Title
Technical Note: CRT with Least Soft-thresholded Squares

Authors
Farrell, Jay A
Roysdon, Paul F.

Publication Date
2019-09-17

Robust GPS-INS Outlier Accommodation: A Soft-thresholded Optimal Estimator

Paul F. Roysdon **Jay A. Farrell** *

* *University of California, Riverside, CA 92521 USA (e-mail: proysdon@ece.ucr.edu, farrell@ece.ucr.edu).*

Abstract: Many highway vehicle applications require reliable, high precision navigation (error less than meter level) while using low-cost consumer-grade inertial and global navigation satellite systems (GNSS). The application environment causes numerous GNSS measurement outliers. Common implementations use a single epoch Extended Kalman Filter (EKF) combined with the Receiver Autonomous Integrity Monitoring (RAIM) for GNSS outlier detection. However, if the linearization point of the EKF is incorrect or if the number of residuals is too low, the outlier detection decisions may be incorrect. False alarms result in good information not being incorporated into the state and covariance estimates. Missed detections result in incorrect information being incorporated into the state and covariance estimates. Either case can cause subsequent incorrect decisions, possibly causing divergence, due to the state and covariance now being incorrect. This article formulates a sliding window estimator containing multiple GNSS epochs. The approach solves the full-nonlinear *Maximum A Posteriori* problem with $\boldsymbol{l_1}$-regularization. By leveraging the resulting window of residuals from the nonlinear optimization, and exploiting the $\boldsymbol{l_1}$-regularization, an improved outlier accommodation strategy is implemented. Experimental sensor data is used to demonstrate the robust sliding window least soft-thresholded squares method and its performance.

Keywords: Moving Horizon Estimation (MHE), Simultaneous Localization and Mapping (SLAM), Extended Kalman Filter (EKF), Global Navigation Satellite Systems (GNSS), Inertial Navigation System (INS), Least Soft-thresholded Squares (LSS)

1. 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. With differential corrections these systems can exhibit submeter accuracy. 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 achieve high reliability, it is critical to detect and remove outlier measurements before they degrade state estimation performance. In GNSS applications such outlier measurements can be caused by multipath, 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 (Brown (1992); Sturza (1988)). 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 (Brown (1992)). The principle of multiple outlier detection has also been well developed over several decades (Brown (1997); Angus (2006)), within a general framework named Fault Detection and Elimination (FDE). The authors of (Hewitson and Wang (2010)) included an inertial measurement unit and a Kalman filter to "extend" the RAIM capabilities. The method is 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.

This work in outlier accommodation is motivated by recent advances in computer vision. It is common in the robotics community to solve state estimation problems by the Maximum A Posteriori (MAP) formulation, but the approach is sensitive to measurements which deviate from their stochastic noise model (i.e., outliers). The authors of (Wright et al. (2009); Mei and Ling (2009); Wang et al. (2015)), using a sparse representation of candidate tracking sets for face recognition, demonstrate that l_1-regularization can exploit the sparseness of outliers in a candidate dataset with redundant measurements to achieve enhanced performance.

Data redundancy is critical to successful outlier accommodation and can be enhanced by considering all GNSS and IMU measurements within a sliding temporal window. The resulting full nonlinear MAP estimator, without outlier detection and removal, is presented in (Zhao et al. (2014)).

This article presents the formulation of the l_1-regularization for a nonlinear sliding window estimator. Because this approach allows real-time analysis of numerous fault scenar-

ios, with real-time error correction, outlier accommodation can be improved. This estimator is demonstrated using real-world data involving urban canyons and overhead foliage.

2. BACKGROUND AND NOTATION

This section introduces Global Positioning System (GPS) aided inertial navigation system (INS) background (Farrell (2008)).

2.1 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

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

where $\boldsymbol{f} : \mathbb{R}^{n_s} \times \mathbb{R}^6 \mapsto \mathbb{R}^{n_s}$ represents the kinematics, and $\mathbf{u} \in \mathbb{R}^6$ is the vector of specific forces and angular rates. The function $\boldsymbol{f}$ is accurately known (see eqn. 11.31-11.33 in Farrell (2008), and Roysdon and Farrell (2017b)). 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 $\tilde{\mathbf{u}}(\tau_i) = \mathbf{u}(\tau_i) + \mathbf{b}(\tau_i) + \boldsymbol{\omega}_u(\tau_i)$, 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 $\hat{\boldsymbol{x}}_i = \phi(\hat{\boldsymbol{x}}_{i-1}, \hat{\mathbf{u}}_{i-1})$, 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 Chapter 4.7 of (Farrell (2008)) that $\hat{\boldsymbol{x}}_k - \boldsymbol{\Phi}(\hat{\boldsymbol{x}}_{k-1}, \mathbf{U}_{k-1}) = \mathbf{w}_k$, and $\mathbf{Q}_{\mathbf{D}k} = \text{Cov}(\mathbf{w}_k)$.

2.2 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.

For $(m+1)$ satellites, $\mathbf{y}_k$ represents the double-differenced code (pseudorange) and Doppler measurement vector, as defined in Section 8.8 of (Farrell (2008)). The double-differenced measurement vector at t_k is modeled as

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

where $\boldsymbol{\eta}_{yk} = [\boldsymbol{\eta}_{\rho k}, \boldsymbol{\eta}_{dk}]$, and $\mathbf{y}_k$, $\boldsymbol{\eta}_{yk}$, $\mathbf{s}_k \in \mathbb{R}^{2m}$. The symbol $\boldsymbol{\eta}_{\rho_k} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}_\rho^2 \mathbf{I})$ represents the pseudorange measurement noise with $\boldsymbol{\sigma}_\rho = 0.1 \sim 3m$, and $\boldsymbol{\eta}_{d_k}^m \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}_d^2 \mathbf{I})$ 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 $\mathbf{R} = \text{Cov}(\boldsymbol{\eta}_{yk})$. The symbol $\mathbf{s}_k$ represents the error due to outliers. Using the state estimate, the GPS measurements 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 $\delta\mathbf{y}_k = \mathbf{y}_k - \hat{\mathbf{y}}_k$.

3. 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 (linear or nonlinear) estimator. The KF can be derived from a variety of perspectives (Maybeck (1979); Jazwinski (1970)): Maximum a Posteriori, or Mean Squared Error. When the time propagation or measurement models are nonlinear, a variety of methods (e.g., the extended Kalman filter, see Maybeck (1979)) are available to solve the sensor fusion problem over a single measurement epoch.

This section presents the MAP estimator solved over a sliding temporal window in the presence of measurement outliers. This approach can be extended to many sensors, e.g. magnetometer, barometric pressure, camera or LiDAR. Without considering outliers, this formulation has been developed extensively in the Simultaneous Localization and Mapping (SLAM) research community (Dellaert and Kaess (2006); Eustice et al. (2006); Kaess et al. (2012)). The approach developed for GNSS and IMU integration in (Zhao et al. (2014)) 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.

3.1 MAP Problem Formulation

Let $\mathbf{X}_k$ 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 epochs, $[\mathbf{y}_{k-L+1}, \ldots, \mathbf{y}_k]$. The window will slide one epoch upon arrival of each new GPS measurement, always keeping L epochs in the window. 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 Kay (2013)):

$$\hat{\mathbf{X}} = \underset{\mathbf{X},\mathbf{S}}{\text{argmax}} \left\{ p(\mathbf{X}, \mathbf{U}, \mathbf{Y}, \mathbf{S}) \right\},$$

where within the time window $\mathbf{U} = \{\mathbf{U}_i \,|\, i \in [k-L, k-1]\}$, $\mathbf{U}_i = \{\tilde{\mathbf{u}}(\tau_\ell) \text{ for } \tau_\ell \in [t_i, t_i + 1]\}$, $\mathbf{Y} = \{\mathbf{y}_j \mid j \in [k - L + 1, k]\}$, and $\mathbf{S} = \{\mathbf{s}_j \mid j \in [k-L+1, k]\}$. The sets $\mathbf{U}$, $\mathbf{Y}$, and

$\mathbf{S}$ will be treated as concatenated vectors in the numerical approach.

The GPS-INS joint probability can be decomposed as

$$p(\mathbf{X},\mathbf{U},\mathbf{Y},\mathbf{S}) = p(\mathbf{S})p(\boldsymbol{x}_{k-L})\prod_{l=k-L}^{k-1} p(\boldsymbol{x}_{l+1}|\boldsymbol{x}_l,\mathbf{U}_l)\prod_{j=k-L+1}^{k} p(\mathbf{y}_j|\boldsymbol{x}_j,\mathbf{s}_j), \quad (4)$$

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, $p(\mathbf{y}_j|\boldsymbol{x}_j,\mathbf{s}_j)$ is the distribution of the GPS measurement noise, $p(\mathbf{S})$ is the distribution which corresponds to the outlier measurements. The derivation of eqn. (4) is in (Roysdon and Farrell (2017b)).

3.2 Solution Formulation

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

The outlier vectors $\mathbf{s}_j$ are assumed to have zero mean Laplacian distributions (Mei and Ling (2009); Wang et al. (2015)):

$$p_{\mathbf{s}_j}(\mathbf{s}) = \frac{1}{2}\exp\left[-\left\|\frac{\mathbf{s}_\rho}{\lambda_\rho}\right\|_1\right]\exp\left[-\left\|\frac{\mathbf{s}_d}{\lambda_d}\right\|_1\right],$$

where $\|\cdot\|_1$ is the l_1-norm, and $\mathbf{s} = [\frac{\mathbf{s}_\rho^\intercal}{\lambda_\rho}, \frac{\mathbf{s}_d^\intercal}{\lambda_d}]^\intercal$. The values λ_ρ and λ_d are selected according to the psuedorange and Doppler measurement distributions, respectively.

Finding $\mathbf{X}$ that maximizes eqn. (4) is identical to minimizing the negative of its natural logarithm. This yields the equivalent nonlinear cost function:

$$\begin{aligned}\mathcal{C}(\mathbf{X},\mathbf{S}) = {} & \|\hat{\boldsymbol{x}}_{k-L} - \boldsymbol{x}(t_{k-L})\|^2_{\mathbf{P}_{(k-L)}} + \sum_{j=k-L+1}^{k}\|\mathbf{s}(t_j)\|_1 \\ & + \sum_{l=k-L}^{k-1}\|\boldsymbol{\Phi}\big(\boldsymbol{x}(t_l),\mathbf{U}_l\big) - \boldsymbol{x}(t_{l+1})\|^2_{\mathbf{Q_D}} \\ & + \sum_{j=k-L+1}^{k}\|\mathbf{y}(t_j) - \mathbf{h}_j\big(\boldsymbol{x}(t_j)\big) - \mathbf{s}(t_j)\|^2_{\mathbf{R}}. \quad (5)\end{aligned}$$

The cost function can be normalized using Cholesky Decomposition. For the positive definite matrix $\mathbf{W}$, define $\boldsymbol{\Sigma}_{\mathbf{W}}$, such that $\mathbf{W}^{-1} = \boldsymbol{\Sigma}_{\mathbf{W}}^\intercal\boldsymbol{\Sigma}_{\mathbf{W}}$. Then, for $\mathbf{b} \triangleq \boldsymbol{\Sigma}_{\mathbf{W}}\mathbf{v}$, $\|\mathbf{v}\|_{\mathbf{W}} = \|\mathbf{b}\|_2$. The minimization problem of eqn. (5) reduces to the optimization of

$$\|\mathbf{z} - \mathbf{g}(\mathbf{X}) - \mathbf{S}\|_2^2 + \|\mathbf{S}\|_1, \quad (6)$$

which will be minimized iteratively. In eqn. (6), the symbol

$$\mathbf{z} = \boldsymbol{\Sigma}_{\mathbf{W}}[\hat{\boldsymbol{x}}_{k-L}^\intercal, \mathbf{0}^\intercal, \mathbf{Y}^\intercal]^\intercal$$

represents the terms that are known at each iteration and

$$\mathbf{g}(\mathbf{X}) = \boldsymbol{\Sigma}_{\mathbf{W}}[\boldsymbol{x}(t_{k-L})^\intercal, \boldsymbol{\Phi}(\mathbf{X},\mathbf{U})^\intercal - \mathbf{X}^\intercal, \mathbf{h}(\mathbf{X})^\intercal]^\intercal$$

represents the terms that are computed based on $\mathbf{X}$. In this expression, $\boldsymbol{\Phi}(\mathbf{X},\mathbf{U}) - \mathbf{X}$ represents the concatenation of the vector terms $\boldsymbol{\phi}\big(\boldsymbol{x}(t_l),\mathbf{U}_l\big) - \boldsymbol{x}(t_{l+1})$, and $\mathbf{h}(\mathbf{X})$ represents the concatenation of the vector $\mathbf{h}_j(\boldsymbol{x}(t_j))$ terms.

4. APPLICATION OF THE SOFT-THRESHOLDING OPERATOR TO A NONLINEAR PROBLEM

The authors of (Wang et al. (2015); Mei and Ling (2009)) provide the derivation of the Soft-Thresholding Operator for the linear case. Equation (6) has the nonlinear form

$$\mathcal{C}(\mathbf{X},\mathbf{S}) = -\frac{1}{2}\|\mathbf{z} - \mathbf{g}(\mathbf{X}) - \mathbf{S}\|_2^2 + \|\mathbf{S}\|_1$$

which can be optimized by methods similar to those of (Wang et al. (2015); Mei and Ling (2009)) as derived below. Because the l_1-norm is not differentiable at its maximum (minimum), the optimal solution to $\mathcal{C}(\mathbf{X},\mathbf{S})$ is solved in two parts.

4.1 Part 1: Estimate $\mathbf{X}$*, given* $\hat{\mathbf{S}}$

Given $\hat{\mathbf{S}}$ and an initial guess $\hat{\mathbf{X}}$, the optimal value of $\mathbf{X}$ is

$$\begin{aligned}\mathbf{X}^* &= \underset{\mathbf{X}}{arg\,max}\ \mathcal{C}(\mathbf{X},\hat{\mathbf{S}}) \\ &= \underset{\mathbf{X}}{arg\,max}\left\{-\frac{1}{2}\|\mathbf{z} - \mathbf{g}(\mathbf{X}) - \hat{\mathbf{S}}\|_2^2 + \|\hat{\mathbf{S}}\|_1\right\} \\ &= \underset{\mathbf{X}}{arg\,max}\left\{-\frac{1}{2}\|(\mathbf{z} - \hat{\mathbf{S}}) - \mathbf{g}(\mathbf{X})\|_2^2\right\}. \quad (7)\end{aligned}$$

This optimization can be solved iteratively using a Taylor series expansion to approximate $\mathbf{g}(\mathbf{X})$,

$$\mathbf{g}(\mathbf{X}) = \mathbf{g}(\hat{\mathbf{X}}) + \mathbf{G}\delta\mathbf{X} + h.o.t's, \quad (8)$$

where the Jacobian $\mathbf{G} = \frac{\partial\mathbf{g}(\mathbf{X})}{\partial\mathbf{X}}\Big|_{\mathbf{X}=\hat{\mathbf{X}}}$, and $\delta\mathbf{X} = \mathbf{X} - \hat{\mathbf{X}}$ where $\delta\mathbf{X} \in \mathbb{R}^{n_e(L+1)}$. 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 (Farrell (2008)). The fact that $n_s = 16$ and $n_e = 15$ is discussed in (Roysdon and Farrell (2017b)). Inserting eqn. (8) into eqn. (7), ignoring the higher order terms ($h.o.t's$), yields

$$\begin{aligned}\delta\mathbf{X}^* &= \underset{\delta\mathbf{X}}{arg\,min}\,\|(\mathbf{z} - \hat{\mathbf{S}}) - (\mathbf{g}(\hat{\mathbf{X}}) + \mathbf{G}\delta\mathbf{X})\|_2^2 \\ &= \underset{\delta\mathbf{X}}{arg\,min}\,\|\mathbf{b} - \mathbf{G}\delta\mathbf{X}\|_2^2\end{aligned}$$

where

$$\mathbf{b} = \mathbf{z} - \hat{\mathbf{S}} - \mathbf{g}(\hat{\mathbf{X}}). \quad (9)$$

Then, the optimal $\delta\hat{\mathbf{X}}$ is the solution of the normal equation:

$$(\mathbf{G}^\intercal\mathbf{G})\delta\hat{\mathbf{X}} = \mathbf{G}^\intercal\mathbf{b}. \quad (10)$$

The optimal update to the trajectory

$$\hat{\mathbf{X}}^+ = \hat{\mathbf{X}}^- + \delta\mathbf{X}^* \quad (11)$$

is obtained by iterating eqns. (8), (9), and (10), to convergence for some user defined stopping conditions. The symbol $(-)$ denotes the prior estimate, whereas $(+)$ is the updated estimate.

4.2 Part 2: Estimate $\mathbf{S}$*, given* $\hat{\mathbf{X}}$

Given an estimate $\hat{\mathbf{X}}$, the optimal value of $\mathbf{S}$ is

$$
\begin{aligned}
\mathbf{S}^* &= \underset{\mathbf{S}}{arg\,max}\ \mathcal{C}(\hat{\mathbf{X}}, \mathbf{S}) \\
&= \underset{\mathbf{S}}{arg\,max} \left\{ -\frac{1}{2}\|\mathbf{z} - \mathbf{g}(\hat{\mathbf{X}}) - \mathbf{S}\|_2^2 + \|\mathbf{S}\|_1 \right\}. \\
&= \underset{\mathbf{S}}{arg\,max} \left\{ -\frac{1}{2}\|\mathbf{c} - \mathbf{S}\|_2^2 + \|\mathbf{S}\|_1 \right\}, \\
&= \underset{\mathbf{S}}{arg\,min} \sum_i \left[(\mathbf{c}_i - \mathbf{S}_i)^2 + |\mathbf{S}_i| \right], \qquad (12)
\end{aligned}
$$

where $\mathbf{c} = \mathbf{z} - \mathbf{g}(\hat{\mathbf{X}})$ is a constant.

Each term in the summation in eqn. (12) only depends on $\mathbf{c}_i$ and $\mathbf{s}_i$, thus each term can be optimized independently

$$
\hat{\mathbf{s}}_i = \underset{\mathbf{s}_i}{arg\,min} \left\{ (\mathbf{c}_i - \mathbf{s}_i)^2 + |\mathbf{s}_i| \right\}. \qquad (13)
$$

The closed-form solution of the optimization problem in eqn. (13) is the soft-thresholding operation

$$
\hat{\mathbf{s}}_i = sign(\mathbf{c}_i) max(|\mathbf{c}_i| - 1, 0), . \qquad (14)
$$

4.3 Summary

Implementation of Sections 4.1 and 4.2 is as summarized in Algorithm 1.

Algorithm 1 Least Soft-threshold Squares Regression

1: Given initial $\hat{\mathbf{X}}$ and $\hat{\mathbf{S}}$. Initialize $i = 0$.
2: **while** $(((\|\delta\hat{\mathbf{X}}\|_2$ & $\|\delta\hat{\mathbf{S}}\|_2) > 1 \times 10^{-3})$ & $(i < 20))$
3: $\quad i = i + 1$
4: $\quad$ Compute $\delta\hat{\mathbf{X}}$ and $\hat{\mathbf{X}}^+ = \hat{\mathbf{X}}^- + \delta\hat{\mathbf{X}}$.
5: $\quad$ **if** $(\|\delta\hat{\mathbf{X}}\|_2 < 1 \times 10^{-3})$
6: $\quad\quad$ Compute $\hat{\mathbf{S}}$ using eqn. (14).
7: $\quad\quad$ Compute $\delta\hat{\mathbf{X}}$ and $\hat{\mathbf{X}}^+ = \hat{\mathbf{X}}^- + \delta\hat{\mathbf{X}}$.
8: $\quad$ **end**
9: **end**

Note that in Section 4.1 the value of the cost function $\mathcal{C}(\mathbf{X}, \mathbf{S})$ is decreased by changing only $\mathbf{X}$, even if multiple nonlinear least squares iterations occur. In Section 4.2 the value of the cost function $\mathcal{C}(\mathbf{X}, \mathbf{S})$ is decreased by changing only $\mathbf{S}$, only a single iteration is required, and $\delta\hat{\mathbf{S}} \triangleq \hat{\mathbf{S}}^+ - \hat{\mathbf{S}}^-$. Therefore, at each iteration, the cost function is decreased and is bounded below. Ultimately at least a local minimum of the cost function will be attained.

A line search is implemented in the direction of $\delta\mathbf{X}^*$ from eqn. (10) to determine the magnitude of the update step in eqn. (11).

The normal equation of eqn. (10) can be compactly expressed as $\mathbf{\Lambda}\delta\mathbf{X} = \boldsymbol{\xi}$ where $\mathbf{\Lambda} = \mathbf{G}^\intercal\mathbf{G}$ is the information matrix, $\boldsymbol{\xi} = \mathbf{G}^\intercal\mathbf{b}$ is the information vector. The matrix $\mathbf{G}$ is sparse; therefore, eqn. (10) can be solved efficiently by employing a sparse matrix library (Dellaert and Kaess (2006); Kaess et al. (2012)). The computational complexity of the algorithm is discussed in (Roysdon and Farrell (2017b)).

The combined operation of Sections 4.1 and 4.2 are equivalent to the Huber Loss function (Huber (1981); Wang et al. (2015)).

5. EXPERIMENTAL RESULTS

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) at $250Hz$, and L1 GPS data (Ublox 6T) at $1Hz$. Differential corrections were obtained from the UCR base-station NTRIP caster (`ntrip.engr.ucr.edu`) in real-time via cellular connection. 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 comparison of the 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 in real-time, using only the data and prior as would be applicable for each approach. The navigation algorithms being compared only use L1 GPS with differential corrections, and IMU data.

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 (Vu and Farrell (2015)).

Due to limited space, only 3D position performance is discussed herein. Velocity and attitude results are provided in (Roysdon and Farrell (2017b)), and are similar to position performance.

5.1 CRT LSS: Fixed Window Example

This subsection discusses the optimization within one fixed CRT window with $L = 20$. The results are shown in Table 1 and Fig. 1.

Table 1 presents the results of each iteration of Alg. 1, which is initialized with $\|\hat{\mathbf{S}}\|_1 = 0$. The $\hat{\mathbf{X}}$ optimization (Alg. 1 row 4) iterates four times to reach $\hat{\mathbf{X}}$ convergence with $\hat{\mathbf{S}} = \mathbf{0}$. After iteration four, $\|\delta\hat{\mathbf{X}}\|_2 < 1 \times 10^{-3}$ and $\hat{\mathbf{S}}$ is computed using eqn. (14). This increases $\|\hat{\mathbf{S}}\|_1$, but decreases both $\|\mathbf{b}\|_2$ and the cost function at iteration

Table 1. CRT LSS Iterations.
Note $\mathbf{b}$, $\hat{\mathbf{S}}$ and $\delta\hat{\mathbf{S}}$ are dimensionless, while each 3-element subvector of $\delta\hat{\mathbf{X}}$ has distinct units (see Section 4.1).

i	$\|\mathbf{b}\|_2$	$\|\hat{\mathbf{S}}\|_1$	$\|\delta\hat{\mathbf{X}}\|_2$	$\|\delta\hat{\mathbf{S}}\|_1$
1	1814564.689	0.000	117.632	0.000
2	104412.888	0.000	1.274	0.000
3	3567.236	0.000	0.064	0.000
4	11.859	0.000	0.001	0.000
4	10.851	2.808	0.077	2.808
5	10.823	2.808	0.003	0.000
6	10.493	2.808	0.000	0.000
6	10.473	2.874	0.005	0.065
7	10.473	2.874	0.004	0.000
7	10.471	2.879	0.000	0.005

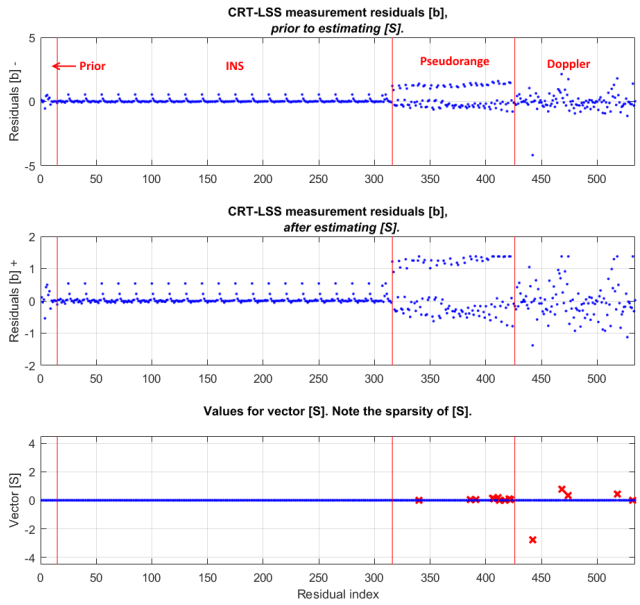

Fig. 1. The CRT-LSS results for residuals $\mathbf{b}$ before and after estimating $\hat{\mathbf{S}}$, and the values of $\hat{\mathbf{S}}$.

five. Then $\delta\hat{\mathbf{X}}$ and $\hat{\mathbf{X}}^+ = \hat{\mathbf{X}}^- + \delta\hat{\mathbf{X}}$ are computed. At this point Algorithm 1 returns to row 2, and the $\hat{\mathbf{X}}$ optimization is repeated until both $\|\delta\hat{\mathbf{X}}\|_2 < 10^{-3}$ and $\|\delta\hat{\mathbf{S}}\|_2 < 10^{-3}$ (the user defined convergence thresholds). Only three additional iterations are required.

Fig. 1 shows the residual $\hat{\mathbf{b}}$ before and after compensation by $\hat{\mathbf{S}}$ during the final iteration. The vectors $\hat{\mathbf{b}}$ and $\hat{\mathbf{S}}$ are structured as follows and separated by red-line boundaries. The prior is the first 15 elements. The INS data are the next 300 elements (20 blocks of 15 elements). The GPS L1 pseudorange data are elements 316-425, and Doppler data are elements 426-534. Under the assumed conditions (white Guassian process and measurement noise with perfect tuning), the vector $\hat{\mathbf{b}}$ would be Gaussian with unit variance. Because this is clearly not the case for the INS and pseudorange portions of $\hat{\mathbf{b}}$, discussion is warranted. The large elements of the INS residuals correspond to the accelerometer and gyro bias errors. Those errors in the INS error model are very slowly time-varying, such that over this $L = 20$ second window, these errors are essentially constant, which cause the 20 peaks in the INS residual. The pseudorange residuals are affected by time-correlated multipath errors. Two satellites are significantly worse than the others. The time-correlation is evident. This correlation could be addressed by state augmentation, which is an interesting topic for future research. The Doppler and prior residuals look reasonable. The vector $\hat{\mathbf{S}}$, shown in the bottom plot, is sparse. As defined in eqn. (4), the non-zero indices in $\hat{\mathbf{S}}$, denoted by a red "x", correspond to the GPS residuals for both pseudorange and Doppler. Comparing the first two graphs, while $\mathbf{b}^-$ has several large elements, the corrections defined by $\hat{\mathbf{S}}$ result in $\mathbf{b}^+$ being magnitude constrained by λ. Pseudorange and Doppler magnitude constrained residuals do not necessarily coincide due to different modalities.

5.2 CRT LSS Example: Sliding Window

Fig. 2 shows the cumulative distribution function (CDF) of the norm of the position error $\|\hat{\mathbf{p}}_k - \mathbf{p}_k\|$ for various estimation algorithms. Position error is computed relative to the ground truth trajectory $\mathbf{p}_k$ discussed earlier. The value of $\hat{\mathbf{p}}_k$ is the *a posteriori* result after convergence of Alg. 1 at the first time when the k-th epoch enters the sliding window. In the legend, EKF and CRT are already defined. IEKF represents the iterated EKF, which 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, curves are included for various values of the window length L.

The CDF shows that the percentage of occurrences where the EKF position error is less than $0.1m$, is roughly 15%. Approximately 90% of the state vectors, as estimated by the EKF, have position error less than $1.0m$. Fig. 2 shows that accuracy improves from the EKF to the IEKF to the CRT. Also, CRT performance (generally) improves with the window length L. For example, in the CRT approach with $L > 5$, 97% of the position errors are less than $1.0m$. The CRT algorithms with $L > 10$ each achieve $1.0m$ position accuracy on 100% of the trajectory. Alternatively, the EKF and IEKF CDF plots do not reach 100% until the position accuracy is over $3.0m$.

The improved performance demonstrated in Fig. 2 is attributed to solving the full nonlinear optimization over a longer window with outlier accommodation. The longer window enhances the redundancy and allows reconsideration of fault decisions, as long as the measurement data is within the sliding window. This enhanced ability to accommodate outliers to achieve reliable performance is one of the major motivations of the CRT approach.

5.3 Comparison: CRT LSS vs. Hypotheses Testing

An alternative to the LSS approach developed and evaluated herein are hypotheses testing approaches (Ferguson (1961)). CRT using the hypotheses testing approach is

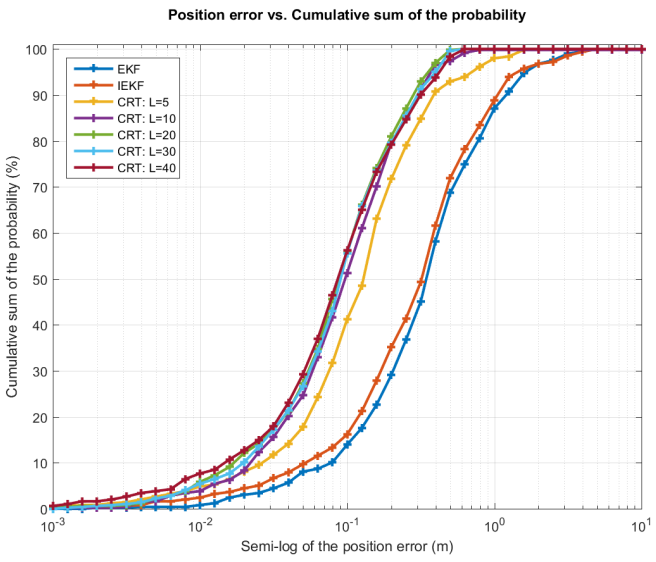

Fig. 2. Position Error Cumulative Distributions.

developed and evaluated in (Roysdon and Farrell (2017a)). Each approach builds on the residual vector $\mathbf{z} - \mathbf{g}(\hat{\mathbf{X}})$ (see eqn. (9)).

The hypotheses testing approach first evaluates the null hypothesis that no outliers exist. If the null hypothesis fails, then the most likely alternative hypothesis is selected from a candidate set of alternative hypotheses. The approach can succeed when the set of candidate hypotheses includes the actual outlier scenario. To evaluate any given 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

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

ways that any number of satellite measurements could fail in any combination at one epoch (see Section 3 of Angus (2006)). 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 to hypotheses testing are required, such as iteratively removing the row with the largest residual and reoptimizing until all residuals pass a threshold test.

The soft-thresholding approach of (Wang et al. (2015); Mei and Ling (2009)) as adapted to the CRT problem in Section 4 does not remove any rows. Instead, it automatically detects which residuals should be soft-thresholded in a manner that guarantees convergence of the MAP optimization for each time window accounting for outliers.

6. CONCLUSION

This article derived and demonstrated a MAP estimator with automated l_1-regularization outlier accommodation using a sliding window smoothing approach. 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 accommodation.

This CRT-LSS framework, through nonlinear optimization, achieves optimal state estimation without linearization assumptions. Real-time implementation is feasible on standard computers. The CRT-LSS performance is demonstrated through direct comparisons of both the accuracy and outlier accommodation abilities of various algorithms using experimental data from a challenging environment. Such methods have utility in autonomous vehicle applications where both accuracy and reliability are critical.

Related areas of interest for future research include accommodation of time correlated errors (e.g., multipath) either by augmented states or non-diagonal covariance matrices (e.g. $\mathbf{R}$), adaptation of the LSS threshold λ to minimize the risk of missing outliers while guaranteeing a desired level of expected accuracy, and adaptation of the window length L again trading off risk and performance.

REFERENCES

Angus, J. (2006). RAIM with Multiple Faults. *J. ION*, 53(4).

Brown, R.G. (1992). A Baseline RAIM Scheme and a Note on the Equivalence of Three RAIM Methods. *ION NTM*, 127–137.

Brown, R.G. (1997). Solution of the Two-Failure GPS RAIM Problem Under Worst Case Bias Conditions: Parity Space Approach. *Navigation, J. ION*, 44(4).

Dellaert, F. and Kaess, M. (2006). Square Root SAM: Simultaneous Localization and Mapping via Square Root Information Smoothing. *Int. J. Rob. Res.*, 25(12), 1181–1203.

Eustice, R., Singh, H., and Leonard, J. (2006). Exactly Sparse Delayed-State Filters for View-Based SLAM. *IEEE T. Robotics,*, 22(6), 1552–3098.

Farrell, J.A. (2008). *Aided Navigation: GPS with High Rate Sensors*. McGraw Hill.

Ferguson, T. (1961). On the Rejection of Outliers. *Berkeley Symposium on Mathematical Statistics and Probability*, 1, 253–287.

Hewitson, S. and Wang, J. (2010). Extended Receiver Autonomous Integrity Monitoring (eRAIM) for GNSS/INS Integration. *J. of Surveying Engineering*, 136(1), 13–22.

Huber, P. (1981). *Robust Statistics*. New York: John Wiley and Sons Inc.

Jazwinski, A.H. (1970). *Stochastic Processes and Filtering*. Academic Press: Mathematics in Sci. and Eng.

Kaess, M., Johannsson, H., Roberts, R., Ila, V., Leonard, J., and Dellaert, F. (2012). iSAM2: Incremental smoothing and mapping using the Bayes tree. *Intl. J. of Robotics Research*, 31(2), 216–235.

Kay, S.M. (2013). *Fundamentals of Statistical Signal Processing, Vol. I - Estimation Theory*. Prentice Hall.

Maybeck, P.S. (1979). *Stochastic Models, Estimation, and Control*. Academic Press: Mathematics in Sci. and Eng.

Mei, X. and Ling, H. (2009). Robust Visual Tracking using L-1 Minimization. *ICCV*.

Roysdon, P.F. and Farrell, J.A. (2017a). GPS-INS Outlier Detection and Elimination using a Sliding Window Filter. *IEEE, In Proc. of American Control Conf.*

Roysdon, P.F. and Farrell, J.A. (2017b). "Technical Note: CRT with Least Soft-thresholded Squares". URL `https://escholarship.org/uc/item/5r88s8dh`.

Sturza, M.A. (1988). Navigation System Integrity Monitoring Using Redundant Measurements. *Navigation, J. of ION*, 35(4).

Vu, A. and Farrell, J.A. (2015). Feature mapping and state estimation for highly automated vehicles. *J. of Control and Decision*, 2(1), 1–25.

Wang, D., Lu, H., and Yang, M. (2015). Robust Visual Tracking via Least Soft-threshold Squares. *IEEE T. on Circuits and Systems for Video Technology*.

Wright, J., Yang, A., Ganesh, A., Sastry, S., and Ma, Y. (2009). Robust Face Recognition via Sparse Representation. *IEEE T. PAMI*, 31(2).

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