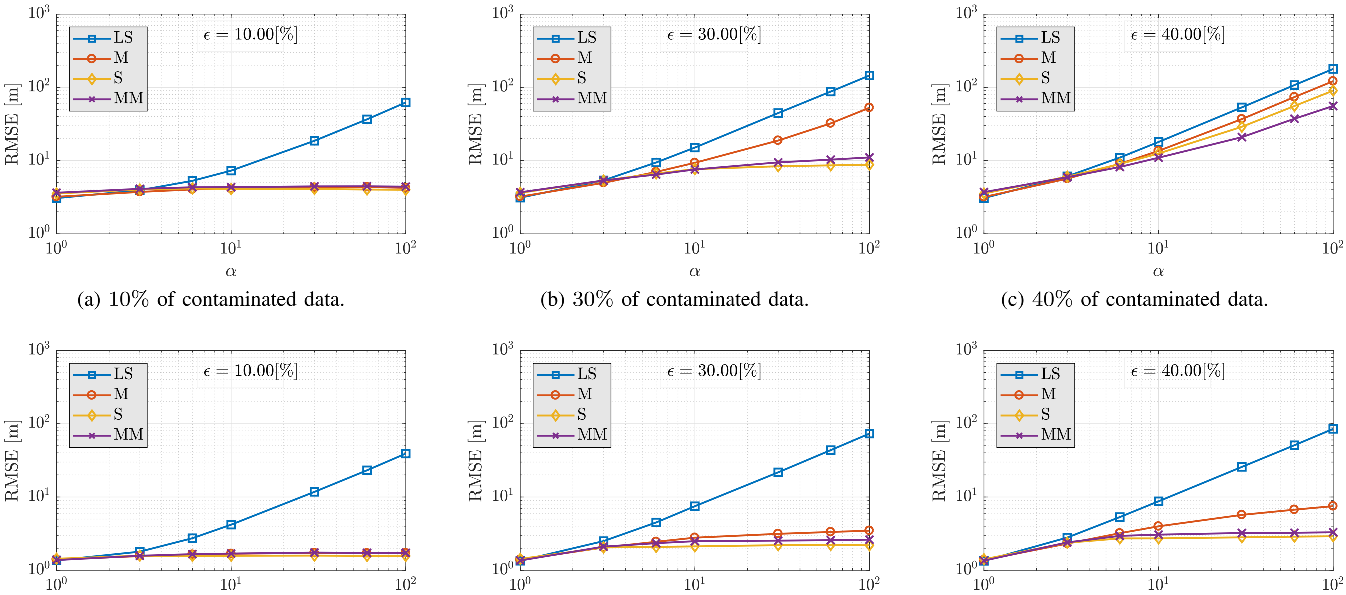


Fig. 2: Sky plot of the tracked satellites for the simulations.

positioning root mean squared error (RMSE) on the ordinate axis, and the magnitude of the outliers is depicted on the abscissa axis. As expected, the LS estimation gets spoiled by the contamined observations, presenting a bias proportional to the size of the outliers. On the other hand, the M,S and MM estimators exhibit certain resilience against the outliers. For a contamination of 10%, the three robust estimators cope perfectly with the contaminated data. Moreover, the performance increases with the size of the outliers, since the detection of these get facilitated by their great impact on the estimation. For large ratios of contaminated data (Figs. 3 (b) and (c)), the characteristics of the estimators become more evident. Since the Huber function applied the M estimation is not redescending – i.e., the effects of the outliers do not get completely eliminated –, the performance of the M estimation rapidly decays. On the contrary, the S and MM estimators make use of the Tukey function, which utterly bounds the effects of the observations presenting the largest residuals. Fig. 3 (c) makes evident the need for observation redundancy to ensure the correct functioning of the robust estimators, since the overall positioning performance gets heavily degraded. Nonetheless, the differences on performance between the M-, S- and MM-estimators support the hypothesis suggested on classical robust theory, for which the S- and MM-estimators pose a higher breakdown point compared to the M-estimator.

Given the prospective scenario in which four GNSS constellations will be fully deployed, it results of great interest the performance characterization of robust estimators under a large number of observations available. Thus, a second experiment is carried out by simulating $n = 40$ satellites (azimuth $\sim \mathcal{U}\left(0, 2\pi\right)$, elevation $\sim \mathcal{U}\left(0, \pi/4\right)$ and distance $\sim \mathcal{N}\left(20.200[\text{km}], 2.000[\text{km}^2]\right)$). The second row of Fig. 3 shows the positioning performance of the evaluated WLS



(a) 10% of contaminated data.

(b) 30% of contaminated data.

(c) 40% of contaminated data.

(d) 10% of contaminated data.

(e) 30% of contaminated data.

(f) 40% of contaminated data.

Fig. 3: RMSE positioning error for $\epsilon \in \{10, 30, 40\}\%$ contamination data (each column) and $n \in \{10, 40\}$ pseudorange observations (each row).

and robust methods. Despite the large of observations, the LS estimation results as spoiled as with a reduced number of measurements, asserting the hypothesis of minimum robustness for classical ML methods. Contrarily, the robust estimators are capable of successfully bounding the effects of outliers, even for the case of 40% data contamination. Again, the S- and MM-estimators manifest the best performance among the robust methods.

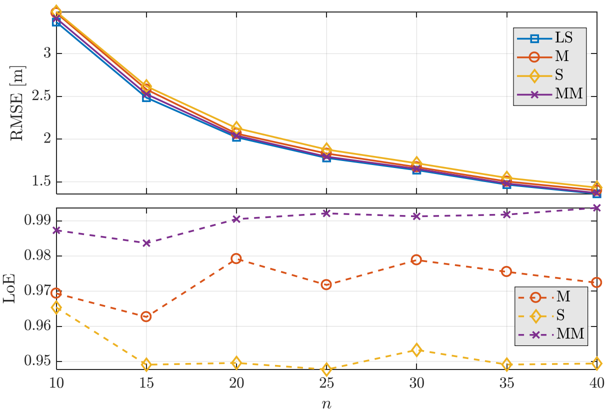


Fig. 4: RMSE performance (top) of LS and robust estimators as a function of the number of measurements and loss-of-efficiency (bottom) comparison of robust methods.

Finally, the Gaussian efficiency of the estimators is studied from the point of view of the loss-of-efficiency (LoE). LoE $\in [0, 1)$ is defined as the ratio between the RMSE of the LS and a particular estimator. Given that LS is an optimal estimator for the normal distributed noise, the higher the LoE for an estimator is, the more efficient at normal distribution such estimator is. Fig. 4 (top) depicts the RMSE of the evaluated methods against the number of observations simulated. While in general the positioning improves with the number of measurements, the LoE of an estimator appears detached from the number of observations, as displayed in Fig. 4 (bottom). The S-estimator presents the lowest LoE and, thus, the poorest Gaussian efficiency. On the other hand, the MM-estimator holds a LoE of approximately 0.99 and it represents the most efficient among the compared methods. Given that MM-estimator poses, together with the S-estimator, the best performance against high percentage of contaminated observations, we might conclude that the MM-estimator resembles the best robust method for the GNSS SPP problem.

V. Outlook and Future Work

This paper provided an overview of robust statistics and how it can be used to enhance the resilience of single point positioning (SPP) solutions in the presence of outliers, caused in practice by multipath propagation or hardware malfunctioning for instance. SPP can be seen as a regression problem, for which this paper presents its robust version leveraging the sound theory of robust statistics. At the same time, the article discusses the specific aspects of applying robust regression to GNSS SPP solvers, and support the discussion with simulation results showing the improvements of such methods as well as their characterization. Future research will provide a better understanding of the loss-of-efficiency incurred by those methods; as well as the relaxation of the i.i.d. assumption among different satellites, and the use of robust techniques in recursive versions that yield to more sophisticated PVT solutions.

REFERENCES

[1] M. G. Amin, P. Closas, A. Broumandan, and J. L. Volakis, "Vulnerabilities, threats, and authentication in satellite-based navigation systems [scanning the issue]," *Proceedings of the IEEE*, vol. 104, no. 6, pp. 1169–1173, 2016.

[2] J. Blanch, T. Walter, P. Enge, Y. Lee, B. Pervan, M. Rippl, and A. Spletter, "Advanced RAIM user algorithm description: integrity support message processing, fault detection, exclusion, and protection level calculation," in *Proceedings of the 25th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS 2012)*, 2012, pp. 2828–2849.

[3] J. Blanch, T. Walter, P. Enge, S. Wallner, F. Amarillo Fernandez, R. Dellago, R. Ioannides, I. Fernandez Hernandez, B. Belabbas, A. Spletter *et al.*, "Critical Elements for a Multi-Constellation Advanced RAIM," *Navigation: Journal of The Institute of Navigation*, vol. 60, no. 1, pp. 53–69, 2013.

[4] A. Angrisano, C. Gioia, S. Gaglione, and G. Del Core, "GNSS reliability testing in signal-degraded scenario," *International Journal of Navigation and Observation*, vol. 2013, 2013.

[5] A. Angrisano, S. Gaglione, and C. Gioia, "RAIM algorithms for aided GNSS in urban scenario," in *2012 Ubiquitous Positioning, Indoor Navigation, and Location Based Service (UPINLBS)*. IEEE, 2012, pp. 1–9.

[6] M. Rippl, G. Schroth, B. Belabbas, and M. Meurer, "A probabilistic assessment on the Range Consensus (RANCO) RAIM Algorithm," in *Proceedings of the ION ITM*, 2009, pp. 248–255.

[7] Z. Jiang, P. D. Groves, W. Y. Ochieng, S. Feng, C. D. Milner, and P. G. Mattos, "Multi-constellation GNSS multipath mitigation using consistency checking," in *Proceedings of the 24th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS 2011)*. Institute of Navigation, 2011, pp. 3889–3902.

[8] F. R. Hampel, "A general qualitative definition of robustness," *The Annals of Mathematical Statistics*, pp. 1887–1896, 1971.

[9] P. J. Huber *et al.*, "Robust regression: asymptotics, conjectures and Monte Carlo," *The Annals of Statistics*, vol. 1, no. 5, pp. 799–821, 1973.

[10] R. A. Maronna, R. D. Martin, V. J. Yohai, and M. Salibián-Barrera, *Robust statistics: theory and methods (with R)*. Wiley, 2018.

[11] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," *IEEE Signal Processing Magazine*, vol. 29, no. 4, pp. 61–80, 2012.

[12] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *International journal of computer vision*, vol. 27, no. 2, pp. 161–195, 1998.

[13] E. Malis and E. Marchand, "Experiments with robust estimation techniques in real-time robot vision," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2006, pp. 223–228.

[14] J. M. Leski, "Robust weighted averaging [of biomedical signals]," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 8, pp. 796–804, 2002.

[15] M. Muma and A. M. Zoubir, "Robust model order selection for corneal height data based on τ estimation," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2011, pp. 4096–4099.

[16] L. Mili, M. G. Cheniae, and P. J. Rousseeuw, "Robust state estimation of electric power systems," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 41, no. 5, pp. 349–358, 1994.

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

[18] H. Kuusniemi, A. Wieser, G. Lachapelle, and J. Takala, "User-level reliability monitoring in urban personal satellite-navigation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 4, pp. 1305–1318, 2007.

[19] K. Fallahi, C.-T. Cheng, and M. Fattouche, "Robust positioning systems in the presence of outliers under weak GPS signal conditions," *IEEE Systems Journal*, vol. 6, no. 3, pp. 401–413, 2012.

[20] J. A. Pozo-Pérez, D. Medina, I. Herrera-Pinzón, A. Heßelbarth, and R. Ziebold, "Robust outlier mitigation in multi-constellation GNSS-based positioning for waterborne applications,," in *Proceedings of the 2017 International Technical Meeting of The Institute of Navigation,*. Institute of Navigation, 2017, pp. 1330–1343.

[21] D. Medina, M. Romanovas, I. Herrera-Pinzón, and R. Ziebold, "Robust position and velocity estimation methods in integrated navigation systems for inland water applications," in *2016 IEEE/ION Position, Location and Navigation Symposium (PLANS)*. IEEE, 2016, pp. 491–501.

[22] O. G. Crespillo, D. Medina, J. Skaloud, and M. Meurer, "Tightly coupled GNSS/INS integration based on robust M-estimators," in *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*. IEEE, 2018, pp. 1554–1561.

[23] H. Pesonen, "Robust estimation techniques for GNSS positioning," in *Proceedings of NAV07-The Navigation Conference and Exhibition, 31.10.-1.11. 2007, London, England*, 2007.

[24] D. Borio and P. Closas, "A fresh look at GNSS anti-jamming," *Inside GNSS*, vol. 12, pp. 54–61, 2017.

[25] D. Borio, "Robust signal processing for GNSS," in *Proc. of the 2017 European Navigation Conference (ENC)*, Lousanne, Switzerland, May 2017, pp. 150–158.

[26] ——, "Myriad non-linearity for GNSS robust signal processing," *IET Radar Sonar and Navigation*, vol. 11, no. 10, pp. 1467–1476, Oct. 2017. [Online]. Available: http://digital-library.theiet.org/content/journals/10.1049/iet-rsn.2016.0610

[27] D. Borio and P. Closas, "Complex signum non-linearity for robust GNSS signal mitigation," *IET Radar Sonar and Navigation*, pp. 1–10, Apr. 2018. [Online]. Available: http://digital-library.theiet.org/content/journals/10.1049/iet-rsn.2017.0552

[28] D. Borio, H. Li, and P. Closas, "Huber's non-linearity for GNSS interference mitigation," *Sensors*, vol. 18, no. 7, p. 2217, 2018.

[29] D. Borio and P. Closas, "Robust transform domain signal processing for GNSS," *Navigation*, vol. 0, no. 0, 2019.

[30] P. J. Huber *et al.*, "Robust estimation of a location parameter," *The annals of mathematical statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[31] P. J. Huber and E. M. Ronchetti, *Robust Statistics*, 2nd ed. Wiley, 2009.

[32] F. R. Hampel *et al.*, *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons, 2011.

[33] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, *Robust Statistics for Signal Processing*. Cambridge University Press, 2018.

[34] P. Rousseeuw and V. Yohai, "Robust regression by means of S-estimators," in *Robust and nonlinear time series analysis*. Springer, 1984, pp. 256–272.

[35] J. S. Subirana, M. Hernandez-Pajares, and J. M. J. Zornoza, *GNSS Data Processing: Fundamentals and Algorithms*. European Space Agency, 2013.

[36] T. Walter and P. Enge, "Weighted RAIM for precision approach," in *PROCEEDINGS OF ION GPS*, vol. 8. Institute of Navigation, 1995, pp. 1995–2004.

[37] R. F. SC-159, *Minimum operational performance standards for global positioning system/wide area augmentation system airborne equipment*. RTCA, 2006.

[38] M.-S. Circiu, M. Meurer, M. Felux, D. Gerbeth, S. Thölert, M. Vergara, C. Enneking, M. Sgammini, S. Pullen, and F. Antreich, "Evaluation of GPS L5 and Galileo E1 and E5a performance for future multifrequency and multiconstellation GBAS," *Navigation: Journal of The Institute of Navigation*, vol. 64, no. 1, pp. 149–163, 2017.

[39] B. Li, "Stochastic modeling of triple-frequency BeiDou signals: estimation, assessment and impact analysis," *Journal of Geodesy*, vol. 90, no. 7, pp. 593–610, 2016.

[40] D. Medina, K. Gibson, R. Ziebold, and P. Closas, "Determination of pseudorange error models and multipath characterization under signal-degraded scenarios," in *Proceedings of the 31st International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2018),*. Institute of Navigation, 2018, pp. 3446–3456.

[41] R. A. Maronna and V. J. Yohai, "Correcting MM estimates for fat data sets," *Computational Statistics & Data Analysis*, vol. 54, no. 12, pp. 3168–3173, 2010.