Error-Covariance Analysis of the Total Least Squares Problem

John L. Crassidis[*]

University at Buffalo, State University of New York, Amherst, NY 14260-4400

Yang Cheng[†]

Mississippi State University, Mississippi State, MS 39762

This paper derives and analyzes the estimate error-covariances associated for both the non-stationary and stationary noise process cases with uncorrelated element-wise components for the total least squares problem. The non-stationary case is derived directly from the associated unconstrained total least squares loss function. The stationary case is derived by using a linear expansion of the total least squares estimate equation, which involves a first order expansion of the associated singular value decomposition matrices. The actual solution for the error-covariance is evaluated at the true variables, which are unknown in practice. Two common approaches to overcome this difficulty are used; the first involves using the measurements directly and the second involves using the estimates which are more accurate than the measurements. This paper shows that using the latter greatly simplifies the error-covariance solution for the stationary case. Simulation results using bearings-only point estimation are shown to quantify the theoretical derivations.

I. Introduction

Total least squares (TLS) expands upon standard least squares by incorporating noise not only in the measurements but also in the basis functions themselves. Several applications

[*]CUBRC Professor in Space Situational Awareness, Department of Mechanical & Aerospace Engineering. Email: johnc@buffalo.edu. Associate Fellow AIAA.

[†]Assistant Professor, Department of Aerospace Engineering. Email: cheng@ae.msstate.edu. Associate Fellow AIAA.

of TLS exist in the real world, such as fuzzy system identification of an industrial gas engine power plant,[1] blind deconvolution problems as encountered in image deblurring when both the image and the blurring function have uncertainty,[2] and applications to astronomy and geodesy.[3] An overview of TLS can be found in Ref. 4.

Since noise exists in the basis functions then the standard least squares solution is not optimal from both a minimum variance and maximum likelihood (ML) point of view. Thus a different loss function must be used other than the standard least squares loss function. The "errors-in-variables" estimator shown in Ref. 5 coincides with the TLS solution. This indicates that the TLS estimate is a strongly consistent estimate for large samples, which leads to an asymptotic unbiasedness property. Ordinary least squares with errors in the basis functions produces biased estimates as the sample size increases. However, the error-covariance of TLS is larger than the ordinary least squares error-covariance, but by increasing the noise in the measurements the bias of ordinary least squares becomes more important and even the dominating term.[6] Also, it is has been shown that weighted least squares and TLS yield asymptotically equivalent results as the perturbation level goes to zero.[7]

The covariance of the estimate errors in the standard linear least squares problem is straightforward to derive. Standard least squares can easily be shown to produce an efficient estimate, i.e. its state error-covariance achieves the Cramér-Rao lower bound (CRLB).[8] The equivalence of the TLS to maximum likelihood estimation is shown in Refs. 9 and 10. A CRLB is derived in Ref. 10. However, the derivation is a mix of Bayesian and non-Bayesian approaches, and is done by introducing nuisance parameters with a prior distribution. Strictly speaking, the likelihood function in Ref. 10 is not the likelihood function because of the use of the prior distribution on the nuisance parameters. As noted by the authors of Ref. 10, "different choices for this prior can lead to likelihood functions whose maxima are not the same as the TLS solution." The derivation in the present paper is different from that of Ref. 10 in that it only uses the standard likelihood function and does not need any prior distributions on the nuisance parameters. It thus avoids subjective choices of prior distributions. The only assumption here is that the errors are element-wise uncorrelated, which is true for many systems. The classic text of Ref. 11 presents first-order covariance approximations for the TLS estimate. A detailed discussion on bias and how the TLS covariance is related to simple least squares error-covariance is also presented. However, the error-covariance is only valid for isotropic noise errors for the basis functions and measurements, i.e. the overall covariance matrix is given by a scalar times identity. In the present paper the derivation of the error-covariance is valid for non-isotropic errors and is also derived directly from the TLS loss function.

In this paper the CRLB is first derived, which is valid for both the non-stationary and stationary noise process cases. Then, a perturbation approach of the TLS loss function is employed to prove that the associated covariance matrix achieves the CRLB to within first-order terms. The TLS estimate for the stationary noise case involves performing a singular value decomposition (SVD) of an augmented matrix involving the basis functions and measurements. The derivation of the error-covariance follows directly from the SVD matrix solution. Unlike the non-stationary noise case, a matrix inverse is not required to compute the error-covariance for the stationary noise case.

The organization of this paper proceeds as follows. First, a review of maximum likelihood estimation is given with a particular emphasis on the linear least squares problem. The Cramér-Rao inequality is also shown. Then, the relationship of TLS to maximum likelihood estimation is shown. Next, error-covariance expressions for both the non-stationary and stationary cases are derived. Finally, simulation results using bearings-only point estimation are shown to validate the derived error-covariance expressions.

II. Maximum Likelihood Estimation

The ML approach yields estimates for the unknown quantities which maximize the probability of obtaining the observed set of data. In this section a review of ML estimation for the standard linear least squares solution is given, which includes a review of the Cramér-Rao inequality. Then the ML formulation for the TLS problem is formally shown. The Cramér-Rao inequality is then derived for the non-stationary noise process case, assuming no correlations exist between element-wise errors of the measurements and basis functions.

A. Linear Least Squares Review

Consider the following linear model:

$$\tilde{\mathbf{y}} = H\,\mathbf{x} + \Delta\mathbf{y} \tag{1}$$

where H is an $m \times n$ matrix which contains no errors and $\Delta\mathbf{y}$ is an $m \times 1$ vector which is a zero-mean Gaussian white-noise process with covariance R. The goal of the least squares problem is to determine an estimate for the $n \times 1$ vector $\mathbf{x}$, with $n \leq m$.

The mean of the $m \times 1$ measurement $\tilde{\mathbf{y}}$, denoted by $\boldsymbol{\mu}$, is computed by taking the expectation of Eq. (1), which gives $\boldsymbol{\mu} = H\,\mathbf{x}$. Then the covariance of $\tilde{\mathbf{y}}$ is given by

$$\text{cov}\left\{\tilde{\mathbf{y}}\right\} \triangleq E\left\{\left(\tilde{\mathbf{y}} - \boldsymbol{\mu}\right)\left(\tilde{\mathbf{y}} - \boldsymbol{\mu}\right)^T\right\} \tag{2}$$

where $E\left\{\,\right\}$ denotes expectation. Carrying out the computation in Eq. (2) gives $\text{cov}\left\{\tilde{\mathbf{y}}\right\} = R$.

Hence the conditional density function of $\tilde{\mathbf{y}}$ given $\mathbf{x}$ is

$$p(\tilde{\mathbf{y}}|\mathbf{x}) = \frac{1}{(2\pi)^{m/2}\left[\det\left(R\right)\right]^{1/2}} \exp\left\{-\frac{1}{2}\left(\tilde{\mathbf{y}} - H\mathbf{x}\right)^T R^{-1}\left(\tilde{\mathbf{y}} - H\mathbf{x}\right)\right\} \tag{3}$$

In the ML approach an estimate of $\mathbf{x}$, denoted by $\hat{\mathbf{x}}$, is sought that maximizes Eq. (3). Due to the monotonic aspect of the function, the ML solution can be accomplished by also taking the natural logarithm of Eq. (3), which yields

$$\ln\left[p(\tilde{\mathbf{y}}|\mathbf{x})\right] = -\frac{1}{2}\left(\tilde{\mathbf{y}} - H\mathbf{x}\right)^T R^{-1}\left(\tilde{\mathbf{y}} - H\mathbf{x}\right) - \frac{m}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln\left[\det\left(R\right)\right] \tag{4}$$

The last two terms of the right-hand side of Eq. (4) can be ignored since they are independent of $\mathbf{x}$. Maximizing Eq. (4) is equivalent to minimizing the negative of it. Therefore, ignoring terms independent of $\mathbf{x}$ leads to the following loss function which is minimized to determine the estimate:

$$J(\hat{\mathbf{x}}) = \frac{1}{2}\left(\tilde{\mathbf{y}} - H\hat{\mathbf{x}}\right)^T R^{-1}\left(\tilde{\mathbf{y}} - H\hat{\mathbf{x}}\right) \tag{5}$$

The solution for this minimization problem leads directly to the classical least squares solution for the estimate:

$$\hat{\mathbf{x}}(\tilde{\mathbf{y}}) \equiv \hat{\mathbf{x}} = (H^T R^{-1} H)^{-1} H^T R^{-1} \tilde{\mathbf{y}} \tag{6}$$

The mean of $\hat{\mathbf{x}}$ is given by $\mathbf{x}$, which means the estimator is unbiased. The error-covariance of $\hat{\mathbf{x}}$ is given by

$$\text{cov}\left\{\hat{\mathbf{x}}\right\} = (H^T R^{-1} H)^{-1} \tag{7}$$

which can be used to develop 3σ bounds on the expected estimate errors.

The Cramér-Rao inequality[12] can be used to provide a lower bound on the expected errors between the estimated quantities and the true values from the known statistical properties of the measurement errors. The theory was proved independently by Cramér and Rao, although it was found earlier by Fisher[13] for the special case of a Gaussian distribution. The Cramér-Rao inequality for an unbiased estimate $\hat{\mathbf{x}}$ is given by

$$P \triangleq E\left\{\left(\hat{\mathbf{x}} - \mathbf{x}\right)\left(\hat{\mathbf{x}} - \mathbf{x}\right)^T\right\} \geq F^{-1} \tag{8}$$

where the Fisher information matrix (FIM), F, is given by

$$F = E\left\{\left(\frac{\partial}{\partial \mathbf{x}}\ln[p(\tilde{\mathbf{y}}|\mathbf{x})]\right)\left(\frac{\partial}{\partial \mathbf{x}}\ln[p(\tilde{\mathbf{y}}|\mathbf{x})]\right)^T\right\} \tag{9}$$

The partial derivatives are assumed to exist and to be absolutely integrable. A formal proof of the Cramér-Rao inequality requires using the "conditions of regularity" (see Ref. 14 for details). It is clear that the estimate in Eq. (6) achieves the CRLB and is thus an efficient estimator.

Maximum likelihood has many desirable properties. A few of the useful ones are now discussed. First, a ML estimator is a consistent estimator, which means $\hat{\mathbf{x}}(\tilde{\mathbf{y}})$ converges in a probabilistic sense to the truth, $\mathbf{x}$, for large samples. This states that the estimate is unbiased for large samples. Second, a ML estimator is asymptotically efficient, which means that $\hat{\mathbf{x}}(\tilde{\mathbf{y}})$ achieves the CRLB for large samples.

Oftentimes, as is seen many times throughout this paper, the estimate equation is nonlinear in both its functional parameters and random errors. To determine the error-covariance matrix, P, in Eq. (8) a classical first-order expansion of the nonlinear functions can be used.[8] This is best illustrated by example. Suppose that a random function is given $\hat{\mathbf{z}} = \mathbf{f}(\hat{\mathbf{p}})$, with $\hat{\mathbf{p}} = \mathbf{p} + \boldsymbol{\delta}\mathbf{p}$, where $\boldsymbol{\delta}\mathbf{p}$ is a zero-mean noise process with covariance denoted by P_{pp}. To within first order the covariance of $\hat{\mathbf{z}}$, denoted by P_{zz}, is computed using the Jacobian of $\mathbf{f}$, and is given by

$$P_{zz} = \left[\left. \frac{\partial \mathbf{f}}{\partial \hat{\mathbf{p}}} \right|_{\text{truth}} \right] P_{pp} \left[\left. \frac{\partial \mathbf{f}}{\partial \hat{\mathbf{p}}} \right|_{\text{truth}} \right]^T \tag{10}$$

This Jacobian is evaluated at the true values, which are replaced with measured or estimated values in practice. It is important to note that Eq. (10) is valid only for an unbiased estimate.

B. Total Least Squares

For the general problem, the TLS model is given by

$$\tilde{\mathbf{y}} = \mathbf{y} + \Delta\mathbf{y} \tag{11a}$$

$$\tilde{H} = H + \Delta H \tag{11b}$$

$$\mathbf{y} = H\,\mathbf{x} \tag{11c}$$

where $\tilde{\mathbf{y}}$ is an $m \times 1$ measurement vector, $\mathbf{y}$ is its respective true value, $\Delta\mathbf{y}$ is the measurement noise, $\tilde{H}$ is an $m \times n$ matrix of basis functions with random errors, H is its respective true value, and ΔH represents the errors to the model H. Define the following $m \times (n+1)$ matrix:

$$\tilde{D} \triangleq \begin{bmatrix} \tilde{H} & \tilde{\mathbf{y}} \end{bmatrix} \tag{12}$$

The TLS problem seeks an optimal estimate of the $n \times 1$ vector $\mathbf{x}$, denoted by $\hat{\mathbf{x}}$ with $\hat{\mathbf{y}} = \hat{H}\hat{\mathbf{x}}$, where $\hat{\mathbf{y}}$ is the estimate of $\mathbf{y}$ and $\hat{H}$ is the estimate of H, which maximizes

$$p(\tilde{D}|D) = \frac{1}{(2\pi)^{m\times(n+1)/2}\,[\det{(R)}]^{1/2}} \exp\left\{ -\frac{1}{2}\text{vec}^T(\tilde{D}^T - D^T)\,R^{-1}\text{vec}(\tilde{D}^T - D^T) \right\} \tag{13}$$

where $D \triangleq [H \;\; \mathbf{y}]$, which satisfies $D\,\mathbf{z} = \mathbf{0}$ with $\mathbf{z} \triangleq [\mathbf{x}^T \;\; -1]^T$, and vec denotes a column vector formed by stacking the consecutive columns of the associated matrix, and R is the covariance matrix. Unfortunately because H now contains errors the constraint $\hat{\mathbf{y}} = \hat{H}\,\hat{\mathbf{x}}$ must also be added to the maximization problem. The negative log-likelihood now leads to the following loss function:

$$J(\hat{D}) = \frac{1}{2}\text{vec}^T(\tilde{D}^T - \hat{D}^T)\,R^{-1}\text{vec}(\tilde{D}^T - \hat{D}^T), \quad \text{s.t.} \quad \hat{D}\,\hat{\mathbf{z}} = \mathbf{0} \tag{14}$$

where $\hat{\mathbf{z}} \triangleq [\hat{\mathbf{x}}^T \;\; -1]^T$ and $\hat{D} \triangleq [\hat{H} \;\; \hat{\mathbf{y}}]$ denotes the estimate of D. For a unique solution it is required that the rank of $\hat{D}$ be n, which means $\hat{\mathbf{z}}$ spans the null space of $\hat{D}$.

III. Error-Covariance Derivation

In this section the estimate error-covariance is derived for two cases in the TLS problem. The first assumes that the errors are element-wise, i.e. the rows of the matrix $\tilde{D}$, uncorrelated but allows the covariance to vary in time, i.e. non-stationary errors. The case covers a wide variety of problems, which is also used to develop a sequential least squares solution for the linear least squares problem.[8] The second case assumes that the errors are element-wise uncorrelated with stationary errors.

A. Element-Wise Uncorrelated and Non-Stationary Case

For this case the covariance matrix is given by the following block diagonal matrix: $R = \text{blkdiag}\begin{bmatrix}\mathcal{R}_1 & \cdots & \mathcal{R}_m\end{bmatrix}$, where each $\mathcal{R}_i$ is an $(n+1)\times(n+1)$ matrix given by

$$\mathcal{R}_i = \begin{bmatrix} \mathcal{R}_{hh_i} & \mathcal{R}_{hy_i} \\ \mathcal{R}_{hy_i}^T & \mathcal{R}_{yy_i} \end{bmatrix} \tag{15}$$

where $\mathcal{R}_{hh_i}$ is an $n \times n$ matrix, $\mathcal{R}_{hy_i}$ is an $n \times 1$ vector and $\mathcal{R}_{yy_i}$ is a scalar. Partition the matrix ΔH and the vector $\Delta\mathbf{y}$ by their rows:

$$\Delta H = \begin{bmatrix} \boldsymbol{\delta}\mathbf{h}_1^T \\ \boldsymbol{\delta}\mathbf{h}_2^T \\ \vdots \\ \boldsymbol{\delta}\mathbf{h}_m^T \end{bmatrix}, \quad \Delta\mathbf{y} = \begin{bmatrix} \delta y_1 \\ \delta y_2 \\ \vdots \\ \delta y_m \end{bmatrix} \tag{16}$$

where each $\boldsymbol{\delta}\mathbf{h}_i$ has dimension $n \times 1$ and each δy_i is a scalar. The partitions in Eq. (15) are then given by

$$\mathcal{R}_{hh_i} = E\left\{\boldsymbol{\delta}\mathbf{h}_i \boldsymbol{\delta}\mathbf{h}_i^T\right\} \tag{17a}$$

$$\mathcal{R}_{hy_i} = E\left\{\delta y_i \boldsymbol{\delta}\mathbf{h}_i\right\} \tag{17b}$$

$$\mathcal{R}_{yy_i} = E\left\{\delta y_i^2\right\} \tag{17c}$$

Note that each $\mathcal{R}_i$ is allowed to be a fully populated matrix so that correlations between the errors in the individual i^{th} row of ΔH and the i^{th} element of $\Delta\mathbf{y}$ can exist. When $\mathcal{R}_{hy_i}$ is zero then no correlations exist.

Partition the matrices $\tilde{D}$, $\hat{D}$ and $\tilde{H}$, and the vector $\tilde{\mathbf{y}}$ by their rows:

$$\tilde{D} = \begin{bmatrix} \tilde{\mathbf{d}}_1^T \\ \tilde{\mathbf{d}}_2^T \\ \vdots \\ \tilde{\mathbf{d}}_m^T \end{bmatrix}, \quad \hat{D} = \begin{bmatrix} \hat{\mathbf{d}}_1^T \\ \hat{\mathbf{d}}_2^T \\ \vdots \\ \hat{\mathbf{d}}_m^T \end{bmatrix}, \quad \tilde{H} = \begin{bmatrix} \tilde{\mathbf{h}}_1^T \\ \tilde{\mathbf{h}}_2^T \\ \vdots \\ \tilde{\mathbf{h}}_m^T \end{bmatrix}, \quad \tilde{\mathbf{y}} = \begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_m \end{bmatrix} \tag{18}$$

where each $\tilde{\mathbf{d}}_i$ and $\hat{\mathbf{d}}_i$ has dimension $(n+1) \times 1$, each $\tilde{\mathbf{h}}_i$ has dimension $n \times 1$ and each $\tilde{y}_i$ is a scalar. For the element-wise uncorrelated and non-stationary case, the constrained loss function in Eq. (14) can be converted to an equivalent unconstrained one.[15,16] Here, a simplified version of this is shown. For the element-wise uncorrelated and non-stationary case, the loss function in Eq. (14) reduces down to

$$J(\hat{\mathbf{d}}_i) = \frac{1}{2}\sum_{i=1}^{m}(\tilde{\mathbf{d}}_i - \hat{\mathbf{d}}_i)^T \mathcal{R}_i^{-1}(\tilde{\mathbf{d}}_i - \hat{\mathbf{d}}_i), \quad \text{s.t.} \quad \hat{\mathbf{d}}_j^T \hat{\mathbf{z}} = 0, \quad j = 1, 2, \ldots, m \tag{19}$$

The loss function is rewritten into an unconstrained one by determining a solution for $\hat{\mathbf{d}}_i$ and substituting its result back into Eq. (19). To accomplish this task the loss function is appended using Lagrange multipliers, which gives the following loss function:

$$J'(\hat{\mathbf{d}}_i) = \lambda_1 \hat{\mathbf{d}}_1^T \hat{\mathbf{z}} + \lambda_2 \hat{\mathbf{d}}_2^T \hat{\mathbf{z}} + \cdots + \lambda_m \hat{\mathbf{d}}_m^T \hat{\mathbf{z}} + \frac{1}{2}\sum_{i=1}^{m}(\tilde{\mathbf{d}}_i - \hat{\mathbf{d}}_i)^T \mathcal{R}_i^{-1}(\tilde{\mathbf{d}}_i - \hat{\mathbf{d}}_i) \tag{20}$$

where each λ_i is a Lagrange multiplier. Taking the partial of Eq. (20) with respect to each $\hat{\mathbf{d}}_i$ leads to the following m necessary conditions:

$$\mathcal{R}_i^{-1}\hat{\mathbf{d}}_i - \mathcal{R}_i^{-1}\tilde{\mathbf{d}}_i + \lambda_i \hat{\mathbf{z}} = \mathbf{0}, \quad i = 1, 2, \ldots, m \tag{21}$$

Left multiplying Eq. (21) by $\hat{\mathbf{z}}^T \mathcal{R}_i$ and using the constraint $\hat{\mathbf{d}}_i^T \hat{\mathbf{z}} = 0$ leads to

$$\lambda_i = \frac{\hat{\mathbf{z}}^T \tilde{\mathbf{d}}_i}{\hat{\mathbf{z}}^T \mathcal{R}_i \, \hat{\mathbf{z}}} \tag{22}$$

Substituting Eq. (22) into Eq. (21) leads to

$$\hat{\mathbf{d}}_i = \left[I_{(n+1)\times(n+1)} - \frac{\mathcal{R}_i \hat{\mathbf{z}} \, \hat{\mathbf{z}}^T}{\hat{\mathbf{z}}^T \mathcal{R}_i \, \hat{\mathbf{z}}} \right] \tilde{\mathbf{d}}_i \tag{23}$$

where $I_{(n+1)\times(n+1)}$ is an $(n+1) \times (n+1)$ identity matrix. If desired the specific estimates for $\mathbf{h}_i$ and y_i, denoted by $\hat{\mathbf{h}}_i$ and $\hat{y}_i$, respectively, are given by

$$\hat{\mathbf{h}}_i = \tilde{\mathbf{h}}_i - \frac{(\mathcal{R}_{hh_i} \hat{\mathbf{x}} - \mathcal{R}_{hy_i}) e_i}{\hat{\mathbf{z}}^T \mathcal{R}_i \, \hat{\mathbf{z}}} \tag{24a}$$

$$\hat{y}_i = \tilde{y}_i - \frac{(\mathcal{R}_{hy_i}^T \hat{\mathbf{x}} - \mathcal{R}_{yy_i}) e_i}{\hat{\mathbf{z}}^T \mathcal{R}_i \, \hat{\mathbf{z}}} \tag{24b}$$

where $e_i \triangleq \tilde{\mathbf{h}}_i^T \hat{\mathbf{x}} - \tilde{y}_i$. Substituting Eq. (23) into Eq. (19) yields the following unconstrained loss function:

$$J(\hat{\mathbf{x}}) = \frac{1}{2} \sum_{i=1}^{m} \frac{(\tilde{\mathbf{d}}_i^T \hat{\mathbf{z}})^2}{\hat{\mathbf{z}}^T \mathcal{R}_i \, \hat{\mathbf{z}}} \tag{25}$$

Note that Eq. (25) represents a non-convex optimization problem. The necessary condition for optimality gives

$$\frac{\partial J(\hat{\mathbf{x}})}{\partial \hat{\mathbf{x}}} = \sum_{i=1}^{m} \frac{e_i \, \tilde{\mathbf{h}}_i}{\hat{\mathbf{x}}^T \mathcal{R}_{hh_i} \hat{\mathbf{x}} - 2\mathcal{R}_{hy_i}^T \hat{\mathbf{x}} + \mathcal{R}_{yy_i}} - \frac{e_i^2 (\mathcal{R}_{hh_i} \hat{\mathbf{x}} - \mathcal{R}_{hy_i})}{(\hat{\mathbf{x}}^T \mathcal{R}_{hh_i} \hat{\mathbf{x}} - 2\mathcal{R}_{hy_i}^T \hat{\mathbf{x}} + \mathcal{R}_{yy_i})^2} = \mathbf{0} \tag{26}$$

1. Fisher Information Matrix Derivation

To derive the FIM for the TLS estimate $\hat{\mathbf{x}}$, it is possible to determine the FIM for the TLS estimate $\hat{D}$ from the likelihood function given by Eq. (13) and then retrieve the FIM for $\hat{\mathbf{x}}$ from it. It is difficult, however, to derive the FIM for $\hat{D}$ because of the constraint $D\,\mathbf{z} = \mathbf{0}$ with $\mathbf{z} \triangleq [\mathbf{x}^T \quad -1]^T$, which explicitly involves $\mathbf{x}$. The FIM for the joint TLS estimate of $\{\mathbf{x}, \;\; H\}$ will be derived instead.

The likelihood function in Eq. (13) is now treated as a function of $\{\mathbf{x}, H\}$:

$$p(\tilde{D}|\mathbf{x}, H) = \frac{1}{(2\pi)^{m/2} \left[\det{(R)}\right]^{1/2}} \exp\left\{ -\frac{1}{2} \mathrm{vec}^T \left(\tilde{D}^T - D^T(\mathbf{x}, \, H) \right) R^{-1} \mathrm{vec} \left(\tilde{D}^T - D^T(\mathbf{x}, \, H) \right) \right\} \tag{27}$$

with $D(\mathbf{x}, \, H) \triangleq [H \quad H\mathbf{x}]$. In the element-wise uncorrelated and non-stationary case, because $\tilde{\mathbf{d}}_i$ and $\tilde{\mathbf{d}}_j$, $i \neq j$, are independent of each other, the likelihood function reduces to

$$
\begin{aligned}
p(\tilde{D}|\mathbf{x}, H) &= \frac{1}{\prod_{i=1}^{m} \left[\det\left(2\pi\mathcal{R}_i\right)\right]^{1/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{m}\left(\tilde{\mathbf{d}}_i - \mathbf{d}_i(\mathbf{x},\, \mathbf{h}_i)\right)^T \mathcal{R}_i^{-1}\left(\tilde{\mathbf{d}}_i - \mathbf{d}_i(\mathbf{x},\, \mathbf{h}_i)\right)\right\} \\
&= \prod_{i=1}^{m} p\left(\tilde{\mathbf{d}}_i|\mathbf{x}, \mathbf{h}_i\right)
\end{aligned}
\tag{28}
$$

with $\mathbf{d}_i(\mathbf{x},\, \mathbf{h}_i) \triangleq [\mathbf{h}_i^T \;\; \mathbf{h}_i^T\mathbf{x}]^T$ and

$$
p(\tilde{\mathbf{d}}_i|\mathbf{x}, \mathbf{h}_i) \triangleq \frac{1}{\left[\det\left(2\pi\mathcal{R}_i\right)\right]^{1/2}} \exp\left\{-\frac{1}{2}\left(\tilde{\mathbf{d}}_i - \mathbf{d}_i(\mathbf{x},\, \mathbf{h}_i)\right)^T \mathcal{R}_i^{-1}\left(\tilde{\mathbf{d}}_i - \mathbf{d}_i(\mathbf{x},\, \mathbf{h}_i)\right)\right\} \tag{29}
$$

Now, the FIM of the likelihood function $p(\tilde{\mathbf{d}}_i|\mathbf{x}, \mathbf{h}_i)$ is derived. Define

$$
\mathbf{a}_i \triangleq \begin{bmatrix} \mathbf{x} \\ \mathbf{h}_i \end{bmatrix}, \quad p(\tilde{\mathbf{d}}_i|\mathbf{a}_i) \triangleq p(\tilde{\mathbf{d}}_i|\mathbf{x}, \mathbf{h}_i), \quad \mathbf{d}_i(\mathbf{a}_i) \triangleq \mathbf{d}_i(\mathbf{x}, \mathbf{h}_i) \tag{30}
$$

The FIM, F_i^a, for $\mathbf{a}_i$ is

$$
F_i^a = E\left\{\left(\frac{\partial}{\partial \mathbf{a}_i}\ln[p(\tilde{\mathbf{d}}_i|\mathbf{a}_i)]\right)\left(\frac{\partial}{\partial \mathbf{a}_i}\ln[p(\tilde{\mathbf{d}}_i|\mathbf{a}_i)]\right)^T\right\} \tag{31}
$$

The natural logarithm of $p(\tilde{\mathbf{d}}_i|\mathbf{a}_i)$ is

$$
\ln[p(\tilde{\mathbf{d}}_i|\mathbf{a}_i)] = -\frac{1}{2}\left(\tilde{\mathbf{d}}_i - \mathbf{d}_i(\mathbf{a}_i)\right)^T \mathcal{R}_i^{-1}\left(\tilde{\mathbf{d}}_i - \mathbf{d}_i(\mathbf{a}_i)\right) - \frac{1}{2}\ln\det(2\pi\mathcal{R}_i) \tag{32}
$$

Taking partials of the natural logarithm of $p(\tilde{\mathbf{d}}_i|\mathbf{a}_i)$ leads to

$$
\frac{\partial}{\partial \mathbf{a}_i}\ln[p(\tilde{\mathbf{d}}_i|\mathbf{a}_i)] = \begin{bmatrix} 0_{n\times n} & \mathbf{h}_i \\ I_{n\times n} & \mathbf{x} \end{bmatrix} \mathcal{R}_i^{-1}\left(\tilde{\mathbf{d}}_i - \mathbf{d}_i(\mathbf{a}_i)\right) \tag{33}
$$

where $0_{n\times n}$ and $I_{n\times n}$ denote the n-dimensional null matrix and identity matrix, respectively. Because $E\left\{\tilde{\mathbf{d}}_i - \mathbf{d}_i(\mathbf{a}_i)\right\} = \mathbf{0}$, then

$$
E\left\{\frac{\partial}{\partial \mathbf{a}_i}\ln[p(\tilde{\mathbf{d}}_i|\mathbf{a}_i)]\right\} = \begin{bmatrix} 0_{n\times n} & \mathbf{h}_i \\ I_{n\times n} & \mathbf{x} \end{bmatrix} \mathcal{R}_i^{-1} E\left\{\left(\tilde{\mathbf{d}}_i - \mathbf{d}_i(\mathbf{a}_i)\right)\right\} = \mathbf{0} \tag{34}
$$

This means the regularity condition

$$E\left\{\frac{\partial}{\partial \mathbf{a}_i}\ln[p(\tilde{\mathbf{d}}_i|\mathbf{a}_i)]\right\} \triangleq \int \frac{\partial}{\partial \mathbf{a}_i}\ln[p(\tilde{\mathbf{d}}_i|\mathbf{a}_i)]\, p(\tilde{\mathbf{d}}_i|\mathbf{a}_i)\, d\tilde{\mathbf{d}}_i = \int \left[\frac{\partial p(\tilde{\mathbf{d}}_i|\mathbf{a}_i)}{\partial \mathbf{a}_i}\right] d\tilde{\mathbf{d}}_i = \mathbf{0} \tag{35}$$

is satisfied, which is prerequisite for the derivation of the CRLB. Post-multiplying $\partial \ln[p(\tilde{\mathbf{d}}_i|\mathbf{a}_i)]/\partial \mathbf{a}_i$ by its transpose leads to

$$\begin{aligned}\left(\frac{\partial}{\partial \mathbf{a}_i}\ln[p(\tilde{\mathbf{d}}_i|\mathbf{a}_i)]\right)\left(\frac{\partial}{\partial \mathbf{a}_i}\ln[p(\tilde{\mathbf{d}}_i|\mathbf{a}_i)]\right)^T &= \begin{bmatrix} 0_{n\times n} & \mathbf{h}_i \\ I_{n\times n} & \mathbf{x}\end{bmatrix} \mathcal{R}_i^{-1}\left(\tilde{\mathbf{d}}_i - \mathbf{d}_i(\mathbf{a}_i)\right) \\ &\times \left(\tilde{\mathbf{d}}_i - \mathbf{d}_i(\mathbf{a}_i)\right)^T \mathcal{R}_i^{-1} \begin{bmatrix} 0_{n\times n} & \mathbf{h}_i \\ I_{n\times n} & \mathbf{x}\end{bmatrix}^T\end{aligned} \tag{36}$$

Taking the expectation and using $E\left\{\left(\tilde{\mathbf{d}}_i - \mathbf{d}_i(\mathbf{a}_i)\right)\left(\tilde{\mathbf{d}}_i - \mathbf{d}_i(\mathbf{a}_i)\right)^T\right\} = \mathcal{R}_i$ leads to

$$F_i^a = E\left\{\left(\frac{\partial}{\partial \mathbf{a}_i}\ln[p(\tilde{\mathbf{d}}_i|\mathbf{a}_i)]\right)\left(\frac{\partial}{\partial \mathbf{a}_i}\ln[p(\tilde{\mathbf{d}}_i|\mathbf{a}_i)]\right)^T\right\} = \begin{bmatrix} 0_{n\times n} & \mathbf{h}_i \\ I_{n\times n} & \mathbf{x}\end{bmatrix} \mathcal{R}_i^{-1} \begin{bmatrix} 0_{n\times n} & \mathbf{h}_i \\ I_{n\times n} & \mathbf{x}\end{bmatrix}^T \tag{37}$$

The next step is to derive the FIM for $\hat{\mathbf{x}}$. The total Fisher information for $\hat{\mathbf{x}}$ will be denoted by F and the Fisher information corresponding to a single measurement $\tilde{\mathbf{d}}_i$ will be denoted by F_i. Because $\tilde{\mathbf{d}}_i$ and $\tilde{\mathbf{d}}_j$ are independent of each other and $\mathbf{h}_i$ and $\mathbf{h}_j$ are different for $i \neq j$, then $F = \sum_{i=1}^m F_i$. To see this, consider the partition of F_i^a:

$$F_i^a \triangleq \begin{bmatrix} F_{xx_i} & F_{xh_i} \\ F_{xh_i}^T & F_{hh_i}\end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 0_{n\times n} & \mathbf{h}_i \end{bmatrix} \mathcal{R}_i^{-1} \begin{bmatrix} 0_{n\times n} \\ \mathbf{h}_i^T \end{bmatrix} & \begin{bmatrix} 0_{n\times n} & \mathbf{h}_i \end{bmatrix} \mathcal{R}_i^{-1} \begin{bmatrix} I_{n\times n} \\ \mathbf{x}^T \end{bmatrix} \\ \begin{bmatrix} I_{n\times n} & \mathbf{x} \end{bmatrix} \mathcal{R}_i^{-1} \begin{bmatrix} 0_{n\times n} \\ \mathbf{h}_i^T \end{bmatrix} & \begin{bmatrix} I_{n\times n} & \mathbf{x} \end{bmatrix} \mathcal{R}_i^{-1} \begin{bmatrix} I_{n\times n} \\ \mathbf{x}^T \end{bmatrix} \end{bmatrix} \tag{38}$$

and the augmented FIM for $[\hat{\mathbf{x}}^T,\, \hat{\mathbf{h}}_1^T,\, \ldots,\, \hat{\mathbf{h}}_m^T]^T$:

$$\mathcal{F} \triangleq \begin{bmatrix} \mathcal{F}_{xx} & \mathcal{F}_{xh} \\ \mathcal{F}_{xh}^T & \mathcal{F}_{hh}\end{bmatrix}^{-1} = \begin{bmatrix} \sum_{i=1}^m F_{xx_i} & F_{xh_1} & \cdots & F_{xh_m} \\ F_{xh_1}^T & F_{hh_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ F_{xh_1}^T & 0 & \cdots & F_{hh_m} \end{bmatrix} \tag{39}$$

Note that each F_{xx_i} is a rank-one matrix, each F_{hh_i} is nonsingular, and for $m \geq n$, $\mathcal{F}$ is nonsingular. Applying the matrix inversion lemma to $\mathcal{F}$ leads to

$$F^{-1} = \mathcal{F}_{xx} = \left(\sum_{i=1}^{m} F_{xx_i} - \begin{bmatrix} F_{xh_1} & \cdots & F_{xh_m} \end{bmatrix} \begin{bmatrix} F_{hh_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & F_{hh_m} \end{bmatrix}^{-1} \begin{bmatrix} F_{xh_1}^T \\ \vdots \\ F_{xh_m}^T \end{bmatrix} \right)^{-1}$$

$$= \left(\sum_{i=1}^{m} \left(F_{xx_i} - F_{xh_i} F_{hh_i}^{-1} F_{xh_i}^T \right) \right)^{-1} \tag{40}$$

or equivalently $F = \sum_{i=1}^{m} F_i$, where $F_i = F_{xx_i} - F_{xh_i} F_{hh_i}^{-1} F_{xh_i}^T$. Note that each F_i is rank-one because $F_i \mathbf{v}_i = \mathbf{0}$ for any $\mathbf{v}_i \perp \mathbf{h}_i$. The set of $\mathbf{v}_i$ form an $(n-1)$-dimensional subspace. Therefore, the rank of F_i must be one.

Partition the inverse of the matrix $\mathcal{R}_i$ in Eq. (15) as

$$\mathcal{R}_i^{-1} \triangleq \begin{bmatrix} \Gamma_i & \boldsymbol{\beta}_i \\ \boldsymbol{\beta}_i^T & \vartheta_i \end{bmatrix} \tag{41}$$

where Γ_i is an $n \times n$ matrix, $\boldsymbol{\beta}_i$ is an $n \times 1$ vector and ϑ_i is a scalar. The elements $\mathcal{R}_{hh_i}$, $\mathcal{R}_{hy_i}$ and $\mathcal{R}_{yy_i}$, written in terms of Γ_i, $\boldsymbol{\beta}_i$ and ϑ_i, are given by

$$\mathcal{R}_{hh_i} = \left(\Gamma_i - \boldsymbol{\beta}_i \boldsymbol{\beta}_i^T / \vartheta_i \right)^{-1} \tag{42a}$$

$$\mathcal{R}_{hy_i} = -\mathcal{R}_{hh_i} \boldsymbol{\beta}_i / \vartheta_i \tag{42b}$$

$$\mathcal{R}_{yy_i} = \frac{1}{\vartheta_i} + \frac{\boldsymbol{\beta}_i^T \mathcal{R}_{hh_i} \boldsymbol{\beta}_i}{\vartheta_i^2} \tag{42c}$$

Because F_i is rank-one, the following equations are equivalent:

$$\mathbf{h}_i^T F_i \mathbf{h}_i = \frac{h_i^4}{\mathbf{z}^T \mathcal{R}_i \, \mathbf{z}} \tag{43a}$$

$$\eta_i^{-1} \triangleq \left(\mathbf{h}_i^T F_{xx_i} \mathbf{h}_i - \mathbf{h}_i^T F_{xh_i} F_{hh_i}^{-1} F_{xh_i}^T \mathbf{h}_i \right)^{-1} = \frac{\mathbf{z}^T \mathcal{R}_i \, \mathbf{z}}{h_i^4} \tag{43b}$$

$$F_i = \frac{\mathbf{h}_i \mathbf{h}_i^T}{\mathbf{z}^T \mathcal{R}_i \, \mathbf{z}} \tag{43c}$$

where $h_i^2 \triangleq \mathbf{h}_i^T \mathbf{h}_i$. The proof of Eq. (43b) is now shown. Define the following variables: $\mathcal{A}_i \triangleq \mathbf{h}_i^T F_{xx_i} \mathbf{h}_i$, $\mathcal{B}_i \triangleq \mathbf{h}_i^T F_{xh_i}$, $\mathcal{C}_i \triangleq F_{xh_i}^T \mathbf{h}_i = \mathcal{B}^T$ and $\mathcal{D}_i \triangleq F_{hh_i}$. Explicitly computing $\mathcal{A}_i$ gives

$$\mathcal{A}_i = \mathbf{h}_i^T \begin{bmatrix} 0_{n \times n} & \mathbf{h}_i \end{bmatrix} \begin{bmatrix} \Gamma_i & \boldsymbol{\beta}_i \\ \boldsymbol{\beta}_i^T & \vartheta_i \end{bmatrix} \begin{bmatrix} 0_{n \times n} \\ \mathbf{h}_i^T \end{bmatrix} \mathbf{h}_i = h_i^4 \, \vartheta_i \tag{44}$$

Explicitly computing $\mathcal{B}_i$ gives

$$\mathcal{B}_i = \mathbf{h}_i^T \begin{bmatrix} 0_{n\times n} & \mathbf{h}_i \end{bmatrix} \begin{bmatrix} \Gamma_i & \boldsymbol{\beta}_i \\ \boldsymbol{\beta}_i^T & \vartheta_i \end{bmatrix} \begin{bmatrix} I_{n\times n} \\ \mathbf{x}^T \end{bmatrix} = h_i^2(\boldsymbol{\beta}_i^T + \vartheta_i \mathbf{x}^T) \tag{45}$$

Explicitly computing $\mathcal{D}_i$ gives

$$\mathcal{D}_i = \begin{bmatrix} I_{n\times n} & \mathbf{x} \end{bmatrix} \begin{bmatrix} \Gamma_i & \boldsymbol{\beta}_i \\ \boldsymbol{\beta}_i^T & \vartheta_i \end{bmatrix} \begin{bmatrix} I_{n\times n} \\ \mathbf{x}^T \end{bmatrix} = \Gamma_i + \mathbf{x}\,\boldsymbol{\beta}_i^T + \boldsymbol{\beta}_i \mathbf{x} + \vartheta_i \mathbf{x}\,\mathbf{x}^T \tag{46}$$

By the matrix inversion lemma, $\left(\mathcal{A}_i - \mathcal{B}_i \mathcal{D}_i^{-1} \mathcal{C}_i\right)^{-1} = \mathcal{A}_i^{-1} + \mathcal{A}_i^{-1} \mathcal{B}_i \left(\mathcal{D}_i - \mathcal{C}_i \mathcal{A}_i^{-1} \mathcal{B}_i\right)^{-1} \mathcal{C}_i \mathcal{A}_i^{-1}$. Substituting Eqs. (44), (45) and (46) into $\mathcal{D}_i - \mathcal{C}_i \mathcal{A}_i^{-1} \mathcal{B}_i$ yields $\mathcal{D}_i - \mathcal{C}_i \mathcal{A}_i^{-1} \mathcal{B}_i = \Gamma_i - \boldsymbol{\beta}_i \boldsymbol{\beta}_i^T / \vartheta_i$. Using Eq. (42a) gives $\left(\mathcal{D}_i - \mathcal{C}_i \mathcal{A}_i^{-1} \mathcal{B}_i\right)^{-1} = \mathcal{R}_{hh_i}$. So, η_i^{-1} in Eq. (43b) is explicitly given by

$$\begin{aligned} \eta_i^{-1} &= \frac{1}{h_i^4\, \vartheta_i} + \frac{1}{\vartheta_i^2} \left(\boldsymbol{\beta}_i^T + \vartheta_i \mathbf{x}^T\right) \mathcal{R}_{hh_i} \left(\boldsymbol{\beta}_i + \vartheta_i \mathbf{x}\right) \\ &= \frac{1}{h_i^4} \left(\mathbf{x}^T \mathcal{R}_{hh_i} \mathbf{x} + \frac{2}{\vartheta_i} \mathbf{x}^T \mathcal{R}_{hh_i} \boldsymbol{\beta}_i + \frac{1}{\vartheta_i} + \frac{1}{\vartheta_i^2} \boldsymbol{\beta}_i^T \mathcal{R}_{hh_i} \boldsymbol{\beta}_i \right) \\ &= \frac{1}{h_i^4} \left(\mathbf{x}^T \mathcal{R}_{hh_i} \mathbf{x} - 2\mathbf{x}^T \mathcal{R}_{hy_i} + \mathcal{R}_{yy_i} \right) \\ &= \frac{\mathbf{z}^T \mathcal{R}_i\, \mathbf{z}}{h_i^4} \end{aligned} \tag{47}$$

Therefore, from Eq. (43c) the FIM is given by

$$F = \sum_{i=1}^{m} \frac{\mathbf{h}_i \mathbf{h}_i^T}{\mathbf{z}^T \mathcal{R}_i\, \mathbf{z}} \tag{48}$$

The error covariance matrix of $\hat{\mathbf{x}}$, for an efficient estimator, is given by $P = F^{-1}$. If $\mathcal{R}_{hh_i}$ and $\mathcal{R}_{hy_i}$ are both zero, meaning no errors exist in the measured basis functions, then the FIM reduces down to

$$F = \sum_{i=1}^{m} \mathcal{R}_{yy_i}^{-1} \mathbf{h}_i \mathbf{h}_i^T \tag{49}$$

which is equivalent to the FIM for the standard least squares problem.

2. Error-Covariance Derivation

The error-covariance is now derived using the approach shown by Eq. (10). First it must be shown that the estimate is unbiased to within first-order terms, which is required for the CRLB. Let the estimate be given by its true value plus a perturbation: $\hat{\mathbf{x}} = \mathbf{x} + \boldsymbol{\delta}\mathbf{x}$. The individual numerator parts of Eq. (25) are then given by

$$
\begin{aligned}
(\tilde{\mathbf{d}}_i^T \hat{\mathbf{z}})^2 &= (\tilde{\mathbf{h}}_i^T \hat{\mathbf{x}} - \tilde{y}_i)^2 \\
&= \bar{e}_i^2 + 2\bar{e}_i(\tilde{\mathbf{h}}_i^T \boldsymbol{\delta}\mathbf{x}) + (\tilde{\mathbf{h}}_i^T \boldsymbol{\delta}\mathbf{x})^2
\end{aligned}
\tag{50}
$$

where $\bar{e}_i \triangleq \tilde{\mathbf{h}}_i^T \mathbf{x} - \tilde{y}_i$. The individual denominator parts of Eq. (25) are given by $\hat{\mathbf{z}}^T \mathcal{R}_i \, \hat{\mathbf{z}} = \mathbf{z}^T \mathcal{R}_i \, \mathbf{z} + \boldsymbol{\delta}\mathbf{x}^T \mathcal{R}_{hh_i} \boldsymbol{\delta}\mathbf{x} + 2\mathbf{b}_i^T \boldsymbol{\delta}\mathbf{x}$, where $\mathbf{b}_i \triangleq \mathcal{R}_{hh_i}\mathbf{x} - \mathcal{R}_{hy_i}$. Using the binomial series for a second-order expansion of $(\hat{\mathbf{z}}^T \mathcal{R}_i \, \hat{\mathbf{z}})^{-1}$ leads to the approximation

$$
\begin{aligned}
(\hat{\mathbf{z}}^T \mathcal{R}_i \, \hat{\mathbf{z}})^{-1} \approx (\mathbf{z}^T \mathcal{R}_i \, \mathbf{z})^{-1} &- (\boldsymbol{\delta}\mathbf{x}^T \mathcal{R}_{hh_i} \boldsymbol{\delta}\mathbf{x} + 2\mathbf{b}_i^T \boldsymbol{\delta}\mathbf{x})(\mathbf{z}^T \mathcal{R}_i \, \mathbf{z})^{-2} \\
&+ (\boldsymbol{\delta}\mathbf{x}^T \mathcal{R}_{hh_i} \boldsymbol{\delta}\mathbf{x} + 2\mathbf{b}_i^T \boldsymbol{\delta}\mathbf{x})^2 (\mathbf{z}^T \mathcal{R}_i \, \mathbf{z})^{-3}
\end{aligned}
\tag{51}
$$

Substituting Eqs. (50) and (51) into Eq. (25), and retaining terms dependent only up to second order in $\boldsymbol{\delta}\mathbf{x}$ yields

$$
\begin{aligned}
J(\boldsymbol{\delta}\mathbf{x}) = \sum_{i=1}^{m} & \frac{2\bar{e}_i(\tilde{\mathbf{h}}_i^T \boldsymbol{\delta}\mathbf{x}) + (\tilde{\mathbf{h}}_i^T \boldsymbol{\delta}\mathbf{x})^2}{\mathbf{z}^T \mathcal{R}_i \, \mathbf{z}} - \frac{\bar{e}_i^2 \boldsymbol{\delta}\mathbf{x}^T \mathcal{R}_{hh_i} \boldsymbol{\delta}\mathbf{x} + 2\bar{e}_i^2 \mathbf{b}_i^T \boldsymbol{\delta}\mathbf{x} + 4\bar{e}_i \boldsymbol{\delta}\mathbf{x}^T \mathbf{b}_i \tilde{\mathbf{h}}_i^T \boldsymbol{\delta}\mathbf{x}}{(\mathbf{z}^T \mathcal{R}_i \, \mathbf{z})^2} \\
&+ \frac{4\bar{e}_i^2 \boldsymbol{\delta}\mathbf{x}^T \mathbf{b}_i \mathbf{b}_i^T \boldsymbol{\delta}\mathbf{x}}{(\mathbf{z}^T \mathcal{R}_i \, \mathbf{z})^3}
\end{aligned}
\tag{52}
$$

Taking the partial with respect to $\boldsymbol{\delta}\mathbf{x}$ and setting to resultant to zero for the necessary condition for optimality gives

$$
\begin{aligned}
\frac{\partial J(\boldsymbol{\delta}\mathbf{x})}{\partial \boldsymbol{\delta}\mathbf{x}} = & \left\{ \sum_{i=1}^{m} \frac{\tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^T}{\mathbf{z}^T \mathcal{R}_i \, \mathbf{z}} - \frac{\bar{e}_i \left[\bar{e}_i \mathcal{R}_{hh_i} + 2(\mathbf{b}_i \tilde{\mathbf{h}}_i^T + \tilde{\mathbf{h}}_i \mathbf{b}_i^T) \right]}{(\mathbf{z}^T \mathcal{R}_i \, \mathbf{z})^2} + \frac{4\bar{e}_i^2 \mathbf{b}_i \mathbf{b}_i^T}{(\mathbf{z}^T \mathcal{R}_i \, \mathbf{z})^3} \right\} \boldsymbol{\delta}\mathbf{x} \\
&+ \sum_{i=1}^{m} \frac{\bar{e}_i \tilde{\mathbf{h}}_i}{\mathbf{z}^T \mathcal{R}_i \, \mathbf{z}} - \frac{\bar{e}_i^2 \mathbf{b}_i}{(\mathbf{z}^T \mathcal{R}_i \, \mathbf{z})^2} = \mathbf{0}
\end{aligned}
\tag{53}
$$

The expected value of the matrix on the left-hand side of Eq. (53) is given by

$$
E\left\{ \sum_{i=1}^{m} \frac{\tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^T}{\mathbf{z}^T \mathcal{R}_i \, \mathbf{z}} - \frac{\bar{e}_i \left[\bar{e}_i \mathcal{R}_{hh_i} + 2(\mathbf{b}_i \tilde{\mathbf{h}}_i^T + \tilde{\mathbf{h}}_i \mathbf{b}_i^T) \right]}{(\mathbf{z}^T \mathcal{R}_i \, \mathbf{z})^2} + \frac{4\bar{e}_i^2 \mathbf{b}_i \mathbf{b}_i^T}{(\mathbf{z}^T \mathcal{R}_i \, \mathbf{z})^3} \right\} = \sum_{i=1}^{m} \frac{\mathbf{h}_i \mathbf{h}_i^T}{\mathbf{z}^T \mathcal{R}_i \, \mathbf{z}}
\tag{54}
$$

where $E\{\bar{e}_i^2\} = \mathbf{z}^T \mathcal{R}_i \, \mathbf{z}$, and $E\{\bar{e}_i \tilde{\mathbf{h}}_i\} = \mathbf{b}_i$ and $E\{\bar{e}_i \tilde{\mathbf{h}}_i^T\} = \mathbf{b}_i^T$ have been used.[17] Also note

$$
E\left\{ \sum_{i=1}^{m} \frac{\bar{e}_i \tilde{\mathbf{h}}_i}{\mathbf{z}^T \mathcal{R}_i \, \mathbf{z}} - \frac{\bar{e}_i^2 \mathbf{b}_i}{(\mathbf{z}^T \mathcal{R}_i \, \mathbf{z})^2} \right\} = \sum_{i=1}^{m} \frac{\mathbf{b}_i}{\mathbf{z}^T \mathcal{R}_i \, \mathbf{z}} - \frac{\mathbf{b}_i}{\mathbf{z}^T \mathcal{R}_i \, \mathbf{z}} = \mathbf{0}
\tag{55}
$$

Equations (53), (54) and (55) indicate that

$$\left[\sum_{i=1}^{m} \frac{\mathbf{h}_i \mathbf{h}_i^T}{\mathbf{z}^T \mathcal{R}_i\, \mathbf{z}}\right] E\left\{\boldsymbol{\delta}\mathbf{x}\right\} = \mathbf{0} \tag{56}$$

As stated previously linearly independent basis functions should be employed in practice. For this case the matrix in Eq. (56) is never singular and $E\{\boldsymbol{\delta}\mathbf{x}\} = \mathbf{0}$ must be true. Thus the TLS estimator produces an unbiased estimate to within first-order terms.

The error-covariance is now derived using Eq. (10), where the estimate follows from Eq. (26). Three error sources are present: the first is $\boldsymbol{\delta}\mathbf{x}$ which is the error on $\hat{\mathbf{x}}$, the second is $\boldsymbol{\delta}\mathbf{h}_i$ which is the error on $\tilde{\mathbf{h}}_i$, and the third is δy_i which is the error on $\tilde{y}_i$. Define the expression in Eq. (26) by $\mathbf{g} \triangleq \partial J(\hat{\mathbf{x}})/\partial\hat{\mathbf{x}} = \mathbf{0}$. The partial of $\mathbf{g}$ with respect to $\hat{\mathbf{x}}$ is given by

$$\begin{aligned}
\frac{\partial \mathbf{g}}{\partial \hat{\mathbf{x}}} &= \sum_{i=1}^{m} \frac{\tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^T}{\mathbf{z}^T \mathcal{R}_i\, \mathbf{z}} - \frac{2\bar{e}_i[\tilde{\mathbf{h}}_i\,(\mathcal{R}_{hh_i}\mathbf{x} - \mathcal{R}_{hy_i})^T + (\mathcal{R}_{hh_i}\mathbf{x} - \mathcal{R}_{hy_i})\,\tilde{\mathbf{h}}_i^T]}{(\mathbf{z}^T \mathcal{R}_i\, \mathbf{z})^2} \\
&\quad - \frac{\mathcal{R}_{hh_i}\bar{e}_i^2}{(\mathbf{z}^T \mathcal{R}_i\, \mathbf{z})^2} + \frac{4\bar{e}_i^2(\mathcal{R}_{hh_i}\mathbf{x} - \mathcal{R}_{hy_i})(\mathcal{R}_{hh_i}\mathbf{x} - \mathcal{R}_{hy_i})^T}{(\mathbf{z}^T \mathcal{R}_i\, \mathbf{z})^3} \\
&\quad + \frac{\mathcal{R}_{hh_i}}{\mathbf{z}^T \mathcal{R}_i\, \mathbf{z}} - \frac{2(\mathcal{R}_{hh_i}\mathbf{x} - \mathcal{R}_{hy_i})(\mathcal{R}_{hh_i}\mathbf{x} - \mathcal{R}_{hy_i})^T}{(\mathbf{z}^T \mathcal{R}_i\, \mathbf{z})^2}
\end{aligned} \tag{57}$$

According to Eq. (10) this partial is evaluated at the true values. Since $y_i = \mathbf{h}_i^T\mathbf{x}$ then

$$\left.\frac{\partial \mathbf{g}}{\partial \hat{\mathbf{x}}}\right|_{\text{truth}} = \sum_{i=1}^{m} \frac{\mathbf{h}_i \mathbf{h}_i^T}{\mathbf{z}^T \mathcal{R}_i\, \mathbf{z}} \tag{58}$$

The partial of $\mathbf{g}$ with respect to $\tilde{\mathbf{h}}_i$ evaluated at the true values is given

$$\left.\frac{\partial \mathbf{g}}{\partial \tilde{\mathbf{h}}_i}\right|_{\text{truth}} = \sum_{i=1}^{m} \frac{\mathbf{h}_i \mathbf{x}^T}{\mathbf{z}^T \mathcal{R}_i\, \mathbf{z}} \tag{59}$$

The partial of $\mathbf{g}$ with respect to $\tilde{y}_i$ evaluated at the true values is given

$$\left.\frac{\partial \mathbf{g}}{\partial \tilde{y}_i}\right|_{\text{truth}} = -\sum_{i=1}^{m} \frac{\mathbf{h}_i}{\mathbf{z}^T \mathcal{R}_i\, \mathbf{z}} \tag{60}$$

Then to within first order the following equation is given:

$$-\left(\sum_{i=1}^{m} \frac{\mathbf{h}_i \mathbf{h}_i^T}{\mathbf{z}^T \mathcal{R}_i\, \mathbf{z}}\right) \boldsymbol{\delta}\mathbf{x} = \sum_{i=1}^{m} \frac{\mathbf{h}_i \mathbf{x}^T \boldsymbol{\delta}\mathbf{h}_i}{\mathbf{z}^T \mathcal{R}_i\, \mathbf{z}} - \frac{\mathbf{h}_i \delta y_i}{\mathbf{z}^T \mathcal{R}_i\, \mathbf{z}} \tag{61}$$

The error-covariance, $P \triangleq E\left\{\boldsymbol{\delta}\mathbf{x}\,\boldsymbol{\delta}\mathbf{x}^T\right\}$, is derived from

$$P = F^{-1} E\left\{\left[\sum_{i=1}^{m} \frac{\mathbf{h}_i(\mathbf{x}^T\boldsymbol{\delta}\mathbf{h}_i - \delta y_i)}{\mathbf{z}^T\mathcal{R}_i\,\mathbf{z}}\right]\left[\sum_{i=1}^{m} \frac{\mathbf{h}_i(\mathbf{x}^T\boldsymbol{\delta}\mathbf{h}_i - \delta y_i)}{\mathbf{z}^T\mathcal{R}_i\,\mathbf{z}}\right]^T\right\} F^{-1} \tag{62}$$

where F is given by Eq. (48). Since element-wise uncorrelated terms are assumed, then the expectation in Eq. (62) reduces down to

$$E\left\{\left[\sum_{i=1}^{m} \frac{\mathbf{h}_i(\mathbf{x}^T\boldsymbol{\delta}\mathbf{h}_i - \delta y_i)}{\mathbf{z}^T\mathcal{R}_i\,\mathbf{z}}\right]\left[\sum_{i=1}^{m} \frac{\mathbf{h}_i(\mathbf{x}^T\boldsymbol{\delta}\mathbf{h}_i - \delta y_i)}{\mathbf{z}^T\mathcal{R}_i\,\mathbf{z}}\right]^T\right\} = E\left\{\sum_{i=1}^{m} \frac{\mathbf{h}_i\mathbf{h}_i^T(\mathbf{x}^T\boldsymbol{\delta}\mathbf{h}_i - \delta y_i)^2}{(\mathbf{z}^T\mathcal{R}_i\,\mathbf{z})^2}\right\} \tag{63}$$

Using $E\{(\mathbf{x}^T\boldsymbol{\delta}\mathbf{h}_i - \delta y_i)^2\} = \mathbf{z}^T\mathcal{R}_i\,\mathbf{z}$ leads to

$$P = F^{-1}\left(\sum_{i=1}^{m} \frac{\mathbf{h}_i\mathbf{h}_i^T}{\mathbf{z}^T\mathcal{R}_i\,\mathbf{z}}\right) F^{-1} = F^{-1} \tag{64}$$

Comparing Eqs. (48) and (64) shows that the CRLB is achieved to within first-order terms.

The FIM is evaluated at the respective true values for $\mathbf{h}_i$ and $\mathbf{x}$, which are not available in practice. Either the estimated or measured values are typically used in their place. The expected errors induced by using the measured values are now shown. The estimate is again written by $\hat{\mathbf{x}} = \mathbf{x} + \boldsymbol{\delta}\mathbf{x}$ where the covariance of $\boldsymbol{\delta}\mathbf{x}$ is given by P. The estimate of the FIM, denoted by $\hat{F}$, using the measured values can now be written as

$$\hat{F} = \sum_{i=1}^{m} \frac{(\mathbf{h}_i + \boldsymbol{\delta}\mathbf{h}_i)(\mathbf{h}_i + \boldsymbol{\delta}\mathbf{h}_i)^T}{\mathbf{z}^T\mathcal{R}_i\,\mathbf{z} + \boldsymbol{\delta}\mathbf{x}^T\mathcal{R}_{hh_i}\boldsymbol{\delta}\mathbf{x} + 2(\mathcal{R}_{hh_i}\mathbf{x} - \mathcal{R}_{hy_i})\boldsymbol{\delta}\mathbf{x}} \tag{65}$$

Using Eq. (51) leads to the approximation

$$\begin{aligned}\hat{F} \approx \sum_{i=1}^{m} & \frac{(\mathbf{h}_i + \boldsymbol{\delta}\mathbf{h}_i)(\mathbf{h}_i + \boldsymbol{\delta}\mathbf{h}_i)^T}{\mathbf{z}^T\mathcal{R}_i\,\mathbf{z}} \\ & - \frac{[\boldsymbol{\delta}\mathbf{x}^T\mathcal{R}_{hh_i}\boldsymbol{\delta}\mathbf{x} + 2(\mathcal{R}_{hh_i}\mathbf{x} - \mathcal{R}_{hy_i})\boldsymbol{\delta}\mathbf{x}](\mathbf{h}_i + \boldsymbol{\delta}\mathbf{h}_i)(\mathbf{h}_i + \boldsymbol{\delta}\mathbf{h}_i)^T}{(\mathbf{z}^T\mathcal{R}_i\,\mathbf{z})^2}\end{aligned} \tag{66}$$

Computing $\delta F \triangleq E\{\hat{F}\} - F$ gives

$$\delta F = \sum_{i=1}^{m} \frac{\mathcal{R}_{hh_i}}{\mathbf{z}^T\mathcal{R}_i\,\mathbf{z}} - \frac{\mathrm{Tr}(P\,\mathcal{R}_{hh_i})(\mathbf{h}_i\mathbf{h}_i^T + \mathcal{R}_{hh_i})}{(\mathbf{z}^T\mathcal{R}_i\,\mathbf{z})^2} \tag{67}$$

Using the measurements of $\mathbf{h}_i$ to compute the FIM will provide an adequate approximation if $F >> \delta F$ is satisfied. If the signal-to-noise ratio is large then $\mathbf{h}_i\mathbf{h}_i^T >> \mathcal{R}_{hh_i}$. This leads to the following simplification for $F >> \delta F$:

$$\sum_{i=1}^{m} \frac{\mathbf{h}_i \mathbf{h}_i^T}{\mathbf{z}^T \mathcal{R}_i \, \mathbf{z}} >> - \sum_{i=1}^{m} \frac{\mathrm{Tr}(P \, \mathcal{R}_{hh_i}) \mathbf{h}_i \mathbf{h}_i^T}{(\mathbf{z}^T \mathcal{R}_i \, \mathbf{z})^2} \tag{68}$$

Using $||\mathbf{h}_i \mathbf{h}_i^T||_{\mathrm{F}} = \mathbf{h}_i^T \mathbf{h}_i$ and looking at each individual component in the summation leads to the following requirement:

$$\mathbf{z}^T \mathcal{R}_i \, \mathbf{z} >> \mathrm{Tr}(P \, \mathcal{R}_{hh_i}) \tag{69}$$

The term $\mathbf{z}^T \mathcal{R}_i \, \mathbf{z}$ is equal to $\mathrm{Tr}(\mathbf{x} \, \mathbf{x}^T \mathcal{R}_{hh_i}) - 2\mathcal{R}_{hy_i}^T \mathbf{x} + \mathcal{R}_{yy_i}$. This indicates that the inequality in Eq. (69) will be satisfied if the signal-to-noise ratio is large and the estimate error-covariance is small. That is, the term $\mathrm{Tr}(P \, \mathcal{R}_{hh_i})$ will be second-order in nature while $\mathbf{z}^T \mathcal{R}_i \, \mathbf{z}$ is first-order in nature. If estimates of $\mathbf{h}_i$ are used in place of measurements, then $\mathcal{R}_{hh_i}$ in Eq. (67) is replaced with the covariance of the estimates.

The covariance of $\hat{\mathbf{h}}_i$ and variance of $\hat{y}_i$ are now derived, which are defined by $P_{hh_i} = E\{(\hat{\mathbf{h}}_i - E\{\hat{\mathbf{h}}_i\})(\hat{\mathbf{h}}_i - E\{\hat{\mathbf{h}}_i\})^T\}$ and $P_{yy_i} = E\{(\hat{y}_i - E\{\hat{y}_i\})(\hat{y}_i - E\{\hat{y}_i\})^T\}$, respectively. The cross-covariance $P_{hy_i} = E\{(\hat{\mathbf{h}}_i - E\{\hat{\mathbf{h}}_i\})(\hat{y}_i - E\{\hat{y}_i\})^T\}$ is also derived which is used to derive the covariance of $\hat{\mathbf{d}}_i$. These covariances are useful for many applications. For example, the estimate $\hat{y}_i$ may be employed in a Kalman filter to provide filtered estimates. The correct variance of $\hat{y}_i$ is need to ensure proper tuning in the Kalman filter design. Using $\hat{\mathbf{x}} = \mathbf{x} + \boldsymbol{\delta}\mathbf{x}$ and Eq. (51), as well as $\tilde{\mathbf{h}}_i = \mathbf{h}_i + \boldsymbol{\delta}\mathbf{h}_i$ and $\tilde{y}_i = y_i + \delta y_i$, in Eq. (24a) gives

$$\begin{aligned} \hat{\mathbf{h}}_i = \mathbf{h}_i + \boldsymbol{\delta}\mathbf{h}_i &- \frac{(\mathcal{R}_{hh_i}\mathbf{x} + \mathcal{R}_{hh_i}\boldsymbol{\delta}\mathbf{x} - \mathcal{R}_{hy_i})(\mathbf{h}_i^T \boldsymbol{\delta}\mathbf{x} + \mathbf{h}_i^T \boldsymbol{\delta}\mathbf{h}_i - \delta y_i)}{\mathbf{z}^T \mathcal{R}_i \, \mathbf{z}} \\ &+ \frac{(\mathcal{R}_{hh_i}\mathbf{x} + \mathcal{R}_{hh_i}\boldsymbol{\delta}\mathbf{x} - \mathcal{R}_{hy_i})(\mathbf{h}_i^T \boldsymbol{\delta}\mathbf{x} + \mathbf{h}_i^T \boldsymbol{\delta}\mathbf{h}_i - \delta y_i)(\boldsymbol{\delta}\mathbf{x}^T \mathcal{R}_{hh_i}\boldsymbol{\delta}\mathbf{x} + 2\mathbf{b}_i^T \boldsymbol{\delta}\mathbf{x})}{(\mathbf{z}^T \mathcal{R}_i \, \mathbf{z})^2} \end{aligned} \tag{70}$$

Retaining terms up to second order only, then

$$E\left\{\hat{\mathbf{h}}_i\right\} = \mathbf{h}_i - \frac{\mathcal{R}_{hh_i} P \, \mathbf{h}_i}{\mathbf{z}^T \mathcal{R}_i \, \mathbf{z}} + \frac{2(\mathbf{h}_i^T P \, \mathbf{b}_i)\mathbf{b}_i}{(\mathbf{z}^T \mathcal{R}_i \, \mathbf{z})^2} \tag{71}$$

The last two terms on the right-hand side of Eq. (71) are second order in nature. Thus, to within first order $E\{\hat{\mathbf{h}}_i\} = \mathbf{h}_i$, which indicates that the estimate is unbiased. Define the following matrices:

$$M_{h_i} \triangleq \left[I_{n\times n} - \frac{\mathbf{b}_i \mathbf{x}^T}{\mathbf{z}^T \mathcal{R}_i\, \mathbf{z}} \quad \frac{\mathbf{b}_i}{\mathbf{z}^T \mathcal{R}_i\, \mathbf{z}} \right] \tag{72a}$$

$$N_{h_i} \triangleq \frac{\mathbf{b}_i \mathbf{h}_i^T}{\mathbf{z}^T \mathcal{R}_i\, \mathbf{z}} \tag{72b}$$

$$M_{y_i} \triangleq \left[-\frac{\beta_i \mathbf{x}^T}{\mathbf{z}^T \mathcal{R}_i\, \mathbf{z}} \quad 1 + \frac{\beta_i}{\mathbf{z}^T \mathcal{R}_i\, \mathbf{z}} \right] \tag{72c}$$

$$N_{y_i} \triangleq \frac{\beta_i \mathbf{h}_i^T}{\mathbf{z}^T \mathcal{R}_i\, \mathbf{z}} \tag{72d}$$

where $\beta_i \triangleq \mathcal{R}_{hy_i}^T \mathbf{x} - \mathcal{R}_{yy_i}$ and $I_{n\times n}$ is an $n \times n$ identity matrix. Then the covariance of $\hat{\mathbf{h}}_i$ up to first-order terms is given by

$$P_{hh_i} = M_{h_i} \mathcal{R}_i M_{h_i}^T + N_{h_i} P\, N_{h_i}^T \tag{73}$$

The term $N_{h_i} P\, N_{h_i}^T$ is often much smaller than $M_{h_i} \mathcal{R}_i M_{h_i}^T$ and can be ignored in most cases. In a similar fashion the expected value of $\hat{y}_i$ can be shown to be given by

$$E\left\{\hat{y}_i\right\} = y_i - \frac{\mathcal{R}_{hy_i}^T P\, \mathbf{h}_i}{\mathbf{z}^T \mathcal{R}_i\, \mathbf{z}} + \frac{2(\mathbf{h}_i^T P\, \mathbf{b}_i)\beta_i}{(\mathbf{z}^T \mathcal{R}_i\, \mathbf{z})^2} \tag{74}$$

As before the estimate is unbiased to within first order. Then the variance of $\hat{y}_i$ up to first-order terms is given by

$$P_{yy_i} = M_{y_i} \mathcal{R}_i M_{y_i}^T + N_{y_i} P\, N_{y_i}^T \tag{75}$$

Also, the cross-covariance is given by

$$P_{hy_i} = M_{h_i} \mathcal{R}_i M_{y_i}^T + N_{h_i} P\, N_{y_i}^T \tag{76}$$

Finally, the covariance of $\hat{\mathbf{d}}_i$, denoted by P_{dd_i}, is given by

$$P_{dd_i} = \begin{bmatrix} P_{hh_i} & P_{hy_i} \\ P_{hy_i}^T & P_{yy_i} \end{bmatrix} \tag{77}$$

The matrices in Eqs. (72) should be computed using the estimated values in practice because they are derived using $\mathbf{h}_i^T \mathbf{x} - \mathbf{y}_i = 0$. The estimates also obey $\hat{\mathbf{h}}_i^T \hat{\mathbf{x}}_i - \hat{\mathbf{y}}_i = 0$ by virtue of the required constraint in Eq. (14), but using the measurements with $\tilde{\mathbf{h}}_i^T \hat{\mathbf{x}}_i - \tilde{\mathbf{y}}_i = 0$ is not zero in practice. Therefore it is more accurate to use the estimates rather than the measurements to compute these matrices. Also, note that P_{dd_i} can be written by

$$P_{dd_i} = \begin{bmatrix} M_{h_i} & N_{h_i} \\ M_{y_i} & N_{y_i} \end{bmatrix} \begin{bmatrix} \mathcal{R}_i & 0_{(n+1)\times n} \\ 0^T_{(n+1)\times n} & P \end{bmatrix} \begin{bmatrix} M_{h_i} & N_{h_i} \\ M_{y_i} & N_{y_i} \end{bmatrix}^T \tag{78}$$

where $0_{(n+1)\times n}$ is an $(n+1)\times n$ matrix of zeros. This shows that P_{dd_i} is a singular matrix, which is due to the constraint $\hat{\mathbf{d}}_i^T \hat{\mathbf{z}} = 0$.

B. Element-Wise Uncorrelated and Stationary Case

For this case R is assumed to have a block diagonal structure of the form $R = \text{blkdiag}\begin{bmatrix} \mathcal{R} & \cdots & \mathcal{R} \end{bmatrix}$, where $\mathcal{R}$ is an $(n+1)\times(n+1)$ matrix. The solution to this problem is presented in Ref. 9. First the Cholesky decomposition of $\mathcal{R}$ is taken: $\mathcal{R} = C^T C$ where C is defined as an upper block diagonal matrix. Partition the inverse as

$$C^{-1} = \begin{bmatrix} C_{11} & \mathbf{c} \\ & \\ \mathbf{0}^T & c_{22} \end{bmatrix} \tag{79}$$

where C_{11} is an $n\times n$ matrix, $\mathbf{c}$ is an $n\times 1$ vector and c_{22} is a scalar. The solution is given by taking the singular value decomposition of the following matrix:

$$\tilde{D}C^{-1} = \tilde{U}\tilde{S}\tilde{V}^T \tag{80}$$

where the reduced form is used, with $\tilde{S} = \text{diag}\begin{bmatrix} \tilde{s}_1 & \cdots & \tilde{s}_{n+1} \end{bmatrix}$, $\tilde{U}$ is an $m\times(n+1)$ matrix and $\tilde{V}$ is an $(n+1)\times(n+1)$ matrix partitioned in a similar manner as the C^{-1} matrix:

$$\tilde{V} = \begin{bmatrix} \tilde{V}_{11} & \tilde{\mathbf{v}} \\ & \\ \tilde{\mathbf{w}}^T & \tilde{v}_{22} \end{bmatrix} \tag{81}$$

The total least squares solution assuming an isotropic error process, i.e. $\mathcal{R}$ is a scalar times an identity matrix, is

$$\hat{\mathbf{x}}_{\text{I}} = -\tilde{v}_{22}^{-1}\tilde{\mathbf{v}} \tag{82}$$

The final solution is then given by

$$\hat{\mathbf{x}} = c_{22}^{-1}(C_{11}\hat{\mathbf{x}}_{\text{I}} - \mathbf{c}) \tag{83}$$

Clearly if the error process is isotropic then $\hat{\mathbf{x}} = \hat{\mathbf{x}}_\mathrm{I}$, because $C_{11} = \sigma^{-2} I_{n\times n}$ where $I_{n\times n}$ is an $n \times n$ identity matrix, $\mathbf{c} = \mathbf{0}$ and $c_{22} = \sigma^{-2}$ where σ^2 is the variance associated with the isotropic process. The estimate for D is given by

$$\hat{D} = \tilde{U}_n \tilde{S}_n \tilde{V}_n^T C \tag{84}$$

where $\tilde{U}_n$ is the truncation of the matrix $\tilde{U}$ to $m \times n$, $\tilde{S}_n$ is the truncation of the matrix $\tilde{S}$ to $n \times n$, and $\tilde{V}_n$ is the truncation of the matrix $\tilde{V}$ to $(n+1) \times n$.

The solution summary is as follows. First form the augmented matrix, $\tilde{D}$, in Eq. (12) and take the Cholesky decomposition of the covariance $\mathcal{R}$. Take the inverse of C and obtain the matrix partitions shown in Eq. (79). Then take the reduced-form singular value decomposition of the matrix $\tilde{D}C^{-1}$, as shown in Eq. (80), and obtain the matrix partitions shown in Eq. (81). Obtain the isotropic solution using Eq. (82) and obtain the final solution using Eq. (83).

The error-covariance for the estimate in Eq. (83) is derived. For this case the FIM in Eq. (48) simplifies to

$$F = \frac{1}{\mathbf{z}^T \mathcal{R}\, \mathbf{z}} \sum_{i=1}^{m} \mathbf{h}_i \mathbf{h}_i^T \tag{85}$$

The covariance requires an inverse of an $n \times n$ matrix. Note that if $\mathcal{R}$ is isotropic then Eq. (85) matches with the result shown in Ref. 10. Since a closed-form solution exists for the element-wise uncorrelated and stationary case, then an approximation for the error-covariance can be derived directly from the solution. The derivation begins by applying perturbations to the vector $\tilde{\mathbf{v}}$ and scalar $\tilde{v}_{22}$:

$$\tilde{\mathbf{v}} = \mathbf{v} + \boldsymbol{\delta}\mathbf{v} \tag{86a}$$

$$\tilde{v}_{22} = v_{22} + \delta v_{22} \tag{86b}$$

where $\mathbf{v}$ is the true value of $\tilde{\mathbf{v}}$, $\boldsymbol{\delta}\mathbf{v}$ is its respective perturbation, v_{22} is the true value of $\tilde{v}_{22}$, and δv_{22} is its respective perturbation. Using the binomial series the first-order expansion of $(v_{22} + \delta v_{22})^{-1}$ is given by

$$(v_{22} + \delta v_{22})^{-1} \approx v_{22}^{-1} - v_{22}^{-2}\, \delta v_{22} \tag{87}$$

Substituting Eqs. (86) and (87) into Eq. (82) and ignoring higher-order terms leads to $\hat{\mathbf{x}}_\mathrm{I} - \mathbf{x}_\mathrm{I} \approx -v_{22}^{-1} \boldsymbol{\delta}\mathbf{v} + v_{22}^{-2} \delta v_{22}\, \mathbf{v}$, where $\mathbf{x}_\mathrm{I} \triangleq -v_{22}^{-1} \mathbf{v}$. Assuming that $\boldsymbol{\delta}\mathbf{v}$ and δv_{22} are random variables leads to the following error-covariance matrix for the isotropic total least squares solution:

$$\begin{aligned} P_{\rm I} &\triangleq E\left\{(\hat{\mathbf{x}}_{\rm I}-\mathbf{x}_{\rm I})(\hat{\mathbf{x}}_{\rm I}-\mathbf{x}_{\rm I})^T\right\} \\ &= v_{22}^{-2}E\left\{\boldsymbol{\delta}\mathbf{v}\,\boldsymbol{\delta}\mathbf{v}^T\right\}+v_{22}^{-4}\mathbf{v}\,\mathbf{v}^TE\left\{\delta v_{22}^2\right\}-v_{22}^{-3}E\left\{\delta v_{22}\boldsymbol{\delta}\mathbf{v}\right\}\mathbf{v}^T-v_{22}^{-3}\mathbf{v}E\left\{\delta v_{22}\boldsymbol{\delta}\mathbf{v}^T\right\} \end{aligned} \tag{88}$$

Using $\mathbf{x}=c_{22}^{-1}(C_{11}\mathbf{x}_{\rm I}-\mathbf{c})$ and Eq. (83) leads to $\hat{\mathbf{x}}-\mathbf{x}=c_{22}^{-1}C_{11}(\hat{\mathbf{x}}_{\rm I}-\mathbf{x})$. Therefore, the error-covariance matrix for the total least squares solution is given by

$$P\triangleq E\left\{(\hat{\mathbf{x}}-\mathbf{x})(\hat{\mathbf{x}}-\mathbf{x})^T\right\}=c_{22}^{-2}C_{11}P_{\rm I}\,C_{11}^T \tag{89}$$

Note that $P_{\rm I}$ is evaluated at the true values, $\mathbf{v}$ and v_{22}, which are not available in practice. These can be replaced with $\tilde{\mathbf{v}}$ and $\tilde{v}_{22}$ in practice, which leads to higher-order error effects that can be ignored for large signal-to-noise ratios as stated previously.

The expectations in Eq. (88) now need to be derived to complete the derivation of the error-covariance. Using Eq. (80) and the fact that the errors are stationary gives

$$C^{-T}E\left\{\tilde{\mathbf{d}}_i\tilde{\mathbf{d}}_i^T\right\}C^{-1}=C^{-T}\mathcal{R}\,C^{-1}=C^{-T}C^TC\,C^{-1}=I_{(n+1)\times(n+1)} \tag{90}$$

where C^{-T} is defined as the transpose of the inverse of C. The goal here is to compute the following quantity:

$$B\triangleq\begin{bmatrix} E\left\{\boldsymbol{\delta}\mathbf{v}\,\boldsymbol{\delta}\mathbf{v}^T\right\} & E\left\{\delta v_{22}\boldsymbol{\delta}\mathbf{v}\right\} \\ E\left\{\delta v_{22}\boldsymbol{\delta}\mathbf{v}^T\right\} & E\left\{\delta v_{22}^2\right\} \end{bmatrix} \tag{91}$$

Let $\tilde{\mathbf{p}}\triangleq[\tilde{\mathbf{v}}^T\ \ \tilde{v}_{22}]^T$. Using the analogy from Eq. (10), together with Eq. (90), the matrix B is approximated using the following matrix:

$$\tilde{B}=\left[\frac{\partial\tilde{\mathbf{p}}}{\partial\text{vec}(\tilde{D}^T)}\right]\left[\frac{\partial\tilde{\mathbf{p}}}{\partial\text{vec}(\tilde{D}^T)}\right]^T \tag{92}$$

where "measured" values are used in place of estimated values, which again leads to higher-order error effects that can be ignored for large signal-to-noise ratios.

A method to compute the Jacobian of the singular value decomposition is shown in Ref. 18, which is reviewed here. The derivatives of the singular values are given by

$$\frac{\partial\tilde{s}_k}{\partial\tilde{d}_{ij}}=\tilde{u}_{ik}\,\tilde{v}_{jk} \tag{93}$$

where $\tilde{s}_k$ is the $k^{\rm th}$ diagonal element of the matrix $\tilde{S}$, $\tilde{d}_{ij}$ is $ij^{\rm th}$ element of $\tilde{D}$, $\tilde{u}_{ik}$ is the $ik^{\rm th}$ element of $\tilde{U}$, and $\tilde{v}_{jk}$ is the $jk^{\rm th}$ element of $\tilde{V}$. To determine the partials of the matrices $\tilde{U}$ and $\tilde{V}$, first the following set of linear equations must be solved for $\omega_{\tilde{U}_{k\ell}}^{ij}$ and

$\omega^{ij}_{\tilde{V}_{k\ell}}$: $\tilde{s}_\ell\, \omega^{ij}_{\tilde{U}_{k\ell}} + \tilde{s}_k\, \omega^{ij}_{\tilde{V}_{k\ell}} = \tilde{u}_{ik}\, \tilde{v}_{j\ell}$ and $\tilde{s}_k\, \omega^{ij}_{\tilde{U}_{k\ell}} + \tilde{s}_\ell\, \omega^{ij}_{\tilde{V}_{k\ell}} = -\tilde{u}_{i\ell}\, \tilde{v}_{jk}$, where $\omega^{ij}_{\tilde{U}_{k\ell}}$ and $\omega^{ij}_{\tilde{V}_{k\ell}}$ are the $k\ell^{\text{th}}$ elements of the skew symmetric matrices $\Omega^{ij}_{\tilde{U}}$ and $\Omega^{ij}_{\tilde{V}}$, respectively. Note because these matrices are skew symmetric then only the upper triangular elements need to be computed to determine the matrices. The partials are then given by

$$\frac{\partial \tilde{U}}{\partial \tilde{d}_{ij}} = \tilde{U}\, \Omega^{ij}_{\tilde{U}} \tag{94a}$$

$$\frac{\partial \tilde{V}}{\partial \tilde{d}_{ij}} = -\tilde{V}\, \Omega^{ij}_{\tilde{V}} \tag{94b}$$

More details can be found in Ref. 18.

The procedure to compute the partials can be computationally expensive. However, for the total least squares problem, only the partial of the last column of $\tilde{V}$, i.e. the vector $\tilde{\mathbf{p}}$, is required which significantly reduces the computations. Specifically only the last column of $\Omega^{ij}_{\tilde{V}}$ is required. The first step is to compute elements of the $(n+1) \times 1$ vector $\boldsymbol{\omega}^{ij} = [\omega^{ij}_1 \;\; \cdots \;\; \omega^{ij}_n \;\; 0]^T$, with

$$\omega^{ij}_k = \frac{1}{\tilde{s}^2_k - \tilde{s}^2_{n+1}} (\tilde{s}_k\, \tilde{u}_{ik}\, \tilde{v}_{j\,n+1} + \tilde{s}_{n+1}\, \tilde{u}_{i\,n+1}\, \tilde{v}_{jk}) \tag{95}$$

for $k = 1, 2\; \ldots, n$. Then the following $(n+1) \times (n+1)$ matrix is formed:

$$\Omega_i \triangleq \left[\boldsymbol{\omega}^{i1} \,|\; \cdots \;|\, \boldsymbol{\omega}^{i\,n+1}\right] \tag{96}$$

Using the block diagonal structure of R allows Eq. (92) to be computed simply by

$$\tilde{B} = \tilde{V} \left[\sum_{i=1}^{m} \Omega_i\, \Omega_i^T\right] \tilde{V}^T \tag{97}$$

Partition the matrix $\tilde{B}$ into

$$\tilde{B} = \begin{bmatrix} \tilde{B}_{11} & \tilde{\mathbf{b}} \\ \tilde{\mathbf{b}}^T & \tilde{b}_{22} \end{bmatrix} \tag{98}$$

where $\tilde{B}_{11}$ is an $n \times n$ matrix, $\tilde{\mathbf{b}}$ is an $n \times 1$ vector and $\tilde{b}_{22}$ is a scalar. Equation (88), evaluated using the tilde quantities, is now given by

$$P_{\text{I}} = \tilde{v}_{22}^{-2} \left[\tilde{B}_{11} + \tilde{v}_{22}^{-2} \tilde{b}_{22} \tilde{\mathbf{v}}\, \tilde{\mathbf{v}}^T - \tilde{v}_{22}^{-1} (\tilde{\mathbf{b}}\, \tilde{\mathbf{v}}^T + \tilde{\mathbf{v}}\, \tilde{\mathbf{b}}^T)\right] \tag{99}$$

Then the error-covariance can now be computed using Eq. (89):

$$P = \tilde{v}_{22}^{-2} c_{22}^{-2} C_{11} \left[\tilde{B}_{11} + \tilde{v}_{22}^{-2} \tilde{b}_{22} \tilde{\mathbf{v}}\, \tilde{\mathbf{v}}^T - \tilde{v}_{22}^{-1} (\tilde{\mathbf{b}}\, \tilde{\mathbf{v}}^T + \tilde{\mathbf{v}}\, \tilde{\mathbf{b}}^T) \right] C_{11}^T \tag{100}$$

In the error-covariance approximation the "measured" quantities are used in place of the true variables. Instead, the estimated values can be used. Note that Eq. (84) is equal to $\hat{D} = \tilde{U}\tilde{\tilde{S}}\tilde{V}^T C$, where $\tilde{\tilde{S}}$ is given by $\tilde{S}$ with $\tilde{s}_{n+1} = 0$. Therefore, Eq. (95) can be approximated by setting $\tilde{s}_{n+1} = 0$, which yields the following expression:

$$\omega_k^{ij} = \tilde{u}_{ik}\, \tilde{v}_{j\,n+1} / \tilde{s}_k \tag{101}$$

Define $\tilde{U}_n$ from Eq. (84) by its rows:

$$\tilde{U}_n = \begin{bmatrix} \tilde{\mathbf{u}}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \tilde{\mathbf{u}}_m^T \end{bmatrix} \tag{102}$$

Using Eqs. (101) and (102) allows Ω_i to be simply written by

$$\Omega_i = \begin{bmatrix} \tilde{\mathbf{v}} \\ \tilde{v}_{22} \end{bmatrix}^T \otimes \begin{bmatrix} \tilde{S}_n^{-1} \tilde{\mathbf{u}}_i \\ 0 \end{bmatrix} \tag{103}$$

where $\tilde{S}_n$ is defined in Eq. (84). Using Eq. (103) to compute Ω_i reduces the computational load while still producing accurate results. The error-covariance in Eq. (100) is valid for any sample size under the small noise assumption. Both Eqs. (85) and (100) require a summation of terms over m, but Eq. (100) does not require a matrix inverse of an $n \times n$ matrix to compute the error-covariance.

IV. Bearings-Only Point Estimation

Total least squares is applied to estimate the two-dimensional location of a stationary target point using passive bearing measurements. The TLS problem is formulated in Ref. 19, however only stationary errors are assumed. Here a more rigorous development is derived. The problem geometry is depicted in Figure 1. The goal is to estimate the point p from bearings-only measurements, denoted by $\tilde{\theta}_i$. The baseline points, denoted by X_i and Y_i, are assumed to be imprecisely known. The bearing measurement model and baseline point

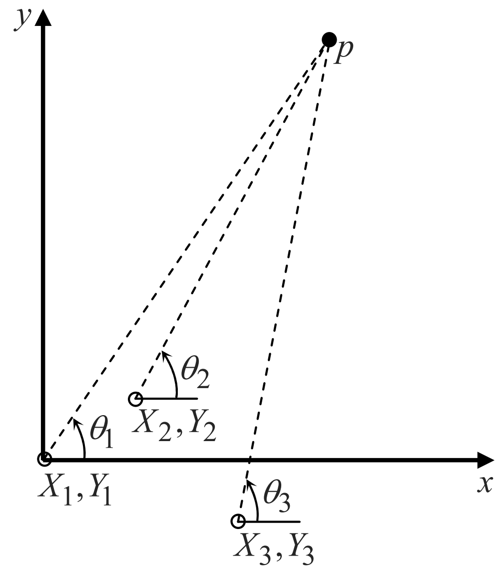


Figure 1. Two-Dimensional Bearings-Only Geometry

models are given by

$$\tilde{\theta}_i = \theta_i + \delta\theta_i \tag{104a}$$

$$\tilde{X}_i = X_i + \delta X_i \tag{104b}$$

$$\tilde{Y}_i = Y_i + \delta Y_i \tag{104c}$$

where $\delta\theta_i$, δX_i and δY_i are zero-mean Gaussian noise processes with variances $\sigma^2_{\theta_i}$, $\sigma^2_{X_i}$ and $\sigma^2_{Y_i}$, respectively. The observations are modeled as

$$\theta_i = \tan^{-1}\left(\frac{y - Y_i}{x - X_i}\right) \tag{105}$$

Taking the tangent of both sides of Eq. (105) leads to $y_i = \mathbf{h}_i^T \mathbf{x}$, with

$$y_i = -X_i \sin(\theta_i) + Y_i \cos(\theta_i) \tag{106a}$$

$$\mathbf{h}_i = [-\sin(\theta_i) \quad \cos(\theta_i)]^T \tag{106b}$$

$$\mathbf{x} = [x \;\; y]^T \tag{106c}$$

Replacing the true values with the measured values and using the first-order approximations $\sin(\theta_i + \delta\theta_i) = \sin(\theta_i) + \delta\theta_i \cos(\theta_i)$ and $\cos(\theta_i + \delta\theta_i) = \cos(\theta_i) - \delta\theta_i \sin(\theta_i)$, yields the following expressions for $\tilde{y}_i$ and $\tilde{\mathbf{h}}_i$:

$$
\begin{aligned}
\tilde{y}_i =& -\tilde{X}_i \sin(\tilde{\theta}_i) + \tilde{Y}_i \cos(\tilde{\theta}_i) \\
=& -X_i \sin(\theta_i) + Y_i \cos(\theta_i) - \delta\theta_i X_i \cos(\theta_i) - \delta X_i \sin(\theta_i) - \delta\theta_i \delta X_i \cos(\theta_i) \\
& - \delta\theta_i Y_i \sin(\theta_i) + \delta Y_i \cos(\theta_i) - \delta\theta_i \delta Y_i \sin(\theta_i)
\end{aligned} \tag{107a}
$$

$$
\begin{aligned}
\tilde{\mathbf{h}}_i &= [-\sin(\tilde{\theta}_i) \quad \cos(\tilde{\theta}_i)]^T \\
&= [-\sin(\theta_i) - \delta\theta_i \cos(\theta_i) \quad \cos(\theta_i) - \delta\theta_i \sin(\theta_i)]^T
\end{aligned} \tag{107b}
$$

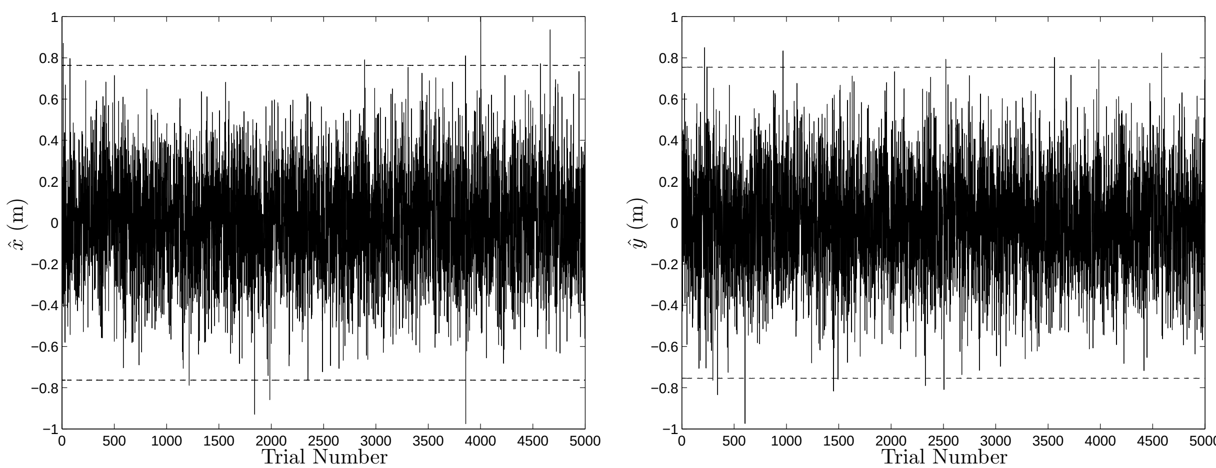
Then the elements of the covariance matrix $\mathcal{R}_i$ are computed to be

$$
\begin{aligned}
\mathcal{R}_{yy_i} =& \sigma_{\theta_i}^2 \{[X_i \cos(\theta_i) + Y_i \sin(\theta_i)]^2 + \sigma_{X_i}^2 \cos^2(\theta_i) + \sigma_{Y_i}^2 \sin^2(\theta_i)\} \\
& + \sigma_{X_i}^2 \sin^2(\theta_i) + \sigma_{Y_i}^2 \cos^2(\theta_i)
\end{aligned} \tag{108a}
$$

$$
\mathcal{R}_{hh_i} = \sigma_{\theta_i}^2 \begin{bmatrix} \cos^2(\theta_i) & \sin(\theta_i)\cos(\theta_i) \\ \sin(\theta_i)\cos(\theta_i) & \sin^2(\theta_i) \end{bmatrix} \tag{108b}
$$

$$
\mathcal{R}_{hy_i} = \sigma_{\theta_i}^2 [X_i \cos(\theta_i) + Y_i \sin(\theta_i)] \begin{bmatrix} \cos(\theta_i) \\ \sin(\theta_i) \end{bmatrix} \tag{108c}
$$

Note that the covariance matrix does not contain the true locations x and y, unlike other approaches to this problem.[20,21]



(a) **Estimate Errors for x and 3σ Bounds** (b) **Estimate Errors for y and 3σ Bounds**

Figure 2. Total Least Squares Bearings-Only Estimation Errors

In the simulation the location of the point p is given at (100, 200) meters. The baseline points are time varying with $X_i = 500\sin(0.01\, t_i)$ and $Y_i = 300\cos(0.2\, t_i)$. The variances are given by $\sigma_{\theta_i}^2 = (1\pi/180)^2$ rad^2 and $\sigma_{X_i}^2 = \sigma_{Y_i}^2 = 25$ m^2 for all i points. The final time of the simulation run is 10 seconds and measurements of $\tilde{\theta}_i$, $\tilde{X}_i$ and $\tilde{Y}_i$ are taken at 0.01 second intervals. Five thousand Monte Carlo runs are executed in order to compare the actual errors with the computed 3σ bounds using Eq. (64). The TLS initial estimate is given by using a standard linear least squares solution with the measurement variance given by Eq. (108a). Figures 2(a) and 2(b) show the errors for the TLS estimates along with their respective computed 3σ bounds. This indicates that Eq. (64) can be used to accurately compute the 3σ bounds. Also, although not shown here, the computed error-covariance using Eq. (64) matches with the sample error-covariance obtained from the Monte Carlo runs. The inequality in Eq. (69) is now checked. Figure 3 shows as a plot of the terms $\mathbf{z}^T \mathcal{R}_i\, \mathbf{z}$ and $\text{Tr}(P\, \mathcal{R}_{hh_i})$ over time. Clearly, the inequality is satisfied by several orders of magnitude. A plot of the output errors and 3σ bounds computed using Eq. (75) for one of the Monte Carlo runs is shown in Figure 4(a). Also plots of the basis function errors and 3σ bounds computed using Eq. (73) are shown in Figures 4(b) and 4(c). Clearly, the derived covariance expressions provide accurate bounds for the actual errors.

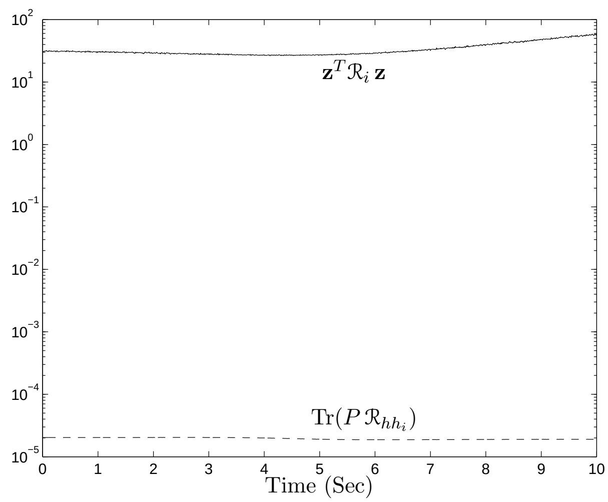
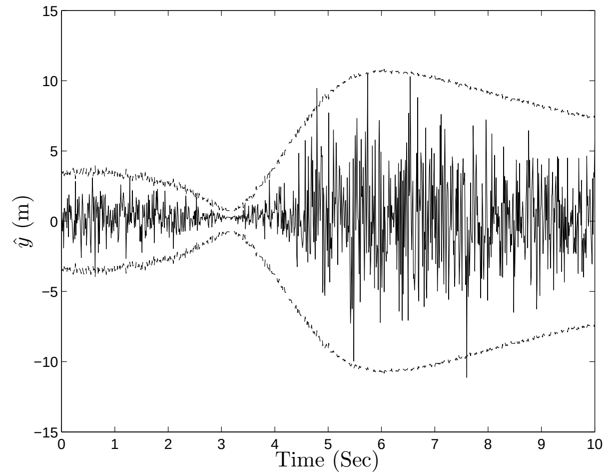
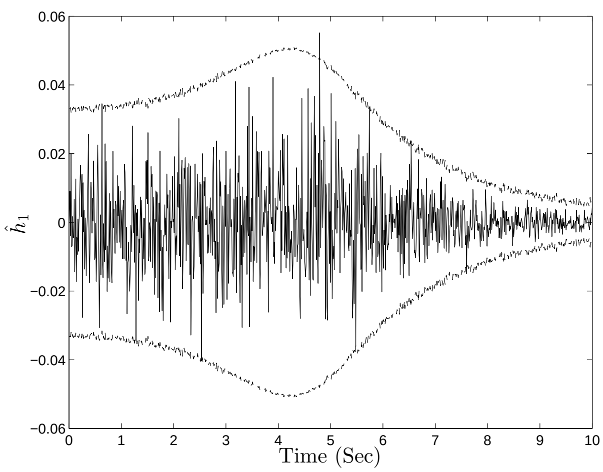


Figure 3. Inequality of Eq. (69)

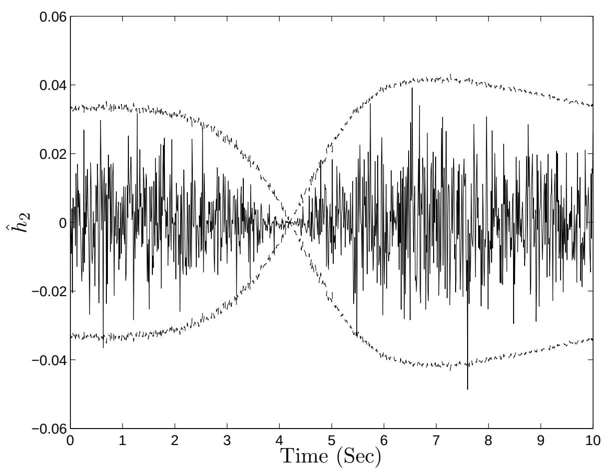
A comparison is made using the standard linear least squares (LLS) solution with its associated error-covariance. Figures 5(a) and 5(b) show the errors for the LLS estimates along with their respective computed 3σ bounds. Comparing these figures to the TLS errors in Figures 2(a) and 2(b) indicates that the LLS solution is not optimal and even biased, as discussed in Ref. 19. A summary of the Monte Carlo runs for the computed mean and three times the standard deviation (std) is shown in Table 1. The values in the parentheses show the computed 3σ bounds derived from the respective error-covariance expressions. Clearly the errors in the basis function matrix can cause significant errors if a LLS solution



(a) **Estimate Errors for y and 3σ Bounds**



(b) **Estimate Errors for h_1 and 3σ Bounds**



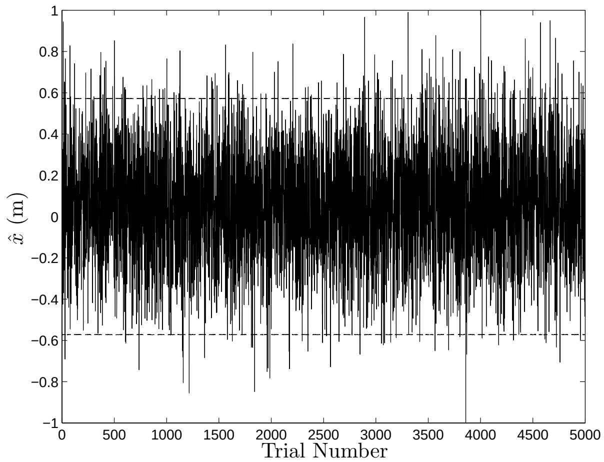
(c) **Estimate Errors for h_2 and 3σ Bounds**

Figure 4. Output and Basis Function Estimation Errors

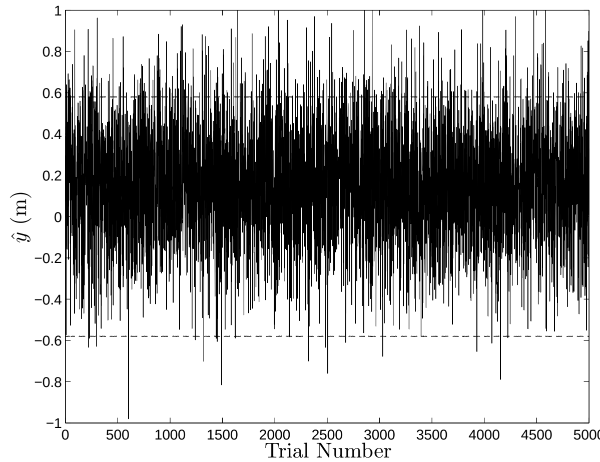
is employed over a TLS solution.

V. Conclusions

The error-covariances derived here for the total least squares problem provide useful measures to quantify the expected errors in the estimates. A perturbation analysis showed that the derived error-covariance from the associated loss function achieves the Cramér-Rao lower bound. Thus the total least squares estimator is an efficient estimator. An expression for



(a) **Estimate Errors for x and 3σ Bounds**



(b) **Estimate Errors for y and 3σ Bounds**

Figure 5. Linear Least Squares Bearings-Only Estimation Errors

Table 1. Summary of Monte Carlo Results

	TLS	LLS
mean$(\hat{x} - x)$	-0.0032	0.0561
$3 \times$ std$(\hat{x} - x)$	0.7681 (0.7636)	0.8222 (0.5720)
mean$(\hat{y} - y)$	-0.0041	0.1482
$3 \times$ std$(\hat{y} - y)$	0.7383 (0.7545)	0.8260 (0.5800)

the error-covariance for stationary errors was derived using a perturbation of the closed-form solution. This expression is useful because it does not require a matrix inverse. Simulation results using bearings-only point estimation showed that the derived error-covariance expressions provide accurate bounds for the estimate errors. Specifically, Monte Carlo runs show that the standard linear least squares solution provides biased estimates and that the computed 3σ bounds do not bound the actual errors, unlike the total least squares solution.

VI. Acknowledgments

The authors wish to thank Dr. Puneet Singla from the University at Buffalo for many helpful comments and suggestions.

References

[1] Jakubeka, S., Hametnerb, C., and Keuth, N., "Total Least Squares in Fuzzy System Identification: An Application to an Industrial Engine," *Engineering Applications of Artificial Intelligence*, Vol. 21, No. 8, Dec. 2008, pp. 1277–1288.

[2] Pruessner, A. and O'Leary, D. P., "Blind Deconvolution Using a Regularized Structured Total Least Norm Algorithm," *SIAM Journal on Matrix Analysis and Applications*, Vol. 24, No. 4, 2003, pp. 1018–1037.

[3] Nievergelt, Y., "A Tutorial History of Least Squares with Applications to Astronomy and Geodesy," *Journal of Computational and Applied Mathematics*, Vol. 121, No. 1, Sept. 2000, pp. 37–72.

[4] Markovsky, I. and Van Huffel, S., "Overview of Total Least-Squares Methods," *Signal Processing*, Vol. 87, No. 10, Oct. 2007, pp. 2283–2302.

[5] Gleser, L. J., "Estimation in a Multivariate Errors-in-Variables Regression Model: Large Sample Results," *Annals of Statistics*, Vol. 9, No. 1, Jan. 1981, pp. 24–44.

[6] Van Huffel, S. and Vandewalle, J., "On the Accuracy of Total Least Squares and Least Squares Techniques in the Presence of Errors on All Data," *Automatica*, Vol. 25, No. 5, Sept. 1989, pp. 765–769.

[7] Stoica, P. and Viberg, M., "Weighted LS and TLS Approaches Yield Asymptotically Equivalent Results," *Signal Processing*, Vol. 45, No. 2, Aug. 1995, pp. 255–259.

[8] Crassidis, J. L. and Junkins, J. L., *Optimal Estimation of Dynamic Systems*, chap. 1, 2 and 3, Chapman & Hall/CRC, Boca Raton, FL, 2nd ed., 2013.

[9] Schuermans, M., Markovsky, I., Wentzell, P. D., and Van Huffel, S., "On the Equivalence Between Total Least Squares and Maximum Likelihood PCA," *Analytica Chimica Acta*, Vol. 544, No. 1-2, 2005, pp. 254–267.

[10] Nestares, O., Fleety, D. J., and Heeger, D. J., "Likelihood Functions and Confidence Bounds for Total-Least-Squares Problems," *Proceedings of the 2000 IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, IEEE Publications, Piscataway, NJ, 2000, pp. 523–530.

[11] Van Huffel, S. and Vandewalle, J., *The Total Least Squares Problem: Computational Aspects and Analysis*, chap. 8, Society for Industial and Applied Mathematics, Philadelphia, PA, 1991.

[12] Cramér, H., *Mathematical Methods of Statistics*, chap. 32-34, Princeton University Press, Princeton, NJ, 1946.

[13] Fisher, R. A., "Theory of Statistical Estimation," *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 22, July 1925, pp. 700–725.

[14] Berry, D. A. and Lingren, B. W., *Statistics, Theory and Methods*, chap. 9, Brooks/Cole Publishing Company, Pacific Grove, CA, 1990.

[15] Markovsky, I., Rastellob, M. L., Premolic, A., Kukush, A., and Van Huffel, S., "The Element-Wise Weighted Total Least-Squares Problem," *Computational Statistics & Data Analysis*, Vol. 50, No. 1, Jan. 2006, pp. 181–209.

[16] Markovsky, I., Schuermans, M., and Van Huffel, S., "An Adapted Version of the Element-Wise Weighted Total Least Squares Method for Applications in Chemometrics," *Chemometrics and Intelligent Laboratory Systems*, Vol. 85, No. 1, Jan. 2007, pp. 40–46.

[17] Crassidis, J. L. and Cheng, Y., "Error-Covariance Analysis of the Total Least Squares Problem," *AIAA Guidance, Navigation, and Control Conference*, Portland, OR, Aug. 2011, AIAA-2011-6620.

[18]Papadopoulo, T. and Lourakis, M. I. A., "Estimating the Jacobian of the Singular Value Decomposition: Theory and Applications," *Lecture Notes in Computer Science, Computer Vision - ECCV*, Vol. 1842, Springer Verlag, Berlin / Heidelberg, 2000, pp. 554–570.

[19]Doğançay, K., "Bearings-Only Target Localization Using Total Least Squares," *Signal Processing*, Vol. 85, No. 9, Sept. 2005, pp. 1695–1710.

[20]Stansfield, R. G., "Statistical Theory of D.F. Fixing," *Journal of the Institution of Electrical Engineers – Part IIIA: Radiocommunication*, Vol. 94, No. 15, March-April 1947, pp. 762–770.

[21]Gavish, M. and Miss, A. J., "Performance Analysis of Bearing-Only Target Location Algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 28, No. 3, July 1992, pp. 817–828.