

# Procrustes Analysis with Deformations: A Closed-Form Solution by Eigenvalue Decomposition

**Fang Bai**

ENCOV, TGI, Institut Pascal, UMR6602 CNRS, Université Clermont Auvergne

Email: Fang.Bai@yahoo.com

**Adrien Bartoli**

Department of Clinical Research and Innovation, CHU de Clermont-Ferrand  
ENCOV, TGI, Institut Pascal, UMR6602 CNRS, Université Clermont Auvergne

Email: Adrien.Bartoli@gmail.com

Received: 1 June 2021 / Accepted: 17 December 2021

Accepted by the *International Journal of Computer Vision*

DOI: 10.1007/s11263-021-01571-8

Corresponding Author: Fang Bai

---

```
@article{bai2022defgpa,  
  title={Procrustes Analysis with Deformations: A Closed-Form Solution by Eigenvalue Decomposition},  
  author={Bai, Fang and Bartoli, Adrien},  
  journal={International Journal of Computer Vision},  
  year={2022},  
  volume={130},  
  number={2},  
  pages={567-593},  
  doi={10.1007/s11263-021-01571-8}  
}
```

Springer version: <https://link.springer.com/article/10.1007/s11263-021-01571-8>

ArXiv version: <http://arxiv.org/abs/2206.14528>

Code repository: <https://bitbucket.org/clermontferrand/deformableprocrustes/src/master/>

# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
<b>2</b>	<b>Modeling Preliminaries</b>	<b>7</b>
2.1	Notation and Terminology . . . . .	7
2.2	Generalized Procrustes Analysis . . . . .	8
2.3	Brockett Cost Function on the Stiefel Manifold . . . . .	9
<b>3</b>	<b>Related Work</b>	<b>10</b>
3.1	Rigid Case . . . . .	10
3.2	Affine Case . . . . .	12
3.3	Deformable Case . . . . .	12
<b>4</b>	<b>Generalized Procrustes Analysis with the Affine Model</b>	<b>14</b>
4.1	Standard Form with the Shape Constraint . . . . .	14
4.1.1	The Shape Constraint $\mathbf{S}\mathbf{1} = \mathbf{0}$ . . . . .	14
4.1.2	The Shape Constraint $\mathbf{S}\mathbf{S}^T = \mathbf{\Lambda}$ . . . . .	14
4.2	Globally Optimal Solution . . . . .	15
4.3	Estimation of the Reference Covariance Prior . . . . .	17
4.4	The Coordinate Transformation of Datum Shapes . . . . .	18
4.5	Connection to Classical Results by Eliminating Translations . . . . .	19
<b>5</b>	<b>Generalized Procrustes Analysis with the Deformation Model</b>	<b>20</b>
5.1	Linear Basis Warps . . . . .	20
5.1.1	Formulation . . . . .	20
5.1.2	Operating on Point Clouds . . . . .	20
5.1.3	Regularization . . . . .	21
5.1.4	Examples of Linear Basis Warps . . . . .	21
5.2	Generalized Procrustes Analysis with Linear Basis Warps . . . . .	22
5.3	Eigenvector Characterization and Globally Optimal Solution . . . . .	23
5.4	Reformulation using the Soft Constraint . . . . .	24
<b>6</b>	<b>Generalized Procrustes Analysis with Partial Shapes</b>	<b>25</b>
6.1	Estimating the Reference Covariance Prior . . . . .	26

6.2	Closed-Form Solution . . . . .	27
6.3	Eigenvector Characterization and Tuning Parameters . . . . .	27
6.4	The Coordinate Transformation of Datum Shapes . . . . .	29
6.5	Reflection . . . . .	29
6.6	Pseudo-Code . . . . .	30
<b>7</b>	<b>Experimental Results</b>	<b>31</b>
7.1	Experimental Setups . . . . .	31
7.2	Evaluation Metrics . . . . .	32
7.2.1	Landmark Residual . . . . .	32
7.2.2	Cross-Validation Error . . . . .	33
7.3	Thin-Plate Spline Smoothing Parameters . . . . .	33
7.4	Results Based on Landmarks . . . . .	35
7.4.1	Statistics . . . . .	35
7.4.2	Visualization . . . . .	35
7.5	Results Based on Image Intensities . . . . .	37
7.6	The Asymmetry in Cost Functions . . . . .	39
<b>8</b>	<b>Conclusion</b>	<b>40</b>
<b>A</b>	<b>The Thin-Plate Spline</b>	<b>40</b>
A.1	General Mapping Definition . . . . .	40
A.2	Standard Form as a Linear Regression Model . . . . .	41
A.3	Regularization by Bending Energy Matrix . . . . .	42
A.4	The Thin-Plate Spline Satisfies $\mathcal{B}(\cdot)^\top \mathbf{x} = \mathbf{1}$ and $\mathbf{Z}\mathbf{x} = \mathbf{0}$ . . . . .	43
A.4.1	$\mathcal{B}(\cdot)^\top \mathbf{x} = \mathbf{1}$ . . . . .	44
A.4.2	$\mathbf{Z}\mathbf{x} = \mathbf{0}$ . . . . .	44
A.5	The Thin-Plate Spline is Invariant to the Coordinate Transformation . . . . .	44
<b>B</b>	<b>Lemmas and Proof of Theorem 2</b>	<b>45</b>
B.1	Lemmas on Positive Semidefinite Matrices . . . . .	45
B.2	Proof of Theorem 2 . . . . .	45
<b>C</b>	<b>Proof of Theorem 3</b>	<b>46</b>

**D Rigid Procrustes Analysis with Two Shapes****47****Updates:**

- The proof of Theorem 1 is given under more general conditions.

## Abstract

Generalized Procrustes Analysis (GPA) is the problem of bringing multiple shapes into a common reference by estimating transformations. GPA has been extensively studied for the Euclidean and affine transformations. We introduce GPA with deformable transformations, which forms a much wider and difficult problem. We specifically study a class of transformations called the Linear Basis Warps (LBWs), which contains the affine transformation and most of the usual deformation models, such as the Thin-Plate Spline (TPS). GPA with deformations is a nonconvex underconstrained problem. We resolve the fundamental ambiguities of deformable GPA using two shape constraints requiring the eigenvalues of the shape covariance. These eigenvalues can be computed independently as a prior or posterior. We give a closed-form and optimal solution to deformable GPA based on an eigenvalue decomposition. This solution handles regularization, favoring smooth deformation fields. It requires the transformation model to satisfy a fundamental property of free-translations, which asserts that the model can implement any translation. We show that this property fortunately holds true for most common transformation models, including the affine and TPS models. For the other models, we give another closed-form solution to GPA, which agrees exactly with the first solution for models with free-translation. We give pseudo-code for computing our solution, leading to the proposed DefGPA method, which is fast, globally optimal and widely applicable. We validate our method and compare it to previous work on six diverse 2D and 3D datasets, with special care taken to choose the hyperparameters from cross-validation.

## 1 Introduction

The problem of Generalized Procrustes Analysis (GPA) is to *register a set of shapes with known correspondences into a single unknown reference shape*, by estimating transformations. The existing literature focuses on GPA with Euclidean (or similarity in case of scaling) and affine transformations, which do not cope with deformations. When present in the datum shapes, the un-modeled deformations are typically considered as noise, which creates biases on the estimation thus causing misfitting to the data. This is undesirable, particularly when the shape is nonrigid and the precision is critical, for example in medical applications.

We consider GPA with deformation models. In specific, we consider a generic class of transformation models, termed Linear Basis Warps (LBWs), which generalize over affine transformation models and most commonly used warp models, like the Thin-Plate Spline (TPS) (Bookstein, 1989; Duchon, 1976), Radial Basis Functions (RBF) (Fornet et al., 2001), and Free-Form Deformations (FFD) (Rueckert et al., 1999; Szeliski and Coughlan, 1997). The LBW shares many similarities with the affine model, where the transformation parameters are linear as weights to a set of possibly nonlinear basis functions. This gives additional modeling flexibility to the LBW which copes with nonlinear deformations, while retaining the computational simplicity of linear models.

We propose GPA with LBWs. Our cost function is formulated in the coordinate frame of the reference shape (which is termed the reference-space cost). This cost function is linear least-squares in optimizable parameters which maximizes the benefit of using the LBW. An important point to bear in mind is that a sufficiently generic deformation model can match any arbitrary shape to another to a good degree of fitness. This makes the estimation of the reference shape highly ambiguous. We resolve these ambiguities by shape constraints, and show that these extra degree of freedoms can be determined by the eigenvalues of the covariance matrix of the reference shape, termed *reference covariance prior*. Importantly, the reference covariance prior is independent of the rest of the computation, thus can be enforced as a prior or posterior. We propose a method to estimate the reference covariance prior based on the similarity to the datum shapes while other options are also possible.

We propose a globally optimal solution in closed-form to the GPA formulation with LBWs. Our solution is based on a specific case of the eigenvalue problem (Theorem 1) and the characterization of LBWs (Theorem 2 and Theorem 3). We present our ideas following a specific-to-general scheme: Section 4 for affine GPA, Section 5 for GPA with LBWs, and Section 6 for GPA with partial datum shapes.

Section 4. We present in Section 4.1 our affine GPA formulation and the ideas behind the shape constraints  $\mathbf{S}\mathbf{1} = \mathbf{0}$  and  $\mathbf{S}\mathbf{S}^T = \mathbf{A}$ , and then in Section 4.2 the theoretical results on how to solve the formulation globally in closed-form. We show that a closed-form solution exists if the matrix to decompose has an all-one eigenvector. In Section 4.3, we present a method to estimate the reference covariance prior  $\mathbf{A}$ . We show in Section 4.4 that the proposed formulation is invariant to coordinate transformations of the datum shapes, where the optimal reference shape will remain the same. Lastly, in Section 4.5, we draw the connection between our method and the classical affine GPA method by eliminating translations explicitly.

Section 5. We first introduce LBWs (Section 5.1) and GPA with LBWs (Section 5.2), and then relate the existence of an all-one eigenvector to the equivalent properties of LBWs (Section 5.3 and Theorem 2 in specific). We introduce the concept of free-translations (Definition 1), and show that to have the desired properties in Theorem 2 it suffices to have a LBW where its translation part is constraint-free (Proposition 4). Both the affine transformation and the TPS warp, which are the two most commonly used LBWs, fall into this category, thus the resulting GPA problems can be solved in closed-form by Theorem 1 and Proposition 1. Lastly, in Section 5.4, we propose a soft-constrained version (using a penalty term) when the condition is not met, which can also be solved in closed-form.

Section 6. We extend the results from full shapes to partial shapes, *i.e.*, the estimation of the reference covariance prior via the shape completion (Section 6.1), the GPA formulation with LBWs for partial shapes (Section 6.2), and the all-one eigenvector characterization (Section 6.3). Section 6.4 shows that the

proposed GPA formulation with LBWs is invariant to coordinate transformations of the datum shapes, where the optimal reference shape will remain the same. Section 6.5 shows how to handle reflections in the solution. Section 6.6 gives pseudo-code to benefit future research. Implementation details are summarized in Algorithm 1 and Algorithm 2.

In a nutshell, we present a closed-form GPA method, termed DefGPA (*i.e.*, for GPA based on LBWs), which is fast, globally optimal and copes with the general problem with regularization and incomplete shapes. The rest of this article is organized as follow. Section 2 introduces our notation, the GPA problem, and the Brockett cost function on the Stiefel manifold. Section 3 reviews the related work. Section 4, 5 and 6 describe our DefGPA method. Section 7 provides experimental results. Section 8 concludes the paper.

## 2 Modeling Preliminaries

### 2.1 Notation and Terminology

**Notation.** We use  $\mathbb{R}$  to denote the set of real numbers,  $\mathbb{R}^n$  the set of column vectors of dimension  $n$ , and  $\mathbb{R}^{m \times n}$  the set of matrices of dimension  $m \times n$ . All the vectors are column majored, denoted by lower-case characters in bold. The matrices are upper-case characters in bold. The scalars are in italics. For matrices, we reserve  $\mathbf{I}$  for identity matrices and  $\mathbf{O}$  for all-zero matrices. For vectors we reserve  $\mathbf{0}$  for all-zero vectors and  $\mathbf{1}$  for all-one vectors. All these special matrices and vectors are assumed to have proper dimensions induced by the context.  $\mathbf{A}^\top$  is the transpose of  $\mathbf{A}$ ,  $\mathbf{A}^{-1}$  the inverse of  $\mathbf{A}$ , and  $\mathbf{A}^\dagger$  the Moore-Penrose pseudo-inverse of  $\mathbf{A}$ . We use the notation  $\mathbf{Range}(\mathbf{A})$  and  $\mathbf{Null}(\mathbf{A})$  to denote the range space and null space of  $\mathbf{A}$ , respectively.  $\mathbf{A} \succeq \mathbf{O}$  means  $\mathbf{A}$  is positive semidefinite.  $\mathbf{A} \succeq \mathbf{B}$  means  $\mathbf{A} - \mathbf{B}$  is positive semidefinite. The symbol  $\|\cdot\|_F$  denotes the matrix Frobenius norm, and  $\|\cdot\|_2$  the vector  $\ell_2$  norm.  $\mathbf{tr}(\mathbf{A})$  is the trace for a square matrix  $\mathbf{A}$ .  $\|\mathbf{A}\|_F^2 = \mathbf{tr}(\mathbf{A}\mathbf{A}^\top) = \mathbf{tr}(\mathbf{A}^\top\mathbf{A})$ . We use  $\mathbf{nnz}(\cdot)$  as a shorthand for “number of non-zeros”.

**The  $d$  top eigenvectors and bottom eigenvectors.** Let  $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$  be the eigenvalue decomposition of a symmetric matrix  $\mathbf{A}$ , where  $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n]$  is a square orthonormal matrix. Let  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$  with  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ . Then we call  $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_d$  in sequence the  $d$  top eigenvectors, and  $\mathbf{q}_n, \mathbf{q}_{n-1}, \dots, \mathbf{q}_{n-d+1}$  in sequence the  $d$  bottom eigenvectors of  $\mathbf{A}$ . The analogous concepts in the Singular Value Decomposition are the  $d$  leftmost and rightmost singular vectors.

**Shape covariance matrix.** We model shapes as point-clouds. Given a  $d$ -dimensional shape of  $m$  points  $\mathbf{S} \in \mathbb{R}^{d \times m}$ , the geometric center, also called *centroid*, is the mean of its columns, defined as  $\mathbf{mean}(\mathbf{S}) = \frac{1}{m}\mathbf{S}\mathbf{1}$ .

The concept *covariance matrix* of the shape  $\mathbf{S}$  is defined by:

$$\text{Cov}(\mathbf{S}) = (\mathbf{S} - \mathbf{mean}(\mathbf{S})\mathbf{1}^\top) (\mathbf{S} - \mathbf{mean}(\mathbf{S})\mathbf{1}^\top)^\top.$$

In particular, if  $\mathbf{S}$  is at the origin of the coordinate frame,  $\mathbf{mean}(\mathbf{S}) = \mathbf{0}$ , the shape covariance becomes  $\text{Cov}(\mathbf{S}) = \mathbf{S}\mathbf{S}^\top$ .

## 2.2 Generalized Procrustes Analysis

A datum shape is represented as a matrix  $\mathbf{D}_i \in \mathbb{R}^{d \times m}$  ( $i \in [1 : n]$ ). Here  $d \in \{2, 3\}$  is the dimension of the points and  $m$  the number of points in the shape. We will index the  $j$ -th point in the shape  $\mathbf{D}_i$  as  $\mathbf{D}_i[j]$ , which is the  $j$ -th column of  $\mathbf{D}_i$ . The shape points are arranged correspondence-wise: the  $j$ -th point in shape  $\mathbf{D}_{i_1}$  and the  $j$ -th point in shape  $\mathbf{D}_{i_2}$  correspond to different observations of the same physical point. Missing data are caused by unobserved points in some shapes. They are modeled by binary visibility variables  $\gamma_{i,j}$  ( $i \in [1 : n]$ ,  $j \in [1 : m]$ ), where  $\gamma_{i,j} = 1$  if and only if the  $j$ -th point occurs in the  $i$ -th shape, and  $\gamma_{i,j} = 0$  otherwise. The reference shape  $\mathbf{S} \in \mathbb{R}^{d \times m}$  is defined accordingly, with its  $j$ -th column  $\mathbf{S}[j]$  corresponding to the  $j$ -th physical point. All the points occur in the reference shape  $\mathbf{S}$ .

We denote  $\mathcal{T}_i$  ( $i \in [1 : n]$ ) the transformation from the datum shape  $\mathbf{D}_i$  to the reference shape  $\mathbf{S}$ . Here  $\mathcal{T}_i$  can be Euclidean, similarity, affine, or non-rigid transformations. Different choices of  $\mathcal{T}_i$  yield different types of GPA problems. The general GPA problem can be written as:

$$\begin{aligned} \arg \min_{\{\mathcal{T}_i\}, \mathbf{S}} \quad & \sum_{i=1}^n \sum_{j=1}^m \gamma_{i,j} \|\mathcal{T}_i(\mathbf{D}_i[j]) - \mathbf{S}[j]\|_2^2 \\ \text{s.t.} \quad & \mathcal{C}(\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_n, \mathbf{S}) = \mathbf{0}, \end{aligned} \tag{1}$$

where  $\mathcal{C}(\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_n, \mathbf{S}) = \mathbf{0}$  denotes the set of constraints used to avoid degeneracy. A typical degenerate case is  $\mathcal{T}_i(\mathbf{p}) = \mathbf{0}$  for any point  $\mathbf{p}$ , and  $\mathbf{S} = \mathbf{O}$ . The detailed choices of these constraints will be provided later on.

The GPA model in formulation (1) holds for many transformation models and handles missing datum points. We give a more compact matrix form which will simplify the derivation of our methods. To this end, we define the diagonal visibility matrix  $\mathbf{\Gamma}_i = \text{diag}(\gamma_{i,1}, \gamma_{i,2}, \dots, \gamma_{i,m})$  ( $i \in [1 : n]$ ), constructed by the visibility variables corresponding to the shape  $\mathbf{D}_i$ . By noting that:

$$\chi_r^2 = \sum_{i=1}^n \sum_{j=1}^m \gamma_{i,j} \|\mathcal{T}_i(\mathbf{D}_i[j]) - \mathbf{S}[j]\|_2^2 = \sum_{i=1}^n \sum_{j=1}^m \|\gamma_{i,j} \mathcal{T}_i(\mathbf{D}_i[j]) - \gamma_{i,j} \mathbf{S}[j]\|_2^2 = \sum_{i=1}^n \|\mathcal{T}_i(\mathbf{D}_i) \mathbf{\Gamma}_i - \mathbf{S} \mathbf{\Gamma}_i\|_F^2,$$



we can write the GPA model in a compact matrix form:

$$\begin{aligned} \arg \min_{\{\mathcal{T}_i\}, \mathbf{S}} \quad & \sum_{i=1}^n \|\mathcal{T}_i(\mathbf{D}_i) \mathbf{\Gamma}_i - \mathbf{S} \mathbf{\Gamma}_i\|_F^2 \\ \text{s.t.} \quad & \mathcal{C}(\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_n, \mathbf{S}) = \mathbf{0}. \end{aligned} \quad (2)$$

In case of full shapes (shapes without missing datum points), we have  $\mathbf{\Gamma}_i = \mathbf{I}$  ( $i \in [1 : n]$ ).

**The reference-space cost and the datum-space cost.** The cost function defined in formulations (1) and (2) is called the *reference-space cost* since the residual error is evaluated in the coordinate frame of the reference shape (Bartoli et al., 2013). In contrast, it is possible to derive a cost function in the *datum-space*:

$$\chi_d^2 = \sum_{i=1}^n \|\mathbf{D}_i \mathbf{\Gamma}_i - \mathcal{T}_i^{-1}(\mathbf{S}) \mathbf{\Gamma}_i\|_F^2. \quad (3)$$

The datum-space cost is a generative model since all the datum shapes  $\mathbf{D}_i$  are generated by the single reference shape  $\mathbf{S}$  with transformations  $\mathcal{T}_i^{-1}$ . In contrast, the reference-space cost is discriminative, as it seeks for the transformations  $\mathcal{T}_i$  that can best match the datum shapes  $\mathbf{D}_i$  with the reference shape  $\mathbf{S}$ . The reference-space cost and the datum-space cost are identical if  $\mathcal{T}_i$  represents Euclidean transformations (Bartoli et al., 2013).

### 2.3 Brockett Cost Function on the Stiefel Manifold

The matrix Stiefel manifold is the set of matrices satisfying:

$$St(d, m) = \left\{ \mathbf{X} \in \mathbb{R}^{m \times d} \mid \mathbf{X}^\top \mathbf{X} = \mathbf{I} \right\}.$$

In particular, we shall use the classical result of the following Stiefel manifold optimization problem:

$$\arg \min_{\mathbf{X} \in St(d, m)} \text{tr}(\mathbf{X}^\top \mathcal{P} \mathbf{X} \mathbf{\Lambda}). \quad (4)$$

The cost in problem (4) is termed the Brockett cost function in the Stiefel manifold optimization literature (Absil et al., 2009; Birtea et al., 2019; Brockett, 1989). Importantly, this problem admits a closed-form solution. The critical points of the Brockett cost function on the Stiefel manifold are the eigenvectors of  $\mathcal{P}$  (Birtea et al., 2019; Brockett, 1989). The global minimum of problem (4) can thus be obtained by arranging the eigenvectors as follow. Let  $(\alpha_j, \boldsymbol{\xi}_j)$  ( $j \in [1 : m]$ ) be the set of eigenvalues and eigenvectors of  $\mathcal{P}$ , such that  $\mathcal{P} \boldsymbol{\xi}_j = \alpha_j \boldsymbol{\xi}_j$ , with  $0 \leq \alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_m$  arranged in the ascending order. Let the diagonal elements of

$\mathbf{\Lambda}$  being arranged in the descending order  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$ . Then  $\mathbf{X}^* = [\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_d]$ , comprising the  $d$  bottom eigenvectors of  $\mathcal{P}$ , is globally optimal to problem (4), with a total cost  $\alpha_1\lambda_1 + \alpha_2\lambda_2 + \dots + \alpha_d\lambda_d$ . Any other combination yields a larger cost which proves the optimality of  $\mathbf{X}^*$ , by a result of Hardy-Littlewood-Polya (Brockett, 1989; Hardy et al., 1952).

**Lemma 1.** *Let  $\mathcal{P}$  be a symmetric matrix with its  $d$ -bottom eigenvectors being  $\mathbf{X} = [\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_d]$ . Let  $\mathbf{\Lambda}$  be a diagonal matrix  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$  with  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ . The globally optimal solution to the optimization problem:*

$$\arg \min_{\mathbf{S}} \text{tr}(\mathbf{S}\mathcal{P}\mathbf{S}^\top) \quad \text{s.t.} \quad \mathbf{S}\mathbf{S}^\top = \mathbf{\Lambda}, \quad (5)$$

is  $\mathbf{S}^* = \sqrt{\mathbf{\Lambda}}\mathbf{X}^\top$ , i.e., obtained by scaling the  $d$  bottom eigenvectors of  $\mathcal{P}$  by  $\sqrt{\mathbf{\Lambda}}$ .

*Proof.* We introduce a matrix  $\mathbf{X}$  so that  $\mathbf{S}^\top = \mathbf{X}\sqrt{\mathbf{\Lambda}}$ . The constraint  $\mathbf{S}\mathbf{S}^\top = \mathbf{\Lambda}$  thus becomes  $\mathbf{X}^\top\mathbf{X} = \mathbf{I}$ . The cost function  $\text{tr}(\mathbf{S}\mathcal{P}\mathbf{S}^\top)$  can be rewritten as  $\text{tr}(\mathbf{S}\mathcal{P}\mathbf{S}^\top) = \text{tr}(\sqrt{\mathbf{\Lambda}}\mathbf{X}^\top\mathcal{P}\mathbf{X}\sqrt{\mathbf{\Lambda}}) = \text{tr}(\mathbf{X}^\top\mathcal{P}\mathbf{X}\mathbf{\Lambda})$ . The globally optimal solution to problem (5) is thus  $\mathbf{S}^{*\top} = \mathbf{X}^*\sqrt{\mathbf{\Lambda}}$ , with  $\mathbf{X}^*$  being the solution of problem (4). In other words,  $\mathbf{S}^*$  is obtained by scaling the  $d$  bottom eigenvectors of  $\mathcal{P}$  by  $\sqrt{\mathbf{\Lambda}}$ .  $\square$

**Remark 1.** *Analogously, the globally optimal solution to the following maximization problem:*

$$\arg \max_{\mathbf{S}} \text{tr}(\mathbf{S}\mathcal{Q}\mathbf{S}^\top) \quad \text{s.t.} \quad \mathbf{S}\mathbf{S}^\top = \mathbf{\Lambda}. \quad (6)$$

is to scale the  $d$  top eigenvectors of the matrix  $\mathcal{Q}$  by  $\sqrt{\mathbf{\Lambda}}$ .

## 3 Related Work

GPA considers shape registration with known correspondences, which abounds in the literature. We classify the literature according to the used transformation model.

### 3.1 Rigid Case

**Two shapes.** The classical approach to the Procrustes analysis problem starts with linear algebraic results of two shape registration using least-squares. In particular, the closed-form solution of registering two point-clouds with orthonormal matrices (termed orthogonal Procrustes) was established half a century ago (Green, 1952; Schönemann, 1966). The result can be extended to tackle translations (Arun et al., 1987) and scale factors (Horn et al., 1988). The result in (Arun et al., 1987; Horn et al., 1988) uses orthonormal matrices to represent rotations. Besides, it is possible to derive the result using unit quaternions to represent rotations (Horn, 1987), and furthermore, using dual quaternions to represent both rotations and translations (Walker

et al., 1991). When confronting large noise, the orthonormal matrix based approaches (Arun et al., 1987; Horn et al., 1988) can give a reflection matrix (with determinant  $-1$ ) as opposed to a valid rotation matrix (Umeyama, 1991). A correction to the special orthogonal group ( $SO(3)$ , which represents valid rotations) is given in (Umeyama, 1991), and a more concise derivation in (Kanatani, 1994). A comparative study of these methods is provided in (Eggert et al., 1997). However from the statistical point of view, these methods assume that noise only occurs in the target point-cloud, which is isotropic, identical and independently Gaussian distributed. This assumption can deviate from real noise schemes. More sophisticated formulations based on anisotropic and inhomogeneous Gaussian noise models are discussed in (Matei and Meer, 1999; Ohta and Kanatani, 1998), where the renormalization technique based on the quaternion parameterization is used to solve the resulting optimization problem effectively. Overall, two shape rigid Procrustes analysis can be considered a solved problem.

**Multiple shapes.** Multiple shape Procrustes analysis is a more difficult problem, which is termed generalized Procrustes analysis (GPA) (Gower, 1975; Kristof and Wingersky, 1971). The work in (Kristof and Wingersky, 1971; Ten Berge, 1977) examined the optimality conditions of the orthogonal GPA (Euclidean GPA without translations). The derived conditions are either necessary or sufficient, however not necessary-sufficient. Actually, due to the existence of rotations, the Euclidean GPA, or similar problems like pose graph optimization (PGO) (Bai et al., 2021) has recently been acknowledged as highly nonlinear and non-convex (Rosen et al., 2019). Therefore up to now, exact techniques for the rigid GPA are all iterative. In early days, using the pairwise rigid Procrustes result as a subroutine, rigid GPA was solved by alternating the estimation of the reference shape and that of the transformations until convergence (Gower, 1975; Rohlf and Slice, 1990; Ten Berge, 1977; Wen et al., 2006). However, such methods are not guaranteed to attain the global optimum of the cost function. The translation part of the rigid GPA problem is free of constraints, thus can be eliminated from the cost function by a variable projection scheme (Golub and Pereyra, 2003). As a result, the rigid GPA can be formulated as a rotation estimation problem on the manifold (Benjemaa and Schmitt, 1998; Krishnan et al., 2005; Williams and Bennamoun, 2001). Sequential techniques were proposed based on the unit quaternion (Benjemaa and Schmitt, 1998) and orthonormal matrices (Williams and Bennamoun, 2001). A complete approach on the  $SO(3)$  Lie group has been provided in (Krishnan et al., 2005). Building upon efficient sparse linear algebra techniques (Davis, 2006) and Lie group theory (Iserles et al., 2000), the Lie group approach can be considered as the current state-of-the-art. An initialization technique to the iterative solver was provided in (Bartoli et al., 2013).

### 3.2 Affine Case

There exists a vast body of research in shape analysis (Dryden and Mardia, 2016; Kendall, 1984; Kendall et al., 2009) where the GPA problem arose in parallel, *e.g.*, the complex arithmetic approach for 2D shapes and its relation to Procrustes analysis (Kent, 1994). The shape statistics, *i.e.*, mean and variability, are defined on aligned shapes (thus being invariant to rotation, translation and scaling). A commonly used alignment technique is the Procrustes analysis (Kendall, 1984) or GPA (Goodall, 1991). The affine GPA has been studied with a closed-form solution proposed (Rohlf and Slice, 1990). In the present notation, following the result in (Rohlf and Slice, 1990), the principal axes of the reference shape are estimated as the  $d$  top eigenvectors of the matrix  $\sum_{i=1}^n \mathbf{D}_i^{\mathbf{T}} (\mathbf{D}_i^{\mathbf{T}})^{\dagger}$ , after centering each  $\mathbf{D}_i$  to the origin of the coordinate frame and eliminating the translation parameters.

We derive similar results in Section 4.5 in the global optimization context instead of the alternation scheme used in (Rohlf and Slice, 1990), which means the scaling factor (*i.e.*, the reference covariance prior used in the proposed shape constraint) is estimated in a different way (Section 4.3). We show in Section 4.5 that this approach gives the same result as the proposed formulation (8), while the proposed formulation without eliminating translations is more general and extensible to the LBWs.

The above affine GPA solution, in our context, is called the reference-space solution. On the other hand, the full shape affine GPA in the datum-space admits a closed-form solution using the Singular Value Decomposition (SVD). Let  $\mathbf{D}_i$  be zero centered. Then the datum-space cost is  $\sum_{i=1}^n \|\mathbf{D}_i - \mathbf{A}_i \mathbf{S}\|_F^2$ . The optimal solution,  $[\mathbf{A}_1^{\mathbf{T}}, \dots, \mathbf{A}_n^{\mathbf{T}}]^{\mathbf{T}} \mathbf{S}$  as a whole, can be determined via the SVD of the matrix  $[\mathbf{D}_1^{\mathbf{T}}, \dots, \mathbf{D}_n^{\mathbf{T}}]^{\mathbf{T}}$ . This idea is in the same spirit of the Tomasi-Kanade factorization (Tomasi and Kanade, 1992) in computer vision, which studied the affine transformation of the orthographic camera model. We term this method as \*AFF\_d and use it as a benchmark algorithm in the experiments.

### 3.3 Deformable Case

In shape analysis, a shape is considered as an element on the so-called *shape manifold* (the set of all possible shapes) (Kendall, 1984; Kilian et al., 2007). The shape deformation is thus studied as the shortest path on the shape manifold, *i.e.*, the geodesic path under a chosen Riemannian metric (Kilian et al., 2007). In the rigid case, after discarding translation and scaling, the shape manifold is a quotient manifold invariant to rotations and the Riemannian metric can be chosen as the Procrustes distance (Kendall, 1984). For deformable cases, the Riemannian metric can be otherwise formulated based on the rigidity or isometry (Kilian et al., 2007). The work (Kendall, 1984; Kilian et al., 2007), as well as this paper, use landmarks (or meshes with fixed connections) to model shapes and define transformations. In addition, recent research

has studied shape analysis based on curves (Joshi et al., 2007; Younes et al., 2008) or surfaces, *e.g.*, level sets (Osher and Fedkiw, 2003), medial surfaces (Bouix et al., 2005), Q-maps (Kurtek et al., 2010, 2011), Square Root Normal Fields (SRNF) (Jermyn et al., 2012; Laga et al., 2017), etc. See the review papers (Laga, 2018; Younes, 2012). There is also a line of work including skeletal structures, *e.g.*, the medial axis representations (M-rep) (Fletcher et al., 2004), and SCAPE (mesh with an articulated skeleton) (Anguelov et al., 2005).

Beyond different shape modeling techniques, deformations based on landmarks/meshes have a rich literature. Typically, the deformation of a mesh is defined piece-wisely, with a local transformation associated to each vertex or triangle (Freifeld and Black, 2012), or with additional geometric or smoothing constraints (Allen et al., 2003; Anguelov et al., 2005; Song et al., 2020; Sumner et al., 2007). It is worth mentioning that such piece-wise models are commonly used in describing template based deformations, where a reference shape is given a priori. In this paper, we use landmarks to model shapes and warps to model deformations. The landmarks can be extracted from the key-points or the samplings of the shape boundary (Cootes et al., 1995). While the proposed method applies to meshes as well, the connection of vertices (or edges of the meshes) is never required. Our analysis is based on a class of warps, called LBWs, which is a linear combination of the nonlinear basis functions (Bartoli et al., 2010; Bookstein, 1989; Fornefett et al., 2001; Rueckert et al., 1999; Szeliski and Coughlan, 1997). A typical example of the LBWs is the well-known TPS warp (Bookstein, 1989) which we will use to demonstrate our results.

There has been work using warps to refine the shapes after solving the rigid registration (Brown and Rusinkiewicz, 2007; Kim et al., 2008). Such methods are not template free in essence as when it comes to the estimation of the warp parameters, a reference shape has been known already. Little attention has been paid to estimate the reference shape and the transformation parameters all together in a unified manner. In this work, we provide a closed-form solution to the unified estimation problem.

Different from classical affine methods (Goodall, 1991; Kendall, 1984; Rohlf and Slice, 1990), our method allows for the possibility to keep the translation parameters during the estimation (Theorem 1 and Proposition 1). This is crucial to GPA with LBWs, as in certain cases, the translation parameters cannot be explicitly identified. In specific, for LBWs, we relate the constraint-free translation (Definition 1) to the existence of an eigenvector of all-ones and the equivalent properties of LBWs (Theorem 2 and Theorem 3), which constitutes the foundation of our closed-form solution.

## 4 Generalized Procrustes Analysis with the Affine Model

### 4.1 Standard Form with the Shape Constraint

We start with the case without missing datum points, *i.e.*,  $\Gamma_i = \mathbf{I}$  in formulation (2). We study the affine-GPA with  $\mathcal{T}_i(\mathbf{D}_i) = \mathbf{A}_i \mathbf{D}_i + \mathbf{t}_i \mathbf{1}^\top$ , with  $\mathbf{A}_i \in \mathbb{R}^{d \times d}$  being linear and  $\mathbf{t}_i \in \mathbb{R}^d$  a translation. We write the affine transformation in homogeneous form as:

$$\mathbf{A}_i \mathbf{D}_i + \mathbf{t}_i \mathbf{1}^\top = \begin{bmatrix} \mathbf{A}_i & \mathbf{t}_i \end{bmatrix} \begin{bmatrix} \mathbf{D}_i \\ \mathbf{1}^\top \end{bmatrix} \stackrel{\text{def}}{=} \tilde{\mathbf{A}}_i \tilde{\mathbf{D}}_i, \quad (7)$$

where we term  $\tilde{\mathbf{D}}_i$  the *homogeneous representation* of  $\mathbf{D}_i$  by completing  $\mathbf{D}_i$  with a row of all ones. In this representation, matrix  $\tilde{\mathbf{A}}_i = [\mathbf{A}_i, \mathbf{t}_i]$  contains all parameters to be estimated.

Following the reference-space cost (2), our proposed formulation is given as:

$$\text{I: } \begin{cases} \arg \min_{\{\tilde{\mathbf{A}}_i\}, \mathbf{S}} \sum_{i=1}^n \|\tilde{\mathbf{A}}_i \tilde{\mathbf{D}}_i - \mathbf{S}\|_F^2 \\ \text{s.t.} \quad \mathbf{S} \mathbf{S}^\top = \mathbf{\Lambda}, \quad \mathbf{S} \mathbf{1} = \mathbf{0}. \end{cases} \quad (8)$$

In formulation (8),  $\mathbf{S} \mathbf{S}^\top = \mathbf{\Lambda}$  and  $\mathbf{S} \mathbf{1} = \mathbf{0}$  are the shape constraints, which concretize the constraint  $\mathcal{C}(\cdot) = \mathbf{0}$  in formulation (2). The matrix  $\mathbf{\Lambda}$ , termed *reference covariance prior*, is a diagonal matrix  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_d)$  driven by  $d$  parameters  $\lambda_1, \lambda_2, \dots, \lambda_d$ . We shall shortly present in Section 4.3 a method to estimate  $\mathbf{\Lambda}$  based on rigidity. The ideas behind these two shape constraints are motivated as follow.

#### 4.1.1 The Shape Constraint $\mathbf{S} \mathbf{1} = \mathbf{0}$

This shape constraint is used to center the reference shape to the origin of the coordinate frame. The role of this constraint is twofold: 1) it provides  $d$  constraints to remove the  $d$  gauge freedoms caused by translations; 2) it reduces the shape covariance matrix of the reference shape to the form  $\text{Cov}(\mathbf{S}) = \mathbf{S} \mathbf{S}^\top$  as in this case  $\text{mean}(\mathbf{S}) = \frac{1}{m} \mathbf{S} \mathbf{1} = \mathbf{0}$ .

#### 4.1.2 The Shape Constraint $\mathbf{S} \mathbf{S}^\top = \mathbf{\Lambda}$

This shape constraint is used to: 1) capture the eigenvalues of the shape covariance matrix  $\text{Cov}(\mathbf{S}) = \mathbf{S} \mathbf{S}^\top$ , and 2) fix the gauge freedom caused by rotations. The main insight here is that the eigenvalues of the shape covariance matrix remain unchanged if the shape undergoes rigid transformations. Moreover, by letting  $\text{Cov}(\mathbf{S}) = \mathbf{S} \mathbf{S}^\top = \mathbf{\Lambda}$  we fix the gauge freedom caused by rotations.

**Lemma 2.**  $\text{Cov}(\mathbf{RS} + \mathbf{t}\mathbf{1}^\top) = \mathbf{R}\text{Cov}(\mathbf{S})\mathbf{R}^\top$  for any arbitrary rotation  $\mathbf{R}$  and translation  $\mathbf{t}$ .

*Proof.* By the fact that  $\text{mean}(\mathbf{RS} + \mathbf{t}\mathbf{1}^\top) = \frac{1}{m}(\mathbf{RS} + \mathbf{t}\mathbf{1}^\top)\mathbf{1} = \mathbf{R}\text{mean}(\mathbf{S}) + \mathbf{t}$ , we thus have  $(\mathbf{RS} + \mathbf{t}\mathbf{1}^\top) - \text{mean}(\mathbf{RS} + \mathbf{t}\mathbf{1}^\top)\mathbf{1}^\top = \mathbf{R}(\mathbf{S} - \text{mean}(\mathbf{S})\mathbf{1}^\top)$ .  $\square$

Let  $\text{Cov}(\mathbf{S}) = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$  be the eigenvalue decomposition of  $\text{Cov}(\mathbf{S})$ . Then  $\mathbf{R}\text{Cov}(\mathbf{S})\mathbf{R}^\top = \mathbf{R}\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top\mathbf{R}^\top$  admits the eigenvalue decomposition of  $\text{Cov}(\mathbf{RS} + \mathbf{t}\mathbf{1}^\top)$  by Lemma 2. This means matrix  $\text{Cov}(\mathbf{RS} + \mathbf{t}\mathbf{1}^\top)$  has the same eigenvalues as matrix  $\text{Cov}(\mathbf{S})$ , which are given as the diagonal elements of  $\mathbf{\Lambda}$ . Moreover,  $\text{Cov}(\mathbf{RS} + \mathbf{t}\mathbf{1}^\top) = \mathbf{\Lambda}$  if and only if  $\mathbf{R} = \mathbf{U}^\top$  which fixes the rotation.

In general, we want the reference shape to look similar to the datum shapes, thus we choose the eigenvalues of  $\mathbf{S}\mathbf{S}^\top$  to be close to those of the datum shape covariance matrices. Based on this idea, we present a method to estimate  $\mathbf{\Lambda}$  in Section 4.3. However, as we shall see in what follows, the matrix  $\mathbf{\Lambda}$  is never used in the intermediate calculation thus can be determined separately as a prior or posterior.

## 4.2 Globally Optimal Solution

Problem (8) is separable. Given  $\mathbf{S}$ , we obtain  $\tilde{\mathbf{A}}_i = \mathbf{S}(\tilde{\mathbf{D}}_i)^\dagger$ , where  $(\tilde{\mathbf{D}}_i)^\dagger = \tilde{\mathbf{D}}_i^\top(\tilde{\mathbf{D}}_i\tilde{\mathbf{D}}_i^\top)^{-1}$  is the Moore-Penrose pseudo-inverse of  $\tilde{\mathbf{D}}_i$ . Substituting  $\tilde{\mathbf{A}}_i = \mathbf{S}(\tilde{\mathbf{D}}_i)^\dagger$  into the cost function of problem (8), we obtain:

$$\sum_{i=1}^n \|\tilde{\mathbf{A}}_i\tilde{\mathbf{D}}_i - \mathbf{S}\|_F^2 = \sum_{i=1}^n \|\mathbf{S}(\tilde{\mathbf{D}}_i)^\dagger\tilde{\mathbf{D}}_i - \mathbf{S}\|_F^2 = \sum_{i=1}^n \text{tr} \left( \mathbf{S} \left( \mathbf{I} - (\tilde{\mathbf{D}}_i)^\dagger\tilde{\mathbf{D}}_i \right) \mathbf{S}^\top \right),$$

where we have used the fact that matrix  $\mathbf{I} - (\tilde{\mathbf{D}}_i)^\dagger\tilde{\mathbf{D}}_i = \mathbf{I} - \tilde{\mathbf{D}}_i^\top(\tilde{\mathbf{D}}_i\tilde{\mathbf{D}}_i^\top)^{-1}\tilde{\mathbf{D}}_i$  is symmetric and idempotent (because it is the orthogonal projection matrix to the null space of  $\tilde{\mathbf{D}}_i$ ). Let us denote  $\mathbf{Q}_I = \sum_{i=1}^n (\tilde{\mathbf{D}}_i)^\dagger\tilde{\mathbf{D}}_i = \sum_{i=1}^n \tilde{\mathbf{D}}_i^\top(\tilde{\mathbf{D}}_i\tilde{\mathbf{D}}_i^\top)^{-1}\tilde{\mathbf{D}}_i$ . By using  $\mathbf{S}\mathbf{S}^\top = \mathbf{\Lambda}$ , the cost function can be simplified as:

$$\|\tilde{\mathbf{A}}_i\tilde{\mathbf{D}}_i - \mathbf{S}\|_F^2 = n\text{tr}(\mathbf{\Lambda}) - \text{tr}(\mathbf{S}\mathbf{Q}_I\mathbf{S}^\top).$$

Therefore after eliminating  $\tilde{\mathbf{A}}_i$  from the optimization, problem (8) reduces to:

$$\arg \max_{\mathbf{S}} \text{tr}(\mathbf{S}\mathbf{Q}_I\mathbf{S}^\top) \quad \text{s.t.} \quad \mathbf{S}\mathbf{S}^\top = \mathbf{\Lambda}, \quad \mathbf{S}\mathbf{1} = \mathbf{0}. \quad (9)$$

**Lemma 3.**  $\mathbf{Q}_I\mathbf{1} = n\mathbf{1}$ . Moreover if we let  $\mathbf{P}_I = \sum_{i=1}^n \left( \mathbf{I} - (\tilde{\mathbf{D}}_i)^\dagger\tilde{\mathbf{D}}_i \right)$  then  $\mathbf{P}_I\mathbf{1} = \mathbf{0}$ .

*Proof.* The matrix  $(\tilde{\mathbf{D}}_i)^\dagger\tilde{\mathbf{D}}_i = \tilde{\mathbf{D}}_i^\top(\tilde{\mathbf{D}}_i\tilde{\mathbf{D}}_i^\top)^{-1}\tilde{\mathbf{D}}_i$  is the orthogonal projection matrix to the range space of  $\tilde{\mathbf{D}}_i^\top$ . Because  $\mathbf{1} \in \text{Range}(\tilde{\mathbf{D}}_i^\top)$  as  $\mathbf{1}$  is the last column of  $\tilde{\mathbf{D}}_i^\top$ , we have  $\tilde{\mathbf{D}}_i^\top(\tilde{\mathbf{D}}_i\tilde{\mathbf{D}}_i^\top)^{-1}\tilde{\mathbf{D}}_i\mathbf{1} = \mathbf{1}$ , thus  $\mathbf{Q}_I\mathbf{1} = \sum_{i=1}^n \tilde{\mathbf{D}}_i^\top(\tilde{\mathbf{D}}_i\tilde{\mathbf{D}}_i^\top)^{-1}\tilde{\mathbf{D}}_i\mathbf{1} = n\mathbf{1}$ . By noticing  $\mathbf{P}_I + \mathbf{Q}_I = n\mathbf{I}$ , we have  $\mathbf{P}_I\mathbf{1} = n\mathbf{1} - \mathbf{Q}_I\mathbf{1} = \mathbf{0}$ .  $\square$

Therefore problem (9) admits a special property that  $\mathcal{Q}_I$  has an eigenvector  $\mathbf{1}$  because  $\mathcal{Q}_I \mathbf{1} = n\mathbf{1}$ . As a result, this problem admits a closed-form solution by the following proposition whose detailed proof requires Theorem 1 which we will present shortly.

**Proposition 1.** *If  $\mathbf{1}$  is an eigenvector of  $\mathcal{Q}$ , then the globally optimal solution to problem*

$$\arg \max_{\mathbf{S}} \operatorname{tr}(\mathbf{S}\mathcal{Q}\mathbf{S}^\top) \quad \text{s.t.} \quad \mathbf{S}\mathbf{S}^\top = \mathbf{\Lambda}, \quad \mathbf{S}\mathbf{1} = \mathbf{0}, \quad (10)$$

is to scale by  $\sqrt{\mathbf{\Lambda}}$  the  $d$  top eigenvectors of  $\mathcal{Q}$  excluding the vector  $\mathbf{1}$ .

*Proof.* Using  $\mathbf{S}^\top = \mathbf{X}\sqrt{\mathbf{\Lambda}}$ , we can rewrite problem (10) into problem (11) with  $\mathbf{u} = \mathbf{1}$ . By Theorem 1, the optimal  $\mathbf{X}$  is the  $d$  top eigenvectors of  $\mathcal{Q}$  excluding the vector  $\mathbf{1}$ .  $\square$

We summarize the globally optimal solution to the affine GPA formulation (8) as follow.

**Summary 1.** *Problem (8) is solved in closed-form. The optimal reference shape  $\mathbf{S}^*$  is obtained by scaling by  $\sqrt{\mathbf{\Lambda}}$  the  $d$  top eigenvectors of  $\mathcal{Q}_I$  excluding the vector  $\mathbf{1}$ . The optimal affine transformations are given by  $\tilde{\mathbf{A}}_i^* = [\mathbf{A}_i^*, \mathbf{t}_i^*] = \mathbf{S}^* \tilde{\mathbf{D}}_i^\top (\tilde{\mathbf{D}}_i \tilde{\mathbf{D}}_i^\top)^{-1}$  ( $i \in [1 : n]$ ).*

Now we present the supporting results to Proposition 1, *i.e.*, Theorem 1 which gives the solution to problem (11). We start with the following Lemma.

**Lemma 4.** *Let  $\mathbf{M} \in \mathbb{R}^{m \times m}$  be a symmetric matrix, and  $(\alpha_j, \boldsymbol{\xi}_j)$  ( $j \in [1 : m]$ ) the set of eigenvalues and unit eigenvectors of  $\mathbf{M}$  such that  $\mathbf{M}\boldsymbol{\xi}_j = \alpha_j \boldsymbol{\xi}_j$ ,  $\|\boldsymbol{\xi}_j\|_2 = 1$ . Then  $\mathbf{M}$  can be written as:  $\mathbf{M} = \sum_{j=1}^m \alpha_j \boldsymbol{\xi}_j \boldsymbol{\xi}_j^\top$ .*

By Lemma 4, given an arbitrary eigenvector  $\boldsymbol{\xi}_k$  of  $\mathbf{M}$ , by adding multiples of  $\boldsymbol{\xi}_k \boldsymbol{\xi}_k^\top$ , saying  $a \boldsymbol{\xi}_k \boldsymbol{\xi}_k^\top$  ( $a$  is an arbitrary real number) to  $\mathbf{M}$ , we have  $\mathbf{M} + a \boldsymbol{\xi}_k \boldsymbol{\xi}_k^\top = \sum_{j=1, j \neq k}^m \alpha_j \boldsymbol{\xi}_j \boldsymbol{\xi}_j^\top + (\alpha_k + a) \boldsymbol{\xi}_k \boldsymbol{\xi}_k^\top$ . Therefore the matrix  $\mathbf{M} + a \boldsymbol{\xi}_k \boldsymbol{\xi}_k^\top$  has exactly the same set of eigenvectors as  $\mathbf{M}$ . In  $\mathbf{M} + a \boldsymbol{\xi}_k \boldsymbol{\xi}_k^\top$ , the eigenvalue of  $\boldsymbol{\xi}_k$  becomes  $\alpha_k + a$ , while the rest remains unchanged as that in  $\mathbf{M}$ .

**Theorem 1.** *If  $\mathbf{u}$  is an eigenvector of  $\mathcal{Q}$ , then the solution to the optimization problem*

$$\arg \max_{\mathbf{X}} \operatorname{tr}(\mathbf{X}^\top \mathcal{Q} \mathbf{X} \mathbf{\Lambda}) \quad \text{s.t.} \quad \mathbf{X}^\top \mathbf{X} = \mathbf{I}, \quad \mathbf{X}^\top \mathbf{u} = \mathbf{0} \quad (11)$$

is the  $d$  top eigenvectors of  $\mathcal{Q}$  excluding the vector  $\mathbf{u}$ .

*Proof.* Since  $\mathbf{X}^\top \mathbf{u} = \mathbf{0}$ , problem (11) is equivalent to the following:

$$\arg \max_{\mathbf{X}} \operatorname{tr}(\mathbf{X}^\top (\mathcal{Q} - c\mathbf{u}\mathbf{u}^\top) \mathbf{X} \mathbf{\Lambda}) \quad \text{s.t.} \quad \mathbf{X}^\top \mathbf{X} = \mathbf{I}, \quad \mathbf{X}^\top \mathbf{u} = \mathbf{0}, \quad (12)$$



where  $c$  is an arbitrary scalar.

Now we consider the following relaxation of problem (12) without constraint  $\mathbf{X}^\top \mathbf{u} = \mathbf{0}$ , and assume  $c$  is positive and sufficiently large:

$$\arg \max_{\mathbf{X}} \operatorname{tr}(\mathbf{X}^\top (\mathbf{Q} - c\mathbf{u}\mathbf{u}^\top) \mathbf{X} \mathbf{\Lambda}) \quad \text{s.t.} \quad \mathbf{X}^\top \mathbf{X} = \mathbf{I}, \quad (13)$$

which admits a standard Brockett cost function on the Stiefel manifold (see Section 2.3). The optimal solution of problem (13), say  $\mathbf{X}_*$ , comprises the  $d$  top eigenvectors of  $\mathbf{Q} - c\mathbf{u}\mathbf{u}^\top$ .

Since  $\mathbf{u}$  is an eigenvector of  $\mathbf{Q}$ , matrix  $\mathbf{Q} - c\mathbf{u}\mathbf{u}^\top$  has the same set of eigenvectors as  $\mathbf{Q}$ . By assuming  $c$  to be positive and sufficiently large, it is always possible to shift  $\mathbf{u}$  to the bottom-eigenvector of  $\mathbf{Q} - c\mathbf{u}\mathbf{u}^\top$ , thus the eigenvector  $\mathbf{u}$  is always excluded from  $\mathbf{X}_*$ . In other words, the  $d$  top eigenvectors of  $\mathbf{Q} - c\mathbf{u}\mathbf{u}^\top$  are the  $d$  top eigenvectors of  $\mathbf{Q}$  excluding the eigenvector  $\mathbf{u}$ .

Lastly, the eigenvectors of  $\mathbf{Q} - c\mathbf{u}\mathbf{u}^\top$  with respect to different eigenvalues are orthogonal thus  $\mathbf{X}_*^\top \mathbf{u} = \mathbf{0}$ . Therefore problem (13) is a tight relaxation of problem (12) if  $c$  is positive and sufficiently large.  $\square$

### 4.3 Estimation of the Reference Covariance Prior

We estimate the reference covariance prior  $\mathbf{\Lambda}$  using the eigenvalues of the datum shape covariance matrices  $\operatorname{Cov}(\mathbf{D}_i \mathbf{D}_i^\top)$  ( $i \in [1 : n]$ ). By zero-centering each datum shape  $\mathbf{D}_i$  as  $\bar{\mathbf{D}}_i = \mathbf{D}_i - \frac{1}{m} \mathbf{D}_i \mathbf{1} \mathbf{1}^\top$ , we have  $\operatorname{Cov}(\mathbf{D}_i \mathbf{D}_i^\top) = \bar{\mathbf{D}}_i \bar{\mathbf{D}}_i^\top$ . Since  $\mathbf{\Lambda}$  is diagonal, we denote  $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_d)$  and define the vector  $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_d]^\top$ . Abusing notations, we collect the eigenvalues of each datum shape covariance  $\mathbf{D}_i \mathbf{D}_i^\top$  by a diagonal matrix  $\mathbf{\Lambda}_i$  and define the vector  $\boldsymbol{\lambda}_i = [\lambda_1^{(i)}, \lambda_2^{(i)}, \dots, \lambda_d^{(i)}]^\top$ . Without loss of generality, we assume that the elements in  $\boldsymbol{\lambda}_i$  have been sorted in the descending (or non-ascending) order such that  $\lambda_1^{(i)} \geq \lambda_2^{(i)} \geq \dots \geq \lambda_d^{(i)} \geq 0$ . The task is now to estimate  $\boldsymbol{\lambda}$  from  $\boldsymbol{\lambda}_i$  ( $i \in [1 : n]$ ).

To proceed, we consider the geometric implication of  $\sqrt{\boldsymbol{\lambda}_i} = \left[ \sqrt{\lambda_1^{(i)}}, \sqrt{\lambda_2^{(i)}}, \dots, \sqrt{\lambda_d^{(i)}} \right]^\top$ , *i.e.*, the vector comprising the  $d$  leftmost singular values of shape  $\bar{\mathbf{D}}_i$ . It has been known in (Horn, 1987) that the scale of the shape  $\mathbf{D}_i$  can be represented by  $\frac{1}{\sqrt{m}} \|\bar{\mathbf{D}}_i\|_F$ . Dropping the common constant  $\frac{1}{\sqrt{m}}$ , we consider:

$$\|\bar{\mathbf{D}}_i\|_F = \sqrt{\operatorname{tr}(\bar{\mathbf{D}}_i \bar{\mathbf{D}}_i^\top)} = \sqrt{\operatorname{tr}(\mathbf{\Lambda}_i)} = \sqrt{\lambda_1^{(i)} + \lambda_2^{(i)} + \dots + \lambda_d^{(i)}} = \|\sqrt{\boldsymbol{\lambda}_i}\|_2.$$

This suggests that  $\sqrt{\boldsymbol{\lambda}_i}$  is a vector with each of its component representing the scale along the corresponding Euclidean axis, and its length the overall shape scale. Given  $\sqrt{\boldsymbol{\lambda}_1}, \sqrt{\boldsymbol{\lambda}_2}, \dots, \sqrt{\boldsymbol{\lambda}_n}$  from  $n$  datum shapes, we are interested in finding  $\sqrt{\boldsymbol{\lambda}} = [\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_d}]^\top$ .

The reference shape is defined up to scale (*i.e.*, up to a similarity transformation), thus in  $\sqrt{\boldsymbol{\lambda}}$ , all that

matters is the proportion of its components, namely the direction of the vector  $\sqrt{\lambda}$ . Therefore we propose to estimate the direction of  $\sqrt{\lambda}$  on the unit ball, denoted by  $\boldsymbol{\theta}$ , by minimizing the angles between  $\boldsymbol{\theta}$  and each  $\sqrt{\lambda_i}$  via maximizing their inner products as:

$$\boldsymbol{\theta}^* = \arg \max_{\boldsymbol{\theta} \in \mathbb{R}^d} \sum_{i=1}^n \left( \boldsymbol{\theta}^\top \frac{\sqrt{\lambda_i}}{\|\sqrt{\lambda_i}\|_2} \right)^2 \quad \text{s.t.} \quad \|\boldsymbol{\theta}\|_2 = 1. \quad (14)$$

Upon defining  $\mathbf{\Pi} = \begin{bmatrix} \frac{\sqrt{\lambda_1}}{\|\sqrt{\lambda_1}\|_2} & \frac{\sqrt{\lambda_2}}{\|\sqrt{\lambda_2}\|_2} & \dots & \frac{\sqrt{\lambda_n}}{\|\sqrt{\lambda_n}\|_2} \end{bmatrix}$ , problem (14) can be rewritten as the Rayleigh quotient optimization:

$$\boldsymbol{\theta}^* = \arg \max_{\boldsymbol{\theta} \in \mathbb{R}^d} \boldsymbol{\theta}^\top \mathbf{\Pi} \mathbf{\Pi}^\top \boldsymbol{\theta} \quad \text{s.t.} \quad \|\boldsymbol{\theta}\|_2 = 1,$$

whose solution is given in closed-form, which is the leftmost left singular vector of the matrix  $\mathbf{\Pi}$ , or equivalently the top eigenvector of the matrix  $\mathbf{\Pi} \mathbf{\Pi}^\top$ . Since  $\mathbf{\Pi} \mathbf{\Pi}^\top$  has non-negative elements, by the Perron–Frobenius theorem, the elements of  $\boldsymbol{\theta}^*$  can be chosen all non-negative.

Given  $\boldsymbol{\theta}^*$ , we can choose  $\sqrt{\lambda} = s\boldsymbol{\theta}^*$  up to a scale factor  $s > 0$ . In practice, it is desirable to choose  $\sqrt{\lambda}$  to be at the similar scale of  $\sqrt{\lambda_i}$ . Therefore, we choose the average scale  $s = \frac{1}{n} \sum_{i=1}^n \|\sqrt{\lambda_i}\|_2$ .

#### 4.4 The Coordinate Transformation of Datum Shapes

We show that by applying (possibly distinct) arbitrary rigid transformations to each datum shape  $D_i$ , the optimal reference shape of problem (8) remains unchanged. Thus formulation (8) is unbiased when facing coordinate transformations.

**Lemma 5.** *Let  $D' = RD + t\mathbf{1}^\top$ . We further denote  $\tilde{D} = \begin{bmatrix} D \\ \mathbf{1}^\top \end{bmatrix}$  and  $\tilde{D}' = \begin{bmatrix} D' \\ \mathbf{1}^\top \end{bmatrix}$ . Then we have  $\mathbf{Range}(\tilde{D}'^\top) = \mathbf{Range}(\tilde{D}^\top)$  and the orthogonal projection matrices thus satisfy:*

$$\tilde{D}'^\top (\tilde{D}' \tilde{D}'^\top)^{-1} \tilde{D}' = \tilde{D}^\top (\tilde{D} \tilde{D}^\top)^{-1} \tilde{D}.$$

*Proof.* The matrices  $\tilde{D}'^\top$  and  $\tilde{D}^\top$  have the same range space, i.e.,  $\mathbf{Range}(\tilde{D}'^\top) = \mathbf{Range}(\tilde{D}^\top)$ , because:

$$\tilde{D}'^\top = \begin{bmatrix} D^\top R^\top + \mathbf{1}t^\top & \mathbf{1} \end{bmatrix} = \tilde{D}^\top \begin{bmatrix} R^\top & \mathbf{0} \\ t^\top & 1 \end{bmatrix}.$$

The orthogonal projection matrices (also called orthogonal projectors) to  $\mathbf{Range}(\tilde{D}'^\top)$  and  $\mathbf{Range}(\tilde{D}^\top)$  are the same by uniqueness (Meyer, 2000).  $\square$

**Proposition 2.** *In problem (8), when we apply arbitrary rigid transformations to each datum shape  $\mathbf{D}_i$ , the matrix  $\mathbf{Q}_I$  remains the same. So does the optimal reference shape.*

*Proof.* By Lemma 5, matrix  $\tilde{\mathbf{D}}_i^\top (\tilde{\mathbf{D}}_i \tilde{\mathbf{D}}_i^\top)^{-1} \tilde{\mathbf{D}}_i$  remains the same when we apply arbitrary rigid transformations to the datum shape  $\mathbf{D}_i$ . Thus  $\mathbf{Q}_I = \sum_{i=1}^n \tilde{\mathbf{D}}_i^\top (\tilde{\mathbf{D}}_i \tilde{\mathbf{D}}_i^\top)^{-1} \tilde{\mathbf{D}}_i$  remains the same as well.  $\square$

#### 4.5 Connection to Classical Results by Eliminating Translations

We recapitulate the key idea of affine GPA in (Rohlf and Slice, 1990) as follows in formulation (15), and show that this approach attains the same result as formulation (8) while the latter is more general. In formulation (8), the optimal  $\mathbf{t}_i$  given  $\mathbf{A}_i$  and  $\mathbf{S}$  is  $\mathbf{t}_i = -\frac{1}{m}(\mathbf{A}_i \mathbf{D}_i - \mathbf{S})\mathbf{1}$ . Moreover if  $\mathbf{S}\mathbf{1} = \mathbf{0}$ , we have  $\mathbf{t}_i = -\frac{1}{m}\mathbf{A}_i \mathbf{D}_i \mathbf{1}$ . Substituting the estimate  $\mathbf{t}_i = -\frac{1}{m}\mathbf{A}_i \mathbf{D}_i \mathbf{1}$  back to formulation (15), we have:

$$\begin{cases} \arg \min_{\{\mathbf{A}_i\}, \mathbf{S}} & \sum_{i=1}^n \|\mathbf{A}_i \bar{\mathbf{D}}_i - \mathbf{S}\|_F^2 \\ \text{s.t.} & \mathbf{S}\mathbf{S}^\top = \mathbf{\Lambda}, \end{cases} \quad (15)$$

where  $\bar{\mathbf{D}}_i = \mathbf{D}_i - \frac{1}{m}\mathbf{D}_i \mathbf{1}\mathbf{1}^\top$  ( $i \in [1 : n]$ ) are zero-centered datum shapes. Let  $\mathbf{\Omega}_o = \sum_{i=1}^n \bar{\mathbf{D}}_i^\top (\bar{\mathbf{D}}_i \bar{\mathbf{D}}_i^\top)^{-1} \bar{\mathbf{D}}_i$ . Following a similar derivation to Section 4.2, we can reduce problem (15) to formulation (6) with  $\mathbf{Q} = \mathbf{\Omega}_o$ , thus the optimal reference shape  $\mathbf{S}$  of problem (15) is to scale the  $d$  top eigenvectors of  $\mathbf{\Omega}_o$  by  $\sqrt{\mathbf{\Lambda}}$ . We drop the constraint  $\mathbf{S}\mathbf{1} = \mathbf{0}$  in formulation (15) because this constraint is automatically satisfied for the optimal  $\mathbf{S}$  of formulation (15), as  $\mathbf{\Omega}_o \mathbf{1} = \mathbf{0}$  thus  $\mathbf{1}$  is orthogonal to the  $d$  top eigenvectors corresponding to nonzero eigenvalues.

**Proposition 3.**  $\mathbf{Q}_I = \frac{n}{m}\mathbf{1}\mathbf{1}^\top + \mathbf{\Omega}_o$ . *The  $d$  top eigenvectors of  $\mathbf{Q}_I$  excluding the vector  $\mathbf{1}$  are the  $d$  top eigenvectors of  $\mathbf{\Omega}_o$ , thus problems (8) and (15) give the same optimal reference shape  $\mathbf{S}$ .*

*Proof.* Let us denote the homogeneous form of the zero-centered datum shape as:

$$\tilde{\mathbf{D}}_i = \begin{bmatrix} \bar{\mathbf{D}}_i \\ \mathbf{1}^\top \end{bmatrix}, \quad \text{thus} \quad \tilde{\mathbf{D}}_i \tilde{\mathbf{D}}_i^\top = \begin{bmatrix} \bar{\mathbf{D}}_i \bar{\mathbf{D}}_i^\top & \mathbf{0} \\ \mathbf{0} & m \end{bmatrix},$$

where we have used the fact that  $\bar{\mathbf{D}}_i \mathbf{1} = \mathbf{0}$ . From Proposition 2, we have  $\mathbf{Q}_I = \sum_{i=1}^n \tilde{\mathbf{D}}_i^\top (\tilde{\mathbf{D}}_i \tilde{\mathbf{D}}_i^\top)^{-1} \tilde{\mathbf{D}}_i = \sum_{i=1}^n \tilde{\mathbf{D}}_i^\top (\tilde{\mathbf{D}}_i \tilde{\mathbf{D}}_i^\top)^{-1} \tilde{\mathbf{D}}_i$ . By a straightforward calculation, it can be verified that:

$$\tilde{\mathbf{D}}_i^\top (\tilde{\mathbf{D}}_i \tilde{\mathbf{D}}_i^\top)^{-1} \tilde{\mathbf{D}}_i = \frac{1}{m}\mathbf{1}\mathbf{1}^\top + \bar{\mathbf{D}}_i^\top (\bar{\mathbf{D}}_i \bar{\mathbf{D}}_i^\top)^{-1} \bar{\mathbf{D}}_i,$$

thus we obtain  $\mathbf{Q}_I = \frac{n}{m}\mathbf{1}\mathbf{1}^\top + \mathbf{\Omega}_o$ . We notice  $\mathbf{\Omega}_o \mathbf{1} = \mathbf{0}$  because of  $\bar{\mathbf{D}}_i \mathbf{1} = \mathbf{0}$ , which means  $\mathbf{1}$  is an eigenvector

of  $\mathfrak{Q}_o$  with eigenvalue 0. Therefore  $\mathfrak{Q}_I$  and  $\mathfrak{Q}_o$  have the same eigenvectors by Lemma 4, while  $\mathbf{1}$  corresponds to the largest eigenvalue  $n$  in  $\mathfrak{Q}_I$  and in contrast to the smallest eigenvalue 0 in  $\mathfrak{Q}_o$ .  $\square$

## 5 Generalized Procrustes Analysis with the Deformation Model

We consider GPA with the deformation model as nonlinear warps. In particular, we consider a class of generalized warps, termed LBWs, whose transformation parameters are linear with respect to the nonlinear basis functions that lift source points to the higher-dimensional feature space. In this section, we consider full shape registration and postpone the discussion of partial shape registration to Section 6.

### 5.1 Linear Basis Warps

A warp is a generalized transformation, that maps a source point  $\mathbf{p} \in \mathbb{R}^d$  ( $d \in \{2, 3\}$ ) to its target point  $\mathbf{p}' \in \mathbb{R}^d$ . The LBW is expressed as a linear combination of a set of basis functions.

#### 5.1.1 Formulation

Let  $\boldsymbol{\beta}(\mathbf{p})$  be a vector of basis functions:

$$\boldsymbol{\beta}(\mathbf{p}) = [\beta_1(\mathbf{p}), \beta_2(\mathbf{p}), \dots, \beta_l(\mathbf{p})]^\top,$$

with each element  $\beta_k(\mathbf{p}) : \mathbb{R}^d \mapsto \mathbb{R}$  ( $k \in [1 : l]$ ) being a scalar basis function. The vectorized basis function  $\boldsymbol{\beta}(\mathbf{p}) : \mathbb{R}^d \mapsto \mathbb{R}^l$  brings a  $d$ -dimensional point  $\mathbf{p}$  to the  $l$ -dimensional feature space, thus is also termed a feature mapping in the context of linear regression models (Bishop, 2006). Given the basis function  $\boldsymbol{\beta}(\mathbf{p})$ , we write a warp model,  $\mathcal{W}(\mathbf{p}, \mathbf{W}) : \mathbb{R}^d \mapsto \mathbb{R}^d$ , as:

$$\mathcal{W}(\mathbf{p}, \mathbf{W}) = \mathbf{W}^\top \boldsymbol{\beta}(\mathbf{p}), \tag{16}$$

with  $\mathbf{W} \in \mathbb{R}^{l \times d}$  being the unknown weight matrix. An example is provided in Appendix A.2, showing how to write the TPS warp in this form.

#### 5.1.2 Operating on Point Clouds

For a point-cloud of  $m$  points in a matrix  $\mathbf{D} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_m] \in \mathbb{R}^{d \times m}$ . We apply the warp  $\mathcal{W}(\mathbf{p}, \mathbf{W})$  to each point in  $\mathbf{D}$  to obtain its warped version. Abusing notations, we write the result as:

$$\mathcal{W}(\mathbf{D}, \mathbf{W}) = \mathbf{W}^\top \mathcal{B}(\mathbf{D}), \tag{17}$$

with  $\mathcal{B}(\mathbf{D}) \in \mathbb{R}^{l \times m}$  collecting the feature of each point in  $\mathbf{D}$  as its columns:

$$\mathcal{B}(\mathbf{D}) = [\beta(\mathbf{p}_1), \beta(\mathbf{p}_2), \dots, \beta(\mathbf{p}_m)].$$

### 5.1.3 Regularization

The warp is often used with a regularization term to avoid over-fitting, for instance, if the dimension of the feature space is greater or equivalent to the number of points in the point-cloud. In the context of deformations, such a term is formed from the partial derivatives of the warp, with different physical implications. In particular, the second-order derivatives are used in the TPS warp:

$$\mathcal{R}(\mathbf{W}) = \int_{\mathbb{R}^d} \left\| \frac{\partial^2}{\partial \mathbf{p}^2} \mathcal{W}(\mathbf{p}, \mathbf{W}) \right\|_F^2 d\mathbf{p}.$$

This is directly proportional to the bending energy. The bending energy term is exactly zero if and only if the warp is affine (Bookstein, 1989). Other possibilities of regularizations include the spring term used in elastic registration (Christensen and He, 2001), and the viscosity term used in fluid registration (Bro-Nielsen and Gramkow, 1996).

For the TPS warp, the integral can be solved in closed-form, as a quadratic form of the transformation parameters:

$$\mathcal{R}(\mathbf{W}) = \|\mathbf{Z}\mathbf{W}\|_F^2. \quad (18)$$

Here  $\mathbf{Z}$  is given as the square root of the bending energy matrix, see Appendix A.3. For other warps, we assume the regularization term can be fairly approximated by the quadratic form as well.

The regularization term is however not compulsory. It is always possible to avoid over-fitting by limiting the dimension of the feature space, by choosing a smaller  $l \ll m$  (Rueckert et al., 1999). For completeness of the discussion, we consider the case with regularization.

### 5.1.4 Examples of Linear Basis Warps

The LBW generalizes over many deformation models, *e.g.* the Free-Form Deformations (FFD) (Rueckert et al., 1999; Szeliski and Coughlan, 1997), and the Radial Basis Functions (RBF) (Bookstein, 1989; Fornefett et al., 2001). Concretely we will use the Thin-Plate Spline (TPS) (Bookstein, 1989; Duchon, 1976), a theoretically principled RBF that minimizes the overall bending energy, in the practical implementation of our theory. An introduction of the TPS warp as an LBW is provided in Appendix A.

The affine transformation is a special case of the LBW without regularization. This can be shown from

the homogeneous form in equation (7), by setting  $\mathbf{W}^\top = \begin{bmatrix} \mathbf{A} & \mathbf{t} \end{bmatrix}$ ,  $\mathcal{B}(D) = \begin{bmatrix} D \\ \mathbf{1}^\top \end{bmatrix}$ .

## 5.2 Generalized Procrustes Analysis with Linear Basis Warps

We consider the case without missing datum points, and use the LBW in formulation (2). We constrain the reference shape to be zero centered by  $\mathbf{S}\mathbf{1} = \mathbf{0}$ . The shape constraint  $\mathbf{S}\mathbf{S}^\top = \mathbf{\Lambda}$  is used to enforce the rigidity of the solution, and the reference covariance prior  $\mathbf{\Lambda}$  is estimated as in Section 4.3. We propose the following formulation for deformable GPA:

$$\text{II} : \begin{cases} \arg \min_{\{\mathbf{W}_i\}, \mathbf{S}} & \sum_{i=1}^n \|\mathbf{W}_i^\top \mathcal{B}_i(D_i) - \mathbf{S}\|_F^2 + \sum_{i=1}^n \mu_i \|\mathbf{Z}_i \mathbf{W}_i\|_F^2 \\ \text{s.t.} & \mathbf{S}\mathbf{S}^\top = \mathbf{\Lambda}, \quad \mathbf{S}\mathbf{1} = \mathbf{0}. \end{cases} \quad (19)$$

Although being termed deformable GPA, formulation (19) includes the affine-GPA formulation (8) in homogeneous form as a special case. The translation part cannot be identified directly from the LBW, thus needs to be estimated jointly inside the transformation parameters.

To proceed, we define the shorthand  $\mathcal{B}_i \stackrel{\text{def}}{=} \mathcal{B}_i(D_i)$ . The problem is separable. Given a reference shape  $\mathbf{S}$ , the estimate of  $\mathbf{W}_i$  is given in closed-form by solving the linear least-squares problem whose solution is:

$$\mathbf{W}_i = (\mathcal{B}_i \mathcal{B}_i^\top + \mu_i \mathbf{Z}_i^\top \mathbf{Z}_i)^{-1} \mathcal{B}_i \mathbf{S}^\top. \quad (20)$$

Substituting equation (20) into problem (19), we obtain an optimization problem with respect to  $\mathbf{S}$  only:

$$\arg \min_{\mathbf{S}} \text{tr}(\mathbf{S} \mathcal{P}_{\text{II}} \mathbf{S}^\top) \quad \text{s.t.} \quad \mathbf{S}\mathbf{S}^\top = \mathbf{\Lambda}, \quad \mathbf{S}\mathbf{1} = \mathbf{0}, \quad (21)$$

where matrix  $\mathcal{P}_{\text{II}}$  is defined as:

$$\mathcal{P}_{\text{II}} = \sum_{i=1}^n \left( \mathbf{I} - \mathcal{B}_i^\top (\mathcal{B}_i \mathcal{B}_i^\top + \mu_i \mathbf{Z}_i^\top \mathbf{Z}_i)^{-1} \mathcal{B}_i \right).$$

We define  $\mathbf{Q}_i = \mathcal{B}_i^\top (\mathcal{B}_i \mathcal{B}_i^\top + \mu_i \mathbf{Z}_i^\top \mathbf{Z}_i)^{-1} \mathcal{B}_i$ , and:

$$\mathcal{Q}_{\text{II}} = \sum_{i=1}^n \mathcal{B}_i^\top (\mathcal{B}_i \mathcal{B}_i^\top + \mu_i \mathbf{Z}_i^\top \mathbf{Z}_i)^{-1} \mathcal{B}_i = \sum_{i=1}^n \mathbf{Q}_i.$$

Then  $\mathcal{P}_{\text{II}} = \sum_{i=1}^n (\mathbf{I} - \mathbf{Q}_i) = n\mathbf{I} - \sum_{i=1}^n \mathbf{Q}_i$ , and  $\text{tr}(\mathbf{S} \mathcal{P}_{\text{II}} \mathbf{S}^\top) = n \text{tr}(\mathbf{\Lambda}) - \text{tr}(\mathbf{S} \mathcal{Q}_{\text{II}} \mathbf{S}^\top)$ . Problem (21) can

thus be equivalently written as a maximization problem:

$$\arg \max_{\mathbf{S}} \operatorname{tr}(\mathbf{S}\mathbf{Q}_{\text{II}}\mathbf{S}^{\top}) \quad \text{s.t.} \quad \mathbf{S}\mathbf{S}^{\top} = \mathbf{\Lambda}, \quad \mathbf{S}\mathbf{1} = \mathbf{0}. \quad (22)$$

If  $\mathbf{1}$  is an eigenvector of  $\mathbf{Q}_{\text{II}}$ , or equivalently if  $\mathbf{1}$  is an eigenvector of  $\mathcal{P}_{\text{II}}$ , then Problem (21) and Problem (22) can be solved globally by Proposition 1. Once we obtain the optimal reference shape  $\mathbf{S}^*$ , the optimal transformation parameters can be calculated by equation (20).

### 5.3 Eigenvector Characterization and Globally Optimal Solution

Now we characterize a class of LBWs for which the resulting  $\mathcal{P}_{\text{II}}$  and  $\mathbf{Q}_{\text{II}}$  have an eigenvector  $\mathbf{1}$ . To this end, we prove that the following statements are equivalent. The proofs have been moved to Appendix B to benefit easy reading.

**Theorem 2.** *The following statements are equivalent:*

- (a)  $\mathcal{P}_{\text{II}}\mathbf{1} = \mathbf{0}$ .
- (b)  $\mathbf{Q}_{\text{II}}\mathbf{1} = n\mathbf{1}$ .
- (c)  $\mathbf{Q}_i\mathbf{1} = \mathbf{1}$ .
- (d) *The cost function  $\operatorname{tr}(\mathbf{S}\mathcal{P}_{\text{II}}\mathbf{S}^{\top})$  is invariant to translations.*
- (e) *There exists  $\mathbf{x}$  such that  $\mathcal{B}_i(\mathbf{D}_i)^{\top}\mathbf{x} = \mathbf{1}$ ; moreover if  $\mu_i > 0$ ,  $\mathbf{x}$  must satisfy  $\mathbf{Z}_i\mathbf{x} = \mathbf{0}$ .*

Theorem 2 relates various aspects of the LBW based GPA to the existence of an eigenvector  $\mathbf{1}$  in  $\mathcal{P}_{\text{II}}$  and  $\mathbf{Q}_{\text{II}}$ . In particular, case (d) states that the reduced problem is invariant to translations, and case (e) stipulates the rules that the LBW as a mapping must follow. While case (e) in Theorem 2 is related to the datum shape  $\mathbf{D}_i$ , we show in what follows that it can be satisfied if the LBW satisfies certain property, making the statement independent of the input datum shapes.

Now we show a sufficient condition to case (e) in Theorem 2 which is that the LBW must contain *free-translations*. We drop the subscript  $i$  and use the notation  $\mathcal{B}(\cdot)$  to indicate that this is a property of the LBW thus is independent of the warp input.

**Definition 1.** *The LBW, given by  $\{\mathbf{W}^{\top}\mathcal{B}(\cdot), \mathbf{Z}\mathbf{W}, \mu\}$  is said to contain free-translations if: there exists  $\mathbf{x}$  such that  $\mathcal{B}(\cdot)^{\top}\mathbf{x} = \mathbf{1}$ ; moreover if  $\mu > 0$ ,  $\mathbf{x}$  must satisfy  $\mathbf{Z}\mathbf{x} = \mathbf{0}$ .*

We now explain the idea behind Definition 1. The LBW contains free-translations, if we can recover the translation vector  $\mathbf{t}$  explicitly by an invertible matrix  $\mathbf{G}$  such that:

$$\mathbf{W}^\top \mathcal{B}(\cdot) = \mathbf{W}^\top \mathbf{G}^{-1} (\mathcal{B}(\cdot)^\top \mathbf{G}^\top)^\top = \underbrace{[\cdots \mathbf{t} \cdots]}_{\mathbf{W}^\top \mathbf{G}^{-1}} \underbrace{[\cdots \mathbf{1} \cdots]}_{\mathcal{B}(\cdot)^\top \mathbf{G}^\top}^\top,$$

where  $\mathbf{W}^\top \mathbf{G}^{-1}$  contains a column vector  $\mathbf{t}$  and  $\mathcal{B}(\cdot)^\top \mathbf{G}^\top$  contains a column vector  $\mathbf{1}$ . Therefore  $\mathbf{1} \in \mathbf{Range}(\mathcal{B}(\cdot)^\top \mathbf{G}^\top) = \mathbf{Range}(\mathcal{B}(\cdot)^\top)$ . Using  $\mathbf{G}$ , the term  $\mathbf{Z}\mathbf{W}$  can be decomposed as:

$$\mathbf{Z}\mathbf{W} = \mathbf{Z}\mathbf{G}^\top (\mathbf{W}^\top \mathbf{G}^{-1})^\top = \mathbf{Z}\mathbf{G}^\top \underbrace{[\cdots \mathbf{t} \cdots]}_{\mathbf{W}^\top \mathbf{G}^{-1}}^\top.$$

Without loss of generality, we assume that in  $\mathcal{B}(\cdot)^\top \mathbf{G}^\top$  the  $k$ -th column vector is  $\mathbf{1}$ , identified by the standard basis vector  $\mathbf{e}_k$ , such that  $\mathcal{B}(\cdot)^\top \mathbf{G}^\top \mathbf{e}_k = \mathbf{1}$ . Then it can be easily verified that the translation  $\mathbf{t}$  is constraint-free in  $\mathbf{Z}\mathbf{W}$  if and only if the  $k$ -th column of  $\mathbf{Z}\mathbf{G}^\top$  is  $\mathbf{0}$ , *i.e.*,  $\mathbf{Z}\mathbf{G}^\top \mathbf{e}_k = \mathbf{0}$ . Since  $\mathcal{B}(\cdot)^\top$  has full column rank, so  $\mathbf{G}^\top \mathbf{e}_k$  is the unique  $\mathbf{x}$  such that  $\mathcal{B}(\cdot)^\top \mathbf{x} = \mathbf{1}$ ,  $\mathbf{Z}\mathbf{x} = \mathbf{0}$ .

**Proposition 4.** *The statements in Theorem 2 are satisfied if the LBW contains free-translations.*

We state that the affine transformation and the TPS warp satisfy Definition 1. The proof for the affine case is straightforward. For the TPS warp, the algebraic proof is given in Appendix A.4. Intuitively, this can be explained because for the TPS warp, the bending energy term only affects the nonlinear part of the warp, which means that the linear part of the TPS warp is constraint-free and so is its translational component.

**Proposition 5.** *The affine transformation and the TPS warp satisfy the statement of Theorem 2.*

The above results are summarized as follow:

**Summary 2.** *Formulation (19) can be solved globally if the the LBW contains free-translations. The optimal reference shape  $\mathbf{S}^*$  is to scale by  $\sqrt{\Lambda}$  the  $d$  top eigenvectors of  $\mathcal{Q}_{\text{II}}$  (or equivalently the  $d$  bottom eigenvectors of  $\mathcal{P}_{\text{II}}$ ) excluding the eigenvector  $\mathbf{1}$ . The optimal transformation parameters are  $\mathbf{W}_i = (\mathcal{B}_i \mathcal{B}_i^\top + \mu_i \mathbf{Z}_i^\top \mathbf{Z}_i)^{-1} \mathcal{B}_i \mathbf{S}^{*\top}$  ( $i \in [1 : n]$ ).*

## 5.4 Reformulation using the Soft Constraint

In Theorem 1, we have proved the equivalence of formulations (11) and (13). Therefore, formulation (22) can be equivalently written as:

$$\arg \max_{\mathbf{S}} \text{tr}(\mathbf{S}(\mathcal{Q}_{\text{II}} - n\bar{\mathbf{I}}\bar{\mathbf{I}}^\top)\mathbf{S}^\top) \quad \text{s.t.} \quad \mathbf{S}\mathbf{S}^\top = \Lambda, \quad (23)$$



where we have used  $\mathcal{Q}_{\text{II}}\bar{\mathbf{1}} = n\bar{\mathbf{1}}$ , with  $\bar{\mathbf{1}}$  being normalized.

In problem (23), we can replace  $n$  with any  $n' \geq n$ . This is because  $(\mathcal{Q}_{\text{II}} - n'\bar{\mathbf{1}}\bar{\mathbf{1}}^\top)\bar{\mathbf{1}} = (n - n')\bar{\mathbf{1}}$ , where the eigenvalue corresponding to the eigenvector  $\bar{\mathbf{1}}$  becomes  $(n - n') \leq 0$ , while the rest of the eigenvalue of  $(\mathcal{Q}_{\text{II}} - \nu\bar{\mathbf{1}}\bar{\mathbf{1}}^\top)$  is nonnegative. Thus the eigenvector  $\bar{\mathbf{1}}$  is excluded in the solution for any  $n' \geq n$ . The rest of the eigenvectors remain unchanged. Expanding the cost of problem (23), we obtain:

$$\arg \max_{\mathbf{S}} \text{tr}(\mathbf{S}(\mathcal{Q}_{\text{II}})\mathbf{S}^\top) - n'\|\mathbf{S}\bar{\mathbf{1}}\|_2^2 \quad \text{s.t.} \quad \mathbf{S}\mathbf{S}^\top = \mathbf{\Lambda}, \quad (24)$$

where  $n' \geq n$ . Finally we substitute  $\mathcal{P}_{\text{II}} = n\mathbf{I} - \mathcal{Q}_{\text{II}}$  and  $\bar{\mathbf{1}} = \frac{1}{m}\mathbf{1}$ , then obtain a minimization problem:

$$\arg \min_{\mathbf{S}} \text{tr}(\mathbf{S}(\mathcal{P}_{\text{II}})\mathbf{S}^\top) + \nu\|\mathbf{S}\mathbf{1}\|_2^2 \quad \text{s.t.} \quad \mathbf{S}\mathbf{S}^\top = \mathbf{\Lambda}, \quad (25)$$

where  $\nu \geq n/m$ . Note that problem (25) is equivalent to problem (21), while the hard constraint  $\mathbf{S}\mathbf{1} = \mathbf{0}$  has been reformulated as a soft constraint in the form of a penalty term.

At last, formulation (19) is equivalent to the following one with the soft constraint:

$$\begin{aligned} \arg \min_{\{\mathbf{W}_i\}, \mathbf{S}} & \sum_{i=1}^n \|\mathbf{W}_i^\top \mathcal{B}_i(\mathbf{D}_i) - \mathbf{S}\|_F^2 + \sum_{i=1}^n \mu_i \|\mathbf{Z}_i \mathbf{W}_i\|_F^2 + \nu\|\mathbf{S}\mathbf{1}\|_2^2 \\ \text{s.t.} & \quad \mathbf{S}\mathbf{S}^\top = \mathbf{\Lambda}. \end{aligned} \quad (26)$$

The equivalence can be easily shown by noting that given  $\mathbf{S}$ , the term  $\nu\|\mathbf{S}\mathbf{1}\|_2^2$  becomes a constant, thus the relation between the estimates of  $\mathbf{W}_i$  and  $\mathbf{S}$  remains the same. Then with an analogous derivation to Section 5.2, after eliminating  $\mathbf{W}_i$ , problem (26) is reduced to problem (25).

It is worth mentioning that formulation (26) is equivalent to formulation (19) for any LBW that satisfies Theorem 2 (for example, the affine transformation and the TPS warp). However formulation (26) is more tractable in terms of solution methods, since there is no need to take care of the hard constraint  $\mathbf{S}\mathbf{1} = \mathbf{0}$ , which can be difficult if Theorem 2 is not satisfied (*e.g.*, the translation part is constrained). In any case, formulation (26) can be reduced to a Brockett cost function, which can be easily solved globally.

## 6 Generalized Procrustes Analysis with Partial Shapes

Now we have equipped with enough insights to discuss GPA with partial shapes. We will follow the soft regularized method discussed in Section 5.4, which generalizes over the cases where Theorem 2 is not satisfied. As the affine transformation is a special case of the LBW, we consider GPA with the LBW only.

**Algorithm 1:** Estimate Reference Covariance Prior

---

```

1 function EstimateReferenceCovariancePrior( $\mathfrak{D}_1, \mathfrak{D}_2, \dots, \mathfrak{D}_n$ )
2   for  $i \in [1 : n]$  do
3      $\bar{\mathfrak{D}}_i = \mathfrak{D}_i - \frac{1}{m} \mathfrak{D}_i \mathbf{1} \mathbf{1}^\top$  /* Zero-centered shapes */
4      $\sqrt{\lambda_i} \leftarrow d\text{-LeftmostSingularValues}(\bar{\mathfrak{D}}_i)$  /* Elements in non-ascending order */
5   end
6    $\mathbf{\Pi} \leftarrow \begin{bmatrix} \frac{\sqrt{\lambda_1}}{\|\sqrt{\lambda_1}\|_2} & \frac{\sqrt{\lambda_2}}{\|\sqrt{\lambda_2}\|_2} & \dots & \frac{\sqrt{\lambda_n}}{\|\sqrt{\lambda_n}\|_2} \end{bmatrix}$ 
7    $\boldsymbol{\theta}^* \leftarrow \text{LeftmostSingularVector}(\mathbf{\Pi})$ 
8    $s \leftarrow \frac{1}{n} \sum_{i=1}^n \|\sqrt{\lambda_i}\|_2$  /* Use average scale */
9    $\sqrt{\lambda} \leftarrow s \boldsymbol{\theta}^*$ 
10   $\sqrt{\Lambda} \leftarrow \text{diag}(\sqrt{\lambda})$  /* Construct diagonal matrix */
11   $\Lambda \leftarrow \sqrt{\Lambda} \sqrt{\Lambda}$ 
12  return  $\Lambda$ 
13 end

```

---

**6.1 Estimating the Reference Covariance Prior**

The reference shape  $\mathbf{S}$  is a full shape. In order to estimate the reference covariance prior  $\Lambda$ , we recover a full shape representation for each  $\mathbf{D}_i$  ( $i \in [1 : n]$ ). Such a process is meant to be cheap, thus we use the classical pairwise similarity GPA to compute the similarity transformation between datum shapes, and then complete the missing points by averaging their occurrence in other shapes.

For each partial shape  $\mathbf{D}_i$ , we compute the pairwise similarity transformation between  $\mathbf{D}_i$  and  $\mathbf{D}_k$  ( $k \in [1 : n]$ ) by:

$$\hat{s}_{ik}, \hat{\mathbf{R}}_{ik}, \hat{\mathbf{t}}_{ik} = \arg \min_{s_{ik}, \mathbf{R}_{ik}, \mathbf{t}_{ik}} \| (s_{ik} \mathbf{R}_{ik} \mathbf{D}_k + \mathbf{t}_{ik} \mathbf{1}^\top - \mathbf{D}_i) \mathbf{\Gamma}_i \mathbf{\Gamma}_k \|_F^2,$$

where  $\mathbf{R}_{ik}$  is an orthonormal matrix,  $s_{ik}$  a non-negative scalar, and  $\mathbf{t}_{ik}$  a  $d$ -dimensional translation vector. This classical Procrustes problem can be solved in closed-form (see Appendix D).

Subsequently we complete the missing points in  $\mathbf{D}_i$  using their occurrence in other shapes by:

$$\mathfrak{D}_i = \mathbf{D}_i \mathbf{\Gamma}_i + \hat{\mathbf{D}}_i \mathbf{\Gamma}_+^{-1} (\mathbf{I} - \mathbf{\Gamma}_i), \quad (27)$$

to obtain a full shape  $\mathfrak{D}_i$ . Here  $\hat{\mathbf{D}}_i = \sum_{k=1}^n \left( \hat{s}_{ik} \hat{\mathbf{R}}_{ik} \mathbf{D}_k + \hat{\mathbf{t}}_{ik} \mathbf{1}^\top \right) \mathbf{\Gamma}_k$  and  $\mathbf{\Gamma}_+ = \sum_{k=1}^n \mathbf{\Gamma}_k$ . Then we estimate the reference covariance prior  $\Lambda$  by the result in Section 4.3 based on the zero-centered full shapes  $\bar{\mathfrak{D}}_i = \mathfrak{D}_i - \frac{1}{m} \mathfrak{D}_i \mathbf{1} \mathbf{1}^\top$ , which is detailed in Algorithm 1.

## 6.2 Closed-Form Solution

We extend the soft-regularized formulation (26) to partial shape GPA as:

$$\text{III} : \begin{cases} \arg \min_{\{\mathbf{W}_i\}, \mathbf{S}} & \sum_{i=1}^n \|\mathbf{W}_i^\top \mathcal{B}_i(D_i) \Gamma_i - \mathbf{S} \Gamma_i\|_F^2 + \sum_{i=1}^n \mu_i \|\mathbf{Z}_i \mathbf{W}_i\|_F^2 + \nu \|\mathbf{S} \mathbf{1}\|_2^2 \\ \text{s.t.} & \mathbf{S} \mathbf{S}^\top = \mathbf{\Lambda}. \end{cases} \quad (28)$$

This formulation includes formulation (26) as a special case, thus is the ultimate form we will implement.

Formulation (28) can be solved by firstly eliminating the transformation parameters, and then solving an optimization problem in  $\mathbf{S}$  by Lemma 1. The key steps are sketched as follow. We define the shorthand  $\mathcal{B}_i \stackrel{\text{def}}{=} \mathcal{B}_i(D_i)$ . Given  $\mathbf{S}$ , problem (28) becomes linear least-squares in  $\mathbf{W}_i$  whose solution is:

$$\mathbf{W}_i = (\mathcal{B}_i \Gamma_i \mathcal{B}_i^\top + \mu_i \mathbf{Z}_i^\top \mathbf{Z}_i)^{-1} \mathcal{B}_i \Gamma_i \mathbf{S}^\top. \quad (29)$$

By substituting equation (29) into problem (28), we obtain:

$$\arg \min_{\mathbf{S}} \text{tr}(\mathbf{S} (\mathcal{P}_{\text{III}} + \nu \mathbf{1} \mathbf{1}^\top) \mathbf{S}^\top) \quad \text{s.t.} \quad \mathbf{S} \mathbf{S}^\top = \mathbf{\Lambda}, \quad (30)$$

where  $\mathcal{P}_{\text{III}}$  is defined as:

$$\mathcal{P}_{\text{III}} = \sum_{i=1}^n \left( \Gamma_i - \Gamma_i \mathcal{B}_i^\top (\mathcal{B}_i \Gamma_i \mathcal{B}_i^\top + \mu_i \mathbf{Z}_i^\top \mathbf{Z}_i)^{-1} \mathcal{B}_i \Gamma_i \right). \quad (31)$$

Problem (30) can be solved in closed-form by Lemma 1. We summarize the above results as:

**Summary 3.** *Formulation (28) can be solved globally. The optimal reference shape  $\mathbf{S}^*$  is obtained as scaling the  $d$  bottom eigenvectors of  $(\mathcal{P}_{\text{III}} + \nu \mathbf{1} \mathbf{1}^\top)$  by  $\sqrt{\mathbf{\Lambda}}$ . The optimal transformation parameters are given by  $\mathbf{W}_i^* = (\mathcal{B}_i \Gamma_i \mathcal{B}_i^\top + \mu_i \mathbf{Z}_i^\top \mathbf{Z}_i)^{-1} \mathcal{B}_i \Gamma_i \mathbf{S}^{*\top}$  ( $i \in [1 : n]$ ).*

## 6.3 Eigenvector Characterization and Tuning Parameters

We rewrite  $\mathcal{P}_{\text{III}}$  as  $\mathcal{P}_{\text{III}} = \sum_{i=1}^n \mathbf{P}_i$ , with:

$$\mathbf{P}_i = \Gamma_i - \Gamma_i \mathcal{B}_i^\top (\mathcal{B}_i \Gamma_i \mathcal{B}_i^\top + \mu_i \mathbf{Z}_i^\top \mathbf{Z}_i)^{-1} \mathcal{B}_i \Gamma_i.$$

Then  $\mathbf{P}_i$  satisfies:  $\mathbf{I} \succeq \Gamma_i \succeq \mathbf{P}_i \succeq \mathbf{O}$ . This can be shown by writing  $\mathbf{P}_i$  as  $\mathbf{P}_i = \Gamma_i \Phi_i \Gamma_i$  with  $\Phi_i = \mathbf{I} - \Gamma_i \mathcal{B}_i^\top (\mathcal{B}_i \Gamma_i \mathcal{B}_i^\top + \mu_i \mathbf{Z}_i^\top \mathbf{Z}_i)^{-1} \mathcal{B}_i \Gamma_i$  where  $\mathbf{I} \succeq \Phi_i \succeq \mathbf{O}$ . As a summation,  $\mathcal{P}_{\text{III}}$  satisfies:  $n\mathbf{I} \succeq \mathcal{P}_{\text{III}} \succeq \mathbf{O}$ .

We extend Theorem 2 from full shapes to partial shapes as follow.

**Theorem 3.** *The following statements are equivalent:*

- (a)  $\mathcal{P}_{\text{III}}\mathbf{1} = \mathbf{0}$ .
- (b)  $P_i\mathbf{1} = \mathbf{0}$ .
- (c) *There exists  $\mathbf{x}$  such that  $\Gamma_i\mathcal{B}_i(D_i)^\top\mathbf{x} = \Gamma_i\mathbf{1}$ ; moreover if  $\mu_i > 0$ ,  $\mathbf{x}$  must satisfy  $Z_i\mathbf{x} = \mathbf{0}$ .*

*Proof.* The proof is analogous to that for Theorem 2. See Appendix C. □

By Definition 1, if the LBW contains free-translations, then the LBW satisfies: “there exists  $\mathbf{x}$  such that  $\mathcal{B}_i(\cdot)^\top\mathbf{x} = \mathbf{1}$ ; moreover if  $\mu_i > 0$ ,  $\mathbf{x}$  must satisfy  $Z_i\mathbf{x} = \mathbf{0}$ ”. Thus by left-multiplying  $\mathcal{B}_i(\cdot)^\top$  with  $\Gamma_i$ , the following statement is also true: “there exists  $\mathbf{x}$  such that  $\Gamma_i\mathcal{B}_i(\cdot)^\top\mathbf{x} = \Gamma_i\mathbf{1}$ ; moreover if  $\mu_i > 0$ ,  $\mathbf{x}$  must satisfy  $Z_i\mathbf{x} = \mathbf{0}$ ”. Therefore the property of free-translations in the LBW is sufficient for case (c) in Theorem 3.

**Proposition 6.** *The statements in Theorem 3 are satisfied if the LBW contains free-translations. The affine transformation and the TPS satisfy the statements in Theorem 3.*

If the statements in Theorem 3 are satisfied, then  $\mathbf{1}$  is an eigenvector of  $(\mathcal{P}_{\text{III}} + \nu\mathbf{1}\mathbf{1}^\top)$  with  $(\mathcal{P}_{\text{III}} + \nu\mathbf{1}\mathbf{1}^\top)\mathbf{1} = m\nu\mathbf{1}$ . Thus if we choose  $\nu \geq n/m$ , the eigenvector  $\mathbf{1}$  is excluded from the solution.

**Proposition 7.** *The tuning parameter  $\nu$  can be safely set to any  $\nu \geq n/m$ .*

The rows of  $\mathbf{S}^*$  are (un-normalized) eigenvectors of  $(\mathcal{P}_{\text{III}} + \nu\mathbf{1}\mathbf{1}^\top)$  obtained by excluding the eigenvector  $\mathbf{1}$ . By the orthogonality of eigenvectors, we know  $\mathbf{S}^*\mathbf{1} = \mathbf{0}$ . Thus we conclude:

**Proposition 8.** *If the statements in Theorem 3 are satisfied and  $\nu \geq n/m$ , the optimal reference shape  $\mathbf{S}^*$  is zero-centered.*

Thus if the statements in Theorem 3 are satisfied, solving formulation (28) with the soft constraint is equivalent to solving the one with the hard constraint:

$$\begin{aligned} \arg \min_{\{\mathbf{W}_i\}, \mathbf{S}} \sum_{i=1}^n \|\mathbf{W}_i^\top \mathcal{B}_i(D_i)\Gamma_i - \mathbf{S}\Gamma_i\|_F^2 + \sum_{i=1}^n \mu_i \|\mathbf{Z}_i\mathbf{W}_i\|_F^2 \\ \text{s.t. } \mathbf{S}\mathbf{S}^\top = \mathbf{A}, \quad \mathbf{S}\mathbf{1} = \mathbf{0}. \end{aligned} \tag{32}$$

However, we would always recommend users to solve formulation (28) since it is always solvable in closed-form, and generalizes to LBWs where the statements in Theorem 3 are not satisfied.

## 6.4 The Coordinate Transformation of Datum Shapes

We apply (possibly distinct) rigid transformations  $(\mathbf{R}_i, \mathbf{t}_i)$  to each datum shape  $\mathbf{D}_i$ , and denote the transformed datum shapes as  $\mathbf{D}'_i = \mathbf{R}_i \mathbf{D}_i + \mathbf{t}_i$  ( $i \in [1 : n]$ ). We denote  $\mathbf{Z}'_i$  the new regularization matrix in replacement of  $\mathbf{Z}_i$  under transformed datum shapes.

**Lemma 6.** *In formulation (28), if there exists an invertible matrix  $\mathbf{H}_i$  such that  $\mathcal{B}_i(\mathbf{D}'_i) = \mathbf{H}_i \mathcal{B}_i(\mathbf{D}_i)$  and  $\mathbf{Z}'_i = \mathbf{Z}_i \mathbf{H}_i^\top$  for each  $i \in [1 : n]$ , the matrix  $\mathcal{P}_{\text{III}}$  remains the same. So does the optimal reference shape  $\mathbf{S}$ .*

*Proof.* The proof is obvious as  $\mathcal{P}_{\text{III}}$  is invariant to such transformations.  $\square$

**Proposition 9.** *For the TPS warp, if we apply the same rigid transformation  $(\mathbf{R}_i, \mathbf{t}_i)$  to both the datum shape  $\mathbf{D}_i$  and the control points in  $\mathbf{D}_i$ , then  $\mathcal{B}_i(\mathbf{D}'_i) = \mathcal{B}_i(\mathbf{D}_i)$  and  $\mathbf{Z}'_i = \mathbf{Z}_i$ .*

*Proof.* See Appendix A.5.  $\square$

Together with the discussion on the affine case in Section 4.4, we have the following conclusion:

**Proposition 10.** *In formulation (28), if the LBW is chosen as the affine transformation or the TSP warp, then the optimal reference shape  $\mathbf{S}$  remains the same when we apply rigid coordinate transformations to the datum shapes ahead.*

The above result indicates that we can parameterize the datum shapes in any coordinate frame, while the solution of formulation (28) will give exactly the same optimal reference shape. The optimal transformations of formulation (28) will automatically accommodate the coordinate transformations.

## 6.5 Reflection

The reflection in the computed reference shape  $\mathbf{S}^*$  can be easily coped with by simply flipping the sign of one row in  $\mathbf{S}^*$ . Let  $\mathbf{S}^* = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_d]^\top$ . It is easy to verify that  $\mathbf{S}^*$  is still globally optimal if we flip the signs of any  $\mathbf{s}_k$  ( $k \in [1 : d]$ ). Assume there are no reflections between the datum shapes. The reflection in  $\mathbf{S}^*$  can be detected by computing an orthogonal Procrustes between  $\mathbf{S}^*$  and any one of  $\mathbf{D}_i$  ( $i \in [1 : n]$ ) by:

$$\hat{\mathbf{R}}, \hat{\mathbf{t}} = \arg \min_{\mathbf{R} \in \text{O}(d), \mathbf{t}} \|(\mathbf{R} \mathbf{D}_i + \mathbf{t} \mathbf{1}^\top - \mathbf{S}^*) \mathbf{\Gamma}_i\|_F^2.$$

The optimal  $\mathbf{t}$  is  $\hat{\mathbf{t}} = -\frac{1}{\text{nnz}(\mathbf{\Gamma}_i)} (\hat{\mathbf{R}} \mathbf{D}_i - \mathbf{S}^*) \mathbf{\Gamma}_i \mathbf{1}$ . Denote  $\mathbf{K}_i = \mathbf{\Gamma}_i - \frac{1}{\text{nnz}(\mathbf{\Gamma}_i)} \mathbf{\Gamma}_i \mathbf{1} \mathbf{1}^\top \mathbf{\Gamma}_i$ . Denote  $\mathbf{E} = \mathbf{D}_i \mathbf{K}_i \mathbf{K}_i^\top \mathbf{S}^{*\top}$  and its SVD as  $\mathbf{E} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top$ . Then the optimal  $\mathbf{R}$  is  $\hat{\mathbf{R}} = \mathbf{V} \mathbf{U}^\top$ .

If  $\det(\hat{\mathbf{R}}) = 1$ , there is no reflection. If  $\det(\hat{\mathbf{R}}) = -1$ , we let  $\mathbf{S}^* = [-\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_d]^\top$ . The correctness of such an approach is shown as follow. The determinants satisfy  $\det(\mathbf{E}) = \det(\mathbf{U}) \det(\mathbf{\Sigma}) \det(\mathbf{V}^\top)$  and

$\det(\hat{\mathbf{R}}) = \det(\mathbf{U}) \det(\mathbf{V}^\top)$ . Because  $\det(\mathbf{\Sigma}) > 0$ , we know that  $\det(\mathbf{E})$  and  $\det(\hat{\mathbf{R}})$  have the same sign. By flipping the sign of one row in  $\mathbf{S}^*$ , we flip the sign of one column in  $\mathbf{E}$ , which causes the flip of the sign of  $\det(\mathbf{E})$ , thus  $\det(\hat{\mathbf{R}})$  as well.

## 6.6 Pseudo-Code

Our algorithm is rather easy to implement. We term the proposed deformable GPA framework as DefGPA. The overall procedure is given as pseudo-code in Algorithm 2. We release our Matlab implementation of DefGPA to foster future research in this direction.

**Code repository:** <https://bitbucket.org/clermontferrand/deformableprocrustes/src/master/>

---

### Algorithm 2: DefGPA – GPA with LBWs

---

```

Data:  $\{D_i \in \mathbb{R}^{d \times m}, \Gamma_i \in \mathbb{R}^{m \times m}\}, i \in [1:n]$  /* Datum shapes and visibility indicators */
Output:  $\mathbf{S}^*$  and  $\{\mathbf{W}_i^*\}, i \in [1:n]$  /* Reference shape and transformation parameters */

1 begin
2   for  $i \in [1:n]$  do /* Users' choice */
3      $\mathcal{B}_i(\cdot), \mathbf{Z}_i \leftarrow \text{InitializeWarpModel}$  /* Choose and initialize LBWs */
4      $\mu_i \leftarrow \text{ChooseSmoothingParameter}$  /* Choose LBWs' smoothing parameters */
5      $\mathcal{B}_i \leftarrow \mathcal{B}_i(D_i)$  /* Lift to feature space */
6   end
7    $\nu \leftarrow m/n$  /* By Proposition 7 */
8   for  $i \in [1:n]$  do /* Shape completion from pairwise Procrustes */
9     if  $\text{nnz}(\Gamma_i) = n$  then /* A full shape */
10       $\mathcal{D}_i \leftarrow D_i$ 
11    else /* A partial shape */
12      for  $k \in [1:n]$  do
13         $\hat{s}_{ik}, \hat{\mathbf{R}}_{ik}, \hat{t}_{ik} \leftarrow \text{PairwiseSimilarityProcrustes}(D_i, D_k)$ 
14      end
15       $\mathcal{D}_i \leftarrow \text{Equation (27)}$  /* Predict missing points and complete shape */
16    end
17  end
18   $\mathbf{\Lambda} \leftarrow \text{EstimateReferenceCovariancePrior}(\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_n)$  /* Algorithm 1 */
19   $\mathcal{P} \leftarrow \text{Equation (31)}$ 
20   $\mathbf{X}^* \leftarrow d\text{-BottomEigenvectors}(\mathcal{P} + \nu \mathbf{1}\mathbf{1}^\top)$  /* Solve for reference shape */
21   $\mathbf{S}^* \leftarrow \sqrt{\mathbf{\Lambda}} \mathbf{X}^{*\top}$ 
22  for  $i \in [1:n]$  do /* Solve for individual transformations */
23     $\mathbf{W}_i^* \leftarrow (\mathcal{B}_i \Gamma_i \mathcal{B}_i^\top + \mu_i \mathbf{Z}_i^\top \mathbf{Z}_i)^{-1} \mathcal{B}_i \Gamma_i \mathbf{S}^{*\top}$ 
24  end
25 end

```

---

Dataset	Dim.	F/P	Landmarks	Shapes	Description
Face	2D	P	43 - 68	10	different facial expressions and camera perspectives
Bag	2D	F	155	8	deforming handbag
Pillow	2D	F	69	10	deforming pillow cover
LiTS	3D	F	54	8	CT scans with fiducial landmarks
Liver	3D	F	4004	10	simulated deformations of a human 3D liver model
ToyRug	3D	F	30	200	2D features and 3D points from a stereo rig

Table 1: List of datasets used for evaluation.

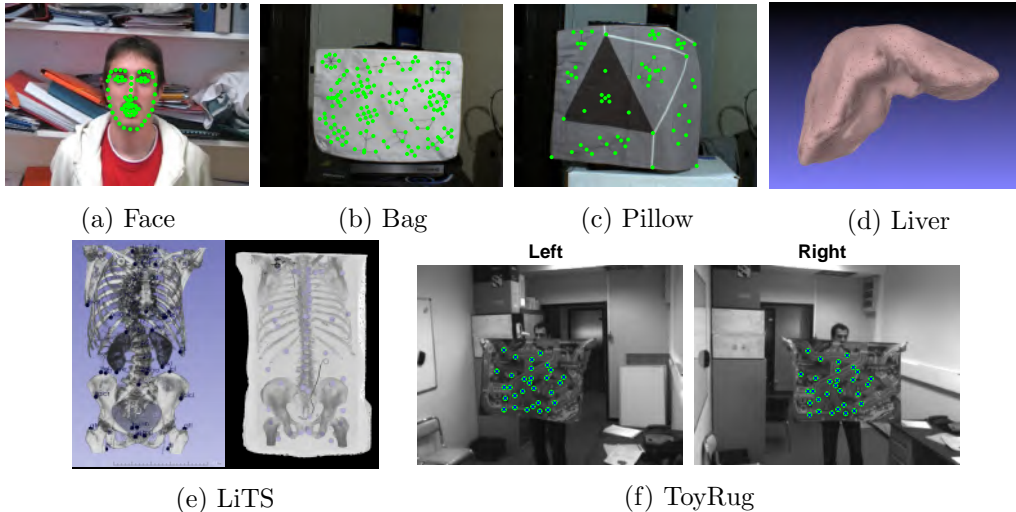


Figure 1: Sample image of each dataset.

## 7 Experimental Results

We provide experimental results with respect to various deformable scenarios. While our method adapts to general LBWs, we use the affine transformation and the TPS warp to show the results. A brief introduction of the TPS warp is provided in Appendix A.

### 7.1 Experimental Setups

The datasets used for evaluation are listed in Table 1, while samples are given in Figure 1. In Table 1, “F” stands for full shapes (without missing datum points), and “P” for partial shapes (with missing datum points). These are public datasets coming from different papers and designed for different problems. These datasets cover the case of structural deformations like facial expressions (Bartoli et al., 2013), deformable objects (Bartoli, 2006; Gallardo et al., 2017), and tissue deformations (Bilic et al., 2019). Both 2D and 3D cases are considered. The Liver dataset is a mesh with 4004 corresponding vertices, while the others are 2D/3D images with corresponding landmarks.

We use Euclidean GPA and Affine GPA with costs defined in the datum-space as benchmark algo-

rithms. These two methods are denoted as \*EUC\_d and \*AFF\_d, where \_d means the methods minimize the datum-space cost. The notion \* indicates they are benchmark methods. For \*EUC\_d and \*AFF\_d, we use the MATLAB implementation by (Bartoli et al., 2013) based on closed-form initialization and Levenberg–Marquardt refinement.

Our methods are denoted by AFF\_r, TPS\_r(3), TPS\_r(5), and TPS\_r(7) respectively, where \_r means the methods minimize the reference-space cost. AFF\_r stands for GPA with the affine model, and TPS\_r(·) stands for GPA with the TPS warp. We choose the control points of the TPS warp evenly along each principal axis of the datum shape. We examine the cases of 3, 5 and 7 points along each principal axis, which results in 9, 25, 49 overall control points for 2D datasets and 27, 125, 343 overall control points for the LiTS and Liver datasets. The ToyRug dataset is almost flat, thus we assign two layers of control points along the first two principal axes which yields 18, 50, 98 overall control points.

Our methods are implemented in MATLAB, which constitutes a fair comparison against \*EUC\_d and \*AFF\_d. The experiments are carried out by an Intel(R) Core(TM) i7-6700K CPU @ 4.00GHz × 8 CPU, running Ubuntu 18.04.5 LTS. The MATLAB version is R2020b.

## 7.2 Evaluation Metrics

### 7.2.1 Landmark Residual

We use RMSE\_r to denote the landmark residual defined in the reference-space, and RMSE\_d the landmark residual in the datum-space. These two metrics are defined as:

$$\text{RMSE}_r = \sqrt{\frac{1}{\kappa} \sum_{i=1}^n \|(\mathcal{T}_i(\mathbf{D}_i) - \mathbf{S}^*) \mathbf{\Gamma}_i\|_F^2}$$

$$\text{RMSE}_d = \sqrt{\frac{1}{\kappa} \sum_{i=1}^n \|(\mathbf{D}_i - \mathcal{T}_i^{-1}(\mathbf{S}^*)) \mathbf{\Gamma}_i\|_F^2},$$

where  $\kappa = \sum_{i=1}^n \mathbf{nnz}(\mathbf{\Gamma}_i)$ . If the transformation model is invertible, it is easy to derive RMSE\_r and RMSE\_d from one another. However, this is typically not the case for LBWs, *e.g.* the TPS warp is not invertible. We propose to use control points and their images as samples to fit an inverse TPS warp. In specific, let  $\mathcal{T}(\cdot)$  be a TPS warp, and  $\mathbf{c}_i$  ( $i \in [1 : l]$ ) be its  $l$  control points. Let the images of these  $l$  control points under  $\mathcal{T}(\cdot)$  be  $\mathbf{c}'_i = \mathcal{T}(\mathbf{c}_i)$  ( $i \in [1 : l]$ ). Then in essence  $\mathcal{T}(\cdot)$  is a regression model obtained by fitting the datum pairs  $(\mathbf{c}_i, \mathbf{c}'_i)$  ( $i \in [1 : l]$ ) with  $\mathbf{c}_i$  being the input and  $\mathbf{c}'_i$  being the output. Therefore the inverse of  $\mathcal{T}(\cdot)$ , denoted by  $\mathcal{T}^{-1}(\cdot)$ , can be defined by fitting the pairs  $(\mathbf{c}'_i, \mathbf{c}_i)$  ( $i \in [1 : l]$ ), with  $\mathbf{c}'_i$  being the input and  $\mathbf{c}_i$  being the output. For the TPS warp, this is realized by letting  $\mathbf{c}'_i$  be the set of control points of  $\mathcal{T}^{-1}(\cdot)$ ,



and computing the warp parameters using the relation  $\mathcal{T}^{-1}(\mathbf{c}'_i) = \mathbf{c}_i$ .

### 7.2.2 Cross-Validation Error

The warp models can overfit the data by using a small enough TPS smoothing parameter. We quantify this behavior by the Cross-Validation Error (CVE) defined as:

$$\text{CVE} = \sqrt{\frac{1}{\kappa} \sum_{i=1}^n \left\| \left( \hat{\mathbf{S}}_i - \mathbf{S}^* \right) \mathbf{\Gamma}_i \right\|_F^2},$$

where the *predicted reference shape*  $\hat{\mathbf{S}}_i$  is computed by the  $G$ -fold cross-validation as follow. We group all  $m$  points indexed as  $1, 2, \dots, m$  as  $G$  mutually exclusive subsets as:

$$\underbrace{\mathbf{p}_1, \dots, \mathbf{p}_N}_{g_1}, \underbrace{\mathbf{p}_{N+1}, \dots, \mathbf{p}_{2N}}_{g_2}, \dots, \underbrace{\mathbf{p}_{(G-1)N+1}, \dots, \mathbf{p}_m}_{g_G}.$$

Each subset has  $N$  points except the subset  $g_G$  which contains the remaining left. We index the points  $g_k$  in shape  $\mathbf{D}_i$  as  $\mathbf{D}_i(g_k)$ , and the points in shape  $\hat{\mathbf{S}}_i$  as  $\hat{\mathbf{S}}_i(g_k)$ . Now for each subset  $g_k$  ( $k = [1 : G]$ ), we solve the GPA with all the points except those in  $g_k$ . This solution is denoted as  $\mathbf{S}_{g_k}^*$  for the reference shape and  $\mathcal{T}_{g_k:i}$  ( $i \in [1 : n]$ ) for the transformations. Then for each datum shape, we predict the positions of points in  $\mathbf{D}_i(g_k)$  by the transformation  $\mathcal{T}_{g_k:i}(\mathbf{D}_i(g_k))$ . We correct the gauge of  $\mathcal{T}_{g_k:i}(\mathbf{D}_i(g_k))$  with the Euclidean Procrustes between  $\mathbf{S}_{g_k}^*$  and  $\mathbf{S}^*$  (by eliminating points  $g_k$  in  $\mathbf{S}^*$ ). Denote such a solution to be  $\mathbf{R}_{g_k}$  and  $\mathbf{t}_{g_k}$ . Finally we set  $\hat{\mathbf{S}}_i(g_k) = \mathbf{R}_{g_k} \mathcal{T}_{g_k:i}(\mathbf{D}_i(g_k)) + \mathbf{t}_{g_k} \mathbf{1}\mathbf{\Gamma}$ . Repeating the above procedure for all  $G$  subsets, we obtain the predicted reference shape  $\hat{\mathbf{S}}_i$ .

The CVE resembles the RMSE, except that it handles overfitting. For the Face, Bag, Pillow, LiTS and ToyRug dataset, we set  $N = 1$ . For the Liver dataset, we set  $N = 40$  to cope with the larger dataset size and dimension.

## 7.3 Thin-Plate Spline Smoothing Parameters

We set the TPS smoothing parameter  $\mu_i$  as  $\mu_i = \mathbf{nnz}(\mathbf{\Gamma}_i)\theta$ , ( $i \in [1 : n]$ ), and adjust  $\theta$  within the range  $[1e - 5, 1e + 5]$ . We report the RMSE\_r, RMSE\_d, and CVE with respect to different  $\theta$  in Figure 2. The  $\theta$  used to generate our results are marked as vertical lines, annotated with the chosen value.

In Figure 2, the RMSE\_r decreases monotonically as  $\theta$  decreases, as a smaller  $\theta$  implies more flexibility. However, a sufficiently small  $\theta$  may cause overfitting. The overfitting in DefGPA is mainly twofold: the bias to the data, and the bias to the optimization criterion (*i.e.*, the cost function). When overfitting happens,

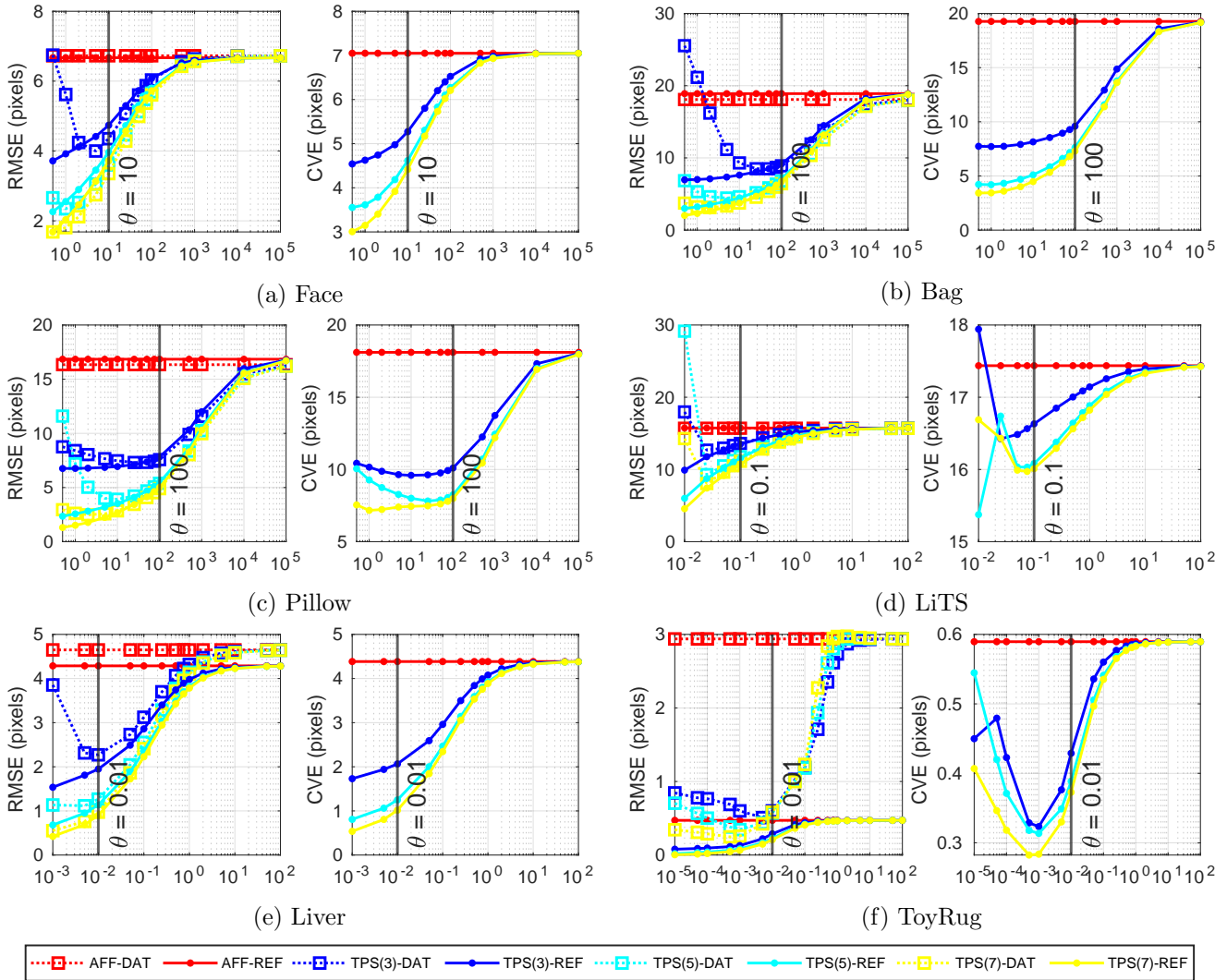


Figure 2: The choice of TPS smoothing parameters. The Root-Mean-Squared-Error (RMSE), Rigidity-Score (RS), and Cross-Validation-Error (CVE) are plotted with respect to different  $\theta$  values (in horizontal axes).

the optimization tends to fit the training data to the cost function as much as possible, thus actually creating misfitting to new data and other criteria. This suggests that we can use the fitness to new data and other criteria to detect overfitting. The idea of using new data is realized by cross-validation given as the CVE metric, and that of using other criteria by the datum-space cost (denoted by  $RMSE_d$ ). In Figure 2, the overfitting is reflected as the divergence of the  $RMSE_r$  and  $RMSE_d$  metric, as well as the slope change of the CVE curve. In general, CVE has a positive slope if  $RMSE_r$  and  $RMSE_d$  agree with one another, and a negative slope if  $RMSE_r$  and  $RMSE_d$  diverge.

For 2D datasets, we choose  $\theta = 10$  for the Face, and  $\theta = 100$  for the Bag and Pillow datasets. For 3D datasets, we choose  $\theta = 0.1$  for the LiTS, and  $\theta = 0.01$  for the Liver and ToyRug datasets. Although the bending energy term has different constants for 2D and 3D cases, the choice of  $\theta$  is roughly stable within each category.

In terms of overfitting, we focus on the agreement of the trend of the RMSE<sub>r</sub> and RMSE<sub>d</sub> curves: whether they both increase or decrease. There is always a gap between the RMSE<sub>r</sub> and RMSE<sub>d</sub> metrics, which stands for the asymmetry between different cost functions. We will examine this asymmetry in Section 7.6.

## 7.4 Results Based on Landmarks

### 7.4.1 Statistics

We report in Table 2 the RMSE<sub>d</sub>, RMSE<sub>r</sub>, CVE and the computational time per GPA problem. The RMSE<sub>d</sub>, RMSE<sub>r</sub> and CVE are evaluated in pixels, and the time in seconds.

Dataset		*EUC <sub>d</sub>	*AFF <sub>d</sub>	AFF <sub>r</sub>	TPS <sub>r</sub> (3)	TPS <sub>r</sub> (5)	TPS <sub>r</sub> (7)
Face	RMSE <sub>d</sub>	7.11	6.52	6.73	4.35	3.60	3.35
	RMSE <sub>r</sub>	7.11	6.64	6.67	4.74	3.96	3.72
	CVE	7.26	7.07	7.05	5.27	4.62	4.42
	Time	0.6326	0.2113	0.0731	0.0759	0.0149	0.0191
Bag	RMSE <sub>d</sub>	30.23	17.73	18.08	8.91	6.80	6.42
	RMSE <sub>r</sub>	30.23	19.78	18.90	9.16	7.12	6.75
	CVE	30.48	20.19	19.27	9.59	7.62	7.30
	Time	0.3785	0.0535	0.0328	0.0303	0.0218	0.0275
Pillow	RMSE <sub>d</sub>	23.44	16.23	16.34	7.58	5.39	4.89
	RMSE <sub>r</sub>	23.44	17.96	16.84	7.82	5.47	5.02
	CVE	24.01	19.26	18.11	10.10	8.25	7.99
	Time	0.1862	0.0339	0.0253	0.0211	0.0140	0.0176
LiTS	RMSE <sub>d</sub>	19.85	15.43	15.71	13.60	11.76	11.13
	RMSE <sub>r</sub>	19.85	15.84	15.74	13.50	11.57	10.77
	CVE	20.50	17.19	17.44	16.63	16.08	16.01
	Time	0.2144	0.0267	0.0245	0.0234	0.0387	0.1284
Liver	RMSE <sub>d</sub>	4.94	4.60	4.65	2.27	1.27	0.97
	RMSE <sub>r</sub>	4.94	4.68	4.29	1.95	1.14	0.90
	CVE	5.01	4.78	4.38	2.06	1.25	1.02
	Time	249.3995	41.8242	9.7071	10.1804	11.0086	13.3759
ToyRug	RMSE <sub>d</sub>	0.88	0.42	2.94	0.60	0.57	0.58
	RMSE <sub>r</sub>	0.88	2.41	0.47	0.29	0.23	0.21
	CVE	0.98	12.01	0.59	0.43	0.39	0.37
	Time	2.2158	0.2969	0.0792	0.1393	0.2122	0.4351

Table 2: The statistics of landmark registration.

### 7.4.2 Visualization

For each case, we visualize the optimal reference shape  $\mathbf{S}^*$ , along with the set of predicted reference shapes  $\hat{\mathbf{S}}_i$  computed by the leave- $N$ -out cross-validation. The details of how to compute each  $\hat{\mathbf{S}}_i$  have been given

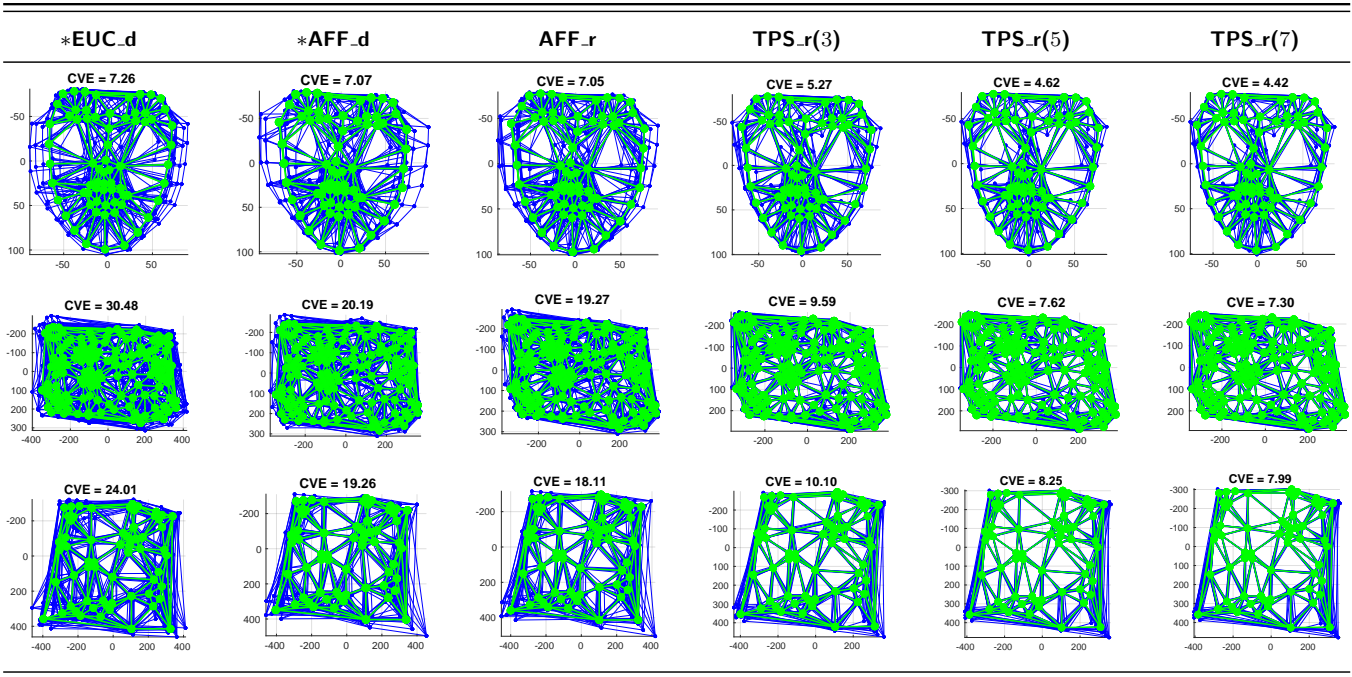


Figure 3: The fitness of each methods visualized by the cross-validation on 2D datasets. The reference shape  $\mathbf{S}^*$  is plotted in green and the predicted reference shapes  $\hat{\mathbf{S}}_i$  in blue.

in Section 7.2. Recall that  $\hat{\mathbf{S}}_i$  is sensitive to overfitting, thus is better to benchmark fitness than using the transformed datum shapes  $\mathcal{T}_i(\mathbf{D}_i)$  directly.

For the 2D datasets, we visualize  $\hat{\mathbf{S}}_i$  and  $\mathbf{S}^*$  directly in Figure 3. The predicted reference shapes  $\hat{\mathbf{S}}_i$  are plotted in blue, and the optimal reference shape  $\mathbf{S}^*$  on top in green. For the 3D datasets, we visualize the points in  $\mathbf{S}^*$ , and encode the CVE with the marker size and color in Fig. 4. The bigger the marker and the redder the color, the larger the associated CVE value.

The result in Figure 3 shows that the landmark fitness is consistently improved by using the affine-GPA (\*AFF\_d and AFF\_r) or the TPS warp based GPA compared with the Euclidean-GPA (\*EUC\_d). Both \*AFF\_d and AFF\_r give similar results. There is a clear improvement of TPS\_r(3) over the affine-GPA. However, the improvement of TPS\_r(7) over TPS\_r(5) is marginal.

The result in Fig. 4 agrees with the result in Figure 3, which improves sequentially by using the \*EUC\_d, AFF\_r, TPS\_r(3), TPS\_r(5) and TPS\_r(7) methods. Both the datum-space method \*AFF\_d and the reference-space method AFF\_r produce similar results on the LiTS and Liver datasets. The \*AFF\_d method gives rather poor fitness on the ToyRug dataset. This is due to the asymmetry in the cost function which be examined in detail in Section 7.6.

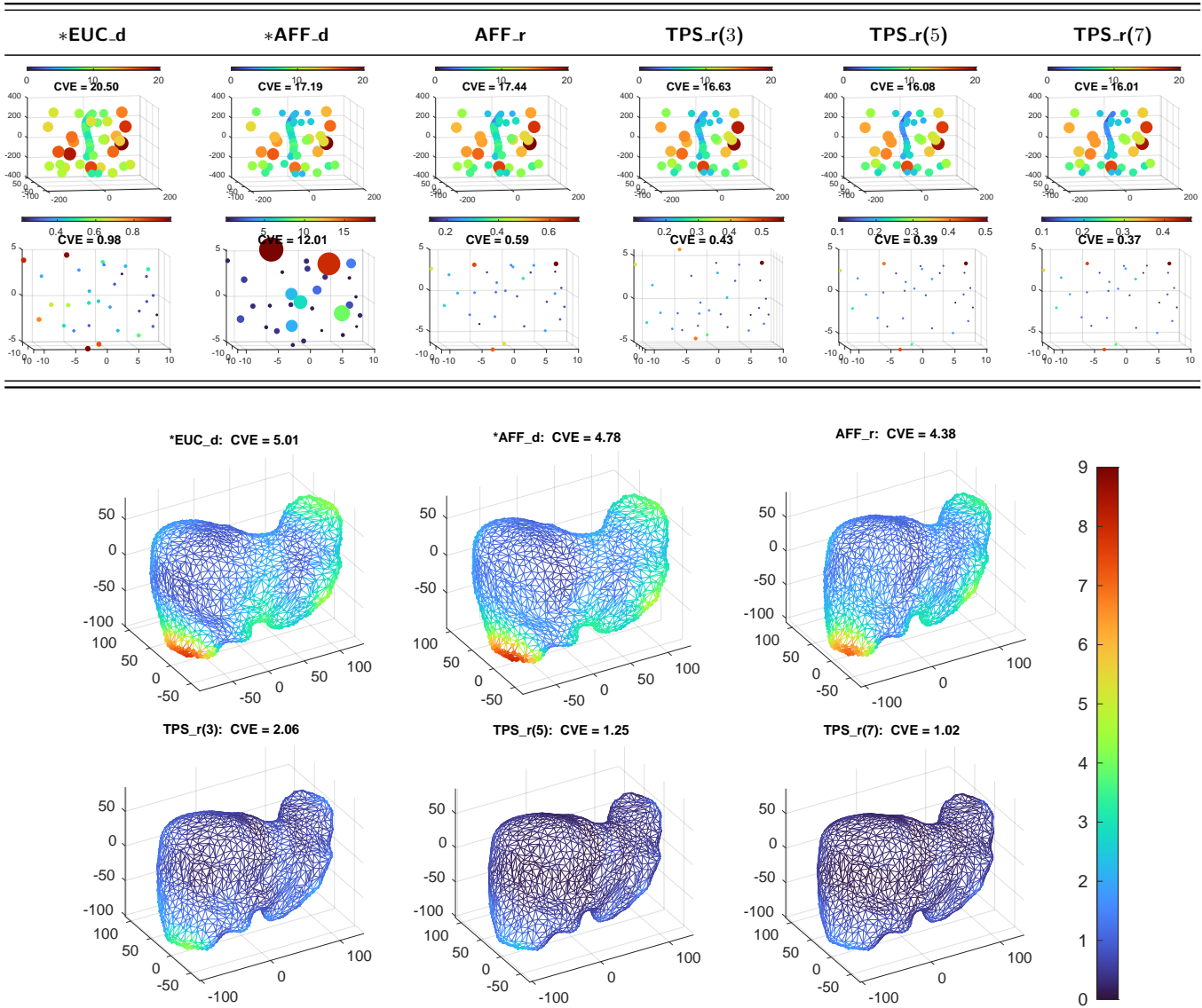


Figure 4: The fitness of each methods visualized by the cross-validation on 3D datasets. We present the reference shape  $\mathcal{S}^*$  and encode the cross-validation error for each point with both the marker sizes and colors. The smaller the marker, the lower the error.

## 7.5 Results Based on Image Intensities

For the 2D datasets, we visualize the image intensities under the computed warps. The image pixel coordinates can be regarded as testing points (in addition to the landmarks) which are immune to the issue of overfitting. We thus evaluate the fitness using the intensities in the transformed pixel coordinates.

Let  $I_i$  be a sample image. For each pixel coordinates  $[u, v]$ , we apply the transformation  $\mathcal{T}_i$  to obtain its target position  $[u', v'] = \mathcal{T}_i([u, v])$ . The *transformed image* (or warped image)  $I'_i$  is defined as  $I'_i(u', v') = I_i(u, v)$ . The warped image  $I'_i$  resides in the coordinate frame of the reference shape, thus can be compared with other warped images. Having computed  $n$  warped sample images  $I'_i$  ( $i \in [1 : n]$ ), we define the *image intensity deviation* as an image  $\Delta I'(u, v) = \mathbf{std} \{I'_1(u, v), I'_2(u, v), \dots, I'_n(u, v)\}$ . In the case of RGB images,

we compute the intensity deviation for all the three channels to obtain an RGB intensity deviation image.

We visualize the image intensity deviation for the Bag and Pillow datasets in Figure 5. The result for  $*AFF\_d$  is similar to  $AFF\_r$ , thus is not shown. In Figure 5, from left to right, the methods  $*EUC\_d$ ,  $AFF\_r$ ,  $TPS\_r(3)$ ,  $TPS\_r(5)$  and  $TPS\_r(7)$  consistently increase the fitness due to the increasing modeling capabilities (from rigid to affine and deformable models), shown as the shrinking of bright regions.

The image intensity deviation of the Face dataset is noisy due to perspective changes, so we visualize a typical warped sample image (with missing correspondences and perspective changes) and the corresponding transformed landmarks in Figure 6. This result helps intuitively understand how each transformation model works: the rigid model preserves the distance; the affine model shears the image; the deformation models can deform the image nonlinearly (*e.g.*, the area around the nose and mouth in Figure 6). The improved fitness using deformation models is obvious without saying.

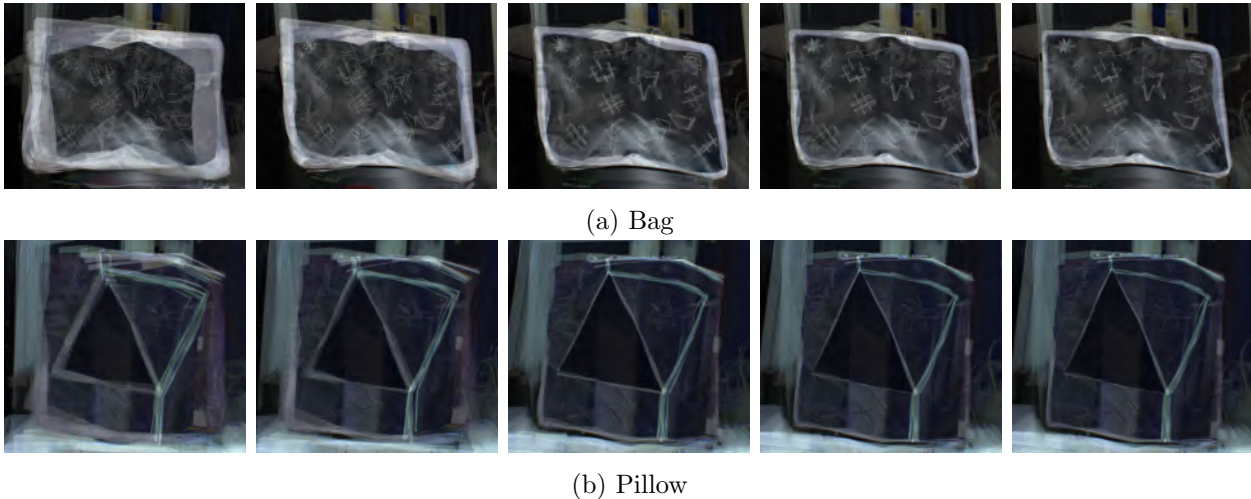


Figure 5: The intensity deviation of the transformed datum images. From left to right are respectively the intensity deviation for  $*EUC\_d$ ,  $AFF\_r$ ,  $TPS\_r(3)$ ,  $TPS\_r(5)$  and  $TPS\_r(7)$ . The bright regions correspond to large intensity deviations, and the dark regions to small intensity deviations.

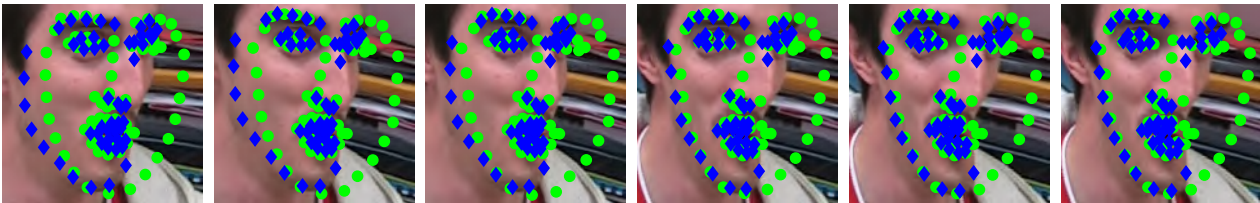


Figure 6: The transformed images using the computed warps. From left to right are respectively the warped images by  $*EUC\_d$ ,  $*AFF\_d$ ,  $AFF\_r$ ,  $TPS\_r(3)$ ,  $TPS\_r(5)$  and  $TPS\_r(7)$ . The transformed landmarks are marked in blue, and the landmarks in the reference shape are marked in green.

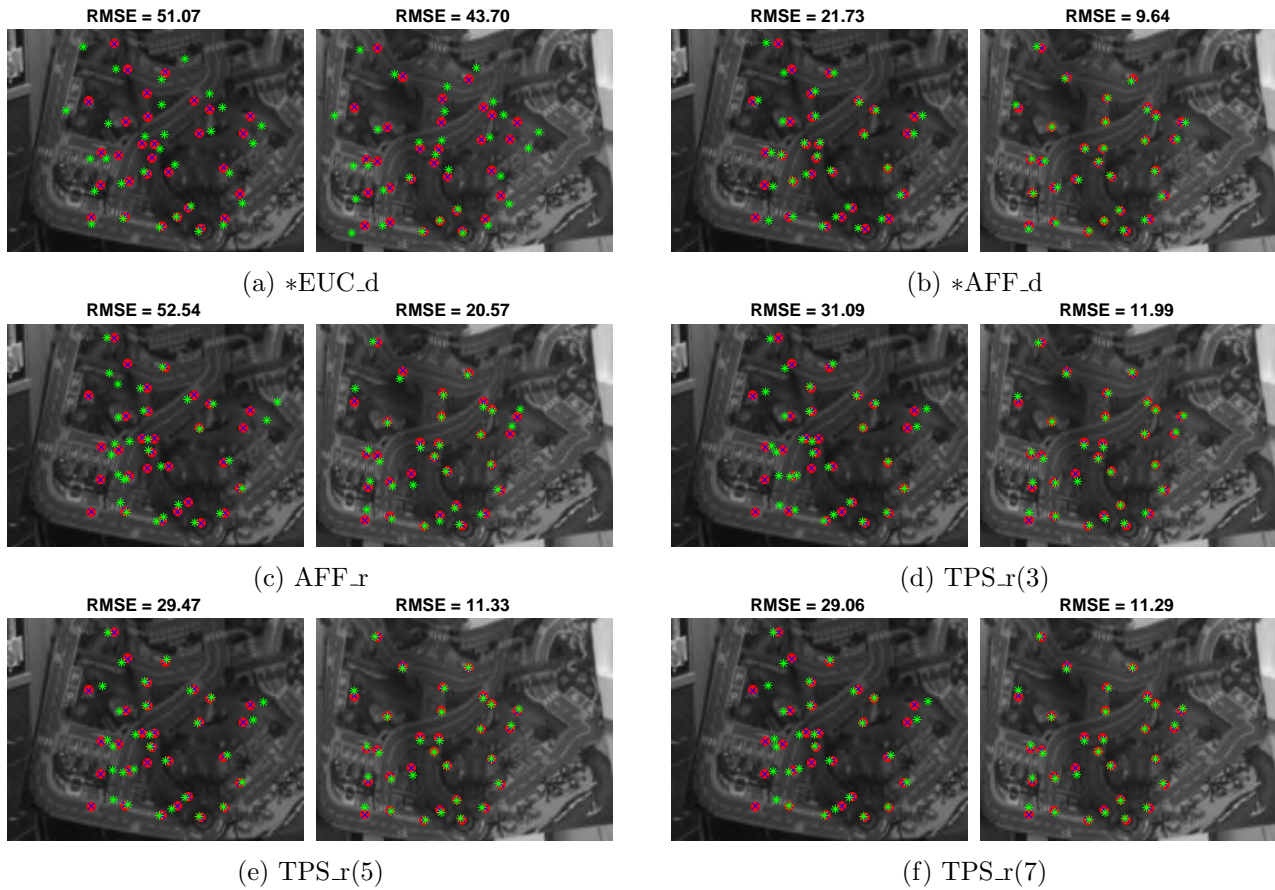


Figure 7: The discrepancy between the generated image features and original image features on the ToyRug dataset. The generated image features are plotted in the green color, and the original image features in the red color. The blue x-markers are reprojected 2D image features from the triangulated 3D datum shapes. The RMSE statistics are measured in pixels.

## 7.6 The Asymmetry in Cost Functions

For the ToyRug dataset, the  $RMSE_r$  and  $RMSE_d$  statistics in Table 2 have shown a strong asymmetry between the reference-space cost and datum-space cost, in particular for the  $*AFF_d$  and  $AFF_r$  methods. This asymmetry is further evidenced in Figure 4 as the  $*AFF_d$  method yields one magnitude larger CVE error than the  $AFF_r$  method. Recall that the datum-space cost aims to find the generative model that best explains all the data; in contrast, the reference-space cost aims to find the reference shape that produces the best fitness to the data. The above results suggest that while being closely related, minimizing the datum-space cost does not necessarily result in a good fitness to the reference shape. Now we report on the other hand that minimizing the reference-space does not necessarily produce a good generative model.

The 3D datum shapes of the ToyRug dataset are obtained by triangulating the image features of a stereo camera rig with known intrinsic and extrinsic parameters. To evaluate GPA methods as generative models, we use the 3D reference shape estimate and transformation estimates to generate 3D datum shapes, and reproject the generated 3D datum shapes to 2D image features. We report the discrepancies between the

generated image features and the original image features in Figure 7, by visualizing the result of the datum shape that produces the largest error.

Contrary to the worst reference-space fitness, the \*AFF\_d method generates the best image features that closely match the original ones as shown in Figure 7, because it minimizes the datum-space cost directly while being more flexible than the \*EUC\_d method. Due to nonrigid deformations, the datum-space method \*EUC\_d does not provide good result. In parallel, because of the metric asymmetry, poor performance is observed for the AFF\_r method which minimizes the reference-space cost. However, although not optimizing the datum-space cost directly, the reference-space methods TPS\_r(3), TPS\_r(5) and TPS\_r(7) are still capable to provide satisfactory results. The quality of the generated features are consistently improved over the TPS\_r(3), TPS\_r(5) and TPS\_r(7) methods. In particular, the features in the bottom right corner generated by the method TPS\_r(7) are better than those generated by the \*AFF\_d method.

This brings us to the final conclusion. The datum-space cost and the reference-cost are two distinct but closely related metrics, while a good result based on one cost does not necessarily guarantee a good result on the other cost due to the metric asymmetry. However it does not mean that the cost functions make a big difference for every dataset: as can be seen in Table 2, the RMSE\_r and RMSE\_d metrics agree most of the time. In adverse cases like the ToyRug dataset, the deformation models like the TPS warp can achieve the best result on both sides, which coins the importance of using DefGPA over classical GPA methods.

## 8 Conclusion

To summarize, we have introduced the problem of GPA with deformations. We have proposed a general problem statement applicable to LBWs, subsuming previous statements made for specific models. We have derived a closed-form solution, applicable to the general case of partial shapes and with regularization. We have extensively validated the proposed solution on various datasets with great care taken to select the regularization weights. Our future work will include studying the consistency of estimates under certain statistical models, and extending our theory to the Nonrigid Structure-from-Motion (NRSfM) problem.

## A The Thin-Plate Spline

### A.1 General Mapping Definition

The Thin-Plate Spline (TPS) is a nonlinear mapping driven by a set of chosen control centers. Let  $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_l$ , with  $\mathbf{c}_i \in \mathbb{R}^d$  ( $d = 2, 3$ ), be the  $l$  control centers of TPS. We present the TPS model in 3D; the 2D case can be analogously derived by removing the  $z$  coordinate and letting the corresponding



TPS kernel function be  $\phi(r) = r^2 \log(r^2)$  (Bookstein, 1997). Strictly speaking, the TPS only occurs in 2D, but is part of a general family of functions called the harmonic splines. For simplicity, we call the function TPS independently of the dimension.

We denote a point in 3D by  $\mathbf{p}^\top = [x, y, z]$ . The TPS function  $\tau(\mathbf{p})$  is a  $\mathbb{R}^3 \mapsto \mathbb{R}$  mapping:

$$\tau(\mathbf{p}) = \sum_{k=1}^l w_k \phi(\|\mathbf{p} - \mathbf{c}_k\|) + a_1 x + a_2 y + a_3 z + a_4$$

where  $\phi(\cdot)$  is the TPS kernel function. In 3D, we have  $\phi(r) = -|r|$ .

We collect the parameters in a single vector  $\boldsymbol{\eta}^\top = [\mathbf{w}^\top, \mathbf{a}^\top]$ , with the coefficient in the nonlinear part as  $\mathbf{w}^\top = [w_1, w_2, \dots, w_l]$ , and that in the linear part as  $\mathbf{a}^\top = [a_1, a_2, a_3, a_4]$ . Let  $\tilde{\mathbf{p}}^\top = [x, y, z, 1]$ . Collect the nonlinear components in a vector:

$$\boldsymbol{\phi}_{\mathbf{p}}^\top = [\phi(\|\mathbf{p} - \mathbf{c}_1\|), \phi(\|\mathbf{p} - \mathbf{c}_2\|), \dots, \phi(\|\mathbf{p} - \mathbf{c}_l\|)].$$

Then the TPS function compacts in the vector form as:

$$\tau(\mathbf{p}) = \boldsymbol{\phi}_{\mathbf{p}}^\top \mathbf{w} + \tilde{\mathbf{p}}^\top \mathbf{a}.$$

We define  $\tilde{\mathbf{c}}_i^\top = [\mathbf{c}_i^\top, 1]$  ( $i \in [1 : l]$ ), and  $\tilde{\mathbf{C}} = [\tilde{\mathbf{c}}_1, \tilde{\mathbf{c}}_2, \dots, \tilde{\mathbf{c}}_l]$ . The transformation parameters in the nonlinear part of the TPS model are constrained with respect to the control centers, *i.e.*,  $\tilde{\mathbf{C}}\mathbf{w} = \mathbf{0}$ . To solve the parameters of TPS, the  $l$  control centers  $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_l$ , are mapped to  $l$  scalar outputs  $y_1, y_2, \dots, y_l$ , by  $\tau(\mathbf{c}_k) = y_k$  ( $k \in [1 : l]$ ). We collect the outputs in a vector  $\mathbf{y}^\top = [y_1, y_2, \dots, y_l]$ .

## A.2 Standard Form as a Linear Regression Model

While there are  $l + 4$  parameters in  $\boldsymbol{\eta}$ , we will show in the following that the 4 constraints in  $\tilde{\mathbf{C}}\mathbf{w} = \mathbf{0}$  can be absorbed in a reparameterization of the TPS model, by using  $l$  parameters only.

To proceed, we define a matrix:

$$\mathbf{K}_\lambda = \begin{bmatrix} \lambda & \phi_{12} & \dots & \phi_{1l} \\ \phi_{21} & \lambda & \dots & \phi_{2l} \\ \vdots & \vdots & & \vdots \\ \phi_{l1} & \phi_{l2} & \dots & \lambda \end{bmatrix},$$

where we use the shorthand  $\phi_{ij} = \phi(\|\mathbf{c}_i - \mathbf{c}_j\|)$ . The original diagonal elements of  $\mathbf{K}_\lambda$  are all zeros,

*i.e.*,  $\phi_{11} = \phi_{22} = \dots = \phi_{ll} = 0$ .  $\lambda$  is an adjustable parameter acting as an internal smoothing parameter to improve the conditioning of the matrix.

The  $l$  TPS mappings of control centers  $\tau(\mathbf{c}_k) = y_k$  ( $k \in [1 : l]$ ), and the 4 parameter constraints  $\tilde{\mathbf{C}}\mathbf{w} = \mathbf{0}$  can be collected in the following matrix form:

$$\begin{bmatrix} \mathbf{K}_\lambda & \tilde{\mathbf{C}}^\top \\ \tilde{\mathbf{C}} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix},$$

which admits a closed-form solution:

$$\begin{bmatrix} \mathbf{w} \\ \mathbf{a} \end{bmatrix} = \boldsymbol{\mathcal{E}}_\lambda \mathbf{y},$$

with the matrix  $\boldsymbol{\mathcal{E}}_\lambda \in \mathbb{R}^{(l+4) \times l}$  defined as:

$$\boldsymbol{\mathcal{E}}_\lambda = \begin{bmatrix} \mathbf{K}_\lambda^{-1} - \mathbf{K}_\lambda^{-1} \tilde{\mathbf{C}}^\top (\tilde{\mathbf{C}} \mathbf{K}_\lambda^{-1} \tilde{\mathbf{C}}^\top)^{-1} \tilde{\mathbf{C}} \mathbf{K}_\lambda^{-1} \\ (\tilde{\mathbf{C}} \mathbf{K}_\lambda^{-1} \tilde{\mathbf{C}}^\top)^{-1} \tilde{\mathbf{C}} \mathbf{K}_\lambda^{-1} \end{bmatrix}.$$

Therefore the TPS function writes:

$$\tau(\mathbf{p}) = \begin{bmatrix} \phi_{\mathbf{p}} \\ \tilde{\mathbf{p}} \end{bmatrix}^\top \begin{bmatrix} \mathbf{w} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \phi_{\mathbf{p}} \\ \tilde{\mathbf{p}} \end{bmatrix}^\top \boldsymbol{\mathcal{E}}_\lambda \mathbf{y} = \mathbf{y}^\top \boldsymbol{\mathcal{E}}_\lambda^\top \begin{bmatrix} \phi_{\mathbf{p}} \\ \tilde{\mathbf{p}} \end{bmatrix}.$$

Let  $\boldsymbol{\beta}(\mathbf{p}) = \boldsymbol{\mathcal{E}}_\lambda^\top \begin{bmatrix} \phi_{\mathbf{p}} \\ \tilde{\mathbf{p}} \end{bmatrix}$ . Then  $\tau(\mathbf{p}) = \mathbf{y}^\top \boldsymbol{\beta}(\mathbf{p})$ .

The 3D TPS warp model  $\mathcal{W}(\mathbf{p}) : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is obtained by stacking 3 TPS functions together, one for each dimension:

$$\mathcal{W}(\mathbf{p}) = \begin{bmatrix} \tau_x(\mathbf{p}) \\ \tau_y(\mathbf{p}) \\ \tau_z(\mathbf{p}) \end{bmatrix} = \begin{bmatrix} \mathbf{y}_x^\top \boldsymbol{\beta}(\mathbf{p}) \\ \mathbf{y}_y^\top \boldsymbol{\beta}(\mathbf{p}) \\ \mathbf{y}_z^\top \boldsymbol{\beta}(\mathbf{p}) \end{bmatrix} = [\mathbf{y}_x, \mathbf{y}_y, \mathbf{y}_z]^\top \boldsymbol{\beta}(\mathbf{p}). \quad (33)$$

This is the standard form of the general warp model defined in equation (16), with  $\mathbf{W} = [\mathbf{y}_x, \mathbf{y}_y, \mathbf{y}_z] \in \mathbb{R}^{l \times 3}$  being a set of new transformation parameters that are free of constraints.

### A.3 Regularization by Bending Energy Matrix

Let  $\bar{\boldsymbol{\mathcal{E}}}_\lambda = \mathbf{K}_\lambda^{-1} - \mathbf{K}_\lambda^{-1} \tilde{\mathbf{C}}^\top (\tilde{\mathbf{C}} \mathbf{K}_\lambda^{-1} \tilde{\mathbf{C}}^\top)^{-1} \tilde{\mathbf{C}} \mathbf{K}_\lambda^{-1}$ . The  $l \times l$  matrix  $\bar{\boldsymbol{\mathcal{E}}}_\lambda$  is the bending energy matrix of the TPS warp (Bookstein, 1989), which satisfies  $\bar{\boldsymbol{\mathcal{E}}}_\lambda \mathbf{K}_\lambda \bar{\boldsymbol{\mathcal{E}}}_\lambda = \bar{\boldsymbol{\mathcal{E}}}_\lambda$ . The bending energy matrix  $\bar{\boldsymbol{\mathcal{E}}}_\lambda$  is positive

semidefinite, with rank  $l - 4$ . The bending energy of a single TPS function is proportional to  $\mathbf{w}^\top \mathbf{K}_\lambda \mathbf{w}$ , with  $\mathbf{w} = \bar{\boldsymbol{\mathcal{E}}}_\lambda \mathbf{y}$ , which is given by:

$$\int_{\mathbb{R}^d} \left\| \frac{\partial^2}{\partial \mathbf{p}^2} \tau(\mathbf{p}) \right\|_F^2 d\mathbf{p} \propto \mathbf{w}^\top \mathbf{K}_\lambda \mathbf{w} = \mathbf{y}^\top \bar{\boldsymbol{\mathcal{E}}}_\lambda \mathbf{y}.$$

For the 3D TPS warp model, the overall bending energy is:

$$\begin{aligned} & \int_{\mathbb{R}^d} \left\| \frac{\partial^2}{\partial \mathbf{p}^2} \mathcal{W}(\mathbf{p}, \mathbf{W}) \right\|_F^2 d\mathbf{p} \\ &= \int_{\mathbb{R}^d} \left( \left\| \frac{\partial^2}{\partial \mathbf{p}^2} \tau_x(\mathbf{p}) \right\|_F^2 + \left\| \frac{\partial^2}{\partial \mathbf{p}^2} \tau_y(\mathbf{p}) \right\|_F^2 + \left\| \frac{\partial^2}{\partial \mathbf{p}^2} \tau_z(\mathbf{p}) \right\|_F^2 \right) d\mathbf{p} \\ &= \mathbf{y}_x^\top \bar{\boldsymbol{\mathcal{E}}}_\lambda \mathbf{y}_x + \mathbf{y}_y^\top \bar{\boldsymbol{\mathcal{E}}}_\lambda \mathbf{y}_y + \mathbf{y}_z^\top \bar{\boldsymbol{\mathcal{E}}}_\lambda \mathbf{y}_z \\ &= \text{tr}(\mathbf{W}^\top \bar{\boldsymbol{\mathcal{E}}}_\lambda \mathbf{W}) = \|\sqrt{\bar{\boldsymbol{\mathcal{E}}}_\lambda} \mathbf{W}\|_F^2 \end{aligned}$$

where  $\sqrt{\bar{\boldsymbol{\mathcal{E}}}_\lambda}$  is the matrix square root of  $\bar{\boldsymbol{\mathcal{E}}}_\lambda$ . Then the bending energy of the TPS warp writes into the standard form  $\mathcal{R}(\mathbf{W}) = \|\sqrt{\bar{\boldsymbol{\mathcal{E}}}_\lambda} \mathbf{W}\|_F^2$ , given in equation (18).

#### A.4 The Thin-Plate Spline Satisfies $\mathcal{B}(\cdot)^\top \mathbf{x} = \mathbf{1}$ and $\mathbf{Z}\mathbf{x} = \mathbf{0}$

Let an arbitrary point cloud  $\mathbf{D}$  be  $\mathbf{D} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_m]$ . Then by equation (33) and the point-cloud operation defined in Section 5.1.2, we write:

$$\begin{aligned} \mathcal{B}(\mathbf{D}) &= [\boldsymbol{\beta}(\mathbf{p}_1), \boldsymbol{\beta}(\mathbf{p}_2), \dots, \boldsymbol{\beta}(\mathbf{p}_m)] \\ &= \boldsymbol{\mathcal{E}}_\lambda^\top \begin{bmatrix} \phi_{\mathbf{p}_1} & \phi_{\mathbf{p}_2} & \cdots & \phi_{\mathbf{p}_m} \\ \tilde{\mathbf{p}}_1 & \tilde{\mathbf{p}}_2 & \cdots & \tilde{\mathbf{p}}_m \end{bmatrix} \\ &= \boldsymbol{\mathcal{E}}_\lambda^\top \begin{bmatrix} \mathbf{M}_D \\ \mathbf{1}^\top \end{bmatrix}. \end{aligned}$$

Hence  $\mathcal{B}(\mathbf{D})^\top$  has structure:

$$\mathcal{B}(\mathbf{D})^\top = \begin{bmatrix} \mathbf{M}_D^\top & \mathbf{1} \end{bmatrix} \boldsymbol{\mathcal{E}}_\lambda.$$

**A.4.1**  $\mathcal{B}(\cdot)^\top \mathbf{x} = \mathbf{1}$ 

It suffices to show such an  $\mathbf{x}$  satisfies  $\mathcal{E}_\lambda \mathbf{x} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}$ . Because:

$$\mathcal{E}_\lambda \tilde{\mathbf{C}}^\top = \begin{bmatrix} \mathbf{O} \\ \mathbf{I} \end{bmatrix} \implies \mathcal{E}_\lambda \tilde{\mathbf{C}}^\top \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{O} \\ \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix},$$

such an  $\mathbf{x}$  indeed exists, which is  $\mathbf{x} = \tilde{\mathbf{C}}^\top \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}$ .

**A.4.2**  $\mathbf{Z}\mathbf{x} = \mathbf{0}$ 

We need to show:

$$\mathbf{Z}\mathbf{x} = \mathbf{0} \iff \|\mathbf{Z}\mathbf{x}\|^2 = \mathbf{x}^\top \mathbf{Z}^\top \mathbf{Z}\mathbf{x} = 0.$$

The proof is immediate by noting that  $\mathbf{Z}^\top \mathbf{Z} = \bar{\mathcal{E}}_\lambda$ , and  $\bar{\mathcal{E}}_\lambda \mathbf{x} = \bar{\mathcal{E}}_\lambda \tilde{\mathbf{C}}^\top \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} = \mathbf{O} \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} = \mathbf{0}$ .

**A.5 The Thin-Plate Spline is Invariant to the Coordinate Transformation**

Given a datum shape  $\mathbf{D}$ , we apply the same rigid transformation  $(\mathbf{R}, \mathbf{t})$  to each datum point  $\mathbf{p}_i$ , and denote the transformed datum point as  $\mathbf{p}'_i = \mathbf{R}\mathbf{p}_i + \mathbf{t}$ . Each control point  $\mathbf{c}_i$  is transformed to  $\mathbf{c}'_i = \mathbf{R}\mathbf{c}_i + \mathbf{t}$ . The matrix  $\mathbf{K}_\lambda$  is invariant by replacing  $\mathbf{c}_i$  with  $\mathbf{c}'_i$  as its elements are radial basis functions. By replacing  $\tilde{\mathbf{C}}$  by  $\tilde{\mathbf{C}}' = [\tilde{\mathbf{c}}'_1, \tilde{\mathbf{c}}'_2, \dots, \tilde{\mathbf{c}}'_l]$ , the matrix  $\mathcal{E}_\lambda$  becomes:

$$\mathcal{E}'_\lambda = \begin{bmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{O} & (\mathbf{T}^\top)^{-1} \end{bmatrix} \mathcal{E}_\lambda, \text{ with } \mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix}.$$

$\phi_{\mathbf{p}_i}$  is also invariant by simultaneously replacing  $\mathbf{c}_i$  with  $\mathbf{c}'_i$  and  $\mathbf{p}_i$  with  $\mathbf{p}'_i$ . Notice that  $\tilde{\mathbf{p}}'_i = \mathbf{T}\tilde{\mathbf{p}}_i$ , thus:

$$\beta(\mathbf{p}'_i) = \mathcal{E}'_\lambda{}^\top \begin{bmatrix} \phi_{\mathbf{p}'_i} \\ \tilde{\mathbf{p}}'_i \end{bmatrix} = \mathcal{E}_\lambda{}^\top \begin{bmatrix} \phi_{\mathbf{p}_i} \\ \tilde{\mathbf{p}}_i \end{bmatrix} = \beta(\mathbf{p}_i).$$

This proves  $\mathcal{B}(\mathbf{D}') = \mathcal{B}(\mathbf{D})$  where  $\mathbf{D}' = \mathbf{R}\mathbf{D} + \mathbf{t}\mathbf{1}^\top$ . Denote  $\bar{\mathcal{E}}'_\lambda$  to be the top  $l \times l$  sub-block of  $\mathcal{E}'_\lambda$ . It is obvious that  $\bar{\mathcal{E}}'_\lambda = \bar{\mathcal{E}}_\lambda$  thus the regularization matrix  $\mathbf{Z}' = \sqrt{\bar{\mathcal{E}}'_\lambda} = \sqrt{\bar{\mathcal{E}}_\lambda} = \mathbf{Z}$  remains unchanged.

## B Lemmas and Proof of Theorem 2

### B.1 Lemmas on Positive Semidefinite Matrices

**Lemma 7.** For any  $M = \sum_{i=1}^n M_i$ , if  $M_i \succeq \mathbf{O}$  for each  $i \in [1 : n]$ , then  $M \succeq \mathbf{O}$ .

**Lemma 8.** For any  $M \succeq \mathbf{O}$ , we have  $\mathbf{x}^\top M \mathbf{x} = 0 \iff M \mathbf{x} = \mathbf{0}$ .

*Proof.* Given  $M \succeq \mathbf{O}$ , there exist a matrix  $\sqrt{M}$ , such that  $M = \sqrt{M}\sqrt{M}$ . Therefore:

$$\mathbf{x}^\top M \mathbf{x} = \|\sqrt{M}\mathbf{x}\|_2^2 = 0 \iff \sqrt{M}\mathbf{x} = \mathbf{0} \implies M\mathbf{x} = \sqrt{M}\sqrt{M}\mathbf{x} = \mathbf{0}.$$

This completes the proof of the sufficiency. The necessity is obvious.  $\square$

**Lemma 9.** For any  $M = \sum_{i=1}^n M_i \succeq \mathbf{O}$ , with each  $M_i \succeq \mathbf{O}$ , we have  $M\mathbf{x} = \mathbf{0} \iff M_i\mathbf{x} = \mathbf{0}$  for each  $i \in [1 : n]$ .

*Proof.* Because  $M \succeq \mathbf{O}$  and  $M_i \succeq \mathbf{O}$ , by Lemma 8, it suffices to show  $\mathbf{x}^\top M \mathbf{x} = 0 \iff \mathbf{x}^\top M_i \mathbf{x} = 0$ , which is true because  $\mathbf{x}^\top M \mathbf{x} = \sum_{i=1}^n \mathbf{x}^\top M_i \mathbf{x}$  with each summand  $\mathbf{x}^\top M_i \mathbf{x} \geq 0$ . Therefore  $\mathbf{x}^\top M \mathbf{x} = 0 \iff \mathbf{x}^\top M_i \mathbf{x} = 0$  for each  $i \in [1 : n]$ .  $\square$

### B.2 Proof of Theorem 2

We assume  $\mu_i$  being a tuning parameter  $\mu_i \geq 0$ . Then the following chain of generalized equality holds:

$$I \succeq \mathcal{B}_i^\top (\mathcal{B}_i \mathcal{B}_i^\top)^{-1} \mathcal{B}_i \succeq \mathcal{B}_i^\top (\mathcal{B}_i \mathcal{B}_i^\top + \mu_i \mathbf{Z}_i^\top \mathbf{Z}_i)^{-1} \mathcal{B}_i \succeq \mathbf{O}.$$

Therefore  $I - Q_i \succeq \mathbf{O}$ , and thus  $\mathcal{P}_{\Pi} = \sum_{i=1}^n (I - Q_i) \succeq \mathbf{O}$ .

*Proof.* We shall prove the result by showing: (a)  $\iff$  (b), (a)  $\iff$  (c), (a)  $\iff$  (d), and (c)  $\iff$  (e).

(a)  $\iff$  (b). Because  $\mathcal{P}_{\Pi} = nI - \mathcal{Q}_{\Pi}$ , thus  $\mathcal{P}_{\Pi}\mathbf{1} = n\mathbf{1} - \mathcal{Q}_{\Pi}\mathbf{1}$ , which means  $\mathcal{P}_{\Pi}\mathbf{1} = \mathbf{0} \iff \mathcal{Q}_{\Pi}\mathbf{1} = n\mathbf{1}$ .

(a)  $\iff$  (c). Note that  $\mathcal{P}_{\Pi} = \sum_{i=1}^n (I - Q_i) \succeq \mathbf{O}$  with  $I - Q_i \succeq \mathbf{O}$ . Therefore by Lemma 9, we have  $\mathcal{P}_{\Pi}\mathbf{1} = \mathbf{0} \iff (I - Q_i)\mathbf{1} = \mathbf{0} \iff Q_i\mathbf{1} = \mathbf{1}$ .

(a)  $\iff$  (d). Applying a translation  $\mathbf{t}$  to  $\mathcal{S}$ , we have:

$$\text{tr}(\mathcal{S}\mathcal{P}_{\Pi}\mathcal{S}^\top) = \text{tr}((\mathcal{S} + \mathbf{t}\mathbf{1}^\top)\mathcal{P}_{\Pi}(\mathcal{S} + \mathbf{t}\mathbf{1}^\top)^\top) \iff -2\text{tr}(\mathcal{S}\mathcal{P}_{\Pi}\mathbf{1}\mathbf{t}^\top) = \text{tr}(\mathbf{t}\mathbf{1}^\top\mathcal{P}_{\Pi}\mathbf{1}\mathbf{t}^\top). \quad (34)$$

Sufficiency: if  $\mathcal{P}_{\Pi}\mathbf{1} = \mathbf{0}$ , then  $-2\text{tr}(\mathcal{S}\mathcal{P}_{\Pi}\mathbf{1}\mathbf{t}^\top) = \text{tr}(\mathbf{t}\mathbf{1}^\top\mathcal{P}_{\Pi}\mathbf{1}\mathbf{t}^\top) = 0$ .

Necessity: note that  $\mathcal{P}_{\text{II}}$  is positive semidefinite, so  $\text{tr}(\mathbf{t}\mathbf{1}^\top\mathcal{P}_{\text{II}}\mathbf{t}^\top) \geq 0$ , which means  $\text{tr}(\mathcal{S}\mathcal{P}_{\text{II}}\mathbf{t}^\top) \leq 0$ . Since equation (34) holds for any  $\mathbf{t}$ , flip the sign of  $\mathbf{t}^\top$ , we obtain  $\text{tr}(\mathcal{S}\mathcal{P}_{\text{II}}\mathbf{t}^\top) \geq 0$ . Therefore:

$$0 \leq \text{tr}(\mathcal{S}\mathcal{P}_{\text{II}}\mathbf{t}^\top) \leq 0 \iff \text{tr}(\mathcal{S}\mathcal{P}_{\text{II}}\mathbf{t}^\top) = 0 \iff \text{tr}(\mathbf{t}\mathbf{1}^\top\mathcal{P}_{\text{II}}\mathbf{t}^\top) = 0.$$

Note that  $\mathbf{1}^\top\mathcal{P}_{\text{II}}\mathbf{1}$  is a scalar, hence:

$$\text{tr}(\mathbf{t}\mathbf{1}^\top\mathcal{P}_{\text{II}}\mathbf{t}^\top) = \mathbf{1}^\top\mathcal{P}_{\text{II}}\mathbf{1} \text{tr}(\mathbf{t}\mathbf{t}^\top) = \mathbf{1}^\top\mathcal{P}_{\text{II}}\mathbf{1}\|\mathbf{t}\|_2^2 = 0,$$

for any  $\mathbf{t}$ , if and only if  $\mathbf{1}^\top\mathcal{P}_{\text{II}}\mathbf{1} = 0 \iff \mathcal{P}_{\text{II}}\mathbf{1} = \mathbf{0}$  by Lemma 8.

(c)  $\iff$  (e). Let  $\bar{\mathbf{Q}}_i = \mathcal{B}_i^\top(\mathcal{B}_i\mathcal{B}_i^\top)^{-1}\mathcal{B}_i$ , and  $(\mathcal{B}_i^\top)^\dagger = (\mathcal{B}_i\mathcal{B}_i^\top)^{-1}\mathcal{B}_i$ . Applying the Woodbury matrix identity, the matrix  $\mathbf{Q}_i$  can be expanded as:

$$\begin{aligned} \mathbf{Q}_i &= \mathcal{B}_i^\top(\mathcal{B}_i\mathcal{B}_i^\top + \mu_i\mathbf{Z}_i^\top\mathbf{Z}_i)^{-1}\mathcal{B}_i \\ &= \bar{\mathbf{Q}}_i - \mu_i \underbrace{\left(\mathbf{Z}_i(\mathcal{B}_i^\top)^\dagger\right)^\top \left(\mathbf{I} + \mu_i\mathbf{Z}_i(\mathcal{B}_i\mathcal{B}_i^\top)^{-1}\mathbf{Z}_i^\top\right)^{-1} \mathbf{Z}_i(\mathcal{B}_i^\top)^\dagger}_{\Delta_i}. \end{aligned}$$

Then we write:

$$\mathbf{I} - \mathbf{Q}_i = (\mathbf{I} - \bar{\mathbf{Q}}_i) + \mu_i\Delta_i \succeq \mathbf{O},$$

with  $\mathbf{I} - \bar{\mathbf{Q}}_i \succeq \mathbf{O}$  and  $\Delta_i \succeq \mathbf{O}$ . Noting that  $\mu_i \geq 0$ , by Lemma 9, we have  $(\mathbf{I} - \mathbf{Q}_i)\mathbf{1} = \mathbf{0}$  if and only if  $(\mathbf{I} - \bar{\mathbf{Q}}_i)\mathbf{1} = \mathbf{0}$  and  $\Delta_i\mathbf{1} = \mathbf{0}$ . In other words,  $\mathbf{Q}_i\mathbf{1} = \mathbf{1}$  if and only if  $\bar{\mathbf{Q}}_i\mathbf{1} = \mathbf{1}$  and  $\Delta_i\mathbf{1} = \mathbf{0}$ .

Note that  $\bar{\mathbf{Q}}_i = \mathcal{B}_i^\top(\mathcal{B}_i\mathcal{B}_i^\top)^{-1}\mathcal{B}_i$  is the orthogonal projection into the range space of  $\mathcal{B}_i^\top$ , which states  $\bar{\mathbf{Q}}_i\mathbf{1} = \mathbf{1} \iff \mathbf{1} \in \text{Range}(\mathcal{B}_i^\top) \iff \exists \mathbf{x}$  such that  $\mathcal{B}_i^\top\mathbf{x} = \mathbf{1}$ . Therefore such an  $\mathbf{x}$  is  $\mathbf{x} = (\mathcal{B}_i^\top)^\dagger\mathbf{1}$ .

Since  $\mathbf{Z}_i(\mathcal{B}_i\mathcal{B}_i^\top)^{-1}\mathbf{Z}_i^\top \succeq \mathbf{O}$ , we know  $\mathbf{I} + \mu_i\mathbf{Z}_i(\mathcal{B}_i\mathcal{B}_i^\top)^{-1}\mathbf{Z}_i^\top$  is exactly positive definite. Therefore  $\Delta_i\mathbf{1} = \mathbf{0} \iff \mathbf{1}^\top\Delta_i\mathbf{1} = 0$  happens if and only if  $\mathbf{Z}_i(\mathcal{B}_i^\top)^\dagger\mathbf{1} = \mathbf{0}$  which is  $\mathbf{Z}_i\mathbf{x} = \mathbf{0}$ .  $\square$

## C Proof of Theorem 3

*Proof.* (b)  $\iff$  (c). We define  $\bar{\mathbf{Q}}_i = \Gamma_i\mathcal{B}_i^\top(\mathcal{B}_i\Gamma_i\mathcal{B}_i^\top)^{-1}\mathcal{B}_i\Gamma_i$ . By the Woodbury matrix identity:

$$\begin{aligned} &\Gamma_i\mathcal{B}_i^\top(\mathcal{B}_i\Gamma_i\mathcal{B}_i^\top + \mu_i\mathbf{Z}_i^\top\mathbf{Z}_i)^{-1}\mathcal{B}_i\Gamma_i \\ &= \bar{\mathbf{Q}}_i - \mu_i \underbrace{\left(\mathbf{Z}_i(\Gamma_i\mathcal{B}_i^\top)^\dagger\right)^\top \left(\mathbf{I} + \mu_i\mathbf{Z}_i(\mathcal{B}_i\Gamma_i\mathcal{B}_i^\top)^{-1}\mathbf{Z}_i^\top\right)^{-1} \mathbf{Z}_i(\Gamma_i\mathcal{B}_i^\top)^\dagger}_{\Delta_i}. \end{aligned}$$

Then  $\mathbf{P}_i$  can be written as:  $\mathbf{P}_i = (\mathbf{\Gamma}_i - \bar{\mathbf{Q}}_i) + \mu_i \mathbf{\Delta}_i$  with  $\mathbf{\Gamma}_i - \bar{\mathbf{Q}}_i \succeq \mathbf{O}$  and  $\mathbf{\Delta}_i \succeq \mathbf{O}$ . Thus  $\mathbf{P}_i \succeq \mathbf{O}$ . By Lemma 9,  $\mathbf{P}_i \mathbf{1} = \mathbf{0}$  if and only if  $(\mathbf{\Gamma}_i - \bar{\mathbf{Q}}_i) \mathbf{1} = \mathbf{0}$  and  $\mathbf{\Delta}_i \mathbf{1} = \mathbf{0}$

Note that  $\bar{\mathbf{Q}}_i$  is the orthogonal projection to  $\mathbf{Range}(\mathbf{\Gamma}_i \mathbf{B}_i^\top)$ . Thus we have:

$$(\mathbf{\Gamma}_i - \bar{\mathbf{Q}}_i) \mathbf{1} = \mathbf{0} \iff \bar{\mathbf{Q}}_i \mathbf{\Gamma}_i \mathbf{1} = \mathbf{\Gamma}_i \mathbf{1} \iff \mathbf{\Gamma}_i \mathbf{1} \in \mathbf{Range}(\mathbf{\Gamma}_i \mathbf{B}_i^\top),$$

if and only if there exists  $\mathbf{x}$  such that  $\mathbf{\Gamma}_i \mathbf{B}_i^\top \mathbf{x} = \mathbf{\Gamma}_i \mathbf{1}$ . Such an  $\mathbf{x}$  is  $\mathbf{x} = (\mathbf{\Gamma}_i \mathbf{B}_i^\top)^\dagger \mathbf{1}$ .

Note that  $\mathbf{\Delta}_i \succeq \mathbf{O}$  and  $\mathbf{I} + \mu_i \mathbf{Z}_i (\mathbf{B}_i \mathbf{\Gamma}_i \mathbf{\Gamma}_i \mathbf{B}_i^\top)^{-1} \mathbf{Z}_i^\top \succ \mathbf{O}$ , thus we have:

$$\mathbf{\Delta}_i \mathbf{1} = \mathbf{0} \iff \mathbf{1}^\top \mathbf{\Delta}_i \mathbf{1} = 0 \iff \mathbf{Z}_i \mathbf{x} = \mathbf{Z}_i (\mathbf{\Gamma}_i \mathbf{B}_i^\top)^\dagger \mathbf{1} = \mathbf{0}.$$

(a)  $\iff$  (b). Since  $\mathbf{P}_i \succeq \mathbf{O}$ , the proof is immediate by Lemma 9.

This completes the proof. □

## D Rigid Procrustes Analysis with Two Shapes

The pairwise Procrustes problem aims to solve for a similarity transformation (a scale factor  $s > 0$ , a rotation/reflection  $\mathbf{R} \in \text{O}(3)$ , and a translation  $\mathbf{t} \in \mathbb{R}^3$ ), that minimizes the registration error between the two point-clouds  $\mathbf{D}_1$  and  $\mathbf{D}_2$  with known correspondences:

$$\arg \min_{s>0, \mathbf{R} \in \text{O}(3)} \|s\mathbf{R}\mathbf{D}_1 + \mathbf{t}\mathbf{1}^\top - \mathbf{D}_2\|_F^2.$$

This problem admits a closed-form solution (Horn, 1987; Horn et al., 1988; Kanatani, 1994).

Let  $\bar{\mathbf{D}}_1 = \mathbf{D}_1 - \frac{1}{m} \mathbf{D}_1 \mathbf{1} \mathbf{1}^\top$ , and  $\bar{\mathbf{D}}_2 = \mathbf{D}_2 - \frac{1}{m} \mathbf{D}_2 \mathbf{1} \mathbf{1}^\top$ . Let the SVD of  $\bar{\mathbf{D}}_1 \bar{\mathbf{D}}_2^\top$  be  $\bar{\mathbf{D}}_1 \bar{\mathbf{D}}_2^\top = \bar{\mathbf{U}} \bar{\mathbf{\Sigma}} \bar{\mathbf{V}}^\top$ . Then the optimal rotation is  $\mathbf{R}^* = \bar{\mathbf{V}} \bar{\mathbf{U}}^\top$  if  $\mathbf{R}^* \in \text{O}(3)$ . If we require  $\mathbf{R}^* \in \text{SO}(3)$ , the optimal rotation is:

$$\mathbf{R}^* = \bar{\mathbf{V}} \text{diag}(1, 1, \det(\bar{\mathbf{V}} \bar{\mathbf{U}}^\top)) \bar{\mathbf{U}}^\top.$$

The optimal scale factor and the optimal translation are then given by:

$$s^* = \frac{\text{tr}(\mathbf{R}^* \bar{\mathbf{D}}_1 \bar{\mathbf{D}}_2^\top)}{\text{tr}(\bar{\mathbf{D}}_1 \bar{\mathbf{D}}_1^\top)}$$

$$\mathbf{t}^* = \frac{1}{m} \mathbf{D}_2 \mathbf{1} - \frac{1}{m} s^* \mathbf{R}^* \mathbf{D}_1 \mathbf{1}.$$

## Acknowledgments

This work was supported by ANR via the TOPACS project. We thank the authors of the public datasets which we could use in our experiments. We appreciate the valuable comments of the anonymous reviewers that have helped improving the quality of the manuscript.

## References

- P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.
- B. Allen, B. Curless, and Z. Popović. The space of human body shapes: reconstruction and parameterization from range scans. *ACM transactions on graphics (TOG)*, 22(3):587–594, 2003.
- D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. SCAPE: shape completion and animation of people. In *ACM SIGGRAPH 2005 Papers*, pages 408–416. 2005.
- K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Transactions on pattern analysis and machine intelligence*, (5):698–700, 1987.
- F. Bai, T. Vidal-Calleja, and G. Grisetti. Sparse pose graph optimization in cycle space. *IEEE Transactions on Robotics*, 37(5):1381–1400, 2021. doi: 10.1109/TRO.2021.3050328.
- A. Bartoli. Towards 3d motion estimation from deformable surfaces. In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, pages 3083–3088. IEEE, 2006.
- A. Bartoli, M. Perriollat, and S. Chambon. Generalized thin-plate spline warps. *International Journal of Computer Vision*, 88(1):85–110, 2010.
- A. Bartoli, D. Pizarro, and M. Loog. Stratified generalized procrustes analysis. *International journal of computer vision*, 101(2):227–253, 2013.
- R. Benjemma and F. Schmitt. A solution for the registration of multiple 3d point sets using unit quaternions. In *European Conference on Computer Vision*, pages 34–50. Springer, 1998.
- P. Bilic, P. F. Christ, E. Vorontsov, G. Chlebus, H. Chen, Q. Dou, C.-W. Fu, X. Han, P.-A. Heng, J. Hesser, et al. The liver tumor segmentation benchmark (lits). *arXiv preprint arXiv:1901.04056*, 2019.
- P. Birtea, I. Cașu, and D. Comănescu. First order optimality conditions and steepest descent algorithm on orthogonal stiefel manifolds. *Optimization Letters*, 13(8):1773–1791, 2019.



- C. M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.
- F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on pattern analysis and machine intelligence*, 11(6):567–585, 1989.
- F. L. Bookstein. *Morphometric tools for landmark data: geometry and biology*. Cambridge University Press, 1997.
- S. Bouix, J. C. Pruessner, D. L. Collins, and K. Siddiqi. Hippocampal shape analysis using medial surfaces. *Neuroimage*, 25(4):1077–1089, 2005.
- M. Bro-Nielsen and C. Gramkow. Fast fluid registration of medical images. In *International Conference on Visualization in Biomedical Computing*, pages 265–276. Springer, 1996.
- R. W. Brockett. Least squares matching problems. *Linear Algebra and its applications*, 122:761–777, 1989.
- B. J. Brown and S. Rusinkiewicz. Global non-rigid alignment of 3-d scans. In *ACM SIGGRAPH 2007 papers*, pages 21–es. 2007.
- G. Christensen and J. He. Consistent nonlinear elastic image registration. In *Proceedings IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA 2001)*, pages 37–43. IEEE, 2001.
- T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.
- T. A. Davis. *Direct methods for sparse linear systems*. SIAM, 2006.
- I. L. Dryden and K. V. Mardia. *Statistical shape analysis: with applications in R*, volume 995. John Wiley & Sons, 2016.
- J. Duchon. Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. *Revue française d’automatique, informatique, recherche opérationnelle. Analyse numérique*, 10(R3):5–12, 1976.
- D. W. Eggert, A. Lorusso, and R. B. Fisher. Estimating 3-d rigid body transformations: a comparison of four major algorithms. *Machine vision and applications*, 9(5-6):272–290, 1997.
- P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE transactions on medical imaging*, 23(8):995–1005, 2004.

- M. Fornefett, K. Rohr, and H. S. Stiehl. Radial basis functions with compact support for elastic registration of medical images. *Image and vision computing*, 19(1-2):87–96, 2001.
- O. Freifeld and M. J. Black. Lie bodies: A manifold representation of 3d human shape. In *European Conference on Computer Vision*, pages 1–14. Springer, 2012.
- M. Gallardo, T. Collins, and A. Bartoli. Dense non-rigid structure-from-motion and shading with unknown albedos. In *Proceedings of the IEEE international conference on computer vision*, pages 3884–3892, 2017.
- G. Golub and V. Pereyra. Separable nonlinear least squares: the variable projection method and its applications. *Inverse problems*, 19(2):R1, 2003.
- C. Goodall. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(2):285–321, 1991.
- J. C. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, 1975.
- B. F. Green. The orthogonal approximation of an oblique structure in factor analysis. *Psychometrika*, 17(4):429–440, 1952.
- G. Hardy, K. M. R. Collection, J. Littlewood, G. Pólya, G. Pólya, and D. Littlewood. *Inequalities*. Cambridge Mathematical Library. Cambridge University Press, 1952. ISBN 9780521358804.
- B. K. Horn. Closed-form solution of absolute orientation using unit quaternions. *Josa a*, 4(4):629–642, 1987.
- B. K. Horn, H. M. Hilden, and S. Negahdaripour. Closed-form solution of absolute orientation using orthonormal matrices. *JOSA A*, 5(7):1127–1135, 1988.
- A. Iserles, H. Z. Munthe-Kaas, S. P. Nørsett, and A. Zanna. Lie-group methods. *Acta numerica*, 9:215–365, 2000.
- I. H. Jermyn, S. Kurtek, E. Klassen, and A. Srivastava. Elastic shape matching of parameterized surfaces using square root normal fields. In *European conference on computer vision*, pages 804–817. Springer, 2012.
- S. H. Joshi, E. Klassen, A. Srivastava, and I. Jermyn. A novel representation for riemannian analysis of elastic curves in rn. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–7. IEEE, 2007.
- K.-i. Kanatani. Analysis of 3-d rotation fitting. *IEEE Transactions on pattern analysis and machine intelligence*, 16(5):543–549, 1994.

- D. G. Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London mathematical society*, 16(2):81–121, 1984.
- D. G. Kendall, D. Barden, T. K. Carne, and H. Le. *Shape and shape theory*, volume 500. John Wiley & Sons, 2009.
- J. T. Kent. The complex Bingham distribution and shape analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(2):285–299, 1994.
- M. Kilian, N. J. Mitra, and H. Pottmann. Geometric modeling in shape space. In *ACM SIGGRAPH 2007 papers*, pages 64–es. 2007.
- Y. J. Kim, S.-T. Chung, B. Kim, and S. Cho. 3d face modeling based on 3D dense morphable face shape model. *International Journal of Computer Science and Engineering*, 2(3):107–113, 2008.
- S. Krishnan, P. Y. Lee, J. B. Moore, S. Venkatasubramanian, et al. Global registration of multiple 3d point sets via optimization-on-a-manifold. In *Symposium on Geometry Processing*, pages 187–196, 2005.
- W. Kristof and B. Wingersky. A generalization of the orthogonal procrustes rotation procedure to more than two matrices. In *Proceedings of the Annual Convention of the American Psychological Association*. American Psychological Association, 1971.
- S. Kurtek, E. Klassen, Z. Ding, and A. Srivastava. A novel riemannian framework for shape analysis of 3d objects. In *2010 IEEE computer society conference on computer vision and pattern recognition*, pages 1625–1632. IEEE, 2010.
- S. Kurtek, E. Klassen, J. C. Gore, Z. Ding, and A. Srivastava. Elastic geodesic paths in shape space of parameterized surfaces. *IEEE transactions on pattern analysis and machine intelligence*, 34(9):1717–1730, 2011.
- H. Laga. A survey on non-rigid 3D shape analysis, 2018.
- H. Laga, Q. Xie, I. H. Jermyn, and A. Srivastava. Numerical inversion of srnf maps for elastic shape analysis of genus-zero surfaces. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2451–2464, 2017.
- B. Matei and P. Meer. Optimal rigid motion estimation and performance evaluation with bootstrap. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, volume 1, pages 339–345. IEEE, 1999.

- C. D. Meyer. *Matrix analysis and applied linear algebra*, volume 71. Siam, 2000.
- N. Ohta and K. Kanatani. Optimal estimation of three-dimensional rotation and reliability evaluation. *IEICE TRANSACTIONS on Information and Systems*, 81(11):1247–1252, 1998.
- S. Osher and R. P. Fedkiw. *Level set methods and dynamic implicit surfaces*, volume 153. Springer, 2003.
- F. J. Rohlf and D. Slice. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic biology*, 39(1):40–59, 1990.
- D. M. Rosen, L. Carlone, A. S. Bandeira, and J. J. Leonard. Se-sync: A certifiably correct algorithm for synchronization over the special euclidean group. *The International Journal of Robotics Research*, 38(2-3):95–125, 2019.
- D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE transactions on medical imaging*, 18(8):712–721, 1999.
- P. H. Schönemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31(1):1–10, 1966.
- J. Song, F. Bai, L. Zhao, S. Huang, and R. Xiong. Efficient two step optimization for large embedded deformation graph based slam. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9419–9425. IEEE, 2020.
- R. W. Sumner, J. Schmid, and M. Pauly. Embedded deformation for shape manipulation. In *ACM SIGGRAPH 2007 papers*, pages 80–es. 2007.
- R. Szeliski and J. Coughlan. Spline-based image registration. *International Journal of Computer Vision*, 22(3):199–218, 1997.
- J. M. Ten Berge. Orthogonal procrustes rotation for two or more matrices. *Psychometrika*, 42(2):267–276, 1977.
- C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *International journal of computer vision*, 9(2):137–154, 1992.
- S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (4):376–380, 1991.

- 
- M. W. Walker, L. Shao, and R. A. Volz. Estimating 3-d location parameters using dual number quaternions. *CVGIP: image understanding*, 54(3):358–367, 1991.
- G. Wen, Z. Wang, S. Xia, and D. Zhu. Least-squares fitting of multiple m-dimensional point sets. *The Visual Computer*, 22(6):387–398, 2006.
- J. Williams and M. Bennamoun. Simultaneous registration of multiple corresponding point sets. *Computer Vision and Image Understanding*, 81(1):117–142, 2001.
- L. Younes. Spaces and manifolds of shapes in computer vision: An overview. *Image and Vision Computing*, 30(6-7):389–397, 2012.
- L. Younes, P. W. Michor, J. M. Shah, and D. B. Mumford. A metric on shape space with explicit geodesics. *Rendiconti Lincei-Matematica e Applicazioni*, 19(1):25–57, 2008.