JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 42, No. 6, June 2019

Maximum Likelihood Analysis of the Total Least Squares Problem with Correlated Errors

John L. Crassidis*
University at Buffalo, State University of New York, Amherst, New York 14260-4400
and
Yang Cheng†
Mississippi State University, Mississippi State, Mississippi 39762

DOI: 10.2514/1.G003815

This paper performs a maximum likelihood analysis of the total least squares problem with Gaussian noise errors and correlated elementwise components in the design matrix. This analysis also includes a derivation of the Fisher information matrix and the error-covariance for the parameter estimates. Furthermore, the error-covariances of the associated coefficient and output estimates are also derived. These error-covariances can yield a much improved covariance approximation than would be achieved using naïve least squares. The results are compared with previously derived results for the uncorrelated elementwise component with nonequal row variance case. Simulation results using three-dimensional bearings-only localization are shown to quantify the theoretical derivations, which show that the derived error-covariances are more consistent than those given by the naïve least squares solution.

I. Introduction

THE total least squares (TLS) problem expands upon the standard least squares problem by incorporating noise not only in the measurements but also in the basis functions themselves. Because noise exists in the design matrix, 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 TLS problem without constraints can basically be broken down into four categories: 1) unweighted or scalar weighting, 2) elementwise uncorrelated with equal variances for the row vectors of the augmented matrix, 3) elementwise uncorrelated with nonequal variances for the row vectors of the augmented matrix, and 4) elementwise correlated noise processes. Here the "elementwise" terminology refers to the rows of the augmented matrix formed using the noisy design matrix and measurement vector.

Problems involving the first category correspond to the case where the covariance matrix of the "vectorized" augmented matrix is an isotropic matrix, that is, a scalar times the identity matrix. A closed-form solution for this case is given by using a singular value decomposition (SVD) approach [1]. Problems involving the second category correspond to the case where the covariance matrix of the vectorized augmented matrix is a block diagonal matrix with each block submatrix being equal to each other. A closed-form solution for this case is also given by using an SVD approach via re-scaling the data [2], which is a slight modification of the solution for problems involving the first category. The third category also involves a block diagonal covariance matrix, but each block submatrix is not restricted to be equal with another. The fourth category is the most general case where the covariance matrix is possibly fully populated. Problems involving the last two categories do not have closed-form solutions, and thus must rely on iterative algorithms.

It is well-known that several TLS estimators have infinite moments [3]. That is, many TLS estimators have no expectation, and therefore the covariance matrix of these TLS estimators is not well defined. The only reason that the estimators have infinite moments is that the values of the estimates can become infinite with positive probability, which occurs when the estimators are singular. For example, the SVD solution, used for the isotropic error case, can become singular when the denominator, the last component of the last singular vector, is zero. In [3] a simple line fitting example is used to show the singularity issue. This work further shows that, in all cases of practical interest, the TLS estimate of the slope coefficient, which has infinite moments, is more accurate than the least squares estimate of the slope coefficient, which has finite moments [4]. Series expansions are commonly used to determine moments. In [3] it is argued that all moments derived from series expansions should be regarded as moments of some approximations, rather than "approximate moments." The analysis performed in the present paper implicitly makes this assumption. From a practical point of view it can be shown that constructing approximate distributions of TLS estimators provides virtually identical results as found from assuming that the singular conditions do not exist, which arise from highly ill-posed cases [4]. These cases involve unrealistic signal-to-noise ratios and/or extremely low probabilities of obtaining a singular solution, which are beyond the scope of this paper.

The error-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; that is, its state error-covariance achieves the Cramér-Rao lower bound (CRLB) [5]. The equivalence of the TLS to maximum likelihood estimation is shown in [2,6]. A CRLB is derived in [6]. 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 [6] is not the likelihood function because of the use of the prior distribution on the nuisance parameters. A more correct form is presented in [7], but this is limited to cases where the covariance of the noise on the coefficient matrix is isotropic. The classic text of [8] presents first-order error-covariance approximations for the TLS estimate. A detailed discussion on how the TLS estimate error-covariance is related to simple least squares error-covariance is also presented. However, like [7], the error-covariance is only valid for isotropic noise errors for the coefficient matrix and measurement vector. Reference [9] derives an ML solution based on a certain structure for the covariance matrix, and shows some interesting cases. Reference [10] derives the CRLB for the elementwise uncorrelated case with nonequal variances for the row vectors. Also, a closed-form expression for the estimate error-covariance for the elementwise uncorrelated equal-variance case is derived using a perturbation approach of the SVD. Furthermore, the error-covariances of the estimated design matrix and output, as well as

Received 9 May 2018; revision received 5 March 2019; accepted for publication 6 March 2019; published online 22 April 2019.

*Samuel P. Capen Chair Professor, Department of Mechanical & Aerospace Engineering; johnc@buffalo.edu. Fellow AIAA.

†Associate Professor, Department of Aerospace Engineering; cheng@ae.msstate.edu. Associate Fellow AIAA.

their cross-covariance terms, for this case are derived. This provides useful results to quantify the expected errors for real-world applications where the truth is not known.

This paper extends the work of Ref. [10] for the most general unconstrained TLS problem: the fully correlated case. The CRLB, given by the inverse of Fisher information matrix, is derived for the parameter estimate vector. This derived result is rigorous in the sense that it is derived using a complete ML approach. Also, the error-covariance of the vectorized form of the augmented matrix formed using the estimated design matrix and output vector is derived using a perturbation approach. It is shown that the elementwise uncorrelated with nonequal row variance case of Ref. [10] is a special case of the derived error-covariance here. All other cases can easily be shown to be a special case of the results presented in this paper. Furthermore, the error-covariances derived in the paper may yield statistical bounds that are more consistent with the actual errors than using results obtained using naïve least squares.

It is also well-known that the estimates from TLS are biased [11]. Several approaches have been developed to "de-bias" the estimate [12]. However, this de-biasing increases the variance of the TLS estimate errors. Therefore, the classic trade-off exists between mitigating the bias effects or providing a less-biased estimate that would have larger variance errors than the biased solution. This is equivalent to the trade-off found in ridge estimation [5]. A modified CRLB can be employed for biased estimates [13]. However, it is assumed here that the bias is small in relation to the magnitude of the estimate. So, the bias is ignored in the present work. Still, the bias is derived here in order for one to make informed decisions of the validity of using the derived error-covariance here for the particular problem at hand. The practical example shown in this paper indicates that the assumption of ignoring the bias is justified, which is also found in many other practical problems such as the one shown in Ref. [12], involving estimating an object point photographed by three terrestrial cameras.

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 the fully correlated TLS problem to maximum likelihood estimation is shown, including a derivation of the Fisher information matrix for the parameter vector. Next, the error-covariance for augmented matrix formed using the estimated coefficient matrix and measurement vector is derived. This is followed by a proof that the derivations shown in this paper reduce to previously derived ones for the elementwise uncorrelated with nonequal row variance case. Finally, simulation results using three-dimensional bearings-only localization are shown to validate the derived CRLB.

II. Linear Least Squares Review

Consider the following linear model: $\tilde{\boldsymbol{y}} = H\boldsymbol{x} + \Delta\boldsymbol{y}$, where H is an $m \times n$ matrix, known as the *design matrix*, which contains no errors and $\Delta\boldsymbol{y}$ is an $m \times 1$ vector, which is a zero-mean Gaussian white-noise process with covariance $\bar{R}_{yy}$. The goal of the least squares problem is to determine an estimate for the $n \times 1$ vector $\boldsymbol{x}$, with $n \le m$. The covariance of $\tilde{\boldsymbol{y}}$ is given by $\text{cov}\{\tilde{\boldsymbol{y}}\} \triangleq E\{(\tilde{\boldsymbol{y}} - \boldsymbol{\mu})(\tilde{\boldsymbol{y}} - \boldsymbol{\mu})^T\} = \bar{R}_{yy}$, where $E\{\}$ denotes expectation, and $\boldsymbol{\mu}$ denotes the mean given by $H\boldsymbol{x}$. The conditional density function of $\tilde{\boldsymbol{y}}$ given $\boldsymbol{x}$ is [5]

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

In the ML approach an estimate of $\boldsymbol{x}$, denoted by $\hat{\boldsymbol{x}}$, is sought that maximizes Eq. (1). Because of the monotonic aspect of the function, the ML solution can be accomplished by also taking the natural logarithm of Eq. (1). The solution for this ML problem leads directly to the classical least squares solution for the estimate: $\hat{\boldsymbol{x}}(\tilde{\boldsymbol{y}}) \equiv \hat{\boldsymbol{x}} = (H^T \bar{R}_{yy}^{-1} H)^{-1} H^T \bar{R}_{yy}^{-1} \tilde{\boldsymbol{y}}$. The mean of $\hat{\boldsymbol{x}}$ is given by $\boldsymbol{x}$, which means that the estimator is unbiased. The error-covariance of $\hat{\boldsymbol{x}}$ is given by $\text{cov}\{\hat{\boldsymbol{x}}\} = (H^T \bar{R}_{yy}^{-1} H)^{-1}$, which can be used to develop 3σ bounds on the expected estimate errors.

The Cramér–Rao inequality [14] 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 Cramér–Rao inequality for an unbiased estimate $\hat{\boldsymbol{x}}$ is given by

$$P \triangleq E\{(\hat{\boldsymbol{x}} - \boldsymbol{x})(\hat{\boldsymbol{x}} - \boldsymbol{x})^T\} \ge F^{-1} \tag{2}$$

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

$$\begin{aligned} F &= E\left\{\left(\frac{\partial}{\partial \boldsymbol{x}} \ln[p(\tilde{\boldsymbol{y}}|\boldsymbol{x})]\right)\left(\frac{\partial}{\partial \boldsymbol{x}} \ln[p(\tilde{\boldsymbol{y}}|\boldsymbol{x})]\right)^T\right\} \\ &= -E\left\{\frac{\partial^2}{\partial \boldsymbol{x} \partial \boldsymbol{x}^T} \ln[p(\tilde{\boldsymbol{y}}|\boldsymbol{x})]\right\} \end{aligned} \tag{3}$$

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 [15] for details). The least squares estimate achieves the CRLB and is thus an efficient estimator [5].

III. Total Least Squares

For the general problem, the TLS model is given by

$$\tilde{\boldsymbol{y}} = \boldsymbol{y} + \Delta\boldsymbol{y} \tag{4a}$$

$$\tilde{H} = H + \Delta H \tag{4b}$$

$$\boldsymbol{y} = H\boldsymbol{x} \tag{4c}$$

where $\tilde{\boldsymbol{y}}$ is an $m \times 1$ measurement vector, $\boldsymbol{y}$ is its respective true value, $\Delta\boldsymbol{y}$ 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 $m \times (n+1)$ matrices:

$$\tilde{D} \triangleq [\tilde{H} \;\; \tilde{\boldsymbol{y}}], \quad D \triangleq [H \;\; \boldsymbol{y}], \quad \Delta D \triangleq [\Delta H \;\; \Delta\boldsymbol{y}], \quad \hat{D} \triangleq [\hat{H} \;\; \hat{\boldsymbol{y}}] \tag{5}$$

where $\hat{H}$ and $\hat{\boldsymbol{y}}$ are the estimates of H and $\boldsymbol{y}$, respectively. The TLS problem seeks the maximum likelihood estimate of the $n \times 1$ vector $\boldsymbol{x}$, denoted by $\hat{\boldsymbol{x}}$, as well as the maximum likelihood estimates of $\boldsymbol{y}$ and H, satisfying $\hat{\boldsymbol{y}} = \hat{H}\hat{\boldsymbol{x}}$ and which maximizes

$$p(\tilde{D}|D) = \frac{1}{(2\pi)^{m(n+1)/2}[\det(R)]^{1/2}} \exp\left\{-\frac{1}{2}\text{vec}^T(\tilde{D}^T - D^T)R^{-1} \times \text{vec}(\tilde{D}^T - D^T)\right\} \tag{6}$$

where $D\boldsymbol{z} = \boldsymbol{0}$ with $\boldsymbol{z} \triangleq [\boldsymbol{x}^T \;\; -1]^T$, vec denotes a column vector formed by stacking the consecutive columns of the associated matrix, and the R is the covariance matrix defined by

$$R \triangleq E\{\text{vec}(\tilde{D}^T - D^T)\text{vec}^T(\tilde{D}^T - D^T)\} \tag{7}$$

The negative log-likelihood now leads to the following loss function:

$$J(\hat{D}, \hat{\boldsymbol{x}}) = \frac{1}{2}\text{vec}^T(\tilde{D}^T - \hat{D}^T)R^{-1}\text{vec}(\tilde{D}^T - \hat{D}^T), \quad \text{s.t.} \;\; \hat{D}\,\hat{\boldsymbol{z}} = \boldsymbol{0} \tag{8}$$

where $\hat{\boldsymbol{z}} \triangleq [\hat{\boldsymbol{x}}^T \;\; -1]^T$. For the estimates $\hat{D}$ and $\hat{\boldsymbol{z}}$ to be unique, it is required that the rank of $\hat{D}$ (and the rank of $\hat{H}$) be n, which means $\hat{\boldsymbol{z}}$ spans the null space of $\hat{D}$.

Equation (8) represents a constrained optimization problem to determine $\hat{D}$ and $\hat{\boldsymbol{x}}$, which can lead to a computational expensive solution. An easier problem involves rewriting the loss function as a function of $\hat{\boldsymbol{x}}$. A solution for the minimization of Eq. (8) is given by using the method of Lagrange multipliers [16], leading to the following Lagrangian:

$$J'(\hat{D},\hat{\boldsymbol{x}}) = \frac{1}{2}\text{vec}^T(\tilde{D}^T - \hat{D}^T)R^{-1}\text{vec}(\tilde{D}^T - \hat{D}^T) + \frac{1}{2}\boldsymbol{\lambda}^T\hat{D}\,\hat{z} \tag{9}$$

where $\boldsymbol{\lambda}$ is an $m \times 1$ Lagrange multiplier vector. Taking the partial derivatives of J' with respect to $\hat{\boldsymbol{x}}$ and $\text{vec}(\hat{D}^T)$, and setting their resultants to zero to obtain the necessary condition for a minimum, gives

$$\frac{\partial J'}{\partial \hat{\boldsymbol{x}}} = \hat{H}^T\boldsymbol{\lambda} = \mathbf{0} \tag{10a}$$

$$\frac{\partial J'}{\partial \text{vec}(\hat{D}^T)} = R^{-1}\text{vec}(\tilde{D}^T - \hat{D}^T) + (I_{m\times m} \otimes \hat{z})\boldsymbol{\lambda} = \mathbf{0} \tag{10b}$$

where $I_{m\times m}$ is an $m \times m$ identity matrix and $\otimes$ is the standard Kronecker product operator. Equation (10a) shows that the Lagrange multiplier must be in the null space of $\hat{H}^T$. Although this is not particularly useful in theory, it may be useful to ensure that a numerical solution has converged to a correct value. Solving Eq. (10b) for $\text{vec}(\hat{D}^T)$, and substituting the resultant into $\hat{D}\,\hat{z} = (I_{m\times m} \otimes \hat{z}^T)\text{vec}(\hat{D}^T)$, with $\hat{D}\,\hat{z} = \mathbf{0}$, yields

$$\boldsymbol{\lambda} = -Q_{\hat{\lambda}}^{-1}(I_{m\times m} \otimes \hat{z}^T)\text{vec}(\tilde{D}^T) \tag{11}$$

where

$$Q_{\hat{\lambda}} \triangleq (I_{m\times m} \otimes \hat{z}^T)R(I_{m\times m} \otimes \hat{z}) \tag{12}$$

Substituting Eq. (11) into Eq. (10b) leads to

$$\text{vec}(\tilde{D}^T - \hat{D}^T) = R(I_{m\times m} \otimes \hat{z})Q_{\hat{\lambda}}^{-1}(I_{m\times m} \otimes \hat{z}^T)\text{vec}(\tilde{D}^T) \tag{13}$$

Substituting Eq. (13) into Eq. (8) gives

$$J(\hat{\boldsymbol{x}}) = \frac{1}{2}\text{vec}^T(\tilde{D}^T)(I_{m\times m} \otimes \hat{z})Q_{\hat{\lambda}}^{-1}(I_{m\times m} \otimes \hat{z}^T)\text{vec}(\tilde{D}^T) \tag{14}$$

which is now only a function of $\hat{\boldsymbol{x}}$. From Eq. (13), $\text{vec}(\hat{D}^T)$ is given by

$$\text{vec}(\hat{D}^T) = [I_{m(n+1)\times m(n+1)} - R(I_{m\times m} \otimes \hat{z})Q_{\hat{\lambda}}^{-1}(I_{m\times m} \otimes \hat{z}^T)]\text{vec}(\tilde{D}^T) \tag{15}$$

Another way to write the probability density function and the loss function is to use $\text{vec}(\tilde{D} - \hat{D})$ instead of $\text{vec}(\tilde{D}^T - \hat{D}^T)$. Define the following matrix:

$$\bar{R} \triangleq \begin{array}{c} \left[\begin{array}{cc} \bar{R}_{hh} & \bar{R}_{hy} \\ \underbrace{\bar{R}_{hy}^T}_{mn} & \underbrace{\bar{R}_{yy}}_{m} \end{array}\right] \end{array} \begin{array}{l} \}mn \\ \}m \end{array} \tag{16}$$

where $\bar{R}_{hh} = E\{\text{vec}(\Delta H)\text{vec}(\Delta H)^T\}$ is the $(mn)\times(mn)$ covariance matrix of $\text{vec}(\Delta H)$, $\bar{R}_{yy} = E\{\text{vec}(\Delta\boldsymbol{y})\text{vec}(\Delta\boldsymbol{y})^T\}$ is the $m\times m$ covariance matrix of $\text{vec}(\Delta\boldsymbol{y})$, and $\bar{R}_{hy} = E\{\text{vec}(\Delta H)\text{vec}(\Delta\boldsymbol{y})^T\}$ is the $(mn)\times m$ cross-correlation matrix. The covariance matrix given in Eq. (16) is useful for the correlated TLS problem because it nicely partitions the covariance elements of ΔH and $\Delta\boldsymbol{y}$. Also, $\text{vec}(D)$ is nicely given by

$$\text{vec}(D) = \begin{bmatrix} \text{vec}(H) \\ \boldsymbol{y} \end{bmatrix} \tag{17}$$

The matrix R and vector $\text{vec}(\tilde{D}^T)$ are related to $\bar{R}$ and $\text{vec}(\tilde{D})$ through

$$R = K_{n+1,m}\bar{R}K_{n+1,m}^T, \qquad \text{vec}(\tilde{D}^T) = K_{n+1,m}\text{vec}(\tilde{D}) \tag{18}$$

where $K_{q,p}$, for any positive scalars p and q, is the commutation matrix [17], also known as the permutation matrix. So, the probability density function can be written as

$$p(\tilde{D}|D) = \frac{1}{(2\pi)^{m(n+1)/2}[\det(\bar{R})]^{1/2}}\exp\left\{-\frac{1}{2}\text{vec}^T(\tilde{D} - D)\bar{R}^{-1} \times \text{vec}(\tilde{D} - D)\right\} \tag{19}$$

where $\det(R) = \det(\bar{R})$ has been used.

Using the identity

$$(I_{m\times m} \otimes \hat{z}^T)\text{vec}(\tilde{D}^T) = (\hat{z}^T \otimes I_{m\times m})\text{vec}(\tilde{D}) \tag{20}$$

in Eq. (14) yields

$$J(\hat{\boldsymbol{x}}) = \frac{1}{2}\text{vec}^T(\tilde{D})(\hat{z} \otimes I_{m\times m})Q_{\hat{\lambda}}^{-1}(\hat{z}^T \otimes I_{m\times m})\text{vec}(\tilde{D}) \tag{21}$$

Also, substituting Eq. (18) into Eq. (12) leads to

$$Q_{\hat{\lambda}} = (\hat{z}^T \otimes I_{m\times m})\bar{R}(\hat{z} \otimes I_{m\times m}) \tag{22}$$

where $\hat{z}^T \otimes I_{m\times m}$ can now be partitioned as $\hat{z}^T \otimes I_{m\times m} = [\hat{\boldsymbol{x}}^T \otimes I_{m\times m} - I_{m\times m}]$. Several iterative algorithms to minimize Eq. (21) are shown in [16].

Substituting $\text{vec}(\hat{D}^T) = K_{n+1,m}\text{vec}(\hat{D})$, $\text{vec}(\tilde{D}^T) = K_{n+1,m}\text{vec}(\tilde{D})$, and Eq. (18) into Eq. (15) leads to

$$\text{vec}(\hat{D}) = [I_{m(n+1)\times m(n+1)} - \bar{R}(\hat{z} \otimes I_{m\times m})Q_{\hat{\lambda}}^{-1}(\hat{z}^T \otimes I_{m\times m})]\text{vec}(\tilde{D}) \tag{23}$$

Note that in practice Eq. (12) would be used for Eq. (15), because it is a function of R, whereas Eq. (22) would be used for Eq. (23), because it is a function of $\bar{R}$.

The loss function in Eq. (8) can be viewed as one of $\text{vec}(H)$ and $\boldsymbol{x}$ as well. The corresponding TLS model is given by

$$\text{vec}(\tilde{H}) = \text{vec}(H) + \text{vec}(\Delta H) \tag{24a}$$

$$\tilde{\boldsymbol{y}} = (\boldsymbol{x}^T \otimes I_{m\times m})\text{vec}(H) + \Delta\boldsymbol{y} \tag{24b}$$

Solving the necessary condition for the estimate of $\text{vec}(H)$ gives

$$\text{vec}(\hat{H}) = \left(\begin{bmatrix} I_{(mn)\times(mn)} \\ \hat{\boldsymbol{x}}^T \otimes I_{m\times m} \end{bmatrix}^T \bar{R}^{-1} \begin{bmatrix} I_{(mn)\times(mn)} \\ \hat{\boldsymbol{x}}^T \otimes I_{m\times m} \end{bmatrix}\right)^{-1} \times \begin{bmatrix} I_{(mn)\times(mn)} \\ \hat{\boldsymbol{x}}^T \otimes I_{m\times m} \end{bmatrix}^T \bar{R}^{-1}\text{vec}(\tilde{D}) \tag{25}$$

Note that the estimate $\hat{\boldsymbol{y}}$ is computed as $\hat{\boldsymbol{y}} = (\hat{\boldsymbol{x}}^T \otimes I_{m\times m})\text{vec}(\hat{H}) = \hat{H}\hat{\boldsymbol{x}}$. Substituting this solution of $\text{vec}(\hat{H})$ to the loss function leads to another equivalent form of $J(\hat{\boldsymbol{x}})$:

$$J(\hat{\boldsymbol{x}}) = \frac{1}{2}\text{vec}^T(\tilde{D})\bar{F}\text{vec}(\tilde{D}) \tag{26}$$

with

$$\bar{F}=\bar{R}^{-1}-\bar{R}^{-1}\begin{bmatrix} I_{(mn)\times(mn)} \\ \hat{\boldsymbol{x}}^T\otimes I_{m\times m}\end{bmatrix}\left(\begin{bmatrix} I_{(mn)\times(mn)} \\ \hat{\boldsymbol{x}}^T\otimes I_{m\times m}\end{bmatrix}^T\bar{R}^{-1}\begin{bmatrix} I_{(mn)\times(mn)} \\ \hat{\boldsymbol{x}}^T\otimes I_{m\times m}\end{bmatrix}\right)^{-1} \times\begin{bmatrix} I_{(mn)\times(mn)} \\ \hat{\boldsymbol{x}}^T\otimes I_{m\times m}\end{bmatrix}^T\bar{R}^{-1} \tag{27}$$

IV. Fisher Information Matrix

Because the ML estimator is asymptotically efficient, the estimate error-covariance matrix can be approximated by the inverse of the FIM when a sufficient number of measurements are included. The Cramér–Rao inequality is only valid for unbiased estimates. It is well-known that the TLS estimate is biased though [18]. A derivation of the TLS bias is shown in Appendix A. It shows that for all practical noise cases, the bias is second-order in nature. Thus, it is assumed here that this bias is neglected in the FIM derivation. To derive the FIM for the TLS estimate $\hat{\boldsymbol{x}}$, it is possible to determine the FIM for the TLS estimate $\hat{D}$ from the likelihood function given by Eq. (6) and then retrieve the FIM for $\hat{\boldsymbol{x}}$ from it. It is difficult, however, to derive the FIM for $\hat{D}$ because of the constraint $D\boldsymbol{z}=\boldsymbol{0}$, which explicitly involves $\boldsymbol{x}$. The FIM for the joint TLS estimate of $\{\boldsymbol{x},H\}$ will be derived instead.

Recall that the associated likelihood function using $\mathrm{vec}(\tilde{D})$ as the measurement is given by

$$p(\tilde{D}|D)=\frac{1}{(2\pi)^{m(n+1)/2}[\det(\bar{R})]^{1/2}}\exp\left\{-\frac{1}{2}\mathrm{vec}^T(\tilde{D}-D)\bar{R}^{-1}\times\mathrm{vec}(\tilde{D}-D)\right\} \tag{28}$$

The likelihood function in Eq. (28) is now treated as a function of $\{\boldsymbol{x},H\}$:

$$p(\tilde{D}|\boldsymbol{x},H)=\frac{1}{(2\pi)^{m(n+1)/2}[\det(\bar{R})]^{1/2}}\exp\left\{-\frac{1}{2}\mathrm{vec}^T[\tilde{D}-D(\boldsymbol{x},H)]\bar{R}^{-1}\times\mathrm{vec}[\tilde{D}-D(\boldsymbol{x},H)]\right\} \tag{29}$$

with $D(\boldsymbol{x},H)\triangleq[H\quad H\boldsymbol{x}]$. Now, the FIM of the likelihood function $p(\tilde{D}|\boldsymbol{x},H)$ is derived. Define

$$\boldsymbol{a}\triangleq\begin{bmatrix}\boldsymbol{x}\\ \mathrm{vec}(H)\end{bmatrix},\quad p(\tilde{D}|\boldsymbol{a})\triangleq p(\tilde{D}|\boldsymbol{x},H),\quad D(\boldsymbol{a})\triangleq D(\boldsymbol{x},H) \tag{30}$$

The FIM, F^a, for $\boldsymbol{a}$ is

$$F^a=E\left\{\left(\frac{\partial}{\partial\boldsymbol{a}}\ln[p(\tilde{D}|\boldsymbol{a})]\right)\left(\frac{\partial}{\partial\boldsymbol{a}}\ln[p(\tilde{D}|\boldsymbol{a})]\right)^T\right\} \tag{31}$$

Taking the partial derivative of the natural logarithm of $p(\tilde{D}|\boldsymbol{a})$ leads to

$$\frac{\partial}{\partial\boldsymbol{a}}\ln[p(\tilde{D}|\boldsymbol{a})]=\begin{bmatrix}0_{n\times(mn)} & H^T\\ I_{(mn)\times(mn)} & (\boldsymbol{x}\otimes I_{m\times m})\end{bmatrix}\bar{R}^{-1}\mathrm{vec}[\tilde{D}-D(\boldsymbol{a})] \tag{32}$$

where $0_{n\times(mn)}$ denotes an $n\times(mn)$ matrix of zeros. Because $E\{\mathrm{vec}[\tilde{D}-D(\boldsymbol{a})]\}=\boldsymbol{0}$, then $E\{(\partial/\partial\boldsymbol{a})\ln[p(\tilde{D}|\boldsymbol{a})]\}=\boldsymbol{0}$. This means that the regularity condition

$$E\left\{\frac{\partial}{\partial\boldsymbol{a}}\ln[p(\tilde{D}|\boldsymbol{a})]\right\}\triangleq\int\frac{\partial}{\partial\boldsymbol{a}}\ln[p(\tilde{D}|\boldsymbol{a})]p(\tilde{D}|\boldsymbol{a})\,\mathrm{dvec}(\tilde{D}) =\int\left[\frac{\partial p(\tilde{D}|\boldsymbol{a})}{\partial\boldsymbol{a}}\right]\mathrm{dvec}(\tilde{D})=\boldsymbol{0} \tag{33}$$

is satisfied, which is prerequisite for the derivation of the CRLB. Postmultiplying the term $\partial\ln[p(\tilde{D}|\boldsymbol{a})]/\partial\boldsymbol{a}$ in Eq. (32) by its transpose, taking the expectation and using $E\{\mathrm{vec}[\tilde{D}-D(\boldsymbol{a})]\mathrm{vec}^T[\tilde{D}-D(\boldsymbol{a})]\}=\bar{R}$, leads to

$$F^a=E\left\{\left(\frac{\partial}{\partial\boldsymbol{a}}\ln[p(\tilde{D}|\boldsymbol{a})]\right)\left(\frac{\partial}{\partial\boldsymbol{a}}\ln[p(\tilde{D}|\boldsymbol{a})]\right)^T\right\} =\begin{bmatrix}0_{n\times(mn)} & H^T\\ I_{(mn)\times(mn)} & (\boldsymbol{x}\otimes I_{m\times m})\end{bmatrix}\bar{R}^{-1}\begin{bmatrix}0_{n\times(mn)} & H^T\\ I_{(mn)\times(mn)} & (\boldsymbol{x}\otimes I_{m\times m})\end{bmatrix}^T \tag{34}$$

Note that F^a is an $(mn+n)\times(mn+n)$ matrix.

The next step is to derive the FIM for $\hat{\boldsymbol{x}}$, denoted by F. The FIM for $\boldsymbol{a}$ is partitioned as follows:

$$F^a\triangleq\begin{bmatrix}F_{xx} & F_{xh}\\ F_{xh}^T & F_{hh}\end{bmatrix} \tag{35}$$

with

$$F_{xx}=[0_{n\times(mn)}\quad H^T]\bar{R}^{-1}\begin{bmatrix}0_{(mn)\times n}\\ H\end{bmatrix} \tag{36a}$$

$$F_{xh}=[0_{n\times(mn)}\quad H^T]\bar{R}^{-1}\begin{bmatrix}I_{(mn)\times(mn)}\\ (\boldsymbol{x}^T\otimes I_{m\times m})\end{bmatrix} \tag{36b}$$

$$F_{hh}=[I_{(mn)\times(mn)}\quad(\boldsymbol{x}\otimes I_{m\times m})]\bar{R}^{-1}\begin{bmatrix}I_{(mn)\times(mn)}\\ (\boldsymbol{x}^T\otimes I_{m\times m})\end{bmatrix} \tag{36c}$$

where F_{xx} is an $n\times n$ matrix, F_{xh} is an $n\times(mn)$ matrix, and F_{hh} is an $(mn)\times(mn)$ matrix. Applying the Sherman-Morrison-Woodbury formula [5] to F^a leads to $F^{-1}=(F_{xx}-F_{xh}F_{hh}^{-1}F_{xh}^T)^{-1}$, or equivalently $F=F_{xx}-F_{xh}F_{hh}^{-1}F_{xh}^T$, which gives the FIM for $\hat{\boldsymbol{x}}$. The FIM for $\mathrm{vec}(\hat{H})$ can be obtained similarly.

Before the expression for F is simplified, it is shown that the same F can result from other parameter vectors $\boldsymbol{a}'$. Suppose

$$\boldsymbol{a}'\triangleq\begin{bmatrix}\boldsymbol{x}\\ \mathrm{vec}(H^T)\end{bmatrix} \tag{37}$$

Since $\mathrm{vec}(H)=K_{n,m}^T\mathrm{vec}(H^T)$ then

$$\boldsymbol{a}=\begin{bmatrix}I_{n\times n} & 0_{n\times(mn)}\\ 0_{(mn)\times n} & K_{n,m}^T\end{bmatrix}\boldsymbol{a}',\quad\frac{\partial\boldsymbol{a}}{\partial\boldsymbol{a}'^T}=\begin{bmatrix}I_{n\times n} & 0_{n\times(mn)}\\ 0_{(mn)\times n} & K_{n,m}^T\end{bmatrix} \tag{38}$$

Hence, the FIM for $\boldsymbol{a}'$ is

$$F^{a'}=\left(\frac{\partial\boldsymbol{a}}{\partial\boldsymbol{a}'^T}\right)^TF^a\left(\frac{\partial\boldsymbol{a}}{\partial\boldsymbol{a}'^T}\right) =\begin{bmatrix}I_{n\times n} & 0_{n\times(mn)}\\ 0_{(mn)\times n} & K_{n,m}\end{bmatrix}F^a\begin{bmatrix}I_{n\times n} & 0_{n\times(mn)}\\ 0_{(mn)\times n} & K_{n,m}^T\end{bmatrix} \tag{39}$$

Substituting Eq. (35) into this equation gives

$$F^{a'}=\begin{bmatrix}F_{xx} & F_{xh}K_{n,m}^T\\ K_{n,m}F_{xh}^T & K_{n,m}F_{hh}K_{n,m}^T\end{bmatrix} \tag{40}$$

Applying the Sherman–Morrison–Woodbury formula to $F^{a'}$ leads to the identical FIM for $\hat{\boldsymbol{x}}$:

$$F = F_{xx} - F_{xh}K_{n,m}^T(K_{n,m}F_{hh}K_{n,m}^T)^{-1}K_{n,m}F_{xh}^T = F_{xx} - F_{xh}F_{hh}^{-1}F_{xh}^T \tag{41}$$

In fact, if $\boldsymbol{a}$ is chosen as $[\boldsymbol{x}^T, \boldsymbol{g}^T(H)]^T$, where $\boldsymbol{g}(H)$ is an invertible transform of H, the FIM for $\hat{\boldsymbol{x}}$ remains the same.

Now the FIM for $\hat{\boldsymbol{x}}$ is simplified. The inverse of $\bar{R}$ can be partitioned into

$$\bar{R}^{-1} \triangleq \left[\begin{array}{cc} \bar{\mathcal{R}}_{hh} & \bar{\mathcal{R}}_{hy} \\ \underbrace{\bar{\mathcal{R}}_{hy}^T}_{mn} & \underbrace{\bar{\mathcal{R}}_{yy}}_{m} \end{array}\right] \begin{array}{l} \}mn \\ \}m \end{array} \tag{42}$$

with

$$\bar{\mathcal{R}}_{hh} = \bar{R}_{hh}^{-1}(I_{(mn)\times(mn)} + \bar{R}_{hy}Z^{-1}\bar{R}_{hy}^T\bar{R}_{hh}^{-1}) \tag{43a}$$

$$\bar{\mathcal{R}}_{hy} = -\bar{R}_{hh}^{-1}\bar{R}_{hy}Z^{-1} \tag{43b}$$

$$\bar{\mathcal{R}}_{yy} = Z^{-1} \tag{43c}$$

$$Z \triangleq \bar{R}_{yy} - \bar{R}_{hy}^T\bar{R}_{hh}^{-1}\bar{R}_{hy} \tag{43d}$$

where $\bar{R}_{hh}$, $\bar{R}_{hy}$, and $\bar{R}_{yy}$ are defined in Eq. (16). Explicitly multiplying the quantities in Eq. (36) yields

$$F_{xx} = H^T\bar{\mathcal{R}}_{yy}H \tag{44a}$$

$$F_{xh} = H^T(\bar{\mathcal{R}}_{hy}^T + \bar{\mathcal{R}}_{yy}Q_x^T) \tag{44b}$$

$$F_{hh} = \bar{\mathcal{R}}_{hh} + Q_x\bar{\mathcal{R}}_{hy}^T + \bar{\mathcal{R}}_{hy}Q_x^T + Q_x\bar{\mathcal{R}}_{yy}Q_x^T \tag{44c}$$

where $Q_x \triangleq \boldsymbol{x} \otimes I_{m\times m}$. Note that F_{xx} is the FIM for the standard least squares estimator when no errors exist on H. Computing $F = F_{xx} - F_{xh}F_{hh}^{-1}F_{xh}^T$ gives

$$F = H^T\mathcal{Q}^{-1}H \tag{45}$$

where

$$\mathcal{Q}^{-1} \triangleq \bar{\mathcal{R}}_{yy} - (\bar{\mathcal{R}}_{hy}^T + \bar{\mathcal{R}}_{yy}Q_x^T)(\bar{\mathcal{R}}_{hh} + Q_x\bar{\mathcal{R}}_{hy}^T + \bar{\mathcal{R}}_{hy}Q_x^T + Q_x\bar{\mathcal{R}}_{yy}Q_x^T)^{-1} \times (\bar{\mathcal{R}}_{hy} + Q_x\bar{\mathcal{R}}_{yy}) \tag{46}$$

The inverse of $\mathcal{Q}^{-1}$ is given by

$$\mathcal{Q} = Z + (Z\bar{\mathcal{R}}_{hy}^T + Q_x^T)\mathcal{G}^{-1}(\bar{\mathcal{R}}_{hy}Z + Q_x) \tag{47}$$

where

$$\begin{aligned}\mathcal{G} &\triangleq \bar{\mathcal{R}}_{hh} + Q_x\bar{\mathcal{R}}_{hy}^T + \bar{\mathcal{R}}_{hy}Q_x^T + Q_xZ^{-1}Q_x^T \\ &\quad - (\bar{\mathcal{R}}_{hy} + Q_xZ^{-1})Z(\bar{\mathcal{R}}_{hy}^T + Z^{-1}Q_x^T) \\ &= \bar{\mathcal{R}}_{hh} - \bar{\mathcal{R}}_{hy}Z\bar{\mathcal{R}}_{hy}^T\end{aligned} \tag{48}$$

Substituting Eq. (43) into Eq. (48) leads simply to $\mathcal{G} = \bar{R}_{hh}^{-1}$. Then substituting this equation and Eq. (43) into Eq. (47) yields

$$\mathcal{Q} = Q_x^T\bar{R}_{hh}Q_x - \bar{R}_{hy}^TQ_x - Q_x^T\bar{R}_{hy} + \bar{R}_{yy} \tag{49}$$

The matrix $Q_{\hat{\lambda}}$ in Eq. (22), evaluated at the true values, is defined by

$$\begin{aligned}Q_\lambda &\triangleq (\boldsymbol{z}^T \otimes I_{m\times m})\bar{R}(\boldsymbol{z} \otimes I_{m\times m}) \\ &= [\,Q_x^T \quad -I_{m\times m}\,]\begin{bmatrix} \bar{R}_{hh} & \bar{R}_{hy} \\ \bar{R}_{hy}^T & \bar{R}_{yy} \end{bmatrix}\begin{bmatrix} Q_x \\ -I_{m\times m} \end{bmatrix}\end{aligned} \tag{50}$$

where Eq. (16) has been used. Explicitly carrying out the matrix multiplications in Eq. (50) gives

$$Q_\lambda = Q_x^T\bar{R}_{hh}Q_x - \bar{R}_{hy}^TQ_x - Q_x^T\bar{R}_{hy} + \bar{R}_{yy} \tag{51}$$

Comparing Eqs. (49) and (51) shows that $\mathcal{Q} = Q_\lambda$. Thus the FIM for $\hat{\boldsymbol{x}}$ is given by

$$F = H^TQ_\lambda^{-1}H \tag{52}$$

where Q_λ is given by Eq. (22) evaluated at the true values:

$$Q_\lambda \triangleq (\boldsymbol{z}^T \otimes I_{m\times m})\bar{R}(\boldsymbol{z} \otimes I_{m\times m}) \tag{53}$$

The FIM equals the Hessian of a special loss function in $\boldsymbol{x}$:

$$\begin{aligned}J(\boldsymbol{x}) &= \frac{1}{2}\mathrm{vec}^T(D)(\boldsymbol{z} \otimes I_{m\times m})Q_\lambda^{-1}(\boldsymbol{z}^T \otimes I_{m\times m})\mathrm{vec}(D) \\ &= \frac{1}{2}\boldsymbol{z}^TD^TQ_\lambda^{-1}D\boldsymbol{z} \\ &= \frac{1}{2}[\boldsymbol{x}^T \quad -1]\begin{bmatrix} H^T \\ \boldsymbol{y}^T \end{bmatrix}Q_\lambda^{-1}[H \quad \boldsymbol{y}]\begin{bmatrix} \boldsymbol{x} \\ -1 \end{bmatrix}\end{aligned} \tag{54}$$

Note that this function does not contain any measurements. Computing the Hessian at the true value of $\boldsymbol{x}$ gives

$$\frac{\partial^2 J}{\partial\boldsymbol{x}\partial\boldsymbol{x}^T} = H^TQ_\lambda^{-1}H \tag{55}$$

Note that Q_λ is a function of $\boldsymbol{x}$, but its partial derivative is not required because the constraint $D\boldsymbol{z} = \boldsymbol{0}$ (or equivalently $H\boldsymbol{x} = \boldsymbol{y}$) negates this portion of the Hessian. Equation (52) is equivalent to Eq. (55).

The FIM can also be obtained using the loss function in Eq. (26) computed at true values:

$$J(\boldsymbol{x}) = \frac{1}{2}\mathrm{vec}^T(D)\bar{F}\mathrm{vec}(D) = \frac{1}{2}\begin{bmatrix} \mathrm{vec}(H) \\ H\boldsymbol{x} \end{bmatrix}^T\bar{F}\begin{bmatrix} \mathrm{vec}(H) \\ H\boldsymbol{x} \end{bmatrix} \tag{56}$$

with $\bar{F}$ given by Eq. (27). Computing the Hessian at the true values gives

$$\begin{aligned}\frac{\partial^2 J}{\partial\boldsymbol{x}\partial\boldsymbol{x}^T} &= [0_{n\times(mn)} \quad H^T]\bar{F}[0_{n\times(mn)} \quad H^T]^T \\ &= F_{xx} - F_{xh}F_{hh}^{-1}F_{xh}^T = F\end{aligned} \tag{57}$$

where the definitions of F_{xx}, F_{xh}, and F_{hh} in Eq. (36) has been used. Note that $\bar{F}$ is a function of $\boldsymbol{x}$, but its partial derivative is not required.

V. Error-Covariance Matrices

The error-covariances of $\hat{\boldsymbol{x}}$, $\mathrm{vec}(\hat{D})$, and $\mathrm{vec}(\hat{D}^T)$ are derived using a first-order Taylor-series expansion approach. The error-covariance of $\mathrm{vec}(\hat{H})$ is also shown to be equivalent to the inverse of the corresponding FIM.

A. Error-Covariance of $\hat{\boldsymbol{x}}$

Applying this first-order expansion to Eq. (23) gives

$$\mathrm{vec}(\hat{D}) = \mathrm{vec}(D) + \bar{G}\frac{\partial \mathrm{vec}(\tilde{D})}{\partial \mathrm{vec}(\tilde{D})}\bigg|_{D} \mathrm{vec}(\Delta D) - \bar{R}\frac{\partial[(\hat{z}\otimes I_{m\times m})Q_{\hat{\lambda}}^{-1}(\hat{z}^T\otimes I_{m\times m})\mathrm{vec}(\tilde{D})]}{\partial \hat{\boldsymbol{x}}}\bigg|_{D,\boldsymbol{x}} \boldsymbol{\delta x} \tag{58}$$

where $\boldsymbol{\delta x} = \hat{\boldsymbol{x}} - \boldsymbol{x}$ and $\bar{G} \triangleq I_{m(n+1)\times m(n+1)} - \bar{R}(z\otimes I_{m\times m}) Q_{\lambda}^{-1}(z^T\otimes I_{m\times m})$. The first partial derivative in Eq. (58) is clearly the identity matrix. There are three terms (i.e., $\hat{z}\otimes I_{m\times m}$, $Q_{\hat{\lambda}}^{-1}$, and $\hat{z}^T\otimes I_{m\times m}$) that depend on $\hat{\boldsymbol{x}}$ in the second partial derivative. Since

$$(\hat{z}^T\otimes I_{m\times m})\mathrm{vec}(\tilde{D}) = \tilde{D}\,\hat{z} \tag{59}$$

then Eq. (58) can be rewritten as

$$\mathrm{vec}(\hat{D}) = \mathrm{vec}(D) + \bar{G}\mathrm{vec}(\Delta D) - \bar{R}\frac{\partial[(\hat{z}\otimes I_{m\times m})Q_{\hat{\lambda}}^{-1}\tilde{D}\,\hat{z}]}{\partial \hat{\boldsymbol{x}}}\bigg|_{D,\boldsymbol{x}} \boldsymbol{\delta x} \tag{60}$$

Because all partial derivatives are evaluated at their true values, the partial derivatives involving first two terms (i.e., $\hat{z}\otimes I_{m\times m}$ and $Q_{\hat{\lambda}}^{-1}$) will lead to zero terms because they satisfy $Dz = \boldsymbol{0}$. Thus, these partial derivatives do not need to be computed. The remaining partial derivative term is given by $(\partial\tilde{D}\,\hat{z}/\partial\hat{\boldsymbol{x}})|_{D,\boldsymbol{x}} = \tilde{H}|_H = H$. Hence, Eq. (60) is given by

$$\mathrm{vec}(\hat{D}) = \mathrm{vec}(D) + \bar{G}\mathrm{vec}(\Delta D) - \bar{R}(z\otimes I_{m\times m})Q_{\lambda}^{-1}H\boldsymbol{\delta x} \tag{61}$$

Since $\hat{\boldsymbol{x}}$ minimizes the loss function $J(\hat{\boldsymbol{x}})$ given by Eq. (21), it can be shown that

$$\boldsymbol{\delta x} \approx -\left(\frac{\partial^2 J}{\partial\boldsymbol{x}\partial\boldsymbol{x}^T}\right)^{-1}\frac{\partial^2 J}{\partial\boldsymbol{x}\partial\mathrm{vec}^T(D)}\mathrm{vec}(\Delta D) \tag{62}$$

where the Hessian $(\partial^2 J/\partial\boldsymbol{x}\partial\boldsymbol{x}^T)$ is given by Eq. (55) and

$$\frac{\partial^2 J}{\partial\boldsymbol{x}\partial\mathrm{vec}^T(D)} = H^T Q_{\lambda}^{-1}(z^T\otimes I_{m\times m}) \tag{63}$$

It follows that

$$\begin{aligned} P &\triangleq E\{\boldsymbol{\delta x}\boldsymbol{\delta x}^T\} \\ &\approx (H^TQ_{\lambda}^{-1}H)^{-1}[H^TQ_{\lambda}^{-1}(z^T\otimes I_{m\times m})]\bar{R}[(z\otimes I_{m\times m})Q_{\lambda}^{-1}H] \\ &\quad\times(H^TQ_{\lambda}^{-1}H)^{-1} \\ &= (H^TQ_{\lambda}^{-1}H)^{-1} \\ &= F^{-1} \end{aligned} \tag{64}$$

where the second equality uses the definition of Q_{λ}.

B. Error-Covariances of $\mathrm{vec}(\hat{D})$ and $\mathrm{vec}(\hat{D}^T)$

From Eq. (62), the cross error-covariance P_{xD} is given by

$$P_{xD} \triangleq E\{\,\boldsymbol{\delta x}\;\mathrm{vec}^T(\Delta D)\,\} = -(H^TQ_{\lambda}^{-1}H)^{-1}H^TQ_{\lambda}^{-1}(z^T\otimes I_{m\times m})\bar{R} \tag{65}$$

Taking the expectation of Eq. (61), and assuming unbiased parameter estimates shows that $E\{\mathrm{vec}(\hat{D})\} = \mathrm{vec}(D)$. Thus, the error-covariance of $\mathrm{vec}(\hat{D})$ is given by

$$\begin{aligned} \mathrm{cov}\{\mathrm{vec}(\hat{D})\} &\triangleq E\{[\mathrm{vec}(\hat{D})-\mathrm{vec}(D)][\mathrm{vec}(\hat{D})-\mathrm{vec}(D)]^T\} \\ &= \bar{G}\bar{R}\bar{G}^T + \bar{R}(z\otimes I_{m\times m})Q_{\lambda}^{-1}H(H^TQ_{\lambda}^{-1}H)^{-1}H^TQ_{\lambda}^{-1}(z^T\otimes I_{m\times m})\bar{R} \\ &\quad - \bar{R}(z\otimes I_{m\times m})Q_{\lambda}^{-1}HP_{xD}\bar{G}^T - [\bar{R}(z\otimes I_{m\times m})Q_{\lambda}^{-1}HP_{xD}\bar{G}^T]^T \end{aligned} \tag{66}$$

The last two terms vanish because

$$\begin{aligned} P_{xD}\bar{G}^T &= -(H^TQ_{\lambda}^{-1}H)^{-1}H^TQ_{\lambda}^{-1}(z^T\otimes I_{m\times m})\bar{R} \\ &\quad\times[I_{m(n+1)\times m(n+1)} - \bar{R}(z\otimes I_{m\times m})Q_{\lambda}^{-1}(z^T\otimes I_{m\times m})]^T \\ &= -(H^TQ_{\lambda}^{-1}H)^{-1}[H^TQ_{\lambda}^{-1}(z^T\otimes I_{m\times m})\bar{R} \\ &\quad - H^TQ_{\lambda}^{-1}(z^T\otimes I_{m\times m})\bar{R}(z\otimes I_{m\times m})Q_{\lambda}^{-1}(z\otimes I_{m\times m})\bar{R}] \\ &= -(H^TQ_{\lambda}^{-1}H)^{-1}\{H^TQ_{\lambda}^{-1}(z^T\otimes I_{m\times m})\bar{R} - H^TQ_{\lambda}^{-1}(z^T\otimes I_{m\times m})\bar{R}\} \\ &= 0 \end{aligned} \tag{67}$$

Hence,

$$\begin{aligned} \mathrm{cov}\{\mathrm{vec}(\hat{D})\} &= \bar{G}\bar{R}\bar{G}^T + \bar{R}(z\otimes I_{m\times m})Q_{\lambda}^{-1}H(H^TQ_{\lambda}^{-1}H)^{-1}H^T \\ &\qquad\times Q_{\lambda}^{-1}(z^T\otimes I_{m\times m})\bar{R} \\ &= \bar{R} + \bar{R}(z\otimes I_{m\times m})[Q_{\lambda}^{-1}HF^{-1}H^TQ_{\lambda}^{-1} - Q_{\lambda}^{-1}] \\ &\qquad\times(z^T\otimes I_{m\times m})\bar{R} \end{aligned} \tag{68}$$

Using $\mathrm{vec}(\hat{D}) = K_{n+1,m}^T\mathrm{vec}(\hat{D}^T)$, and a number of matrix identities, the error-covariance of $\mathrm{vec}(\hat{D}^T)$ can likewise be shown to be given by

$$\begin{aligned} \mathrm{cov}\{\mathrm{vec}(\hat{D}^T)\} &\triangleq E\{[\mathrm{vec}(\hat{D}^T)-\mathrm{vec}(D^T)][\mathrm{vec}(\hat{D}^T)-\mathrm{vec}(D^T)]^T\} \\ &= GRG^T + R(I_{m\times m}\otimes z)Q_{\lambda}^{-1}H(H^TQ_{\lambda}^{-1}H)^{-1}H^TQ_{\lambda}^{-1}(I_{m\times m}\otimes z^T)R \\ &= R + R(I_{m\times m}\otimes z)[Q_{\lambda}^{-1}HF^{-1}H^TQ_{\lambda}^{-1} - Q_{\lambda}^{-1}](I_{m\times m}\otimes z^T)R \end{aligned} \tag{69}$$

where

$$G \triangleq I_{m(n+1)\times m(n+1)} - R(I_{m\times m}\otimes z)Q_{\lambda}^{-1}(I_{m\times m}\otimes z^T) \tag{70}$$

As with their respective estimates, in practice Eq. (12) would be used for Eq. (68), because it is a function of R, whereas Eq. (22) would be used for Eq. (69), because it is a function of $\bar{R}$. Also, in practice the estimates or measurements must be employed to compute the error-covariances of $\mathrm{vec}(\hat{D})$ and $\mathrm{vec}(\hat{D}^T)$.

If $m = n$ and H is invertible, then $\mathrm{cov}\{\mathrm{vec}(\hat{D})\} = \bar{R}$ and $\mathrm{cov}\{\mathrm{vec}(\hat{D}^T)\} = R$ because

$$[Q_{\lambda}^{-1}H(H^TQ_{\lambda}^{-1}H)^{-1}H^TQ_{\lambda}^{-1} - Q_{\lambda}^{-1}] = Q_{\lambda}^{-1} - Q_{\lambda}^{-1} = 0 \tag{71}$$

in this case. In general, the rank of $\mathrm{cov}\{\mathrm{vec}(\hat{D})\} - \bar{R}$ or $\mathrm{cov}\{\mathrm{vec}(\hat{D}^T)\} - R$ is $(m-n)$. That is because

$$\begin{aligned} \mathrm{rank}\{\mathrm{cov}\{\mathrm{vec}(\hat{D})\} - \bar{R}\} &= \mathrm{rank}\{\mathrm{cov}\{\mathrm{vec}(\hat{D}^T)\} - R\} \\ &= \mathrm{rank}\{Q_{\lambda}^{-1}H(H^TQ_{\lambda}^{-1}H)^{-1}H^TQ_{\lambda}^{-1} - Q_{\lambda}^{-1}\} \end{aligned} \tag{72}$$

which is because $\bar{R}$, R, $z\otimes I_{m\times m}$, and $I_{m\times m}\otimes z$ in Eqs. (68) and (69) are full-rank matrices. Since

$$[Q_{\lambda}^{-1}H(H^TQ_{\lambda}^{-1}H)^{-1}H^TQ_{\lambda}^{-1} - Q_{\lambda}^{-1}]H = Q_{\lambda}^{-1}H - Q_{\lambda}^{-1}H = 0 \tag{73a}$$

$$H^T[Q_{\lambda}^{-1}H(H^TQ_{\lambda}^{-1}H)^{-1}H^TQ_{\lambda}^{-1} - Q_{\lambda}^{-1}] = H^TQ_{\lambda}^{-1} - H^TQ_{\lambda}^{-1} = 0 \tag{73b}$$

and the $m\times n$ matrix H has rank n (for the TLS problem to have a unique solution), the rank-deficient $m\times m$ matrix $[Q_{\lambda}^{-1}H(H^TQ_{\lambda}^{-1}H)^{-1}H^TQ_{\lambda}^{-1} - Q_{\lambda}^{-1}]$ has rank $(m-n)$. Therefore, the rank of

$\text{cov}\{\text{vec}(\hat{D})\} - \bar{R}$ or $\text{cov}\{\text{vec}(\hat{D}^T)\} - R$ is $(m - n)$. The $(mn + m) \times (mn + m)$ error-covariance matrices $\text{cov}\{\text{vec}(\hat{D})\}$ and $\text{cov}\{\text{vec}(\hat{D}^T)\}$ are rank-deficient, too, and have rank $(mn + n)$, the minimal number of parameters to represent $\hat{D}$.

C. Error-Covariance of vec$(\hat{H})$

The quantities $\text{vec}(\hat{H})$ and $\hat{\boldsymbol{y}}$ can be extracted from $\text{vec}(\hat{D})$ using $\text{vec}(\hat{H}) = \mathcal{C}\text{vec}(\hat{D})$ and $\hat{\boldsymbol{y}} = \mathcal{D}\text{vec}(\hat{D})$, where

$$\mathcal{C} \triangleq [\, I_{(mn)\times(mn)} \quad 0_{(mn)\times m} \,] \quad \text{and} \quad \mathcal{D} \triangleq [\, 0_{m\times(mn)} \quad I_{m\times m} \,] \tag{74}$$

The error-covariance matrices of $\text{vec}(\hat{H})$ and $\hat{\boldsymbol{y}}$ can be readily retrieved from either $\text{cov}\{\text{vec}(\hat{D})\}$ or $\text{cov}\{\text{vec}(\hat{D}^T)\}$:

$$\begin{aligned}\text{cov}\{\text{vec}(\hat{H})\} &= \mathcal{C}\text{cov}\{\text{vec}(\hat{D})\}\mathcal{C}^T \\ &= [\mathcal{C}K^T_{(mn+m)\times(mn+m)}]\text{cov}\{\text{vec}(\hat{D}^T)\}[\mathcal{C}K^T_{(mn+m)\times(mn+m)}]^T\end{aligned} \tag{75}$$

$$\begin{aligned}\text{cov}\{\hat{\boldsymbol{y}}\} &= \mathcal{D}\text{cov}\{\text{vec}(\hat{D})\}\mathcal{D}^T \\ &= [\mathcal{D}K^T_{(mn+m)\times(mn+m)}]\text{cov}\{\text{vec}(\hat{D}^T)\}[\mathcal{D}K^T_{(mn+m)\times(mn+m)}]^T\end{aligned} \tag{76}$$

It is now proven that the error-covariance $\text{cov}\{\text{vec}(\hat{H})\}$ can be approximated as the inverse of the FIM for $\text{vec}(\hat{H})$:

$$\text{cov}\{\text{vec}(\hat{H})\} \approx (F_{hh} - F^T_{xh}F^{-1}_{xx}F_{xh})^{-1} \tag{77}$$

with the right-hand-side matrices defined by Eq. (36). Using the Sherman–Morrison–Woodbury formula on Eq. (77) gives

$$\text{cov}\{\text{vec}(\hat{H})\} \approx F^{-1}_{hh} + F^{-1}_{hh}F^T_{xh}PF_{xh}F^{-1}_{hh} \tag{78}$$

where $P \approx F^{-1} = (F^{-1}_{xx} - F_{xh}F^{-1}_{hh}F^T_{xh})^{-1}$ has been used. The error-covariance matrices given by Eqs. (75) and (78) are identical. That is,

$$\mathcal{C}\text{cov}\{\text{vec}(\hat{D})\}\mathcal{C}^T = F^{-1}_{hh} + F^{-1}_{hh}F^T_{xh}PF_{xh}F^{-1}_{hh} \tag{79}$$

This result is now proved. Using Eq. (68) with $P = (H^T Q^{-1}_\lambda H)^{-1}$ gives

$$\begin{aligned}\mathcal{C}\text{cov}\{\text{vec}(\hat{D})\}\mathcal{C}^T = \mathcal{C}\{&\bar{R} + \bar{R}(\boldsymbol{z} \otimes I_{m\times m})[Q^{-1}_\lambda HPH^T Q^{-1}_\lambda - Q^{-1}_\lambda] \\ &\times (\boldsymbol{z}^T \otimes I_{m\times m})\bar{R}\}\mathcal{C}^T\end{aligned} \tag{80}$$

Rearranging terms gives

$$\begin{aligned}\mathcal{C}\text{cov}\{\text{vec}(\hat{D})\}\mathcal{C}^T &= \mathcal{C}[\bar{R} - \bar{R}(\boldsymbol{z} \otimes I_{m\times m})Q^{-1}_\lambda(\boldsymbol{z}^T \otimes I_{m\times m})\bar{R}]\mathcal{C}^T \\ &+ [\mathcal{C}\bar{R}(\boldsymbol{z} \otimes I_{m\times m})Q^{-1}_\lambda H]P[\mathcal{C}\bar{R}(\boldsymbol{z} \otimes I_{m\times m})Q^{-1}_\lambda H]^T\end{aligned} \tag{81}$$

Therefore, the following sufficient conditions for Eq. (79) to be true are given:

$$\mathcal{C}[\bar{R} - \bar{R}(\boldsymbol{z} \otimes I_{m\times m})Q^{-1}_\lambda(\boldsymbol{z}^T \otimes I_{m\times m})\bar{R}]\mathcal{C}^T = F^{-1}_{hh} \tag{82a}$$

$$\mathcal{C}\bar{R}(\boldsymbol{z} \otimes I_{m\times m})Q^{-1}_\lambda H = F^{-1}_{hh}F^T_{xh} \tag{82b}$$

Note that $[\bar{R} - \bar{R}(\boldsymbol{z} \otimes I_{m\times m})Q^{-1}_\lambda(\boldsymbol{z}^T \otimes I_{m\times m})\bar{R}] = \bar{G}\bar{R}$. Equation (82a) is now proved. Premultiplying Eq. (23) by $\mathcal{C}$ gives

$$\text{vec}(\hat{H}) = \mathcal{C}[I_{m(n+1)\times m(n+1)} - \bar{R}(\hat{\boldsymbol{z}} \otimes I_{m\times m})Q^{-1}_{\hat{\lambda}}(\hat{\boldsymbol{z}}^T \otimes I_{m\times m})]\text{vec}(\tilde{D}) \tag{83}$$

On the other hand, substituting the definition of F^{-1}_{hh} into Eq. (25) gives

$$\text{vec}(\hat{H}) = F^{-1}_{hh}\begin{bmatrix} I_{(mn)\times(mn)} \\ \hat{\boldsymbol{x}}^T \otimes I_{m\times m}\end{bmatrix}^T \bar{R}^{-1}\text{vec}(\tilde{D}) \tag{84}$$

Since the two estimates of $\text{vec}(H)$ are identical, then

$$\begin{aligned}\mathcal{C}[I_{m(n+1)\times m(n+1)} &- \bar{R}(\boldsymbol{z} \otimes I_{m\times m})Q^{-1}_\lambda(\boldsymbol{z}^T \otimes I_{m\times m})] \\ &= F^{-1}_{hh}\begin{bmatrix} I_{(mn)\times(mn)} \\ \boldsymbol{x}^T \otimes I_{m\times m}\end{bmatrix}^T \bar{R}^{-1}\end{aligned} \tag{85}$$

Postmultiplying both sides of Eq. (85) by $\bar{R}\mathcal{C}^T$ gives

$$\begin{aligned}\mathcal{C}[\bar{R} &- \bar{R}(\boldsymbol{z} \otimes I_{m\times m})Q^{-1}_\lambda(\boldsymbol{z}^T \otimes I_{m\times m})\bar{R}]\mathcal{C}^T \\ &= F^{-1}_{hh}\begin{bmatrix} I_{(mn)\times(mn)} \\ \boldsymbol{x}^T \otimes I_{m\times m}\end{bmatrix}^T \mathcal{C}^T = F^{-1}_{hh}\end{aligned} \tag{86}$$

where the last equality is derived using

$$\begin{aligned}\begin{bmatrix} I_{(mn)\times(mn)} \\ \boldsymbol{x}^T \otimes I_{m\times m}\end{bmatrix}^T \mathcal{C}^T &= [\, I_{(mn)\times(mn)} \quad \boldsymbol{x} \otimes I_{m\times m} \,]\begin{bmatrix} I_{(mn)\times(mn)} \\ 0_{m\times(mn)}\end{bmatrix} \\ &= I_{(mn)\times(mn)}\end{aligned} \tag{87}$$

This proves Eq. (82a). Postmultiplying both sides of Eq. (85) by $[\, 0_{m\times(mn)} \quad H^T \,]^T$ and using the definition of F^T_{xh} gives

$$\mathcal{C}\begin{bmatrix} 0_{(mn)\times m} \\ H\end{bmatrix} - \mathcal{C}\bar{R}(\boldsymbol{z} \otimes I_{m\times m})Q^{-1}_\lambda(\boldsymbol{z}^T \otimes I_{m\times m})\begin{bmatrix} 0_{(mn)\times m} \\ H\end{bmatrix} = F^{-1}_{hh}F^T_{xh} \tag{88}$$

Substituting

$$\mathcal{C}\begin{bmatrix} 0_{(mn)\times m} \\ H\end{bmatrix} = 0_{(mn)\times m} \tag{89a}$$

$$(\boldsymbol{z}^T \otimes I_{m\times m})\begin{bmatrix} 0_{(mn)\times m} \\ H\end{bmatrix} = [\boldsymbol{x}^T \otimes I_{m\times m} \quad -I_{m\times m}]\begin{bmatrix} 0_{(mn)\times m} \\ H\end{bmatrix} = -H \tag{89b}$$

into the left-hand-side of Eq. (88) gives Eq. (82b). This completes of the proof of Eq. (79). Finally, it is pointed out that the error-covariance $\text{cov}\{\hat{\boldsymbol{y}}\}$ can be derived from

$$\begin{aligned}\hat{\boldsymbol{y}} &= \hat{H}\hat{\boldsymbol{x}} = (\hat{\boldsymbol{x}}^T \otimes I_{m\times m})\text{vec}(\hat{H}) \\ &= (\hat{\boldsymbol{x}}^T \otimes I_{m\times m})\mathcal{C}[I_{m(n+1)\times m(n+1)} - \bar{R}(\hat{\boldsymbol{z}} \otimes I_{m\times m})Q^{-1}_{\hat{\lambda}} \\ &\quad \times (\hat{\boldsymbol{z}}^T \otimes I_{m\times m})]\text{vec}(\tilde{D})\end{aligned} \tag{90}$$

where Eq. (83) has been used. Linearizing around the true values gives

$$\begin{aligned}\hat{\boldsymbol{y}} \approx \boldsymbol{y} &+ [(\boldsymbol{x}^T \otimes I_{m\times m})\mathcal{C}\bar{G}]\text{vec}(\Delta D) \\ &+ [I_{m\times m} - (\boldsymbol{x}^T \otimes I_{m\times m})\mathcal{C}\bar{G}(\boldsymbol{z} \otimes I_{m\times m})Q^{-1}_\lambda]H\boldsymbol{\delta x}\end{aligned} \tag{91}$$

It can be shown that

$$(\boldsymbol{x}^T \otimes I_{m\times m})\mathcal{C}\bar{G} = \mathcal{D}\bar{G} \tag{92a}$$

$$I_{m\times m} - (\boldsymbol{x}^T \otimes I_{m\times m})\mathcal{C}\bar{G}(\boldsymbol{z} \otimes I_{m\times m})Q^{-1}_\lambda = -\mathcal{D}\bar{R}(\boldsymbol{z} \otimes I_{m\times m})Q^{-1}_\lambda \tag{92b}$$

So, $\hat{\boldsymbol{y}}$ can be rewritten as $\hat{\boldsymbol{y}} = \mathcal{D}\text{vec}(\hat{D})$. Therefore, the error-covariance $\text{cov}\{\hat{\boldsymbol{y}}\} = \mathcal{D}\text{cov}\{\text{vec}(\hat{D})\}\mathcal{D}^T$.

VI. Elementwise Uncorrelated Case

Equation (14) is useful for analysis purposes to relate the general TLS to the elementwise uncorrelated case shown in [10]. For this case the covariance matrix is given by the following block diagonal matrix: $R = \text{blkdiag}[\mathcal{R}_1 \quad \mathcal{R}_2 \quad \cdots \quad \mathcal{R}_m]$, where each $\mathcal{R}_i$ is an $(n+1)\times(n+1)$ matrix given by

$$\mathcal{R}_i = \begin{bmatrix} \mathcal{R}_{hh_i} & \mathcal{R}_{hy_i} \\ \mathcal{R}_{hy_i}^T & \mathcal{R}_{yy_i} \end{bmatrix} \tag{93}$$

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 $\Delta\boldsymbol{y}$ by their rows:

$$\Delta H = \begin{bmatrix} \boldsymbol{\delta h}_1^T \\ \boldsymbol{\delta h}_2^T \\ \vdots \\ \boldsymbol{\delta h}_m^T \end{bmatrix}, \qquad \Delta\boldsymbol{y} = \begin{bmatrix} \delta y_1 \\ \delta y_2 \\ \vdots \\ \delta y_m \end{bmatrix} \tag{94}$$

where each $\boldsymbol{\delta h}_i$ has dimension $n\times 1$ and each δy_i is a scalar. The partitions in Eq. (93) are then given by

$$\mathcal{R}_{hh_i} = E\{\boldsymbol{\delta h}_i \boldsymbol{\delta h}_i^T\} \tag{95a}$$

$$\mathcal{R}_{hy_i} = E\{\delta y_i \boldsymbol{\delta h}_i\} \tag{95b}$$

$$\mathcal{R}_{yy_i} = E\{\delta y_i^2\} \tag{95c}$$

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

A. Loss Function Analysis

This section shows that the loss function in Eq. (14) is equivalent to the one derived in [10] for the elementwise uncorrelated case. Since R is now block diagonal, then the quantity $R(I_{m\times m}\otimes\hat{z})$ in the matrix $Q_{\hat{\lambda}}$ is explicitly given by

$$R(I_{m\times m}\otimes\hat{z}) = \begin{bmatrix} \mathcal{R}_1\hat{z} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathcal{R}_2\hat{z} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathcal{R}_m\hat{z} \end{bmatrix} \tag{96}$$

Then $Q_{\hat{\lambda}}$ reduces down to

$$Q_{\hat{\lambda}} = \text{diag}[\,\hat{z}^T\mathcal{R}_1\hat{z} \quad \hat{z}^T\mathcal{R}_2\hat{z} \quad \cdots \quad \hat{z}^T\mathcal{R}_m\hat{z}\,] \tag{97}$$

Also, the quantity $(I_{m\times m}\otimes\hat{z})Q_{\hat{\lambda}}^{-1}(I_{m\times m}\otimes\hat{z}^T)$ in Eq. (14) is given by

$$(I_{m\times m}\otimes\hat{z})Q_{\hat{\lambda}}^{-1}(I_{m\times m}\otimes\hat{z}^T) = Q_{\hat{\lambda}}^{-1}\otimes(\hat{z}\hat{z}^T) = \text{blkdiag}\left[\frac{\hat{z}\hat{z}^T}{\hat{z}^T\mathcal{R}_1\hat{z}} \; \frac{\hat{z}\hat{z}^T}{\hat{z}^T\mathcal{R}_2\hat{z}} \; \cdots \; \frac{\hat{z}\hat{z}^T}{\hat{z}^T\mathcal{R}_m\hat{z}}\right] \tag{98}$$

Partition the matrix $\tilde{D}$, $\hat{D}$, and D by their rows:

$$\tilde{D} = \begin{bmatrix} \tilde{\boldsymbol{d}}_1^T \\ \tilde{\boldsymbol{d}}_2^T \\ \vdots \\ \tilde{\boldsymbol{d}}_m^T \end{bmatrix}, \qquad \hat{D} = \begin{bmatrix} \hat{\boldsymbol{d}}_1^T \\ \hat{\boldsymbol{d}}_2^T \\ \vdots \\ \hat{\boldsymbol{d}}_m^T \end{bmatrix}, \qquad D = \begin{bmatrix} \boldsymbol{d}_1^T \\ \boldsymbol{d}_2^T \\ \vdots \\ \boldsymbol{d}_m^T \end{bmatrix} \tag{99}$$

where each of $\tilde{\boldsymbol{d}}_i$, $\hat{\boldsymbol{d}}_i$, and $\boldsymbol{d}_i$ has dimension $(n+1)\times 1$. Then the loss function in Eq. (14) reduces down to

$$J(\hat{\boldsymbol{x}}) = \frac{1}{2}\sum_{i=1}^{m}\frac{(\tilde{\boldsymbol{d}}_i^T\hat{z})^2}{\hat{z}^T\mathcal{R}_i\hat{z}} \tag{100}$$

which is equivalent to the one derived in [10].

For the elementwise uncorrelated case Q_λ is given by Eq. (97) evaluated at its true values:

$$Q_\lambda = \text{diag}[\,z^T\mathcal{R}_1 z \quad z^T\mathcal{R}_2 z \quad \cdots \quad z^T\mathcal{R}_m z\,] \tag{101}$$

Substituting Eq. (101) into Eq. (52) gives

$$F = \sum_{i=1}^{m}\frac{\boldsymbol{h}_i\boldsymbol{h}_i^T}{z^T\mathcal{R}_i z} \tag{102}$$

where $\boldsymbol{h}_i^T$ is the ith row of H. Equation (102) is equivalent to the FIM for the elementwise uncorrelated case derived in [10]. If $\mathcal{R}_{hh_i}$ and $\mathcal{R}_{hy_i}$ are both zero, meaning that no errors exist in the measured basis functions, then the FIM reduces down to $F = \sum_{i=1}^{m}\mathcal{R}_{yy_i}^{-1}\boldsymbol{h}_i\boldsymbol{h}_i^T$, which is equivalent to the FIM for the standard least squares problem.

B. Error-Covariance Analysis

This section shows that the error-covariance in Eq. (69) is equivalent to the one derived in [10] for the elementwise uncorrelated case, which is first summarized here. Define the following matrices:

$$M_{h_i} \triangleq \begin{bmatrix} I_{n\times n} - \dfrac{\boldsymbol{b}_i\boldsymbol{x}^T}{z^T\mathcal{R}_i z} & \dfrac{\boldsymbol{b}_i}{z^T\mathcal{R}_i z} \end{bmatrix} \tag{103a}$$

$$N_{h_i} \triangleq \frac{\boldsymbol{b}_i\boldsymbol{h}_i^T}{z^T\mathcal{R}_i z} \tag{103b}$$

$$M_{y_i} \triangleq \begin{bmatrix} -\dfrac{\beta_i\boldsymbol{x}^T}{z^T\mathcal{R}_i z} & 1 + \dfrac{\beta_i}{z^T\mathcal{R}_i z} \end{bmatrix} \tag{103c}$$

$$N_{y_i} \triangleq \frac{\beta_i\boldsymbol{h}_i^T}{z^T\mathcal{R}_i z} \tag{103d}$$

where $\beta_i \triangleq \mathcal{R}_{hy_i}^T\boldsymbol{x} - \mathcal{R}_{yy_i}$ and $\boldsymbol{b}_i \triangleq \mathcal{R}_{hh_i}\boldsymbol{x} - \mathcal{R}_{hy_i}$. The error-covariance of $\hat{\boldsymbol{h}}_i$ is given by

$$P_{hh_i} = M_{h_i}\mathcal{R}_i M_{h_i}^T + N_{h_i} P N_{h_i}^T \tag{104}$$

The error-variance of $\hat{y}_i$ is given by

$$P_{yy_i} = M_{y_i}\mathcal{R}_i M_{y_i}^T + N_{y_i} P N_{y_i}^T \tag{105}$$

Also, the cross-covariance is given by

$$P_{hy_i} = M_{h_i}\mathcal{R}_i M_{y_i}^T + N_{h_i} P N_{y_i}^T \tag{106}$$

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

$$P_{dd_i} = \begin{bmatrix} P_{hh_i} & P_{hy_i} \\ P_{hy_i}^T & P_{yy_i} \end{bmatrix} \tag{107}$$

The matrices in Eq. (103) should be computed using the estimated values in practice because they are derived using $\boldsymbol{h}_i^T \boldsymbol{x} - \boldsymbol{y}_i = 0$. The estimates also obey $\hat{\boldsymbol{h}}_i^T \hat{\boldsymbol{x}}_i - \hat{\boldsymbol{y}}_i = 0$ by virtue of the required constraint in Eq. (8), but using the measurements with $\tilde{\boldsymbol{h}}_i^T \hat{\boldsymbol{x}}_i - \tilde{\boldsymbol{y}}_i = 0$ 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

$$P_{dd_i} = \begin{bmatrix} M_{h_i} & N_{h_i} \\ M_{y_i} & N_{y_i} \end{bmatrix} \begin{bmatrix} \mathcal{R}_i & 0_{(n+1)\times n} \\ 0_{(n+1)\times n}^T & P \end{bmatrix} \begin{bmatrix} M_{h_i} & N_{h_i} \\ M_{y_i} & N_{y_i} \end{bmatrix}^T \tag{108}$$

The ith block diagonal matrix of Eq. (69) is given by

$$P_{dd_i} = \mathcal{R}_i - \frac{\mathcal{R}_i z z^T \mathcal{R}_i}{z^T \mathcal{R}_i z} + \frac{\mathcal{R}_i z \boldsymbol{h}_i^T P \boldsymbol{h}_i z^T \mathcal{R}_i}{(z^T \mathcal{R}_i z)^2} \tag{109}$$

The quantity $\mathcal{R}_i z$ can be written using the definitions of $\boldsymbol{b}_i$ and β_i by $\mathcal{R}_i z = [\boldsymbol{b}_i^T \quad \beta_i]^T$. Therefore, P_{hh_i}, P_{hy_i}, and P_{yy_i} are given by

$$P_{hh_i} = \mathcal{R}_{hh_i} - \frac{\boldsymbol{b}_i \boldsymbol{b}_i^T}{z^T \mathcal{R}_i z} + \frac{\boldsymbol{b}_i \boldsymbol{h}_i^T P \boldsymbol{h}_i \boldsymbol{b}_i^T}{(z^T \mathcal{R}_i z)^2} \tag{110a}$$

$$P_{hy_i} = \mathcal{R}_{hy_i} - \frac{\beta_i \boldsymbol{b}_i}{z^T \mathcal{R}_i z} + \frac{\beta_i \boldsymbol{b}_i \boldsymbol{h}_i^T P \boldsymbol{h}_i}{(z^T \mathcal{R}_i z)^2} \tag{110b}$$

$$P_{yy_i} = \mathcal{R}_{yy_i} - \frac{\beta_i^2}{z^T \mathcal{R}_i z} + \frac{\beta_i^2 \boldsymbol{h}_i^T P \boldsymbol{h}_i}{(z^T \mathcal{R}_i z)^2} \tag{110c}$$

The next step is to show that the expressions given by Eqs. (107) and (108) are equivalent to the ones given by Eq. (110). Explicitly multiplying terms in Eq. (108) to determine P_{hh_i} gives

$$\begin{aligned} P_{hh_i} = \mathcal{R}_{hh_i} &- \frac{(\boldsymbol{b}_i \boldsymbol{x}^T \mathcal{R}_{hh_i} - \boldsymbol{b}_i \mathcal{R}_{hy_i}^T + \mathcal{R}_{hh_i} \boldsymbol{x} \boldsymbol{b}_i^T - \mathcal{R}_{hy_i} \boldsymbol{b}_i^T)}{z^T \mathcal{R}_i z} \\ &+ \frac{(\boldsymbol{b}_i \boldsymbol{x}^T \mathcal{R}_{hh_i} \boldsymbol{x} \boldsymbol{b}_i^T - 2\boldsymbol{b}_i \mathcal{R}_{hy_i}^T \boldsymbol{x} \boldsymbol{b}_i^T + \boldsymbol{b}_i \mathcal{R}_{yy_i} \boldsymbol{b}_i^T)}{(z^T \mathcal{R}_i z)^2} + \frac{\boldsymbol{b}_i \boldsymbol{h}_i^T P \boldsymbol{h}_i \boldsymbol{b}_i^T}{(z^T \mathcal{R}_i z)^2} \end{aligned} \tag{111}$$

From the definition of $\boldsymbol{b}_i$, and using $z^T \mathcal{R}_i z = \boldsymbol{x}^T \mathcal{R}_{hh_i} \boldsymbol{x} - 2\mathcal{R}_{hy_i}^T \boldsymbol{x} + \mathcal{R}_{yy_i}$, it can easily be seen that Eq. (111) is equivalent to Eq. (110a). Explicitly multiplying terms in Eq. (108) to determine P_{hy_i} gives

$$\begin{aligned} P_{hy_i} = \mathcal{R}_{hy_i} &- \frac{(\beta_i \mathcal{R}_{hh_i} \boldsymbol{x} - \beta_i \mathcal{R}_{hy_i} + \boldsymbol{b}_i \mathcal{R}_{hy_i}^T \boldsymbol{x} - \boldsymbol{b}_i \mathcal{R}_{yy_i})}{z^T \mathcal{R}_i z} \\ &+ \frac{(\beta_i \boldsymbol{b}_i \boldsymbol{x}^T \mathcal{R}_{hh_i} \boldsymbol{x} - 2\beta_i \boldsymbol{b}_i \mathcal{R}_{hy_i}^T \boldsymbol{x} + \beta_i \boldsymbol{b}_i \mathcal{R}_{yy_i})}{(z^T \mathcal{R}_i z)^2} + \frac{\beta_i \boldsymbol{b}_i \boldsymbol{h}_i^T P \boldsymbol{h}_i}{(z^T \mathcal{R}_i z)^2} \end{aligned} \tag{112}$$

From the definitions of $\boldsymbol{b}_i$ and β_i, and using $z^T \mathcal{R}_i z = \boldsymbol{x}^T \mathcal{R}_{hh_i} \boldsymbol{x} - 2\mathcal{R}_{hy_i}^T \boldsymbol{x} + \mathcal{R}_{yy_i}$, it can easily be seen that Eq. (112) is equivalent to Eq. (110b). Explicitly multiplying terms in Eq. (108) to determine P_{yy_i} gives

$$\begin{aligned} P_{yy_i} = \mathcal{R}_{yy_i} &- \frac{2(\beta_i \mathcal{R}_{hy_i}^T \boldsymbol{x} - \beta_i \mathcal{R}_{yy_i})}{z^T \mathcal{R}_i z} \\ &+ \frac{(\beta_i^2 \boldsymbol{x}^T \mathcal{R}_{hh_i} \boldsymbol{x} - 2\beta_i^2 \mathcal{R}_{hy_i}^T \boldsymbol{x} + \beta_i^2 \mathcal{R}_{yy_i})}{(z^T \mathcal{R}_i z)^2} + \frac{\beta_i^2 \boldsymbol{h}_i^T P \boldsymbol{h}_i}{(z^T \mathcal{R}_i z)^2} \end{aligned} \tag{113}$$

From the definition of β_i, and using $z^T \mathcal{R}_i z = \boldsymbol{x}^T \mathcal{R}_{hh_i} \boldsymbol{x} - 2\mathcal{R}_{hy_i}^T \boldsymbol{x} + \mathcal{R}_{yy_i}$, it can easily be seen that Eq. (113) is equivalent to Eq. (110c). This completes the proofs. It should be noted that using Eq. (109) or Eq. (111) to compute P_{dd_i} or its submatrices involves less computations than using Eq. (108).

Substituting $R = \text{blkdiag}[\mathcal{R}_1 \quad \mathcal{R}_2 \quad \cdots \quad \mathcal{R}_m]$ into Eq. (70) and using Eq. (98) evaluated at the true values leads to

$$G = I_{m(n+1)\times m(n+1)} - \text{blkdiag}\left[\frac{\mathcal{R}_1 z z^T}{z^T \mathcal{R}_1 z} \quad \frac{\mathcal{R}_2 z z^T}{z^T \mathcal{R}_2 z} \quad \cdots \quad \frac{\mathcal{R}_m z z^T}{z^T \mathcal{R}_m z}\right] \tag{114}$$

Therefore, GRG^T in Eq. (69) is given by

$$GRG^T = \text{blkdiag}[G_1 \mathcal{R}_1 G_1^T \quad G_2 \mathcal{R}_2 G_2^T \quad \cdots \quad G_m \mathcal{R}_m G_m^T] \tag{115}$$

where

$$G_i = I_{n+1} - \frac{\mathcal{R}_i z z^T}{z^T \mathcal{R}_i z} \tag{116}$$

This shows that the matrix GRG^T has a block diagonal structure. However, the matrix

$$R(I_{m\times m} \otimes z) Q_\lambda^{-1} H P H^T Q_\lambda^{-1} (I_{m\times m} \otimes z^T) R \tag{117}$$

does not have a block diagonal structure, even with a block diagonal R matrix. This means that even though the measurements are elementwise uncorrelated, their corresponding estimates are correlated. The ith estimate $\hat{\boldsymbol{d}}_i$ is given by [10]

$$\hat{\boldsymbol{d}}_i = \left[I_{(n+1)\times(n+1)} - \frac{\mathcal{R}_i \hat{z} \hat{z}^T}{\hat{z}^T \mathcal{R}_i \hat{z}}\right] \tilde{\boldsymbol{d}}_i \tag{118}$$

The cross-covariance terms are computed using

$$P_{d_i d_j} \triangleq E\{(\hat{\boldsymbol{d}}_i - E\{\hat{\boldsymbol{d}}_i\})(\hat{\boldsymbol{d}}_j - E\{\hat{\boldsymbol{d}}_j\})^T\} \tag{119}$$

For the elementwise uncorrelated case $E\{\tilde{\boldsymbol{d}}_i \tilde{\boldsymbol{d}}_j^T\} = \boldsymbol{d}_i \boldsymbol{d}_j^T$. Then the only contribution in the cross-covariance terms comes from the term $\mathcal{R}_i \hat{z} \hat{z}^T \boldsymbol{d}_i / (\hat{z}^T \mathcal{R}_i \hat{z})$. The estimate $\hat{z}$ is related to its true value z through

$$\hat{z} = z + \boldsymbol{\delta z} \tag{120}$$

where $\boldsymbol{\delta z} \triangleq [\boldsymbol{\delta x}^T \quad 0]^T$. Substituting Eq. (120) into $\mathcal{R}_i \hat{z} \hat{z}^T \boldsymbol{d}_i$ and ignoring second-order terms gives

$$\begin{aligned} \mathcal{R}_i \hat{z} \hat{z}^T \boldsymbol{d}_i &\approx \mathcal{R}_i z z^T \boldsymbol{d}_i + \mathcal{R}_i z \boldsymbol{\delta z}^T \boldsymbol{d}_i + \mathcal{R}_i \boldsymbol{\delta z} z^T \boldsymbol{d}_i \\ &= \mathcal{R}_i z \boldsymbol{\delta z}^T \boldsymbol{d}_i \end{aligned} \tag{121}$$

where the identity $\boldsymbol{d}_i^T z = 0$ has been used. Using the binomial series for a first-order expansion of $(\hat{z}^T \mathcal{R}_i \hat{z})^{-1}$ leads to the approximation

$$(\hat{z}^T \mathcal{R}_i \hat{z})^{-1} \approx (z^T \mathcal{R}_i z)^{-1} - 2(z^T R_i \boldsymbol{\delta z})(z^T \mathcal{R}_i z)^{-2} \tag{122}$$

Then, to within first-order $\mathcal{R}_i \hat{z} \hat{z}^T \boldsymbol{d}_i / (\hat{z}^T \mathcal{R}_i \hat{z})$ is given by

$$\frac{\mathcal{R}_i \hat{z} \hat{z}^T \boldsymbol{d}_i}{\hat{z}^T \mathcal{R}_i \hat{z}} \approx \frac{\mathcal{R}_i z \boldsymbol{d}_i^T \boldsymbol{\delta z}}{z^T \mathcal{R}_i z} = \frac{\mathcal{R}_i z \boldsymbol{h}_i^T}{z^T \mathcal{R}_i z} \boldsymbol{\delta x} \tag{123}$$

where the identity $\boldsymbol{d}_i^T \boldsymbol{\delta z} = \boldsymbol{h}_i^T \boldsymbol{\delta x}$ has been used. Equation (123) indicates that $E\{\hat{\boldsymbol{d}}_i\} = \boldsymbol{d}_i$ because the errors on $\tilde{\boldsymbol{d}}_i$ are zero mean. Using $E\{\boldsymbol{\delta x} \quad \boldsymbol{\delta x}^T\} = P$ and the fact that $E\{\hat{\boldsymbol{d}}_i\} = \boldsymbol{0}$ yields

$$P_{d_i d_j} = \frac{\mathcal{R}_i z \boldsymbol{h}_i^T P \boldsymbol{h}_j z^T \mathcal{R}_j}{(z^T \mathcal{R}_i z)(z^T \mathcal{R}_j z)} \tag{124}$$

Setting $m = 1$ in Eq. (69) gives $(I_{m\times m} \otimes z) = z$. Then the cross-covariance terms computed using Eq. (69) are given by

$$P_{d_i d_j} = \mathcal{R}_i z Q_{\lambda_i}^{-1} \boldsymbol{h}_i^T P \boldsymbol{h}_j Q_{\lambda_j}^{-1} z^T \mathcal{R}_j \tag{125}$$

where $Q_{\lambda_i} = z^T \mathcal{R}_i z$ and $Q_{\lambda_j} = z^T \mathcal{R}_j z$. This shows that Eq. (124) is indeed identical to Eq. (125). Therefore, the elementwise uncorrelated error-covariance expressions are identical to the ones derived from the generalized correlated case shown in this paper.

VII. Bearings-Only Localization

Total least squares is applied to estimate the three-dimensional location of a stationary target point using passive bearing measurements. The problem geometry is depicted in Fig. 1. The goal is to estimate the point p from bearings-only measurements, denoted by $\tilde{\theta}_i$ and $\tilde{\phi}_i$. The baseline points, denoted by X_i, Y_i, and Z_i, are assumed to be imprecisely known. The bearing measurement model and baseline point models are given by

$$\tilde{\theta}_i = \theta_i + \delta\theta_i \tag{126a}$$

$$\tilde{\phi}_i = \phi_i + \delta\phi_i \tag{126b}$$

$$\tilde{X}_i = X_i + \delta X_i \tag{126c}$$

$$\tilde{Y}_i = Y_i + \delta Y_i \tag{126d}$$

$$\tilde{Z}_i = Z_i + \delta Z_i \tag{126e}$$

where $\delta\theta_i$, $\delta\phi_i$, δX_i, δY_i, and δZ_i are zero-mean Gaussian noise processes with variances $\sigma^2_{\theta_i}$, $\sigma^2_{\phi_i}$, $\sigma^2_{X_i}$, $\sigma^2_{Y_i}$, and $\sigma^2_{Z_i}$, respectively. It is assumed that all these errors are uncorrelated with each other. The observations are modeled as

$$\theta_i = \tan^{-1}\left(\frac{y - Y_i}{x - X_i}\right) \tag{127a}$$

$$\phi_i = \sin^{-1}\left(\frac{z - Z_i}{\sqrt{(x - X_i)^2 + (y - Y_i)^2 + (z - Z_i)^2}}\right) \tag{127b}$$

Taking the tangent of both sides of Eq. (127a) leads to $y_{\theta_i} = \boldsymbol{h}^T_{\theta_i}\boldsymbol{x}$, with

$$y_{\theta_i} = -X_i \sin(\theta_i) + Y_i \cos(\theta_i) \tag{128a}$$

$$\boldsymbol{h}_{\theta_i} = [-\sin(\theta_i) \quad \cos(\theta_i) \quad 0]^T \tag{128b}$$

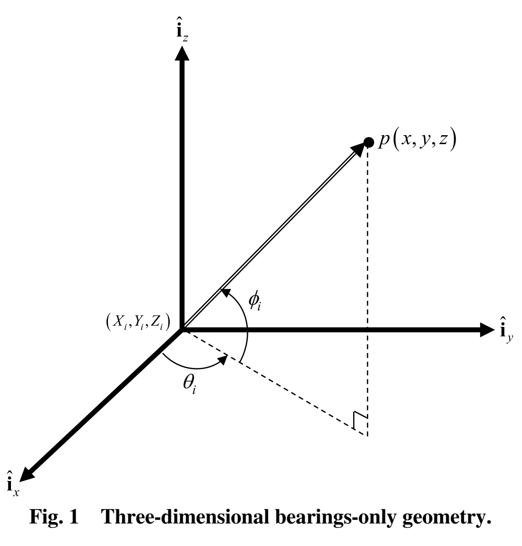


Fig. 1 Three-dimensional bearings-only geometry.

$$\boldsymbol{x} = [x \quad y \quad z]^T \tag{128c}$$

The linear model form involving ϕ_i is derived in [19] for the case of $X_i = Y_i = Z_i = 0$. Here the derivation is shown for nonzero values of X_i, Y_i, and Z_i. Taking the sine and squaring both sides of Eq. (127b) leads to

$$[(x - X_i)^2 + (y - Y_i)^2 + (z - Z_i)^2]\sin^2(\phi_i) = (z - Z_i)^2 \tag{129}$$

Substituting $y - Y_i = \tan(\theta_i)(x - X_i)$ in Eq. (129) gives

$$[1 + \tan^2(\theta_i)](x - X_i)^2\sin^2(\phi_i) = [1 - \sin^2(\phi_i)](z - Z_i)^2 \tag{130}$$

or

$$\frac{1}{\cos^2(\theta_i)}(x - X_i)^2\sin^2(\phi_i) = \cos^2(\phi_i)(z - Z_i)^2 \tag{131}$$

Taking the square root of both sides of Eq. (131) leads to $y_{\phi_i} = \boldsymbol{h}^T_{\phi_i}\boldsymbol{x}$, with

$$y_{\phi_i} = -X_i \sin(\phi_i) + Z_i \cos(\theta_i)\cos(\phi_i) \tag{132a}$$

$$\boldsymbol{h}_{\phi_i} = [-\sin(\phi_i) \ 0 \ \cos(\theta_i) \ \cos(\phi_i)]^T \tag{132b}$$

Note that derivation of Eq. (132) involves the process of squaring an equation and then taking the square root. This equation is correct because the straight line satisfies $((x - X_i)/\cos(\phi_i)\cos(\theta_i)) = ((y - Y_i)/\cos(\phi_i)\sin(\theta_i)) = ((z - Z_i)/\sin(\phi_i))$. The matrix H and vector $\boldsymbol{y}$ are given by

$$H = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_m \end{bmatrix}, \qquad \boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_m \end{bmatrix} \tag{133}$$

where

$$H_i = \begin{bmatrix} \boldsymbol{h}^T_{\theta_i} \\ \boldsymbol{h}^T_{\phi_i} \end{bmatrix} = \begin{bmatrix} -\sin(\theta_i) & \cos(\theta_i) & 0 \\ -\sin(\phi_i) & 0 & \cos(\theta_i)\cos(\phi_i) \end{bmatrix} \tag{134a}$$

$$\boldsymbol{y}_i = \begin{bmatrix} y_{\theta_i} \\ y_{\phi_i} \end{bmatrix} = \begin{bmatrix} -X_i \sin(\theta_i) + Y_i \cos(\theta_i) \\ -X_i \sin(\phi_i) + Z_i \cos(\theta_i)\cos(\phi_i) \end{bmatrix} \tag{134b}$$

The matrix D is given by the usual TLS form of $D = [H \quad \boldsymbol{y}]$. The matrix $\tilde{D}$ is formed by using the measurements in place of D, and the matrix $\hat{D}$ is formed by using the estimates in place of D. The linear form of Eq. (127) is preferred over the nonlinear form for two reasons. First, a linear least squares approach can be used to determine an initial guess of the point p. Second, the nonlinear form will result in a linearized H matrix that involves noise in its elements, which leads to a nonlinear TLS problem. It is important to note that the solution to the TLS problem is only an approximate one because the Gaussian measurement errors are no longer Gaussian through the transformation to a TLS form. Still, the approximate TLS solution is good for the realistic case of having fairly large signal-to-noise ratios, as will be seen by the results.

The two-row elements of $\tilde{D}$ are denoted by $\tilde{D}_i$, and are given by $\tilde{D}_i = [\tilde{H}_i \quad \tilde{\boldsymbol{y}}_i]$. The covariance matrix of the vector $\mathrm{vec}(\tilde{D}^T_i)$, denoted by R_i, is now shown. The procedure is shown in [10]. Only one term will be derived here, and the same procedure to derive all other terms follows along the same lines. The 1-7 element of R_i is given by $R_{17_i} = E\{-\sin(\tilde{\theta}_i)\cos(\tilde{\theta}_i)\cos(\tilde{\phi}_i)\} + \sin(\theta_i)\cos(\theta_i)\cos(\phi_i)$. The first-order approximations $\sin(\alpha + \delta\alpha) = \sin(\alpha) + \delta\alpha\cos(\alpha)$ and $\cos(\alpha + \delta\alpha) = \cos(\alpha) - \delta\alpha\sin(\alpha)$ for any angle α and small-angle perturbation $\delta\alpha$ are now used. Using the associated angles with these approximations gives

$$R_{17_i} = E\{\delta\theta_i \sin^2(\theta_i)\cos(\phi_i) - \delta\theta_i \cos^2(\theta_i)\cos(\phi_i) + \delta\theta_i^2 \sin(\theta_i)\cos(\theta_i)\cos(\phi_i) - \delta\theta_i\delta\phi_i \sin^2(\theta_i)\sin(\phi_i) + \delta\theta_i\delta\phi_i \cos^2(\theta_i)\sin(\phi_i) - \delta\theta_i^2\delta\phi_i \sin(\theta_i)\cos(\theta_i)\sin(\phi_i) + \delta\phi_i \sin(\theta_i)\cos(\theta_i)\cos(\phi_i)\} \tag{135}$$

Because $\delta\theta_i$ and $\delta\phi_i$ are each zero-mean processes and are also uncorrelated, then $R_{17_i} = \sigma_{\theta_i}^2 \sin(\theta_i)\cos(\theta_i)\cos(\phi_i)$. The matrix R_i is given by

$$R_i = \begin{bmatrix} R_{11_i} & R_{12_i} & 0 & R_{14_i} & 0 & 0 & R_{17_i} & R_{18_i} \\ R_{12_i} & R_{22_i} & 0 & R_{24_i} & 0 & 0 & R_{27_i} & R_{28_i} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ R_{14_i} & R_{24_i} & 0 & R_{44_i} & 0 & 0 & R_{47_i} & R_{48_i} \\ 0 & 0 & 0 & 0 & R_{55_i} & 0 & R_{57_i} & R_{58_i} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ R_{17_i} & R_{27_i} & 0 & R_{47_i} & R_{57_i} & 0 & R_{77_i} & R_{78_i} \\ R_{18_i} & R_{28_i} & 0 & R_{48_i} & R_{58_i} & 0 & R_{78_i} & R_{88_i} \end{bmatrix} \tag{136}$$

The diagonal elements are given by

$$R_{11_i} = \sigma_{\theta_i}^2 \cos^2(\theta_i) \tag{137a}$$

$$R_{22_i} = \sigma_{\theta_i}^2 \sin^2(\theta_i) \tag{137b}$$

$$R_{44_i} = \sigma_{\theta_i}^2 \{[X_i \cos(\theta_i) + Y_i \sin(\theta_i)]^2 + \sigma_{X_i}^2 \cos^2(\theta_i) + \sigma_{Y_i}^2 \sin^2(\theta_i)\} + \sigma_{X_i}^2 \sin^2(\theta_i) + \sigma_{Y_i}^2 \cos^2(\theta_i) \tag{137c}$$

$$R_{55_i} = \sigma_{\phi_i}^2 \cos^2(\phi_i) \tag{137d}$$

$$R_{77_i} = \sigma_{\theta_i}^2 \sin^2(\theta_i)\cos^2(\phi_i) + \sigma_{\phi_i}^2 \cos^2(\theta_i)\sin^2(\phi_i) + \sigma_{\theta_i}^2\sigma_{\phi_i}^2 \sin^2(\theta_i)\sin^2(\phi_i) \tag{137e}$$

$$R_{88_i} = \sigma_{\theta_i}^2 \sin^2(\theta_i)\cos^2(\phi_i)[Z_i^2 + \sigma_{\phi_i}^2 + \sigma_{Z_i}^2] + \sigma_{\phi_i}^2 \{[X_i \cos(\phi_i) + Z_i \cos(\theta_i)\sin(\phi_i)]^2 + \sigma_{X_i}^2 \cos^2(\phi_i) + \sigma_{Z_i}^2 \cos^2(\theta_i)\sin^2(\phi_i)\} + \sigma_{X_i}^2 \sin^2(\phi_i) + \sigma_{Z_i}^2 \cos^2(\theta_i)\cos^2(\phi_i) + \sigma_{\theta_i}^2\sigma_{\phi_i}^2\sigma_{Z_i}^2 \sin^2(\theta_i)\sin^2(\phi_i) \tag{137f}$$

The off-diagonal elements are given by

$$R_{12_i} = \sigma_{\theta_i}^2 \sin(\theta_i)\cos(\theta_i) \tag{138a}$$

$$R_{14_i} = \sigma_{\theta_i}^2 \cos(\theta_i)[X_i \cos(\theta_i) + Y_i \sin(\theta_i)] \tag{138b}$$

$$R_{17_i} = \sigma_{\theta_i}^2 \sin(\theta_i)\cos(\theta_i)\cos(\phi_i) \tag{138c}$$

$$R_{18_i} = \sigma_{\theta_i}^2 Z_i \sin(\theta_i)\cos(\theta_i)\cos(\phi_i) \tag{138d}$$

$$R_{24_i} = \sigma_{\theta_i}^2 \sin(\theta_i)[X_i \cos(\theta_i) + Y_i \sin(\theta_i)] \tag{138e}$$

$$R_{27_i} = \sigma_{\theta_i}^2 \sin^2(\theta_i)\cos(\phi_i) \tag{138f}$$

$$R_{28_i} = \sigma_{\theta_i}^2 Z_i \sin^2(\theta_i)\cos(\phi_i) \tag{138g}$$

$$R_{47_i} = \sigma_{\theta_i}^2 \sin(\theta_i)\cos(\phi_i)[X_i \cos(\theta_i) + Y_i \sin(\theta_i)] \tag{138h}$$

$$R_{48_i} = \sigma_{\theta_i}^2 Z_i \sin(\theta_i)\cos(\phi_i)[X_i \cos(\theta_i) + Y_i \sin(\theta_i)] + \sigma_{X_i}^2 \sin(\theta_i)\sin(\phi_i) \tag{138i}$$

$$R_{57_i} = \sigma_{\phi_i}^2 \cos(\theta_i)\sin(\phi_i)\cos(\phi_i) \tag{138j}$$

$$R_{58_i} = \sigma_{\phi_i}^2 \cos(\phi_i)[X_i \cos(\phi_i) + Z_i \cos(\theta_i)\sin(\phi_i)] \tag{138k}$$

$$R_{78_i} = \sigma_{\theta_i}^2 Z_i \sin^2(\theta_i)[\cos^2(\phi_i) + \sigma_{\phi_i}^2 \sin^2(\phi_i)] + \sigma_{\phi_i}^2 \cos(\theta_i)\sin(\phi_i)[X_i \cos(\phi_i) + Z_i \cos(\theta_i)\sin(\phi_i)] \tag{138l}$$

The measured values can be used in place of the true values, which leads to second-order error effects for large signal-to-noise ratios. The matrix R is then given by

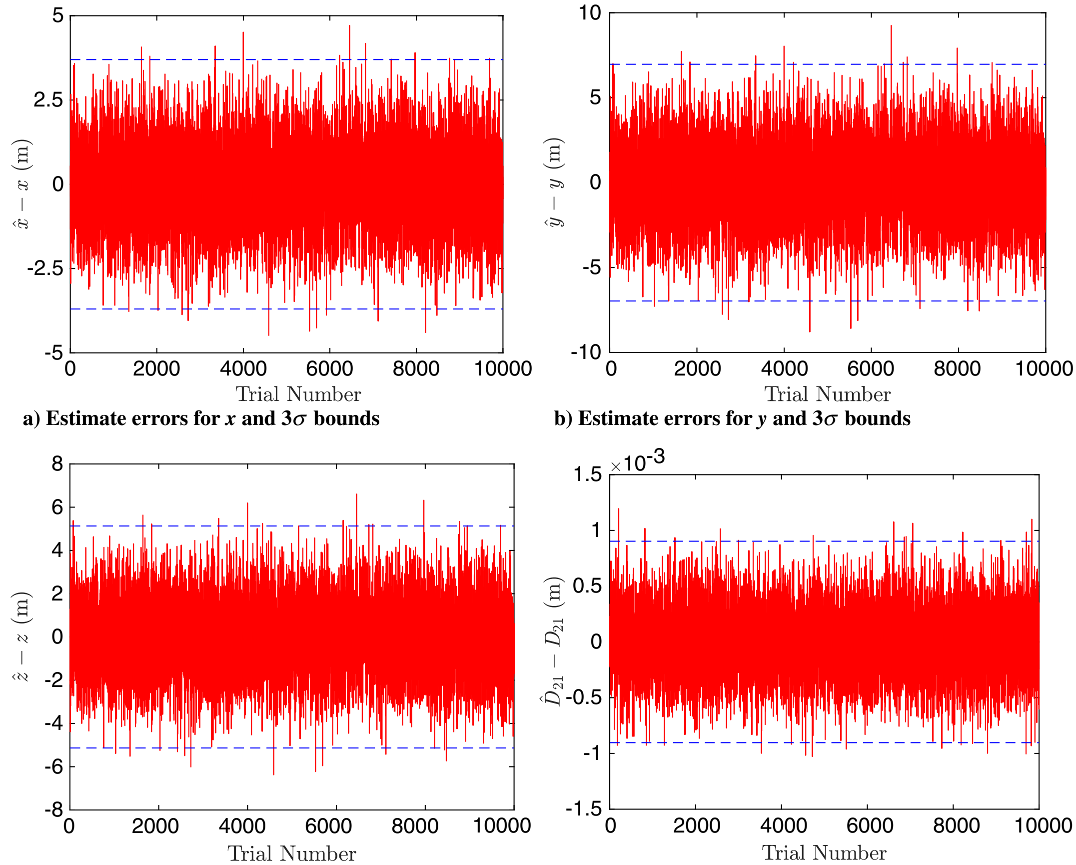
$$R = \text{blkdiag}[\,R_1 \quad R_2 \quad \cdots \quad R_m\,] \tag{139}$$

The matrix $\bar{R}$ used in the TLS solution can be obtained using $\bar{R} = K_{n+1,m}^T R K_{n+1,m}$.

In the simulation the location of the point p is given at (1000, 2000, 1500) m. The baseline points are time varying with $X_i = 500\sin(0.01t_i)$, $Y_i = 300\cos(0.2t_i)$, and $Z_i = 200\sin(0.2t_i)$. The variances are given by $\sigma_{\theta_i}^2 = \sigma_{\phi_i}^2 = (0.05\pi/180)^2\ \text{rad}^2$ and $\sigma_{X_i}^2 = \sigma_{Y_i}^2 = \sigma_{Z_i}^2 = 1\ \text{m}^2$ for all i points. The final time of the simulation run is 10 s, and measurements of $\tilde{\theta}_i$, $\tilde{\phi}_i$, $\tilde{X}_i$, $\tilde{Y}_i$, and $\tilde{Z}_i$ are taken at 0.05-s intervals. The number of measurements is thus $m = 201$. Ten thousand Monte Carlo runs are executed in order to compare the actual errors with the computed 3σ bounds using the inverse of Eq. (52). The iterative algorithm shown in Appendix B is used to determine the TLS estimate. The TLS initial estimate is given by using a standard linear least squares solution with the measurement covariance given by $\bar{R}_{yy}$. The dimension of $\bar{R}$ is 1608×1608, but its rank is only 804. This poses no issues in the TLS solution because the matrix Q_λ, which has dimension 402×402, is indeed full rank. Thus, it is invertible. The solution may not be guaranteed to be unique, though, because $\bar{R}$ is positive semidefinite. Methods to overcome this issue are discussed in [20], but are not required for this particular problem because the estimates are shown to be consistent.

Figure 2 shows the errors for the TLS estimates along with their respective computed 3σ bounds. This indicates that using the inverse of Eq. (52) can be used to accurately compute the 3σ bounds, and that the actual errors are consistent with the 3σ bounds. Also, Fig. 2d shows the estimate errors for the D_{21} element of the matrix D, which corresponds to $-\sin(\phi_1)$, the first element of $\boldsymbol{h}_{\phi_1}$, along with its computed 3σ bounds using Eq. (68). Again the estimate errors are consistent with their associated 3σ bounds. Plots of the standard least squares errors are shown in Fig. 3. The number of x-errors that are outside of their respective 3σ bounds is 2172 out of 10,000; the number of y-errors that are outside of their respective 3σ bounds is 2129 out of 10,000; and the number of z-errors that are outside of their respective 3σ bounds is 2160 out of 10,000. This shows that naïvely using the noisy design matrix in standard least squares can produce inconsistent estimates. But the total least squares solution overcomes this issue.

Another simulation is used to check the analytical expression for the bias shown in Eq. (A5). All parameters are the same as the previous simulation; however, the measurements are taken at 0.5-s intervals. Ten million Monte Carlo runs are executed in this simulation. The bias computed by Eq. (A5) is given by $\boldsymbol{\beta}_x = [\,0.0110 \quad 0.0208 \quad 0.0153\,]^T$. The TLS bias computed from the Monte Carlo runs is given by $[\,0.0119 \quad 0.0229 \quad 0.0172\,]^T$, which is fairly close to the theoretical value. The bias produced using standard least squares is given by $[\,-0.4132 \quad -0.7851 \quad -0.5753\,]^T$, which is an order of magnitude larger than the TLS estimates.



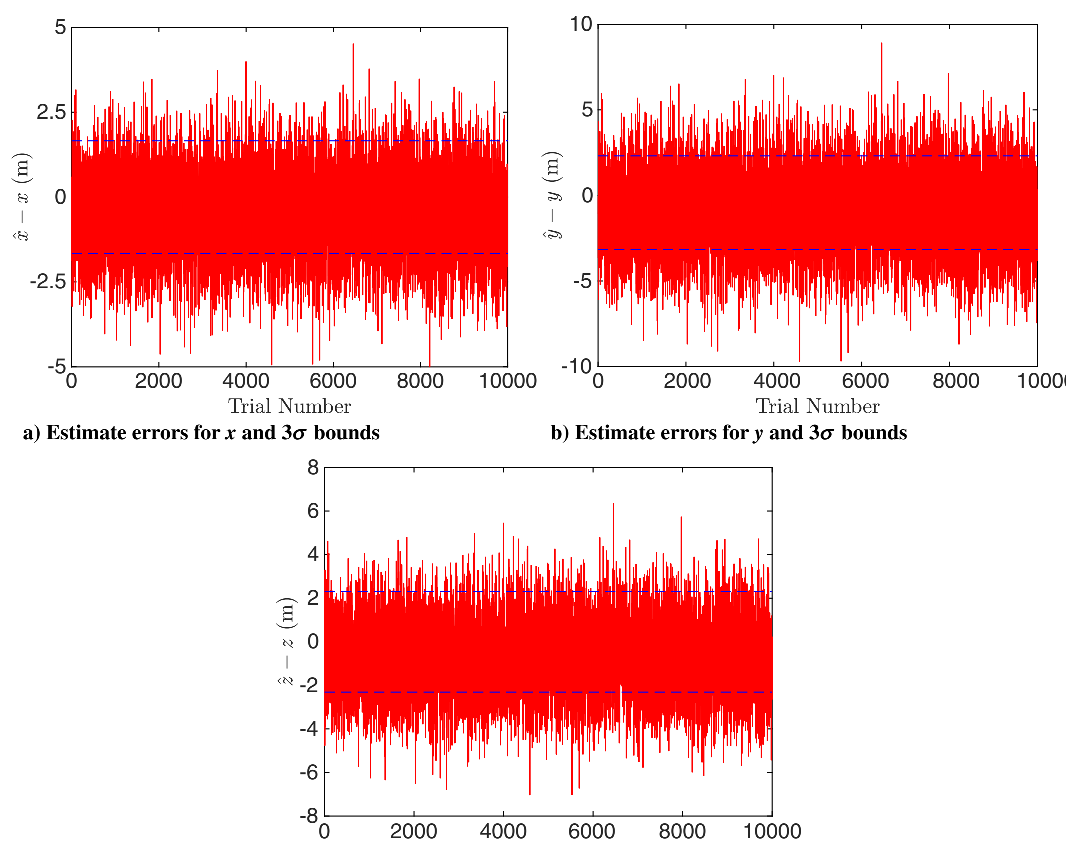
a) Estimate errors for x and 3σ bounds

b) Estimate errors for y and 3σ bounds

c) Estimate errors for z and 3σ bounds

d) Estimate errors for D_{21} and 3σ bounds

Fig. 2 Total least squares bearings-only estimation errors.



a) Estimate errors for x and 3σ bounds

b) Estimate errors for y and 3σ bounds

c) Estimate errors for z and 3σ bounds

Fig. 3 Standard least squares bearings-only estimation errors.

Thus, not only are the TLS estimates more consistent than the standard least squares estimates, but also they are less biased.

VIII. Conclusions

The maximum likelihood analysis presented here applies to the most general case for the total least squares problem where all the errors are correlated. It is, however, assumed here that the associated bias in the estimate is small enough to be ignored in the analysis presented here. With this assumption the Fisher information matrix has been derived, which is used to provide the Cramér–Rao lower bound. Although achieving this bound cannot be easily shown in theory, because no closed-form solution for the solution of the most general total least squares problem exists, the usual maximum likelihood properties can be invoked. These include the invariance principle, asymptotic consistency, and the asymptotic efficient property. Furthermore, the derived error-covariance expressions are useful to analyze the quality of the estimates. Simulation results using three-dimensional bearings-only localization showed that the derived Fisher information matrix can be used to provide accurate bounds for the estimate errors.

Appendix A: Bias in Total Least Squares Estimation

The bias $\boldsymbol{\beta}^a$ in the TLS estimate of $\boldsymbol{a} = [\boldsymbol{x}^T \quad \mathrm{vec}^T(H)]^T$ is zero to first order in measurement noise and model errors. The second-order bias can be derived using the approach shown in [21], which is given by

$$\boldsymbol{\beta}^a = \begin{bmatrix} \boldsymbol{\beta}_x \\ \boldsymbol{\beta}_h \end{bmatrix} \triangleq E\{\hat{\boldsymbol{a}}\} - \boldsymbol{a} = -P^a \begin{bmatrix} H^T \bar{\mathcal{R}}_{yy} \\ \bar{\mathcal{R}}_{hy} + \boldsymbol{x} \otimes \bar{\mathcal{R}}_{yy} \end{bmatrix} \boldsymbol{m} \tag{A1}$$

where the matrices are given by

$$P^a = \begin{bmatrix} P_{xx} & P_{xh} \\ P_{xh}^T & P_{hh} \end{bmatrix} \triangleq (F^a)^{-1} \tag{A2a}$$

$$\bar{\mathcal{R}}_{hy} = -\bar{R}_{hh}^{-1} \bar{R}_{hy} (\bar{R}_{yy} - \bar{R}_{hy}^T \bar{R}_{hh}^{-1} \bar{R}_{hy})^{-1} \tag{A2b}$$

$$\bar{\mathcal{R}}_{yy} = (\bar{R}_{yy} - \bar{R}_{hy}^T \bar{R}_{hh}^{-1} \bar{R}_{hy})^{-1} \tag{A2c}$$

In the matrix P^a, $P_{xx} \triangleq P$ and P_{xh} and P_{hh} are sub-blocks of P_{xD} and $\mathrm{cov}\{\mathrm{vec}(\hat{D})\}$. The components m_i of the $m \times 1$ column vector $\boldsymbol{m}$ are given by

$$m_i = \sum_{j=1}^{n} P_{H_{ij}, x_j} \tag{A3}$$

where P_{H_{ij},x_j} denotes the $[n+i+(j-1)m, \ j]$ or $[j, \ n+i+(j-1)m]$ element of P^a associated with H_{ij} and x_j. It is the covariance between H_{ij} and x_j (to first order). Another way to compute the ith element of $\boldsymbol{m}$ is to first form the following matrix:

$$\mathcal{M} = P_{xh} K_{m,n} \equiv [\mathcal{M}_1 \quad \mathcal{M}_2 \quad \cdots \quad \mathcal{M}_m] \tag{A4}$$

where $\mathcal{M}_i$ is the ith $n \times n$ submatrix of $\mathcal{M}$. The ith element of $\boldsymbol{m}$ is then given by the trace of $\mathcal{M}_i$.

It follows from Eq. (A1) that the bias in the estimate of $\boldsymbol{x}$ is

$$\boldsymbol{\beta}_x = -[PH^T \bar{\mathcal{R}}_{yy} + P_{xh}(\bar{\mathcal{R}}_{hy} + \boldsymbol{x} \otimes \bar{\mathcal{R}}_{yy})]\boldsymbol{m} \tag{A5}$$

The estimation bias is small if the signal-to-noise/error ratio is large and the problem is well conditioned. In these cases, the matrix P is a good approximation of the covariance matrix. If both H and $\boldsymbol{x}$ reduce to scalars and $\bar{R}_{hy} = 0$, the total least squares estimate $\hat{x} = \tilde{y}/\tilde{H}$. It can be shown that (to second order)

$$P = \frac{\bar{R}_{yy}}{H^2} + \frac{x^2 \bar{R}_{hh}}{H^2} \tag{A6a}$$

$$P_{xh} = -\frac{x}{H} \bar{R}_{hh} \tag{A6b}$$

Substituting Eqs. (A6) into Eq. (A5) gives

$$\beta_x = \frac{\bar{R}_{hh}}{H^2} x \tag{A7}$$

Clearly, the bias is a small portion of the true value of x when the noise variance is small and the true value of H is large.

Appendix B: Iteration Solution for the Total Least Squares Solution

The iteration process described in [22] is used for the TLS solution, which is reviewed here.

1) Begin with the initial estimate using least squares: $\hat{\boldsymbol{x}}^0 = (\tilde{H}^T \bar{R}_{yy}^{-1} \tilde{H})^{-1} \tilde{H}^T \bar{R}_{yy}^{-1} \tilde{\boldsymbol{y}}$.

2) Compute $\hat{\boldsymbol{z}}^{i-1} = [(\hat{\boldsymbol{x}}^{i-1})^T - 1]^T$ with the current estimate, and then compute the following quantities:

$$\begin{aligned}
Q_{\hat{\lambda}}^i &= [(\hat{\boldsymbol{z}}^{i-1})^T \otimes I_{m\times m}] \bar{R} (\hat{\boldsymbol{z}}^{i-1} \otimes I_{m\times m}) \\
\hat{E}_H^i &= \mathrm{Ivec}\{-[\bar{R}_{hh} \quad \bar{R}_{hy}](\hat{\boldsymbol{z}}^{i-1} \otimes I_{m\times m})(Q_{\hat{\lambda}}^i)^{-1}(\tilde{\boldsymbol{y}} - \tilde{H}\hat{\boldsymbol{x}}^{i-1})\} \\
N^i &= (\tilde{H} - \hat{E}_H^i)^T (Q_{\hat{\lambda}}^i)^{-1} (\tilde{H} - \hat{E}_H^i) \\
\boldsymbol{n}^i &= (\tilde{H} - \hat{E}_H^i)^T (Q_{\hat{\lambda}}^i)^{-1} (\tilde{\boldsymbol{y}} - \hat{E}_H^i \hat{\boldsymbol{x}}^{i-1}) \\
\hat{\boldsymbol{x}}^i &= (N^i)^{-1} \boldsymbol{n}^i
\end{aligned}$$

where Ivec denotes the "inverse" vec operator, which forms an $m \times n$ matrix for $\hat{E}_H^i$.

3) End when $\|\hat{\boldsymbol{x}}^{i-1} - \hat{\boldsymbol{x}}^i\| < \epsilon$ for a small threshold ϵ, $\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}^i$.

This iterative algorithm provides a linear convergence rate based on a Gauss-Newton approach.

Acknowledgments

The authors wish to thank Christopher Nebelecky from CUBRC and Puneet Singla from Pennsylvania State University for many helpful comments and suggestions.

References

[1] Golub, G. H., and Van Loan, C. F., "An Analysis of the Total Least Squares Problem," *SIAM Journal on Numerical Analysis*, Vol. 17, No. 6, 1980, pp. 883–893. doi:10.1137/0717073

[2] 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, Nos. 1–2, 2005, pp. 254–267. doi:10.1016/j.aca.2004.12.059

[3] Anderson, T. W., "Estimation of Linear Functional Relationships: Approximate Distributions and Connections with Simultaneous Equations in Econometrics," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 38, No. 1, 1976, pp. 1–20. doi:10.1111/rssb.1976.38.issue-1

[4] Al-Sharadqah, A., and Chernov, N., "Error Analysis for Circle Fitting Algorithms," *Electronic Journal of Statistics*, Vol. 3, 2009, pp. 886–911. doi:10.1214/09-EJS419

[5] Crassidis, J. L., and Junkins, J. L., *Optimal Estimation of Dynamic Systems*, 2nd ed., Chapman & Hall/CRC, Boca Raton, FL, 2012, Chaps. 1 and 2.

[6] 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 Publ., Piscataway, NJ, 2000, pp. 523–530. doi:10.1109/CVPR.2000.855864

[7] Wiesel, A., Eldar, Y. C., and Yeredor, A., "Linear Regression with Gaussian Model Uncertainty: Algorithms and Bounds," *IEEE Transactions on Signal Processing*, Vol. 56, No. 6, 2008, pp. 2194–2205. doi:10.1109/TSP.2007.914323

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

[9] Beck, A., and Eldar, Y. C., "Structured Total Maximum Likelihood: An Alternative to Structured Total Least Squares," *SIAM Journal on Matrix Analysis and Applications*, Vol. 31, No. 5, 2010, pp. 2623–2649. doi:10.1137/090756338

[10] Crassidis, J. L., and Cheng, Y., "Error-Covariance Analysis of the Total Least-Squares Problem," *Journal of Guidance, Control, and Dynamics*, Vol. 37, No. 4, 2014, pp. 1053–1063. doi:10.2514/1.62959

[11] Markovsky, I., and Van Huffel, S., "Overview of Total Least-Squares Methods," *Signal Processing*, Vol. 87, No. 10, 2007, pp. 2283–2302. doi:10.1016/j.sigpro.2007.04.004

[12] Tong, X., Jin, Y., Zhang, S., and Li, L., "Bias-Corrected Weighted Total Least-Squares Adjustment of Condition Equations," *Journal of Surveying Engineering*, Vol. 141, No. 2, 2015, Paper 04014013. doi:10.1061/(ASCE)SU.1943-5428.0000140

[13] Eldar, Y. C., "Minimum Variance in Biased Estimation: Bounds and Asymptotically Optimal Estimators," *IEEE Transactions on Signal Processing*, Vol. 52, No. 7, 2004, pp. 1915–1930. doi:10.1109/TSP.2004.828929

[14] Cramér, H., *Mathematical Methods of Statistics*, Princeton Univ. Press, Princeton, NJ, 1946, Chaps. 32–34.

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

[16] Fang, X., "Weighted Total Least Squares: Necessary and Sufficient Conditions, Fixed and Random Parameters," *Journal of Geodesy*, Vol. 87, No. 8, 2013, pp. 733–749. doi:10.1007/s00190-013-0643-2

[17] Magnus, J. R., and Neudecker, H., "The Commutation Matrix: Some Properties and Applications," *The Annals of Statistics*, Vol. 7, No. 2, 1979, pp. 381–394. doi:10.1214/aos/1176344621

[18] Xu, P., Liu, J., and Shi, C., "Total Least Squares Adjustment in Partial Errors-in-Variables Models: Algorithm and Statistical Analysis," *Journal of Geodesy*, Vol. 86, No. 8, 2012, pp. 661–675. doi:10.1007/s00190-012-0552-9

[19] Grzymisch, J., and Fichte, W., "Observability Criteria and Unobservable Maneuvers for In-Orbit Bearings-Only Navigation," *Journal of Guidance, Control, and Dynamics*, Vol. 37, No. 4, 2014, pp. 1250–1259. doi:10.2514/1.62476

[20] Jazaeri, S., Schaffrin, B., and Snow, K., "On Weigthed Total Least-Squares Adjustment with Multiple Constraints and Singular Dispersion Matrices," *zfv—Zeitschrift für Geodäsie, Geoinformation und Landmanagement*, Vol. 139, No. 4, 2014, pp. 229–240. doi:10.12902/zfv-0017-2014

[21] Box, M. J., "Bias in Nonlinear Estimation," *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 33, No. 2, 1971, pp. 171–190. doi:10.1111/rssb.1971.33.issue-2

[22] Fang, X., "A Structured and Constrained Total Least-Squares Solution with Cross-Covariances," *Studia Geophysica et Geodaetica*, Vol. 58, No. 1, 2014, pp. 1–16. doi:10.1007/s11200-012-0671-z