Eidgenössische Technische Hochschule Zürich

Ecole polytechnique fédérale de Zurich
Politecnico federale di Zurigo
Swiss Federal Institute of Technology Zurich

Optimal image alignment with random projections of manifolds: algorithm and geometric analysis

E. Kokiopoulou, D. Kressner and P. Frossard*

Research Report No. 2009-41
December 2009

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

*Signal Processing Laboratory (LTS4), Ecole Polytechnique Fédérale de Lausanne, Switzerland

Optimal image alignment with random projections of manifolds: algorithm and geometric analysis

Effrosyni Kokiopoulou, Daniel Kressner and Pascal Frossard

***Abstract*—This paper addresses image alignment based on random measurements. Image alignment consists in estimating the relative transformation between a query image and a reference image. We consider the specific problem where the query image is not given exactly, but rather provided in a compressed form with linear measurements captured by a vision sensor. According to the theory behind compressed sensing, image alignment can still be performed effectively in this case, provided that the number of measurements is sufficiently large. We cast the alignment problem as a manifold distance minimization problem in the linear subspace defined by the measurements. We then show that, when the reference image is sparsely represented over parametric dictionaries, the corresponding objective function can be decomposed as the difference of two convex functions (DC). Thus the optimization problem becomes a DC program, which in turn can be solved globally optimally by, e.g., a cutting plane method. The quality of the solution is typically affected by the number of random measurements and the condition number of the manifold that describes the transformations of the reference image. We show that the manifold condition number remains bounded in our image alignment problem, which means that the relative transformation between two images can be determined optimally in a reduced subspace.**

***Index Terms*—Pattern transformations, transformation manifolds, sparse representations, random projections, manifold condition number.**

I. Introduction

The problem of image alignment is of paramount importance and enjoys numerous applications in various fields including pattern recognition, computer vision and medical image analysis, to name just a few [1]. The comparison of two visual patterns is generally only meaningful if they are aligned first, so that their distance reflects their structural and geometric differences. Image alignment consists of estimating the relative transformation between patterns. The transformed version of a pattern can be described as a point of a submanifold in a high dimensional space, which is usually called the *transformation manifold*. The manifold distance (MD) is the minimum distance between the query image p and the manifold generated by the reference image s, see Figure 1.

In certain applications, we might not have access to the full query image or it might be computationally too expensive to deal with the complete image. In this paper, we only work with a few random projections of the query image. According to the theory of compressed sensing (CS), a few random projections of a sparse (or nearly sparse) signal are sufficient to preserve its salient information. Moreover, in [2], [3], [4] it is shown that random projections of signal manifolds result into approximately isometric embeddings, i.e., pairwise Euclidean distances are nearly preserved in the reduced space.

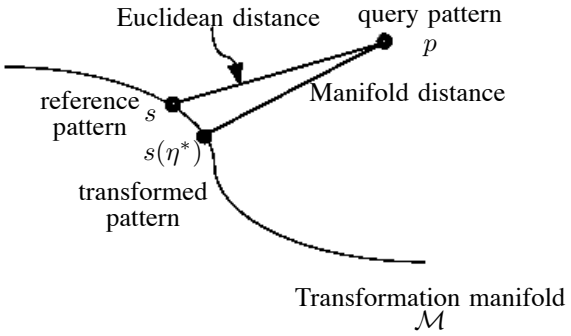

Fig. 1. Manifold distance is the minimum distance from a query point p to the transformation manifold $\mathcal{M}$ spanned by the transformed versions of s.

In the context of image alignment, we use linear measurements in order to estimate the relative geometric transformation between a query and a reference image. While the use of linear measurements permits to reduce the computational complexity of the estimation, it further permits to extend the image alignment problem to imperfect settings where the representation of the query image is provided in a compressed form by low complexity vision sensors.

In this paper, we propose a new method for image alignment, which estimates the globally optimal transformation between a query image and a reference image s, using a sufficient number of random measurements. For this purpose, we represent s as a sparse linear combination of geometric primitives, called atoms, which are chosen from a parametric, possibly redundant dictionary. This representation permits to build a parametric definition of the transformation manifold, which describes the possible transformations of the reference image. The image alignment problem can then be cast as a manifold distance minimization problem. Building on our previous work [5], [6], we then formulate the pattern alignment problem with random measurements as a DC program by showing that the objective function can be represented as a *difference of convex functions* (DC). DC programs are non-convex problems that can be solved by efficient globally optimal algorithms by exploiting their special structure. Our approach therefore provides a feasible way to perform image

E. Kokiopoulou and D. Kressner are with the Seminar for Applied Mathematics, Department of Mathematics, ETH Zurich, CH-8092 Zurich email: {effrosyni.kokiopoulou,daniel.kressner}@sam.math.ethz.ch

P. Frossard is with the Signal Processing Laboratory (LTS4), Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne e-mail: pascal.frossard@epfl.ch.

alignment with random measurements.

At the same time, the results from [4] suggest that the number of measurements necessary for proper alignment depends on the condition number of the pattern transformation manifold. We perform a geometric analysis of this manifold, showing that it is well conditioned and providing an explicit upper bound on the condition number. Moreover, we estimate this number numerically by establishing an efficient procedure for computing the principal curvature at a certain point on the manifold. These results confirm that the required number of measurements is bounded in our problem, and that image alignment can be solved efficiently in a reduced subspace. In summary, the contribution of this paper consists of providing (1) a globally optimal approach to image alignment with random projections and (2) theoretical as well as numerical insights into the geometric properties of the pattern transformation manifolds that influence the number of necessary measurements.

The rest of this paper is organized as follows. In Sec. II we formulate the problem of image alignment from random measurements. In Sec. III we discuss the representation of transformation manifolds using sparse geometric expansions. We then show in Sec. IV that the distance between the query pattern and the transformation manifold is a DC function of the transformation parameters. In Sec. V we provide a geometric analysis of transformation manifolds which confirms that a limited number of measurements is sufficient for accurate alignment. Finally, experimental results are presented in Sec. VI.

II. Problem formulation

In this paper, we are interested in estimating the relative transformation η^* that matches best two visual patterns. This problem is very common in image processing or computer vision applications that rely on image alignment or matching of views. In particular, we consider the problem where only a compressed version of the query pattern is available, which is given by linear measurements captured by the vision sensor. Formally, we consider that images undergo transformations described by the parameter η. We further consider that all the transformed versions $s(\eta)$ of the reference pattern s can be represented by a low dimensional transformation manifold $\mathcal{M}$. Then, we assume that we have m random projections of the query pattern p, obtained by computing inner products with m random signals $z_1, \ldots, z_m$. The image alignment problem is equivalent to a manifold distance minimization problem in the linear subspace defined by the measurement vectors. It can be formulated as a parameter estimation problem as follows.

Transformation estimation problem

$$\eta^* = \arg\min_{\eta} f(\eta), \quad \text{where}$$
$$f(\eta) = \sum_{i=1}^{m} |\langle s(\eta), z_i\rangle - \langle p, z_i\rangle|. \tag{1}$$

The optimization problem (1) for determining the best transformation parameters η^* is typically non-convex [7]. This makes it hard to solve using traditional methods, such as steepest descent or Newton-type methods due to their local convergence property and the presence of an unknown number of local minima. We show in the next section how the transformation manifold could be described in a parametric form. This enables to write the objective function as a difference of two convex (DC) functions, as shown in Section IV. This in turn allows us to formulate the optimization problem as a DC program and solve it globally by a cutting plane method [8, Thm 5.3].

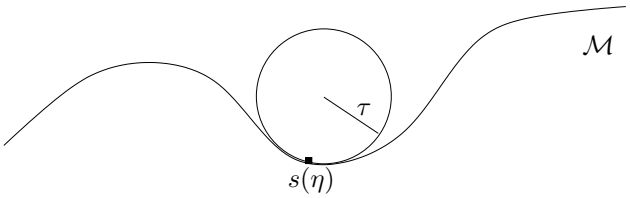

Fig. 2. A large τ corresponds to a well conditioned manifold, which has a low curvature.

One still has to choose the number of measurements that leads to efficient transformation estimation. The optimal number of random projections is hard to define in practice. Suppose that we project the transformation manifolds spanned by two distinct patterns on m random vectors $z_1, \ldots, z_m$. In order to make sure that matching points in the reduced space is close to matching the corresponding points in the initial high-dimensional space, the embedding should be nearly isometric, that is, pairwise Euclidean distances should be nearly preserved. Only if this is the case, one can reliably perform image alignment in the reduced space and estimate the unknown transformation.

Recently, Baraniuk and Wakin [2] provide an estimate of m that is linear in d and logarithmic in n, the number of pixels in the image. We revisit the main result from [2].

***Theorem** 1:* Let $\mathcal{M}$ be a compact d-dimensional manifold in $\mathbb{R}^n$ having condition number $1/\tau$, volume V, and geodesic covering regularity R. Fix $0 < \epsilon < 1$ and $0 < \rho < 1$. Let Z be a random orthoprojector from $\mathbb{R}^n$ to $\mathbb{R}^m$ and

$$m \geq O\Big(\frac{d \log(nVR\tau^{-1}\epsilon^{-1}) \log(\rho^{-1})}{\epsilon^2}\Big). \tag{2}$$

Suppose $m < n$. Then, with probability exceeding $1-\rho$, the following statement holds: For every pair of points $x, y \in \mathcal{M}$,

$$(1-\epsilon)\sqrt{\frac{m}{n}} \leq \frac{\|Zx - Zy\|_2}{\|x-y\|_2} \leq (1+\epsilon)\sqrt{\frac{m}{n}}. \tag{3}$$

Roughly speaking, Theorem 1[1] is proved by determining a high-resolution sampling on the manifold and then applying the Johnson-Lindenstrauss lemma [9] to the sampled points. The above theorem implies that besides d and n, the number m depends logarithmically on other properties of the manifold, such as its condition number $1/\tau$, volume V and geodesic covering regularity R. Note that R is closely related to the condition number [2] and we will omit its definition. Intuitively, the condition number of $\mathcal{M}$ is defined as $1/\tau$, where

[1]Note that even if $f(\eta)$ in (1) is based on the 1-norm distance in our formulation, Theorem 1 still provides insights on the required number of measurements, by the equivalence of norms in finite-dimensional spaces (i.e., a good match in the 1-norm yields a good match in the 2-norm).

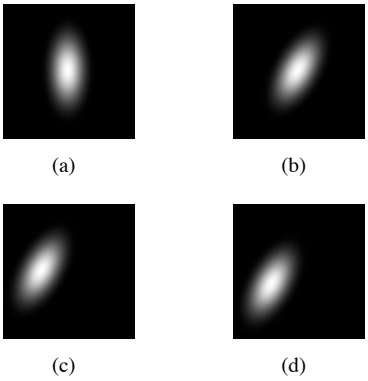

Fig. 3. Sample Gaussian atoms; (a) $b_x = 0$, $b_y = 0$, $a_x = 10$, $a_y = 5$, $\omega = 0$, (b) $b_x = 0$, $b_y = 0$, $a_x = 10$, $a_y = 5$, $\omega = -\pi/6$, (c) $b_x = 0$, $b_y = -10$, $a_x = 10$, $a_y = 5$, $\omega = -\pi/6$ (d) $b_x = 5$, $b_y = -10$, $a_x = 10$, $a_y = 5$, $\omega = -\pi/6$

τ is the maximum radius of a sphere that, when placed tangent to any point in $\mathcal{M}$, it intersects $\mathcal{M}$ only at that point (see [10] for a precise definition). This is illustrated graphically in Fig. 2 showing that a large τ implies a well conditioned manifold. Therefore, the condition number is an important property of $\mathcal{M}$ towards characterizing the number m of required random measurements. We show in Section V that our transformation manifold is well conditioned, which means that the number of necessary measurements is clearly bounded in the problem (1).

III. Transformation manifolds

A. Visual pattern representation

We show in this section how one could build a parametric transformation manifold for a reference pattern s. We first explain the representation of the pattern as a linear combination of geometric functions (usually called *atoms*), taken from a structured parametric and possibly redundant dictionary $\mathcal{D} = \{\phi_\gamma,\ \gamma \in \Gamma\}$ spanning the input space. This representation aims at capturing the most prominent features of the pattern. The atoms in a *parametric* dictionary are constructed by applying geometric transformations to a generating mother function denoted by ϕ. A geometric transformation $\gamma \in \Gamma$ can be represented by the action of an operator $U(\gamma)$ and therefore the parametric dictionary takes the form

$$\mathcal{D} = \{\ \phi_\gamma = U(\gamma)\phi,\ \gamma \in \Gamma\}. \tag{4}$$

A transformation γ_i, defining the ith atom, is composed of elementary transformations of the following three types.

- *Translation* by $b_i = [b_{ix}\ b_{iy}]^\top$. $U(b_i)$ moves the generating function across the image, i.e., $U(b_i)\phi(x,y) = \phi(x - b_{ix}, y - b_{iy})$.
- *Rotation* by ω_i. $U(\omega_i)$ rotates the generating function by the angle ω_i, i.e., $U(\omega_i)\phi(x,y) = \phi(\cos(\omega_i)x + \sin(\omega_i)y, \cos(\omega_i)y - \sin(\omega_i)x)$.
- *Anisotropic scaling* by $a_i = [a_{ix}\ a_{iy}]^\top$. $U(a_i)$ scales the generating function anisotropically, i.e., $U(a_i)\phi(x,y) = \phi(\frac{x}{a_{ix}}, \frac{y}{a_{iy}})$.

These elementary transformations yield a transformation $\gamma_i = (b_i, a_i, \omega_i) \in \Gamma$ as a synthesis of translations, anisotropic scalings, and rotations. It can be observed that applying a transformation on the mother function is equivalent to transforming the coordinate system from $\{x, y\}$ to $\{\tilde{x}, \tilde{y}\}$ before applying $\phi(\cdot)$. In particular, the ith atom $\phi_{\gamma_i} = U(\gamma_i)\phi(x,y)$ with $\gamma_i = (b_i, a_i, \omega_i) \in \Gamma$ can be regarded as the pullback

$$\phi_{\gamma_i}(x,y) = \phi(\Psi_{\gamma_i}(x,y)), \tag{5}$$

where $(\tilde{x}, \tilde{y}) := \Psi_{\gamma_i}(x,y)$ satisfies

$$\begin{aligned}\begin{bmatrix}\tilde{x}\\ \tilde{y}\end{bmatrix} &= \underbrace{\begin{bmatrix}\frac{1}{a_{ix}} & 0\\ 0 & \frac{1}{a_{iy}}\end{bmatrix}}_{A}\underbrace{\begin{bmatrix}\cos\omega_i & \sin\omega_i\\ -\sin\omega_i & \cos\omega_i\end{bmatrix}}_{R(\omega_i)}\underbrace{\begin{bmatrix}x - b_{ix}\\ y - b_{iy}\end{bmatrix}}_{t}\\ &= AR(\omega_i)t.\end{aligned} \tag{6}$$

The approximation of a pattern s with atoms from the dictionary $\mathcal{D}$ can be obtained in different ways. Even if finding the sparsest approximation of s is generally a hard problem, effective sub-optimal solutions are usually sufficient to capture the salient and geometric structure of the pattern with only a few atoms. In this work, we have chosen to use Orthogonal Matching Pursuit (OMP) [11, Sec. 9.5.3], which is a simple yet effective algorithm for computing sparse approximations in practice.

Initially, OMP chooses the residual $r_0 = s$ and then proceeds iteratively by selecting in the kth step the atom ϕ_{γ_k} that best matches the residual r_{k-1} i.e., $\gamma_k = \arg_{\gamma\in\Gamma}\max|\langle r_{k-1}, \phi_\gamma\rangle|$. Then γ_k is removed from the residual by projection: $r_k = (I - P_k)r_{k-1}$, where P_k is the orthogonal projector onto span$\{\phi_{\gamma_k}\}$. After K steps of OMP, the pattern s is approximated by a sparse linear combination of K atoms:

$$s \approx \sum_{k=1}^{K} \xi_k \phi_{\gamma_k}. \tag{7}$$

We propose the use of a dictionary of two-dimensional atoms capturing the geometric information in an image. The generating function ϕ of $\mathcal{D}$ used in this paper is the Gaussian

$$\phi(x,y) = \frac{1}{\rho}\exp(-(x^2 + y^2)). \tag{8}$$

Figure 3 shows a few sample Gaussian atoms corresponding to various geometric transformations γ. In addition, Figure 4 illustrates the progressive approximation of a human face from a Gaussian dictionary using OMP. Observe that already very few atoms are sufficient to capture the main geometric characteristics of the pattern and that the representation (7) does not need to be very accurate before it is useful for alignment purposes.

B. Transformation manifolds

In the following, we show how all the geometric transformations of the reference image s build a parametric transformation manifold. We restrict scalings to be isotropic, i.e., the geometric transformation η takes the form $\eta = (b, \alpha, \omega)$ consisting of a translation $b = [b_x, b_y]$, an isotropic scaling α, and a rotation ω. The manifold $\mathcal{M}$ of all such transformed images can be expressed mathematically as

$$\mathcal{M} = \big\{s(\eta) := U(\eta)s,\ \eta = (b, \alpha, \omega)\big\}. \tag{9}$$

Fig. 4. Progressive OMP approximation of a human face (leftmost) with 20, 50, 80, 110 and 140 Gaussian atoms (from left to right).

Fig. 5. Samples from the transformation manifold of the human face. From left to right, the samples correspond to rotation angles from 0 to 2π with step $\pi/4$.

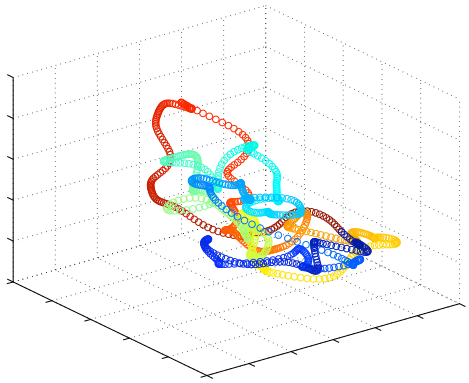

Fig. 6. Random projections in 3D of the rotation manifold of Figure 5. The samples correspond to rotation angles from 0 to 2π with step $\pi/500$. The color/shading is a linear map of the rotation angles.

Although the manifold resides in a high-dimensional space, its intrinsic dimension d is rather small and equal to the number of transformation parameters, which is 4. Figure 5 shows a few samples from the transformation manifold of a human face, with the transformation restricted to a rotation. The random projections of the resulting manifold are illustrated in Figure 6.

In general, all possible transformations η forms a group, the so called similitude group SIM(2) of the plane. As in (6), we denote

$$R(\omega) = \begin{bmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{bmatrix}, \quad 0 \leq \omega < 2\pi,$$

as the rotation matrix for the angle ω. If (b, α, ω) and (b', α', ω') are two elements of the SIM(2) group, then the group law [12] is given by

$$(b, \alpha, \omega) \circ (b', \alpha', \omega') = (b + \alpha R(-\omega)b', \alpha\alpha', \omega' + \omega). \quad (10)$$

In the following, we replace the reference image s by its approximation (7). Using the pullback interpretation (5) it follows that the transformation η applied to s results in

$$s(\eta) = U(\eta)s = \sum_{k=1}^{K} \xi_k U(\eta)\phi_{\gamma_k} = \sum_{k=1}^{K} \xi_k \phi_{\eta\circ\gamma_k}, \quad (11)$$

where $\eta \circ \gamma_k$ is a composition of transformations. In other words, the transformation is applied to each constituent atom individually. Furthermore, the group law (10) can be employed to determine the updated parameters of the transformed atoms. Let us emphasize the importance of equation (11): it allows to express the manifold equation (9) in closed form with respect to the transformation parameters η. The definition of the manifold in (11) is used in the optimization problem (1). We show that the resulting manifold distance problem can then be defined as the difference of two convex functions, which permits the use of the DC programming methodology that is proposed in the next section.

IV. DC DECOMPOSITION

A. Properties of DC functions

The purpose of this section is to show that the objective function (1) is DC. We start with some definitions and basic properties about DC functions [8], [13], [14]. Let $X \subseteq \mathbb{R}^n$ be convex. A function $f : X \to \mathbb{R}$ is called DC on X if there exist two convex functions $g, h : X \to \mathbb{R}$ such that

$$f(x) = g(x) - h(x). \quad (12)$$

A representation of this form is called *DC decomposition* of f. DC decompositions are clearly not unique; for any convex function $c(x)$, the decomposition $f(x) = (g(x) + c(x)) - (h(x) + c(x))$ is also DC. We will make use of the following two properties.

***Proposition** 1 (Properties of DC functions [13, Sec 4.2]):* Let $f = g - h$ and $f_i = g_i - h_i,\ i = 1 \ldots, m$ be DC functions. Then the following functions are also DC:

(a) $\sum_{i=1}^{m} \lambda_i f_i = \left[\sum_{\{i:\lambda_i \geq 0\}} \lambda_i g_i - \sum_{\{i:\lambda_i < 0\}} \lambda_i h_i\right] - \left[\sum_{\{i:\lambda_i \geq 0\}} \lambda_i h_i - \sum_{\{i:\lambda_i < 0\}} \lambda_i g_i\right]$.

(b) $|f| = 2\max\{g, h\} - (g + h)$.

B. DC form of the objective function

We now combine Proposition 1 with our previous results [6] to prove the main result of this paper.

***Theorem** 2:* The objective function f in (1) is DC.

Proof: Recall that

$$\begin{aligned} f(\eta) &= \sum_{i=1}^{m} |\langle s(\eta), z_i\rangle - \langle p, z_i\rangle| \\ &= \sum_{i=1}^{m} \Big| \sum_{k=1}^{K} \xi_k \langle \phi_{\eta_k}, z_i\rangle - \langle p, z_i\rangle \Big|, \end{aligned} \tag{13}$$

where $\eta_k = \eta \circ \gamma_k$. In [6] we have shown that (i) the transformed generating functions ϕ_{η_k} are DC, (ii) the inner products $\langle \phi_{\eta_k}, z\rangle$ between the atoms and a fixed pattern z are DC, and (iii) the inner product $\langle s(\eta), z\rangle = \sum_{k=1}^{K} \xi_k \langle \phi_{\eta_k}, z\rangle$ is also a DC function of η.

In particular, each function $\langle s(\eta), z_i\rangle = \sum_{k=1}^{K} \xi_k \langle \phi_{\eta_k}, z_i\rangle$ corresponding to a measurement vector z_i, with $1 \leq i \leq m$, is DC. Note that $f_i(\eta) := \langle s(\eta), z_i\rangle - \langle p, z_i\rangle$ remains DC since the second term is constant and does not depend on η. Assume now that the DC decomposition of each function f_i is given by $f_i(\eta) = g_i(\eta) - h_i(\eta)$.

By Proposition 1(b), the absolute value of a DC function is DC and hence

$$|f_i(\eta)| = 2\max\{g_i, h_i\} - (g_i + h_i) = \tilde{g}_i(\eta) - \tilde{h}_i(\eta),$$

is also DC. Finally, the objective function in (13) is DC since it is simply a sum of M DC functions:

$$\begin{aligned} f(\eta) &= \sum_{i=1}^{m} |f_i(\eta)| = \sum_{i=1}^{m} (\tilde{g}_i(\eta) - \tilde{h}_i(\eta)) \\ &= \underbrace{\sum_{i=1}^{m} \tilde{g}_i(\eta)}_{g(\eta)} - \underbrace{\sum_{i=1}^{m} \tilde{h}_i(\eta)}_{h(\eta)}. \end{aligned}$$

□

C. DC programs

An optimization problem is called a DC program if it takes the form

$$\begin{aligned} \min_{x} \quad & f(x) = g(x) - h(x), \\ \text{s.t.} \quad & x \in X = \{x \in \mathbb{R}^n : \delta(x) \leq 0\}, \end{aligned} \tag{14}$$

where $g, h : X \to \mathbb{R}$ are convex functions and $\delta : \mathbb{R}^n \to \mathbb{R}$ is a convex function. Assume that (14) is solvable and denote its global minimum by ω^*. The next proposition provides an optimality condition for (14).

***Proposition** 2 ([8]):* The point $x^* \in X$ is an optimal solution to the DC problem (14) if and only if there exists $t^* \in \mathbb{R}$ such that

$$0 = \inf\{-h(x) + t : \ x \in X, \ t \in \mathbb{R}, \ g(x) - t \leq g(x^*) - t^*\}. \tag{15}$$

In this work, we have chosen to solve the DC Program (14) by the outer approximation cutting plane algorithm proposed in [8, Sec 5.3], for its simplicity and also due to the fact that the parameter space in our problem is four-dimensional. However, we should mention that our framework could also be combined with other DC solvers such as Branch-and-Bound schemes [8, Sec 5.1, Sec 5.2] and DCA [15].

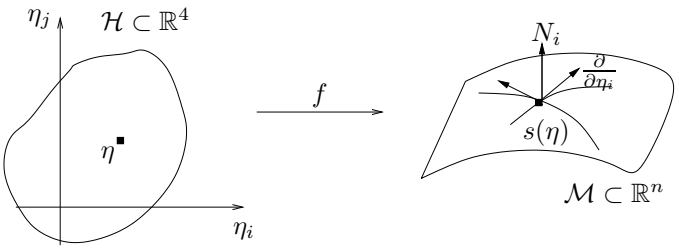

Fig. 7. The parameter space $\mathcal{H}$ provides a parametrization of the transformation manifold $\mathcal{M}$.

V. Geometric analysis of transformation manifolds

We have seen in Section II that the condition number of the manifold $\mathcal{M}$ is an important factor towards characterizing the number m of random measurements needed. At the same time, it is also known that the condition number is closely related to classical notions of curvature in differential geometry via the second fundamental form. In particular, P. Niyogi et al in [10, Proposition 6.1] show that the condition number $1/\tau$ is an upper bound of the principal curvature (defined below) at any point on the manifold.

In this section, we first derive an upper bound of the principal curvature of parametric transformation manifolds $\mathcal{M}$ defined in (9). This upper bound can be used instead of the condition number for characterizing the manifold. It further indicates that the transformation manifold is well conditioned, which means that the number of measurements required to solve the Problem (1) is clearly bounded. Furthermore, based on the obtained developments, we additionally provide an efficient numerical algorithm for computing in practice the principal curvature at a certain point on $\mathcal{M}$.

For notational convenience, we will denote the transformation parameters as follows: $\eta = (\eta_1, \eta_2, \eta_3, \eta_4) = (b_x, b_y, \alpha, \omega)$. The metric tensor $G \in \mathbb{R}^{4\times 4}$ is then given by

$$\big[G\big]_{ij} = \langle t_i, t_j\rangle, \tag{16}$$

where t_i, t_j are the ith and jth tangent vectors, defined as

$$t_i = \frac{\partial s(\eta)}{\partial \eta_i}$$

and assumed to be linearly independent. The tangent space $T_\eta\mathcal{M}$ at point $s(\eta) \in \mathcal{M}$ is defined as

$$T_\eta\mathcal{M} = \operatorname{span}\{t_1, t_2, t_3, t_4\}.$$

Note that $d = \dim T_\eta\mathcal{M} = 4$ although the transformation manifold $\mathcal{M}$ is a submanifold of $\mathbb{R}^n$. The codimension of $T_\eta\mathcal{M}$ is therefore given by $n - 4$. Consider the direct sum

$$\mathbb{R}^n = T_\eta\mathcal{M} \oplus T_\eta\mathcal{M}^\perp$$

and let $\{N_1, N_2, \ldots, N_{n-4}\}$ be an orthonormal basis of $T_\eta\mathcal{M}^\perp$. Then any (unit) normal vector can be written as

$$N = \sum_{i=1}^{n-4} \zeta^i N_i, \tag{17}$$

with coefficients $\zeta^i = \langle N_i, N\rangle$. Figure 7 provides a graphical illustration of the scenario that we consider.

In what follows, we show how one can compute the linear operator $L_\zeta : T_\eta\mathcal{M} \to T_\eta\mathcal{M}$ associated with the second fundamental form. According to the standard definition [16, Proposition 2.3],

$$L_\zeta(X) = -(\nabla_X N)^T, \tag{18}$$

where $X \in T_\eta\mathcal{M}$, ∇_X denotes the covariant derivative in $\mathbb{R}^n$ and $(\ \cdot\ )^T$ denotes the projection on the tangent space. Take a tangent vector $X = \sum_{j=1}^4 x^j t_j$. Then it holds that

$$-\nabla_X N = -\nabla_{\sum_{j=1}^4 x^j t_j} N = \sum_{j=1}^4 x^j(-\nabla_{t_j} N). \tag{19}$$

Hence, it suffices to study the operator $\nabla_{t_j} N$. It holds that

$$\begin{aligned} -\nabla_{t_j} N &= -\nabla_{t_j}\left(\sum_{i=1}^{n-4}\zeta^i N_i\right) \\ &= -\sum_{i=1}^{n-4}\nabla_{t_j}\zeta^i N_i - \sum_{i=1}^{n-4}\zeta^i\nabla_{t_j} N_i. \end{aligned}$$

Now observe that

$$(-\nabla_{t_j} N)^T = \sum_{i=1}^{n-4}\zeta^i(-\nabla_{t_j} N_i)^T. \tag{20}$$

The covariant derivative $\nabla_{t_j} N_i \in \mathbb{R}^n$ can be decomposed as

$$-\nabla_{t_j} N_i = \frac{\partial N_i}{\partial \eta_j} = \sum_{k=1}^4 L_{ij}^k t_k + \sum_{k=1}^{n-4} P_{ij}^k N_k \tag{21}$$

for some coefficients L_{ij}^k with $i = 1,\dots,n-4$ and $j,k = 1,\dots,4$. This directly gives

$$(-\nabla_{t_j} N_i)^T = \sum_{k=1}^4 L_{ij}^k t_k. \tag{22}$$

In what follows, we show how one can compute the coefficients L_{ij}^k in (21), from which we obtain

$$-\langle\frac{\partial N_i}{\partial \eta_j}, t_l\rangle = \sum_{k=1}^4 L_{ij}^k g_{kl}.$$

Unfortunately, it is not easy to compute $\frac{\partial N_i}{\partial \eta_j}$ in practice, as the normal vectors $N_i,\ i = 1,\dots,n-4$ are typically obtained by a Gram-Schmidt orthogonalization process. However, it is known [17] (the proof is provided in the appendix for the sake of completeness) that

$$\langle\frac{\partial N_i}{\partial \eta_j}, t_l\rangle = -\langle N_i, t_{lj}\rangle,$$

where t_{lj} is the mixed partial derivative, i.e.,

$$t_{lj} = \frac{\partial^2 s(\eta)}{\partial\eta_l\partial\eta_j}.$$

This has the advantage that t_{lj} is much easier to compute in practice than $\frac{\partial N_i}{\partial \eta_j}$. Therefore, for fixed i,j, the coefficients $L_{ij}^k,\ k = 1,\dots,4$, can be obtained by solving a 4×4 linear system

$$G\begin{bmatrix} L_{ij}^1 \\ \vdots \\ L_{ij}^4 \end{bmatrix} = \begin{bmatrix} \langle N_i, t_{j1}\rangle \\ \vdots \\ \langle N_i, t_{j4}\rangle \end{bmatrix},$$

where G is the 4×4 metric tensor defined in (16). In more compact form:

$$L_{ij}^k = \sum_{l=1}^4 g^{kl}\langle N_i, t_{jl}\rangle, \tag{23}$$

where $g^{kl} = \left[G^{-1}\right]_{kl}$ denotes the (k,l) entry of the inverse of the metric tensor.

Combining equations (20) and (22) yields

$$\begin{aligned} (-\nabla_{t_j} N)^T &= \sum_{i=1}^{n-4}\zeta^i\left(\sum_{k=1}^4 L_{ij}^k t_k\right) \\ &= \sum_{k=1}^4\underbrace{\left(\sum_{i=1}^{n-4}\zeta^i L_{ij}^k\right)}_{=:\tilde{L}_j^k} t_k = \sum_{k=1}^4 \tilde{L}_j^k t_k. \end{aligned} \tag{24}$$

Using (24), Equation (19) becomes

$$\begin{aligned} -(\nabla_X N)^T &= \sum_{j=1}^4 x^j(-\nabla_{t_j} N)^T = \sum_{j=1}^4 x^j\sum_{k=1}^4\tilde{L}_j^k t_k \\ &= \sum_{k=1}^4\left(\sum_{j=1}^4\tilde{L}_j^k x^j\right)t_k. \end{aligned} \tag{25}$$

The above equation implies that $Y = -(\nabla_X N)^T =: \sum_{k=1}^4 y^k t_k$ with components

$$y^k = \sum_{j=1}^4\tilde{L}_j^k x^j.$$

Therefore, the linear operator L_ζ has the following matrix representation

$$L_\zeta = \begin{bmatrix} \tilde{L}_1^1 & \tilde{L}_2^1 & \tilde{L}_3^1 & \tilde{L}_4^1 \\ \tilde{L}_1^2 & \tilde{L}_2^2 & \tilde{L}_3^2 & \tilde{L}_4^2 \\ \tilde{L}_1^3 & \tilde{L}_2^3 & \tilde{L}_3^3 & \tilde{L}_4^3 \\ \tilde{L}_1^4 & \tilde{L}_2^4 & \tilde{L}_3^4 & \tilde{L}_4^4 \end{bmatrix}. \tag{26}$$

It is important to mention at this point that the operator L_ζ is self-adjoint with respect to the induced metric in the tangent space [16] and therefore its eigenvalues are real. The maximum eigenvalue of L_ζ is usually called the *principal curvature*.

In what follows, we provide an upper bound on the principal curvature.

***Proposition** 3:*

$$\lambda_{\max}(L_\zeta) \le 4\frac{\sup_{l,j}\|t_{lj}\|}{\sigma_{\min}(G)}. \tag{27}$$

Proof: It is well known that

$$\lambda_{\max}(L_\zeta) \le \|L_\zeta\|_2. \tag{28}$$

Algorithm 1 Numerical estimation of the principal curvature

1: **Input**: normal direction $\zeta^1, \ldots, \zeta^{n-4}$
2: **Output**: linear operator L_ζ, estimate λ_ζ of the principal curvature.
3: Compute the tangent vectors $t_i = \frac{\partial s}{\partial \eta_i},\ i = 1, \ldots, 4$.
4: Compute the mixed second order partial derivatives

$$t_{ij} = \frac{\partial^2 s}{\partial \eta_i \partial \eta_j},\ i, j = 1, \ldots, 4.$$

5: Compute the metric tensor $\big[G\big]_{ij} = \langle t_i, t_j \rangle$ as well as its inverse g^{ij}.
6: Build an orthonormal basis $\{N_1, N_2, \ldots, N_{n-4}\}$ of $T_\eta \mathcal{M}^\perp$ using Gram-Schmidt orthogonalization
7: Compute $N = \sum_{i=1}^{n-4} \zeta^i N_i$.
8: **for** $j = 1, \ldots, 4$ **do**
9: **for** $k = 1, \ldots, 4$ **do**
10: Compute $\tilde{L}_j^k = \sum_{l=1}^{4} g^{kl} \langle N, t_{jl} \rangle$.
11: **end for**
12: **end for**
13: Set $L_\zeta = [\tilde{L}_j^k]_{j,k=1,\ldots,4}$.
14: Compute the maximum eigenvalue λ_ζ of L_ζ.

In light of (23), the entries $\tilde{L}_j^k = \sum_{i=1}^{m} \zeta^i L_{ij}^k$ of L_ζ satisfy

$$\begin{aligned} \tilde{L}_j^k &= \sum_{i=1}^{m} \zeta^i \sum_{l=1}^{4} g^{kl} \langle N_i, t_{jl} \rangle \\ &= \sum_{l=1}^{4} g^{kl} \sum_{i=1}^{m} \zeta^i \langle N_i, t_{jl} \rangle \\ &= \sum_{l=1}^{4} g^{kl} \langle N, t_{jl} \rangle, \end{aligned} \tag{29}$$

where we have used (17). Hence, $L_\zeta = G^{-1}\mathcal{N}$ with

$$\mathcal{N} = \begin{bmatrix} \langle N, t_{11} \rangle & \cdots & \langle N, t_{41} \rangle \\ \vdots & & \vdots \\ \langle N, t_{14} \rangle & \cdots & \langle N, t_{44} \rangle \end{bmatrix}.$$

This implies

$$\begin{aligned} \|L_\zeta\|_2 &\leq \|G^{-1}\|_2 \|\mathcal{N}\|_F \leq \frac{4}{\sigma_{\min}(G)} \sup_{l,j} |\langle N, t_{lj} \rangle| \\ &\leq \frac{4}{\sigma_{\min}(G)} \sup_{l,j} \|t_{lj}\|, \end{aligned}$$

using the fact that N is a unit vector. ∎

Observe that $\|t_{jl}\|$ is finite, since the mother function ϕ is in C^∞. Also, G is full rank, which implies that $\sigma_{\min}(G) > 0$. Therefore, the upper bound (27) is finite. This bound on the principal curvature can be used instead of the condition number for analyzing dimensionality reduction of manifolds with random measurements. This implies that the transformation manifolds (9), which we consider in this work, cannot be too much curved and are generally expected to be well behaved. This is also verified in practice as we will show in the experiments section below.

Based on the developments above, Algorithm 1 provides an efficient numerical procedure for computing the principal curvature at a point $s(\eta)$ on the manifold along a certain normal direction ζ. The algorithm makes use of the compact equation (29) for computing the entries of L_ζ. It is important to stress that the tangent vectors t_i as well as the mixed partial derivatives t_{ij} can be computed analytically (i.e., without any approximation) thanks to the closed-form expression of the manifold equation; see (11). Hence, one completely avoids the drawbacks from a finite difference approximation, such as noise sensitivity. The details of the computation are given in the Appendix.

VI. Experimental Results

A. Alignment with random measurements

a) Face manifold: The pattern s is the facial image shown in Fig. 4. We build a pattern model of s using OMP with $K = 40$ Gaussian atoms. Observe that already a few atoms are sufficient to capture the main geometric structure of the pattern. In our experiments we consider η to be a synthesis of an isotropic scaling $\alpha \in [0.5, 1.5]$ and rotation $\omega \in [0, 2\pi)$.

We run 40 random experiments with random transformations η and different random realizations of the measurement matrix Z. Each query image p is built using (11) by applying the exact geometric transformation η^* to s. Then, for each random experiment, 150 iterations of the cutting plane method are employed to align s with p. We compute the alignment error of the estimated transformation $\hat{\eta}$ as follows

$$e = \min\{2\pi - |\hat{\omega} - \omega^*|, |\hat{\omega} - \omega^*|\} + |\hat{\alpha} - \alpha^*|. \tag{30}$$

The entries of the measurement matrices Z follow a standard Gaussian distribution $\mathcal{N}(0, 1)$.

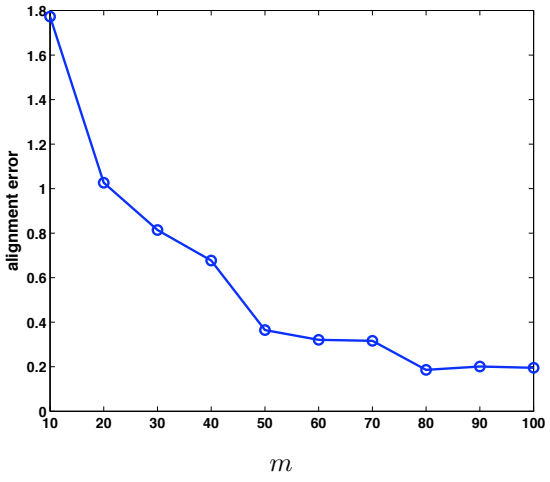

Fig. 8. Median of the alignment error (30) for the face manifold as a function of the number of measurements.

Figure 8 shows the statistics of the alignment error e defined in (30) with respect to the number of random projections m used for image alignment. For each distinct value of m we run 40 random experiments and report the median value of the alignment error. Observe that the alignment error drops quickly when the number of random measurements increases and then it saturates. Furthermore, the experimental results show that for

Fig. 9. Progressive approximation of an MRI brain image (leftmost) with the number of atoms ranging from 20 to 140 with step 20.

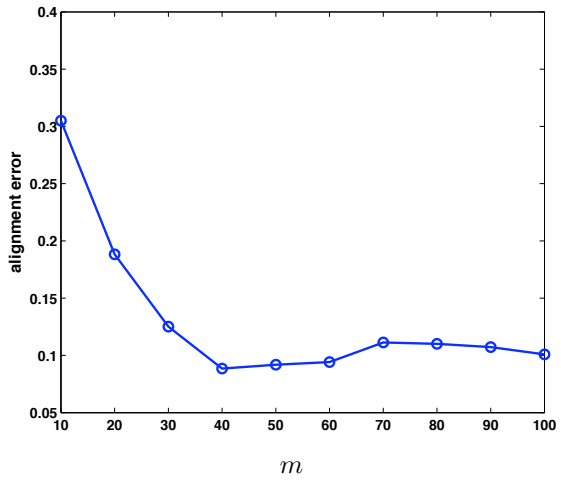

Fig. 10. Median of the alignment error (30) for the MRI brain image manifold as a function of the number of measurements.

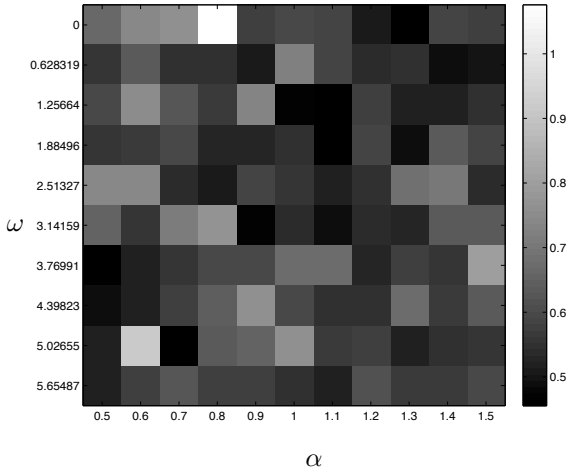

Fig. 11. Principal curvature of the face manifold.

this manifold about 50 measurements are sufficient to enable the cutting plane method to reach the vicinity of the exact transformation in the vast majority of cases.

b) MRI Brain image manifold: We consider now a typical MRI brain image, which is shown in Fig. (9) (leftmost panel). The figure also shows the progressive OMP approximation of the image with increasing number of Gaussian atoms. Notice once more that a few atoms are sufficient to capture the main geometric features of the brain image, although brain images have in general quite different characteristics than facial images.

The pattern s is now the brain image. We repeat the alignment experiment that was conducted on the facial image above. Similarly, we use $K = 40$ Gaussian atoms and perform 40 random experiments with random transformations and random realizations of the measurement matrix Z. Fig. (10) shows the statistics of the alignment error e with respect to the number of random projections m. Observe that the alignment error drops quickly and saturates after 40 random measurements.

To summarize, for both manifolds a few random measurements are in general sufficient to reach the vicinity of the global minimizer with the cutting plane method.

B. Numerical estimation of the principal curvature

The purpose of this section is to estimate numerically the principal curvature of the two manifolds used in the above experiments, using the numerical algorithm presented in Section V. We uniformly discretize the transformation parameter space $[0, 2\pi) \times [0.5, 1.5]$ using 10 rotation angles and 11 scaling levels respectively. Then, for each grid point (ω, α)

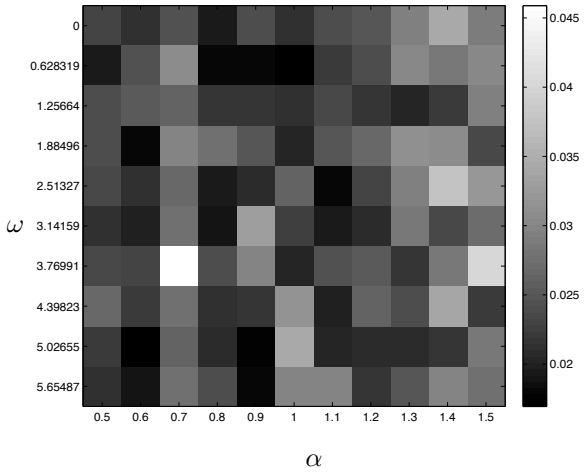

Fig. 12. Principal curvature of the MRI brain image manifold.

we estimate the principal curve using 40 random realizations of the normal direction ζ.

Figures 11 and 12 show the obtained values for each grid point of the face and brain image manifolds respectively. The maximum over all these values is 1.08 for the facial manifold and 0.046 for the brain image manifold (see also the difference in the numerical values of the colorbar on the right of each figure). This implies that the brain image manifold is less curved and therefore better conditioned than the face manifold.

Our numerical experience also indicates that the bound (2) does not in general allow for an accurate prediction of the required measurements in practice, mostly due to its asymptotic

nature. Hence, in principle, this does not permit to reliably draw the conclusion that the difference in the numerical values of the principal curvatures implies the difference in the behavior of the alignment experiments for the two manifolds. Note that empirical algorithms are typically used to determine m directly (see, e.g., [3]).

VII. Related work

In what follows, we review the most relevant work from the literature and put it in perspective with the work proposed in this paper.

We first mention the approaches in [18] and [19] for image alignment based on random projections. In particular, the approach in [19] is used in the context of compressive classification, where the images are considered to be geometrically transformed and the problem is to perform invariant classification based on their random projections. Although the two approaches in [18] and [19] may look different at a first glance, they are both based on exhaustive search, by discretizing the transformation parameter space and estimating the transformation by nearest neighbor search among the images after random projection. This not only leads to high memory requirements, but it also provides no theoretical guarantee for the optimality of the attained solution. In contrast, our DC approach enjoys global optimality guarantees.

An upper bound somewhat similar to (27) has been derived by L. Jacques and C. De Vleeschouwer in [20], where it is used towards alleviating the dictionary discretization effects in Matching Pursuit algorithms. Note, however, that the bound in [20] assumes a different definition of curvature and only holds for the case when the manifold is simply a parametric dictionary of the form (4).

The authors in [3] propose a linear dimensionality reduction methodology based on random projections. They show that a few random projections are sufficient to estimate the intrinsic dimension of the manifold. They also provide an empirical procedure for estimating the number of necessary random projections in the context of ISOMAP. However, their procedure is particularly designed for the ISOMAP algorithm and does not easily extend to other manifold learning algorithms, or more generally, to other image analysis problems.

Finally, the authors in [21] study the non-differentiability of manifolds spanned by natural images. They show that non-differentiability arises due to the movement of sharp edges (causing global non-differentiability) and due to the occlusion (causing local non-differentiability). The manifolds considered in this work are differentiable, thanks to their parametric nature and the smoothness of the mother function.

VIII. Conclusions

We have proposed a globally optimal method for image alignment with random projections. We build on previous work and use sparse geometric expansions to represent the transformation manifold, which describes the transformed versions of a pattern. We formulate the image alignment problem with random projections as a DC program, by proving that the objective function is DC. In addition, we provide theoretical as well as numerical insights on the geometric properties of transformation manifolds, by deriving an upper bound on the principal curvature as well as establishing an efficient numerical algorithm for computing it in practice. We show that the transformation manifolds are well conditioned, so that the image alignment problem can be solved with a bounded number of measurements. This is confirmed by experimental results where the proposed method is shown to be successful in finding the global minimizer in practice, even with a small number of random projections.

IX. Acknowledgments

The authors would like to thank Dr. Xavier Claeys for insightful discussions on differential geometry and Elif Vural for very fruitful discussions on the alignment problem and helpful comments on the manuscript.

X. Appendix

A. Proof of $\langle \frac{\partial N_i}{\partial \eta_j}, t_l \rangle = -\langle N_i, t_{lj} \rangle$

Proof: Observe first that $\langle N_i, \frac{\partial s}{\partial \eta_l} \rangle = 0$, by definition of the normal vector. Therefore,

$$\begin{aligned} 0 &= \frac{\partial}{\partial \eta_j}\langle N_i, \frac{\partial s}{\partial \eta_l}\rangle = \langle \frac{\partial N_i}{\partial \eta_j}, \frac{\partial s}{\partial \eta_l}\rangle + \langle N_i, \frac{\partial^2 s}{\partial \eta_l \partial \eta_j}\rangle \\ &= \langle \frac{\partial N_i}{\partial \eta_j}, t_l\rangle + \langle N_i, t_{lj}\rangle, \end{aligned}$$

which implies the result. ∎

B. Computation of t_j

By definition $s = \sum_{k=1}^{K} \xi_k \phi_{\gamma_k}$, and therefore

$$\frac{\partial s}{\partial \eta_i} = \sum_{k=1}^{K} \xi_k \frac{\partial \phi_{\eta \circ \gamma_k}}{\partial \eta_i}.$$

Recall that $\phi_{\eta\circ\gamma_k} = \phi(\tilde{x}, \tilde{y})$, where $\tilde{x}, \tilde{y}$ are the transformed coordinates, and therefore

$$\frac{\partial \phi}{\partial \eta_i} = \frac{\partial \phi}{\partial \tilde{x}}\frac{\partial \tilde{x}}{\partial \eta_i} + \frac{\partial \phi}{\partial \tilde{y}}\frac{\partial \tilde{y}}{\partial \eta_i}.$$

In the case that ϕ is Gaussian,

$$\frac{\partial \phi}{\partial \tilde{x}} = -2\tilde{x}\exp(-(\tilde{x}^2+\tilde{y}^2)),\ \frac{\partial \phi}{\partial \tilde{y}} = -2\tilde{y}\exp(-(\tilde{x}^2+\tilde{y}^2)).$$

C. Computation of t_{ij}

We have

$$\begin{aligned} \frac{\partial^2 \phi}{\partial \eta_i \partial \eta_j} &= \left[\frac{\partial^2 \phi}{\partial \tilde{x}^2}\frac{\partial \tilde{x}}{\partial \eta_j} + \frac{\partial^2 \phi}{\partial \tilde{x}\partial \tilde{y}}\frac{\partial \tilde{y}}{\partial \eta_j}\right]\frac{\partial \tilde{x}}{\partial \eta_i} + \frac{\partial \phi}{\partial \tilde{x}}\frac{\partial^2 \tilde{x}}{\partial \eta_i \partial \eta_j} + \\ &\quad \left[\frac{\partial^2 \phi}{\partial \tilde{y}\partial \tilde{x}}\frac{\partial \tilde{x}}{\partial \eta_j} + \frac{\partial^2 \phi}{\partial \tilde{y}^2}\frac{\partial \tilde{y}}{\partial \eta_j}\right]\frac{\partial \tilde{y}}{\partial \eta_i} + \frac{\partial \phi}{\partial \tilde{y}}\frac{\partial^2 \tilde{y}}{\partial \eta_i \partial \eta_j}. \end{aligned}$$

In the case that ϕ is Gaussian,

$$\begin{aligned} \frac{\partial^2 \phi}{\partial \tilde{x}^2} &= (4\tilde{x}^2 - 2)\exp(-(\tilde{x}^2+\tilde{y}^2)), \\ \frac{\partial^2 \phi}{\partial \tilde{y}^2} &= (4\tilde{y}^2 - 2)\exp(-(\tilde{x}^2+\tilde{y}^2)), \\ \frac{\partial^2 \phi}{\partial \tilde{x}\partial \tilde{y}} &= \frac{\partial^2 \phi}{\partial \tilde{y}\partial \tilde{x}} = (4\tilde{x}\tilde{y})\exp(-(\tilde{x}^2+\tilde{y}^2)). \end{aligned}$$

D. Computation of $\frac{\partial \tilde{x}}{\partial \eta_i}$ and $\frac{\partial y}{\partial \eta_i}$

Recall that $\eta = (b, \alpha, \omega)$ denotes transformation parameters while $\gamma_i = (b_i, a_i, \omega_i)$ denotes atom parameters. The group law is given by

$$\eta \circ \gamma = (b + \alpha R(-\omega) b_i, \alpha a_i, \omega + \omega_i). \tag{31}$$

Denoting

$$A = \begin{bmatrix} 1/a_{ix} & 0 \\ 0 & 1/a_{iy} \end{bmatrix}, \; R(\omega_i) = \begin{bmatrix} \cos(\omega_i) & \sin(\omega_i) \\ -\sin(\omega_i) & \cos(\omega_i) \end{bmatrix},$$

we define the matrix $C := AR(\omega_i)$, which does not depend on the transformation parameters. Then the transformed coordinates of the ith atom are given by

$$\begin{aligned} \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} &= C\frac{1}{\alpha}R(\omega)\left[\begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} b_x \\ b_y \end{pmatrix} - \alpha R(-\omega)\begin{pmatrix} b_{ix} \\ b_{iy} \end{pmatrix}\right] \\ &= C\left[\frac{1}{\alpha}R(\omega)\begin{pmatrix} x - b_x \\ y - b_y \end{pmatrix} - \begin{pmatrix} b_{ix} \\ b_{iy} \end{pmatrix}\right] \\ &= C\begin{bmatrix} \frac{1}{a}(\cos(\omega)(x-b_x) + \sin(\omega)(y-b_y)) - b_{ix} \\ \frac{1}{a}(-\sin(\omega)(x-b_x) + \cos(\omega)(y-b_y)) - b_{iy} \end{bmatrix}. \end{aligned}$$

Differentiation gives

$$\begin{aligned} \begin{bmatrix} \partial\tilde{x}/\partial\omega \\ \partial\tilde{y}/\partial\omega \end{bmatrix} &= C\begin{bmatrix} \frac{1}{a}(-\sin(\omega)(x-b_x) + \cos(\omega)(y-b_y)) \\ \frac{1}{a}(-\cos(\omega)(x-b_x) - \sin(\omega)(y-b_y)) \end{bmatrix}, \\ \begin{bmatrix} \partial\tilde{x}/\partial b_x \\ \partial\tilde{y}/\partial b_x \end{bmatrix} &= C\begin{bmatrix} -\frac{1}{a}\cos(\omega) \\ \frac{1}{a}\sin(\omega) \end{bmatrix}, \\ \begin{bmatrix} \partial\tilde{x}/\partial b_y \\ \partial\tilde{y}/\partial b_y \end{bmatrix} &= C\begin{bmatrix} -\frac{1}{a}\sin(\omega) \\ -\frac{1}{a}\cos(\omega) \end{bmatrix}, \\ \begin{bmatrix} \partial\tilde{x}/\partial\alpha \\ \partial\tilde{y}/\partial\alpha \end{bmatrix} &= C\begin{bmatrix} -\frac{1}{a^2}(-\cos(\omega)(x-b_x) + \sin(\omega)(y-b_y)) \\ -\frac{1}{a^2}(-\sin(\omega)(x-b_x) + \cos(\omega)(y-b_y)) \end{bmatrix}. \end{aligned}$$

E. Computation of $\frac{\partial^2 x}{\partial \eta_i \partial \eta_j}$

$$\begin{aligned} \begin{bmatrix} \partial^2\tilde{x}/\partial\omega^2 \\ \partial^2\tilde{y}/\partial\omega^2 \end{bmatrix} &= C\begin{bmatrix} \frac{1}{a}(-\cos(\omega)(x-b_x) - \sin(\omega)(y-b_y)) \\ \frac{1}{a}(\sin(\omega)(x-b_x) - \cos(\omega)(y-b_y)) \end{bmatrix}, \\ \begin{bmatrix} \partial^2\tilde{x}/\partial\omega\partial b_x \\ \partial^2\tilde{y}/\partial\omega\partial b_x \end{bmatrix} &= C\begin{bmatrix} \frac{1}{a}\sin(\omega) \\ \frac{1}{a}\cos(\omega) \end{bmatrix}, \\ \begin{bmatrix} \partial^2\tilde{x}/\partial\omega\partial b_y \\ \partial^2\tilde{y}/\partial\omega\partial b_y \end{bmatrix} &= C\begin{bmatrix} -\frac{1}{a}\cos(\omega) \\ \frac{1}{a}\sin(\omega) \end{bmatrix}, \\ \begin{bmatrix} \partial^2\tilde{x}/\partial\omega\partial\alpha \\ \partial^2\tilde{y}/\partial\omega\partial\alpha \end{bmatrix} &= C\begin{bmatrix} -\frac{1}{a^2}(-\sin(\omega)(x-b_x) + \cos(\omega)(y-b_y)) \\ -\frac{1}{a^2}(-\cos(\omega)(x-b_x) - \sin(\omega)(y-b_y)) \end{bmatrix}, \\ \begin{bmatrix} \partial^2\tilde{x}/\partial b_x^2 \\ \partial^2\tilde{y}/\partial b_x^2 \end{bmatrix} &= 0, \\ \begin{bmatrix} \partial^2\tilde{x}/\partial b_x\partial b_y \\ \partial^2\tilde{y}/\partial b_x\partial b_y \end{bmatrix} &= 0, \\ \begin{bmatrix} \partial^2\tilde{x}/\partial b_x\partial\alpha \\ \partial^2\tilde{y}/\partial b_x\partial\alpha \end{bmatrix} &= C\begin{bmatrix} \frac{1}{a^2}\cos(\omega) \\ -\frac{1}{a}\sin(\omega) \end{bmatrix}, \\ \begin{bmatrix} \partial^2\tilde{x}/\partial b_y^2 \\ \partial^2\tilde{y}/\partial b_y^2 \end{bmatrix} &= 0, \\ \begin{bmatrix} \partial^2\tilde{x}/\partial b_y\partial\alpha \\ \partial^2\tilde{y}/\partial b_y\partial\alpha \end{bmatrix} &= C\begin{bmatrix} \frac{1}{a^2}\sin(\omega) \\ \frac{1}{a^2}\cos(\omega) \end{bmatrix}, \\ \begin{bmatrix} \partial^2\tilde{x}/\partial\alpha^2 \\ \partial^2\tilde{y}/\partial\alpha^2 \end{bmatrix} &= C\begin{bmatrix} \frac{2}{a^3}(\cos(\omega)(x-b_x) + \sin(\omega)(y-b_y)) \\ \frac{2}{a^3}(-\sin(\omega)(x-b_x) + \cos(\omega)(y-b_y)) \end{bmatrix}. \end{aligned}$$

REFERENCES

[1] B. Zitová and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, 2003.

[2] R. G. Baraniuk and M. B. Wakin. Random projections of smooth manifolds. *Foundations of Computational Mathematics*, 9(1):51–77, February 2009.

[3] C. Hegde, M. B. Wakin, and R. G. Baraniuk. Random projections for manifold learning. *Neural Information Processing Systems (NIPS)*, 2007.

[4] M. B. Wakin and R. G. Baraniuk. Random projections of signal manifolds. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, V:941–944, 2006.

[5] E. Kokiopoulou, D. Kressner, and P. Frossard. Optimal image alignment with random measurements. *17th European Signal Processing Conference (EUSIPCO)*, 2009.

[6] E. Kokiopoulou and P. Frossard. Minimum distance between pattern transformation manifolds: Algorithm and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(7):1225–1238, June 2009.

[7] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 1999.

[8] R. Horst and N. V. Thoai. DC Programming: Overview. *Journal of Optimization Theory and Applications*, 102(1):1–43, October 1999.

[9] S. Dasgupta and A. Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. *Random structures and algorithms*, 22(1):60–65, 2003.

[10] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete and Computational Geometry*, 39(1-3):419–441, March 2008.

[11] S. Mallat. *A Wavelet Tour of Signal Processing, 2nd edn.* Academic Press, 1998.

[12] R. Figueras i Ventura, P. Vandergheynst, and P. Frossard. Low rate and flexible image coding with redundant representations. *IEEE Transactions on Image Processing*, 15(3):726–739, March 2006.

[13] R. Horst, P. M. Pardalos, and N. V. Thoai. *Introduction to Global Optimization*, volume 48 of *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers, 2nd edition, 2000.

[14] R. Horst and P. M. Pardalos. *Handbook of Global Optimization*, volume 2 of *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers, 1995.

[15] L. Thi Hoai An and P. Dinh Tao. The DC (Difference of Convex Functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. *Annals of Operations Research*, 133:23–46, 2005.

[16] M. P. Do Carmo. *Riemannian Geometry*. Birkhäuser, 1992.

[17] S. Lang. *Fundamentals of Differential Geometry*. Springer, 1999.

[18] D. M. Healy and G. K. Rohde. Fast global image registration using random projections. *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 476–479, 2007.

[19] M. A. Davenport, M. F. Duarte, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly, and R. G. Baraniuk. The smashed filter for compressive classification and target recognition. *Proceedings of SPIE*, 6498, 2007.

[20] L. Jacques and C. De Vleeschouwer. A geometrical study of matching pursuit parametrization. *IEEE Transactions on Signal Processing*, 56(7):2835–2848, July 2008.

[21] M. B. Wakin, D. L. Donoho, H. Choi, and R. G. Baraniuk. The multiscale structure of non-differentiable image manifolds. *SPIE Wavelets XI*, July 2005.

Research Reports

No.	Authors/Title
09-41	*E. Kokiopoulou, D. Kressner, N. Paragios, P. Frossard* Optimal image alignment with random projections of manifolds: algorithm and geometric analysis
09-40	*P. Benner, P. Ezzatti, D. Kressner, E.S. Quintana-Ortí, A. Remón* A mixed-precision algorithm for the solution of Lyapunov equations on hybrid CPU-GPU platforms
09-39	*V. Wheatley, P. Huguenot, H. Kumar* On the role of Riemann solvers in discontinuous Galerkin methods for magnetohydrodynamics
09-38	*E. Kokiopoulou, D. Kressner, N. Paragios, P. Frossard* Globally optimal volume registration using DC programming
09-37	*F.G. Fuchs, A.D. McMurray, S. Mishra, N.H. Risebrom, K. Waagan* Approximate Riemann solvers and stable high-order finite volume schemes for multi-dimensional ideal MHD
09-36	*Ph. LeFloch, S. Mishra* Kinetic functions in magnetohydrodynamics with resistivity and hall effects
09-35	*U.S. Fjordholm, S. Mishra* Vorticity preserving finite volume schemes for the shallow water equations
09-34	*S. Mishra, E. Tadmor* Potential based constraint preserving genuinely multi-dimensional schemes for systems of conservation laws
09-33	*S. Mishra, E. Tadmor* Constraint preserving schemes using potential-based fluxes. III. Genuinely multi-dimensional central schemes for for MHD equations
09-32	*S. Mishra, E. Tadmor* Constraint preserving schemes using potential-based fluxes. II. Genuinely multi-dimensional central schemes for systems of conservation laws
09-31	*S. Mishra, E. Tadmor* Constraint preserving schemes using potential-based fluxes. I. Multidimensional transport equations
09-30	*D. Braess, S. Sauter, C. Schwab* On the justification of plate models