Technical Note: CRT with Hypothesis Testing

Paul F. Roysdon[†] Jay A. Farrell[‡] [*]

July 31, 2017

Due to slight variations in nomenclature, this Technical Note contains the following independent sections:

1. INS Temporal Propagation,
2. INS State Error Model,
3. INS Noise Propagation,
4. Computational Cost of the CRT Estimator,
5. CRT-HT Outlier Accommodation Derivations & Proofs,
6. CRT-HT Position, Velocity & Attitude Results.

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

Technical Note: INS Temporal Propagation

Paul F. Roysdon[†] Jay A. Farrell[‡]

***Abstract*— Due to space limitations in [1] and [2], this Technical Note is supplied to explain an approach for the time propagation of the state vector estimate $\hat{x}$ by the function $\boldsymbol{f}(\boldsymbol{x}(t), \mathbf{u}(t))$ in an inertial navigation system (INS).**

I. Introduction

While many reference frames may be used, the Earth-centered Earth-fixed (ECEF) reference frame is a convenient reference frame for GPS-aided INS, simply because the computation of satellite navigation solutions are resolved in the ECEF reference frame.

In this article we define the e-frame as the ECEF-frame (or Earth-frame), and the i-frame as the Earth-Centered-Inertial (ECI) frame (or inertial-frame). The direction cosine matrix (DCM), or rotation matrix, from body-frame to Earth-frame is $\mathbf{R}_b^e$. The time derivative of position in the body-frame b with respect to the Earth-frame e resolved about the Earth-frame is $\dot{\mathbf{r}}_{eb}^e$, and similarly the time derivative of velocity is $\dot{\mathbf{v}}_{eb}^e$. The time derivative of the rotation matrix from body-frame to Earth-frame is $\dot{\mathbf{R}}_b^e$.

Expanding eqn. (1) in both [1] and [2], the INS kinematic equations defining $\boldsymbol{f}(\boldsymbol{x}(t), \mathbf{u}(t))$ in the ECEF frame (see Section 11.2.2 of [3]) are

$$\dot{\mathbf{r}}_{eb}^e = \mathbf{v}_{eb}^e \tag{1}$$

$$\dot{\mathbf{v}}_{eb}^e = \mathbf{R}_b^e \mathbf{f}_{ib}^b + \mathbf{g}_b^e - 2\boldsymbol{\Omega}_{ie}^e \mathbf{v}_{eb}^e \tag{2}$$

$$\dot{\mathbf{R}}_b^e = \mathbf{R}_b^e (\boldsymbol{\Omega}_{ib}^e - \boldsymbol{\Omega}_{ie}^e). \tag{3}$$

The inputs $\mathbf{u}$, are specific force $\mathbf{f}_{ib}^b$ and angular rate $[\boldsymbol{\omega}_{ib}^b\times] = \boldsymbol{\Omega}_{ib}^b$, with respect to the inertial-frame i. The local gravity vector is $\mathbf{g}^e$, and the Earth-rotation rate is $[\boldsymbol{\omega}_{ie}^e\times] = \boldsymbol{\Omega}_{ie}^e$. The notation $[\mathbf{a}\times]$ represents the skew-symmetric matrix corresponding to the vector $\mathbf{a}$.

Fig. 1 is a block diagram showing how the angular-rate and specific-force measurements, of an Inertial Measurement Unit (IMU), are used to update the Earth-referenced attitude, velocity, and position states at t_i.

II. INS Time Propagation Equations

A. Attitude Update

Using eqn. (2.54) in [3], define the time derivative of the rotation matrix of the ECEF-frame with respect to body-frame, represented in the body-frame,

$$\dot{\mathbf{R}}_b^e = \mathbf{R}_b^e \boldsymbol{\Omega}_{eb}^b, \tag{4}$$

where $\boldsymbol{\Omega}_{eb}^b = [\boldsymbol{\omega}_{eb}^b\times]$, and the rotation from body to ECEF-frame is $\mathbf{R}_b^e$.

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

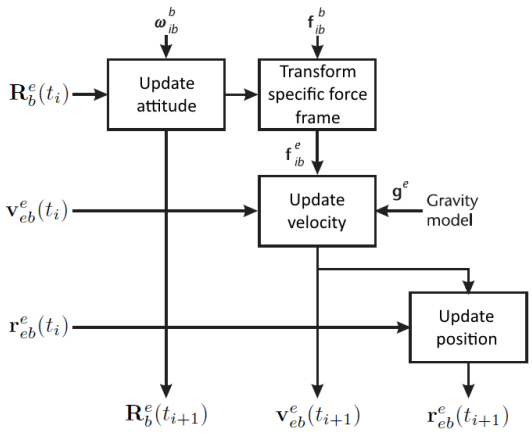

Fig. 1. Block diagram of the ECEF referenced INS equations.

Let the attitude increment be defined as $\boldsymbol{\alpha}_{ib}^b = \boldsymbol{\omega}_{ib}^b \tau_i$, for the IMU measurement interval τ_i. Using the small-angle approximation (see eqns. (2.50)-(2.54) of [3]), truncating the expansions to first-order, and assuming the IMU angular-rate measurement is constant over the integration interval, eqn. (4) is represented as

$$\mathbf{R}_b^e(t_{i+1}) = \mathbf{R}_b^e(t_i)\mathbf{R}_{b(t)}^{b(t+\tau_i)} - \boldsymbol{\Omega}_{ie}^b \mathbf{R}_b^e(t_i)\tau_i + \mathbf{R}_b^e(t_i). \tag{5}$$

The variable $\mathbf{R}_{b(t)}^{b(t+\tau_i)}$ represents the small angle rotation of $\boldsymbol{\Omega}_{ib}^b$ over the interval $(t+\tau_i)$, and $\mathbf{R}_b^e(t_i)$ and $\mathbf{R}_b^e(t_{i+1})$ are the prior and updated rotation matrix, respectively.

By eqn. (2.64) of [3], the attitude update matrix in terms of the attitude increment is

$$\mathbf{R}_{b(t)}^{b(t+\tau_i)} = \mathbf{R}_{b-}^{b+} = \exp[\boldsymbol{\alpha}_{ib}^b\times], \tag{6}$$

where the power-series expansion of the matrix exponential is

$$\exp[\boldsymbol{\alpha}_{ib}^b\times] = \sum_{r=0}^{\infty} \frac{[\boldsymbol{\alpha}_{ib}^b\times]^r}{r!}. \tag{7}$$

As shown in eqns. (2.66) and (2.67) of [3], the odd and even powers of the attitude increment skew-symmetric matrix are,

$$[\boldsymbol{\alpha}_{ib}^b\times]^{2r+1} = (-1)^r ||\boldsymbol{\alpha}_{ib}^b||^r [\boldsymbol{\alpha}_{ib}^b\times] \tag{8}$$

$$[\boldsymbol{\alpha}_{ib}^b\times]^{2r} = (-1)^r ||\boldsymbol{\alpha}_{ib}^b||^{2r} [\boldsymbol{\alpha}_{ib}^b\times]^2, \tag{9}$$

where $r = 1, 2, 3, \cdots$.

Expanding eqn. (6) using eqns. (7)-(9),

$$\mathbf{R}_{b-}^{b+} = \mathbf{I}_3 + \left(\sum_{r=0}^{\infty}(-1)^r \frac{||\boldsymbol{\alpha}_{ib}^b||^{2r}}{(2r+1)!}\right)[\boldsymbol{\alpha}_{ib}^b\times] + \left(\sum_{r=0}^{\infty}(-1)^r \frac{||\boldsymbol{\alpha}_{ib}^b||^{2r}}{(2r+2)!}\right)[\boldsymbol{\alpha}_{ib}^b\times]^2. \quad (10)$$

Eqn. (10) is equivalent to (see eqn. (2.69) of [3], and [5]–[7]),

$$\mathbf{R}_{b-}^{b+} = \mathbf{I}_3 + \frac{\sin(||\boldsymbol{\alpha}_{ib}^b||)}{||\boldsymbol{\alpha}_{ib}^b||}[\boldsymbol{\alpha}_{ib}^b\times] + \frac{1-\cos(||\boldsymbol{\alpha}_{ib}^b||)}{||\boldsymbol{\alpha}_{ib}^b||^2}[\boldsymbol{\alpha}_{ib}^b\times]^2. \quad (11)$$

The fourth-order approximation of eqn. (10) is

$$\mathbf{R}_{b-}^{b+} = \mathbf{I}_3 + \left(1 - \frac{||\boldsymbol{\alpha}_{ib}^b||^2}{6}\right)[\boldsymbol{\alpha}_{ib}^b\times] + \left(\frac{1}{2} - \frac{||\boldsymbol{\alpha}_{ib}^b||^2}{24}\right)[\boldsymbol{\alpha}_{ib}^b\times]^2. \quad (12)$$

Therefore eqn. (5) becomes

$$\mathbf{R}_b^e(t_{i+1}) = \mathbf{R}_b^e(t_i)\mathbf{R}_{b-}^{b+} - \boldsymbol{\Omega}_{ie}^b\mathbf{R}_b^e(t_i)\tau_i + \mathbf{R}_b^e(t_i), \quad (13)$$

using the definition of $\mathbf{R}_{b-}^{b+}$ from either eqn. (11) or eqn. (12).

B. Specific-Force Frame Transformation

The frame transformation of specific-force takes the form

$$\mathbf{f}_{ib}^e = \mathbf{R}_b^e\mathbf{f}_{ib}^b. \quad (14)$$

Incorporating the updated rotation matrix $\mathbf{R}_b^e(t_{i+1})$ from eqn. (13), the specific force transformation is

$$\mathbf{f}_{ib}^e = \mathbf{R}_b^e(t_{i+1})\mathbf{f}_{ib}^b. \quad (15)$$

C. Velocity Update

Neglecting acceleration in the ECEF frame, assume the ECI frame instantaneously coincides with the ECEF frame, such that $\ddot{\mathbf{r}}_{ib}^e = \ddot{\mathbf{r}}_{eb}^e$, $\dot{\mathbf{r}}_{ib}^e = \dot{\mathbf{r}}_{eb}^e$, and $\mathbf{r}_{ib}^e = \mathbf{r}_{eb}^e$. The velocity update in the ECEF frame is

$$\mathbf{v}_{eb}^e(t_{i+1}) = \mathbf{v}_{eb}^e(t_i) + \left(\mathbf{f}_{ib}^e + \mathbf{g}_b^e(\mathbf{r}_{eb}^e(t_i)) - 2\boldsymbol{\Omega}_{ie}^e\mathbf{v}_{eb}^e(t_i)\right)\tau_i \quad (16)$$

where $\mathbf{r}_{eb}^e(t_i)$ and $\mathbf{v}_{eb}^e(t_i)$ are the prior position and velocity, respectively, and $\mathbf{v}_{eb}^e(t_{i+1})$ is the updated velocity for the IMU interval τ_i.

D. Position Update

Using eqn. (1), the position update is derived as follows:

$$\mathbf{r}_{eb}^e(t_{i+1}) = \mathbf{r}_{eb}^e(t_i) + (\mathbf{v}_{eb}^e(t_i) + \mathbf{v}_{eb}^e(t_{i+1}))\frac{\tau_i}{2}, \quad (17)$$

$$= \mathbf{r}_{eb}^e(t_i) + \mathbf{v}_{eb}^e(t_i)\tau_i + \left(\mathbf{f}_{ib}^e + \mathbf{g}^e(\mathbf{r}_{eb}^e(t_i)) - 2\boldsymbol{\Omega}_{ie}^e\mathbf{v}_{eb}^e(t_i)\right)\frac{\tau_i^2}{2}. \quad (18)$$

where $\mathbf{r}_{eb}^e(t_i)$ and $\mathbf{r}_{eb}^e(t_{i+1})$ are the prior and updated positions, respectively.

E. Gravity Model

A precise gravity model [8] formulated in the ECEF frame, is defined as

$$\boldsymbol{\gamma}_{ib}^e = -\frac{\mu}{||\mathbf{r}_{eb}^e||^3} \times \left\{\mathbf{r}_{eb}^e + \frac{3}{2}J_2\frac{R_0^2}{||\mathbf{r}_{eb}^e||^2}\begin{bmatrix}(1-5(\mathbf{r}_{eb,z}^e)^2/||\mathbf{r}_{eb}^e||)\mathbf{r}_{eb,x}^e \\ (1-5(\mathbf{r}_{eb,z}^e)^2/||\mathbf{r}_{eb}^e||)\mathbf{r}_{eb,y}^e \\ (3-5(\mathbf{r}_{eb,z}^e)^2/||\mathbf{r}_{eb}^e||)\mathbf{r}_{eb,z}^e\end{bmatrix}\right\}, \quad (19)$$

where the Earth's second gravitational constant $J_2 = 1.082627E^{-3}$ (m^3/s^2), the Equatorial radius is $R_0 = 6.378137E^{+6}$ (m), the gravitational constant $\mu = 3.986004418E^{+14}$ (m^3/s^2), and the Earth-rotation rate is $\omega_{ie} = 7.292115E^{-5}$ (rad/s).

References

[1] P. F. Roysdon and J. A. Farrell, "GPS-INS Outlier Detection and Elimination using a Sliding Window Filter," *American Control Conference, In Presc.*, 2017.

[2] ——, "Robust GPS-INS Outlier Accomodation using a Sliding Window Filter," *22th IFAC World Congress*, 2017.

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

[4] P. D. Groves, *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems*. Artech House, 2013.

[5] F. L. Markley and J. Crassidis, "Attitude Estimation Using Modified Rodrigues Parameters," *Proceedings of the Flight Mechanics/Estimation Theory Symposium*, 1996.

[6] M. Shuster, "A survey of attitude representations," *Journal of the Astronautical Sciences*, vol. 41, no. 4, pp. 439–517, 1993.

[7] ——, "The nature of the quaternion," *Journal of the Astronautical Sciences*, vol. 56, no. 3, p. 359, 2010.

[8] M. Wei and K. P. Schwarz, "A strapdown inertial algorithm using earth-fixed cartesian frame," *Navigation: JION*, vol. 371, no. 2, pp. 153–167, 1990.

Technical Note: INS State Error Model

Paul F. Roysdon[†] Jay A. Farrell[‡]

***Abstract*— Due to space limitations in [1] and [2], this Technical Note is supplied to explain why the dimension of the state vector $\boldsymbol{x} \in \mathbb{R}^{16}$ and the dimension of error state vector $\delta\boldsymbol{x} \in \mathbb{R}^{15}$. This Technical Note also describes the additive and multiplicative operations required to use $\delta\boldsymbol{x}$ to correct $\boldsymbol{x}$.**

I. Introduction

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

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

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, and $\mathbf{q} \in \mathbb{R}^4$ represents the attitude quaternion ($n_s = 16$), each at time t. Let $\hat{\boldsymbol{x}} \in \mathbb{R}^{n_s}$ denote the estimate of the rover state vector:

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

The error between $\boldsymbol{x}(t)$ and $\hat{\boldsymbol{x}}(t)$ is denoted as $\delta\boldsymbol{x}$. The error vector is

$$\delta\boldsymbol{x} = [\delta\mathbf{p}^\intercal, \delta\mathbf{v}^\intercal, \boldsymbol{\rho}^\intercal, \delta\mathbf{b}_a^\intercal, \delta\mathbf{b}_g^\intercal]^\intercal \in \mathbb{R}^{n_e},$$

where $\delta\mathbf{p}$ and $\delta\mathbf{v}$, each in $\mathbb{R}^3$, represent the error between the true and computed position and velocity, respectively. The small-angle error state, denoted as $\boldsymbol{\rho} \in \mathbb{R}^{3\times 1}$, is defined in Section 2.5.5 of [3], and discussed in Section II-B. The errors $\delta\mathbf{b}_a$ and $\delta\mathbf{b}_g$, each in $\mathbb{R}^3$, represent the accelerometer bias, and gyro bias errors, respectively. Therefore $\delta\boldsymbol{x} \in \mathbb{R}^{15}$ (i.e. $n_e = 15$). The fact that $n_s = 16$ and $n_e = 15$ is discussed in Section II-B.

II. State Correction

Let $\delta\hat{\boldsymbol{x}}$ denote an estimate of $\delta\boldsymbol{x}$. The state correction to the state vector $\hat{\boldsymbol{x}}$ is denoted as

$$\hat{\boldsymbol{x}}^+ = \hat{\boldsymbol{x}}^- \oplus \delta\hat{\boldsymbol{x}}.$$

The symbol $(-)$ denotes the prior estimate, whereas $(+)$ is the updated estimate. The symbol $\oplus$ is discussed in Sections II-A and II-B

A. Position, Velocity, and Bias Updates

Position, velocity, accelerometer bias and gyro bias, each have corrections which are additive. The state correction step is

$$\begin{aligned}
\hat{\mathbf{p}}^+ &= \hat{\mathbf{p}}^- + \delta\mathbf{p} \\
\hat{\mathbf{v}}^+ &= \hat{\mathbf{v}}^- + \delta\mathbf{v} \\
\hat{\mathbf{b}}_a^+ &= \hat{\mathbf{b}}_a^- + \delta\mathbf{b}_a \\
\hat{\mathbf{b}}_g^+ &= \hat{\mathbf{b}}_g^- + \delta\mathbf{b}_g.
\end{aligned}$$

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

B. Attitude Update

When the attitude error is sufficiently small (see Section 2.5.5 of [3]), the attitude can be represented as a set of small-angle planar rotations $\{\rho_x, \rho_y, \rho_z\}$ about three orthogonal axes $\{x, y, z\}$, thus the attitude error can be defined in $\mathbb{R}^3$.

1) Rotation Matrix: Let $\mathbf{R}_b^n \in \mathbb{R}^{3\times 3}$ represent the true rotation from body-frame (b) to navigation-frame (n) that is equivalent to $\mathbf{q}(t)$ (see eqn. D.13 in [3]). Let $\hat{\mathbf{R}}_b^n \in \mathbb{R}^{3\times 3}$ represent the computed rotation that is equivalent to $\hat{\mathbf{q}}(t)$. The error between the true and computed rotation is

$$\mathbf{R}_{\hat{n}}^n = (\mathbf{R}_b^n)(\hat{\mathbf{R}}_n^b),$$

where $\mathbf{R}_{\hat{n}}^n$ represents the rotation matrix from the computed to actual navigation frame. When the error between the true and computed rotation is zero, then $\mathbf{R}_{\hat{n}}^n = \mathbf{I}$. Otherwise, as discussed in Section 2.6.1 of [3],

$$\mathbf{R}_{\hat{n}}^n = [\mathbf{I} - \mathbf{P}]$$

where $\mathbf{P} = [\boldsymbol{\rho}\times]$, and $\boldsymbol{\rho} = [\rho_x, \rho_y, \rho_z]^\intercal \in \mathbb{R}^3$ (see eqn. 10.28 of [3]).

Using this notation, the attitude update (as defined in eqn. 10.29 of [3]) is

$$(\mathbf{R}_b^n)^+ = [\mathbf{I} - \mathbf{P}](\hat{\mathbf{R}}_b^n)^-.$$

Note that the attitude correction is multiplicative.

2) Quaternion: A similar approach to Section II-B.1 is valid when the attitude error is represented by a quaternion. Let $\mathbf{q}_b^n$ represent the true quaternion from b-frame to n-frame. Let $\hat{\mathbf{q}}_b^n$ represent the computed quaternion. The error may be represented as

$$\mathbf{q}_{\hat{n}}^n = \mathbf{q}_b^n \otimes \hat{\mathbf{q}}_n^b$$

where $\mathbf{q}_{\hat{n}}^n$ represents the quaternion from the computed to actual navigation frame. The symbol $\otimes$ represents the quaternion multiplication operation defined in Section D of [3]. When the error between the true and computed rotation is zero, then $\mathbf{q}_{\hat{n}}^n = [1, 0, 0, 0]^\intercal$, otherwise $\mathbf{q}_{\hat{n}}^n$ may be represented as

$$\mathbf{q}_{\hat{n}}^n = \begin{bmatrix} \hat{\mathbf{q}}_s \\ \hat{\mathbf{q}}_v \end{bmatrix} = \begin{bmatrix} \sqrt{1 - \left\|\frac{1}{2}\boldsymbol{\rho}\right\|_2^2} \\ \frac{1}{2}\boldsymbol{\rho} \end{bmatrix} \approxeq \begin{bmatrix} 1 \\ \frac{1}{2}\boldsymbol{\rho} \end{bmatrix}, \quad (1)$$

where the scalar part of the quaternion is $\hat{\mathbf{q}}_s = 1$, and the vector part is $\hat{\mathbf{q}}_v = \boldsymbol{\rho}$. The approximation on the right-hand side of eqn. (1) is shown in Appendix I.

Using this notation, the multiplicative quaternion update is

$$\hat{\mathbf{q}}_b^{n\ +} = \mathbf{q}_{\hat{n}}^n \otimes \hat{\mathbf{q}}_b^{n\ -}.$$

Quaternion operations are defined in Section D of [3].

APPENDIX I
QUATERNION UPDATE APPROXIMATION

Let $\boldsymbol{f}(\boldsymbol{\rho}) = \sqrt{1 - \left\|\frac{1}{2}\boldsymbol{\rho}\right\|_2^2} \in \mathbb{R}^1$, and $\delta\boldsymbol{\rho} = \boldsymbol{\rho} - \mathbf{0} \in \mathbb{R}^{3\times 1}$. By first-order Taylor series expansion of $\boldsymbol{f}(\boldsymbol{\rho})$, assuming small-angle $\boldsymbol{\rho}$, the quantity $\mathbf{q}_{\hat{n}}^n$ is

$$\begin{aligned}
\mathbf{q}_{\hat{n}}^n &= \begin{bmatrix} \boldsymbol{f}(\boldsymbol{\rho}) \\ \frac{1}{2}\boldsymbol{\rho} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{f}(\mathbf{0}) + \frac{\partial \boldsymbol{f}(\boldsymbol{\rho})}{\partial \boldsymbol{\rho}}\Big|_{\boldsymbol{\rho}=\mathbf{0}} \delta\boldsymbol{\rho} + \delta\boldsymbol{\rho}^\intercal \frac{\partial^2 \boldsymbol{f}(\boldsymbol{\rho})}{2\ \partial \boldsymbol{\rho}^2}\Big|_{\boldsymbol{\rho}=\mathbf{0}} \delta\boldsymbol{\rho} + \cdots \\ \frac{1}{2}\boldsymbol{\rho} \end{bmatrix} \\
&= \begin{bmatrix} 1 + \left[\frac{-\boldsymbol{\rho}}{2\sqrt{1-\left\|\frac{1}{2}\boldsymbol{\rho}\right\|_2^2}}\right]\Big|_{\boldsymbol{\rho}=\mathbf{0}} \delta\boldsymbol{\rho} \\ \frac{1}{2}\boldsymbol{\rho} \end{bmatrix} \\
&= \begin{bmatrix} 1 \\ \frac{1}{2}\boldsymbol{\rho} \end{bmatrix},
\end{aligned}$$

where the derivation of the gradient and Hessian is provided in Appendix II. Note:

- $\boldsymbol{f}(\mathbf{0}) = \sqrt{1 - \left\|\frac{1}{2}\mathbf{0}\right\|_2^2} = 1$.
- For $\boldsymbol{\rho} \ll 1$, then $\delta\boldsymbol{\rho}^\intercal \frac{\partial^2 \boldsymbol{f}(\boldsymbol{\rho})}{2\ \partial \boldsymbol{\rho}^2}\Big|_{\boldsymbol{\rho}=\mathbf{0}} \delta\boldsymbol{\rho} \approxeq 0$.
- $\hat{\mathbf{q}}_v$ is linear already.

APPENDIX II
GRADIENT AND HESSIAN DERIVATION

Let $\mathbf{h}(\boldsymbol{x}) = (\boldsymbol{x}^\intercal \boldsymbol{x})^{1/2}$ where $\boldsymbol{x} \in \mathbb{R}^{3\times 1}$. The Jacobian of $\mathbf{h}(\boldsymbol{x})$ is

$$\frac{\partial \mathbf{h}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\boldsymbol{x}^\intercal}{(\boldsymbol{x}^\intercal \boldsymbol{x})^{1/2}}.$$

The Hessian of $\mathbf{h}(\boldsymbol{x})$ is

$$\begin{aligned}
\frac{\partial^2 \mathbf{h}(\boldsymbol{x})}{\partial \boldsymbol{x}^2} &= \frac{\mathbf{I}}{(\boldsymbol{x}^\intercal \boldsymbol{x})^{1/2}} - \left(\frac{1}{2}\right)(2)\frac{\boldsymbol{x}\boldsymbol{x}^\intercal}{(\boldsymbol{x}^\intercal \boldsymbol{x})^{3/2}} \\
&= \frac{\boldsymbol{x}^\intercal \boldsymbol{x}\ \mathbf{I} - \boldsymbol{x}\boldsymbol{x}^\intercal}{(\boldsymbol{x}^\intercal \boldsymbol{x})^{3/2}}.
\end{aligned}$$

REFERENCES

[1] P. F. Roysdon and J. A. Farrell, "GPS-INS Outlier Detection and Elimination using a Sliding Window Filter," *American Control Conference, In Presc.*, 2017.

[2] ——, "Robust GPS-INS Outlier Accomodation using a Sliding Window Filter," *22th IFAC World Congress*, 2017.

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

Technical Note: INS Noise Propagation

Paul F. Roysdon[†] Jay A. Farrell[‡]

***Abstract*— Due to space limitations in [1] and [2], this Technical Note is supplied to explain the state and noise propagation for an inertial navigation system (INS), between two aiding measurement times. The temporal propagation of the state error and noise is required for optimal state estimation.**

I. Introduction

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 ($n_s = 16$).

Let $\boldsymbol{x}_v(t) = [\mathbf{p}^\intercal(t), \mathbf{v}^\intercal(t), \mathbf{q}^\intercal(t)]^\intercal \in \mathbb{R}^{10}$ represent the vehicle state position, velocity and attitude. Let $\boldsymbol{x}_c(t) = [\mathbf{b}_a^\intercal(t), \mathbf{b}_g^\intercal(t)]^\intercal \in \mathbb{R}^6$ represent the IMU calibration terms: accelerometer bias and gyro bias. Then $\boldsymbol{x}(t)$ can be represented as $\boldsymbol{x}(t) = [\boldsymbol{x}_v^\intercal(t), \boldsymbol{x}_c^\intercal(t)]^\intercal$.

Let τ_i denote the time instants of the i^{th} IMU measurements of $\mathbf{u}$, as defined in Section II.A in [1] and [2]. Let $\boldsymbol{x}_i = \boldsymbol{x}(\tau_i)$ and $\mathbf{u}_i = \mathbf{u}(\tau_i)$.

Let the state estimate time propagation be represented as

$$\hat{\boldsymbol{x}}_{i+1} \doteq \boldsymbol{\phi}(\hat{\boldsymbol{x}}_i, \tilde{\mathbf{u}}_i),$$

where the vehicle state estimate is $\hat{\boldsymbol{x}}_{v,i+1} \doteq \boldsymbol{\phi}_v(\hat{\boldsymbol{x}}_{v,i}, \tilde{\mathbf{u}}_i)$.

Let the true state time propagation be represented as

$$\boldsymbol{x}_{i+1} = \boldsymbol{\phi}(\boldsymbol{x}_i, \mathbf{u}_i),$$

and the true vehicle state as $\boldsymbol{x}_{v,i+1} = \boldsymbol{\phi}_v(\boldsymbol{x}_{v,i}, \mathbf{u}_i)$.

Define the state error as

$$\delta\boldsymbol{x}_i = \boldsymbol{x}_i \ominus \hat{\boldsymbol{x}}_i \in \mathbb{R}^{n_e},$$

where the symbol '$\ominus$', which is discussed in [3], represents the subtraction operation for position, velocity and bias states, and the multiplication operation of the attitude states. The fact that $n_s = 16$ and $n_e = 15$ is discussed in [3]. Let $\delta\boldsymbol{x}_{v,i} \in \mathbb{R}^9$ represent the vehicle state error for position, velocity and attitude. Let $\delta\boldsymbol{x}_{c,i} \in \mathbb{R}^6$ represent the error in the IMU calibration terms: accelerometer bias and gyro bias. Then $\delta\boldsymbol{x}_i$ can be represented as $\delta\boldsymbol{x}_i = [\delta\boldsymbol{x}_{v,i}^\intercal, \delta\boldsymbol{x}_{c,i}^\intercal]^\intercal$.

Let the IMU measurement be defined as

$$\tilde{\mathbf{u}}(\tau_i) \triangleq \mathbf{u}(\tau_i) - \mathbf{b}(\tau_i) - \boldsymbol{\omega}_u(\tau_i) \in \mathbb{R}^6,$$

with additive stochastic errors $\boldsymbol{\omega}_u(\tau_i) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q_d})$ and $\mathbf{b} = [\mathbf{b}_a^\intercal, \mathbf{b}_g^\intercal]^\intercal$. The sensor bias $\mathbf{b}$ represents time correlated measurement errors, and $\boldsymbol{\omega}_u(\tau_i)$ represents the white measurement errors. Let the estimate $\hat{\mathbf{u}}_i \triangleq \tilde{\mathbf{u}}_i + \hat{\mathbf{b}}_i$, where $\hat{\mathbf{b}}_i$ is the estimate of $\mathbf{b}_i$ (i.e. $\hat{\boldsymbol{x}}_{c,i}$). Let the measurements $\tilde{\mathbf{u}}(\tau_i)$ be defined for IMU measurement times i, between aiding measurement times k, such that $\tau_i \in [t_{k-1}, t_k]$.

Define

$$\begin{aligned}\delta\mathbf{u}_i &\triangleq \mathbf{u}_i - \hat{\mathbf{u}}_i \\ &= \mathbf{u}_i - \tilde{\mathbf{u}}_i - \hat{\mathbf{b}}_i \\ &= \mathbf{u}_i - (\mathbf{u}_i - \mathbf{b}_i - \boldsymbol{\omega}_{u,i}) - \hat{\mathbf{b}}_i \\ &= \delta\mathbf{b}_i + \boldsymbol{\omega}_{u,i},\end{aligned}$$

where $\delta\mathbf{b}_i \triangleq \mathbf{b}_i - \hat{\mathbf{b}}_i$, and $\delta\mathbf{b}_i$ (i.e. $\delta\mathbf{b}_i = \delta\boldsymbol{x}_{c,i}$) is a state calibration term.

Linearization of the state error, using Taylor series to first order, yields

$$\begin{aligned}\delta\boldsymbol{x}_{v,i+1} &= \boldsymbol{\phi}_v(\boldsymbol{x}_{v,i}, \mathbf{u}_i) - \boldsymbol{\phi}_v(\hat{\boldsymbol{x}}_{v,i}, \hat{\mathbf{u}}_i) \\ &= \boldsymbol{\phi}_v(\hat{\boldsymbol{x}}_{v,i}, \hat{\mathbf{u}}_i) + \left.\frac{\partial\boldsymbol{\phi}_v}{\partial\boldsymbol{x}_{v,i}}\right|_{\hat{\boldsymbol{x}}_{v,i}} \delta\boldsymbol{x}_{v,i} \\ &\quad + \left.\frac{\partial\boldsymbol{\phi}_v}{\partial\mathbf{u}_i}\right|_{\hat{\mathbf{u}}_i} \delta\mathbf{u}_i - \boldsymbol{\phi}_v(\hat{\boldsymbol{x}}_{v,i}, \hat{\mathbf{u}}_i) \\ &= \left.\frac{\partial\boldsymbol{\phi}_v}{\partial\boldsymbol{x}_{v,i}}\right|_{\hat{\boldsymbol{x}}_{v,i}} \delta\boldsymbol{x}_{v,i} + \left.\frac{\partial\boldsymbol{\phi}_v}{\partial\mathbf{u}_i}\right|_{\hat{\mathbf{u}}_i} \delta\mathbf{u}_i \\ &= \mathbf{A}_i\delta\boldsymbol{x}_{v,i} + \mathbf{B}_i\delta\mathbf{u}_i \\ &= \mathbf{A}_i\delta\boldsymbol{x}_{v,i} + \mathbf{B}_i\delta\mathbf{b}_i + \mathbf{B}_i\boldsymbol{\omega}_{u,i}, \qquad (1)\end{aligned}$$

where $\mathbf{A}_i = \left.\frac{\partial\boldsymbol{\phi}_v}{\partial\boldsymbol{x}_{v,i}}\right|_{\hat{\boldsymbol{x}}_{v,i},\hat{\mathbf{u}}_i} \in \mathbb{R}^{9\times9}$, and $\mathbf{B}_i = \left.\frac{\partial\boldsymbol{\phi}_v}{\partial\mathbf{u}_i}\right|_{\hat{\boldsymbol{x}}_{v,i},\hat{\mathbf{u}}_i} \in \mathbb{R}^{9\times6}$.

Let the model of the sensor bias be defined as a first-order Gauss-Markov process

$$\delta\mathbf{b}_{i+1} = \mathbf{F}_b\delta\mathbf{b}_i + \boldsymbol{\nu}, \qquad (2)$$

where $\mathbf{F_b} \in \mathbb{R}^{6\times6}$ is selected such that the bias errors are modeled as either random constants or random walk plus constants (see eqns. 11.106 and 11.107 of [4]), and $\boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0}, \sigma_\nu\mathbf{I})$.

Rewriting eqns. (1) and (2) in matrix form:

$$\begin{bmatrix}\delta\boldsymbol{x}_{v,i+1}\\ \delta\boldsymbol{x}_{c,i+1}\end{bmatrix} = \begin{bmatrix}\mathbf{A}_i & \mathbf{B}_i\\ \mathbf{0} & \mathbf{F_b}\end{bmatrix}\begin{bmatrix}\delta\boldsymbol{x}_{v,i}\\ \delta\mathbf{b}_i\end{bmatrix} + \begin{bmatrix}\mathbf{B}_i & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{\omega}_{u,i}\\ \boldsymbol{\nu}\end{bmatrix}. \qquad (3)$$

For analysis, in the following section let $\mathbf{F_b} = \mathbf{I}$.

II. Propagation of State Error

This section analyzes the error accumulation over the time interval $t \in [k-1, k]$ using superposition.

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

A. Propagation of Initial State Error

Consider eqn. (3) over the interval $t \in [k-1, k]$, where $\boldsymbol{\omega}_{u,i-1} = \mathbf{0}$ and $\boldsymbol{\nu} = \mathbf{0}$. Without loss of generality let $k = 1$, such that $t \in [0, 1]$. For each time instant, eqn. (3) can be represented in terms of $\mathbf{A}_i$, and $\mathbf{B}_i$, with initial condition errors $\delta\boldsymbol{x}_{v,0}$, and $\delta\mathbf{b}_0$:

$$\begin{aligned}
\delta\boldsymbol{x}_1 &= \begin{bmatrix} \mathbf{A}_0 & \mathbf{B}_0 \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \delta\boldsymbol{x}_{v,0} \\ \delta\mathbf{b}_0 \end{bmatrix} \\
\delta\boldsymbol{x}_2 &= \begin{bmatrix} \mathbf{A}_1 & \mathbf{B}_1 \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \delta\boldsymbol{x}_{v,1} \\ \delta\mathbf{b}_1 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{A}_1 & \mathbf{B}_1 \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}_0 & \mathbf{B}_0 \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \delta\boldsymbol{x}_{v,0} \\ \delta\mathbf{b}_0 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{A}_1\mathbf{A}_0 & \mathbf{A}_1\mathbf{B}_0 + \mathbf{B}_1 \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \delta\boldsymbol{x}_{v,0} \\ \delta\mathbf{b}_0 \end{bmatrix} \\
\delta\boldsymbol{x}_3 &= \begin{bmatrix} \mathbf{A}_2 & \mathbf{B}_2 \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \delta\boldsymbol{x}_{v,2} \\ \delta\mathbf{b}_2 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{A}_2 & \mathbf{B}_2 \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A}_1\mathbf{A}_0 & \mathbf{A}_1\mathbf{B}_0 + \mathbf{B}_1 \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \delta\boldsymbol{x}_{v,0} \\ \delta\mathbf{b}_0 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{A}_2\mathbf{A}_1\mathbf{A}_0 & \mathbf{A}_2\mathbf{A}_1\mathbf{B}_0 + \mathbf{A}_2\mathbf{B}_1 + \mathbf{B}_2 \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \delta\boldsymbol{x}_{v,0} \\ \delta\mathbf{b}_0 \end{bmatrix}. \quad (4)
\end{aligned}$$

Define F_s as the sample frequency of the sensor (e.g. IMU). As defined in Section II.A of both [1] and [2], let $\mathbf{U}_k = \{\tilde{\mathbf{u}}(\tau_i) \text{ for } \tau_i \in [t_{k-1}, t_k]\}$. As defined in Section III.A of both [1] and [2], let $\mathbf{X}_k = [\boldsymbol{x}(t_{k-L})^\intercal, \ldots, \boldsymbol{x}(t_k)^\intercal]^\intercal \in \mathbb{R}^{n_s(L+1)}$ denote the vehicle trajectory over a sliding time window that contains L one second GPS measurement epochs: $[\mathbf{y}_{k-L+1}, \ldots, \mathbf{y}_k]$. After F_s IMU time steps (i.e. $F_{s,k=1}$)

$$\begin{aligned}
\delta\boldsymbol{x}_{F_s} &= \left\{ \prod_{i=1}^{F_s} \begin{bmatrix} \mathbf{A}_i & \mathbf{B}_i \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \right\} \begin{bmatrix} \delta\boldsymbol{x}_{v,0} \\ \delta\mathbf{b}_0 \end{bmatrix} \quad (5) \\
&= \boldsymbol{\Upsilon}(\hat{\mathbf{X}}_k, \mathbf{U}_k)[\delta\boldsymbol{x}_{v,0}, \delta\mathbf{b}_0]^\intercal, \quad (6)
\end{aligned}$$

where the operator $\boldsymbol{\Upsilon}(\hat{\mathbf{X}}_k, \mathbf{U}_k)$ in eqn. (6) represents the product operation in eqn. (5), and $\hat{\mathbf{X}}_k$ is the estimate of $\mathbf{X}_k$. The product operation in eqn. (5) must follow the order of multiplications shown in eqn. (4).

B. Noise Propagation

Again consider eqn. (3) over the interval $t \in [k-1, k]$. Here we will analyze the effect of the noise terms $\boldsymbol{\omega}_u$ and $\boldsymbol{\nu}$, with $\delta\boldsymbol{x}_{v,0}$ and $\delta\mathbf{b}_0$ both zero.

To simplify notation, let

$$\mathbf{C}_i \triangleq \begin{bmatrix} \mathbf{A}_i & \mathbf{B}_i \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \quad \mathbf{D}_i \triangleq \begin{bmatrix} \mathbf{B}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix},$$

and

$$\delta\boldsymbol{x}_i \triangleq \begin{bmatrix} \delta\boldsymbol{x}_{v,i} \\ \delta\boldsymbol{x}_{c,i} \end{bmatrix}, \quad \mathbf{n}_i \triangleq \begin{bmatrix} \boldsymbol{\omega}_{u,i} \\ \boldsymbol{\nu} \end{bmatrix}.$$

Defining eqn. (3) using the terms above, we have

$$\delta\boldsymbol{x}_{i+1} = \mathbf{C}_i\delta\boldsymbol{x}_i + \mathbf{D}_i\mathbf{n}_i. \quad (7)$$

Performing operations on eqn. (7) (similar to the operations leading up to eqn. (4)),

$$\begin{aligned}
\delta\boldsymbol{x}_1 &= \mathbf{C}_0\delta\boldsymbol{x}_0 + \mathbf{D}_0\mathbf{n}_0 \\
\delta\boldsymbol{x}_2 &= \mathbf{C}_1\delta\boldsymbol{x}_1 + \mathbf{D}_1\mathbf{n}_1 \\
&= \mathbf{C}_1(\mathbf{C}_0\delta\boldsymbol{x}_0 + \mathbf{D}_0\mathbf{n}_0) + \mathbf{D}_1\mathbf{n}_1 \\
&= \mathbf{C}_1\mathbf{C}_0\delta\boldsymbol{x}_0 + \mathbf{C}_1\mathbf{D}_0\mathbf{n}_0 + \mathbf{D}_1\mathbf{n}_1 \\
\delta\boldsymbol{x}_3 &= \mathbf{C}_2\delta\boldsymbol{x}_2 + \mathbf{D}_2\mathbf{n}_2 \\
&= \mathbf{C}_2(\mathbf{C}_1\mathbf{C}_0\delta\boldsymbol{x}_0 + \mathbf{C}_1\mathbf{D}_0\mathbf{n}_0 + \mathbf{D}_1\mathbf{n}_1) + \mathbf{D}_2\mathbf{n}_2 \\
&= \mathbf{C}_2\mathbf{C}_1\mathbf{C}_0\delta\boldsymbol{x}_0 + \mathbf{C}_2\mathbf{C}_1\mathbf{D}_0\mathbf{n}_0 + \mathbf{C}_2\mathbf{D}_1\mathbf{n}_1 + \mathbf{D}_2\mathbf{n}_2. \quad (8)
\end{aligned}$$

For $i = F_s$, and $\delta\boldsymbol{x}_0 = \mathbf{0}$, the terms in eqn. (8) can be defined as

$$\begin{aligned}
\mathbf{w}_{k-1} &= \left\{ \sum_{i=0}^{F_s-2} \left(\prod_{j=i+1}^{F_s-1} \mathbf{C}_j \right) \mathbf{D}_i\mathbf{n}_i \right\} + \mathbf{D}_{F_s-1}\mathbf{n}_{F_s-1} \quad (9) \\
&= \boldsymbol{\Gamma}\boldsymbol{\eta}.
\end{aligned}$$

Let the product of $\mathbf{C}_j$ in eqn. (9) be defined as

$$\mathbf{C}_j^p \triangleq \begin{cases} \prod_{j=q}^{p} \mathbf{C}_j = \mathbf{C}_p \cdots \mathbf{C}_{q-1}\mathbf{C}_q & \text{for } q \neq p \\ \mathbf{C}_q & \text{for } q = p \end{cases} \quad (10)$$

where product operation in eqn. (10) must follow the order of operations shown in eqn. (8). Let $\boldsymbol{\Gamma}$ and $\boldsymbol{\eta}$ be defined as

$$\begin{aligned}
\boldsymbol{\Gamma} &\triangleq \left[\mathbf{C}_1^{F_s-1}\mathbf{D}_0, \mathbf{C}_2^{F_s-1}\mathbf{D}_1, \ldots, \mathbf{C}_{F_s-1}^{F_s-1}\mathbf{D}_{F_s-2}, \mathbf{D}_{F_s-1} \right] \\
\boldsymbol{\eta} &\triangleq [\mathbf{n}_0, \mathbf{n}_1, \ldots, \mathbf{n}_{F_s-1}].
\end{aligned}$$

C. Summary

Combining the results from Sections II-A and II-B, the linear state transition error model over $t \in [t_{k-1}, t_k]$ is

$$\delta\boldsymbol{x}_k = \boldsymbol{\Upsilon}_{k-1}\delta\boldsymbol{x}_{k-1} + \mathbf{w}_{k-1}. \quad (11)$$

with

$$\begin{aligned}
\mathbf{Q}_\mathbf{D} &= \mathrm{Cov}(\mathbf{w}_{k-1}) \in \mathbb{R}^{n_e \times n_e} \\
&= E\left\langle \boldsymbol{\Gamma}\boldsymbol{\eta}\boldsymbol{\eta}^\intercal\boldsymbol{\Gamma}^\intercal \right\rangle \\
&= E\left\langle \boldsymbol{\Gamma} \begin{bmatrix} \boldsymbol{\eta}_0 \\ \boldsymbol{\eta}_1 \\ \vdots \\ \boldsymbol{\eta}_{F_s-1} \end{bmatrix} [\boldsymbol{\eta}_0 \ \boldsymbol{\eta}_1 \ \cdots \ \boldsymbol{\eta}_{F_s-1}] \boldsymbol{\Gamma}^\intercal \right\rangle \\
&= \boldsymbol{\Gamma} \begin{bmatrix} \mathbf{Q}_{\mathbf{d},0} & & \\ & \ddots & \\ & & \mathbf{Q}_{\mathbf{d},F_s-1} \end{bmatrix} \boldsymbol{\Gamma}^\intercal, \quad (12)
\end{aligned}$$

where $\mathbf{Q}_{\mathbf{d},i} = \boldsymbol{\eta}_i\boldsymbol{\eta}_i^\intercal$. The stochastic properties of eqn. (12) are well understood, and can be found in Sections 4.7 and 7.2.5.2 of [4].

References

[1] P. F. Roysdon and J. A. Farrell, "GPS-INS Outlier Detection and Elimination using a Sliding Window Filter," *American Control Conference, In Presc.*, 2017.

[2] ——, "Robust GPS-INS Outlier Accomodation using a Sliding Window Filter," *22th IFAC World Congress*, 2017.

[3] ——. (2017, Feb) "Technical Note: CRT with Hypothesis Testing". [Online]. Available: https://escholarship.org/uc/item/5r88s8dh

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

Technical Note: Computational Cost of the CRT Estimator

Paul F. Roysdon[†] Jay A. Farrell[‡]

I. Introduction

In this Technical Note we consider the computational cost (number of floating-point operations (FLOPs)) for each update of the CRT estimator. We assume that the implementation uses a sparse matrix library, and only compute and store the upper-triangular elements of all symmetric positive-definite matrices used in the CRT estimator. First we define the computational cost of the Kalman Filter (KF) [1], then evaluate the most expensive operation, the computation of $\boldsymbol{\delta}\mathbf{X}$ in the CRT estimator [2].

Consider the general matrices/vectors, $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{v} \in \mathbb{R}^{n\times 1}$, upper-triangular $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$, and $\mathbf{B} \in \mathbb{R}^{m\times n}$, with $m > n$. The computational cost for basic linear algebra operations on dense matrices [3] is shown in the first-half of Table I. If $\mathbf{A}$ and $\mathbf{B}$ are sparse, a lower-bound can be determined using sparse matrix operations (see [4]), the results are shown in the second-half of Table I.

The computational cost for three matrix decompositions is provided in Table II.

The computational cost for three Least-Square (LS) algorithms is provided in Table III.

TABLE I

Computational cost for dense and sparse matrices.

Operation	Computational Cost
Dense matrices	
$\mathbf{A}^{-1}$	$\approx n^3$
$\mathbf{AA}$	n^3
$\mathbf{Av}$	n^2
$\mathbf{AB}^\intercal$	mn^2
$\mathbf{B}^\intercal\mathbf{B}$	mn^2
$\mathbf{Bv}$	mn
$\boldsymbol{\Sigma}\mathbf{v}$	$\frac{n^2-n}{2}+n$
$\boldsymbol{\Sigma}\mathbf{A}$	$\frac{n^3-n^2}{2}+n^2$
$\boldsymbol{\Sigma}\boldsymbol{\Sigma}$	$\frac{n^3}{3}$
Sparse matrices	
$\mathbf{A}^{-1}$	n
$\mathbf{AA}$	n
$\mathbf{Av}$	n
$\mathbf{AB}^\intercal$	n
$\mathbf{B}^\intercal\mathbf{B}$	n
$\mathbf{Bv}$	n

II. Computational cost for KF

For the KF state estimate $\hat{\boldsymbol{x}} \in \mathbb{R}^n$, and measurement $\mathbf{y} \in \mathbb{R}^m$ with $m > n$, the computational cost with dense matrices is presented in Table IV.

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

TABLE II

Computational cost for matrix decompositions.

Decomposition	Computational Cost
Cholesky (Section 4.2.1 [3])	$\frac{n^3}{3}$
Householder QR (Section 5.2.1 [3])	$4\left(m^2n - mn^2 + \frac{n^3}{3}\right)$
Givens QR (Section 5.2.3 [3])	$3n^2\left(m - \frac{n}{3}\right)$

TABLE III

Computational cost for LS algorithms (Section 5.5.9 [3]).

LS Algorithm	Computational Cost
Normal Equations	$mn^2 + \frac{n^3}{3}$
Householder QR Orthogonalization	$2mn^2 + \frac{2n^3}{3}$
Givens QR Orthogonalization	$3mn^2 - n^3$

TABLE IV

Computational cost for the KF (Section 3.3.1 [5]).

Operation	Computational Cost
System Propagation	
$\hat{\boldsymbol{x}}_k^- = \boldsymbol{\Phi}_{k-1}\hat{\boldsymbol{x}}_{k-1}^+$	n^2
$\mathbf{P}_k^- = \boldsymbol{\Phi}_{k-1}\mathbf{P}_{k-1}^-\boldsymbol{\Phi}_{k-1}^\intercal + \mathbf{Qd}_{k-1}$	$2n^3$
Measurement update	
$\mathbf{K}_k = \mathbf{P}_k^-\mathbf{H}_k^\intercal(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^\intercal + \mathbf{R}_k)^{-1}$	$2mn^2 + 2mn$
$\hat{\boldsymbol{x}}_k^+ = \hat{\boldsymbol{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\hat{\boldsymbol{x}}_k^-)$	$2mn$
$\mathbf{P}_k^+ = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^-$	$2mn^2$

III. Computational cost for CRT using Normal Equations

This section is split into two primary parts: first compute $\boldsymbol{\delta}\mathbf{X}$ by dense matrices, then compute $\boldsymbol{\delta}\mathbf{X}$ by sparse matrices. In both cases we only consider one iteration of the optimization. Because the calculation of $\boldsymbol{\delta}\mathbf{X}$ is the most expensive operation in the CRT estimator, assume the state transition matrix $\boldsymbol{\Phi}$, measurement matrix $\mathbf{H}$, measurement prediction $\hat{\mathbf{y}}$, integrated state vector $\boldsymbol{x}_i$, and optimized state vector $\boldsymbol{x}^*$ are each given (refer to Sections 2 & 3 of [2]) and are therefore not included in the cost to compute $\boldsymbol{\delta}\mathbf{X}$.

A. Dense Jacobian

Given the number of error states n_e, the number of measurements n_m, and CRT window length L, the components of the normalized Jacobian matrix $\mathbf{J}$ and residual vector $\mathbf{r}$, are defined below with a mapping to the number of computations required for dense matrices. Both $\mathbf{J}$ and $\mathbf{r}$

are normalized using Cholesky factorization[1], which has computational complexity $\frac{n^3}{3}$. Define $\boldsymbol{\delta}\mathbf{X}_O$, $\mathbf{r}_O$, and $\mathbf{J}_O$ as the number of operations to compute $\boldsymbol{\delta}\mathbf{X}$, $\mathbf{r}$, and $\mathbf{J}$, respectively.

The components of the Jacobian are,

$$\begin{aligned}
\mathbf{J}_{\mathbf{P}_0} &\triangleq \boldsymbol{\Sigma}_{\mathbf{P}_0}[\mathbf{I}, \mathbf{0}] \mapsto \left(\frac{n_e^3}{3} + n_e^3\right) \\
\mathbf{J}_{\mathbf{Qd}} &\triangleq \boldsymbol{\Sigma}_{\mathbf{Qd}}[\boldsymbol{\Phi}, -\mathbf{I}] \mapsto \left(\frac{n_e^3}{3} + 2n_e^3\right) L \\
\mathbf{J}_{\mathbf{R}} &\triangleq \boldsymbol{\Sigma}_{\mathbf{R}}\mathbf{H} \mapsto \left(\frac{n_m^3}{3} + n_m n_e^2\right) L \\
\mathbf{J} &\triangleq \begin{bmatrix} \mathbf{J}_{\mathbf{P}_0} \\ \mathbf{J}_{\mathbf{Qd}} \\ \mathbf{J}_{\mathbf{R}} \end{bmatrix} \mapsto \begin{bmatrix} \frac{n_e^3}{3} + n_e^3 \\ \left(\frac{n_e^3}{3} + 2n_e^3\right) L \\ \left(\frac{n_m^3}{3} + n_m n_e^2\right) L \end{bmatrix},
\end{aligned}$$

where, $\mathbf{J}_{\mathbf{P}_0} \in \mathbb{R}^{n_e \times n_e}$, $\mathbf{J}_{\mathbf{Qd}} \in \mathbb{R}^{n_e L \times n_e L}$, $\mathbf{J}_{\mathbf{R}} \in \mathbb{R}^{n_m L \times n_e L}$, and $\mathbf{J} \in \mathbb{R}^{(n_e(L+1)+n_m L) \times (n_e(L+1))}$. The number of operations to compute $\mathbf{J}$ is

$$\mathbf{J}_O = \frac{n_e^3}{3} + n_e^3 + \left(\frac{n_e^3}{3} + 2n_e^3 + \frac{n_m^3}{3} + n_m n_e^2\right) L.$$

The components of the residual vector are,

$$\begin{aligned}
\mathbf{r}_{\mathbf{P}_0} &\triangleq \boldsymbol{\Sigma}_{\mathbf{P}_0}[\boldsymbol{\mu} - \boldsymbol{x}^*] \mapsto n_e^2 \\
\mathbf{r}_{\mathbf{Qd}} &\triangleq \boldsymbol{\Sigma}_{\mathbf{Qd}}[\boldsymbol{x}^* - \boldsymbol{x}_i] \mapsto n_e^2 L \\
\mathbf{r}_{\mathbf{R}} &\triangleq \boldsymbol{\Sigma}_{\mathbf{R}}\boldsymbol{\delta}\mathbf{y} \mapsto n_m n_e L \\
\mathbf{r} &\triangleq \begin{bmatrix} \mathbf{r}_{\mathbf{P}_0} \\ \mathbf{r}_{\mathbf{Qd}} \\ \mathbf{r}_{\mathbf{R}} \end{bmatrix} \mapsto \begin{bmatrix} n_e^2 \\ n_e^2 L \\ n_m n_e L \end{bmatrix},
\end{aligned}$$

where, $\mathbf{r}_{\mathbf{P}_0} \in \mathbb{R}^{n_e \times 1}$, $\mathbf{r}_{\mathbf{Qd}} \in \mathbb{R}^{n_e L \times 1}$, $\mathbf{r}_{\mathbf{R}} \in \mathbb{R}^{n_m L \times 1}$, and $\mathbf{r} \in \mathbb{R}^{(n_e(L+1)+n_m L) \times 1}$. Note, the $\frac{n_e^3}{3}$ and $\frac{n_m^3}{3}$ are not included in the components $\mathbf{r}$, because the cost to compute the Cholesky decomposition is included in the components of $\mathbf{J}$. The number of operations to compute $\mathbf{r}$ is

$$\mathbf{r}_O = n_e^2 + (n_e^2 + n_m n_e)L.$$

Given the residual vector $\mathbf{r} \in \mathbb{R}^{p \times 1}$ and (dense) Jacobian $\mathbf{J} \in \mathbb{R}^{p \times q}$ for $p > q$, where $p = (n_e \times (L+1)) + (n_m \times L)$ and $q = (n_e \times (L+1))$, the number of operations to compute $\boldsymbol{\delta}\mathbf{X}$ by the Normal Equations is

$$\boldsymbol{\delta}\mathbf{X} = (\mathbf{J}^\intercal\mathbf{J})^{-1}\mathbf{J}^\intercal\mathbf{r} \mapsto pq^2 + \frac{q^3}{3} = \boldsymbol{\delta}\mathbf{X}_O.$$

Expanding p and q:

$$\begin{aligned}
\boldsymbol{\delta}\mathbf{X}_O = n_e^3 &\left(\frac{4}{3}L^3 + 4L^2 + 4L + \frac{4}{3}\right) \\
&+ n_e^2 n_m (L^3 + 2L^2 + L).
\end{aligned}$$

[1]Householder QR, or Givens QR factorization may be used instead (see Table II).

Define $\boldsymbol{\delta}\mathbf{X}_{TO}$ as the total number of operations to compute $\boldsymbol{\delta}\mathbf{X}$ by the Normal Equations.

$$\begin{aligned}
\boldsymbol{\delta}\mathbf{X}_{TO} &= \boldsymbol{\delta}\mathbf{X}_O + \mathbf{r}_O + \mathbf{J}_O \\
&= n_e^3 \left(\frac{4}{3}L^3 + 4L^2 + 4L + \frac{4}{3}\right) \\
&\quad + n_e^2 n_m (L^3 + 2L^2 + L). \\
&\quad + n_e^2 + \left(n_e^2 + n_m n_e\right) L \\
&\quad + \frac{n_e^3}{3} + n_e^3 + \left(\frac{n_e^3}{3} + n_e^2 + \frac{n_m^3}{3} + n_m n_e^2\right) L \\
&= n_e^3 \left(\frac{4}{3}L^3 + 4L^2 + \frac{19}{3}L + \frac{8}{3}\right) \\
&\quad + n_e^2 (n_m L^3 + 2n_m L^2 + 2n_m L + L + 1) \\
&\quad + n_e n_m L + \frac{n_m^3}{3} L.
\end{aligned}$$

Therefore the computation for a single iteration of the dense $\boldsymbol{\delta}\mathbf{X}$ by the Normal Equations is $\mathcal{O}(n_e^3 L^3)$.

B. Sparse Jacobian

Presently, common sparse libraries do not support LS by the Normal Equations because the operations are not efficient (see [3] and [4]). By inspection, the operation $\mathbf{J}^\intercal\mathbf{J}$ produces a dense matrix, and the inversion of $\mathbf{J}^\intercal\mathbf{J}$ is $\mathcal{O}(n_e^3 L^3)$.

IV. Computational Cost for CRT using QR Factorization

This section first introduces the theory for solving systems of linear equations using QR factorization, then follows an outline similar to Section III. The same assumptions of Section III apply here.

A. Theory

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m \geq n$ and $\mathbf{b} \in \mathbb{R}^m$. Suppose that the orthogonal matrix $\mathbf{Q} \in \mathbb{R}^{m \times m}$ is computed such that

$$\mathbf{Q}^\intercal\mathbf{A} = \mathbf{R} = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{0} \end{bmatrix}$$

is upper-triangular, with $\mathbf{R}_1 \in \mathbb{R}^{n \times n}$ and $\mathbf{0} \in \mathbb{R}^{(m-n) \times n}$. If

$$\mathbf{Q}^\intercal\mathbf{b} = \begin{bmatrix} \mathbf{c} \\ \mathbf{d} \end{bmatrix}$$

where $\mathbf{c} \in \mathbb{R}^{n \times 1}$ and $\mathbf{d} \in \mathbb{R}^{(m-n) \times 1}$, then

$$\begin{aligned}
\|\mathbf{A}\boldsymbol{x} - \mathbf{b}\|_2^2 &= \|\mathbf{Q}^\intercal\mathbf{A}\boldsymbol{x} - \mathbf{Q}^\intercal\mathbf{b}\|_2^2 \\
&= \|\mathbf{R}_1\boldsymbol{x} - \mathbf{c}\|_2^2 + \|\mathbf{d}\|_2^2 \qquad (1)
\end{aligned}$$

for any $\boldsymbol{x} \in \mathbb{R}^n$. If $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{R}_1) = n$, then the Least-Square estimate $\hat{\boldsymbol{x}}$ is defined by the upper-triangular system $\mathbf{R}_1\hat{\boldsymbol{x}} = \mathbf{c}$. Given $\mathbf{R}_1$, solving this system requires $2n^2\left(m - \frac{n}{3}\right)$ FLOPs. While $\mathcal{O}(mn)$ are required to update $\mathbf{c}$, and $\mathcal{O}(n^2)$ are required for back-substitution (see Section 5.3.2 of [3]), the most expensive operation is the QR factorization of $\mathbf{A}$, which requires at least $3n^2\left(m - \frac{n}{3}\right)$ FLOPs (assuming Givens QR).

Modifying eqn. (1) for the CRT estimator, define $\mathbf{J} \in \mathbb{R}^{p\times q}$, $\mathbf{r} \in \mathbb{R}^{p\times 1}$, $\boldsymbol{\delta}\mathbf{X} \in \mathbb{R}^{q\times 1}$, $\mathbf{R}_1 \in \mathbb{R}^{q\times q}$, $\mathbf{c} \in \mathbb{R}^{q\times 1}$ and $\mathbf{d} \in \mathbb{R}^{(p-q)\times 1}$. For

$$\mathbf{J}\boldsymbol{\delta}\mathbf{X} = \mathbf{r}$$

the Least-Square estimate $\boldsymbol{\delta}\hat{\mathbf{X}}$ of $\boldsymbol{\delta}\mathbf{X}$ is found by

$$\|\mathbf{J}\boldsymbol{\delta}\mathbf{X} - \mathbf{r}\|_2^2 = \|\mathbf{R}_1\boldsymbol{\delta}\mathbf{X} - \mathbf{c}\|_2^2 + \|\mathbf{d}\|_2^2$$

where $\boldsymbol{\delta}\hat{\mathbf{X}}$ is defined by the upper-triangular system $\mathbf{R}_1\boldsymbol{\delta}\hat{\mathbf{X}} = \mathbf{c}$.

B. Dense Jacobian

Let $\mathbf{J}_{QR,d}$ denote the number of operations to compute the QR factorization of (dense) $\mathbf{J}$. Using the previously defined values for p and q, and the values in Table II,

$$\begin{aligned}\mathbf{J}_{QR,d} &= 3q^2\left(p - \frac{q}{3}\right)\\ &= 2n_e^3(L^3 + 3L^2 + 3L + 1)\\ &\quad + 3n_e^2 n_m(L^3 + 2L^2 + L).\end{aligned}$$

Let $\boldsymbol{\delta}\hat{\mathbf{X}}_{QR,d}$ denote the number of operations to compute the Least-Square estimate $\boldsymbol{\delta}\hat{\mathbf{X}}$ with Givens QR of (dense) $\mathbf{J}$ (see Table III). Then

$$\begin{aligned}\boldsymbol{\delta}\hat{\mathbf{X}}_{QR,d} &= 2q^2\left(p - \frac{q}{3}\right)\\ &= 2n_e^3(L^3 + 3L^2 + 3L + 1)\\ &\quad + 3n_e^2 n_m(L^3 + 2L^2 + L).\end{aligned}$$

Therefore the computation for a single iteration of the dense $\boldsymbol{\delta}\mathbf{X}$ using Givens QR is $\mathcal{O}(n_e^3 L^3)$.

C. Sparse Jacobian

Let $\mathbf{J}_{QR,s}$, $\mathbf{r}_{QR,s}$, and $\boldsymbol{\delta}\mathbf{X}_{QR,s}$, denote the number of operations to compute the QR factorization of (sparse) $\mathbf{J}$, $\mathbf{r}$, and $\boldsymbol{\delta}\mathbf{X}$, respectively. The components of the sparse Jacobian are,

$$\begin{aligned}\mathbf{J}_{\mathbf{P}_0} &\mapsto \left(\frac{n_e^3}{3} + \frac{n_e^2 - n_e}{2} + n_e\right)\\ \mathbf{J}_{\mathbf{Qd}} &\mapsto \left(\frac{n_e^3}{3} + \frac{n_e^3 - n_e^2}{2} + n_e^2 + \frac{n_e^2 - n_e}{2} + n_e\right)L\\ \mathbf{J}_{\mathbf{R}} &\mapsto \left(\frac{n_m^3}{3} + 3n_m\right)L.\end{aligned}$$

The number of operations to compute $\mathbf{J}$ is

$$\begin{aligned}\mathbf{J}_{QR,s} &= \frac{n_e^3}{3} + \frac{n_e^2 - n_e}{2} + n_e\\ &\quad + \left(\frac{n_e^3}{3} + \frac{n_e^3 - n_e^2}{2} + n_e^2 + \frac{n_e^2 - n_e}{2} + n_e\right)L\\ &\quad + \left(\frac{n_m^3}{3} + 3n_m\right)L.\end{aligned}$$

The components of the residual vector are,

$$\begin{aligned}\mathbf{r}_{\mathbf{P}_0} &\mapsto \left(\frac{n_e^2 - n_e}{2} + n_e\right)\\ \mathbf{r}_{\mathbf{Qd}} &\mapsto \left(\frac{n_e^2 - n_e}{2} + n_e\right)L\\ \mathbf{r}_{\mathbf{R}} &\mapsto n_m L.\end{aligned}$$

The number of operations to compute $\mathbf{r}$ is

$$\mathbf{r}_{QR,s} = \frac{n_e^2 - n_e}{2} + n_e + \left(\frac{n_e^2 - n_e}{2} + n_e + n_m\right)L$$

Using a sparse library [4], the LS solution via QR factorization is proportional to the number of non-zero elements in the Jacobian (see [3] and [4]). Therefore

$$\boldsymbol{\delta}\mathbf{X}_{QR,s} = n_e(L+1) + n_m L$$

Define $\boldsymbol{\delta}\mathbf{X}_{QR,TO}$ as the total number of operations to compute $\boldsymbol{\delta}\mathbf{X}$ by the sparse QR factorization.

$$\begin{aligned}\boldsymbol{\delta}\mathbf{X}_{QR,TO} &= \boldsymbol{\delta}\mathbf{X}_O + \mathbf{r}_O + \mathbf{J}_O\\ &= n_e^3\left(\frac{1}{3} + \frac{1}{3}L\right)\\ &\quad + n_e^2\left(\frac{n_e - 1}{2}L + 1\right)\\ &\quad + n_e(3L + 3)\\ &\quad + n_m\left(\frac{1}{3}n_m^2 L + 5L\right).\end{aligned}$$

Therefore the computation for a single iteration of the sparse $\boldsymbol{\delta}\mathbf{X}$ using QR factorization is $\mathcal{O}(n_e^3 L)$.

V. Discussion

A. Normal Equations vs. QR

In Section 5.3.8 of [3], the authors identify the challenge with selecting the "right" algorithm to solve a linear system of equations.

- Condition Number: while the Normal Equations effectively square the condition number, QR does not.
- Arithmetic: the Normal Equations use roughly half of the arithmetic as QR when $m \gg n$, while also requiring less storage (memory).
- Applicability: unlike the Normal Equations, QR is more widely applicable because the errors that arise in $\mathbf{A}^\intercal\mathbf{A}$ of the Normal Equations "break down" more quickly than the process on $\mathbf{Q}^\intercal\mathbf{A} = \mathbf{R}$ of LS by QR.

B. Sparse vs. Dense

Considering only LS solutions using QR, Sections IV-B and IV-C clearly define the cost savings using sparse matrices which are $\mathcal{O}(n_e^3 L)$, versus dense matrices which are $\mathcal{O}(n_e^3 L^3)$.

However, on modern computers with dual-precision (64-bit) floating point processors, the primary factor in computational cost is not FLOPs, rather in the storage and memory access of the data [4]. Identifying the computational complexity is important to understand the amount of storage required by a particular algorithm, as sparse representations use less memory and require fewer memory accesses.

C. Estimator Comparison

The computational cost for the calculation of $\boldsymbol{\delta}\mathbf{X}$ in the Iterated Extended KF (IEKF), e.g. CRT with $L = 1$, is $\mathcal{O}(n_e^3)$, whereas the most expensive computation of the KF is the state covariance propagation, $\mathcal{O}(n_e^3)$. In contrast, the CRT estimator, with window length L, scales linearly in L with $\mathcal{O}(n_e^3 L)$ when using sparse matrices.

REFERENCES

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

[2] P. F. Roysdon and J. A. Farrell, "GPS-INS Outlier Detection and Elimination using a Sliding Window Filter," *American Control Conference, In Presc.*, 2017.

[3] G. H. Golub and C. F. van Loan, *Matrix Computations*, 3rd ed. The Johns Hopkins University Press, 1996.

[4] J. R. Gilbert, C. Moler, and R. Schreiber, "Sparse matricies in Matlab: Design and Implementation," *The Mathworks Incorporated*, 1991.

[5] P. D. Groves, *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems*. Artech House, 2013.

Technical Note: CRT-HT Derivations & Proofs

Paul F. Roysdon[†] Jay A. Farrell[‡]

I. Introduction

In this Technical Note we consider the hypothesis testing method used at each update of the CRT estimator. First the general theory in fault detection and removal (FDR) is presented, and then proofs for key steps is provided.

We make the assumption that the reader is already familiar with the CRT method. For clarity, only the final results are repeated here:

The Gauss-Newton step of the nonlinear (iterated) optimization problem can be derived by solving the normal equation,

$$\mathbf{J}^\intercal \mathbf{J} \delta \mathbf{X} = \mathbf{J}^\intercal \mathbf{r}, \tag{1}$$

where $\mathbf{J}$ is the normalized Jacobian matrix, $\mathbf{r}$ is the normalized residual vector, and $\delta\mathbf{X}$ is the measurement correction to be computed. For some noise $\boldsymbol{\eta}$, equation (1) can be rearranged to yield

$$\mathbf{r}(\mathbf{X}) = -\mathbf{J}\delta\mathbf{X} + \boldsymbol{\eta}, \tag{2}$$

where $\mathbf{X}$ is the state vector for the entire CRT window.

II. FDR Background

Beard is credited for early work in the field of outlier detection in state estimation systems. Beard, in his 1971 PhD thesis [1], defines two stages in FDR, namely residual generation and decision making.

Residual generation can be performed by many methods common to state estimation, e.g. least-squares [2]–[4], recursive least-squares [5], or parity space [6]–[15].

The most common challenge to outlier detection is the issue of multiple outliers in a given dataset. Hampel estimates that a routine dataset contain 1-10% (or more) outliers [16]. A study, performed by Rousseeuw [17], identified that in many cases outliers go unnoticed, causing serious effects in estimation or model selection. The most popular statistical modeling method is least-squares [17], which is very sensitive to outliers, so sensitive in-fact a single outlier may cause the estimate to fail completely.

A common method for outlier detection is the *leave-one-out* approach [18]. In this iterative approach, the *i*'th measurement is left out, and the residual is generated for the remaining measurements. If there are multiple outliers in the remaining measurements, and the remaining outliers result in a large standard deviation, then the *i*'th measurement does not appear as a outlier; this is called *masking* [18]. If the outliers cause the norm of the residual to be large for the non-outlier case of the *i*'th measurement, the result is called *swamping* [18]. The effects of masking and swamping are the primary challenge to multiple outlier detection, making the selection of the decision making process a crucial step in FDR.

The decision making process can be performed by a variety of methods. In 1925, Fisher proposed a hypothesis test to evaluate measurements against their expected value, for a fixed threshold [19]. After Fisher, many threshold tests were developed to evaluate measurements against a *normal* distribution [20], a *hypothesized* distribution [21], or an *empirically derived* distribution [22], [23]. However it was Neyman and Pearson who identified the limitations of a fixed threshold, particularly if based on an incorrect distribution. They proposed the "Neyman-Pearson Lemma;" an adaptive threshold hypothesis test, based on a generalized likelihood ratio test (GLRT) [10], [24]. (Derivations of both the NP-Lemma and the GLRT are provided in Sections IV-G and IV-H, respectively.)

Recently, Aggarwal published a text on outlier analysis [25], introducing new methods from the field of Robust Statistics, which involves alternative loss (cost) functions, and thresholding. For detection of multiple outliers simultaneously, Huber proposed a class of robust or resistant regressors based on maximum likelihood methods, called *M*-estimators [26], while Rousseeuw proposed a method based on Least Trimmed Squares (LTS) [17], and Serbert described a clustering-based technique [27].

Finally, robustness to faulty measurement can be achieved by both sensor and analytical redundancy [28]–[36], with further improvements to modeling (less truncation of higher order terms in Taylor series expansions of nonlinear functions) or initialization errors, by using nonlinear optimization [8], [37], [38].

III. 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 eqns. (1) and (2).

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

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

$$\mathbf{y} = \mathbf{H}\boldsymbol{x} + \boldsymbol{\eta}, \tag{3}$$

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

$$\mathbf{y} = \mathbf{H}\boldsymbol{x} + \boldsymbol{\eta} + \mu_i \mathbf{e}_i. \tag{4}$$

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

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., $\mu_i = 0$.

The number of *degrees-of-freedom* (DOF) is $(m-n)$, the difference between the total number of measurements m and the number of unknown variables n. the number of DOF indicates the redundancy of the data, where a larger DOF may yield a higher capability to detect outliers.

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

From eqn. (3), derive the minimum-variance unbiased estimator (MVUE) [39] as

$$\hat{\boldsymbol{x}} = (\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{y}. \tag{5}$$

To analyze the effect of $\boldsymbol{\eta}$, substitute eqn. (3) into eqn. (5)

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

where $\mathbf{H}^* \triangleq (\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal$. The state error due to noise is

$$\begin{aligned}\delta\boldsymbol{x} &= \boldsymbol{x} - \hat{\boldsymbol{x}},\\ &= \mathbf{H}^*\boldsymbol{\eta}.\end{aligned}$$

From the Gaussian noise assumption, the expected value of the state error is

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

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

$$\begin{aligned}\mathbf{r} &= \mathbf{y} - \hat{\mathbf{y}} \qquad (6)\\ &= \mathbf{H}\boldsymbol{x} + \boldsymbol{\eta} - \mathbf{H}\hat{\boldsymbol{x}}\\ &= \mathbf{H}\boldsymbol{x} + \boldsymbol{\eta} - \mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal(\mathbf{H}\boldsymbol{x} + \boldsymbol{\eta})\\ &= \mathbf{H}\boldsymbol{x} + \boldsymbol{\eta} - \mathbf{H}\boldsymbol{x} - \mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\boldsymbol{\eta}\\ &= (\mathbf{I} - \mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal)\boldsymbol{\eta}\\ &= (\mathbf{I} - \mathbf{P})\boldsymbol{\eta},\end{aligned}$$

where $\mathbf{P} \triangleq \mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal$ is the projection matrix onto the range-space of $\mathbf{H}$, i.e. $\mathcal{C}(\mathbf{H})$, which is symmetric, idempotent, and $\text{rank}(\mathbf{P}) = n$ (see Section IV-B).

Let the real, symmetric, idempotent matrix $\mathbf{Q} \triangleq (\mathbf{I} - \mathbf{P})$, where $\mathbf{Q}$ is a projection matrix onto the left null-space of $\mathbf{H}$, i.e. $LN(\mathbf{H}) = N(\mathbf{H}^\intercal)$, with eigenvalues equal to 0 or 1, and the trace is equal to the number of non-zero eigenvalues. The rank of $\mathbf{Q}$ can be derived through the singular value decomposition (SVD) (see Section IV-D).

The mean and covariance of the residual are

$$\begin{aligned}\mathrm{E}\left\langle \mathbf{r}\right\rangle &= \mathrm{E}\left\langle \mathbf{Q}\boldsymbol{\eta}\right\rangle\\ &= \mathbf{Q}\ \mathrm{E}\left\langle \boldsymbol{\eta}\right\rangle\\ &= \mathbf{0} \qquad (7)\\ \mathrm{Cov}\left\langle \mathbf{r}\right\rangle &= \mathrm{E}\left\langle (\mathbf{r} - \mathrm{E}\left\langle \mathbf{r}\right\rangle)(\mathbf{r} - \mathrm{E}\left\langle \mathbf{r}\right\rangle)^\intercal\right\rangle\\ &= \mathrm{E}\left\langle \mathbf{r}\mathbf{r}^\intercal\right\rangle\\ &= \mathrm{E}\left\langle (\mathbf{Q}\boldsymbol{\eta})(\mathbf{Q}\boldsymbol{\eta})^\intercal\right\rangle\\ &= \mathrm{E}\left\langle \mathbf{Q}\boldsymbol{\eta}\boldsymbol{\eta}^\intercal\mathbf{Q}^\intercal\right\rangle\\ &= \mathbf{Q}\mathrm{E}\left\langle \boldsymbol{\eta}\boldsymbol{\eta}^\intercal\right\rangle\mathbf{Q}^\intercal\\ &= \mathbf{Q}(\boldsymbol{\sigma}_y^2\mathbf{I})\mathbf{Q}^\intercal\\ &= \boldsymbol{\sigma}_y^2\mathbf{Q}. \qquad (8)\end{aligned}$$

The final step is achieved because $\mathbf{Q}$ is idempotent, $\mathbf{Q}\mathbf{Q}^\intercal = \mathbf{Q}$ (see Section IV-C). The mean square error (MSE) [39] is

$$\begin{aligned}\mathrm{E}\left\langle \|\mathbf{r}\|^2\right\rangle &= \mathrm{E}\left\langle \mathbf{r}^\intercal\mathbf{r}\right\rangle = \mathrm{E}\left\langle tr\{\mathbf{r}\mathbf{r}^\intercal\}\right\rangle\\ &= \mathrm{E}\left\langle tr\{\mathbf{Q}\boldsymbol{\eta}\boldsymbol{\eta}^\intercal\mathbf{Q}^\intercal\}\right\rangle\\ &= \mathrm{E}\left\langle tr\{\boldsymbol{\eta}\boldsymbol{\eta}^\intercal\mathbf{Q}\}\right\rangle\\ &= tr\{\mathrm{E}\left\langle \boldsymbol{\eta}\boldsymbol{\eta}^\intercal\right\rangle\mathbf{Q}\}\\ &= tr\{\mathbf{Q}\}\boldsymbol{\sigma}_y^2\\ &= (m-n)\boldsymbol{\sigma}_y^2, \qquad (9)\end{aligned}$$

where $\mathrm{E}\left\langle \boldsymbol{\eta}\boldsymbol{\eta}^\intercal\right\rangle = \boldsymbol{\sigma}_y^2\mathbf{I}$, and $tr\{\cdot\}$ is the trace operator. By the Idempotent Rule for the matrix $\mathbf{Q}$, $tr\{\mathbf{Q}\} = \text{rank}(\mathbf{Q}) = m - n$. The last step is from the SVD of $\mathbf{Q}$.

Suppose that the MAP optimization in eqn. (1) converges to an optimal estimate $(\hat{\mathbf{X}})$. From eqn. (6) we can design a test to detect outliers by evaluating the norm of $\mathbf{r}$ and normalizing it by the MSE. Under normal conditions, this test will evaluate to 1. We can define the a test statistic [20]–[23], $\boldsymbol{\Gamma}_{\hat{\mathbf{X}}}$ based on eqns. (6) and (9), such that

$$\begin{aligned}\boldsymbol{\Gamma}_{\hat{\mathbf{X}}} &= \frac{\|\mathbf{r}(\hat{\mathbf{X}})\|^2}{\mathrm{E}\left\langle \|\mathbf{r}(\hat{\mathbf{X}})\|^2\right\rangle}\\ &= \frac{\|\mathbf{r}(\hat{\mathbf{X}})\|^2}{(m-n)\boldsymbol{\sigma}_y^2}, \qquad (10)\end{aligned}$$

where m is the total number of measurements (constraints), and n is the total number of variables to be estimated. This is simply the mean-square weighted deviation, also known as the reduced chi-square statistic [40].

In the variance analysis of general least-square estimation, eqn. (10) is also referred to as the *a-posteriori* variance factor for least-square adjustment [41]. Under normal conditions (no outliers), $\boldsymbol{\Gamma}_{\hat{\mathbf{X}}}$ is referred to as the *a-posteriori* variance of unit weight [30], [41].

The test statistic in eqn. (10) is quadratic in $\mathbf{r}$, and therefore is Chi-square distributed. To detect the existence of outliers, the test statistic calculated by eqn. (10) is tested against the one-tailed Chi-square test [40] with respect to a

significance level α [19], [40], normalized by the number of DOF

$$\mathbf{\Gamma}_{\hat{\mathbf{X}}} < \frac{\boldsymbol{\chi}^2_{\alpha/2,(m-n)}}{(m-n)}. \tag{11}$$

In practice, an outlier will result in values of $\mathbf{\Gamma}_{\hat{\mathbf{X}}} > 1$, therefore we test for values in the shaded region of Fig. 1.

The value for $\boldsymbol{\chi}^2_{\alpha/2,(m-n)}$ is determined from a look-up table for α versus DOF [19], [40]. The significance level α is chosen by the designer for some probability of success. For example, $\alpha = 0.05$ indicates a 95% confidence level.

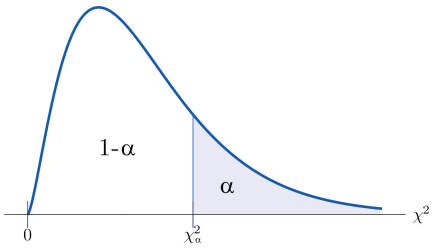

Fig. 1. Chi-square distribution. The shaded area represents the region defined by the significance level α for the test described by right side of eqn. (11).

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 , $\mathcal{H}_i$

By eqn. (4) the outlier is deterministic, and it can be shown that the residual lies in the left-null-space (the Residual Space) of $\mathbf{H}$ (the proof is provided in Section IV-F).

To analyze the effect of $\boldsymbol{\varepsilon}_i$ on the state error, substitute eqn. (4) into eqn. (5)

$$\begin{aligned}\hat{\boldsymbol{x}} &= ((\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal)(\mathbf{H}\boldsymbol{x} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}_i)\\ &= (\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{H}\boldsymbol{x} + (\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\boldsymbol{\eta} + \boldsymbol{\varepsilon}_i\\ &= \mathbf{I}\boldsymbol{x} + (\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal(\boldsymbol{\eta} + \boldsymbol{\varepsilon}_i).\end{aligned} \tag{12}$$

Therefore,

$$\delta\boldsymbol{x} = \mathbf{H}^*(\boldsymbol{\eta} + \boldsymbol{\varepsilon}_i).$$

The expected value of the state error due to the outlier is

$$\mathrm{E}\left\langle\delta\boldsymbol{x}\right\rangle = \mathbf{H}^*\boldsymbol{\varepsilon}_i.$$

To analyze the effect of the outlier on the residual, substitute (4) and (12) into (6):

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

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

$$\begin{aligned}\mathrm{E}\left\langle\mathbf{r}\right\rangle &= \mathrm{E}\left\langle\mathbf{Q}(\boldsymbol{\eta} + \boldsymbol{\varepsilon}_i)\right\rangle\\ &= \mathbf{Q}\boldsymbol{\varepsilon}_i\end{aligned} \tag{13}$$

$$\begin{aligned}\mathrm{Cov}\left\langle\mathbf{r}\right\rangle &= \mathrm{E}\left\langle(\mathbf{r} - \mathrm{E}\left\langle\mathbf{r}\right\rangle)(\mathbf{r} - \mathrm{E}\left\langle\mathbf{r}\right\rangle)^\intercal\right\rangle\\ &= \mathrm{E}\left\langle(\mathbf{r} - \mathbf{Q}\boldsymbol{\varepsilon}_i)(\mathbf{r} - \mathbf{Q}\boldsymbol{\varepsilon}_i)^\intercal\right\rangle\\ &= \mathrm{E}\left\langle(\mathbf{Q}\boldsymbol{\eta} + \mathbf{Q}\boldsymbol{\varepsilon}_i - \mathbf{Q}\boldsymbol{\varepsilon}_i)(\mathbf{Q}\boldsymbol{\eta} + \mathbf{Q}\boldsymbol{\varepsilon}_i - \mathbf{Q}\boldsymbol{\varepsilon}_i)^\intercal\right\rangle\\ &= \mathrm{E}\left\langle(\mathbf{Q}\boldsymbol{\eta})(\mathbf{Q}\boldsymbol{\eta})^\intercal\right\rangle\\ &= \mathrm{E}\left\langle\mathbf{Q}\boldsymbol{\eta}\boldsymbol{\eta}^\intercal\mathbf{Q}^\intercal\right\rangle\\ &= \mathbf{Q}\mathrm{E}\left\langle\boldsymbol{\eta}\boldsymbol{\eta}^\intercal\right\rangle\mathbf{Q}^\intercal\\ &= \mathbf{Q}\boldsymbol{\sigma}_y^2\mathbf{I}\mathbf{Q}^\intercal\\ &= \boldsymbol{\sigma}_y^2\mathbf{Q}.\end{aligned} \tag{14}$$

Comparing eqn. (8) with eqn. (14), 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. (7) and (13), and demonstrated in Fig. 2. 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 $\mathbf{r} \sim \mathcal{N}(\mathbf{Q}\boldsymbol{\varepsilon}_i, \boldsymbol{\sigma}_y^2\mathbf{Q})$.

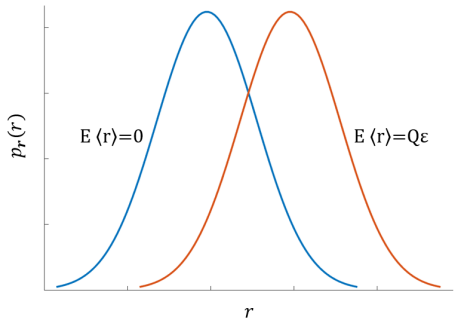

Fig. 2. Assume a normal distribution for two sets of random variables with equal variance. One set contains an outlier and the other set does not. The difference between the two sets will be a shift of the mean in the set which contains the outlier.

Consider the parity vector [11], [42]

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

where $\mathbf{U}_2^\intercal$ is defined in Section IV-D. By analysis

$$\begin{aligned}\mathbf{p} &= \mathbf{U}_2^\intercal\mathbf{r}\\ &= \mathbf{U}_2^\intercal(\mathbf{y} - \mathbf{H}\hat{\boldsymbol{x}})\\ &= \mathbf{U}_2^\intercal\mathbf{Q}(\boldsymbol{\eta} + \mu_i\mathbf{e}_i)\\ &= \mathbf{U}_2^\intercal(\mathbf{U}_2\mathbf{U}_2^\intercal)(\boldsymbol{\eta} + \mu_i\mathbf{e}_i)\\ &= \mathbf{U}_2^\intercal(\boldsymbol{\eta} + \mu_i\mathbf{e}_i)\\ &= (\mathbf{U}_2^\intercal\mathbf{e}_i)\mu_i + \mathbf{U}_2^\intercal\boldsymbol{\eta},\end{aligned}$$

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

Then the magnitude of the outlier μ_i in eqn. (4) can be

estimated as (see Section 5 of [40]),

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

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

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

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. (1) is repeated.

C. Comparison of DOF's

This section theoretically compares a few algorithms in terms of their available number of degrees of freedom $(m - n)$ for outlier detection. In this section, the state dimension is denoted by n_s. The variable m_k is the number of satellite pseudoranges available at epoch k. The variable n denotes the total number of real variables to be estimated. The variable 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 only 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

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 m residuals in each GPS epoch, there are

$$\sum_{k=1}^{m} \left(\begin{array}{c} m \\ k \end{array} \right) = \sum_{k=1}^{m} \frac{m!}{(m-k)!\ k!}$$

ways in which k outliers could occur (see Section 3 of [43]). For example: if a maximum of 9 satellites are available after double-differencing, $m = 9$ and $k = 5$, results in 511 possible combinations required at each epoch. For the CRT window $L = 20$, this requires $10,220$ possible combinations for each window, which is not feasible in real-time. Therefore, simplified approaches to hypotheses testing are required, such as iteratively removing the row with the largest residual and re-optimizing until all residuals pass a threshold test.

E. Application to GPS-INS

In [42], the Parity Vector method used in RAIM, is shown to be equivalent to the Residual Space method presented in Section III-A. The Parity Vector for GPS outlier detection is extensively researched [5], [13]–[15], [44]–[50], and based on the foundational work in geodesy fault detection by Baarda [30], [31]. Early work in RAIM considered only a single satellite failure in one epoch [4], [14], [47], and later extended to multiple failures within a single epoch [48]–[50].

Here we consider three FDR cases: GPS only, GPS-INS, and GPS-INS with a window of measurements.

1) GPS-only FDR: GPS positioning requires a minimum of four satellites to solve both position and clock errors [51]. If RAIM methods are to be used in a single epoch linear system, a minimum of five satellites are necessary [14]. Therefore, in the case where only five satellites are available, the DOF in eqn. (11) is equal to 1. If outliers exist in more than one satellite residual, a 3D position calculation is not possible.

A 3D position can be calculated from GPS pseudoranges using iterative least-squares, see Section 8.2.2 of [51]. Assuming GPS-only positioning, the residuals vector generated by RAIM is equivalent to a single row of $\mathbf{r}$ in eqn. (1), e.g.

$$\boldsymbol{\Sigma}_{\mathbf{R}_\rho}^{1} \left(\mathbf{h}_1^1 \left(\hat{\boldsymbol{x}}(t_1) \right) - \boldsymbol{\rho}^1(t_1) \right).$$

The limited redundancy within a single epoch of GPS-only residuals, severely limits the capability to correctly detect outliers, as discussed in Section III-A. Additional redundancy can be gained by including INS measurements.

2) GPS-INS FDR: Recently, [15], [52], [53] proposed a method called extended RAIM (eRAIM), utilizing the residuals of the measurement update step in an Extended Kalman Filter (EKF), see Section 5.4 of [51], where the *a priori* residuals vector is defined as $\mathbf{r} = \tilde{\mathbf{y}} - \hat{\mathbf{y}}$. Thus for a single GPS epoch, eRAIM provides additional DOF, and therefore increased robustness to outliers.

In the CRT framework where $L = 1$, and optimization is limited to one iteration, the CRT estimator is equivalent to the EKF. Under these constraints, the standard method

for eRAIM utilize a prior, as well as the IMU and GPS measurements of $\mathbf{r}$ in eqn. (1), e.g.

$$\begin{aligned}&\boldsymbol{\Sigma}_{\mathbf{P}_0}\left(\hat{\boldsymbol{x}}(t_0)-\boldsymbol{x}_0\right),\\&\boldsymbol{\Sigma}_{\mathbf{Q}_0}\left(\bar{\boldsymbol{x}}_1-\hat{\boldsymbol{x}}_1\right),\\&\boldsymbol{\Sigma}_{\mathbf{R}_\rho}^{\frac{1}{1}}\left(\mathbf{h}_1^1\big(\hat{\boldsymbol{x}}(t_1)\big)-\boldsymbol{\rho}^1(t_1)\right).\end{aligned}$$

However, eRAIM is still limited to a single epoch of measurements. If fewer than five satellites are available after FDR, the EKF will not perform an optimal correction. Additional redundancy can be gained by considering a temporal window of measurements.

3) GPS-INS window FDR: We propose a novel method of outlier detection and removal based on the CRT estimator, we call interval RAIM (iRAIM). Leveraging the full window of *a posteriori* measurement residuals, e.g. all rows of $\mathbf{r}$ in eqn. (1), iRAIM has several advantages over current RAIM and eRAIM techniques:

- In contrast to the EKF, the CRT estimator has the advantage that both current and past GPS-INS measurements can evaluated, or re-evaluated and removed if necessary. This method enables the unique opportunity to re-evaluate old measurements against new measurements, for the purpose of reducing the likelihood of outlier masking and swamping effects.
- By solving the full nonlinear MAP, the potential of false alarms is minimized because the residuals are the result of an optimized solution.
- By increasing the CRT window length, the DOF in eqn. (11) can be maximized, thereby reducing the probability of false alarm and increasing the probability of outlier detection.
- For GPS epochs within the CRT window, where fewer than five satellites are available after outlier FDR, the CRT estimator (unlike the EKF) yields the optimal result.
- The numerical implementation of the CRT window, can be easily modified to utilize a cost function which combines an L-1 and L-2 norm, with soft-thresholding, as proposed in [54]. This has the advantage that the outliers are retained and magnitude-constrained, instead of removed, and the exhaustive search of the residuals vector in Section III-B is eliminated. This is an area of our ongoing research.

IV. Proofs

A. Proof of idempotent P

For the matrix $\mathbf{P}$ to be idempotent, it must be the case that $\mathbf{P}=\mathbf{P}^\intercal\mathbf{P}=\mathbf{PP}$, where $\mathbf{P}\triangleq\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal$, and $\mathbf{H}\in\mathbb{R}^{m\times n}$, with $m>n$. Thus we can show:

$$\begin{aligned}\mathbf{P}^\intercal&=(\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal)^\intercal\\&=\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\\&=\mathbf{P}\\\mathbf{PP}&=\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\\&=\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\\&=\mathbf{P}\\\therefore\mathbf{P}^\intercal\mathbf{P}&=\mathbf{PP}=\mathbf{P}.\end{aligned}$$

■

B. Proof of rank P

We can prove that $\text{rank}(\mathbf{P})=n$. First recall that $\mathbf{H}\in\mathbb{R}^{m\times n}$, with $m>n$ and full column rank, i.e. $\text{rank}(\mathbf{H})=n$. Let the SVD of $\mathbf{H}$ be defined as

$$\begin{aligned}\mathbf{H}&=\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\intercal\\&=[\mathbf{U}_1,\mathbf{U}_2]\left[\begin{pmatrix}\boldsymbol{\Sigma}_1\\\boldsymbol{\Sigma}_0\end{pmatrix}\right]\mathbf{V}^\intercal\end{aligned}$$

where $\boldsymbol{\Sigma}\in\mathbb{R}^{m\times m}$, $\boldsymbol{\Sigma}_1=\text{diag}(\sigma_1,\ldots,\sigma_n)\in\mathbb{R}^{n\times n}$, and $\boldsymbol{\Sigma}_0=\mathbf{0}\in\mathbb{R}^{(m-n)\times n}$, where σ_i for $i=1,\ldots,n$ are the singular values of $\mathbf{H}$. Both $\mathbf{U}\in\mathbb{R}^{m\times m}$ and $\mathbf{V}\in\mathbb{R}^{n\times n}$ are unitary matrices, therefore $\mathbf{UU}^\intercal=\mathbf{U}^\intercal\mathbf{U}=\mathbf{I}\in\mathbb{R}^{m\times m}$ and $\mathbf{VV}^\intercal=\mathbf{V}^\intercal\mathbf{V}=\mathbf{I}\in\mathbb{R}^{m\times m}$. The columns of $\mathbf{U}_1\in\mathbb{R}^{m\times n}$ form an orthonormal basis for the range-space of $\mathbf{H}$, and the columns of $\mathbf{U}_2\in\mathbb{R}^{m\times(m-n)}$ form the null-space of $\mathbf{H}^\intercal$. Similarly the first n columns of $\mathbf{V}$ form an orthonormal basis for the range of $\mathbf{H}^\intercal$, and the $m-n$ columns of $\mathbf{V}$ form an orthonormal basis for the null-space of $\mathbf{H}$. Finally, the eigenvectors $\mathbf{V}$ of the matrix $\mathbf{H}^\intercal\mathbf{H}$ are the right singular values of $\mathbf{H}$, and the singular values of $\mathbf{H}$ squared are the corresponding nonzero eigenvalues: $\sigma_i=\sqrt{\lambda_i(\mathbf{H}^\intercal\mathbf{H})}$. Similarly, the eigenvectors of $\mathbf{HH}^\intercal$ are the left singular vectors $\mathbf{U}$ of matrix $\mathbf{H}$, and the singular values of $\mathbf{H}$ squared are the nonzero eigenvalues of $\mathbf{HH}^\intercal$: $\sigma_i=\sqrt{\lambda_i(\mathbf{HH}^\intercal)}$.

Define $\mathbf{P}$ in terms of the SVD of $\mathbf{H}$:

$$\begin{aligned}\mathbf{P}&=\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\\&=(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\intercal)(\mathbf{V}\boldsymbol{\Sigma}^\intercal\mathbf{U}^\intercal\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^t)^{-1}(\mathbf{V}\boldsymbol{\Sigma}^\intercal\mathbf{U}^\intercal)\\&=(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\intercal)(\mathbf{V}\boldsymbol{\Sigma}^\intercal\boldsymbol{\Sigma}\mathbf{V}^\intercal)^{-1}(\mathbf{V}\boldsymbol{\Sigma}^\intercal\mathbf{U}^\intercal)&(15)\\&=(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\intercal)(\mathbf{V}\boldsymbol{\Sigma}_1^2\mathbf{V}^\intercal)^{-1}(\mathbf{V}\boldsymbol{\Sigma}^\intercal\mathbf{U}^\intercal)&(16)\\&=(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\intercal)(\mathbf{V})^{-1}(\boldsymbol{\Sigma}_1^2)^{-1}(\mathbf{V}^\intercal)^{-1}(\mathbf{V}\boldsymbol{\Sigma}^\intercal\mathbf{U}^\intercal)&(17)\\&=\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\intercal\mathbf{V}\boldsymbol{\Sigma}_1^{-2}\mathbf{V}^\intercal\mathbf{V}\boldsymbol{\Sigma}^\intercal\mathbf{U}^\intercal&(18)\\&=\mathbf{U}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_1^{-2}\boldsymbol{\Sigma}_1^\intercal\mathbf{U}^\intercal&(19)\\&=\mathbf{U}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_1^\intercal\mathbf{U}^\intercal&(20)\\&=\mathbf{U}\mathbf{I}_{n\times n}\mathbf{U}^\intercal\\&=[\mathbf{U}_1\mathbf{U}_2]\begin{bmatrix}\mathbf{I}_{n\times n}&\mathbf{0}\\\mathbf{0}&\mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{U}_1^\intercal\\\mathbf{U}_2^\intercal\end{bmatrix}\\&=\mathbf{U}_1\mathbf{U}_1^\intercal.\end{aligned}$$

The middle product in eqn. (16) can be separated because it is an $n\times n$ matrix with rank n, and it is non-singular. In eqns. (16)-(20), we need only consider $\boldsymbol{\Sigma}_1$ as $\boldsymbol{\Sigma}_0$ drops out.

The rank of matrix $\mathbf{P}$ is defined as the number of non-zero singular values of $\mathbf{P}$. Thus, $\text{rank}(\mathbf{P}) = n$. Similarly, because $\mathbf{P}$ is idempotent, $\text{rank}(\mathbf{P}) = tr(\mathbf{P})$, then $\text{rank}(\mathbf{P}) = n$. ■

C. Proof of idempotent Q

For the matrix $\mathbf{Q}$ to be idempotent, it must be the case that $\mathbf{Q} = \mathbf{Q}^\intercal\mathbf{Q} = \mathbf{Q}\mathbf{Q}$, where $\mathbf{Q} \triangleq (\mathbf{I} - \mathbf{P})$, and $\mathbf{P} \in \mathbb{R}^{m\times m}$. Thus we can show:

$$\begin{aligned}
\mathbf{Q}\mathbf{Q} &= (\mathbf{I}-\mathbf{P})(\mathbf{I}-\mathbf{P}) \\
&= \mathbf{I}-\mathbf{P} \\
&= \mathbf{Q} \\
\mathbf{Q}^\intercal\mathbf{Q} &= (\mathbf{I}-\mathbf{P})^\intercal(\mathbf{I}-\mathbf{P}) \\
&= (\mathbf{I}-\mathbf{P}^\intercal)(\mathbf{I}-\mathbf{P}) \\
&= \mathbf{I}-\mathbf{P}-\mathbf{P}^\intercal+\mathbf{P}^\intercal\mathbf{P}, \quad \mathbf{P} = \mathbf{P}^\intercal\mathbf{P} \\
&= \mathbf{I}-\mathbf{P}-\mathbf{P}^\intercal+\mathbf{P}, \quad \mathbf{P}^\intercal = \mathbf{P} \\
&= \mathbf{I}-\mathbf{P} \\
&= \mathbf{Q} \\
\therefore \mathbf{Q}^\intercal\mathbf{Q} &= \mathbf{Q}\mathbf{Q} = \mathbf{Q}
\end{aligned}$$

■

D. Proof of rank Q

We can prove that $\text{rank}(\mathbf{Q}) = m - n$ by the SVD of $\mathbf{H}$. Apply the result from the proof for the rank of $\mathbf{P}$ in Section IV-B, where $\mathbf{P} \in \mathbb{R}^{m\times m}$ and $\mathbf{I} \in \mathbb{R}^{m\times m}$. Using the inner product we can define $\mathbf{I}$ in terms of $\mathbf{U}$

$$\begin{aligned}
\mathbf{I} &= \mathbf{U}\mathbf{U}^\intercal \\
&= [\mathbf{U}_1\mathbf{U}_2]\begin{bmatrix}\mathbf{U}_1^\intercal \\ \mathbf{U}_2^\intercal\end{bmatrix} \\
&= \mathbf{U}_1\mathbf{U}_1^\intercal + \mathbf{U}_2\mathbf{U}_2^\intercal.
\end{aligned}$$

Alternatively, by the outer product we can define

$$\begin{aligned}
\mathbf{I} &= \mathbf{U}^\intercal\mathbf{U} \\
&= \begin{bmatrix}\mathbf{U}_1^\intercal \\ \mathbf{U}_2^\intercal\end{bmatrix}[\mathbf{U}_1\mathbf{U}_2] \\
&= \begin{bmatrix}\mathbf{U}_1^\intercal\mathbf{U}_1 & \mathbf{U}_1^\intercal\mathbf{U}_2 \\ \mathbf{U}_2^\intercal\mathbf{U}_1 & \mathbf{U}_2^\intercal\mathbf{U}_2\end{bmatrix}
\end{aligned}$$

where $\mathbf{U}_1^\intercal\mathbf{U}_1 = \mathbf{I} \in \mathbb{R}^{n\times n}$, $\mathbf{U}_2^\intercal\mathbf{U}_2 = \mathbf{I} \in \mathbb{R}^{(m-n)\times(m-n)}$. Finally, $\mathbf{U}_1\mathbf{U}_1^\intercal = \mathbf{P} \in \mathbb{R}^{n\times n}$ as shown in Section IV-B, and $\mathbf{U}_2\mathbf{U}_2^\intercal = \mathbf{Q} \in \mathbb{R}^{m\times m}$ which is derived below.

Now define $\mathbf{Q}$ as

$$\begin{aligned}
\mathbf{Q} &= \mathbf{I}-\mathbf{P} \\
&= (\mathbf{U}_1\mathbf{U}_1^\intercal + \mathbf{U}_2\mathbf{U}_2^\intercal) - \mathbf{U}_1\mathbf{U}_1^\intercal \\
&= \mathbf{U}_2\mathbf{U}_2^\intercal.
\end{aligned}$$

The rank of matrix $\mathbf{Q}$ is defined as the number of non-zero singular values of $\mathbf{Q}$. Thus, for $\mathbf{Q} \triangleq (\mathbf{I}-\mathbf{P})$, and $\text{rank}(\mathbf{P}) = n$, the number of non-zero singular values of $\mathbf{Q}$ is at most $m - n$, and therefore the $\text{rank}(\mathbf{Q}) = m - n$. ■

E. Physical Interpretation of P & Q

The physical interpretation for $\mathbf{P}$ and $\mathbf{Q}$ is a mapping of the measurement and the residual, as shown in Fig. 3. $\mathbf{P}\mathbf{y}$ projects $\mathbf{y}$ onto the range$(\mathbf{P})$ along the direction of $\mathbf{y}$. The complementary projector is $\mathbf{Q}$, where $\mathbf{Q}\mathbf{y}$ projects $\mathbf{y}$ onto the range$(\mathbf{Q})$ which is orthogonal to the range$(\mathbf{P})$.

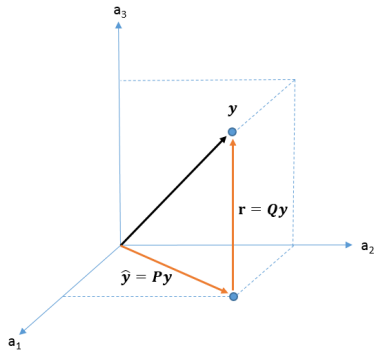

Fig. 3. For a general space in $\mathbb{R}^3$, the mapping $\mathbf{P}\mathbf{y} = \hat{\mathbf{y}}$ is the estimate for $\mathbf{y}$, and $\mathbf{Q}\mathbf{y} = \mathbf{r}$ is the estimation residual for $\mathbf{y}$.

From the SVD of $\mathbf{H}$ we have the relations:

1) $\mathbf{V}_1\mathbf{V}_1^\intercal$ is the orthogonal projector onto $[N(\mathbf{H})]^\perp = R(\mathbf{H}^\intercal)$.
2) $\mathbf{V}_2\mathbf{V}_2^\intercal$ is the orthogonal projector onto $N(\mathbf{H})$.
3) $\mathbf{U}_1\mathbf{U}_1^\intercal$ is the orthogonal projector onto $R(\mathbf{H})$.
4) $\mathbf{U}_2\mathbf{U}_2^\intercal$ is the orthogonal projector onto $[R(\mathbf{H})]^\perp = N(\mathbf{H}^\intercal)$.

F. Proof of left-null-space for r

Claim: The residual must lie in the subspace orthogonal to the column space spanned by $\mathbf{H}$. We know the residual lies in the left-null-space (the Residual Space) of $\mathbf{H}$, where the null-space of $\mathbf{H}$ is defined as

$$N(\mathbf{H}) = \{\boldsymbol{x} \in \mathbb{R}^m : \mathbf{H}\boldsymbol{x} = \mathbf{0}\}.$$

The left-null-space of $\mathbf{H}$ is defined as

$$N(\mathbf{H}^\intercal) = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{H}^\intercal\mathbf{y} = \mathbf{0}\}.$$

Thus, for $\mathbf{r}$ to lie in the left-null-space of $\mathbf{H}$, requires $\mathbf{H}^\intercal\mathbf{r} = \mathbf{0}$. **The proof follows:** First derive $\mathbf{r}$ in eqn. (6) for $\mathbf{y}$

$$\begin{aligned}
\mathbf{r} &= \mathbf{y} - \mathbf{H}\hat{\boldsymbol{x}} \\
&= \mathbf{y} - \mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{y} \\
&= (\mathbf{I} - \mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal)\mathbf{y}.
\end{aligned}$$

Then apply this to the claim for $\mathbf{H}^\intercal\mathbf{r} = \mathbf{0}$.

$$\begin{aligned}
\mathbf{H}^\intercal\mathbf{r} &= \mathbf{H}^\intercal(\mathbf{I} - \mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal)\mathbf{y} \\
\mathbf{H}^\intercal\mathbf{r} &= (\mathbf{H}^\intercal - \mathbf{H}^\intercal\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal)\mathbf{y} \\
\mathbf{H}^\intercal\mathbf{r} &= (\mathbf{H}^\intercal - \mathbf{H}^\intercal)\mathbf{y} \\
\mathbf{H}^\intercal\mathbf{r} &= \mathbf{0}.
\end{aligned}$$

■

G. Definition of the Neyman-Pearson Lemma

Consider two densities $\mathrm{p}(\mathbf{y}|\mathcal{H}_0)$ and $\mathrm{p}(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta}_1)$, for the parameter $\mathbf{y}$ and random variable $\boldsymbol{\theta}$, where $\mathcal{H}_0$ is the *null-hypothesis*, and $\mathcal{H}_1$ is the *alternate-hypothesis*. A constrained optimization problem can be formulated and solved by Lagrange multipliers, to maximize the probability of detection, P_D, of event $\mathcal{H}_1$, given the probability of false alarm, $P_{FA} = \alpha$.

For $\mathbf{y} \in \mathcal{X}$, the subspace $\mathcal{X}_1$, where $\mathcal{H}_1$ is decided, is found by maximizing

$$P_D = \int_{\mathcal{X}_1} \mathrm{p}(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta}) d\mathbf{y}$$

under the constraint

$$P_{FA} = \int_{\mathcal{X}_1} \mathrm{p}(\mathbf{y}|\mathcal{H}_0) d\mathbf{y} = \alpha \tag{21}$$

where $0 < \alpha < 1$.

Define an objective function (the Lagrangian) using the Lagrangian multiplier $\boldsymbol{\gamma}$, such that

$$\begin{aligned}\mathcal{L} &= P_D - \boldsymbol{\gamma}(P_{FA} - \alpha) \\ &= \int_{\mathcal{X}_1} \mathrm{p}(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta}) d\mathbf{y} - \boldsymbol{\gamma} \left[\int_{\mathcal{X}_1} \mathrm{p}(\mathbf{y}|\mathcal{H}_0) d\mathbf{y} - \alpha \right] \\ &= \int_{\mathcal{X}_1} \left[\mathrm{p}(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta}) - \boldsymbol{\gamma} \mathrm{p}(\mathbf{y}|\mathcal{H}_0) \right] d\mathbf{y} + \boldsymbol{\gamma}\alpha.\end{aligned}$$

For any given value $\boldsymbol{\gamma}$, the region $\mathcal{X}_1$ that maximizes $\mathcal{L}$, and hence P_D, under the constraint $P_{FA} = \alpha$, is given by

$$\mathcal{X}_1 = \{\mathbf{y} \in \mathcal{X} \mid \mathrm{p}(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta}) > \boldsymbol{\gamma} \mathrm{p}(\mathbf{y}|\mathcal{H}_0)\}. \tag{22}$$

Equation (22) yields the likelihood ratio test (LRT) [24] which is the *uniformly most powerful* (UMP) test [24] with $P_{FA} = \alpha$,

$$\Lambda(\mathbf{y}) = \frac{\mathrm{p}(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta})}{\mathrm{p}(\mathbf{y}|\mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \boldsymbol{\gamma},$$

where $\boldsymbol{\gamma}$ is determined from the constraint $P_{FA} \leq \alpha$ in eqn. (21).

1) Illustrative Example: Consider the P_D versus the P_{FA} for given threshold $\boldsymbol{\gamma}$ for two signal distributions shown in Fig. 4. Distribution *a* is *noise-only*, while distribution *b* is *signal-plus-noise*. Both distributions in Fig. 4 have the same distance between the peaks. The height of the distributions represent how often a signal is present, and the spread of the distributions indicate the magnitude of noise present; less noise reduces the spread of the distributions, while more noise increases the spread.

Consider the internal response to a signal detection system. In Fig. 5, the vertical line represents the threshold $\boldsymbol{\gamma}$, which splits the figure into four regions: a *hit* is defined as the *signal-plus-noise* region greater than (to the right of) $\boldsymbol{\gamma}$, and a *miss* is the *signal-plus-noise* region less than (to the left of) $\boldsymbol{\gamma}$. *False alarms* represent the *noise-only* region greater than $\boldsymbol{\gamma}$, while *correct rejection* represents the *noise-only* region less than $\boldsymbol{\gamma}$.

Suppose that a low threshold is chosen, then the *signal-plus-noise* will likely be detected, and therefore the system will have a very high hit rate, at the cost of an increased number of false alarms. Conversely, if the threshold is chosen to be high, then the number of false alarms will be reduced, at the cost of an increased miss rate. This is demonstrated by an example of threshold-shifting, shown in Fig. 6.

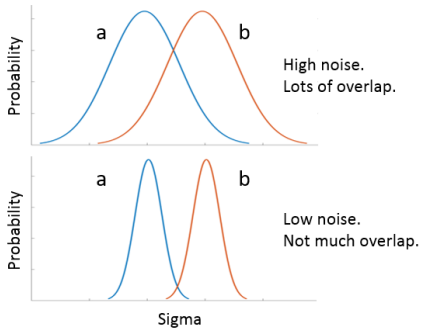

Fig. 4. Signal spread due to noise, and resulting overlap.

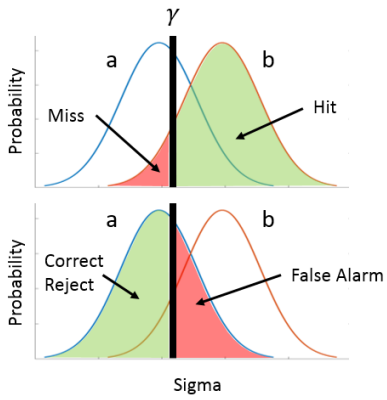

Fig. 5. The four regions of hypothesis testing, defined by two overlapping distributions (*a* & *b*) and the vertical line of the decision threshold $\boldsymbol{\gamma}$.

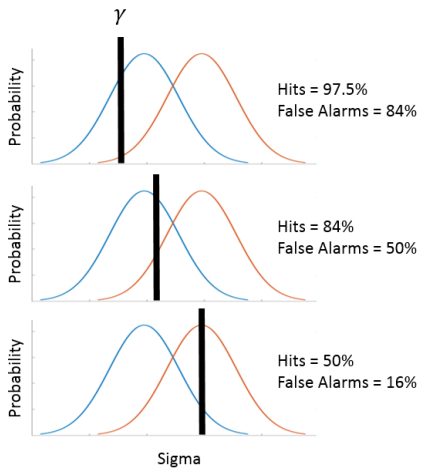

Fig. 6. An example result of *hits* versus *false alarms*, due to shifting the threshold $\boldsymbol{\gamma}$.

Consider the following three conditions:

1) As the region $\mathcal{X}_1$ shrinks ($\boldsymbol{\gamma}$ tends toward infinity),

both P_D and P_{FA} shrink toward zero.
2) As the region $\mathcal{X}_1$ grows (γ tends toward zero), both P_D and P_{FA} grow toward unity.
3) The case where $P_D = 1$ and $P_{FA} = 0$ will never occur if the conditional PDF's $\mathrm{p}(\boldsymbol{x}|\mathcal{H}_0)$ and $\mathrm{p}(\boldsymbol{x}|\mathcal{H}_1, \boldsymbol{\theta})$ overlap as in Fig. 4.

Item 3 in the list above represents the fundamental trade-off in hypothesis testing, and motivates the N-P Lemma.

2) Relation to ROC Plots: In the 1940's, the Allied forces in England sought to make sense of the signals received from new Radar technology [40]. Specifically, they needed a graphical way to represent and determine a good signal from random noise. This graphical method is called the Receiver Operating Characteristic (ROC) plot, with the P_D equal to one minus the Probability of Missed Detection (P_{MD}) on the vertical axis (shown as *Hits* in Fig. 7), versus the P_{FA} on the horizontal axis (shown as *False Alarms* in Fig. 7). The *Discriminability index* (d'), where $d' = 0$, is the *Line-of-No-Discrimination*, indicating that any signal along or below this line cannot be discerned from random noise. The curves above this line, $d' = 1 \cdots 4$, represent a detected signal with varying thresholds and/or discriminability. A perfectly detected signal lies on the vertical axis.

The discriminability of a signal depends both on the separation and the spread of the *noise-only* and *signal-plus-noise* curves. Discriminability is made easier either by increasing the separation (stronger signal) or by decreasing the spread (less noise). The number, d', is often referred to as an estimate of the signal strength [40].

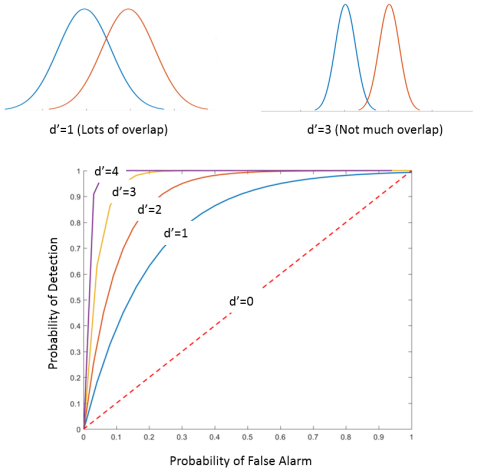

Fig. 7. ROC plot for Hits (P_D) versus False Alarms (P_{FA}).

H. Definition of the GLRT

Consider a test for a signal present in Gaussian additive noise with non-zero mean [55]. A binary test can be performed for a random sample from a population that is normally distributed and has known variance. Based on the Neyman-Pearson (N-P) Lemma for binary hypothesis testing [24], [39], consider

$$\mathcal{H}_0 : \mathbf{y} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}^2\mathbf{I}) \tag{23}$$
$$\mathcal{H}_1 : \mathbf{y} \sim \mathcal{N}(\mathbf{H}\boldsymbol{\theta}, \boldsymbol{\sigma}^2\mathbf{I}) \tag{24}$$

for measurement $\mathbf{y} \in \mathbb{R}^{m\times 1}$, where $\boldsymbol{\sigma}^2 > 0$ is known, $\mathbf{H} \in \mathbb{R}^{m\times n}$ is known, and the unknown $\boldsymbol{\theta} \in \mathbb{R}^{n\times 1}$. The standard, or *null-hypothesis*, with known mean is defined as $\mathcal{H}_0$, and the *alternate-hypothesis* with unknown mean is defined as $\mathcal{H}_1$.

The Likelihood Ratio Test (LRT) [40] compares the model in $\mathcal{H}_1$ to the model in $\mathcal{H}_0$, for threshold $\boldsymbol{\gamma}$, such that

$$\frac{\mathrm{p}(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta})}{\mathrm{p}(\mathbf{y}|\mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \boldsymbol{\gamma},$$

where

$$\mathrm{p}(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta}) = \frac{1}{(2\pi\boldsymbol{\sigma}^2)^{k/2}} e^{\left(-\frac{1}{2\boldsymbol{\sigma}^2}(\mathbf{y}-\mathbf{H}\boldsymbol{\theta})^\intercal(\mathbf{y}-\mathbf{H}\boldsymbol{\theta})\right)}$$
$$\mathrm{p}(\mathbf{y}|\mathcal{H}_0) = \frac{1}{(2\pi\boldsymbol{\sigma}^2)^{k/2}} e^{\left(-\frac{1}{2\boldsymbol{\sigma}^2}(\mathbf{y}^\intercal\mathbf{y})\right)}.$$

When $\mathcal{H}_1$ is decided:

- if $\mathcal{H}_1$ is valid, this is a *correct detection*,
- if $\mathcal{H}_1$ not valid, this is a *false alarm*.

When $\mathcal{H}_0$ is decided:

- if $\mathcal{H}_1$is valid, this is a *missed detection*,
- if $\mathcal{H}_1$ not valid, this is a *correct rejection*.

The log likelihood ratio test is

$$\ln\left(\Lambda(\mathbf{y})\right) = \ln\left(\frac{\mathrm{p}(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta}_1)}{\mathrm{p}(\mathbf{y}|\mathcal{H}_0)}\right) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \boldsymbol{\gamma}' \tag{25}$$

where $\boldsymbol{\gamma}' = \ln(\boldsymbol{\gamma})$.

Defining eqn. (25) in terms of eqn. (23) & (24) yields

$$\begin{aligned}\ln\left(\Lambda(\mathbf{y})\right) &= -\frac{1}{2\boldsymbol{\sigma}^2}\Big((\mathbf{y}-\mathbf{H}\boldsymbol{\theta})^\intercal(\mathbf{y}-\mathbf{H}\boldsymbol{\theta}) - \mathbf{y}^\intercal\mathbf{y}\Big) \\ &= -\frac{1}{2\boldsymbol{\sigma}^2}(-\mathbf{y}^\intercal\mathbf{H}\boldsymbol{\theta} - \boldsymbol{\theta}^\intercal\mathbf{H}^\intercal\mathbf{y} + \boldsymbol{\theta}^\intercal\mathbf{H}^\intercal\mathbf{H}\boldsymbol{\theta}) \\ &= -\frac{1}{2\boldsymbol{\sigma}^2}(-2\boldsymbol{\theta}^\intercal\mathbf{H}^\intercal\mathbf{y} + \boldsymbol{\theta}^\intercal\mathbf{H}^\intercal\mathbf{H}\boldsymbol{\theta}).\end{aligned} \tag{26}$$

The simplification in eqn. (26) is possible because: $\mathbf{y}^\intercal\mathbf{H}\boldsymbol{\theta} = \mathbf{y}\bullet(\mathbf{H}\boldsymbol{\theta}) = (\mathbf{H}\boldsymbol{\theta})^\intercal\mathbf{y}$. Because $\boldsymbol{\theta}$ is unknown, eqn. (26) cannot be evaluated to implement a test.

The Generalized Likelihood Ratio Test (GLRT) [40] compares the *most likely* model in $\mathcal{H}_1$ to the *most likely* model in $\mathcal{H}_0$, for threshold $\boldsymbol{\gamma}$, such that

$$\frac{\max_{\boldsymbol{\theta}} \mathrm{p}(\mathbf{y}|\mathcal{H}_1, \boldsymbol{\theta})}{\max_{\boldsymbol{\theta}} \mathrm{p}(\mathbf{y}|\mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \boldsymbol{\gamma}.$$

The GLRT is determined by finding the Maximum Likelihood Estimate (MLE) of $\boldsymbol{\theta}$. The MLE estimates $\hat{\boldsymbol{\theta}}$ by finding the value of $\boldsymbol{\theta}$ that maximizes $\hat{\Lambda}(\boldsymbol{\theta}; \mathbf{y})$ [39], for $i = \{0, 1\}$:

$$\hat{\boldsymbol{\theta}}_i \triangleq \underset{\boldsymbol{\theta}}{arg\,max}\ \mathrm{p}(\mathbf{y}|\mathcal{H}_i, \boldsymbol{\theta}).$$

For the alternate-hypothesis, the $\boldsymbol{\theta}$ that makes $\mathbf{y}$ most likely is

$$\begin{aligned}\hat{\boldsymbol{\theta}}_1 &= \underset{\boldsymbol{\theta}}{arg\,max}\ \mathrm{p}(\mathbf{y}|\mathcal{H}_1,\boldsymbol{\theta}) \\ &= \underset{\boldsymbol{\theta}}{arg\,max}\ \frac{1}{(2\pi\boldsymbol{\sigma}^2)^{k/2}} e^{-\frac{1}{2\boldsymbol{\sigma}^2}(\mathbf{y}-\mathbf{H}\boldsymbol{\theta})^\intercal(\mathbf{y}-\mathbf{H}\boldsymbol{\theta})} \quad (27)\\ &= \underset{\boldsymbol{\theta}}{arg\,max}\ -\frac{1}{2\boldsymbol{\sigma}^2}(\mathbf{y}-\mathbf{H}\boldsymbol{\theta})^\intercal(\mathbf{y}-\mathbf{H}\boldsymbol{\theta}) \quad (28)\\ &= \underset{\boldsymbol{\theta}}{arg\,min}\ (\mathbf{y}-\mathbf{H}\boldsymbol{\theta})^\intercal(\mathbf{y}-\mathbf{H}\boldsymbol{\theta}) \quad (29)\\ &= \underset{\boldsymbol{\theta}}{arg\,min}\ (\mathbf{y}^\intercal\mathbf{y}-2\boldsymbol{\theta}^\intercal\mathbf{H}^\intercal\mathbf{y}+\boldsymbol{\theta}^\intercal\mathbf{H}^\intercal\mathbf{H}\boldsymbol{\theta}). \quad (30)\end{aligned}$$

The exponential function of $\boldsymbol{\theta}$ is an increasing function. Eqn. (27) can be reduced to eqn. (28) because the log of the exponent does not change the maximization of the exponent over $\boldsymbol{\theta}$. Eqn. (28) can be reduced to eqn. (29) because $\frac{1}{2\boldsymbol{\sigma}^2}$ is independent of $\boldsymbol{\theta}$, which will not change the maximum relative to $\boldsymbol{\theta}$. Accounting for the negative value in eqn. (28) changes the problem from a maximization over $\boldsymbol{\theta}$, to an equivalent minimization over $\boldsymbol{\theta}$, in eqn. (29). Finally eqn. (30) is simply algebra.

To find $\hat{\boldsymbol{\theta}}_1$, take the partial derivative of eqn. (30) and set it equal to zero:

$$\begin{aligned}&\frac{\partial}{\partial\boldsymbol{\theta}}(\mathbf{y}^\intercal\mathbf{y}-\boldsymbol{\theta}^\intercal\mathbf{H}^\intercal\mathbf{y}+\boldsymbol{\theta}^\intercal\mathbf{H}^\intercal\mathbf{H}\boldsymbol{\theta})=0\\ &0-2\mathbf{H}^\intercal\mathbf{y}+2\mathbf{H}^\intercal\mathbf{H}\boldsymbol{\theta}=0\\ &\hat{\boldsymbol{\theta}}_1=(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{y}. \quad (31)\end{aligned}$$

Substituting eqn. (31) into eqn. (26) yields the analytical form of the GLRT

$$\begin{aligned}\ln\left(\hat{\Lambda}(\mathbf{y})\right) &= -\frac{1}{2\boldsymbol{\sigma}^2}\Big(-2\mathbf{y}^\intercal\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{y}\\ &\quad +\mathbf{y}^\intercal\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{y}\Big)\\ &= -\frac{1}{2\boldsymbol{\sigma}^2}\Big(-2\mathbf{y}^\intercal\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{y}\\ &\quad +\mathbf{y}^\intercal\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{y}\Big)\\ &= -\frac{1}{2\boldsymbol{\sigma}^2}\Big(-2\mathbf{y}^\intercal\mathbf{P}\mathbf{y}+\mathbf{y}^\intercal\mathbf{P}\mathbf{y}\Big)\\ &= \frac{1}{\boldsymbol{\sigma}^2}\Big(\mathbf{y}^\intercal\mathbf{P}\mathbf{y}-\frac{1}{2}\mathbf{y}^\intercal\mathbf{P}\mathbf{y}\Big)\\ &= \frac{1}{2\boldsymbol{\sigma}^2}\mathbf{y}^\intercal\mathbf{P}\mathbf{y}\ \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}}\ \boldsymbol{\gamma}', \quad (32)\end{aligned}$$

where $\mathbf{P}\triangleq\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal$.

From the result in eqn. (32), we can now determine the relation of the GLRT to the Probability of False Alarm (P_{FA}) and the Chi-square distribution.

1) GLRT relation to P_{FA} and χ^2: The objective is to choose $\boldsymbol{\gamma}'$ for the desired P_{FA} by evaluating eqn. (32) for the binary hypothesis. First, consider $\mathbf{y}^\intercal\mathbf{P}\mathbf{y}$ under $\mathcal{H}_0$. Define $\mathbf{H}$ in terms of the "thin" *QR* factorization [56], e.g. $\mathbf{H}=\mathbf{Q}_1\mathbf{R}_1$:

$$\begin{aligned}\mathbf{H} &= \mathbf{Q}\mathbf{R}\\ &= [\mathbf{Q}_1\ \mathbf{Q}_2]\begin{bmatrix}\mathbf{R}_1\\ \mathbf{0}\end{bmatrix}\\ &= \mathbf{Q}_1\mathbf{R}_1\end{aligned}$$

where $\mathbf{Q}\in\mathbb{R}^{m\times m}$ is a basis for the column space of $\mathbf{H}$, and $\mathbf{R}\in\mathbb{R}^{m\times n}$ with $m>n$. Both $\mathbf{Q}_1$ and $\mathbf{Q}_2$ have orthogonal columns, where $\mathbf{Q}_1\in\mathbb{R}^{m\times n}$, $\mathbf{Q}_2\in\mathbb{R}^{m\times(m-n)}$. The parameter $\mathbf{R}_1\in\mathbb{R}^{n\times n}$ is an invertible upper triangular matrix, and the zeros matrix, $\mathbf{0}\in\mathbb{R}^{(m-n)\times n}$. For full column-rank $\mathbf{H}$, i.e. rank$(\mathbf{H})=n$, then both $\mathbf{Q}_1$ and $\mathbf{R}_1$ are unique.

Using the *QR* factorization of $\mathbf{H}$ allows analysis of $\mathbf{P}$ as

$$\begin{aligned}\mathbf{P} &= \mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\\ &= \mathbf{Q}_1\mathbf{R}_1(\mathbf{R}_1^\intercal\mathbf{Q}_1^\intercal\mathbf{Q}_1\mathbf{R}_1)^{-1}\mathbf{R}_1^\intercal\mathbf{Q}_1^\intercal\\ &= \mathbf{Q}_1\mathbf{R}_1(\mathbf{R}_1^\intercal\mathbf{I}\mathbf{R}_1)^{-1}\mathbf{R}_1^\intercal\mathbf{Q}_1^\intercal\\ &= \mathbf{Q}_1\mathbf{R}_1\mathbf{R}_1^{-1}(\mathbf{R}_1^\intercal)^{-1}\mathbf{R}_1^\intercal\mathbf{Q}_1^\intercal\\ &= \mathbf{Q}_1\mathbf{Q}_1^\intercal \quad (33)\end{aligned}$$

where $\mathbf{Q}_1^\intercal\mathbf{Q}_1=\mathbf{I}$. Substituting eqn. (33) into eqn. (32), the decision statistic is

$$\begin{aligned}\ln\left(\hat{\Lambda}(\mathbf{y})\right) &= \frac{1}{2\boldsymbol{\sigma}^2}\mathbf{y}^\intercal\mathbf{P}\mathbf{y}\\ &= \frac{1}{2\boldsymbol{\sigma}^2}\mathbf{y}^\intercal\mathbf{Q}_1\mathbf{Q}_1^\intercal\mathbf{y}\\ &= \frac{1}{2\boldsymbol{\sigma}^2}\mathbf{z}^\intercal\mathbf{z}\end{aligned}$$

where $\mathbf{z}=\mathbf{Q}_1^\intercal\mathbf{y}\in\mathbb{R}^{n\times 1}$ is a Gaussian random variable with m degrees of freedom, and $\ln\left(\hat{\Lambda}(\mathbf{y})\right)$ is a $\chi^2_{(m-n)}$ random variable with $m-n$ degrees-of-freedom: the degrees-of-freedom of a Chi-square random variable is the number of measurements m, minus the number of parameters n.

Under the alternate hypothesis, $\mathcal{H}_1:\mathbf{y}_1\sim\mathcal{N}(\mathbf{H}\boldsymbol{\theta},\boldsymbol{\sigma}^2\mathbf{I})$, the expected value of $\mathbf{z}_1$ is

$$\begin{aligned}E\langle\mathbf{z}_1\rangle &= E\langle\mathbf{Q}_1^\intercal\mathbf{y}\rangle\\ &= \mathbf{Q}_1^\intercal E\langle\mathbf{y}\rangle\\ &= \mathbf{Q}_1^\intercal\mathbf{H}\boldsymbol{\theta},\end{aligned}$$

and covariance of $\mathbf{z}_1$ is

$$\begin{aligned}E\langle\mathbf{z}_1\mathbf{z}_1^\intercal\rangle &= E\langle\mathbf{Q}_1^\intercal\mathbf{y}\mathbf{y}^\intercal\mathbf{Q}_1\rangle\\ &= \mathbf{Q}_1^\intercal(\boldsymbol{\sigma}^2\mathbf{I}_n)\mathbf{Q}_1\\ &= \sigma^2\mathbf{I}_n.\end{aligned}$$

Therefore, under $\mathcal{H}_1$, $\mathbf{z}_1\sim\mathcal{N}(\mathbf{H}\boldsymbol{\theta},\boldsymbol{\sigma}^2\mathbf{I}_n)\in\mathbb{R}^{n\times 1}$.

Under the null hypothesis, $\mathcal{H}_0:\mathbf{y}_0\sim\mathcal{N}(\mathbf{0},\boldsymbol{\sigma}^2\mathbf{I})$, the expected value of $\mathbf{z}_0$ is

$$\begin{aligned}E\langle\mathbf{z}_0\rangle &= E\langle\mathbf{Q}_1^\intercal\mathbf{y}\rangle \quad (34)\\ &= \mathbf{Q}_1^\intercal E\langle\mathbf{y}\rangle \quad (35)\\ &= \mathbf{0}, \quad (36)\end{aligned}$$

and covariance of $\mathbf{z}_0$ is

$$\begin{aligned} E\left\langle \mathbf{z}_0 \mathbf{z}_0^\intercal \right\rangle &= E\left\langle \mathbf{Q}_1^\intercal \mathbf{y}\mathbf{y}^\intercal \mathbf{Q}_1 \right\rangle \\ &= \mathbf{Q}_1^\intercal (\boldsymbol{\sigma}^2 \mathbf{I}_n) \mathbf{Q}_1 \\ &= \sigma^2 \mathbf{I}_n. \end{aligned}$$

Therefore, under $\mathcal{H}_0$, $\mathbf{z}_0 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}^2 \mathbf{I}_n) \in \mathbb{R}^{n \times 1}$.

From eqn. (32), the test statistic is

$$\frac{\mathbf{z}^\intercal \mathbf{z}}{2\boldsymbol{\sigma}^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \boldsymbol{\gamma}'. \tag{37}$$

Given the P_{FA} constraint, the optimum decision threshold $\boldsymbol{\gamma}'$ is found by applying the inverse CDF of the Chi-square distribution with $m-n$ degrees-of-freedom. Thus, under $\mathcal{H}_0$, we can define the P_{FA} in terms of the GLRT

$$P_{FA} = \mathrm{p}(\chi^2_{(m-n)} > \boldsymbol{\gamma}).$$

In statistics literature, eqn. (37) is referred to as Wilks Theorem [57].

REFERENCES

[1] R. V. Beard, "Failure Accommodation in Linear Systems Through Self-Reorganization," Ph.D. dissertation, Massachusetts Inst. of Technology, Cambridge, MA, 1971.

[2] J. Chen, "Robust residual generation for model-based fault diagnosis of dynamic systems," Ph.D. dissertation, University of York., 1995.

[3] P. M. Frank and X. Ding, "Survey of robust residual generating and evaluation methods in observer-based fault detection systems," *Journal of Process Control*, vol. 7, no. 6, p. 403424, 1997.

[4] J. F. Magni and P. Mouyon, "On residual generation by observer and parity space approaches," *IEEE Transactions on Automatic Control*, vol. 39, no. 2, pp. 441–447, 1994.

[5] P. Y. Hwang and R. G. Brown, "From RAIM to NIORAIM: A New Integrity Approach to Integrated Multi-GNSS Systems," *Inside GNSS*, May-June, 2008.

[6] J. Gertler and D. Singer, "A new structural framework for parity equation-based failure detection and isolation," *Automatica*, vol. 26, no. 2, pp. 381–388, 1990.

[7] J. Gertler, "Fault detection and isolation using parity relations," *Control Engineering Practice*, vol. 5, no. 5, p. 653661, 1997.

[8] J. Gertler and M. Staroswiecki, "Structured fault diagnosis in mildly nonlinear systems: parity space and input-output formulations," *In Preprints of the 15th IFAC World Congress*, 2002.

[9] A. Medvedev, "Fault Detection and Isolation by a Continuous Parity Space Method," *Automatica*, vol. 31, no. 7, pp. 1039–1034, 1995.

[10] R. J. Patton and J. Chen, "A review of parity space approaches to fault diagnosis for aerospace systems," *Journal of Guidance, Control and Dynamics*, vol. 17, no. 2, pp. 278–285, 1994.

[11] M. A. Sturza, "Navigation System Integrity Monitoring Using Redundant Measurements," *Navigation, Journal of The Institute of Navigation*, vol. 35, no. 4, Winter 1988-89.

[12] P. J. G. Teunissen, "Quality Control in Integrated Navigation Systems," *IEEE Aerospace and Electronic Systems Magazine*, vol. 5, no. 7, pp. 35–41, August 1990.

[13] U. I. Bhatti and W. Y. Ochieng, "Detecting Multiple failures in GPS/INS integrated system: A Novel architecture for Integrity Monitoring," *Journal of Global Positioning Systems*, vol. 8, no. 1, pp. 26–42, 2006.

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

[15] J. Diesel and J. King, "Integration of Navigation Systems for Fault Detection, Exclusion, and Integrity Determination - Without WAAS," *Proceedings of the 1995 National Technical Meeting of The Institute of Navigation, Anaheim, CA*, pp. 683–692, 1995.

[16] F. Hampel, E. Ronchetti, P. Rousseeuw, and W. Stahel, *Robust Statistics: The approach Based on Influence Functions*. New York: John Wiley and Sons Inc., 1986.

[17] P. Rousseeuw and A. Leroy, *Robust regression and outlier detection*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, New York: John Wiley and Sons Inc., 1987.

[18] S. Weisberg, *Applied Linear Regression*. 2nd ed, New York, Wiley, 1985.

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

[20] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3, pp. 591–611, 1965.

[21] T. Anderson and D. Darling, "A Test of Goodness-of-Fit," *Journal of the American Statistical Association*, vol. 49, no. 765-769, 1954.

[22] N. Smirnov, "Table for estimating the goodness of fit of empirical distributions," *Annals of Mathematical Statistics*, vol. 19, pp. 279–281, 1948.

[23] M. A. Stephens, "EDF Statistics for Goodness of Fit and Some Comparisons," *Journal of the American Statistical Association (American Statistical Association)*, vol. 69, no. 374, pp. 730–737, 1974.

[24] J. Neyman and E. S. Pearson, "The testing of statistical hypotheses in relation to probabilities a priori," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 29, pp. 492–510, 1933.

[25] C. C. Aggarwal, *Outlier Analysis*. Springer, 2013.

[26] P. Huber, *Robust Statistics*. New York: John Wiley and Sons Inc., 1986.

[27] D. Serbert, D. Montgomery, and D. Rollier, "A clustering algorithm for identifying multiple outliers in linear regression," *Computational Statistics and Data Analysis*, vol. 27, pp. 461–484, 1998.

[28] A. S. Willsky and E. Y. Chow, "Analytical Redundancy and the Design of Robust Failure Detection Systems," *IEEE Transactions of Automatic Control*, vol. AC-29, no. 7, pp. 603–614, 1984.

[29] W. Förstner, "Reliability and Discernibility of Extended Gauss-Markov Models," *DGK, A. Reihe, ed., Heft 98, Munich, Germany, (translated from German into English)*, 1998.

[30] W. Baarda, "Statistical Concepts in Geodesy," *Netherlands Geod. Comm.*, vol. 2, no. 4, 1967.

[31] ——, "A Testing Procedure for Use in Geodetic Networks," *Netherlands Geod. Comm.*, vol. 2, no. 5, 1968.

[32] P. M. Frank, "Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy a survey and some new results," *Automatica*, vol. 26, no. 3, p. 459474, 1990.

[33] J. Gertler, "Analytical Redundancy Methods in Fault Detection and Isolation," *In Preprints of IFAC/IMACS Symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS91*, pp. 9–21, 1991.

[34] X. Lou, A. S. Willsky, and G. C. Verghese, "Optimally Robust Redundancy Relations for Failure Detection in Uncertain Systems," *Automatica*, vol. 22, no. 3, pp. 333–344, 1986.

[35] R. J. Patton, "Fault detection and diagnosis in aerospace systems using analytical redundancy," *Computing and Control Engineering Journal*, vol. 2, no. 3, pp. 127–136, 1991.

[36] M. Staroswiecki and G. Comtet-Varga, "Analytical redundancy relations for fault detection and isolation in algebraic dynamic systems," *Automatica*, vol. 37, no. 5, pp. 687–699, 2001.

[37] P. M. Frank, "On-line fault detection in uncertain nonlinear systems using diagnostic observers: A survey," *International Journal Systems Science*, vol. 25, no. 12, p. 21292154, 1994.

[38] E. A. Garcia and P. M. Frank, "Deterministic nonlinear observer-based approaches to fault diagnosis: A survey," *Control Engineering Practice*, vol. 5, no. 5, p. 663670, 1997.

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

[40] ——, *Fundamentals of Statistical Signal Processing, Vol. II - Detection Theory*. Prentice Hall PTR, 1998.

[41] E. M. Mikhail, *Observations and Least Squares*. Thomas Y. Crowell Company, Inc., 1976.

[42] R. G. Brown, "A Baseline RAIM Scheme and a Note on the Equivalence of Three RAIM Methods," *Proceedings of the 1992 National Technical Meeting of The Institute of Navigation, San Diego, CA*, pp. 127–137, January 27 - 29, 1992.

[43] J. Angus, "RAIM with Multiple Faults," *Journal of The Institute of Navigation*, vol. 53, no. 4, 2006.

[44] M. Brenner, "Integrated GPS/Inertial Fault Detection Availability," *Navigation, Journal of The Institute of Navigation*, vol. 43, no. 2, 1996.

[45] S. Castaldo, A. Angrisano, S. Gaglione, and S. Troisi, "P-RanSaC: An Integrity Monitoring Approach for GNSS Signal Degraded Scenario," *International Journal of Navigation and Observation*, 2014.
[46] J. L. Farrell, "Full Integrity Testing for GPS/INS," *Navigation, Journal of The Institute of Navigation*, vol. 53, no. 1, Spring 2006.
[47] S. Hewitson, H. K. Lee, and J. Wang, "Localizability Analysis for GPS/Galileo Receiver Autonomous Integrity Monitoring," *Navigation, Journal of The Institute of Navigation*, vol. 57, pp. 245–259, 2004.
[48] S. Hewitson and J. Wang, "GNSS Receiver Autonomous Integrity Monitoring (RAIM) Performance Analysis," *GPS Solutions, Springer-Verlag*, vol. 10, no. 3, pp. 155–170, 2006.
[49] J. Wang and P. B. Ober, "On the Availability of Fault Detection and Exclusion in GNSS Receiver Autonomous Integrity Monitoring," *Navigation, Journal of The Institute of Navigation*, vol. 62, pp. 251–261, 2009.
[50] W. Zhang and M. Ghogho, "GPS Signal Detection Using Hypothesis Testing Analysis," *Journal of Global Positioning Systems*, vol. 10, no. 1, pp. 125–135, 2011.
[51] J. A. Farrell, *Aided Navigation: GPS with High Rate Sensors*. McGraw Hill, 2008.
[52] S. Hewitson and J. Wang, "Extended Receiver Autonomous Integrity Monitoring (eRAIM) for GNSS/INS Integration," *Journal of Surveying Engineering*, vol. 136, no. 1, pp. 13–22, 2010.
[53] ——, "GNSS Receiver Autonomous Integrity Monitoring (eRAIM) for Multiple Outliers," *Journal of Geodesy*, vol. 87, no. 2, pp. 13–22, 2012.
[54] D. Wang, H. Lu, and M. Yang, "Robust Visual Tracking via Least Soft-threshold Squares," *IEEE Transactions on Circuits and Systems for Video Technology*, 2015.
[55] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523–531, 1967.
[56] G. H. Golub and C. F. Van Loan, *Matrix Computations (3rd ed.)*. Johns Hopkins, 1996.
[57] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *Annals of Mathematical Statistics*, vol. 9, no. 1, pp. 60–62, 1938.

Technical Note: CRT-HT Position, Velocity & Attitude Results

Paul F. Roysdon[†] Jay A. Farrell[‡]

I. Comparison: CRT HT Results

For each algorithm in [1], 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 eqn. (6) in [1] 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. Similar results are shown in Figs. 2 & 3 for velocity and attitude error, respectively.

The CDF in Fig. 1 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. For the CRT with $L > 5$, 100% of the position errors are less than $1.0m$. 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 $3.0m$.

The CDF in Fig. 2 shows that the percentage of occurrences where the EKF velocity error is less than $0.01m/s$, is roughly 15%. Roughly 85% of the trajectory, as estimated by the EKF, has errors less than $0.1m/s$. For the CRT with $L > 10$, 30% of the velocity errors are less than $0.01m/s$ and 100% of the velocity errors are less than $0.1m/s$.

The CDF in Fig. 3 shows that the percentage of occurrences where the EKF attitude error is less than 0.01^o, is roughly 10%. Roughly 98% of the trajectory, as estimated by the EKF, has errors less than 0.1^o. For the CRT with $L > 10$, 15% of the attitude errors are less than 0.01^o and 100% of the attitude errors are less than 0.1^o.

Figs. 1-3 indicate that accuracy improves from the EKF to the IEKF to the CRT. Also, CRT performance (generally) improves with the window length L.

References

[1] P. F. Roysdon and J. A. Farrell, "GPS-INS Outlier Detection and Elimination using a Sliding Window Filter," *American Control Conference, In Presc.*, 2017.

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

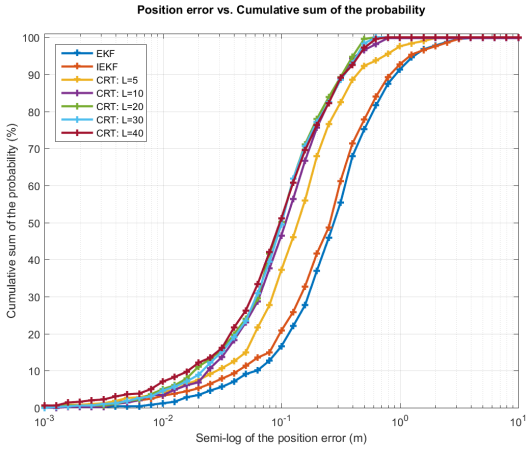

Fig. 1. Cumulative distribution of position error for each algorithm.

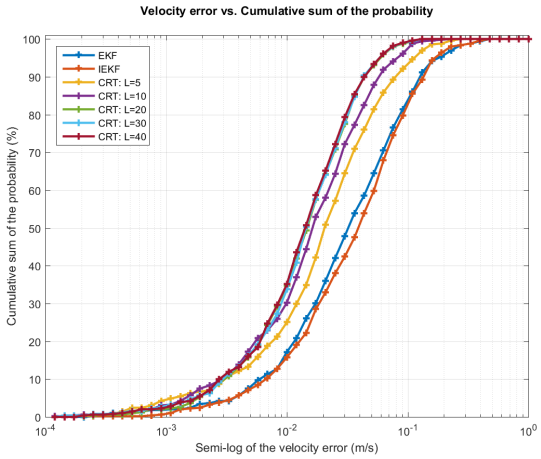

Fig. 2. Cumulative distribution of velocity error for each algorithm.

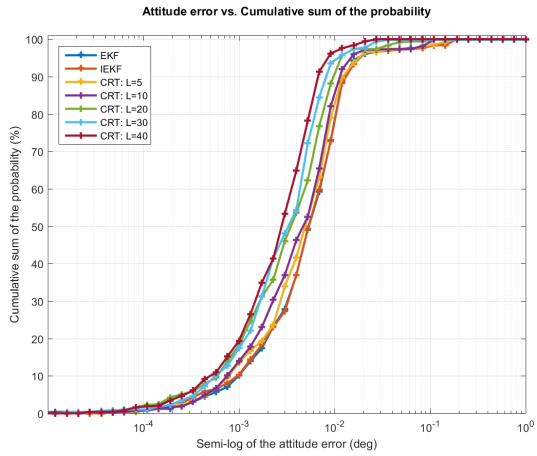

Fig. 3. Cumulative distribution of attitude error for each algorithm.