Uncertainty characterization of the orthogonal Procrustes problem with arbitrary covariance matrices

Pedro Lourenço, Bruno Guerreiro, Pedro Batista, Paulo Oliveira,
and Carlos Silvestre

Pattern Recognition, vol. 61, pp. 210-220, January 2017

https://doi.org/10.1016/j.patcog.2016.07.037

Uncertainty Characterization of the Orthogonal Procrustes Problem with Arbitrary Covariance Matrices

Pedro Lourenço[a,*], Bruno J. Guerreiro[a,b], Pedro Batista[a,b], Paulo Oliveira[d,c,a], Carlos Silvestre[e,a]

[a] *Institute for Systems and Robotics, Laboratory of Robotics and Systems in Engineering and Science, Portugal.*
[b] *Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal*
[c] *Department of Mechanical Engineering, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal*
[d] *Institute of Mechanical Engineering, Associated Laboratory for Energy, Transports and Aeronautics, Lisbon, Portugal*
[e] *Department of Electrical and Computer Engineering, Faculty of Science and Technology of the University of Macau*

Abstract

This paper addresses the weighted orthogonal Procrustes problem of matching stochastically perturbed point clouds, formulated as an optimization problem with a closed-form solution. A novel uncertainty characterization of the solution of this problem is proposed resorting to perturbation theory concepts, which admits arbitrary transformations between point clouds and individual covariance and cross-covariance matrices for the points of each cloud. The method is thoroughly validated through extensive Monte Carlo simulations, and particularly interesting cases where nonlinearities may arise are further analysed.

Keywords: Weighted Procrustes Statistics, Perturbation Theory,

Uncertainty Characterization, Map Transformation

*Corresponding author

Email addresses: plourenco@isr.tecnico.ulisboa.pt (Pedro Lourenço), bguerreiro@isr.tecnico.ulisboa.pt (Bruno J. Guerreiro), pbatista@isr.tecnico.ulisboa.pt (Pedro Batista), pjcro@isr.tecnico.ulisboa.pt (Paulo Oliveira), csilvestre@umac.mo (Carlos Silvestre)

1. Introduction

The problem of finding the similarity transformation between two sets of points in n-dimensional space appears commonly in many applications of computer vision, robotics, statistics, and other fields of research. The study of this family of problems is usually known as the Procrustes analysis [1], which includes the statistical characterization of the transformation between the shape of objects [2]. One particularly important problem in this family is the so-called orthogonal Procrustes problem, which can be traced back to the work presented in [3], and consists in extracting the orthogonal transformation that maps one set of points into a second set of points, with known associations between them. It is closely related to Wahba's problem [4] and to the Kabsch algorithm [5]. The generalization for rotation, translation, and scaling has been the subject of extensive research in areas such as computer vision, and can be traced back to [6], [7], and [8]. While initially the problem was posed without any restrictions on the transformation between the sets, i.e., rotations and reflections were allowed, a more evolved strategy appeared restricting the transformation to the special orthogonal group, as detailed in [6] and [9]. Furthermore, [10] demonstrated that the previous solutions are optimal even when both data sets are perturbed with isotropic and identical Gaussian noise.

The statistical characterization of the Procrustes analysis has also been the subject of study in works such as [2], [9], [11], and [12]. Using perturbation theory, the nonlinear problem of characterizing the uncertainty was addressed with some limiting options, such as the absence of weighting of the point sets, the use of small rotations, or the same covariance for all points.

Within the field of medical imaging, the work presented in [13] also resorts to perturbation theory to present a statistical characterization of a target position, considering small rotations, isotropic uncertainty, and equal weights for each point. More recently, the work presented in [14] extends these results for anisotropic uncertainty in the components of the point space. This is achieved by considering the same covariance matrix for all points, which may weigh each component of the point space independently. The authors of [15] further expand this by considering different noise levels for each point, while keeping the linearized model for the rotation matrix. An interesting advance in the study of the uncertainty is the first order error propagation proposed in [16]. The optimization problem that is considered is not weighted and therefore identical isotropic noise is assumed for all the points. The author defines a first order error model that is propagated through the solution, while assuming independent and identically distributed points (no longer necessarily isotropic). It is noted that the findings of the aforementioned works are all restricted to three-dimensional points. In [17] a different optimization problem is proposed that accounts for independent anisotropic noise affecting rotated-only point sets also in three dimensions. The authors determine the theoretical lower bound for the covariance of the rotation error in that case, and through an iterative solution recurring to quaternion representation reach the theoretical bound. Besides the iterative solution, some shortcomings of this work are its limitation to the tridimensional problem with rotation-only, and the fact that, although anisotropic, the input covariances are normalized and share a common normalizing factor. Regarding the stability of the solution, [18] addresses the study of this issue when the algo-

rithm is exposed to perturbed data sets, concluding that the singular values of the matrices composed with the points in each set are closely related to the conditioning of the problem, whilst finding a bound for the perturbation on the rotation matrix when the input perturbations are bounded. In related directions of research that demonstrate the relevance of pattern point matching, and, consequently, of point registration problems such as the Procrustes problem, the authors of [19] and [20] propose algorithms that exploit different approaches to registration and matching. Furthermore, the latter is an iterative algorithm that, assuming a rotation-scaling-translation transformation between two sets of points, finds the point correspondences and a variational Bayesian approximation for the distribution of the transformation.

This paper addresses the n-dimensional (n-D) extended orthogonal Procrustes problem considering a transformation composed of a rotation and a translation (no scaling). The problem is posed with individual scalar weights for each pair of points, and a closed-form solution is presented. Data association is assumed to be performed a priori. Founded on perturbation theory, a novel and general uncertainty description for the solution of the optimization problem is proposed. Building on the results presented in [13], [14], and [16], and assuming a stochastic perturbation model for the point sets with individual covariance matrices for each point, as well as cross-covariances for each pair of points, the first and second moments of the resulting translation and rotation are computed. This is achieved considering arbitrary rotations and translations, individual weights, and full covariance matrices for both point sets. As a by-product of this work, an application to robotics was proposed in [21] and [22] within the scope of simultaneous localization

and mapping [23]. In this application, if a landmark map (or set of points) is available in a coordinate frame attached to the robot, it is possible to compute the transformation between that frame and another frame fixed to the initial position of the robot. Following this idea, an online Earth-fixed trajectory and map estimation algorithm based on the Procrustes problem was proposed and its uncertainty characterization derived, making full use of the methodology proposed in this paper. This builds on the previous works by the authors, where globally asymptotically stable filters for simultaneous localization and mapping in a sensor-based or robocentric framework were proposed for bidimensional [24] and tridimensional mission scenarios [25]. The performance and consistency of the overall algorithm are validated in a real world environment for both dimensionalities, showing that the algorithm provides accurate and consistent estimates, and, therefore, also providing an experimental validation of the uncertainty characterization proposed in this paper.

The contributions of this paper are: i) the full uncertainty characterization of the optimization problem of obtaining the transformation between corresponding n-dimensional point sets and its closed-form solution, while considering point sets perturbed by anisotropic noise, and points that are not required to be independent nor identically distributed; and ii) a thorough validation of the uncertainty characterization, using extensive Monte Carlo simulations to study the main properties of the proposed methodology. This paper builds on the preliminary versions of this work presented in [21] and [22], by reformulating the problem of obtaining the pose of the vehicle, while extending the derivation therein to points of arbitrary dimensions. In

contrast with the latter work, this paper provides new theoretical results, generalizes the proposed uncertainty characterization to $\mathbb{R}^n$, and provides statistical validation through extensive Monte Carlo simulations for several dimensions and a multitude of parameter combinations.

The applications of Procrustes analysis are found in a wide variety of fields, which can benefit from the proposed approach, including rigid body motion, vibration tests of large complex structures [26], structural and system identification, factor analysis in n-D (e.g. checking whether two matrices are equivalent), similarity evaluation in statistical data sets [27, Chapter 20], medical imaging [14], photogrammetry [28], shape comparison (generalized Procrustes analysis) [29], and quantitative psychology [30] (where the problem was initially solved). In recent years several algorithms were developed in the field of computer vision that availed themselves of the Procrustes problem, from shape matching and retrieval [31] to similarity search in image collections [32], among others. Shape matching is in fact a more complex problem, as the problem of finding the transformations is coupled with the problem of finding the reference shape to which all the measured shapes relate. In [33] the authors propose a unifying framework that has a closed-form computation for affine, similarity or Euclidean transformations between a set of shapes, while allowing to find the underlying shape and accounting for missing pairs of points. All this is performed considering noise in the measured shapes and not in the reference-space as is customary. Other applications include non-negative matrix factorization [34], and phase FIR filter bank design [35], whereas the work in [36] underlines the importance of addressing the problem in less common dimensionalities, such as four-

dimensional shapes. Another possible application of the Procrustes problem lies in iterative closest point algorithms such as [37], even tough most use quaternions to parametrize the rotation of the sets. If the registration is performed in each step with a constrained least squares approach, it can benefit from the characterization here proposed. Another interesting application of Procrustes analysis is manifold alignment [38] in the area of machine learning. In this n-dimensional technique, it is argued that it is possible to model the underlying structure of most datasets by manifolds, whose alignment then allows for knowledge transfer across datasets. The authors of [38] demonstrate the validity of this approach by applying the idea to learning transfer in reinforcement learning with Markov Decision Processes, alignment of the tertiary structure of proteins, cross lingual information alignment, within others. These demonstrate the real world relevance of the n-dimensional Procrustes problem in several fields even for dimensionalities outside of the 2-D/3-D common problems. Furthermore, given the noisy nature of these problems, the proposed uncertainty characterization can be useful to compute the reliability of the alignment resulting from the Procrustes procedure.

Paper Structure. The paper is organized as follows. Section 2 presents a brief overview of some mathematical concepts needed in the course of this paper. Section 3 presents the formulation and closed-form solution of the weighted orthogonal Procrustes problem. A novel uncertainty characterization of this problem is derived in Section 4 and validated in Section 5 through extensive Monte Carlo simulations. Finally, Section 6 provides concluding remarks. Further figures depicting the simulations detailed in this section are provided in the supplementary material, along with detailed proofs to the results.

2. Preliminary definitions

This section serves the purpose of introducing the notation used in this paper, as well as a few definitions and properties needed for the mathematical derivations in the sequel.

2.1. Notation

Throughout the paper, vectors and matrices are represented in small and capital boldface letters, respectively. Scalar symbols are expressed in italic: constants by capital letters, and scalar variables in small letters. Particularly, the symbol $\mathbf{0}_{n\times m}$ denotes an $n \times m$ matrix of zeros (if only one subscript is present, the matrix is square), $\mathbf{I}_n$ is an identity matrix with dimension $n \times n$, and $\text{diag}\left(\mathbf{A}_1, \ldots, \mathbf{A}_n\right)$ is a block diagonal matrix. The determinant of a generic square matrix is denoted by $|\mathbf{A}|$, and, for a generic matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$, the Frobenius norm is adopted, i.e., $\|\mathbf{A}\| = \sqrt{\text{tr}\left(\mathbf{A}\mathbf{A}^T\right)}$. The operator skew : $\mathbb{R}^{n\times n} \to \mathfrak{so}(n)$ yields the skew-symmetric component of a square matrix, $\text{skew}(\mathbf{A}) = \frac{1}{2}\left(\mathbf{A} - \mathbf{A}^T\right)$. Finally, the expectation operator is denoted as $\langle . \rangle$, and the covariance matrix between two generic stochastic vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ is denoted by $\mathbf{\Sigma}_{ab} = \left\langle \left(\mathbf{a} - \langle \mathbf{a} \rangle\right)\left(\mathbf{b} - \langle \mathbf{b} \rangle\right)^T \right\rangle$ or $\mathbf{\Sigma}_a$, if $\mathbf{a} = \mathbf{b}$. The Orthogonal Group is denoted by $\text{O}(n) := \left\{\mathbf{X} \in \mathbb{R}^{n\times n} : \mathbf{X}\mathbf{X}^T = \mathbf{X}^T\mathbf{X} = \mathbf{I}\right\}$, and the Special Orthogonal Group is denoted by $\text{SO}(n) := \{\mathbf{X} \in \text{O}(n) : |\mathbf{X}| = 1\}$.

2.2. Definitions and properties

In this paper, except when explicitly stated, the dimension of the space $\mathbb{R}^n$ is arbitrary, i.e., all the derivations are valid for point clouds in $\mathbb{R}^n$ for all $n \geq 2$, and the rotations are expressed in the special orthogonal group $\text{SO}(n)$. Note that the term rotation, and rotation matrix, applies to all the

orthogonal matrices of unitary determinant. This is related to the Lie algebra $\mathfrak{so}(n)$ comprised of skew-symmetric matrices that can be mapped to $\mathrm{SO}(n)$ through the exponential map.

Definition 1. *The matrix $\mathbf{S}[\boldsymbol{\omega}] \in \mathfrak{so}(n)$ is a skew-symmetric matrix parametrized by the vector $\boldsymbol{\omega} := \begin{bmatrix}\omega_1 & \cdots & \omega_{n_p}\end{bmatrix}^T \in \mathbb{R}^{n_p}$, with $n_p = \frac{n(n-1)}{2}$. One possible parametrization is given by*

$$\mathbf{S}[\boldsymbol{\omega}] = \begin{bmatrix} 0 & (-1)\omega_{n_p} & (-1)^2\omega_{n_p-1} & \cdots & (-1)^{n-2}\omega_{2n-3} & (-1)^{n-1}\omega_{n-1} \\ * & 0 & (-1)\omega_{n_p-2} & \cdots & (-1)^{n-3}\omega_{2n-4} & (-1)^{n-2}\omega_{n-2} \\ * & * & 0 & & \vdots & \vdots \\ * & * & * & \ddots & (-1)\omega_n & (-1)^2\omega_2 \\ * & * & * & * & 0 & (-1)\omega_1 \\ * & * & * & * & * & 0 \end{bmatrix}. \quad (1)$$

The elements under the diagonal are automatically defined by $\mathbf{S}[\boldsymbol{\omega}] = -\mathbf{S}^T[\boldsymbol{\omega}]$, and are therefore omitted to avoid cluttering the reading. Also note that, in general, $n_p \geq n$, except in the bidimensional case when $n_p = 1$ and the whole vector $\boldsymbol{\omega}$ collapses to ω_1. The unskew operator $\mathbf{S}^{-1} : \mathfrak{so}(n) \to \mathbb{R}^{n_p}$ is related to this matrix as it extracts the vector that parametrizes a skew-symmetric matrix, i.e., $\mathbf{S}^{-1}(\mathbf{S}[\boldsymbol{\omega}]) = \boldsymbol{\omega}$.

Notice that, given $\boldsymbol{\omega} \in \mathbb{R}^{n_p}$, other parametrizations could be used to obtain $\mathbf{S}[\boldsymbol{\omega}]$. This form was chosen without loss of generality, as the results could be derived for any parametrization. For the tridimensional case ($n = 3$), this matrix is also known to encode the cross-product, as in $\mathbf{S}[\mathbf{a}]\,\mathbf{b} = \mathbf{a} \times \mathbf{b}$. Due to this fact, there is a possible anti-commutation between the two vectors given by $\mathbf{S}[\mathbf{a}]\,\mathbf{b} = -\mathbf{S}[\mathbf{b}]\,\mathbf{a} = \mathbf{S}^T[\mathbf{b}]\,\mathbf{a}$. Even though these properties are

solely applicable to the tridimensional case, a possible extension to all the dimensions is the matrix $\bar{\mathbf{S}}[.]$ defined hereafter.

Definition 2. *The generalized anti-commutation matrix* $\bar{\mathbf{S}}[\mathbf{a}] \in \mathbb{R}^{n_p \times n}$ *parametrized by the vector* $\mathbf{a} \in \mathbb{R}^n$*, when associated to the parametrization of* $\mathbf{S}[\boldsymbol{\omega}]$ *in* (1)*, can be defined recursively by*

$$\bar{\mathbf{S}}[\mathbf{a}_{1:i}] = \left[\begin{array}{cccc|c} 0 & \cdots & 0 & -a_i & (-1)^0 a_{i-1} \\ \vdots & & (-1)^2 a_i & 0 & (-1)^1 a_{i-2} \\ 0 & \iddots & & \vdots & \vdots \\ (-1)^{i-1} a_i & 0 & \cdots & 0 & (-1)^{i-2} a_1 \\ \hline & \bar{\mathbf{S}}[\mathbf{a}_{1:i-1}] & & & \mathbf{0}_{\frac{(i-1)(i-2)}{2} \times 1} \end{array}\right] \tag{2}$$

for all $i = 3, \ldots, n$*, with* $\mathbf{a}_{1:i} = \begin{bmatrix} a_1 & \cdots & a_i \end{bmatrix}^T$ *and* $\bar{\mathbf{S}}[\mathbf{a}_{1:2}] = \begin{bmatrix} -a_2 & a_1 \end{bmatrix}$.

Note that for vectors in three dimensions, this matrix degenerates in the skew-symmetric matrix, i.e., $\bar{\mathbf{S}}[\mathbf{a}_{1:3}] = \mathbf{S}[\mathbf{a}_{1:3}]$. Besides this fact, the matrix (2) exhibits a series of properties that are used in this paper. These are summarized in the following lemma.

Lemma 1. *The generalized anti-commutation matrix has the following properties for every* $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$*,* $\boldsymbol{\omega} \in \mathbb{R}^{n_p}$*, constants* $\alpha, \beta \in \mathbb{R}$ *and integer* $n \geq 2$*.*

1. *Linearity:* $\bar{\mathbf{S}}[\alpha\mathbf{a} + \beta\mathbf{b}] = \alpha\bar{\mathbf{S}}[\mathbf{a}] + \beta\bar{\mathbf{S}}[\mathbf{b}]$
2. *Anti-commutativity:* $\bar{\mathbf{S}}[\mathbf{a}]\,\mathbf{b} = -\bar{\mathbf{S}}[\mathbf{b}]\,\mathbf{a}$
3. *Anti-commutation with a skew-symmetric matrix:* $\mathbf{S}[\boldsymbol{\omega}]\,\mathbf{a} = \bar{\mathbf{S}}^T[\mathbf{a}]\,\boldsymbol{\omega}$
4. *Relation to the unskew operator:* $\mathbf{S}^{-1}\big(\mathbf{b}\mathbf{a}^T - \mathbf{a}\mathbf{b}^T\big) = \bar{\mathbf{S}}[\mathbf{a}]\,\mathbf{b}$.

Proof. Due to the recursive nature of $\bar{\mathbf{S}}[\mathbf{a}]$ and $\mathbf{S}[\boldsymbol{\omega}]$, it is possible to express all the sides of these identities recursively and thus these properties can be proven using mathematical induction. The proof is presented in Appendix A in the supplementary material. □

3. Procrustes optimization problem and closed-form solution

This section presents the formulation and closed-form solution of the weighted extended orthogonal Procrustes problem, featuring individual weights for each pair of points, related by a translation and rotation.

Consider the existence of two point sets in $\mathbb{R}^n$, $\mathcal{S}_A$ and $\mathcal{S}_B$, which contain, respectively, the points expressed in an arbitrary frame $\{A\}$ and the same points expressed in some other frame $\{B\}$. Each point $\mathbf{a}_i \in \mathcal{S}_A$ nominally corresponds to a point $\mathbf{b}_i \in \mathcal{S}_B$, with $i \in \mathcal{S} := \{1, \ldots, m\}$ and $m \geq n$, and that correspondence is expressed by $\mathbf{a}_i = \mathbf{R}\mathbf{b}_i + \mathbf{t}$, where the pair $(\mathbf{R}, \mathbf{t}) \in \mathrm{SO}(n) \times \mathbb{R}^n$ fully defines the transformation from frame $\{B\}$ to frame $\{A\}$, as it represents the rotation and translation from $\{B\}$ to $\{A\}$. Given the relation between the two sets, it is possible to define the error function $\mathbf{e}_i = \mathbf{a}_i - \mathbf{R}\mathbf{b}_i - \mathbf{t}$, that represents the error between the i-th point estimate in $\mathcal{S}_A$ and its homologous in $\mathcal{S}_B$, rotated and translated into frame $\{A\}$. Obtaining the pair $(\mathbf{R}, \mathbf{t})$ is the purpose of the optimization problem

$$(\mathbf{R}^*, \mathbf{t}^*) = \arg \min_{\substack{\mathbf{R} \in \mathrm{SO}(n) \\ \mathbf{t} \in \mathbb{R}^n}} G(\mathbf{R}, \mathbf{t}), \tag{3}$$

where the function $G(\mathbf{R}, \mathbf{t})$ is defined as

$$G(\mathbf{R}, \mathbf{t}) := \frac{1}{m} \sum_{i=1}^{m} \sigma_i^{-2} \left\| \mathbf{a}_i - \mathbf{R}\mathbf{b}_i - \mathbf{t} \right\|^2 = \frac{1}{m} \left\| \left(\mathbf{Y} - \mathbf{R}\mathbf{X} - \mathbf{t}\mathbf{1}^T \right) \boldsymbol{\Sigma}_e^{-1/2} \right\|^2,$$

where $\sigma_i^2 > 0$, $i \in \mathcal{S}$, accounts for the intrinsic uncertainty of each point pair, $\mathbf{Y} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_m \end{bmatrix} \in \mathbb{R}^{n \times m}$ and $\mathbf{X} = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_m \end{bmatrix} \in \mathbb{R}^{n \times m}$ are, respectively, the concatenation of the point vectors expressed in frames $\{A\}$ and $\{B\}$, $\mathbf{1} = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}^T \in \mathbb{R}^m$ is a vector of ones, and the weight matrix $\mathbf{\Sigma}_e$ is a diagonal matrix whose entries are the weights $\sigma_1^2, \ldots, \sigma_m^2$ that model the point uncertainty. If, for example, one of the points in a pair is missing, the corresponding σ_i^{-2} can be set to zero, thus allowing the cost function to account for missing data. Given that the true $\mathbf{\Sigma}_e$ is not known, these can be conservatively defined as $\sigma_i^2 = \lambda_{max}\left(\mathbf{\Sigma}_{a_i}\right) + \lambda_{max}\left(\mathbf{\Sigma}_{b_i}\right) \geq \lambda_{max}\left(\mathbf{\Sigma}_{a_i} + \mathbf{R}\mathbf{\Sigma}_{b_i}\mathbf{R}^T\right)$, denoting $\lambda_{max}(.)$ as the maximum eigenvalue. This weight matrix allows the use of the information regarding the different degrees of uncertainty of each point pair.

The optimization problem (3) has a closed-form, numerically robust, and computationally efficient solution based on the work presented in [6] and [2]. The weighted statistical properties of the point sets $\mathcal{S}_A$ and $\mathcal{S}_B$, in the form of their weighted centroids and covariances, can be expressed in matrix form using the symmetric weight matrix $\mathbf{W} := \mathbf{\Sigma}_e^{-1} - \frac{1}{N_W}\mathbf{\Sigma}_e^{-1}\mathbf{1}\mathbf{1}^T\mathbf{\Sigma}_e^{-1} \in \mathbb{R}^{m \times m}$, where $N_W := \sum_{i=1}^{m} \sigma_i^{-2} = \mathbf{1}^T\mathbf{\Sigma}_e^{-1}\mathbf{1}$. The resulting expressions for the weighted centroids of the sets $\mathcal{S}_A$ and $\mathcal{S}_B$, respectively $\boldsymbol{\mu}_A, \boldsymbol{\mu}_B \in \mathbb{R}^n$, are

$$\boldsymbol{\mu}_A := \frac{1}{N_W}\sum_{i=1}^{m} \sigma_i^{-2}\mathbf{a}_i = \frac{1}{N_W}\mathbf{Y}\mathbf{\Sigma}_e^{-1}\mathbf{1}, \quad \text{and} \quad \boldsymbol{\mu}_B := \frac{1}{N_W}\mathbf{X}\mathbf{\Sigma}_e^{-1}\mathbf{1},$$

whereas the weighted covariance $\mathbf{\Sigma}_{AB} \in \mathbb{R}^{n \times n}$ is given by

$$\mathbf{\Sigma}_{AB} := \frac{1}{N_W}\sum_{i=1}^{m} \sigma_i^{-2}\left(\mathbf{a}_i - \boldsymbol{\mu}_A\right)\left(\mathbf{b}_i - \boldsymbol{\mu}_B\right)^T = \frac{1}{N_W}\mathbf{Y}\mathbf{W}\mathbf{X}^T := \frac{1}{N_W}\mathbf{B}^T.$$

Consider now the singular value decomposition of $\mathbf{B}^T$

$$\mathbf{U}\mathbf{D}\mathbf{V}^T = \text{svd}(\mathbf{B}^T). \tag{4}$$

The optimal rotation matrix, from the optimization problem (3), is given by

$$\mathbf{R}^* = \mathbf{U}\,\text{diag}\,(1, \ldots, 1, |\mathbf{U}|\,|\mathbf{V}|)\,\mathbf{V}^T, \tag{5}$$

which in fact is the optimal rotation between the two sets when their weighted centroids coincide. This solution is valid as long as the covariance matrix $\mathbf{\Sigma}_{AB}$ has rank $n-1$, as reported in [6]. The optimal translation vector is

$$\mathbf{t}^* = \frac{1}{N_W}\,(\mathbf{Y} - \mathbf{R}^*\mathbf{X})\,\mathbf{\Sigma}_e^{-1}\mathbf{1} = \boldsymbol{\mu}_A - \mathbf{R}^*\boldsymbol{\mu}_B. \tag{6}$$

Notice that the optimal translation is the vector that translates the weighted centroid of the points in $\mathcal{S}_B$ rotated to frame $\{A\}$ to the weighted centroid of the points in $\mathcal{S}_A$.

Some remarks on applications. The derivation here presented and followed in the next section is detailed for the matching of two point clouds. Nevertheless, the problem is completely equivalent to the general problem of matching two different matrices, considering that each point is a column of the corresponding matrix. When dealing with generalized Procrustes analysis, for shape matching/registration, the problem at hand is extended to include the estimation of a reference shape. This is not the objective of the optimization problem (3). However, once a reference shape is chosen or found (for example using the the classical alternation approach [1, Chapter 9] or the more evolved approaches proposed in [33] and [29]), the problem of finding the Euclidean transformation is exactly expressed by (3) and therefore all the derivations in this and the following sections holds.

4. Uncertainty characterization

The goal of this section is to compute accurate approximations of the uncertainty associated with the extended Procrustes problem solution obtained above. The study of the statistical properties of the orthogonal Procrustes problem can be traced back to [11] and [2], which used perturbation theory to find the approximate distribution of the cost functional $G(\mathbf{R}, \mathbf{t})$, assuming $\boldsymbol{\Sigma}_e^{-1} = \mathbf{I}$, small rotations, and the same covariance matrix for each landmark. These results have been extensively applied to the field of point-based medical image registration problems, as detailed in [13], [14], and [15]. The work in [16] eliminates the assumptions on the rotation, but still addresses the optimization problem without weights and considers that the points are independent and identically distributed, albeit with anisotropic uncertainty. The analysis presented hereafter is based on the aforementioned works and aims at providing approximate uncertainty descriptions for the transformation parameters, $\mathbf{R}^*$ and $\mathbf{t}^*$. This uncertainty characterization is achieved while considering arbitrary rotations and translations and using individual weights as well as individual covariance matrices for each landmark and cross covariance terms.

Within the scope of perturbation theory, the error models of the known variables are defined as

$$\mathbf{a}_i = \mathbf{a}_i^{(0)} + \epsilon\, \mathbf{a}_i^{(1)} + \mathcal{O}(\epsilon^2) \tag{7a}$$

$$\mathbf{b}_i = \mathbf{b}_i^{(0)} + \epsilon\, \mathbf{b}_i^{(1)} + \mathcal{O}(\epsilon^2) \tag{7b}$$

for all $i \in \mathcal{S}$, where ϵ is the smallness parameter, the notation $\mathcal{O}(\epsilon^m)$ stands for the remaining terms of order m or higher, the zero order terms are the

true values, hence $\langle \mathbf{a}_i^{(0)} \rangle = \mathbf{a}_i^{(0)}$ and $\langle \mathbf{b}_i^{(0)} \rangle = \mathbf{b}_i^{(0)}$, whereas the first order terms, $(.)^{(1)}$, are assumed to follow a known distribution with zero mean and covariance matrices defined by $\mathbf{\Sigma}_{a_{ij}} := \langle \mathbf{a}_i^{(1)} \mathbf{a}_j^{(1)^T} \rangle$ and $\mathbf{\Sigma}_{b_{ij}} := \langle \mathbf{b}_i^{(1)} \mathbf{b}_j^{(1)^T} \rangle$, respectively, for all $i, j \in \mathcal{S}$. Note that each of these clouds has cross covariance terms between their points, as suggested in the above covariance matrices expressions. The optimal translation vector is assumed to have an error model with a similar structure to that of (7),

$$\mathbf{t}^* = \mathbf{t}^{(0)} + \epsilon\, \mathbf{t}^{(1)} + \mathcal{O}(\epsilon^2). \tag{8}$$

4.1. Rotation uncertainty

The rotation matrix obtained through the optimization process described before is restricted to the special orthogonal group $\mathrm{SO}(n)$. However, nothing is known about the components of its error model. The following lemma is introduced to address these quantities.

Lemma 2. *Consider the generic error model*

$$\mathbf{M} = \mathbf{M}^{(0)} + \epsilon \mathbf{M}^{(1)} + \mathcal{O}(\epsilon^2). \tag{9}$$

If the matrix $\mathbf{M}$ *belongs to the orthogonal group* $\mathrm{O}(n)$ *then* $\mathbf{M}^{(0)}$ *belongs to the same matrix space, its determinant is equal to the determinant of* $\mathbf{M}$*, and* $\mathbf{M}^{(1)}$ *has the special structure* $\mathbf{M}^{(1)} = \mathbf{S}[\boldsymbol{\omega}]\, \mathbf{M}^{(0)} = \mathbf{M}^{(0)} \mathbf{S}[\boldsymbol{\varpi}]$*, with* $\mathbf{S}[\boldsymbol{\omega}], \mathbf{S}[\boldsymbol{\varpi}] \in \mathfrak{so}(n)$ *and* $\boldsymbol{\omega}, \boldsymbol{\varpi} \in \mathbb{R}^{\frac{n(n-1)}{2}}$*. Furthermore, if* $\mathbf{M} \in \mathrm{SO}(n)$*, then* $\mathbf{M}^{(0)} \in \mathrm{SO}(n)$.

Proof. The proof is made by exploring the algebraic constraints imposed by the matrix spaces $\mathrm{O}(n)$ and $\mathrm{SO}(n)$ and using error model (9). The identity

$\mathbf{M}^T\mathbf{M} = \mathbf{I}$ is expanded, resulting in a unperturbed part that yields $\mathbf{M}^{(0)} \in \mathrm{O}(n)$ and a perturbed part that implies that $\mathbf{M}^{(1)} = \mathbf{S}[\boldsymbol{\omega}]\,\mathbf{M}^{(0)} = \mathbf{M}^{(0)}\mathbf{S}[\boldsymbol{\varpi}]$. Finally, the determinant of $\mathbf{M}$ noting the structure of $\mathbf{S}[\boldsymbol{\omega}]$ is expanded. The proof is presented in Appendix B in the supplementary material. □

From the solution to the optimization problem in [6, Lemma], it is known that $\mathbf{R}^* \in \mathrm{SO}(n)$. If (9) is applied to $\mathbf{R}^*$, the following error model results

$$\mathbf{R}^* = \left[\mathbf{I} + \epsilon\, \mathbf{S}[\boldsymbol{\omega}] + \mathcal{O}(\epsilon^2)\right] \mathbf{R}^{(0)}. \tag{10}$$

To ascertain whether $\mathbf{R}^{(0)} \in \mathrm{SO}(n)$ corresponds to the true rotation, consider that matrix $\mathbf{B}$, used to compute the estimated rotation, can be described in terms of its error model, using that of matrices $\mathbf{X}$ and $\mathbf{Y}$, which are a generalization of (7b) and (7a), respectively. This results in

$$\mathbf{B} = \mathbf{B}^{(0)} + \epsilon\, \mathbf{B}^{(1)} + \mathcal{O}(\epsilon^2), \tag{11}$$

with $\mathbf{B}^{(0)} = \mathbf{X}^{(0)}\mathbf{W}\mathbf{Y}^{(0)^T}$ and $\mathbf{B}^{(1)} = \mathbf{X}^{(1)}\mathbf{W}\mathbf{Y}^{(0)^T} + \mathbf{X}^{(0)}\mathbf{W}\mathbf{Y}^{(1)^T}$. Furthermore, the singular value decomposition (4) of $\mathbf{B}$ can itself be expanded defining similar models for each of its components [39], i.e., $\mathbf{U} = \mathbf{U}^{(0)} + \epsilon\mathbf{U}^{(1)} + \mathcal{O}(\epsilon^2)$, $\mathbf{D} = \mathbf{D}^{(0)} + \epsilon\mathbf{D}^{(1)} + \mathcal{O}(\epsilon^2)$, and $\mathbf{V} = \mathbf{V}^{(0)} + \epsilon\mathbf{V}^{(1)} + \mathcal{O}(\epsilon^2)$, which yields

$$\mathbf{B} = \mathbf{U}^{(0)}\mathbf{D}^{(0)}\mathbf{V}^{(0)^T} + \mathcal{O}(\epsilon).$$

Comparison with the error model (11) yields $\mathbf{X}^{(0)}\mathbf{W}\mathbf{Y}^{(0)^T} = \mathbf{U}^{(0)}\mathbf{D}^{(0)}\mathbf{V}^{(0)^T}$, thus confirming that these terms are composed of true quantities only. Combining the error model of the optimal rotation matrix with its solution (5) yields

$$\mathbf{R}^{(0)} + \mathcal{O}(\epsilon) = \mathbf{U}^{(0)}\,\mathrm{diag}\left(1, \ldots, 1, \left|\mathbf{U}^{(0)}\right|\left|\mathbf{V}^{(0)}\right|\right)\mathbf{V}^{(0)^T} + \mathcal{O}(\epsilon).$$

Note that, as $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices, Lemma 2 applies, and, therefore, not only $\mathbf{U}^{(0)}$ and $\mathbf{V}^{(0)}$ are orthogonal matrices, but their determinants correspond to the determinants of the perturbed versions, a fact that was used in the last expression. From this, it follows that $\mathbf{R}^{(0)}$ is the true rotation as it is obtained from the singular value decomposition of $\mathbf{B}^{(0)}$, itself computed using the true terms $\mathbf{a}_i^{(0)}$ and $\mathbf{b}_i^{(0)}$, for all $i \in \{1, \ldots, m\}$. As $\mathbf{R}^{(0)}$ is the true rotation and given the structure of (10), the vector $\boldsymbol{\omega}$ can be seen as a rotation error around the true rotation matrix.

From the proof of [6, Lemma], it is known that the matrix $\mathbf{B}\mathbf{R}^*$ is symmetric, and thus, using the error models (10) and (11) one can write,

$$\operatorname{skew}\left(\mathbf{B}\mathbf{R}^*\right) = \operatorname{skew}\left(\mathbf{B}^{(0)}\mathbf{R}^{(0)}\right) + \epsilon \operatorname{skew}\left(\mathbf{B}^{(1)}\mathbf{R}^{(0)} + \mathbf{B}^{(0)}\mathbf{S}[\boldsymbol{\omega}]\,\mathbf{R}^{(0)}\right) = \mathbf{0}.$$

This implies that each skew operator is null, meaning that $\mathbf{B}^{(0)}\mathbf{R}^{(0)}$ is also symmetric and that, after left multiplication by $\mathbf{R}^{(0)}$, right multiplication by ${\mathbf{R}^{(0)}}^T$, and using the identity $\mathbf{S}[\boldsymbol{\omega}] = -\mathbf{S}^T[\boldsymbol{\omega}]$,

$$\operatorname{skew}\left(\mathbf{R}^{(0)}\mathbf{B}^{(0)}\mathbf{S}^T[\boldsymbol{\omega}]\right) = \operatorname{skew}\left(\mathbf{R}^{(0)}\mathbf{B}^{(1)}\right). \tag{12}$$

Further manipulation of this formula is needed in order to obtain a tractable expression for $\boldsymbol{\omega}$. Consider the matrix $\mathbf{R}^{(0)}\mathbf{B}^{(0)}\mathbf{S}^T[\boldsymbol{\omega}]$ expressed as a summation of point terms, noting that $\mathbf{a}_i^{(0)} = \mathbf{r}_i^{(0)} + \mathbf{t}^{(0)}$ with $\mathbf{r}_i^{(.)} := \mathbf{R}^{(0)}\mathbf{b}_i^{(.)}$ for all $i \in \mathcal{S}$,

$$\mathbf{R}^{(0)}\mathbf{B}^{(0)}\mathbf{S}^T[\boldsymbol{\omega}] = \sum_{i=1}^{m} \sigma_i^{-2} \left[\mathbf{r}_i^{(0)}{\mathbf{r}_i^{(0)}}^T\mathbf{S}^T[\boldsymbol{\omega}] - \frac{1}{N_W}\sum_{j=1}^{m} \sigma_j^{-2}\mathbf{r}_i^{(0)}{\mathbf{r}_j^{(0)}}^T\mathbf{S}^T[\boldsymbol{\omega}]\right], \tag{13}$$

The same can be done with $\mathbf{R}^{(0)}\mathbf{B}^{(1)}$, yielding

$$\mathbf{R}^{(0)}\mathbf{B}^{(1)} = \sum_{i=1}^{m} \sigma_i^{-2}\left[\mathbf{r}_i^{(0)}{\bar{\mathbf{a}}_i^{(1)}}^T + \mathbf{r}_i^{(1)}{\bar{\mathbf{a}}_i^{(0)}}^T\right], \tag{14}$$

where $\bar{\mathbf{a}}_i^{(.)} := \mathbf{a}_i^{(.)} - \boldsymbol{\mu}_A^{(.)}$.

Due to the skew-symmetric nature of (12), it is possible to apply the unskew operator to both sides. This enables to apply Properties (4) and (2) of Lemma 1 combined with (13) and (14), yielding

$$\begin{aligned}\mathbf{S}^{-1}\Big(2\,\mathrm{skew}\left(\mathbf{R}^{(0)}\mathbf{B}^{(0)}\mathbf{S}^{T}[\boldsymbol{\omega}]\right)\Big) = \frac{1}{N_W}\sum_{i,j=1}^{m}\sigma_i^{-2}\sigma_j^{-2}\bar{\mathbf{S}}\left[\mathbf{r}_i^{(0)}\right]\mathbf{S}[\boldsymbol{\omega}]\,\mathbf{r}_j^{(0)} \\ - \sum_{i=1}^{m}\sigma_i^{-2}\bar{\mathbf{S}}\left[\mathbf{r}_i^{(0)}\right]\mathbf{S}[\boldsymbol{\omega}]\,\mathbf{r}_i^{(0)},\end{aligned} \tag{15}$$

whereas the right-hand side becomes

$$\mathbf{c} := \mathbf{S}^{-1}\Big(2\,\mathrm{skew}\left(\mathbf{R}^{(0)}\mathbf{B}^{(1)}\right)\Big) = \sum_{i=1}^{m}\sigma_i^{-2}\left(\bar{\mathbf{S}}\left[\bar{\mathbf{a}}_i^{(0)}\right]\mathbf{r}_i^{(1)} - \bar{\mathbf{S}}\left[\mathbf{r}_i^{(0)}\right]\bar{\mathbf{a}}_i^{(1)}\right). \tag{16}$$

Using Property (3) of Lemma 1 in (15), one obtains a linear matrix equation after applying the unskew operator to both sides of (12), given by $\mathbf{A}\boldsymbol{\omega} = \mathbf{c}$, where the matrix $\mathbf{A} \in \mathbb{R}^{n_p \times n_p}$ is defined as

$$\mathbf{A} := -\sum_{i=1}^{m}\sigma_i^{-2}\bar{\mathbf{S}}\left[\mathbf{r}_i^{(0)}\right]\bar{\mathbf{S}}^{T}\left[\mathbf{r}_i^{(0)}\right] + \frac{1}{N_W}\sum_{i,j=1}^{m}\sigma_i^{-2}\sigma_j^{-2}\bar{\mathbf{S}}\left[\mathbf{r}_i^{(0)}\right]\bar{\mathbf{S}}^{T}\left[\mathbf{r}_j^{(0)}\right].$$

From the linear equation now derived it is possible to obtain $\boldsymbol{\omega}$, as long as $\mathbf{A}$ is invertible. It can be shown that this result degenerates into the main result of [16] when assuming independent and identically distributed points in $\mathbb{R}^3$. However, the existence and uniqueness of the solution is not addressed there. That is the focus of the forthcoming theorem, addressing the conditions under which the invertibility of $\mathbf{A}$ in n-dimensional space holds.

Theorem 1. *The matrix* $\mathbf{A}$ *is invertible if and only if there are at least* n *points in the set* $\mathcal{S}_B$ *whose connecting vectors span* $\mathbb{R}^{n-1}$.

Proof. The proof is made by manipulating $\mathbf{u}^T\mathbf{A}\mathbf{u} = 0$ so that a clear relation between the landmarks and $\mathbf{u}$ appears, allowing to explore the nature of $\mathbf{S}[.]$ to analyse the solutions of $\mathbf{u}^T\mathbf{A}\mathbf{u} = 0$ for $\|\mathbf{u}\| = 1$ when $\mathbf{A}$ is singular. It can be found in Appendix C in the supplementary material. □

By observation of (16), it can be seen that $\langle \mathbf{c} \rangle = 0$, as its uncertain elements are only $\mathbf{r}_i^{(0)}$ and $\bar{\mathbf{a}}_i^{(0)}$, which have zero mean. Therefore $\langle \boldsymbol{\omega} \rangle = \mathbf{A}^{-1}\langle \mathbf{c} \rangle$ is also zero, and the corresponding covariance matrix is given by

$$\begin{aligned}\boldsymbol{\Sigma}_\omega = \langle \boldsymbol{\omega}\boldsymbol{\omega}^T \rangle =& \mathbf{A}^{-1}\left[\sum_{i,j=1}^{m} \sigma_i^{-2}\sigma_j^{-2}\bar{\mathbf{S}}\left[\bar{\mathbf{a}}_i^{(0)}\right]\mathbf{R}^{(0)}\boldsymbol{\Sigma}_{b_{ij}}\mathbf{R}^{(0)^T}\bar{\mathbf{S}}^T\left[\bar{\mathbf{a}}_j^{(0)}\right]\right.\\ &+ \sum_{i,j=1}^{m} \sigma_i^{-2}\sigma_j^{-2}\bar{\mathbf{S}}\left[\mathbf{r}_i^{(0)}\right]\langle \bar{\mathbf{a}}_i^{(1)}\bar{\mathbf{a}}_j^{(1)^T}\rangle\bar{\mathbf{S}}\left[\mathbf{r}_j^{(0)}\right]\\ &- \sum_{i,j=1}^{m} \sigma_i^{-2}\sigma_j^{-2}\bar{\mathbf{S}}\left[\bar{\mathbf{a}}_i^{(0)}\right]\langle \mathbf{r}_i^{(1)}\bar{\mathbf{a}}_j^{(1)^T}\rangle\bar{\mathbf{S}}\left[\mathbf{r}_j^{(0)}\right]\\ &\left.- \sum_{i,j=1}^{m} \sigma_i^{-2}\sigma_j^{-2}\bar{\mathbf{S}}\left[\mathbf{r}_i^{(0)}\right]\langle \bar{\mathbf{a}}_i^{(1)}\mathbf{r}_j^{(1)^T}\rangle\bar{\mathbf{S}}\left[\bar{\mathbf{a}}_j^{(0)}\right]\right]\mathbf{A}^{-1^T}.\end{aligned}$$

The individual covariances between points of the two sets are detailed in Appendix D of the supplementary material.

4.2. Translation uncertainty

The optimal translation between frames is given by (6) and the associated error model is assumed to be (8). Using this information along with the error models for the points in both sets, defined in (7), it is possible to expand (6) to obtain expressions for the error model components, the true translation $\mathbf{t}^{(0)} = \frac{1}{N_W}\sum_{i=1}^{m} \sigma_i^{-2}\left[\mathbf{a}_i^{(0)} - \mathbf{R}^{(0)}\mathbf{b}_i^{(0)}\right]$, and the first order perturbed part

$$\mathbf{t}^{(1)} = \frac{1}{N_W}\sum_{i=1}^{m} \sigma_i^{-2}\left(\mathbf{a}_i^{(1)} - \mathbf{S}[\boldsymbol{\omega}]\,\mathbf{r}_i^{(0)} - \mathbf{r}_i^{(1)}\right). \tag{17}$$

It is confirmed that $\mathbf{t}^{(1)}$ has zero mean, noting that all the perturbation quantities involved have zero mean. Therefore, the covariance matrix of the position estimate $\boldsymbol{\Sigma}_t$ is simply given by $\boldsymbol{\Sigma}_t = \langle \mathbf{t}^{(1)} \mathbf{t}^{(1)^T} \rangle$, and it can be computed by expanding $\mathbf{t}^{(1)}$ according to (17), using Property (3) of Lemma 1 to extract $\boldsymbol{\omega}$ from the skew-symmetric matrix. This yields

$$\begin{aligned}\boldsymbol{\Sigma}_t = \sum_{i,j=1}^{m} \frac{\sigma_i^{-2}\sigma_j^{-2}}{N_W^2} \Big(& \boldsymbol{\Sigma}_{a_{ij}} + \mathbf{R}^{(0)} \boldsymbol{\Sigma}_{b_{ij}} \mathbf{R}^{(0)^T} + \bar{\mathbf{S}}^T \Big[\mathbf{r}_i^{(0)}\Big] \boldsymbol{\Sigma}_\omega \bar{\mathbf{S}} \Big[\mathbf{r}_j^{(0)}\Big] \\ & - \bar{\mathbf{S}}^T \Big[\mathbf{r}_i^{(0)}\Big] \langle \boldsymbol{\omega} \mathbf{a}_j^{(1)^T} \rangle - \langle \mathbf{a}_i^{(1)} \boldsymbol{\omega}^T \rangle \bar{\mathbf{S}} \Big[\mathbf{r}_j^{(0)}\Big] - \mathbf{R}^{(0)} \boldsymbol{\Sigma}_{b_i a_j} \\ & + \bar{\mathbf{S}}^T \Big[\mathbf{r}_i^{(0)}\Big] \langle \boldsymbol{\omega} \mathbf{r}_j^{(1)^T} \rangle + \langle \mathbf{r}_i^{(1)} \boldsymbol{\omega}^T \rangle \bar{\mathbf{S}} \Big[\mathbf{r}_j^{(0)}\Big] - \boldsymbol{\Sigma}_{a_i b_j} \mathbf{R}^{(0)^T} \Big),\end{aligned}$$

4.3. Cross translation-rotation uncertainty

The optimal translation estimate is obtained using the optimal rotation, and for that reason there is a correlation between the two. The corresponding cross covariance is given by

$$\begin{aligned}\boldsymbol{\Sigma}_{\omega t} = & \mathbf{A}^{-1} \sum_{i,j=1}^{m} \frac{\sigma_i^{-2}\sigma_j^{-2}}{N_W} \bar{\mathbf{S}} \Big[\bar{\mathbf{a}}_i^{(0)}\Big] \Big(\mathbf{R}^{(0)} \boldsymbol{\Sigma}_{b_i a_j} - \langle \mathbf{r}_i^{(1)} \boldsymbol{\omega}^T \rangle \bar{\mathbf{S}} \Big[\mathbf{r}_j^{(0)}\Big] - \mathbf{R}^{(0)} \boldsymbol{\Sigma}_{b_i j} \mathbf{R}^{(0)^T} \Big) \\ & + \mathbf{A}^{-1} \sum_{i,j=1}^{m} \frac{\sigma_i^{-2}\sigma_j^{-2}}{N_W} \bar{\mathbf{S}} \Big[\mathbf{r}_i^{(0)}\Big] \Big(\boldsymbol{\Sigma}_{a_i b_j} \mathbf{R}^{(0)^T} + \langle \bar{\mathbf{a}}_i^{(1)} \boldsymbol{\omega}^T \rangle \bar{\mathbf{S}} \Big[\mathbf{r}_j^{(0)}\Big] - \langle \bar{\mathbf{a}}_i^{(1)} \mathbf{a}_j^{(1)^T} \rangle \Big),\end{aligned}$$

where the cross covariances between points in either sets and the rotation error are computed in the same manner, and are detailed in Appendix D in the supplementary material. Note that $\boldsymbol{\Sigma}_{a_i b_j}$ may in most cases be zero due to the possible independence of the sets.

Remark 1. *As the true quantities* $\mathbf{R}^{(0)}$, $\mathbf{a}_i^{(0)}$, *and* $\mathbf{b}_i^{(0)}$ *are unknown for all* $i \in \mathcal{S}$, *a possible approximation is to use the optimal values for the rotation and translation and the original perturbed points* $\mathbf{a}_i$ *and* $\mathbf{b}_j$ *instead. This applies to all the covariance expressions derived in this section.*

5. Algorithm validation

In this section, the proposed uncertainty characterization is validated through extensive Monte Carlo simulations. Despite the general derivation presented above, the numerical validation requires the particularization for one or several dimensions. Although bidimensional and tridimensional problems are the most common in the literature, as in [24] and [25], the study of rotations and other quantities in seven-dimensional space has also been addressed in [40]. For these reasons and to represent meaningful examples, this section will focus on those dimensionalities.

As the optimization and uncertainty description presented in this paper are intended to work for any rotation and translation, and taking into account that some approximations were introduced, a Monte Carlo simulation with $N = 1000$ samples was performed for each of $M = 500$ random configurations, to validate the proposed technique. In each of these configurations, the transformation is uniformly distributed, with each component of the translation vector ranging from -10 to 10 m, the locations of the points from 0 and 10 m in each component, and the rotation matrix is built using the exponential map $\mathbf{R} = \mathrm{e}^{\mathbf{S}[\theta\mathbf{u}]}$, where θ is uniformly distributed between $-\pi$ and π radians and $\mathbf{u}$ is an uniformly distributed random unit vector. In order to comprehensively test the properties of the proposed method, the covariance matrices used to generate the normally distributed location errors of the points in each set are complete covariances with cross-correlations between all the points and different individual covariances. In the two-dimensional case, the eigenvalues of the covariance matrices are $(0.01^2, 0.05^2)$. For the three-dimensional case, the

eigenvalues are $(0.01^2, 0.03^2, 0.05^2)$, and the seven-dimensional eigenvalues are $(0.01^2, 0.017^2, 0.023^2, 0.03^2, 0.037^2, 0.043^2, 0.05^2)$. In each of these cases, all the permutations of eigenvalues are used an equal number of times. The full covariance for each point cloud is then $\mathbf{\Sigma}^{nom} = \mathbf{U}\,\mathrm{diag}\,(\lambda_1, \ldots, \lambda_{nm})\,\mathbf{U}^T$, where the eigenvector matrix $\mathbf{U}$ is an $nm \times nm$ rotation matrix built following the same reasoning used to build $\mathbf{R}$.

The meaningful comparison between two different models, usually denoted as the null model and the alternate model, is a difficult problem which is frequently performed using likelihood ratio tests. In short, the likelihood ratio expresses how many times more likely the data is explained under one model than the other, by obtaining the probability distribution of the test, assuming the null model to be true (null hypothesis). Thus, the null hypothesis is rejected if its probability is below a desired significance value, typically 0.01 or 0.05. However, a central limitation for the usage of these tests is the assumption that the stochastic variable follows a specific distribution, usually Gaussian. Indeed, and as a consequence of the considered arbitrary rotations, the translation tends to be Gaussian distributed as the number of points increases, but may fail to be so for a reduced number of landmarks, which is in accordance with the central limit theorem.

Figure 1 and Tables 1 and 2 summarize the likelihood ratio tests for the comparison between the sample covariance matrices resulting from the Monte Carlo simulations, $\mathbf{\Sigma}_{\omega}^{sim}$, $\mathbf{\Sigma}_{t}^{sim}$, and the joint error covariance $\mathbf{\Sigma}_{\xi}^{sim}$ where $\boldsymbol{\xi} = \left[\boldsymbol{\omega}^T \mathbf{t}^T\right]^T$, and the covariance matrices resulting from the proposed methodology and respective uncertainty approximation, $\mathbf{\Sigma}_{\omega}^{opt}$, $\mathbf{\Sigma}_{t}^{opt}$, and $\mathbf{\Sigma}_{\xi}^{opt}$. The definition of each ratio depends on whether the quantity is a scalar or a

Table 1: Worst-case covariance likelihood ratio tests for $n = 2$ (left) and $n = 3$ (right)

m	H_ω [%]	H_t [%]	H_ξ [%]	H_ω^{100} [%]	H_t^{100} [%]	H_ξ^{100} [%]	H_ω [%]	H_t [%]	H_ξ [%]	H_ω^{100} [%]	H_t^{100} [%]	H_ξ^{100} [%]
2	77.2	27.8	29.2	100	71.4	73.2						
3	94.4	51.8	55.4	100	90.2	90.4	67.6	13.4	9.8	95.4	60.0	60.2
4	96.6	65.4	70.8	100	96.2	95.4	88.4	32.8	27.8	98.0	84.8	85.8
5	96.8	79.4	82.8	100	97.6	97.8	93.8	50.0	47.4	98.8	93.8	93.6
7	96.4	87.6	89.2	100	97.0	97.0	96.0	76.6	75.2	99.0	97.8	98.8
10	98.8	93.2	94.8	100	98.2	98.4	98.2	92.0	91.2	97.8	96.8	97.2
20	97.4	95.8	96.6	100	99.0	98.8	97.8	96.6	96.0	99.0	97.8	97.6
50	98.2	98.6	98.4	100	99.6	99.2	98.8	98.0	97.8	99.2	99.0	98.2
100	99.0	98.4	98.6	100	99.0	99.6	98.4	98.0	97.6	98.6	99.4	99.4

Table 2: Worst-case covariance likelihood ratio tests for $n = 7$

m	H_ω [%]	H_t [%]	H_ξ [%]	H_ω^{100} [%]	H_t^{100} [%]	H_ξ^{100} [%]
7	61.4	1.6	0.0	84.4	32.2	7.4
10	95.2	28.6	9.0	89.0	89.8	56.4
15	98.0	74.6	65.2	93.0	97.2	70.2
20	98.2	90.2	86.0	90.8	98.2	73.4
30	98.4	96.6	94.0	93.0	98.6	75.8
50	98.6	97.6	97.4	93.4	98.4	76.0
100	98.4	97.8	98.2	93.0	99.0	79.6

matrix. For $n = 2$, the rotation error is a scalar, and therefore its covariance will be too. Then, the ratio for the null hypothesis $H_\omega^{n=2} : \Sigma_\omega^{sim} = \Sigma_\omega^{opt}$ is defined as

$$\lambda_\omega^* := N^* \frac{\Sigma_\omega^{sim}}{\Sigma_\omega^{opt}} \sim \chi^2_{N^*}$$

which is asymptotically χ^2 distributed with $N^* := N - 1$ degrees of freedom and, thus, can be compared with a predefined significance level, which in the presented results is considered to be 0.01. For the multivariable cases of the null hypothesis of the general case for rotation $H_\omega^{n\geq 3} : \mathbf{\Sigma}_\omega^{sim} = \mathbf{\Sigma}_\omega^{opt}$, the translation $H_t : \mathbf{\Sigma}_t^{sim} = \mathbf{\Sigma}_t^{opt}$ and the joint error covariance $H_\xi : \mathbf{\Sigma}_\xi^{sim} = \mathbf{\Sigma}_\xi^{opt}$, the specific likelihood ratio test can be found in [41, Section 10.8]. The

likelihood ratio for these cases is defined as

$$\lambda_1^* = \left(\frac{\mathrm{e}}{N^*}\right)^{\frac{1}{2}pN^*} \left|\mathbf{B}^{sim}\boldsymbol{\Sigma}^{opt\,-1}\right|^{\frac{1}{2}N^*} \mathrm{e}^{-\frac{1}{2}\,\mathrm{tr}\left(\mathbf{B}^{sim}\boldsymbol{\Sigma}^{opt\,-1}\right)},$$

where $\mathbf{B}^{sim} := N^*\boldsymbol{\Sigma}^{sim}$, and the comparison with the significance level is achieved by noting that $-2\log\lambda_1^*$ is asymptotically χ^2 distributed with $\frac{1}{2}p(p+1)$ degrees of freedom where p is the dimension of the vector (n for the translation, n_p for the rotation error, and $n+n_p$ for the joint error vector $\boldsymbol{\xi}$). Hence, for the likelihood ratio tests, λ_ω^* is compared with a threshold $t_\omega^{n=2} := F^{-1}_{\chi^2_{N^*}}(1-\alpha)$ in the bidimensional case, and the remaining likelihood tests are made by comparing $-2\log\lambda_1^*$ with similar thresholds $t_\omega^{n\geq 3} := F^{-1}_{\chi^2_{n_p}}(1-\alpha)$, $t_t := F^{-1}_{\chi^2_n}(1-\alpha)$ and $t_\xi := F^{-1}_{\chi^2_{n_p+n}}(1-\alpha)$, all with $\alpha = 0.01$.

As the true values of the coordinates of the points and of the rotation matrix are not available in real-world scenarios, the covariances resulting from the uncertainty description of the previous section must be computed using the perturbed quantities instead, therefore depending on the actual values of the simulation. The covariances and respective likelihood ratios are computed for each sample of the Monte Carlo simulations, that is, for each configuration, one $\boldsymbol{\Sigma}^{sim}$ and N instances of $\boldsymbol{\Sigma}^{opt}$ are computed, resulting in $N \times M$ covariances, and consequently, the same quantity of ratios and tests. The highest ratio in each configuration is used for the hypothesis tests presented in Tables 1 and 2, thus representing the worst-case scenarios (in terms of likelihood ratios) for each configurations and number of landmarks. The values shown in the tables for these tests denote the percentage of successful tests among the M different configurations for $n = 2$, $n = 3$, and $n = 7$, respectively. As mentioned before, the Gaussian assumption may not be valid when dealing with an arbitrary rotation $\mathbf{R}$, which, together with the fact

that the ratio expressions and the degrees of freedom of their approximation depend on the number of Monte Carlo samples, the tests are expected to fail more pronouncedly when the number of Monte Carlo samples increases. For this reason, an additional set of likelihood ratio tests are presented using a smaller number of samples, $N = 100$, and denoted as H^{100}. From the values shown in these tables, it can be concluded that, for more than 5 points in 2-D, 7 points in 3-D, and 15 points in 7-D, the obtained covariance matrices pass more than 70% of the tests in the worst-case, and can therefore be considered a good approximation.

The cumulative distributions of the ratios depicted in Figure 1, however, show that the worst-case scenario in terms of likelihood ratio is not representative of the majority of the computed covariance matrices. The left side of this figure presents the cumulative distribution of the total $N \times M$ likelihood ratios, normalized by the respective thresholds, and therefore depict the quantity of samples that is explained by the model according to the null hypothesis (area to the left of the vertical dashed line) as well as its variation with the number of available points for 2-D. Comparing with the worst cases provided in the previous tables, it can be seen that when using the covariances built with every sample in each configuration the results are significantly better even when a low ratio of points per dimension is present. In particular, even with the minimum amount of points ($m/n = 1$), it is possible to achieve more than 70% of positive rotation and translation tests for all the dimensions except for $n = 7$, where 10 points are needed to surpass that threshold. In fact, in most cases the growth of the cumulative function is quite fast, reaching high values of cumulative percentage for rel-

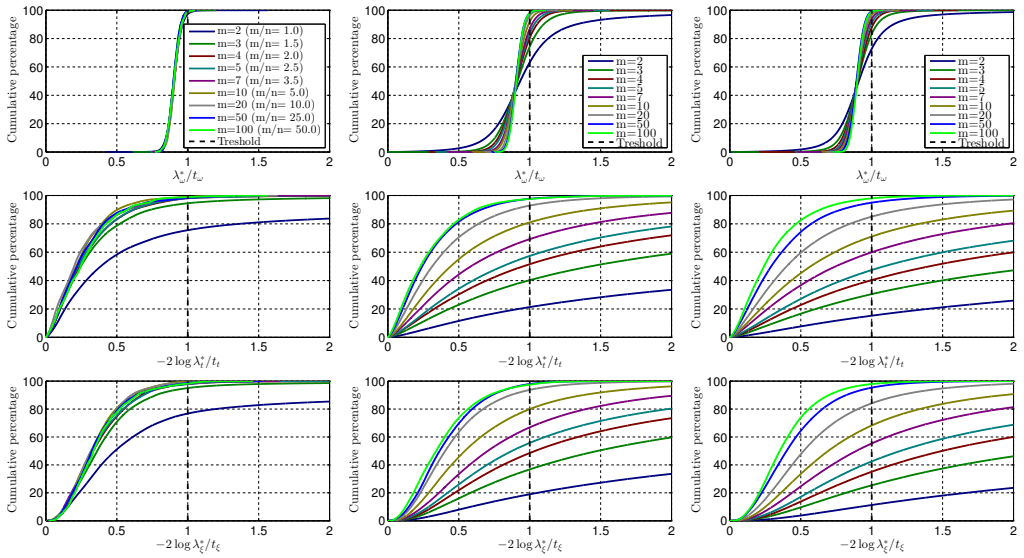

Figure 1: The cumulative distribution of the likelihood ratios for the 2-D case including all the covariances built with each particular simulation values and different input noise covariance profiles (profile 1 in the left, 3 in the center, and 4 in the right columns). Rotation error $\boldsymbol{\omega}$ (top row), translation error (center row) and joint error vector (bottom row).

atively low values of the ratio. It is also confirmed that adding a few points to the $m/n = 1$ case leads to a relevant improvement of the test results for all the three n, noting that even the joint error ratio displays success values above 90% for each dimensionality, respectively with $m = 3$, $m = 5$, and $m = 10$. Nonetheless, the $m/n = 1$ test results are degraded with the increase of n (from 79% in 2-D to 64% in 3-D and 22% in 7-D for the joint error), a tendency less apparent for the tests with $m/n \geq 2$. To better explore the various aspects of the uncertainty characterization, and achieve a broader validation, additional simulations for $n = 2$ were performed with several noise scales, namely: 1) nominal noise; 2) $\mathcal{S}_B$ noise scaled by 10^2; 3) $\mathcal{S}_A$ noise scaled by 10^2; and 4) both multiplied by 10^2 which are depicted in Figure 1. Profiles 2 and 3 are very similar and only one is shown. One im-

portant fact to note prior to this analysis is that, as the covariances are built with the noise-perturbed point sets and the estimated rotation, increasing the noise will inevitably lead to worse test results in general. Notwithstanding this fact, for a reasonable number of points ($m = 10$) more than 70% of the samples pass the test for all the quantities tested in all 4 profiles. Regardless of the noise profile, the rotation error covariances perform much better than the translation in these tests, as 3 points are enough to have more than 70% of positive tests for the rotation, while for the translation 7 points are needed for profiles 2 and 3, and 10 points for the last profile. Furthermore, with 2 points the performance may be below 20% in the translation for profile 4, while the rotation is never below 60% in any profile. Even though the left and center columns of Figure 1 are very similar, the translation tests are worse in the former (around 1.4% less for each m), when the higher noise is on $\mathcal{S}_B$, while the rotation tests are better (around 0.7% higher for each m). This may be explained by the fact that, in the mathematical expressions derived in Section 4.2, the actual values of the points in that set are always multiplied by the rotation, hence amplifying the influence of the noise in that set. When both input covariances are scaled, the results worsen as every distribution crosses the threshold with around 10% less samples than in profiles 2 and 3. Note however that in realistic situations, these noise levels would imply that 99% of the noisy points could be in a ellipse with a major axis of 1.5 meters around the true value (3σ) which is not a usual situation. Nevertheless, to further assess the influence of the noise in the tests, new simulations were made where both nominal input covariances are multiplied by 100^2. In these tests, the rotation still performs adequately (60% for 20

points, 80% for 50 points), but the translation is far from that level (with 100 points only 3.37% of the samples pass the test). Nevertheless, when the true values are used to compute the covariances in these harsh noise conditions the performance is much better (55% for rotation, 44% for translation, and 12% for the joint error vector, all for $m = 2$), confirming that it is also the fact that the covariances must be built with the perturbed values that leads to the worsening of the results. This is intrinsic to the nonlinear problem at hand, and not specific to the proposed characterization.

To provide a better understanding of the underlying problem, Figure 2 is provided, depicting the spatial distribution of the translation error (red dots) with the corresponding 99% bound in dashed black, along with 3σ bounds given by the median ratio covariance matrix in solid blue, and the simulation covariance in dashed green. Here a difficult case is presented, where all the points in $\mathcal{S}_B$ are evenly distributed along the x-axis, there is a translation of 2 units in each axis and a 180° rotation between sets, with narrow input covariances on both sets, $\mathbf{\Sigma}_b = \mathbf{\Sigma}_a = \text{diag}\left(10^{-5}, 9 \times 10^{-3}, 4 \times 10^{-5}, 16 \times 10^{-3}\right)$ for each two points. The baseline (Δ), i.e., the maximum distance between points in $\mathcal{S}_B$, increases from the leftmost figure to the rightmost. When moving downwards, the number of equidistant points increases, maintaining the baseline. A careful analysis of the information contained in Figure 2 indicates that the baseline is the dominant factor that can lead to highly nonlinear uncertainty distributions. As these nonlinearities are not captured by the proposed uncertainty computation, its performance will be degraded in difficult cases. Although a short baseline leads to highly nonlinear error distributions, as the number of points increases these distributions become

more conventional, although not as notorious as with the increase of the baseline. In fact, even with only 2 points, a baseline of 4 meters leads to a median covariance quite similar to the experimental covariance, and conservative to the actual shape of the 99% simulation bound.

In a further simulation, the rotation, translation and the shape of the input noise covariance (narrow versus round) were varied for only 2 points with a fixed baseline. It was observed that the influence of the rotation and translation is not evidently visible in the results, further supporting that the baseline, the number of points, and the shape and size of the input noise distributions are the most significant parameters in both the results of the optimization problem and the validity of the uncertainty characterization. In fact, these two aspects of the whole problem are closely related: the uncertainty characterization here proposed is not adequate when the distribution of the error cannot be sufficiently characterized by its first two moments, mean and covariance, even though it demonstrates great accuracy when the error distribution is more conventional. In sum, narrow shapes of input noise, low baselines, and low number of points lead to non-conventional shapes of error distributions, and bettering any of these three parameters leads to more natural error distributions, especially for the case of the baseline. Even though this places restrictions on the baseline and number of points available for the optimization, it is by no means a very harsh limitation, as in the vast majority of applications it is common practice to use a relatively large number of points and reasonable baselines when compared to noise levels.

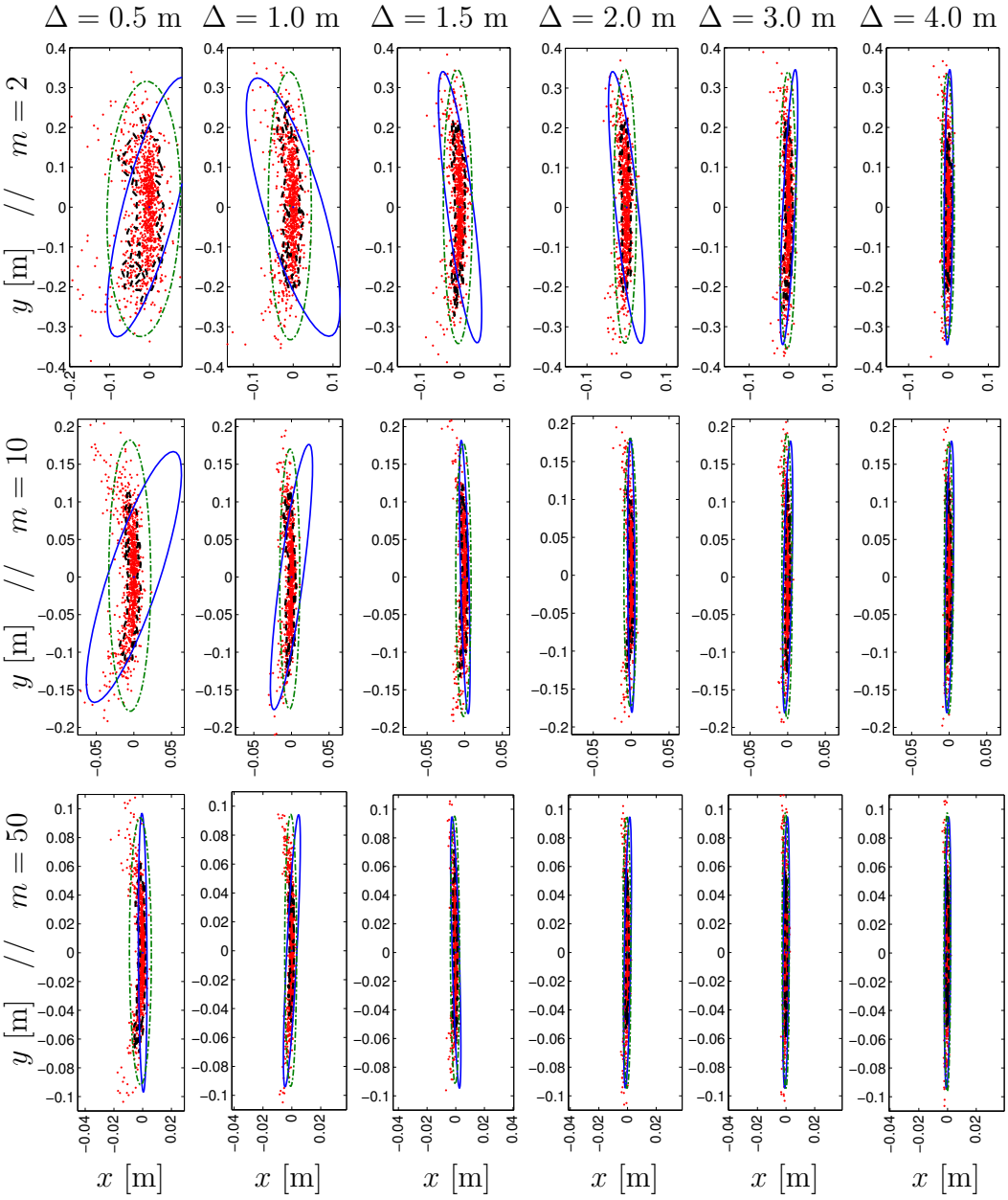

Figure 2: Translation error and uncertainty for a series of combinations of baseline-number of points. Translation error Monte Carlo samples in red, 99% bound in dashed black, 3σ simulation covariance ellipse in dashed green, and the 3σ covariance ellipse resulting from the uncertainty characterization with the median likelihood ratio. The baseline ($\Delta = \max \|\mathbf{b}_i - \mathbf{b}_j\|$ for all $i \neq j$, $i, j \in \mathcal{S}$) increases in each figure from left to right, whereas the number of points used in the optimization increases from top to down. Note that the scaling changes from line to line as the number of points increases (ellipses in the bottom are smaller than those at the top).

6. Conclusion

In this paper, the uncertainty involved in the weighted orthogonal Procrustes problem for stochastically perturbed n-dimensional point clouds was studied thoroughly and analytical expressions were derived for the first and second moments of the stochastic outputs, the translation and rotation, as well as cross terms that characterize the anisotropic uncertainty of the problem, not imposing assumptions on the actual rotation and translation. These were obtained after assuming an error model based on perturbation theory for each point in the point clouds, and advancing similar error models both for the translation and the rotation matrix. This novel uncertainty characterization for the Procrustes problem was validated through extensive Monte Carlo simulations exploring the general framework of the algorithm, using arbitrary rotations and translations, as well as full covariance matrices for each point. For a relatively low number of points, and reasonable noise conditions, the proposed uncertainty characterization performed very well in likelihood ratio tests that encompass a wide variety of configurations varying in translation, rotation, and spatial distribution of points. A thorough analysis of the influence of several parameters on the error distribution and the validity of the results was performed, resulting in the conclusion that with a reasonable baseline the uncertainty characterization performs accurately. These results are consistent with the assumptions, and further validate this approach.

Possible directions of future work include applying this methodology to similar optimization problems (e.g. LiDAR calibration from 3-D point clouds [42]), and research on the use of the uncertainty characterization in iterative Procrustes or ICP algorithms.

Acknowledgments

This work was supported by the Fundação para a Ciência e a Tecnologia (FCT) through ISR under LARSyS UID/EEA/50009/2013, and through IDMEC, under LAETA UID/EMS/50022/2013 contracts, by the University of Macau Project MYRG2015-00126-FST, and by the Macao Science and Technology Development Fund under Grant FDCT/048/2014/A1. The work of P. Lourenço and B. Guerreiro was supported, respectively, by the PhD. Grant SFRH/BD/89337/2012 and by the Post-Doc grant SFRH/BPD/110416/2015 from FCT.

[1] J. C. Gower, G. B. Dijksterhuis, Procrustes Problems, Oxford Statistical Science Series, Oxford University Press, 2004.

[2] C. Goodall, Procrustes Methods in the Statistical Analysis of Shape, Journal of the Royal Statistical Society. Series B (Methodological) 53 (2) (1991) 285–339.

[3] P. H. Schönemann, A generalized solution of the orthogonal procrustes problem, Psychometrika 31 (1) (1966) 1–10.

[4] G. Wahba, Problem 65-1: A least squares estimate of satellite attitude, SIAM Review 7 (3) (1965) 409.

[5] W. Kabsch, A solution for the best rotation to relate two sets of vectors, Acta Crystallographica Section A 32 (5) (1976) 922–923.

[6] S. Umeyama, Least-squares estimation of transformation parameters between two point patterns, IEEE Transactions On Pattern Analysis and Machine Intelligence 13 (4) (1991) 376–380.

[7] B. K. P. Horn, H. Hilden, S. Negahdaripour, Closed-form solution of absolute orientation using orthonormal matrices, Journal of the Optical Society America 5 (7) (1988) 1127–1135.

[8] K. Arun, T. S. Huang, S. D. Blostein, Least-Squares Fitting of Two 3-D Point Sets, IEEE Transactions on Pattern Analysis and Machine Intelligence 9 (5) (1987) 698–700.

[9] K. Kanatani, Analysis of 3-D rotation fitting, IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (5) (1994) 543–549.

[10] D. Goryn, S. Hein, On the estimation of rigid body rotation from noisy

data, IEEE Transactions on Pattern Analysis and Machine Intelligence 17 (12) (1995) 1219–1220.

[11] R. Sibson, Studies in the robustness of multidimensional scaling: Perturbational analysis of classical scaling, Journal of the Royal Statistical Society. Series B (Methodological) 41 (2) (1979) 217–229.

[12] R. Sibson, Studies in the robustness of multidimensional scaling: Procrustes statistics, Journal of the Royal Statistical Society. Series B (Methodological) 40 (2) (1978) 234–238.

[13] J. Fitzpatrick, J. West, The Distribution of Target Registration Error in Rigid-body Point-based Registration, IEEE Transactions on Medical Imaging 20 (9) (2001) 917–927.

[14] A. D. Wiles, A. Likholyot, D. D. Frantz, T. M. Peters, A Statistical Model for Point-Based Target Registration Error With Anisotropic Fiducial Localizer Error, IEEE Transactions on Medical Imaging 27 (3) (2008) 378–390.

[15] M. H. Moghari, P. Abolmaesumi, Distribution of target registration error for anisotropic and inhomogeneous fiducial localization error, IEEE Transactions on Medical Imaging 28 (6) (2009) 799–813.

[16] L. Dorst, First order error propagation of the procrustes method for 3D attitude estimation, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2) (2005) 221–229.

[17] N. Ohta, K. Kanatani, Optimal estimation of three-dimensional rotation and reliability evaluation, IEICE Transactions on Information and Systems E81-D (11) (1998) 1247–1252.

[18] I. Söderkvist, Perturbation analysis of the orthogonal procrustes problem, BIT Numerical Mathematics 33 (4) (1993) 687–694.

[19] J. Tang, L. Shao, X. Zhen, Robust point pattern matching based on spectral context, Pattern Recognition 47 (3) (2014) 1469–1484.

[20] J. Christmas, R. Everson, J. Bell, C. Winlove, Inexact bayesian point pattern matching for linear transformations, Pattern Recognition 47 (10) (2014) 3265–3275.

[21] B. J. Guerreiro, Sensor-based control and localization of autonomous systems in unknown environments, Ph.D. thesis, Instituto Superior

Técnico, University of Lisbon (December 2013).

[22] P. Lourenço, B. J. Guerreiro, P. Batista, P. Oliveira, C. Silvestre, 3-D Inertial Trajectory and Map Online Estimation: Building on a GAS Sensor-based SLAM filter, in: Proc. of the 2013 European Control Conference, Zurich, Switzerland, 2013, pp. 4214–4219.

[23] H. Durrant-Whyte, T. Bailey, Simultaneous Localisation and Mapping (SLAM): Part I The Essential Algorithms, IEEE Robotics & Automation Magazine 13 (2) (2006) 99–110.

[24] B. J. Guerreiro, P. Batista, C. Silvestre, P. Oliveira, Globally Asymptotically Stable Sensor-based Simultaneous Localization and Mapping, IEEE Transactions on Robotics 29 (6) (2013) 1380–1395.

[25] P. Lourenço, B. J. Guerreiro, P. Batista, P. Oliveira, C. Silvestre, Simultaneous Localization and Mapping for Aerial Vehicles: a 3-D sensor-based GAS filter, Autonomous Robots 40 (2016) 881–902.

[26] C. Beattie, S. Smith, Optimal matrix approximants in structural identification, Journal of Optimization Theory and Applications 74 (1) (1992) 23–56.

[27] I. Borg, P. Groenen, Modern Multidimensional Scaling: Theory and Applications, Springer Series in Statistics, Springer, 2007.

[28] F. Crosilla, A. Beinat, Use of generalised procrustes analysis for the photogrammetric block adjustment by independent models, {ISPRS} Journal of Photogrammetry and Remote Sensing 56 (3) (2002) 195–209.

[29] L. Igual, X. Perez-Sala, S. Escalera, C. Angulo, F. D. la Torre, Continuous generalized procrustes analysis, Pattern Recognition 47 (2) (2014) 659–671.

[30] P. Schönemann, R. Carroll, Fitting One Matrix to Another Under Choice of a Central Dilation and a Rigid Motion, Psychometrika 35 (1970) 245–255.

[31] G. McNeill, S. Vijayakumar, Hierarchical procrustes matching for shape retrival, in: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition (CVPR), CVPR, San Diego, CA, USA, 2006, pp. 885–894.

[32] Y. Gong, S. Lazebnik, A. Gordo, F. Perronnin, Iterative Quantization:

A Procrustean Approach to Learning Binary Codes for Large-Scale Image Retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (12) (2013) 2916–2929.

[33] A. Bartoli, D. Pizarro, M. Loog, Stratified Generalized Procrustes Analysis, International Journal of Computer Vision 101 (2) (2012) 227–253.

[34] K. Huang, N. Sidiropoulos, A. Swami, Non-Negative Matrix Factorization Revisited: Uniqueness and Algorithm for Symmetric Decomposition, IEEE Transactions on Signal Processing 62 (1) (2014) 211–224.

[35] B.-K. Ling, N. Tian, C.-F. Ho, W.-C. Siu, K.-L. Teo, Q. Dai, Maximally decimated paraunitary linear phase fir filter bank design via iterative svd approach, IEEE Transactions on Signal Processing 63 (2) (2015) 466–481.

[36] K. Lekadir, N. Keenan, D. Pennell, G.-Z. Yang, An Inter-Landmark Approach to 4-D Shape Extraction and Interpretation: Application to Myocardial Motion Assessment in MRI, IEEE Transactions on Medical Imaging 30 (1) (2011) 52–68.

[37] S. Kaneko, T. Kondo, A. Miyamoto, Robust matching of 3d contours using iterative closest point algorithm improved by m-estimation, Pattern Recognition 36 (9) (2003) 2041–2047.

[38] C. Wang, S. Mahadevan, Manifold Alignment Using Procrustes Analysis, in: Proceedings of the 25th International Conference on Machine Learning, ACM, New York, NY, USA, 2008, pp. 1120–1127.

[39] J. Liu, X. Liu, X. Ma, First-order perturbation analysis of singular vectors in singular value decomposition, IEEE Transactions on Signal Processing 56 (7) (2008) 3044–3049.

[40] W. S. Massey, Cross products of vectors in higher dimensional euclidean spaces, The American Mathematical Monthly 90 (10) (1983) 697–701.

[41] T. Anderson, An Introduction to Multivariate Statistical Analysis, Wiley Series in Probability and Statistics - Applied Probability and Statistics Section Series, Wiley, 1984.

[42] B. J. Guerreiro, C. Silvestre, P. Oliveira, Automatic 2-D LiDAR geometric calibration of installation bias, Robotics and Autonomous Systems 62 (8) (2014) 1116–1129.