Hypothesis Testing: Derivations & Proofs

Paul F. Roysdon, Ph.D.

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, $\mathcal{H}_i$:

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

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.

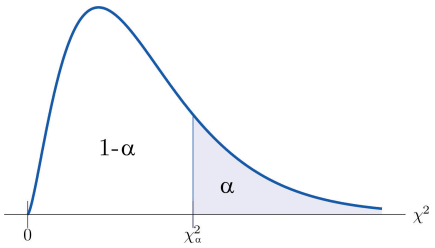

Fig. 1. Chi-square distribution. The shaded area represents the region defined by the significance level α for the test described by right side of eqn. (11).

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 $\mathbf{r} \sim \mathcal{N}(\mathbf{Q}\boldsymbol{\varepsilon}_i, \boldsymbol{\sigma}_y^2\mathbf{Q})$.

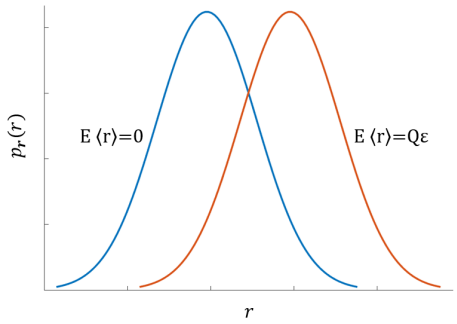

Fig. 2. Assume a normal distribution for two sets of random variables with equal variance. One set contains an outlier and the other set does not. The difference between the two sets will be a shift of the mean in the set which contains the outlier.

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$(\mathbf{P})$.

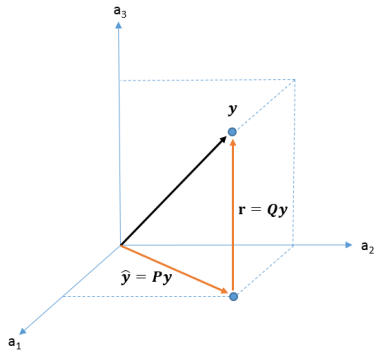

Fig. 3. For a general space in $\mathbb{R}^3$, the mapping $\mathbf{P}\mathbf{y} = \hat{\mathbf{y}}$ is the estimate for $\mathbf{y}$, and $\mathbf{Q}\mathbf{y} = \mathbf{r}$ is the estimation residual for $\mathbf{y}$.

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.

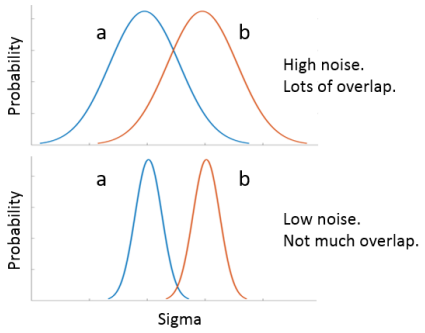

Fig. 4. Signal spread due to noise, and resulting overlap.

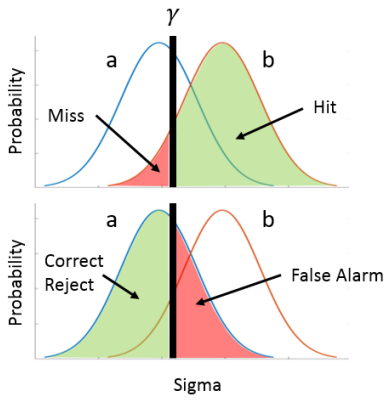

Fig. 5. The four regions of hypothesis testing, defined by two overlapping distributions (*a* & *b*) and the vertical line of the decision threshold $\boldsymbol{\gamma}$.

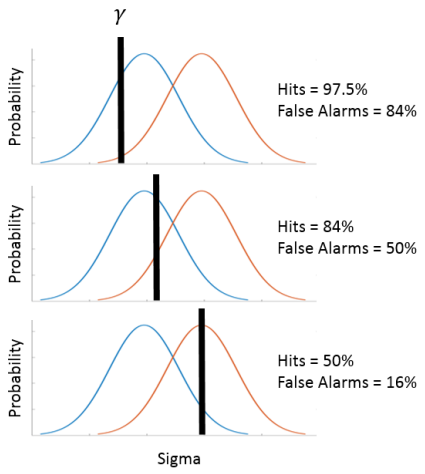

Fig. 6. An example result of *hits* versus *false alarms*, due to shifting the threshold $\boldsymbol{\gamma}$.

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].

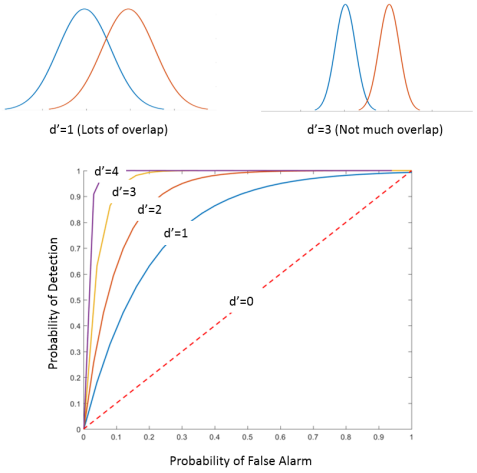

Fig. 7. ROC plot for Hits (P_D) versus False Alarms (P_{FA}).

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.