Total Least Squares for Optimal Pose Estimation

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

Yang Cheng†
Mississippi State University, Mississippi State, MS, 39762

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

Matthias Schmid§
Clemson University, Clemson, SC, 29634

This work provides a theoretical framework for the pose estimation problem using total least squares for vector observations from landmark features. First, the optimization framework is formulated with observation vectors extracted from point cloud features. Then, error-covariance expressions are derived. The attitude and position solutions obtained via the derived optimization framework are proven to reach the bounds defined by the Cramér-Rao lower bound under the small-angle approximation of attitude errors. The measurement data for the simulation of this problem is provided through a series of vector observation scans, and a fully populated observation noise-covariance matrix is assumed as the weight in the cost function to cover the most general case of the sensor uncertainty. Here, previous derivations are expanded for the pose estimation problem to include more generic correlations in the errors than previous cases involving an isotropic noise assumption. The proposed solution is simulated in a Monte-Carlo framework to validate the error-covariance analysis.

I. Introduction

In localization problems, such as simultaneous localization and mapping (SLAM), the part that determines the vehicle’s orientation can be stated as an attitude estimation problem. The attitude estimation can be categorized in two ways. The first involves determining the orientation of the body frame with respect to a reference frame by taking multiple (at least two non-collinear) measurements in the form of vector observations from nearby features or the vehicle itself. The second involves filtering this observation information with models outlining the motion of the vehicle in a state estimation framework [1]. The advantage of the first is that an attitude solution is provided at each time-point in a deterministic manner without the transients or even divergence associated with filtering approaches. The advantage of the second is that attitude estimates can still be provided, even when one vector is available, and filtered estimates are given as part of the solution. One of the early derivations of the attitude determination for the first involved approaches that solve Wahba’s problem [2], some of which have closed-form solutions [3]. Wahba’s problem also has connections to the attitude and position determination problem [4], which is called pose estimation. This pose estimation problem, which involves both attitude and position estimation, is addressed in this paper.

Pose estimation based on imagery data has various applications for human-computer interaction in virtual environments and augmented reality systems, in which monocular or stereo cameras are used for hand pose estimation [5] with several degrees of freedom. Image-based relative pose determination is categorized in two classes of general algorithms [6]: model-based and non-model-based techniques. Model-based solutions use an initial known model of the object that includes dimensions, shape or texture, and matches the visual features to determine the model’s unknown parameters. Template matching [7] and contour tracking [8] are examples for this class of pose estimation solutions. Non-model-based solutions do not employ a model of the object but rather determine the pose from a sequence of
images such as structure from motion problem [9]. Deep-learning approaches are also another series of solutions for relative pose estimation from image data sets [10–12].

There is another category of pose estimation solutions that rely on Lidar data sets, which use point cloud data structures [13] to illustrate the objects in the environment. Reference [14] provides a solution for Lidar-based pose estimation for non-cooperative objects. The iterative closest point (ICP) approach [15] is one of the well-known approaches to match the data structures from different Lidar frames, which is an essential problem for the alignment of 3D point cloud data from different scans. Segmentation of the object of interest from its surroundings is a crucial task in point cloud pre-processing [16]. Connected components [17] and super-point graphs [18] are two examples of such methods. Feature extraction is another primary step to provide the necessary information from the point cloud segmented objects. A class of stable algorithms exist for this purpose, such as SIFT [19] and SURF [20]. Feature matching and correspondence of different features in several Lidar scans is another issue associated with SLAM applications. One solution for the SLAM problem involves matching the histograms extracted from the features [21]. Reference [22] utilizes a point as well as the local properties of its neighborhood, e.g. the curvature and normal direction, to align point clouds.

It is assumed in this paper that the segmentation, feature extraction, and feature matching steps for data sets are already solved. The focus of the present work is to solve the pose estimation from matched features as vector observations. The pose estimation problem from observation vectors is well-studied [23]. However, little work exists on efficiently determining the associated error-covariance expressions, i.e. ones that achieve the Cramér-Rao lower bound [24]. Reference [25] derives efficient error-covariance expressions, but assumes isotropic matrices for the vector-measurement covariance. In this paper, the problem of the most general case of sensor uncertainty that includes fully-populated measurement covariance matrices, which also account for the correlations between the measurement vectors, is solved. Furthermore, error-covariance expressions are derived that achieve the Cramér-Rao lower bound, and reduce to the expressions shown in [25] when the measurement covariance matrices are isotropic. An example that can be benefited from the analysis shown in this paper involves medical image databases [26].

II. Problem Statement

The pose estimation problem is described as finding the pose of a sensor attached to a vehicle with respect to another sensor or a reference frame. The pose itself consists of two components: 1) a translation vector that connects the center of the two frames, and 2) an attitude matrix for the relative orientation of the unit vectors of the coordinate systems. A geometric model relating the pose of the reference frame to the body frame is shown in Figure 1. The associated geometric constraint acts as a measurement model. The vectors \( \mathbf{r}_i \) and \( \mathbf{b}_i \) are the coordinates of a point of interest, i.e. an observation feature of a landmark in the environment, with respect to the reference and body frames, respectively, and \( \mathbf{p} \) refers to the translation vector between the two coordinates.

The estimation approach provides an estimate of the attitude as well as the translation vector based on the measurement
information, denoted by \( \tilde{r}_i \) and \( \tilde{b}_i \). The estimate error, which comprises the difference between the transformed version of \( \tilde{r}_i \) and \( \tilde{b}_i \) is given by

\[
e_i(\hat{\boldsymbol{A}}, \hat{\boldsymbol{p}}) = \tilde{b}_i - \hat{\boldsymbol{A}}\tilde{r}_i + \hat{\boldsymbol{p}}
\]  

(1)

where \( \hat{\boldsymbol{A}} \) is the estimated attitude matrix, and \( \hat{\boldsymbol{p}} \) is the estimated translation vector. In case of perfect measurements, the error samples \( e_i \) are all zero, and the problem can be solved with the measurement model of the form

\[
b_i = \hat{\boldsymbol{A}}r_i - \hat{\boldsymbol{p}}
\]  

(2)

where \( b_i \) and \( r_i \) denote the true values of the observation vectors. But in an actual applications, these errors are not zero, which leads to an optimization problem derived from a constrained maximum likelihood approach [24], given by

\[
\min_{\hat{\boldsymbol{A}}, \hat{\boldsymbol{p}}} J = \frac{1}{2} \sum_{i=1}^{n} (\tilde{b}_i - \hat{\boldsymbol{A}}\tilde{r}_i + \hat{\boldsymbol{p}})^T R_i^{-1} (\tilde{b}_i - \hat{\boldsymbol{A}}\tilde{r}_i + \hat{\boldsymbol{p}})
\]

subject to: \( \hat{\boldsymbol{A}}^T \hat{\boldsymbol{A}} = I_{3x3}, \det(\hat{\boldsymbol{A}}) = 1 \)

(3)

where \( I_{3x3} \) is an identity matrix, and \( R_i \) is the measurement covariance matrix that accounts for errors on both \( \tilde{b}_i \) and \( \tilde{r}_i \), as well as correlations that exist between them. The determinant condition is required so that \( \hat{\boldsymbol{A}} \) is a proper attitude matrix. The observation errors are defined as

\[
\Delta b_i = \tilde{b}_i - b_i
\]

\[
\Delta r_i = \tilde{r}_i - r_i
\]

(4a)

(4b)

It is assumed that zero-mean Gaussian measurement errors exist, with

\[
\mathcal{E}\{\Delta b_i \Delta b_i^T\} = R_{b_i}
\]

\[
\mathcal{E}\{\Delta r_i \Delta r_i^T\} = R_{r_i}
\]

\[
\mathcal{E}\{\Delta r_i \Delta b_i^T\} = \mathcal{E}\{\Delta b_i \Delta r_i^T\}^T = R_{r b_i}
\]

\[
R_i = \mathcal{E}\left\{\begin{bmatrix} \Delta r_i \\ \Delta b_i \end{bmatrix}\begin{bmatrix} \Delta r_i^T \\ \Delta b_i^T \end{bmatrix}\right\} = \begin{bmatrix} R_{r_i} & R_{r b_i} \\ R_{r b_i}^T & R_{b_i} \end{bmatrix}
\]

(5a)

(5b)

(5c)

(5d)

where \( \mathcal{E} \) denotes expectation. The optimization problem in Eq. (3) can be shown to be related to a total least squares (TLS) problem [25].

This paper solves the pose estimation problem in Eq. (3) with the fully populated noise covariance matrices \( R_{r_i}, R_{b_i} \) and \( R_{r b_i} \) using TLS. Note that although there are nine components in the attitude matrix \( A \), only three of them are independent, so the attitude estimation solution can be accomplished with a minimum of three parameters, which can be Euler angles or any other minimal attitude parameterization [27]. Regardless of the attitude parameterization used in the optimization process, the estimated attitude can be related to the true attitude using an attitude error-vector involving small roll, pitch and yaw angles, as will be seen in section II.C.

The derivation begins with an introductory formulation of TLS in section II.A. A cost function is developed in section II.C based on the pose estimation problem from Eq. (3). The necessary conditions for efficient pose estimation are derived, and the cost function will be written in the attitude-only format for the sake of simplicity in the proceeding derivations. A linear approximation of the attitude error is then derived to approximate the cost function within second-order terms in the attitude errors. The efficient estimations for the attitude and translation vectors are used to derive the error-covariance of the estimate errors and residuals in section II.D which are beneficial for control purposes in a sense that the most accurate pose estimates are provided to the control logic to prevent extra effort. Finally, a numerical verification of the proposed TLS estimator is shown in the context of a Monte-Carlo simulation in section III.

A. Overview of Linear Least Squares and Total Least Squares

This section briefly introduces linear and total least squares, how they are related, and their differences. For a more in-depth review of the TLS, the reader is referred to [28–30]. Consider the measurement model of the form

\[
\tilde{y} = Hx + \Delta y
\]

(6)
where $H$ is an $m \times n$ deterministic matrix with no errors, $x$ is the $n \times 1$ vector of unknowns, $\tilde{y}$ is the $m \times 1$ measurement vector, and $Ay$ is the $m \times 1$ measurement error-vector. The least squares estimate of $x$ is given by solving the following problem:

$$
\min_{\hat{x}} \ J = \frac{1}{2} \Delta y^T \Delta y \\
\text{subject to: } \hat{y} = H\hat{x}
$$

(7)

where the number of measurement samples stacked vertically in $\tilde{y}$ should be more than the number of unknowns for the problem to be observable. The main underlying assumption in the statistical analysis of least squares is that $\tilde{y}$ has a Gaussian distribution with the conditional likelihood function given by

$$
p(\tilde{y}|x) = \frac{1}{(2\pi)^{\frac{m}{2}} |\text{det}(R_{yy})|^\frac{1}{2}} \exp \left(-\frac{1}{2} (\tilde{y} - Hx)^T R_{yy}^{-1} (\tilde{y} - Hx) \right)
$$

(8)

where the distribution mean is denoted by $Hx$ and the covariance is $R_{yy}$. Because of the properties of the exponential function, maximizing the likelihood function is equivalent to minimizing the negative of the log-likelihood. The mean and error-covariance of the estimate are given by

$$
E\{\hat{x}\} = \left( H^T R_{yy}^{-1} H \right)^{-1} H^T R_{yy}^{-1} \tilde{y} \\
cov\{\hat{x}\} = \left( H^T R_{yy}^{-1} H \right)^{-1}
$$

(9a, 9b)

which shows that the least squares estimate is unbiased.

As stated previously, the design matrix $H$ in the least squares measurement model in Eq. (6) has no errors. If this underlying assumption does not exist anymore, which happens in many applications, as will be seen in the SLAM problem in section II.C, then another formulation must be used to consider the errors in the design matrix, which leads to the TLS problem, with parameters defined by

$$
\tilde{y} = y + Ay \\
y = Hx \\
\tilde{H} = H + \Delta H
$$

(10a, 10b, 10c)

where $\Delta H$ shows the errors in the design matrix. Consider the following augmented matrix:

$$
\tilde{D} = \begin{bmatrix} H & y \end{bmatrix}
$$

(11)

The conditional likelihood function of the TLS problem is defined by

$$
p(\tilde{D}|D) = \frac{1}{(2\pi)^{\frac{m}{2}} |\text{det}(R)|^\frac{1}{2}} \exp \left(-\frac{1}{2} \text{vec}(\tilde{D} - D)^T R^{-1} \text{vec}(\tilde{D} - D) \right)
$$

(12)

where vec operator stacks all columns of a matrix in a single column. The maximum likelihood approach for this cost function leads to the minimization of the log-likelihood function as

$$
J(\tilde{D}) = \frac{1}{2} \text{vec}(\tilde{D} - D)^T R^{-1} \text{vec}(\tilde{D} - D) \\
\text{subject to: } \tilde{D} \hat{z} = 0
$$

(13)

where $\hat{z} = [\tilde{y}^T - 1]^T$. A unique solution for this problem can be obtained if $\text{rank}(D) = n$. Also, $R$ is the covariance matrix that accounts for the errors in both $\tilde{y}$ and $\tilde{H}$. Although the TLS solution is known to be biased, the TLS problem is proven to reach the Cramér-Rao lower bound (CRLB) for the estimate error-covariance to within first-order error-terms, and therefore is an efficient estimator. Closed-form solutions for the TLS problem are possible only when $R$ is an isotropic matrix.
B. Pose Estimation Sensor Model

The relation between the true vectors $b_i$ and $r_i$ is given by

$$b_i = A r_i - p$$

$$= \begin{bmatrix} r_i^T & 0_{1\times 3} & 0_{1\times 3} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} - p$$

$$= H_i x - p$$

$\equiv y_i$

in which Eq. (108) has been used. The matrix $H_i$ is the individual sensor model design matrix and $a_i$, $i = 1, 2, 3$, are the columns of the attitude matrix $A$. However, the perfect measurement model is not realistic because of noise in the design matrix as well as the observation vectors of Eq. (14). Therefore, in the actual version of the sensor model in Eq. (14), the following relation is used:

$$\tilde{b}_i - \Delta b_i = A (\tilde{r}_i - \Delta r_i) - p$$

$$= (\tilde{H}_i - \Delta H_i) x - p$$

$\equiv \tilde{y}_i - \Delta y_i$

where the design matrix $\tilde{H}_i$ and the observation vector $\tilde{y}_i$ have the errors of $\Delta H_i$ and $\Delta y_i$, respectively. Because the model is linear in terms of the unknowns $x$ and the translation vector $p$, then the problem can be posed using a TLS formulation with the constraint

$$\tilde{b}_i = A \tilde{r}_i - \tilde{p}$$

which is equivalent to

$$\tilde{D}_i \tilde{z} - \tilde{p} = 0$$

where $\tilde{z} = [z^T - 1]^T$ and $\tilde{D}_i = [\tilde{H}_i \ \tilde{y}_i]$.

C. Total Least Squares Derivation for Pose Determination

The TLS cost function is given by

$$J(\tilde{D}_i) = \frac{1}{2} \sum_{i=1}^{n} \text{vec}(\tilde{D}_i - \tilde{D}_i)^T R_{D_i}^{-1} \text{vec}(\tilde{D}_i - \tilde{D}_i)$$

(18)

where $R_{D_i}$ is the covariance of vec($\tilde{D}_i$). The augmented cost function that includes the linear constraint in Eq. (17) is given by

$$J_a(\tilde{D}_i, \Lambda_i) = \frac{1}{2} \sum_{i=1}^{n} \text{vec}(\tilde{D}_i - \tilde{D}_i)^T R_{D_i}^{-1} \text{vec}(\tilde{D}_i - \tilde{D}_i) + \Lambda_i^T (\tilde{D}_i \tilde{z} - \tilde{p})$$

(19)

where $n$ is the number of features in each sensor scan. Taking the partial derivative of the augmented cost function with respect to vec($\tilde{D}_i$) and utilizing Eq. (108) gives

$$\frac{\partial J_a}{\partial \text{vec}(\tilde{D}_i)} = -R_{D_i}^{-1} \text{vec}(\tilde{D}_i - \tilde{D}_i) + (\tilde{z}^T \otimes I_{3\times 3})^T \Lambda_i$$

(20)

where $\otimes$ is the Kronecker product [33]. From the necessary condition, this partial derivative should be a zero vector, and therefore

$$\text{vec}(\tilde{D}_i) = \text{vec}(\tilde{D}_i) - R_{D_i}(\tilde{z}^T \otimes I_{3\times 3})^T \Lambda_i$$

(21)

Using Eq. (108), the constraint in Eq. (17) can be written as

$$(\tilde{z}^T \otimes I_{3\times 3}) \text{vec}(\tilde{D}_i) - \tilde{p} = 0$$

(22)
Substituting Eq. (21) into the constraint gives

\[
(\hat{z}^T \otimes I_{3x3}) \left[ \text{vec}(\hat{D}_i) - R_{D_i}(\hat{z}^T \otimes I_{3x3})^T \lambda_i \right] - \hat{p} = 0
\]  

(23)

Solving for \(\lambda_i\) leads to

\[
\lambda_i = Q_{\lambda_i}^{-1} \left[ (\hat{z}^T \otimes I_{3x3}) \text{vec}(\hat{D}_i) - \hat{p} \right]
\]

(24)
in which \(Q_{\lambda_i} = (\hat{z}^T \otimes I_{3x3}) R_{D_i}(\hat{z}^T \otimes I_{3x3})^T\). Note that in this paper, \(Q_{\lambda_i}\) is considered to be a positive-definite matrix, though it might be singular in some sensor models, such as the Quest Measurement Model \([34]\). This problem can be solved using a similar approach to the eigenvalue decomposition in \([35]\). Using the necessary condition for \(\hat{D}_i\) from Eq. (21), and substituting the Lagrangian multiplier \(\lambda_i\) from Eq. (24) gives

\[
\text{vec}(\hat{D}_i) - \text{vec}(\hat{D}_i) = R_{D_i}(\hat{z}^T \otimes I_{3x3})^T Q_{\lambda_i}^{-1} \left[ (\hat{z}^T \otimes I_{3x3}) \text{vec}(\hat{D}_i) - \hat{p} \right]
\]

\[
= R_{D_i}(\hat{z}^T \otimes I_{3x3})^T Q_{\lambda_i}^{-1} (\hat{D}_i \hat{z} - \hat{p})
\]

\[
= -R_{D_i}(\hat{z}^T \otimes I_{3x3})^T Q_{\lambda_i}^{-1} (\hat{p} - \hat{D}_i \hat{z})
\]

(25)

Substituting Eq. (25) into the original cost function in Eq. (18) yields

\[
J(\hat{z}, \hat{p}) = \frac{1}{2} \sum_{i=1}^{n} (\hat{p} - \hat{D}_i \hat{z})^T Q_{\lambda_i}^{-1} (\hat{p} - \hat{D}_i \hat{z})
\]

(26)

This results in a formulation, satisfying the constraint in Eq. (17) in the cost function in Eq. (18). Then in terms of the observation vectors \(\bar{b}_i\) and \(\bar{r}_i\), and from \(\hat{p} - \hat{D}_i \hat{z} = \bar{b}_i - \hat{A} \bar{r}_i + \hat{p}\), the following expression is given:

\[
J(\hat{A}, \hat{p}) = \frac{1}{2} \sum_{i=1}^{n} (\bar{b}_i - \hat{A} \bar{r}_i + \hat{p})^T Q_{\lambda_i}^{-1} (\bar{b}_i - \hat{A} \bar{r}_i + \hat{p})
\]

(27)

For simplicity in the proceeding derivations, the cost function can be written only in terms of the attitude matrix, which is known as the attitude-only cost function. For this purpose, \(\hat{p}\) needs to be eliminated from the cost function in Eq. (27). The necessary condition for the translation vector \(\hat{p}\) results in

\[
\hat{p} = - \left( \sum_{i=1}^{n} Q_{\lambda_i}^{-1} \right)^{-1} \left[ \sum_{i=1}^{n} Q_{\lambda_i}^{-1} (\bar{b}_i - \hat{A} \bar{r}_i) \right]
\]

(28)

Substituting the optimal value of \(\hat{p}\) into Eq. (27), the attitude-only cost function becomes

\[
J(\hat{A}) = \frac{1}{2} \sum_{i=1}^{n} (\bar{b}_i - \hat{A} \bar{r}_i)^T Q_{\lambda_i}^{-1} (\bar{b}_i - \hat{A} \bar{r}_i) - \frac{1}{2} \left[ \sum_{i=1}^{n} Q_{\lambda_i}^{-1} (\bar{b}_i - \hat{A} \bar{r}_i) \right]^T \tilde{Q}_{\lambda} \left[ \sum_{i=1}^{n} Q_{\lambda_i}^{-1} (\bar{b}_i - \hat{A} \bar{r}_i) \right]
\]

(29)

in which

\[
\tilde{Q}_{\lambda} = \left( \sum_{i=1}^{n} Q_{\lambda_i}^{-1} \right)^{-1}
\]

(31)

are the observation vectors with their corresponding covariance shown in Eq. (30a), (30b) and (30c). Also it is proven in \([25]\) that the weight matrix \(Q_{\lambda_i}\) can be derived as a function of the attitude matrix and the observation noise covariance matrices as

\[
Q_{\lambda_i} = \hat{A} R_{r_i} \hat{A}^T - \hat{A} R_{r_{hi}} - R_{r_{hi}}^T \hat{A}^T + R_{hi}
\]

(32)
D. Covariance Analysis of the Estimates and Residuals

The covariance expressions for the attitude as well as the translation and observation vector estimates are now derived. Note that the cost function can be written in terms of the attitude error \( \delta \alpha \), since only 3 independent components exist inside of the attitude matrix. The relation between the true and the estimated attitude matrix can be expressed as

\[
\hat{A} = \exp(-[\delta \alpha \times])A
\]

where \([. \times]\) denotes the cross product matrix of a vector \([36]\). Using a small angle assumption and a first-order approximation of the attitude error, the attitude estimate can be written as

\[
\hat{A} \approx (I_{3 \times 3} - [\delta \alpha \times])A
\]

The cost function needs be derived up to second-order in terms of attitude error \( \delta \alpha \), since covariance expressions for a first-order approximation of the unknown errors are sought. The derivation begins with the approximation of the error-terms inside the cost function of Eq. (29). The attitude approximation in Eq. (34), Eqs. (108) and (105) from the appendix, as well as Eqs. (30a) and (30b), are utilized for the formulation of the observation vectors, yielding

\[
\hat{b}_i - \hat{A}r_i = b_i + \Delta b_i - (I_{3 \times 3} - [\delta \alpha \times])A(r_i + \Delta r_i)
\]

\[
\approx b_i - Ar_i + \Delta b_i - A[Ar_i \times]\delta \alpha
\]

(35)

The following abbreviations are introduced for simplicity:

\[
\Delta a_i \equiv \Delta b_i - A\Delta r_i
\]

(36a)

\[
p = Ar_i - b_i
\]

(36b)

\[
\nu_i \equiv -\Delta a_i + p
\]

(36c)

\[
\mathcal{A}_i \equiv [Ar_i \times]
\]

(36d)

This allows for the reformulation of the error-term \( \hat{b}_i - \hat{A}r_i \) as

\[
\hat{b}_i - \hat{A}r_i \approx -\nu_i - \mathcal{A}_i \delta \alpha
\]

(37)

Note that \( Q_{\hat{A}_i} \) in Eq. (37) is also a function of the attitude estimate \( \hat{A} \) and subsequently of the attitude error \( \delta \alpha \). As the neighboring terms are already a function of the first-order attitude estimate, any other term besides the attitude-error dependent ones in \( Q_{\hat{A}_i} \) (terms that are independent of \( \delta \alpha \)) are not kept for the second-order approximation of the cost function. The matrix \( Q_{\hat{A}_i} \) is now written as

\[
Q_{\hat{A}_i} = \hat{A}R_{r_i}\hat{A}^T - \hat{A}R_{r_{b_i}} - R_{r_{b_i}}^T\hat{A}^T + R_{b_i}
= (I_{3 \times 3} - [\delta \alpha \times])AR_{r_i}A^T(I_{3 \times 3} - [\delta \alpha \times])^T - (I_{3 \times 3} - [\delta \alpha \times])AR_{r_{b_i}} - R_{r_{b_i}}^T(35)
\]

Further decomposition of \( Q_{\hat{A}_i} \) in terms of \( \delta \alpha \) then yields

\[
Q_{\hat{A}_i} = Q_{\hat{A}_i} + \delta Q_{\hat{A}_i} + \delta^2 Q_{\hat{A}_i}
\]

(39)

where

\[
Q_{\hat{A}_i} = AR_{r_i}A^T - AR_{r_{b_i}} - R_{r_{b_i}}^T A^T + R_{b_i}
\]

(40a)

\[
\delta Q_{\hat{A}_i}(\delta \alpha) = \mathcal{K}^{T}[\delta \alpha \times] + [\delta \alpha \times]^T \mathcal{K}^{T}
\]

(40b)

\[
\delta^2 Q_{\hat{A}_i} = [\delta \alpha \times]AR_{r_i}A^T [\delta \alpha \times]^T
\]

(40c)

with

\[
\mathcal{K}^{T} = AR_{r_i}A^T - R_{r_{b_i}}^T A^T
\]

(41)

The inverse of \( Q_{\hat{A}_i} \) is approximated by

\[
Q_{\hat{A}_i}^{-1} = Q_{\hat{A}_i}^{-1} - Q_{\hat{A}_i}^{-1} \delta Q_{\hat{A}_i} Q_{\hat{A}_i}^{-1} - Q_{\hat{A}_i}^{-1} \delta^2 Q_{\hat{A}_i} Q_{\hat{A}_i}^{-1}
\]

(42)
Therefore, the matrix $Q_{\hat{A}}^{-1}$ is the only portion of $Q_{\hat{A}}$ that is not a function of the attitude error. This portion will be used for building the second-order cost function. The second summation of the cost function in Eq. (29) contains $\hat{Q}_{A}$, which itself is a function of $\delta \alpha$. The portion of $\hat{Q}_{A}$ that is not dependent on the attitude error needs to be extracted. This way, the corresponding part of the cost function ignores higher-order terms. The expansion of $\hat{Q}_{A}$ is given as

$$
\hat{Q}_{A} = \left( \sum_{i=1}^{n} Q_{\hat{A}_{i}}^{-1} \right) + \left( \sum_{i=1}^{n} Q_{\hat{A}_{i}}^{-1} \right) \left[ \sum_{i=1}^{n} Q_{\hat{A}_{i}}^{-1} \left( \delta Q_{\hat{A}_{i}}^{-1} + \delta^2 Q_{\hat{A}_{i}}^{-1} \right) Q_{\hat{A}_{i}}^{-1} \right] \left( \sum_{i=1}^{n} Q_{\hat{A}_{i}}^{-1} \right)^{-1}
$$

The summations in the above equation can be abbreviated as

$$
S_{\hat{A}} = \left( \sum_{i=1}^{n} Q_{\hat{A}_{i}}^{-1} \right)^{-1}
$$

$$
\delta S_{\hat{A}}(\delta \alpha) = S_{\hat{A}} \left( \sum_{i=1}^{n} Q_{\hat{A}_{i}}^{-1} \delta Q_{\hat{A}_{i}}^{-1} Q_{\hat{A}_{i}}^{-1} \right) S_{\hat{A}}
$$

$$
\delta^2 S_{\hat{A}}(\delta \alpha^T \delta \alpha) = S_{\hat{A}} \left( \sum_{i=1}^{n} Q_{\hat{A}_{i}}^{-1} \delta^2 Q_{\hat{A}_{i}}^{-1} Q_{\hat{A}_{i}}^{-1} \right) S_{\hat{A}}
$$

Hence, $S_{\hat{A}}$ emerges as the only term that is not a function of the attitude error. Utilizing the first-order errors in Eq. (37), the matrix $Q_{\hat{A}_{i}}$ in Eq. (39), and $S_{\hat{A}}$ in Eq. (44a), the following approximation of the cost function is given:

$$
L(\delta \alpha, \delta \alpha^T \delta \alpha) = \frac{1}{2} \left[ L_{1}(\delta \alpha) + L_{2}(\delta \alpha^T \delta \alpha) \right]
$$

where the first- and second-order terms yield

$$
L_{1}(\delta \alpha) = \delta \alpha^T g
$$

$$
L_{2}(\delta \alpha^T \delta \alpha) = \delta \alpha^T \mathcal{H} \delta \alpha
$$

with

$$
g = \left( \sum_{i=1}^{n} \mathcal{A}_{i} Q_{\hat{A}_{i}}^{-1} \right) S_{\hat{A}} \left( \sum_{i=1}^{n} Q_{\hat{A}_{i}}^{-1} v_{i} \right) + \sum_{i=1}^{n} \mathcal{A}_{i}^T Q_{\hat{A}_{i}}^{-1} v_{i}
$$

$$
\mathcal{H} = \left( \sum_{i=1}^{n} \mathcal{A}_{i}^T Q_{\hat{A}_{i}}^{-1} \mathcal{A}_{i} \right) - \left( \sum_{i=1}^{n} Q_{\hat{A}_{i}}^{-1} \mathcal{A}_{i} \right)^T S_{\hat{A}} \left( \sum_{i=1}^{n} Q_{\hat{A}_{i}}^{-1} \mathcal{A}_{i} \right)
$$

From the necessary condition for the extremum of the cost function $L$ with respect to $\delta \alpha$, i.e.

$$
\frac{\partial L}{\partial \delta \alpha} = \frac{\partial}{\partial \delta \alpha} (L_{1} + L_{2}) = 0
$$

the vector estimate of the attitude-error error emanates as

$$
\delta \alpha = -\mathcal{H}^{-1} g
$$

Employing Eq. (36c) allows for expanding the above expression to

$$
\delta \alpha = -\mathcal{H}^{-1} \left[ \sum_{i=1}^{n} \mathcal{A}_{i} Q_{\hat{A}_{i}}^{-1} \right] S_{\hat{A}} \left[ \sum_{i=1}^{n} Q_{\hat{A}_{i}}^{-1} (p - \Delta a_{i}) \right] + \sum_{i=1}^{n} \mathcal{A}_{i}^T Q_{\hat{A}_{i}}^{-1} (p - \Delta a_{i})
$$

At the same time, the following relation holds:

$$
\left( \sum_{i=1}^{n} \mathcal{A}_{i} Q_{\hat{A}_{i}}^{-1} \right) S_{\hat{A}} \left( \sum_{i=1}^{n} Q_{\hat{A}_{i}}^{-1} p \right) + \sum_{i=1}^{n} \mathcal{A}_{i}^T Q_{\hat{A}_{i}}^{-1} p = 0
$$
Given that $\mathbf{A}_i^T = -\mathbf{A}_i$, as well as that $Q_{ij}^{-1}$, $S_i$ and $\mathbf{H}$ are symmetric, the terms containing the translation vector $\mathbf{p}$ cancel, thus yielding

$$
\delta \alpha = \mathbf{H}^{-1} \left[ \left( \sum_{i=1}^{n} \mathbf{A}_i Q_{ij}^{-1} \right) S_i \left( \sum_{i=1}^{n} Q_{ij}^{-1} \Delta \alpha_i \right) - \sum_{i=1}^{n} \mathbf{A}_i Q_{ij}^{-1} \Delta \alpha_i \right]
$$

This attitude-error now allows for the derivation of the error-covariance expressions. The estimation error-covariance of the attitude is defined as

$$
P_{\delta \alpha} = \mathcal{E} \{ \delta \alpha \Delta \alpha^T \}
$$

Further expansion of the individual terms leads to

$$
P_{\delta \alpha} = \mathbf{H}^{-1} \left[ -\sum_{i=1}^{n} (\mathbf{A}_i Q_{ij}^{-1} \mathcal{E}(\Delta \alpha_i \Delta \alpha_i^T) Q_{ij}^{-1} \mathbf{A}_i) \right.
\left. + \left( \sum_{i=1}^{n} \mathbf{A}_i Q_{ij}^{-1} \right) S_i \left( \sum_{i=1}^{n} Q_{ij}^{-1} \mathcal{E}(\Delta \alpha_i \Delta \alpha_i^T) Q_{ij}^{-1} \right) S_i \left( \sum_{i=1}^{n} Q_{ij}^{-1} \mathbf{A}_i^T \right) \right]
$$

Employing the fact that

$$
P_{\Delta \alpha_i} = \mathcal{E}(\Delta \alpha_i \Delta \alpha_i^T) = Q_{ji}
$$

$$
\mathcal{E}(\Delta \alpha_i \Delta \alpha_j^T) = 0_{3 \times 3}, j \neq i
$$

the attitude error-covariance can be rewritten as

$$
P_{\delta \alpha} = \mathbf{H}^{-1} \left[ -\sum_{i=1}^{n} (\mathbf{A}_i Q_{ij}^{-1} Q_{ji}^{-1} \mathbf{A}_i) \right.
\left. + \left( \sum_{i=1}^{n} \mathbf{A}_i Q_{ij}^{-1} \right) S_i \left( \sum_{i=1}^{n} Q_{ij}^{-1} Q_{ji} Q_{ji}^{-1} \right) S_i \left( \sum_{i=1}^{n} Q_{ij}^{-1} \mathbf{A}_i^T \right) \right]
$$

The last two contributions in the above sum cancel out each other, thus yielding

$$
P_{\delta \alpha} = \mathbf{H}^{-1} \left[ -\sum_{i=1}^{n} (\mathbf{A}_i Q_{ij}^{-1} \mathbf{A}_i) + \left( \sum_{i=1}^{n} \mathbf{A}_i Q_{ij}^{-1} \right) S_i \left( \sum_{i=1}^{n} Q_{ij}^{-1} \right) S_i \left( \sum_{i=1}^{n} Q_{ij}^{-1} \mathbf{A}_i^T \right) \right] \mathbf{H}^{-T}
$$

With Eqs. (44a) and (47b), the above expression further simplifies to

$$
P_{\delta \alpha} = \mathbf{H}^{-1}
$$
which verifies that the estimate error-covariance of the attitude error is equal to the Hessian of the cost function. Note that a more detailed discussion of this observation is provided later in the context of the Fisher information matrix (FIM) for the cost function in Eq. (29).

The estimation error $\delta p$ for the translation vector is now derived, which begins with

$$\hat{p} = p + \delta p$$  \hspace{1cm} (59)

Decomposition of Eq. (28) by utilizing Eq. (37) to separate the first-order terms in the attitude error yields

$$p + \delta p = -\left( \sum_{i=1}^{n} Q_{A_i}^{-1} \right)^{-1} \left[ \sum_{i=1}^{n} Q_{A_i}^{-1} (-v_i - A_i \delta \alpha) \right]$$  \hspace{1cm} (60)

Thus, the estimate-error $\delta p$ emerges as

$$\delta p = -S_A \left( \sum_{i=1}^{n} Q_{A_i}^{-1} (-v_i - A_i \delta \alpha + p) \right)$$  \hspace{1cm} (61)

From the definition of $v_i$ in Eq. (36c) gives

$$\delta p = -S_A \sum_{i=1}^{n} Q_{A_i}^{-1} (\Delta a_i - A_i \delta \alpha)$$  \hspace{1cm} (62)

The error-covariance of the translation vector within the first-order of estimation errors is given by

$$\text{cov}(\hat{p}) \equiv E\left\{ (\hat{p} - E(\hat{p}))(\hat{p} - E(\hat{p}))^T \right\} = E\{\delta p \delta p^T\}$$  \hspace{1cm} (63)

in which the fact that the estimation for the translation vector is unbiased within the first-order terms of error is used, and $E(\hat{p}) = p$. Then, the transnational error-covariance is computed as

$$\text{cov}(\hat{p}) = E\left\{ \left[ -S_A \sum_{i=1}^{n} Q_{A_i}^{-1} (\Delta a_i - A_i \delta \alpha) \right] \left[ -S_A \sum_{i=1}^{n} Q_{A_i}^{-1} (\Delta a_i - A_i \delta \alpha) \right]^T \right\}$$  \hspace{1cm} (64)

It is observed that the cross-covariance of the attitude errors and $\Delta a_i$ is required for the translation error-covariance, which is computed as

$$E\{\delta \alpha \Delta a_i^T\} = H^{-1}(A_i - \bar{A})$$  \hspace{1cm} (65)

where

$$\bar{A} = S_A \sum_{i=1}^{n} Q_{A_i}^{-1} A_i$$  \hspace{1cm} (66)

Using Eqs. (55a), (55b) and (65), and from the attitude estimate error-covariance in Eq. (58), the translation vector error-covariance is now given by

$$\text{cov}(\hat{p}) = S_A + \bar{A} P_{\delta \alpha} \bar{A}^T$$  \hspace{1cm} (67)

Because there are estimates for the observation vectors from the TLS formulation in Eq. (21), their associated covariance expressions can be derived. First an expression for their corresponding first-order residuals is found, and then this residual approximation is used to construct an analytical covariance formulation. Using Eq. (21) for the observation vectors and using the derivation in [25], it can be shown that the estimates of the observation vectors are

$$\hat{b}_i = \bar{b}_i + (R_{rb_i}^T \bar{A}^T - R_{bi}) Q_{A_i}^{-1} (\hat{b}_i - \hat{\Delta} \bar{r}_i + \hat{p})$$  \hspace{1cm} (68a)

$$\hat{r}_i = \bar{r}_i + (R_{ri}^T \bar{A}^T - R_{ri}) Q_{A_i}^{-1} (\hat{b}_i - \hat{\Delta} \bar{r}_i + \hat{p})$$  \hspace{1cm} (68b)
Define the estimate-error for the observation vectors:
\[
\delta b_i = \hat{b}_i - b_i
\]  
\[
\delta r_i = \hat{r}_i - r_i
\]  
(69a) (69b)
The residual errors using Eq. (30b) and (30a) are \( \delta b_i - \Delta b_i \) and \( \delta r_i - \Delta r_i \). Then deducting both sides of Eq. (68a) by \( b_i \) and Eq. (68b) by \( r_i \) leads to
\[
\delta b_i - \Delta b_i = (R_{r b_i}^T \hat{A}^T - R_{b_i}) Q_{\hat{A} i}^{-1} (\hat{b}_i - \hat{A} \hat{r}_i + \hat{p})
\]  
(70a)
\[
\delta r_i - \Delta r_i = (R_{r \hat{r}_i}^T - R_{r b_i}) Q_{\hat{A} i}^{-1} (\hat{b}_i - \hat{A} \hat{r}_i + \hat{p})
\]  
(70b)
Their corresponding first-order approximations are given by
\[
\delta b_i - \Delta b_i \approx C_i (\Delta a_i - G_i \delta f)
\]  
(71a)
\[
\delta r_i - \Delta r_i \approx D_i (\Delta a_i - G_i \delta f)
\]  
(71b)
where
\[
C_i = (R_{r b_i}^T A^T - R_{b_i}) Q_{\hat{A} i}^{-1}
\]  
(72a)
\[
D_i = (R_{r \hat{r}_i}^T - R_{r b_i}) Q_{\hat{A} i}^{-1}
\]  
(72b)
\[
G_i = \left[ \begin{array}{c} A_i \\ A_i - I_{3 \times 3} \end{array} \right]
\]  
(72c)
\[
\delta f = \left[ \begin{array}{c} \delta \alpha \\ \delta \bar{p} \end{array} \right]_{6 \times 1}
\]  
(72d)
Using
\[
P_{\Delta a_i} = Q_{\hat{A} i}
\]  
(73a)
\[
P_{\Delta a_i, \Delta a_i} = S_{\hat{A}}
\]  
(73b)
\[
P_{\Delta \delta a} = 0_{3 \times 3}
\]  
(73c)
and from Eq. (65), the covariances of the measurement residuals are given by
\[
\text{cov}(\hat{b}_i - b_i) = \mathcal{E}\{ (\delta b_i - \Delta b_i)(\delta b_i - \Delta b_i)^T \}
\]  
\[
= C_i (Q_{\hat{A} i} G_i P_t G_i^T) C_i^T
\]  
(74a)
\[
\text{cov}(\hat{r}_i - r_i) = \mathcal{E}\{ (\delta r_i - \Delta r_i)(\delta r_i - \Delta r_i)^T \}
\]  
\[
= D_i (Q_{\hat{A} i} G_i P_t G_i^T) D_i^T
\]  
(74b)
where
\[
P_t \equiv \text{cov}(\delta f) = \left( \sum_{i=1}^{n} G_i^T Q_{\hat{A} i}^{-1} G_i \right)^{-1}
\]  
(75)
Equations (71a) and (71b) can now be written as
\[
\delta b_i \approx \Delta b_i + C_i (\Delta a_i - G_i \delta f)
\]  
(76a)
\[
\delta r_i \approx \Delta r_i + D_i (\Delta a_i - G_i \delta f)
\]  
(76b)
Their corresponding estimate covariances are given by
\[
P_{\hat{b}_i} = \mathcal{E}\{ \delta b_i \delta b_i^T \}
\]  
\[
= \mathcal{E}\{ (\hat{b}_i - b_i)(\hat{b}_i - b_i)^T \}
\]  
\[
= R_{b_i} + \text{cov}(\hat{b}_i - \hat{b}_i)
\]  
\[
+ C_i R_{b_i} - C_i G_i P_t G_i^T Q_{\hat{A} i}^{-1} R_{b_i}
\]  
\[
+ \left( C_i R_{b_i} - C_i G_i P_t G_i^T Q_{\hat{A} i}^{-1} R_{b_i} \right)^T
\]  
(77)
\[ P_r = \mathcal{E}\{\delta r_i \delta r_i^T\} \]
\[ = \mathcal{E}\{(\hat{r}_i - r_i)(\hat{r}_i - r_i)^T\} \]
\[ = R_{r_i} + \text{cov}(\hat{r}_i - r_i) \]
\[ - D_iAR_{r_i} + D_iG_iP_fG_i^T Q^{-1}_{\lambda_i} AR_{r_i} \]
\[ + \left(-D_iAR_{r_i} + D_iG_iP_fG_i^T Q^{-1}_{\lambda_i} AR_{r_i}\right)^T \]

**E. Isotropic Noise Covariance**

The previous section provided analytical expressions for the covariance of estimates and residuals for the most generic case of observation noise covariance, which is a fully-populated symmetric positive-definite matrix as denoted in Eq. (5d). Now a particular case of previous covariance derivations is elaborated. In some sensors, there is a simplifying assumption on the distribution of noise having the same characteristics in the different space coordinates of \( x, y \) and \( z \). This presumption results in a particular form of the noise covariance called *isotropic*. For the pose estimation problem in this work, this isotropic covariance of observation errors is denoted by

\[ R_i = \begin{bmatrix} \sigma^2_{r_i} I_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & \sigma^2_{b_i} I_{3 \times 3} \end{bmatrix} \]  

(79)

in which \( \sigma_{r_i} \) and \( \sigma_{b_i} \) indicate the standard deviation of noise for the sensors in the reference and body frames, respectively. The cross-correlation between the reference and body sensor noise will be zero, as shown in the off-diagonal blocks. For the isotropic case, the weight matrix \( Q_{\lambda_i} \) in the cost function from Eq. (38) shrinks to

\[ Q_{\lambda_i} = \hat{A}R_{r_i}\hat{A}^T + R_{b_i} \]
\[ = \hat{A}\sigma^2_{r_i} I_{3 \times 3}\hat{A}^T + \sigma^2_{b_i} I_{3 \times 3} \]
\[ = \left(\sigma^2_{r_i} + \sigma^2_{b_i}\right) I_{3 \times 3} \]
\[ = \sigma^2 I_{3 \times 3} \]  

(80)

As a result, the second weight \( \bar{Q}_{\lambda} \) of the cost function in Eq. (31) becomes

\[ \bar{Q}_{\lambda} = \bar{\sigma}^{-2} I_{3 \times 3} \]  

(82a)

\[ \bar{\sigma} = \sqrt{\sum_{i=1}^{n} \sigma^{-2}_i} = \sqrt{\frac{1}{\sum_{i=1}^{n} \sigma^{-2}_i + \sigma^{-2}_{b_i}}} \]  

(82b)

Then the attitude-only cost function in Eq. (29) yields

\[ J(\hat{\lambda}) = \frac{1}{2} \left[ \sum_{i=1}^{n} \sigma^{-2}_i \| \vec{b}_i - \hat{\lambda}\vec{b}_i \| ^2 \right] - \frac{1}{2} \bar{\sigma}^{-2} \| \vec{b} - \hat{\lambda}\vec{b} \|^2 \]  

(83a)

\[ \vec{b} = \bar{\sigma}^{-2} \sum_{i=1}^{n} \sigma^{-2}_i \vec{b}_i \]  

(83b)

\[ \hat{\lambda} = \bar{\sigma}^{-2} \sum_{i=1}^{n} \sigma^{-2}_i \hat{\lambda}_i \]  

(83c)

It is to be noted that in the isotropic case the matrix \( Q_{\lambda} \) is not a function of attitude matrix. Therefore, the attitude error avoids several complications in the derivation. This simplicity comes with the cost of ignoring the possible cross-covariances in the reference and body frames and assuming the same statistical properties of noise along all space coordinates. The cost function in Eq. (83a) is already second-order in terms of the attitude matrix and there is no need for trimming the higher-order terms, which is another simplicity resulting from the isotropic assumption.
The optimal attitude from Eq. (52) yields
\[
\delta \alpha = \mathcal{H}^{-1} \left[ \sum_{j=1}^{n} \sigma_j^{-2} A_j \bar{\Delta} - \sum_{i=1}^{n} \sigma_i^{-2} A_i \bar{\Delta} \right]
\]
\[
= \mathcal{H}^{-1} \left[ \sum_{i=1}^{n} \sigma_i^{-2} \bar{\Delta} A_i - \sum_{i=1}^{n} \sigma_i^{-2} A_i \bar{\Delta} \right]
\]
\[
= \mathcal{H}^{-1} \left[ \sum_{i=1}^{n} \sigma_i^{-2} \left( \bar{\Delta} - A_i \right) \Delta_i \right]
\]
in which
\[
\mathcal{A}_i = [A r, \times]
\]
\[
\bar{\mathcal{A}} = \bar{\sigma}^{-2} \sum_{i=1}^{n} \sigma_i^{-2} \left[ A r, \times \right]
\]
and the Hessian matrix \( \mathcal{H} \) will be
\[
\mathcal{H} = \sum_{i=1}^{n} \sigma_i^{-2} A_i^T A_i - \left( \sum_{i=1}^{n} \sigma_i^{-2} A_i \right)^T \bar{\sigma}^{-2} \left( \sum_{i=1}^{n} \sigma_i^{-2} A_i \right)
\]
\[
= - \sum_{i=1}^{n} \sigma_i^{-2} \mathcal{A}_i^2 - \left( \bar{\sigma}^{-2} \sum_{i=1}^{n} \sigma_i^{-2} \mathcal{A}_i \right)^T \bar{\sigma}^{-2} \left( \bar{\sigma}^{-2} \sum_{i=1}^{n} \sigma_i^{-2} \mathcal{A}_i \right)
\]
\[
= - \sum_{i=1}^{n} \sigma_i^{-2} \mathcal{A}_i^2 + \bar{\sigma}^2 \bar{\mathcal{A}}^2
\]

The skew-symmetric property of matrices \( \mathcal{A}_i \) and \( \bar{\mathcal{A}} \) is employed in the above derivation, which originates from the skew-symmetric nature of the cross product matrices.

Now the attitude error-covariance can be computed. From Eq. (58) and Eq. (86), this becomes
\[
P_{\delta \alpha} = \left[ - \sum_{i=1}^{n} \sigma_i^{-2} \mathcal{A}_i^2 + \bar{\sigma}^2 \bar{\mathcal{A}}^2 \right]^{-1}
\]

Regarding the position estimates, the estimate errors are derived from Eq. (62) as
\[
\delta \mathbf{p} = -\bar{\sigma}^{-2} \sum_{i=1}^{n} \sigma_i^{-2} \left( \Delta \mathbf{a}_i - \mathcal{A}_i \delta \alpha \right)
\]
\[
= -\Delta \mathbf{a} + \bar{\mathcal{A}} \delta \alpha
\]
where
\[
\Delta \mathbf{a} \equiv \bar{\sigma}^{-2} \sum_{i=1}^{n} \sigma_i^{-2} \Delta \mathbf{a}_i
\]
The resulting error-covariance is obtained from Eq. (67). For the isotropic case, it will be
\[
\text{cov} \{ \hat{\mathbf{p}} \} = \bar{\sigma}^{-2} I_{3 \times 3} - \bar{\mathcal{A}} \left[ - \sum_{i=1}^{n} \sigma_i^{-2} \mathcal{A}_i^2 + \bar{\sigma}^2 \bar{\mathcal{A}}^2 \right]^{-1} \bar{\mathcal{A}}
\]
The observation vector estimates can be computed from Eq. (68a) and Eq. (68b), and for the isotropic covariance will result in
\[
\hat{b}_i = \hat{b}_i - \frac{\sigma_{b_i}^2}{\sigma_{b_i}^2 + \sigma_{r_i}^2} \left( \hat{b}_i - \bar{\mathcal{A}} r_i + \hat{\mathbf{p}} \right)
\]
\[
(91a)
\]
\[
\hat{r}_i = \tilde{r}_i + \frac{\sigma_{r_i}^2}{\sigma_{r_i}^2 + \sigma_{\tilde{r}_i}^2} \hat{A}^T (\hat{b}_i - \hat{\tilde{r}}_i + \hat{\hat{b}})
\]  

(91b)

From the first-order approximation of observation residuals in Eqs. (71a) and (71b), the observation residuals are given as

\[
\delta b_i - \Delta b_i \approx -\frac{\sigma_{b_i}^2}{\sigma_{b_i}^2 + \sigma_{\tilde{b}_i}^2} (\Delta a_i - G_i \delta f)
\]  

(92a)

\[
\delta r_i - \Delta r_i \approx \frac{\sigma_{r_i}^2}{\sigma_{r_i}^2 + \sigma_{\tilde{r}_i}^2} A^T (\Delta a_i - G_i \delta f)
\]  

(92b)

with the definitions of \(\Delta a_i\), \(G_i\), and \(\delta f\) in Eqs. (36a), (72c) and (72d), respectively. Then, the observation residual covariances become

\[
cov(\tilde{b}_i - \hat{b}_i) = \frac{\sigma_{b_i}^2}{\sigma_{b_i}^2 + \sigma_{\tilde{b}_i}^2} \left[ (\sigma_{b_i}^2 + \sigma_{\tilde{b}_i}^2) I_{3\times3} - [A_i - I_{3\times3}] \left( \sum_{i=1}^{n} \frac{1}{\sigma_{b_i}^2 + \sigma_{\tilde{b}_i}^2} \begin{bmatrix} -A_i^2 & A_i \\ A_i^T & I_{3\times3} \end{bmatrix} \right)^{-1} \begin{bmatrix} A_i^T \\ -I_{3\times3} \end{bmatrix} \right]
\]  

(93a)

\[
cov(\tilde{r}_i - \hat{r}_i) = \frac{\sigma_{r_i}^2}{\sigma_{r_i}^2 + \sigma_{\tilde{r}_i}^2} A^T \left[ (\sigma_{r_i}^2 + \sigma_{\tilde{r}_i}^2) I_{3\times3} - [A_i - I_{3\times3}] \left( \sum_{i=1}^{n} \frac{1}{\sigma_{r_i}^2 + \sigma_{\tilde{r}_i}^2} \begin{bmatrix} -A_i^2 & A_i \\ A_i^T & I_{3\times3} \end{bmatrix} \right)^{-1} \begin{bmatrix} A_i^T \\ -I_{3\times3} \end{bmatrix} \right] A
\]  

(93b)

The estimate covariances of the observation vectors are simplified from Eqs. (77) and (78) as

\[
P_{b_i} = \sigma_{b_i}^2 I_{3\times3} + \text{cov}(\tilde{b}_i - \hat{b}_i)
\]

\[
-2 \frac{\sigma_{b_i}^3}{\sigma_{b_i}^2 + \sigma_{\tilde{b}_i}^2} I_{3\times3} + \frac{\sigma_{b_i}^3}{\sigma_{b_i}^2 + \sigma_{\tilde{b}_i}^2} \left[ A_i - I_{3\times3} \left( \sum_{i=1}^{n} \frac{1}{\sigma_{b_i}^2 + \sigma_{\tilde{b}_i}^2} \begin{bmatrix} -A_i^2 & A_i \\ A_i^T & I_{3\times3} \end{bmatrix} \right)^{-1} \begin{bmatrix} A_i^T \\ -I_{3\times3} \end{bmatrix} \right]
\]

\[
+ \frac{\sigma_{b_i}^3}{\sigma_{b_i}^2 + \sigma_{\tilde{b}_i}^2} \left[ A_i - I_{3\times3} \left( \sum_{i=1}^{n} \frac{1}{\sigma_{b_i}^2 + \sigma_{\tilde{b}_i}^2} \begin{bmatrix} -A_i^2 & A_i \\ A_i^T & I_{3\times3} \end{bmatrix} \right)^{-1} \begin{bmatrix} A_i^T \\ -I_{3\times3} \end{bmatrix} \right] A
\]  

(94)

\[
P_{r_i} = \sigma_{r_i}^2 I_{3\times3} + \text{cov}(\tilde{r}_i - \hat{r}_i)
\]

\[
-2 \frac{\sigma_{r_i}^3}{\sigma_{r_i}^2 + \sigma_{\tilde{r}_i}^2} I_{3\times3} + \frac{\sigma_{r_i}^3}{\sigma_{r_i}^2 + \sigma_{\tilde{r}_i}^2} A^T \left[ A_i - I_{3\times3} \left( \sum_{i=1}^{n} \frac{1}{\sigma_{r_i}^2 + \sigma_{\tilde{r}_i}^2} \begin{bmatrix} -A_i^2 & A_i \\ A_i^T & I_{3\times3} \end{bmatrix} \right)^{-1} \begin{bmatrix} A_i^T \\ -I_{3\times3} \end{bmatrix} \right] A
\]

\[
+ \frac{\sigma_{r_i}^3}{\sigma_{r_i}^2 + \sigma_{\tilde{r}_i}^2} A^T \left[ A_i - I_{3\times3} \left( \sum_{i=1}^{n} \frac{1}{\sigma_{r_i}^2 + \sigma_{\tilde{r}_i}^2} \begin{bmatrix} -A_i^2 & A_i \\ A_i^T & I_{3\times3} \end{bmatrix} \right)^{-1} \begin{bmatrix} A_i^T \\ -I_{3\times3} \end{bmatrix} \right] A
\]  

(95)

All of the covariance expressions reduce down to the ones derived in [25].

**F. Fisher Information Matrix**

From the analysis in section II.D the estimate error-covariances for the attitude error \(\delta a\), the translation vector \(p\), the residuals as well as the estimate covariances for the observation vectors \(\tilde{b}_i\) and \(\tilde{r}_i\) have been derived. An efficiency proof for the estimate error-covariances of the attitude and translation vector is now shown based on the FIM and the CRLB [21] defined for the estimation covariances. For an unbiased estimator \(\hat{f}\), the estimate error-covariance has a lower bound as

\[
cov(\hat{f}) \geq F^{-1} \equiv \mathbf{E} \left[ -\frac{\partial^2}{\partial \theta \partial \hat{f}} \rho(\hat{y} | \hat{f}) \right]^{-1}
\]  

(96)
The term inside of the expectation shows the Hessian of the the negative log-likelihood function, which is given in Eq. (45). For an optimal estimator, the equality in the Eq. (96) is given, and the estimator is efficient. The Hessian of the second-order approximated cost function \( L \) is the FIM. From Eqs. (37) and Eq. (59), the following are given:

\[
\hat{\mathbf{b}}_t - \hat{\mathbf{A}}_t \hat{\mathbf{r}}_t + \hat{\mathbf{p}} \approx -\mathbf{p} + \Delta \mathbf{a}_t - \mathbf{A}_t \delta \mathbf{a} + \mathbf{p} + \delta \mathbf{p} \\
= \Delta \mathbf{a}_t - \mathbf{A}_t \delta \mathbf{a} + \delta \mathbf{p} \tag{97}
\]

The second-order cost function involving \( \delta \mathbf{p} \) is now given by

\[
L = \frac{1}{2} \sum_{i=1}^{n} (\Delta \mathbf{a}_t - \mathbf{A}_t \delta \mathbf{a} + \delta \mathbf{p})^T Q_{\Delta i}^{-1} (\Delta \mathbf{a}_t - \mathbf{A}_t \delta \mathbf{a} + \delta \mathbf{p}) \tag{98}
\]

The FIM, denoted by \( F \), will be

\[
F = \begin{bmatrix}
\sum_{i=1}^{n} \mathbf{A}_i^T Q_{\Delta i}^{-1} \mathbf{A}_i & - \sum_{i=1}^{n} \mathbf{A}_i^T Q_{\Delta i}^{-1} \\
- \sum_{i=1}^{n} Q_{\Delta i}^{-1} \mathbf{A}_i & \sum_{i=1}^{n} Q_{\Delta i}^{-1}
\end{bmatrix} \tag{99}
\]

The block matrices of the FIM are

\[
F = \begin{bmatrix}
F_{11} & F_{12} \\
F_{21} & F_{22}
\end{bmatrix} \tag{100}
\]

The inverse of FIM, denoted by \( \mathcal{F} \), is given by

\[
\mathcal{F} = \begin{bmatrix}
\mathcal{F}_{11} & \mathcal{F}_{12} \\
\mathcal{F}_{21} & \mathcal{F}_{22}
\end{bmatrix} \tag{101}
\]

The \( \mathcal{F}_{11} \) block of this matrix is calculated by the Sherman–Morrison–Woodbury lemma [37]:

\[
\mathcal{F}_{11} = \left( F_{11} - F_{12} F_{22}^{-1} F_{21} \right)^{-1}
\]

\[
= \left[ \sum_{i=1}^{n} \mathbf{A}_i^T Q_{\Delta i}^{-1} \mathbf{A}_i - \left( \sum_{i=1}^{n} \mathbf{A}_i^T Q_{\Delta i}^{-1} \right) S_A \left( - \sum_{i=1}^{n} Q_{\Delta i}^{-1} \mathbf{A}_i \right)^T \right]^{-1} \tag{102}
\]

This has already been proven in Eq. (58), which is the CRLB for the attitude error. The term \( \mathcal{F}_{22} \) is also derived by using matrix inversion lemma as

\[
\mathcal{F}_{22} = \left( F_{22} - F_{21} F_{11}^{-1} F_{12} \right)^{-1}
\]

\[
= \left[ \sum_{i=1}^{n} Q_{\Delta i}^{-1} - \left( \sum_{i=1}^{n} Q_{\Delta i}^{-1} \mathbf{A}_i \right) \left( \sum_{i=1}^{n} \mathbf{A}_i^T Q_{\Delta i}^{-1} \mathbf{A}_i \right)^{-1} \left( - \sum_{i=1}^{n} \mathbf{A}_i^T Q_{\Delta i}^{-1} \mathbf{A}_i \right) \right]^{-1} \tag{103}
\]

Using Eq. (58) leads to

\[
\mathcal{F}_{22} = S_A + \mathbf{A} \left[ \sum_{i=1}^{n} \mathbf{A}_i^T Q_{\Delta i}^{-1} \mathbf{A}_i - \sum_{i=1}^{n} \mathbf{A}_i^T Q_{\Delta i}^{-1} S_A \sum_{i=1}^{n} Q_{\Delta i}^{-1} \mathbf{A}_i \right]^{-1} \mathbf{A}^T \tag{104}
\]

This proves the CRLB for covariance of \( \{ \delta \mathbf{p} \} \). Note that from Eq. (58) it can be concluded that the CRLB holds for the attitude estimate because the estimate error-covariance is equal to the inverse of the Hessian of the cost function in Eq. (45).
III. Numerical Validation

A pose estimation problem is solved here with two scans of a vector-observation-enabled sensor, which can be a Lidar, with three vector observations per scan. The ground truth values for the attitude matrix \( A \), translation vector \( \mathbf{p} \) and vector observations \( \mathbf{b}_i, i = 1, 2, 3 \), as well as the measurement covariances, are given as

\[
A = I_{3 \times 3}
\]

\[
\mathbf{p} = \begin{bmatrix} 0.3 & -0.4 & 0.5 \end{bmatrix}^T
\]

\[
\mathbf{b}_1 = \begin{bmatrix} 0 & 9.7590 \times 10^{-2} & -1.4833 \times 10^{-1} \end{bmatrix}^T
\]

\[
\mathbf{b}_2 = \begin{bmatrix} 0 & 1.9518 \times 10^{-1} & -1.2855 \times 10^{-2} \end{bmatrix}^T
\]

\[
\mathbf{b}_3 = \begin{bmatrix} 1 & 9.7590 \times 10^{-1} & 9.8885 \times 10^{-1} \end{bmatrix}^T
\]

\[
R_1 = 10^{-6} \times \begin{bmatrix} 0.1902 & 0.0228 & -0.0190 & -0.0345 & -0.0079 & 0.0225 \\ 0.0228 & 0.2288 & -0.0003 & 0.0145 & 0.0483 & -0.0161 \\ -0.0190 & -0.0003 & 0.3554 & 0.0765 & -0.0180 & 0.1386 \\ -0.0345 & 0.0145 & 0.0765 & 0.2566 & -0.0201 & 0.0408 \\ -0.0079 & 0.0483 & -0.0180 & -0.0201 & 0.2621 & -0.0800 \\ 0.0225 & -0.0161 & 0.1386 & 0.0408 & -0.0800 & 0.3349 \end{bmatrix}
\]

\[
R_2 = 10^{-6} \times \begin{bmatrix} 0.1981 & 0.0213 & 0.0021 & -0.0519 & -0.0218 & -0.0231 \\ 0.0213 & 0.1980 & -0.0264 & 0.0023 & -0.0116 & 0.0030 \\ 0.0021 & -0.0264 & 0.2040 & -0.0456 & 0.0273 & -0.0152 \\ -0.0519 & 0.0023 & -0.0456 & 0.2481 & 0.0025 & 0.0258 \\ -0.0218 & -0.0116 & 0.0273 & 0.0025 & 0.1933 & 0.0069 \\ -0.0231 & 0.0030 & -0.0152 & 0.0258 & 0.0069 & 0.1851 \end{bmatrix}
\]

\[
R_3 = 10^{-6} \times \begin{bmatrix} 0.1705 & -0.0071 & -0.0154 & -0.0247 & 0.0081 & 0.0049 \\ -0.0071 & 0.2036 & 0.0038 & 0.0259 & -0.0311 & 0.0064 \\ -0.0154 & 0.0038 & 0.1910 & 0.0376 & 0.0085 & 0.0166 \\ -0.0247 & 0.0259 & 0.0376 & 0.2738 & -0.0153 & 0.0170 \\ 0.0081 & -0.0311 & 0.0085 & -0.0153 & 0.1850 & -0.0114 \\ 0.0049 & 0.0064 & 0.0166 & 0.0170 & -0.0114 & 0.2049 \end{bmatrix}
\]

The true observation vectors \( \mathbf{r}_i \) are generated by the constraint in Eq. (14). Both measurement and actual observation vectors have a meter unit. A Monte-Carlo simulation with 10,000 samples is performed here to showcase how well the 3σ bounds generated by the covariance expressions in Eqs. (58) and (67) for the attitude and translation vectors, respectively, Eqs. (77) and (78) for estimated observations, and Eqs. (74a) and (74b) for residuals of observations are bounding their corresponding estimate errors or residuals. Artificial noise is generated from a Gaussian distribution with zero mean and covariance of \( R_{r_1}, R_{b_1} \) and \( R_{r_2} \) to produce \( \tilde{\mathbf{r}}_i \) and \( \tilde{\mathbf{b}}_i \) samples. The covariance matrices of the measurements \( R_{r_1}, R_{b_1} \) and \( R_{r_2} \) are generated randomly with a signal-to-noise ratio of around \( 10^{-4} \). Note that by default, the measurement covariance is selected to be positive-definite while being random. Singularities in the measurement covariance matrix can be handled if they exist, but this is not the focus of this paper. Figure 2(a) shows the attitude errors in terms of roll, pitch and yaw angles in degrees from the Monte-Carlo samples. The blue line depicts the estimation errors, and the red lines are the 3σ bounds computed from the estimate error-covariances. Figure 2(b) shows the translation vector estimate error in the x, y and z directions, respectively. It is seen that the estimate errors are well-bounded by their corresponding 3σ bounds. Figure 3(a) shows the estimation errors, and Figure 3(b) shows the residual errors, for observation vector \( \mathbf{b}_1 \). Figures 4(a) and 4(b) show the same results for the observation vector \( \mathbf{r}_1 \), respectively. It is seen that observation vectors are also bounded by their corresponding 3σ bounds, provided by the covariances of estimates as well as the residuals.
IV. Conclusions

This paper develops an analytical framework for an efficient pose estimation problem within the first-order approximation of estimate errors. Pose estimation is central to simultaneous localization and mapping (SLAM) problems. Efficient estimation makes the controller policy easier to track the desired signals to obtain a more accurate estimate of the states, which in turn consumes lower energy levels in the control action. The static SLAM problem is shown to be solved as a total least squares (TLS) problem. A quadratic cost function based on the TLS formulation is introduced for taking into account the attitude matrix and translation vector. The weight matrix in the cost function is extracted from the most generic positive-definite fully populated matrix to include the correlations between the observation vectors in the most general case. This work expands previously derived pose estimation solutions where an isotropic assumption is applied to the measurement noise covariance. The cost function is then written in the attitude-only format. The
 covariance expression of the attitude error is provided within the small-angle assumption. This assumption leads to a
second-order approximation of the cost function in terms of the attitude error. Covariance expressions for the translation
vector as well as the estimates and residuals of the observation vectors are obtained analytically. The Fisher information
matrix (FIM) is derived, and the covariance expressions are proven to be the block inverses of FIM, proving the equality
in the Cramér-Rao lower bound and thus the efficiency of estimates within the first-order attitude errors. A simulation
framework showcases the efficacy of the covariance analyses by simulating observation vectors in a pose estimation
problem with 10,000 Monte-Carlo samples. Also, an advantage of the proposed solution is that it provides an estimate of
the residual for the observation vectors as well as their covariance. These quantities are useful to assess the performance
of the pose estimation solution when no access to the ground truth data in an actual SLAM application is available.

Appendix

A list of some preliminary equations used in the paper is shown here. For the cross product of two $3 \times 1$ vectors $\mathbf{a}$
and $\mathbf{b}$, the following relation is given:

$$\mathbf{a} \times \mathbf{b} = [\mathbf{a} \times] \mathbf{b} = -\mathbf{b} \times \mathbf{a} = -[\mathbf{b} \times] \mathbf{a}$$

(105)

The cross product matrix of a vector $\mathbf{a}$ is a skew-symmetric matrix, so that

$$[\mathbf{a} \times]^T = -[\mathbf{a} \times]$$

(106)

where $[\mathbf{a} \times]$ denotes the $3 \times 3$ cross product matrix constructed from the vector $\mathbf{a}$ as

$$[\mathbf{a} \times] = \begin{bmatrix}
0 & -a_3 & a_2 \\
-a_3 & 0 & -a_1 \\
a_2 & a_1 & 0
\end{bmatrix}$$

(107)

The Kronecker product \[33\] also can be combined with the vec operator in the following identity, where the notation
vec(.) \[38\] for an $m \times n$ matrix $A$ results in an $(mn) \times 1$ matrix vec($A$) that is constructed by stacking the columns of $A$:

$$A\mathbf{z} = (\mathbf{x}^T \otimes I_{m \times m}) \text{vec}(A)$$

(108)

References

[1] Crassidis, J. L., Markley, F. L., and Cheng, Y., “Survey of Nonlinear Attitude Estimation Methods,” Journal of Guidance,
Control, and Dynamics, Vol. 30, No. 1, 2007, pp. 12–28. [https://doi.org/10.2514/1.22452]
[23] Hashim, H. A., “Attitude Determination and Estimation Using Vector Observations: Review, Challenges and Comparative Results,” arXiv preprint arXiv:2001.03787, 2020.

[24] Crassidis, J. L., Optimal estimation of dynamic systems John L. Crassidis, John L. Junkins., 2nd ed., Chapman & Hall/CRC applied mathematics and nonlinear science series, Chapman and Hall/CRC, Boca Raton, Fla, 2011, Chap. 2. https://doi.org/10.1201/b11154

[25] Cheng, Y., and Crassidis, J. L., “Optimal Pose Estimation with Error-Covariance Analysis,” AIAA Scitech 2021 Forum, 2021, p. 1758. https://doi.org/10.2514/6.2021-1758

[26] Estépar, R. S. J., Brun, A., and Westin, C.-F., “Robust Generalized Total Least Squares Iterative Closest Point Registration,” International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2004, pp. 234–241. https://doi.org/10.1007/978-3-540-30135-6_29

[27] Shuster, M. D., “A Survey of Attitude Representations,” The Journal of the Astronautical Sciences, Vol. 41, No. 4, 1993, pp. 439–517.

[28] Golub, G. H., “Some Modified Matrix Eigenvalue Problems,” Siam Review, Vol. 15, No. 2, 1973, pp. 318–334. https://doi.org/10.1137/1015032

[29] Markovsky, I., and Van Huffel, S., “Overview of Total Least-Squares Methods,” Signal processing, Vol. 87, No. 10, 2007, pp. 2283–2302. https://doi.org/10.1016/j.sigpro.2007.04.004

[30] 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. https://doi.org/10.1137/0717073

[31] Cramér, H., Mathematical Methods of Statistics, Vol. 43, Princeton university press, 1999. https://doi.org/10.1015/9781400883868

[32] Crassidis, J. L., and Cheng, Y., “Maximum Likelihood Analysis of the Total Least Squares Problem with Correlated Errors,” Journal of Guidance, Control, and Dynamics, Vol. 42, No. 6, 2019, pp. 1204–1217. https://doi.org/10.2514/6.2019-1931

[33] Steeb, W.-H., and Hardy, Y., Matrix Calculus and Kronecker Product: A Practical Approach to Linear and Multilinear Algebra, World Scientific Publishing Company, 2011. https://doi.org/10.1142/8030

[34] Shuster, M. D., and Oh, S. D., “Three-Axis Attitude Determination from Vector Observations,” Journal of guidance and Control, Vol. 4, No. 1, 1981, pp. 70–77. https://doi.org/10.2514/3.19717

[35] Amiri-Simkooei, A., “Weighted Total Least Squares with Singular Covariance Matrices subject to Weighted and Hard Constraints,” Journal of Surveying Engineering, Vol. 143, No. 4, 2017, p. 04017018. https://doi.org/10.1061/(ASCE)SU.1943-5428.0000239

[36] Markley, F. L., and Crassidis, J. L., Fundamentals of Spacecraft Attitude Determination and Control, Springer, 2014.

[37] Sherman, J., and Morrison, W. J., “Adjustment of an Inverse Matrix Corresponding to a Change in One Element of a Given Matrix,” The Annals of Mathematical Statistics, Vol. 21, No. 1, 1950, pp. 124–127. https://doi.org/10.1214/aoms/1177729895

[38] Brewer, J., “Kronecker Products and Matrix Calculus in System Theory,” IEEE Transactions on Circuits and Systems, Vol. 25, No. 9, 1978, pp. 772–781. https://doi.org/10.1109/TCS.1978.1084534