Title
Technical Note: CRT with Hypothesis Testing

Author
Farrell, Jay A

Publication Date
2017-02-28

GPS-INS Outlier Detection & Elimination using a Sliding Window Filter

Paul F. Roysdon[†] Jay A. Farrell[‡]

***Abstract*— Many applications require reliable, high precision navigation (sub-meter accuracy) while using low-cost inertial and global navigation satellite systems (GNSS). Success requires optimal state estimate while mitigating measurement outliers. Common implementations use an Extended Kalman Filter (EKF) combined with the Receiver Autonomous Integrity Monitoring (RAIM) on a single epoch. However, if the linearization point of the EKF is incorrect or if the number of residuals is too low, outlier detection decisions may be incorrect. False alarms result in good information not being incorporated. Missed detections result in incorrect information being incorporated. Either case can cause subsequent incorrect decisions in the future, possibly causing divergence, due to the state and covariance now being incorrect.**

This article formulates a sliding window estimator that solves the full-nonlinear *Maximum A Posteriori* estimate in real-time. By leveraging the resulting window of residuals, an improved fault detection and removal strategy is implemented. Sensor data is used to demonstrate the interval RAIM (iRAIM) performance improvement.

I. Introduction

The past decade has seen the rapid rise and adoption of navigation systems on automobiles, unmanned vehicles, and personal mobile devices such as smartphones. These systems can exhibit very good accuracy (e.g. sub-meter error). However, further improvements in the reliability and continuity of this accuracy are required to fully support autonomous vehicle operations, especially in urban environments, where variations in the operating conditions and direct signal path can have critical effects. To design a reliable, high-performance system, it is critical to detect and remove outlier measurements before they degrade performance. In GNSS applications such outlier measurements can be caused by multi-path, non-line of sight signals, or overhead foliage.

RAIM is a set of techniques to cope with GNSS receiver outlier measurements, based on measurement residual generation techniques equivalent to least-squares [1], [2]. Integrity is a measure of the trust that can be placed in the correctness of the information supplied by the total system. Often, RAIM is designed assuming only one outlier occurs and that there is enough measurement redundancy to detect and identify the source [1]. The principle of multiple outlier detection has also been well developed over several decades [3], [4]. The authors of [5] included an inertial measurement unit and a Kalman filter to "extend" the RAIM capabilities, a method called eRAIM. However, both RAIM and eRAIM are based on measurements from a single epoch, limiting data redundancy. Furthermore, the residual generation algorithm in RAIM and eRAIM assumes a linear system.

[†]Ph.D. student, [‡]Professor at the Dept. of Electrical & Computer Engineering, UC Riverside. {proysdon, farrell}@ece.ucr.edu.

Data redundancy is critical to successful outlier detection and removal and can be enhanced by considering all GNSS and inertial measurement unit (IMU) data within a sliding temporal window. The resulting full nonlinear Maximum A Posteriori (MAP) estimator, without outlier detection and removal, is presented in [6]. This article extends that approach with methods to detect and remove outlier measurements within the temporal interval and is therefore referred to as interval RAIM (iRAIM). Because this approach allows real-time analysis of numerous fault scenarios, with real-time error correction, outlier detection and removal can be improved. This estimator is demonstrated in using real-world data involving urban canyons and overhead foliage.

II. Background and Notation

This section introduces Global Positioning System (GPS) aided inertial navigation system (INS) background [7].

A. Aided Inertial Navigation

Let $\boldsymbol{x} \in \mathbb{R}^{n_s}$ denote the rover state vector, where

$$\boldsymbol{x}(t) = [\mathbf{p}^\intercal(t), \mathbf{v}^\intercal(t), \mathbf{q}^\intercal(t), \mathbf{b}_a^\intercal(t), \mathbf{b}_g^\intercal(t)]^\intercal \in \mathbb{R}^{n_s},$$

where $\mathbf{p}$, $\mathbf{v}$, $\mathbf{b}_a$, $\mathbf{b}_g$ each in $\mathbb{R}^3$ represent the position, velocity, accelerometer bias and gyro bias vectors, respectively, $\mathbf{q} \in \mathbb{R}^4$ represents the attitude quaternion ($n_s = 16$).

The kinematic equations for the rover state are

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t)), \tag{1}$$

where $\boldsymbol{f} : \mathbb{R}^{n_s} \times \mathbb{R}^6 \mapsto \mathbb{R}^{n_s}$ represents the kinematics, and $\boldsymbol{u} \in \mathbb{R}^6$ is the vector of specific forces and angular rates. The function $\boldsymbol{f}$ is accurately known (see eqn. 11.31-11.33 in [7], and [8]). Nature integrates eqn. (1) to produce $\boldsymbol{x}(t)$.

Let τ_i denote the time instants at which IMU measurements are valid. Assume there is a prior for the initial state: $\boldsymbol{x}(t_0) \sim \mathcal{N}(\boldsymbol{x}_0, \mathbf{P}_0)$. Given the initial condition $\boldsymbol{x}_0$ and the IMU measurements $\tilde{\mathbf{u}}(\tau_i) = \mathbf{u}(\tau_i) + \mathbf{b}(\tau_i) + \boldsymbol{\omega}_u(\tau_i)$ of $\mathbf{u}(\tau_i)$, with additive stochastic errors $\boldsymbol{\omega}_u(\tau_i) \sim \mathcal{N}(\mathbf{0}, \mathbf{Qd})$ and $\mathbf{b} = [\mathbf{b}_a^\intercal, \mathbf{b}_g^\intercal]^\intercal$, a navigation system propagates an estimate of the vehicle state as the solution of

$$\dot{\hat{\boldsymbol{x}}}(t) = \boldsymbol{f}(\hat{\boldsymbol{x}}(t), \tilde{\mathbf{u}}(t)), \tag{2}$$

where $\hat{\boldsymbol{x}}(t)$ denotes the real-time estimate of $\boldsymbol{x}(t)$.

The solution of (2) over the interval $t \in [\tau_{i-1}, \tau_i]$ from the initial condition $\boldsymbol{x}_{i-1}$ is represented as the operator:

$$\phi(\boldsymbol{x}_{i-1}, \mathbf{u}_{i-1}) = \boldsymbol{x}_{i-1} + \int_{\tau_{i-1}}^{\tau_i} \boldsymbol{f}(\boldsymbol{x}(\tau), \mathbf{u}(\tau))\, d\tau \tag{3}$$

where $\hat{\boldsymbol{x}}_{i+1} = \phi(\hat{\boldsymbol{x}}_{i-1}, \hat{\mathbf{u}}_{i-1})$, with $\hat{\mathbf{u}}_{i-1} = \tilde{\mathbf{u}}_{i-1} - \hat{\mathbf{b}}_{i-1}$. Define $\mathbf{U}_{k-1} = \{\tilde{\mathbf{u}}(\tau_i) \text{ for } \tau_i \in [t_{k-1}, t_k]\}$. The integral operator in (3) can be iterated for all IMU measurements in $\mathbf{U}_k$ to propagate the state from t_{k-1} to t_k: $\hat{\boldsymbol{x}}_k = \boldsymbol{\Phi}(\hat{\boldsymbol{x}}_{k-1}, \mathbf{U}_{k-1})$. It is shown in [8] that $\hat{\boldsymbol{x}}_k - \boldsymbol{\Phi}(\hat{\boldsymbol{x}}_{k-1}, \mathbf{U}_{k-1}) = \mathbf{w}_k$ can be modeled with covariance $\mathbf{Q}_{\mathbf{D}k}$.

B. GPS Model

For notational simplicity, it is assumed that the double difference approach removes all common-mode errors (e.g., ionosphere, troposphere, satellite clock and ephemeris errors), as well as the receiver clock biases. Let $t_k = kT$ denote the time instants at which GPS measurements are valid, and $\boldsymbol{x}_k$ denote the state at $\boldsymbol{x}(kT)$. It is typically the case that $T \gg [\tau_i - \tau_{i-1}]$. Therefore, there are numerous IMU measurements available between GPS epochs.

The double-differenced code (pseudorange) and Doppler measurement vector at t_k is modeled as

$$\mathbf{y}_k = \mathbf{h}_k(\boldsymbol{x}_k) + \boldsymbol{\eta}_{yk},$$

with $\mathbf{y}_k$, $\boldsymbol{\eta}_{yk} = [\boldsymbol{\eta}_{\rho k}, \boldsymbol{\eta}_{dk}] \in \mathbb{R}^{2m_k}$. In the theoretical portions of this article, we assume that $m_k = m$, where m is the number of satellites. For $j \in 1, \ldots, m$, the j^{th} component of the vector function $\mathbf{h}_k$ is the geometric distance $\|\mathbf{p}(t_k) - \mathbf{p}^j(t_k)\|_2$ between the rover position $\mathbf{p} \in \mathbb{R}^3$ and the known position of the j^{th} satellite $\mathbf{p}^j \in \mathbb{R}^3$. For $j \in m+1, \ldots, 2m$, the j^{th} component of the vector function $\mathbf{h}_k$ is the projection of the rover velocity onto the satellite line-of-sight vector. The symbol $\boldsymbol{\eta}_{\rho_k} \sim \mathcal{N}(0, \sigma_\rho^2 \mathbf{I})$ represents the pseudorange measurement noise with $\sigma_\rho = 0.1 \sim 3m$, and $\boldsymbol{\eta}_{d_k}^m \sim \mathcal{N}(0, \sigma_d^2 \mathbf{I})$ represents the Doppler measurement noise with $\boldsymbol{\sigma}_d = 0.1 \sim 0.5m/s$. Depending on receiver design, environmental factors and the performance of multipath mitigation techniques. The noise level $\boldsymbol{\sigma}_\rho$ and $\boldsymbol{\sigma}_d$ can vary for each available satellite. The symbol $\mathbf{R}$ will represent the block diagonal matrix of $\boldsymbol{\sigma}_\rho \mathbf{I}$ and $\boldsymbol{\sigma}_d \mathbf{I}$. Using the state estimate, the GPS pseudorange at t_k are predicted to be $\hat{\mathbf{y}}_k = \mathbf{h}_k(\hat{\boldsymbol{x}}_k)$. The GPS measurement residual vector is computed as $\delta \mathbf{y}_k = \mathbf{y}_k - \hat{\mathbf{y}}_k$, with variance $\mathbf{R}$.

III. Estimation Theory

For a known linear system with white, normally distributed, and mutually uncorrelated process and measurement noise vectors with known covariance, the Kalman filter (KF) is the optimal estimator. When the time propagation or measurement models are nonlinear, a variety of methods (e.g., the extended Kalman filter [9]) are available to solve the sensor fusion problem over a single GPS epoch.

This section reviews the *Maximum A Posteriori* estimator [10] solved over a sliding temporal window in real-time. This approach has been developed extensively in the Simultaneous Localization and Mapping (SLAM) research community [11], [12]. The approach developed for GNSS and IMU integration in [6] is referred to as a Contemplative Real Time (CRT) method due to its enhanced ability to detect and remove outliers. That ability has not yet been demonstrated, but is developed and demonstrated herein.

A. Theoretical Solution

Let $\mathbf{X}$ denote the vehicle trajectory over a sliding time window $\mathbf{X} = [\boldsymbol{x}(t_{k-L})^\intercal, \ \ldots \ , \boldsymbol{x}(t_k)^\intercal]^\intercal$, where L is the length of the window, and contains L GPS measurement epochs, $[\mathbf{y}_{k-L+1}, \ldots, \mathbf{y}_k]$. We assume that the window will slide one epoch upon arrival of each new GPS measurement. For presentation purposes only, we assume that each GPS epoch aligns with an IMU measurement time. The results in the experimental section relax this assumption.

Estimation of the vehicle trajectory $\mathbf{X}$ can be formulated as a MAP problem (see Ch. 11.5 of [10]):

$$\hat{\mathbf{X}} = \underset{\mathbf{X}}{\operatorname{argmax}} \left\{ p(\mathbf{X}, \mathbf{U}, \mathbf{Y}) \right\}, \tag{4}$$

where within the time window $\mathbf{U} = \{\mathbf{U}_i \,|\, i \in [k-L, k-1]\}$, and $\mathbf{Y} = \{\mathbf{y}_j \mid j \in [k-L+1, k]\}$ is the set of GPS measurements over the time window for satellites $1, ..., m$. The joint probability for the GPS-INS problem, $p(\mathbf{X}, \mathbf{U}, \mathbf{Y})$, can be decomposed as

$$\begin{aligned} &p(\mathbf{X}, \mathbf{U}, \mathbf{Y}) \\ &= p(\boldsymbol{x}_{k-L}) \prod_{l=k-L}^{k-1} p(\boldsymbol{x}_{l+1} | \boldsymbol{x}_l, \mathbf{U}_l) \prod_{j=k-L+1}^{k} p(\mathbf{y}_j | \boldsymbol{x}_j), \end{aligned} \tag{5}$$

where $p(\boldsymbol{x}_{k-L})$ is the distribution of the initial condition for the time window, $p(\boldsymbol{x}_{l+1} | \boldsymbol{x}_l, \mathbf{U}_{l+1})$ is the distribution of the IMU measurement noise, $p(\mathbf{y}_j | \boldsymbol{x}_j)$ is the distribution of the pseudorange measurement noise $\boldsymbol{\eta}_{\boldsymbol{\rho} k}$.

B. Numerical Solution

Assume that $\boldsymbol{x}(t_0)$, $\boldsymbol{\omega}_u$, and $\boldsymbol{\eta}_y$ have Gaussian distributions with positive definite covariance matrices $\mathbf{P}_{(k-L)}$, $\mathbf{Q}_{\mathbf{D}}$, and $\mathbf{R}$, respectively. Let $\mathbf{W} = \text{blkdiag}(\mathbf{P}_{(k-L)}, \mathbf{Q}_{\mathbf{D}}, \mathbf{R})$. Then $\|\mathbf{v}\|_{\mathbf{W}}^2 = \mathbf{v}^\intercal \mathbf{W}^{-1} \mathbf{v}$ represents the squared Mahalanobis norm.

Finding $\mathbf{X}$ that maximizes eqn. (5) is identical to minimizing the negative of its natural logarithm. This yields the equivalent nonlinear cost function:

$$\begin{aligned} \|\mathbf{v}(\mathbf{X})\|_{\mathbf{W}}^2 = &\|\hat{\boldsymbol{x}}_{k-L} - \boldsymbol{x}(t_{k-L})\|_{\mathbf{P}_{(k-L)}}^2 \\ &+ \sum_{l=k-L}^{k-1} \|\boldsymbol{\Phi}\big(\boldsymbol{x}(t_l), \mathbf{U}_l\big) - \boldsymbol{x}(t_{l+1})\|_{\mathbf{Q}_{\mathbf{D}}}^2 \\ &+ \sum_{j=k-L+1}^{k} \|\mathbf{y}(t_j) - \mathbf{h}_j\big(\boldsymbol{x}(t_j)\big)\|_{\mathbf{R}}^2. \end{aligned} \tag{6}$$

The cost function can be normalized using Cholesky Decomposition. For the positive definite matrix $\mathbf{W}$, defining $\boldsymbol{\Sigma}_{\mathbf{W}}$, such that $\mathbf{W}^{-1} = \boldsymbol{\Sigma}_{\mathbf{W}}^\intercal \boldsymbol{\Sigma}_{\mathbf{W}}$. Then, for $\mathbf{r} \triangleq \boldsymbol{\Sigma}_{\mathbf{W}} \mathbf{v}$, $\|\mathbf{v}\|_{\mathbf{W}} = \|\mathbf{r}\|_2$. The minimization problem of eqn. (6) reduces to the standard nonlinear least squares optimization

$$\min_{\mathbf{X} \in \mathbb{R}^{n_s(L+1)}} \|\boldsymbol{r}(\mathbf{X})\|_2^2$$

which will be solved iteratively.

Consider the l^{th} iteration of the optimization, where l is a positive integer. Given an estimate of the solution

$$\hat{\mathbf{X}}^l = \left[\hat{\boldsymbol{x}}^l(t_{k-L})^\intercal, \ \ldots \ , \hat{\boldsymbol{x}}^l(t_k)^\intercal\right]^\intercal,$$

which is treated as a vector in $\mathbb{R}^{n_s(L+1)}$. The optimization algorithm computes an error vector $\delta\mathbf{X}^l \in \mathbb{R}^{n_e(L+1)}$, which corrects $\hat{\mathbf{X}}^l$ to yield an improved solution $\hat{\mathbf{X}}^{l+1}$ to eqn. (6). The dimension of the error state vector is n_e. The error state vector is $\delta\boldsymbol{x} = [\delta\mathbf{p}^\intercal,\ \delta\mathbf{v}^\intercal,\ \delta\boldsymbol{\theta}^\intercal,\ \delta\mathbf{b}_a{}^\intercal,\ \delta\mathbf{b}_g{}^\intercal]^\intercal \in \mathbb{R}^{15}$, where $\delta\mathbf{p}$, $\delta\mathbf{v}$, $\delta\boldsymbol{\theta}$, $\delta\mathbf{b}_a$, and $\delta\mathbf{b}_g$ each in $\mathbb{R}^3$ are the position, velocity, attitude, accelerometer bias and gyro bias error vectors, respectively. The dynamics and stochastic properties of this estimation error vector are well understood, and can be found in Section 11.4 of [7]. The fact that $n_s = 16$ and $n_e = 15$ is discussed in [8]. The optimization approach is formulated in the following section.

C. Optimization: Iterated Solution

In the l-th iteration, the *residual* $\mathbf{r}$ linearized around the current estimate $\hat{\mathbf{X}}^l$ is

$$\mathbf{r}(\mathbf{X}) = \mathbf{J}(\hat{\mathbf{X}}^l)\delta\mathbf{X}^l + \boldsymbol{\eta}_\mathbf{r}, \tag{7}$$

where $\mathbf{J}(\hat{\mathbf{X}}^l)$ is the Jacobian of $\mathbf{r}(\mathbf{X})$ evaluated at $\hat{\mathbf{X}}^l$, and $\boldsymbol{\eta}_\mathbf{r} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. In the following, without loss of generality, we consider one iteration of the optimization method and drop the iteration counter 'l'.

With eqn. (7), a convex optimization problem can be formulated by a quadratic approximation $\mathbf{L}(\delta\mathbf{X})$ to the cost function $\mathscr{C}(\mathbf{X}) \triangleq \|\mathbf{r}(\hat{\mathbf{X}})\|_2^2$:

$$\mathbf{L}(\delta\mathbf{X}) = \frac{1}{2}\|\mathbf{r}(\hat{\mathbf{X}}) - \mathbf{J}(\hat{\mathbf{X}})\delta\mathbf{X}\|_2^2. \tag{8}$$

By minimizing $\mathbf{L}(\delta\mathbf{X})$, a candidate step for the estimation error $\delta\mathbf{X}$ is obtained. A line search in the direction of $\delta\mathbf{X}$ is used to update the state estimate.

The Gauss-Newton step is the solution of the normal equation,

$$\mathbf{J}^\intercal\mathbf{J}\delta\mathbf{X} = \mathbf{J}^\intercal\mathbf{b}, \tag{9}$$

where $\mathbf{b} \triangleq \mathbf{r}(\hat{\mathbf{X}})$ and $\mathbf{J} = \mathbf{J}(\hat{\mathbf{X}}^l)$ as defned in (7). Eqn. (9) can be compactly expressed as

$$\boldsymbol{\Lambda}\delta\mathbf{X} = \boldsymbol{\xi}, \tag{10}$$

where $\boldsymbol{\Lambda} = \mathbf{J}^\intercal\mathbf{J}$ is the information matrix, $\boldsymbol{\xi} = \mathbf{J}^\intercal\mathbf{b}$ is the information vector. As shown in [11], $\mathbf{J}$ is a sparse, block diagonal, matrix. Eqn. (10) can be solved efficiently by many methods, e.g. Cholesky, or QR. Further computational gains can be achieved by employing a sparse matrix library as discussed in [11] and [12]. The computational complexity of the algorithm is discussed in [8].

IV. Faulty Measurement Removal

This section examines the Residual Space method for faulty measurement detection and removal. The discussion in this section uses the standard notation in the literature. For application to the window based smoother, $\mathbf{y} = \mathbf{r}(\mathbf{X})$ and $\mathbf{H} = \mathbf{J}$. See eqn. (7).

After the final optimization process iteration, consider the following hypotheses related to the linearized residual $\mathbf{r}(\mathbf{X})$:

- Null Hypothesis, $\mathcal{H}_0$:

$$\mathbf{y} = \mathbf{H}\boldsymbol{x} + \boldsymbol{\eta}, \tag{11}$$

- Alternative Hypothesis, $\mathcal{H}_i$:

$$\mathbf{y} = \mathbf{H}\boldsymbol{x} + \boldsymbol{\eta} + \mu_i\mathbf{e}_i. \tag{12}$$

In both hypotheses, $\mathbf{y} \in \mathbb{R}^{m\times 1}$ is the measurement vector, $\mathbf{H} \in \mathbb{R}^{m\times n}$, $m > n$, $\text{rank}(\mathbf{H}) = n$ is the measurement matrix, $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$ is the vector to be estimated, and the measurement noise $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}) \in \mathbb{R}^m$, where $\mathbf{C} = \sigma_y^2\mathbf{I}$. For the alternative hypothesis, the error vector is $\mathbf{e}_i = [0, \ldots, 0, 1, 0, \ldots, 0]^\intercal \in \mathbb{R}^{m\times 1}$, such that only the i^{th} element is 1. The magnitude of the error is $\mu_i \in \mathbb{R}^{1\times 1}$. To simplify notation in the following equations, let

$$\boldsymbol{\varepsilon}_i \triangleq \mu_i\mathbf{e}_i.$$

In the case of the i^{th} alternate-hypothesis, when the quantity μ_i is nonzero, the i^{th} measurement is called an *outlier*. The magnitude μ_i will affect the ability to detect such outliers. The null-hypothesis assumes no outliers, i.e., $\mu_i = 0$.

A. Null-hypothesis, $\mathcal{H}_0$

From eqn. (11), the minimum-variance unbiased estimator (MVUE) for $\boldsymbol{x}$ is [10]

$$\hat{\boldsymbol{x}} = (\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{y}. \tag{13}$$

To analyze the effect of $\boldsymbol{\eta}$, substitute eqn. (11) into eqn. (13)

$$\hat{\boldsymbol{x}} = ((\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal)(\mathbf{H}\boldsymbol{x} + \boldsymbol{\eta}) = \mathbf{I}\boldsymbol{x} + \mathbf{H}^*\boldsymbol{\eta},$$

where $\mathbf{H}^* \triangleq (\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal$. The state error $\delta\boldsymbol{x} = \boldsymbol{x} - \hat{\boldsymbol{x}}$ due to noise is $\delta\boldsymbol{x} = \mathbf{H}^*\boldsymbol{\eta}$. From the zero mean Gaussian noise assumption, the expected value of the state error is

$$\mathrm{E}\langle\delta\boldsymbol{x}\rangle = \mathbf{0}.$$

Consider the residual $\mathbf{r}$, where $\hat{\mathbf{y}} = \mathbf{H}\hat{\boldsymbol{x}}$. Then

$$\mathbf{r} \triangleq \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{P})\boldsymbol{\eta}, \tag{14}$$

where $\mathbf{P} \triangleq \mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal \in \mathbb{R}^{m\times m}$ is the projection matrix onto the range-space of $\mathbf{H}$, i.e. $\mathcal{C}(\mathbf{H})$. The matrix $\mathbf{P}$ is symmetric, idempotent, and $\text{rank}(\mathbf{P}) = n$.

Similarly, the matrix $\mathbf{Q} \triangleq (\mathbf{I}-\mathbf{P}) \in \mathbb{R}^{m\times m}$ is a real, symmetric, and idempotent matrix. The matrix $\mathbf{Q}$ is a projection matrix onto the left null-space of $\mathbf{H}$, i.e. $LN(\mathbf{H}) = N(\mathbf{H}^\intercal)$. It has eigenvalues equal to 0 or 1, and its trace is equal to the number of non-zero eigenvalues: $(m - n)$.

The mean and covariance of the residual are

$$\begin{aligned}\mathrm{E}\langle\mathbf{r}\rangle &= \mathrm{E}\langle\mathbf{Q}\boldsymbol{\eta}\rangle = \mathbf{0} &(15)\\ \mathrm{Cov}\langle\mathbf{r}\rangle &= \mathrm{E}\langle(\mathbf{r} - \mathrm{E}\langle\mathbf{r}\rangle)(\mathbf{r} - \mathrm{E}\langle\mathbf{r}\rangle)^\intercal\rangle \\ &= \mathbf{Q}\mathrm{E}\langle\boldsymbol{\eta}\boldsymbol{\eta}^\intercal\rangle\,\mathbf{Q}^\intercal = \boldsymbol{\sigma}_y^2\mathbf{Q}. &(16)\end{aligned}$$

The final step is valid because $\mathbf{Q}$ is idempotent, $\mathbf{Q}\mathbf{Q}^\intercal = \mathbf{Q}$, and $\mathrm{E}\langle\boldsymbol{\eta}\boldsymbol{\eta}^\intercal\rangle = \boldsymbol{\sigma}_y^2\mathbf{I}$. The mean square error (MSE) [10] is

$$\begin{aligned}\mathrm{E}\langle\|\mathbf{r}\|^2\rangle &= \mathrm{E}\langle\mathbf{r}^\intercal\mathbf{r}\rangle = \mathrm{E}\langle tr\{\mathbf{r}\mathbf{r}^\intercal\}\rangle \\ &= \mathrm{E}\langle tr\{\mathbf{Q}\boldsymbol{\eta}\boldsymbol{\eta}^\intercal\mathbf{Q}^\intercal\}\rangle = tr\{\mathbf{Q}\}\boldsymbol{\sigma}_y^2 \\ &= (m-n)\boldsymbol{\sigma}_y^2, &(17)\end{aligned}$$

where $tr\{\cdot\}$ is the trace operator.

The standard test statistic $\mathbf{\Gamma}_{\hat{\mathbf{X}}}$ for the validty of $\mathcal{H}_0$ is based on eqns. (14) and (17):

$$\mathbf{\Gamma}_{\hat{\mathbf{X}}} = \frac{\|\mathbf{r}(\hat{\mathbf{X}})\|^2}{\mathrm{E}\left\langle \|\mathbf{r}(\hat{\mathbf{X}})\|^2 \right\rangle} = \frac{\|\mathbf{r}(\hat{\mathbf{X}})\|^2}{(m-n)\boldsymbol{\sigma}_y^2}. \tag{18}$$

Under normal conditions, this test will evaluate to 1.

This is a reduced chi-square statistic [13]. Therefore, to test the hypothesis (i.e., detect the existence of outliers), the test statistic calculated by eqn. (18) is evaluated relative to a threshold computed using the one-tailed Chi-square distribution with respect to a significance level α, normalized by the number of DOF [13], [14]

$$\mathbf{\Gamma}_{\hat{\mathbf{X}}} < \frac{\boldsymbol{\chi}^2_{\alpha/2,(m-n)}}{(m-n)}. \tag{19}$$

The value for $\boldsymbol{\chi}^2_{\alpha/2,(m-n)}$ is determined from a look-up table for α versus DOF. The significance level α is chosen by the designer for some probability of success. For example, $\alpha = 0.05$ indicates a 95% confidence level.

If the test succeeds, $(\hat{\mathbf{X}})$ is finalized as the optimal estimate. Otherwise, outlier identification executes, as discussed in the following section.

B. Alternate-hypothesis, $\mathcal{H}_i$

To analyze the effect of the outlier $\boldsymbol{\varepsilon}_i$ on the state error, substitute eqn. (12) into eqn. (13)

$$\begin{aligned}\hat{\boldsymbol{x}} &= ((\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal)(\mathbf{H}\boldsymbol{x} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}_i) \\ &= \mathbf{I}\boldsymbol{x} + (\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal(\boldsymbol{\eta} + \boldsymbol{\varepsilon}_i).\end{aligned} \tag{20}$$

Therefore, $\delta\boldsymbol{x} = \mathbf{H}^*(\boldsymbol{\eta}+\boldsymbol{\varepsilon}_i)$. The expected value of the state error due to the outlier is $\mathrm{E}\langle\delta\boldsymbol{x}\rangle = \mathbf{H}^*\boldsymbol{\varepsilon}_i$.

To analyze the effect of the outlier on the residual, substitute (12) and (20) into (14):

$$\begin{aligned}\mathbf{r} &= \mathbf{H}\boldsymbol{x} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}_i - \mathbf{H}\hat{\boldsymbol{x}} \\ &= \mathbf{H}\boldsymbol{x} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}_i - \mathbf{H}\left(\mathbf{I}\boldsymbol{x} + (\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal(\boldsymbol{\eta} + \boldsymbol{\varepsilon}_i)\right) \\ &= \mathbf{Q}(\boldsymbol{\eta} + \boldsymbol{\varepsilon}_i).\end{aligned}$$

Note that the residual still lies in the left-null-space of $\mathbf{H}$. The mean and covariance of $\mathbf{r}$ due to the outlier, are

$$\mathrm{E}\langle\mathbf{r}\rangle = \mathrm{E}\langle\mathbf{Q}(\boldsymbol{\eta} + \boldsymbol{\varepsilon}_i)\rangle = \mathbf{Q}\boldsymbol{\varepsilon}_i \tag{21}$$

$$\mathrm{Cov}\langle\mathbf{r}\rangle = \mathrm{E}\langle(\mathbf{r} - \mathbf{Q}\boldsymbol{\varepsilon}_i)(\mathbf{r} - \mathbf{Q}\boldsymbol{\varepsilon}_i)^\intercal\rangle = \boldsymbol{\sigma}_y^2\mathbf{Q}. \tag{22}$$

Comparing eqn. (16) with eqn. (22), we see that both cases with and without the outlier have the same covariance. The difference between the two cases is the mean of the distributions, as shown in eqns. (15) and (21). The difference in the means is important, because it provides the basis for identifying outliers. The decision statistic under the alternate-hypothesis $\mathcal{H}_i$, is based on the distribution of $\mathbf{r} \sim \mathcal{N}(\mathbf{Q}\boldsymbol{\varepsilon}_i, \boldsymbol{\sigma}_y^2\mathbf{Q})$.

Consider the parity vector [1], [2]

$$\mathbf{p} \triangleq \mathbf{U}_2^\intercal\mathbf{r} \in \mathbb{R}^{(m-n)},$$

where $\mathbf{U}_2 \in \mathbb{R}^{m\times(m-n)}$ may be found by the SVD of $\mathbf{H}$. Therefore,

$$\mathbf{p} = \mathbf{U}_2^\intercal(\mathbf{y} - \mathbf{H}\hat{\boldsymbol{x}}) = (\mathbf{U}_2^\intercal\mathbf{e}_i)\mu_i + \mathbf{U}_2^\intercal\boldsymbol{\eta};$$

therefore, $\mathbf{p} \sim \mathcal{N}(\mu_i\mathbf{U}_2^\intercal\mathbf{e}_i,\ \boldsymbol{\sigma}_y^2\mathbf{I}_{m-n})$.

Then the magnitude of the outlier μ_i in eqn. (12) can be estimated as (see Section 5 of [13]),

$$\begin{aligned}\hat{\mu_i} &= \left((\mathbf{U}_2^\intercal\mathbf{e}_i)^\intercal(\boldsymbol{\sigma}_y^2\mathbf{I})^{-1}(\mathbf{U}_2^\intercal\mathbf{e}_i)\right)^{-1}(\mathbf{U}_2^\intercal\mathbf{e}_i)^\intercal(\boldsymbol{\sigma}_y^2\mathbf{I})^{-1}\mathbf{p} \\ &= \boldsymbol{\sigma}_y^2\left(\mathbf{e}_i^\intercal\mathbf{U}_2\mathbf{U}_2^\intercal\mathbf{e}_i\right)^{-1}\frac{1}{\boldsymbol{\sigma}_y^2}\mathbf{e}_i^\intercal\mathbf{U}_2\mathbf{p} = \frac{\mathbf{e}_i^\intercal\mathbf{Q}^\intercal\mathbf{r}}{\mathbf{e}_i^\intercal\mathbf{Q}^\intercal\mathbf{e}_i},\end{aligned}$$

where $\mathbf{U}_2\mathbf{U}_2^\intercal = \mathbf{Q} = (\mathbf{I} - \mathbf{P})$, and the covariance of $\hat{\mu_i}$ is

$$\mathrm{Cov}\langle\hat{\mu_i}\rangle = (\mathbf{e}_i^\intercal\mathbf{Q}^\intercal\mathbf{e}_i)^{-1}.$$

Outlier identification is executed iteratively for each $\mu_i\mathbf{e}_i$ from $i = 1, ..., m$. Each μ_i is compared against a threshold $\boldsymbol{\gamma}$, such that any $\mu_i > \boldsymbol{\gamma}$ is considered an outlier.

After completion of the identification process, if an outlier is identified, it's measurement is removed from the measurement-set and the optimization step in eqn. (9) is repeated.

C. Comparison of DOF's

This section theoretically compares the number of degrees of freedom $(m - n)$ available for outlier detection between algorithms. In this section, n_s is the state dimension; m_k is the number of satellite pseudoranges available at epoch k; n is the total number of real variables to be estimated; and m denotes the total number of available constraints.

The EKF at any time step has $n = n_s$ variables to estimate (one state vector) and $m = n_s + m_k$ constraints (GPS and prior); therefore, the DOF is m_k. The DOF of the Iterated EKF (IEKF) is the same: the IEKF is the same as the CRT with $L = 1$. The advantage of the IEKF is its ability to perform a nonlinear iterative correction.

For the CRT algorithm with window length L, the number of variables to be estimated is $n = (L + 1)n_s$. The number of constraints is $m = (L + 1)n_s + \sum_{j=(k-L+1)}^{k} m_j$. The DOF is therefore, $\sum_{j=(k-L+1)}^{k} m_j$.

Both the outlier detection capability and the amount of required computation are expected to increase with L.

D. Complexity

To evaluate each alternative hypothesis, the rows corresponding to the faulty measurements are removed from both the residual vector and Jacobian matrix, then the nonlinear optimization process is repeated and its likelihood computed. For a large number of alternative hypotheses, this becomes computationally expensive.

Given (mL) residuals in each CRT window, there are

$$\sum_{k=1}^{mL}\begin{pmatrix} mL \\ k \end{pmatrix} = \sum_{k=1}^{mL}\frac{(mL)!}{((mL)-k)!\ k!},$$

ways that any number of satellite measurements could fail in any combination at one epoch (see Section 3 of [4]). For the EKF or IEKF (with $L = 1$ and $m = 9$) this results in 511

hypotheses, which is too large for full consideration. For the CRT with $L = 20$, consideration of all hypothesis is even more infeasible for real-time implementation. Therefore, simplified approaches are required. The approach used herein is to remove all rows with residuals greater than a user defined threshold and re-optimize, possibly repeating until all residuals pass the threshold test.

V. Illustrative Example

Real-world performance is evaluated using data from a drive-test around University of California, Riverside using a consumer-grade GPS antenna (Antcomm ANN-MS-0-005) mounted on the vehicle roof. During driving, the sensor data is time-stamped and stored. The sensor data includes consumer-grade: Quartz-MEMS IMU data (Epson M-G320) used in eqn. (2) at 250Hz, and L1 Differential GPS data (Ublox 6T) used in eqn. (6) at 1Hz. This trajectory contains a variety of real-world automotive conditions that adversely affect GPS receiver performance, e.g. tall buildings and trees.

To allow direct performance of various algorithms, using the identical input data, the results of this section are computed during post-processing. Even though running in post-processing for this evaluation, each algorithm is written in C++ to run on the navigation system in a real-time fashion, using only the data and prior as would be applicable for each approach. All algorithms in this evaluation are capable of real-time implementation on standard computers.

The *ground truth* trajectory is found by solving a nonlinear optimization problem over the entire (600 second) trajectory, formulated in the maximum a posteriori perspective. This smoother uses integer resolved carrier phase DGPS and IMU measurements, to achieve centimeter level accuracy [15].

Due to limited space, only 3D position performance is discussed herein. Velocity and attitude results are provided in [8], and are similar to position performance.

A. Performance After Outlier Removal

For each algorithm, Fig. 1 shows the cumulative distribution function (CDF) of the position error norm $\|\hat{\mathbf{p}}_k - \mathbf{p}_k\|$ where the ground truth trajectory is used as $\mathbf{p}_k$. The value of $\hat{\mathbf{p}}_k$ is the *a posteriori* result after the optimization of (6) at the first time when the k-th epoch enters the sliding window. For outlier removal, the threshold was computed using a significance level of $\alpha = 0.05$. The CRT algorithm curves are included for various window lengths L.

The CDF shows that the percentage of occurrences where the EKF position error is less than $0.1m$, is roughly 18%. Roughly 90% of the trajectory, as estimated by the EKF, has errors less than $1.0m$. This is as expected for a double-difference L1 pseudorange-only GPS-INS with an EKF.

Figure 1 indicates that accuracy improves from the EKF to the IEKF to the CRT. Also, CRT performance (generally) improves with the window length L. For the CRT with $L > 5$, 100% of the position errors are less that $1.0m$. CRT algorithms with $L > 20$ each achieve $0.6m$ position accuracy on 100% of the trajectory. The EKF and IEKF CDF plots do not reach 100% until the position accuracy is over $3.0m$.

B. Sensitivity Analysis

The improved performance demonstrated in Fig. 1 is attributed to solving the full nonlinear optimization over a longer window, which enhances the redundancy and allows reconsideration of fault decisions, as long as the measurement data is within the sliding window. This enhanced ability to detect and remove outliers to achieve reliable performance is one of the major motivations of the CRT approach. This section considers the robustness to outliers in greater detail.

Fig. 1 considered a single detection threshold γ, which is a decreasing function of the significance level α (see eqn. (19), where $\gamma = \frac{\chi^2_{\alpha/2,(m-n)}}{(m-n)}$, and $\alpha = 0.05$). Fig. 2 presents a receiver operating characteristic (ROC) curve for each algorithm. Each ROC curve plots the Probability of Detection, P_D, versus Probability of False Alarm, P_{FA} as a function of γ.

Since ground truth is available, the correct outlier decisions are known for each satellite at each epoch. To construct the ROC curve for each algorithm, the procedure of Section V-A is repeated once for each specified value of γ. The outlier detection decisions for each algorithm and each value of γ are compared with the ground truth decisions to compute P_D and P_{FA}. Each run for a single value of γ generates one point on the ROC curve.

For example, to evaluate the EKF using eRAIM methods, a significance threshold α is chosen, γ is computed and held constant, and the EKF is run for the entire trajectory. At the completion of the trajectory, the probability of detection (of outliers) is calculated (against ground truth), as well as the probability of false alarm, and the values are recorded. This provides one point on the EKF ROC curve. The trajectory estimation is repeated for a vector of α values, using the EKF. This set of values provides the EKF ROC curve.

This process is then repeated for each estimator: IEKF, and CRT with $L = \{5, 10, 20, 30, 40\}$. In the case where a window of data is evaluated, such as the CRT estimator with

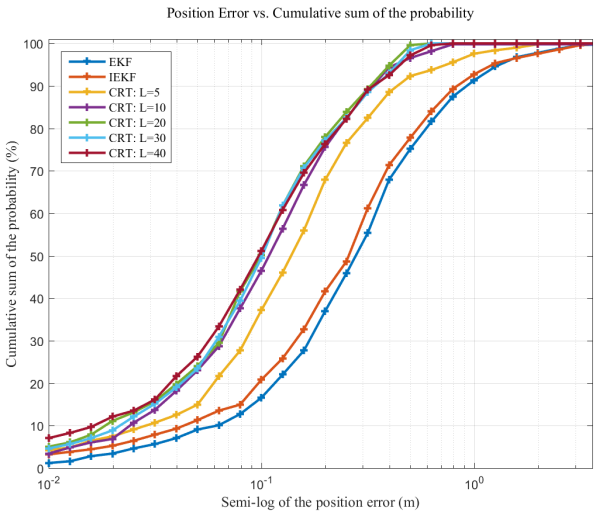

Fig. 1. Cumulative distribution of position error for each algorithm.

$L > 1$, the outlier detection algorithm evaluates the entire residual vector after convergence to the optimal trajectory for the hypothesized set of valid measurements. If outliers are detected, the corresponding measurements are removed, and the optimization is repeated for one additional iteration. Upon sliding the window, the prior outliers are ignored and the detection procedure searches the entire residual vector. More advanced processes both for accommodating outliers or reconsidering past outlier hypotheses will be considered in future research.

Each curve in Fig. 2 displays the expected P_D vs. P_{FA} trends for α. As γ increases, the next point on each curve will be below and to the left of the last point, as P_D and P_{FA} both decrease with γ. As Fig. 2 shows, the rates of decrease are very different for the different algorithms. Curves closer to the upper left corner show improved ability to detect outliers with lower probability of injecting false alarms. In particular, the CRT algorithms with $L > 10$ each have a P_{FA} below 10% with P_D greater than 90%, whereas both EKF and IEKF have a P_{FA} below 10% only when P_D is less than 60%. In fact, the CRT with $L = 40$ is shown to have a P_{FA} less than 10% with P_D rate greater than 95%.

These results confirm the claim that the ability of the CRT algorithm to discriminate outliers from valid data is enhanced with the length of the sliding window, which increases the redundancy as quantified by the number of degrees of freedom. The reliability of achieving a specified level of accuracy increases with the ability to remove outliers.

VI. Conclusion

This article presented a method to enhance the level of redundancy in a GNSS and IMU based navigation system to facilitate the accommodation of outlier measurements. Over a multiple epoch sliding window of data the algorithm performs MAP estimation within a nonlinear optimization framework, while maintaining a real-time estimate as necessary for control and planning purposes. Increasing the duration L of the sliding window enhances redundancy at the expense of increased computation. Enhancing redundancy improves the reliability of achieving any given accuracy specification, by better outlier removal. The MAP framework, through real-time nonlinear optimization, achieves optimal state estimation without linearization assumptions. The enhanced performance of these methods is demonstrated through direct comparisons of both the accuracy and outlier detection abilities of various algorithms using experimental data from a challenging environment.

Current research in the fields of machine learning and robust statistics allow new approaches for multiple outlier detection and removal, including residual generation performed with non-quadratic cost functions and thresholding [16], [17]. Our future research will investigate such outlier accommodation algorithms within the sliding window MAP framework.

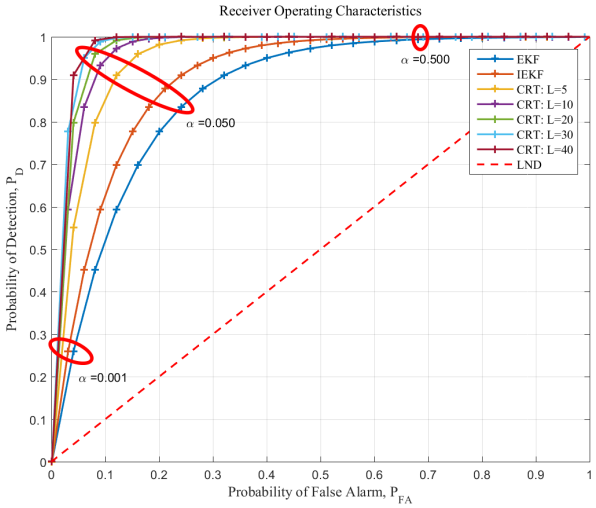

Fig. 2. Receiver Operating Characteristic curves for each algorithm with varying α. The red dotted line is the Line-of-No-Discrimination.

VII. Acknowledgments

The authors thank Seiko-Epson USA for providing the M-G320 Quartz-MEMS IMU for this research.

References

[1] R. G. Brown, "A Baseline RAIM Scheme and a Note on the Equivalence of Three RAIM Methods," *Proc. Nat. Tech. Mtng. ION, San Diego, CA*, pp. 127–137, 1992.

[2] M. A. Sturza, "Navigation System Integrity Monitoring Using Redundant Measurements," *Navigation, J. ION*, vol. 35, no. 4, 1988-89.

[3] R. G. Brown, "Solution of the Two-Failure GPS RAIM Problem Under Worst Case Bias Conditions: Parity Space Approach," *Navigation, J. of ION*, vol. 44, no. 4, 1997-98.

[4] J. Angus, "RAIM with Multiple Faults," *J. ION*, vol. 53, no. 4, 2006.

[5] S. Hewitson and J. Wang, "Extended Receiver Autonomous Integrity Monitoring (eRAIM) for GNSS/INS Integration," *J. of Surveying Engineering*, vol. 136, no. 1, pp. 13–22, 2010.

[6] S. Zhao, Y. Chen, H. Zhang, and J. A. Farrell, "Differential GPS Aided Inertial Navigation: A Contemplative Realtime Approach," *19th IFAC World Congress*, pp. 8959–8964, 2014.

[7] J. A. Farrell, *Aided Navigation: GPS with High Rate Sensors*. McGraw Hill, 2008.

[8] P. F. Roysdon and J. A. Farrell. (2017, Feb) "Technical Note: CRT with Hypothesis Testing". [Online]. Available: http://www.ee.ucr.edu/~farrell/TN_CRT_HT.pdf

[9] P. S. Maybeck, *Stochastic Models, Estimation, and Control*. Academic Press: Mathematics in Science and Engineering, 1979.

[10] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. I - Estimation Theory*. Prentice Hall PTR, 2013.

[11] F. Dellaert and M. Kaess, "Square Root SAM: Simultaneous Localization and Mapping via Square Root Information Smoothing," *Int. J. Rob. Res.*, vol. 25, no. 12, pp. 1181–1203, 2006.

[12] M. Kaess, A. Ranganathan, and F. Dellaert, "iSAM: Incremental Smoothing and Mapping," *IEEE T. Rob.*, vol. 24, no. 6, pp. 1365–1378, 2008.

[13] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. II - Detection Theory*. Prentice Hall PTR, 1998.

[14] R. A. Fisher, "Statistical Methods for Research Workers," *Edinburgh, UK: Oliver and Boyd*, p. 43, 1925.

[15] A. Vu and J. A. Farrell, "Feature mapping and state estimation for highly automated vehicles," *J. of Control and Decision*, vol. 2, no. 1, pp. 1–25, 2015.

[16] P. Huber, *Robust Statistics*. NY: John Wiley and Sons Inc., 1986.

[17] D. Wang, H. Lu, and M. Yang, "Robust Visual Tracking via Least Soft-threshold Squares," *IEEE T. CAS for VT*, 2015.