Semiparametric estimation of rigid transformations on compact Lie groups

Jérémie Bigot, Jean-Michel Loubes, Myriam Vimond

Semiparametric estimation of rigid transformations on compact Lie groups

Jérémie Bigot[1], Jean-Michel Loubes[1] and Myriam Vimond[2]

[1] Institut de Mathématiques de Toulouse, Université de Toulouse, 31062 Toulouse Cedex 9, France,

[2] Université Haute Bretagne, Equipe de Statistique, Place du Recteur H. Le Moal, CS 24307, 35043 Rennes Cedex, France

Abstract. We study a simple model for the estimation of rigid transformations between noisy images. The transformations are supposed to belong to a compact Lie group, and a new matching criteria based on the Fourier transform is proposed. Consistency and asymptotic normality of the resulting estimators are studied. Some simulations are used to illustrate the methodology, and we describe potential applications of this approach to various image registration problems.

1 Introduction

Originating in Grenander's pattern theory, transformation Lie groups are commonly used to model the deformations of images. The study of the properties and intrinsic geometries of such deformation groups is now an active field of research (see e.g. [3], [10], [12], [13], [14], [18]). In this setting, an important problem is the estimation of the deformations that may exist between similar images in the presence of additive noise. Rigid displacements such as translations, rotations or affine transformations can be modeled by finite dimensional Lie groups. Hence, the estimation of rigid deformation parameters can be formulated as a semi-parametric estimation problem which is an important field of research in mathematical statistics [4]. Indeed, semiparametric modeling is concerned with statistical problems where the parameters of interest are finite dimensional but where their observation is blurred by an infinite dimensional parameter. Here, the finite-dimensional parameters are the Lie group elements, and the infinite-dimensional parameter is an unknown 2D or 3D image, which is warped to obtain the different deformations. This image, that has to be recovered, is often called a template. The semiparametric framework provides optimal recovery of the warping parameters, contrary to nonparametric methods, leading to a better estimation of the template obtained by aligning the observed images.

As Lie groups are typically nonlinear spaces, an important question is the development of information geometry tools to extend classical notions such as asymptotic normality and efficiency, or the Cramer-Rao bound originally proposed for parameters lying in an Euclidean space. In the context of parametric statistics, several generalizations of these concepts to arbitrary manifolds have been proposed (see e.g. [17]). However, in the more general situation of semi-parametric models, there is not so much work dealing with the estimation of parameters lying in a Lie group.

A first attempt in this direction has been proposed in [8], [20] for the simple problem of recovering shifts between one-dimensional noisy curves observed on an interval. The goal of this paper is extend such an approach to the more general case where the shift parameters belong to a compact Lie group. Thanks to a general shift property of the Fourier transform on compact Lie groups, a matching criterion based on the Fourier transform of the data can be defined, and we study some statistical properties of the resulting estimators.

In Section 2, a simple model for shifts on groups is introduced. Some properties on the Fourier transform are briefly reviewed (see [16] for further details), and are then used to define a general matching criterion on compact Lie groups. In Section 3, the consistency and the asymptotic normality of the estimators are established. Some numerical simulations are presented in Section 4, and extensions of the model are given in Section 5 for the registration of spherical images. Finally, we give some perspectives for future work. Note that the proofs of our theorems are based on the theory of M-estimation (see e.g. [19]), but are quite long and technical, and are thus not given in this manuscript but can be found in [2].

2 A shift model on Lie groups and a matching criteria based on the Fourier transform

2.1 The Fourier transform on compact Lie groups

Let G be a compact Lie group. Denote by e the identity element, and by hg the binary operation between two elements $h, g \in G$. Let $\mathbb{L}^2(G)$ be the Hilbert space of complexed valued, square integrable functions on the group G with respect to the Haar measure dg.

To define a Fourier transform on $\mathbb{L}^2(G)$, a fundamental tool is the theory of group representations, which aims at studying the properties of groups via their representations as linear transformation of vector spaces. More precisely, a representation is an homomorphism from the group G to the automorphism group of a vector space. So let V be a finite-dimensional vector space, a *representation* of G in V is thus defined as a continuous homomorphism $\pi : G \to \mathrm{GL}(V)$, where $\mathrm{GL}(V)$ denotes the set of automorphisms of V. Hence it provides a linear transformation which depends on the vector space on which the group acts.

Every irreducible representation π of a compact group G in a vector space V is finite dimensional, so we denote by d_π the dimension of V. By choosing a basis for V, it is often convenient to identify $\pi(g)$ with a matrix of size $d_\pi \times d_\pi$ with complex entries. The function $g \mapsto \mathrm{Tr}\,\pi(g)$ is called the *character* of π. The characters carry the essential information about the group representation. Moreover, the fundamental theorem of *Schur orthogonality* states that the characters form an orthonormal system in $\mathbb{L}^2(G)$ when π ranges over the dual set $\hat{G}$ of irreducible and equivalent representations of G. In the case of compact groups, the dual $\hat{G}$ is a countable set, and the *Peter-Weyl Theorem* states that the characters are dense in $\mathbb{L}^2(G)$. Indeed, if π is a finite dimensional representation of G in the vector space V, then one can define, for every $f \in \mathbb{L}^2(G)$, the linear mapping $\pi(f) : V \to V$ by

$$\pi(f)v = \int_G f(g)\overline{\pi(g)}^T v dg, \text{ for } v \in V.$$

The matrix $\pi(f)$ is the generalization to the case of compact group of the usual notion of Fourier coefficient. Then, Peter-Weyl Theorem implies that (for simplicity the same notation is used for π and its equivalence class in $\hat{G}$)

$$f(g) = \sum_{\pi \in \hat{G}} d_\pi \mathrm{Tr}\ (\pi(g)\pi(f)) \text{ and } \|f\|^2_{\mathbb{L}^2(G)} = \sum_{\pi \in \hat{G}} d_\pi \mathrm{Tr}\ \left(\pi(f)\overline{\pi(f)}^T\right) \qquad (1)$$

In the sequel, we will also denote by $\langle A, B\rangle_{HS} = \mathrm{Tr}\ (\overline{A}^T B)$ the Hilbert-Schmidt inner product between two finite dimensional $d_\pi \times d_\pi$ matrices A and B.

2.2 A simple shift model on groups

Let $\mathcal{X}$ be a subset of $\mathbb{R}^d$ (with $d = 2, 3$ in our applications) and G be a compact Lie group acting on $\mathcal{X}$. A general deformation model for a set of J noisy signals would be

$$Y_j(x) = f^*({g_j}^{-1} \cdot x) + W_j(x) \text{ for } x \in \mathcal{X}, \qquad (2)$$

where $f^* : \mathcal{X} \mapsto \mathbb{R}$ is an unknown template, $W_j(x)$ is some additive noise, and g_j are the deformations to estimate. Our interest is to provide a very general framework for image registration under warping effect given by model (2), and thus to deal with the general problem of estimation of the elements of a Lie Group G acting on a space $\mathcal{X}$. In this paper, a simplest model for which $\mathcal{X} = G$ is studied to give the main ideas of our approach. An example for which $\mathcal{X} \neq G$ is given in Section 5 to show that our approach can be extended to more complex situations, and we discuss possible extensions to more complex situations in the concluding section.

Now, consider the following white noise model : for $j = 1, \ldots, J$ and $g \in G$

$$dY_j(g) = f_j(g)dg + \epsilon dW_j(g), \qquad (3)$$

where $f_j(g) = f^*({h_j^*}^{-1} g)$. The function $f^* : G \to \mathbb{R}$ is the unknown common shape of the observed images Y_j. The parameter $h_j^*, j = 1, \ldots, J$ are the unknown shift parameters that we wish to estimate. $W_j, j = 1, \ldots, J$ are independent standard Brownian sheets on the topological space G with measure dg, ϵ is an unknown noise level parameter which will tend to zero in our asymptotic considerations. Note that the white noise model (3) is a continuous model which is a very useful tool for the theoretical study of statistical problem in image analysis. In practice, the noisy images Y_j are typically discretely sampled on a regular grid, but the model (3) can be shown to lead to comparable asymptotic theory in a sampled data model [5].

Obviously, without any further restriction on the set of possible shifts, the model (3) is not identifiable. Indeed, if s is an element of G with $s \neq e$, then one can replace the h_j^*'s in equation (3) by $\tilde{h}_j = h_j^* s$ and f^* by $\tilde{f}(g) = f^*(sg)$ without changing the formulation of the model. To ensure identification, we further assume that the set of parameters G^J is reduced to the subset $\mathcal{A}_0 \subset \mathcal{A}$ such that

$$\mathcal{A}_0 = \{(h_1, \dots, h_J) \in \mathcal{A}, h_1 = e\}. \tag{4}$$

This choice implies that the first image will be a reference and all the other images will be warped onto this particular choice. Now, remark that since $\overline{\pi(g)}^T = \pi(g^{-1})$, one has that $\pi(f_j) = \pi(f)\pi(h_j^{*-1})$ for all $j = 1, \dots, J$. This equality is classically referred to as the *shift property* of the Fourier transform, and is at the heart of our estimation procedure to exhibit the shift parameters h_j^*.

2.3 A matching criterion based on the Fourier transform

For $h = (h_1, \dots, h_J) \in \mathcal{A}_0$, we propose to minimize the following criterion inspired by recent results in [8] and [20] for the estimation of shifts between curves:

$$M(h_1, \dots, h_J) = \frac{1}{J} \sum_{j=1}^{J} \left\| f_j \circ T_{h_j} - \frac{1}{J} \sum_{j'=1}^{J} f_{j'} \circ T_{h_{j'}} \right\|_{\mathbb{L}^2(G)}^2, \tag{5}$$

where $T_h : g \in G \to hg \in G$. Using the Parseval-Plancherel formula, the criterion may be rewritten in the Fourier domain as:

$$M(h) = M(h_1, \dots, h_J) = \frac{1}{J} \sum_{j=1}^{J} \sum_{\pi \in \hat{G}} d_\pi \left\| \pi(f_j)\pi(h_j) - \frac{1}{J} \sum_{j'=1}^{J} \pi(f_{j'})\pi(h_{j'}) \right\|_{HS}^2, \tag{6}$$

for $h = (h_1, \dots, h_J) \in \mathcal{A}_0$. Given that $\pi(f_j) = \pi(f^*)\pi(h_j^{*-1})$, the criterion M has a minimum at $h^* = (h_1^*, \dots, h_J^*)$ such that $M(h^*) = 0$.

2.4 The empirical criterion

Our estimation method is then based on the Fourier Transform of the noisy data given by model (3). Let π an irreducible representation of G into V. We consider the following linear mappings from V to V which are defined from the model (3):

$$\pi(Y_j) = \int_G \pi(g^{-1}) dY_j(g) = \pi(f_j) + \epsilon \pi(W_j), \quad j = 1 \dots J,$$

where $\pi(W_j) = \int_G \pi(g^{-1}) dW_j(g), j = 1 \dots J$. Let us denote by $(\pi_{kl}(W_j))$ the matrix coefficients of $\pi(W_j) : \pi_{kl}(W_j) = \int_G \pi_{kl}(g^{-1}) dW_j(g)$. Using the Schur orthogonality and the fact that W_j is a standard Brownian sheet on G, one obtains that the complex variables $\pi_{kl}(W_j)$ are independent identically distributed Gaussian variables $\mathcal{N}_{\mathbb{C}}(0, d_\pi^{-1})$.

Let $\hat{G}_\epsilon$ be a finite subset of $\hat{G}$ such that the sequence $\hat{G}_\epsilon$ increases when ϵ tends to 0 and $\cup_{\epsilon>0}\hat{G}_\epsilon = \hat{G}$. Practical choices for the set $\hat{G}_\epsilon$ will be discussed later on for the case of Abelian groups and the non-commutative group SO(3). Then, we consider the following criterion:

$$M_\epsilon(h_1, \ldots, h_J) = \frac{1}{J}\sum_{j=1}^{J}\sum_{\pi\in\hat{G}_\epsilon} d_\pi \left\| \pi(Y_j)\pi(h_j) - \frac{1}{J}\sum_{j'=1}^{J}\pi(Y_{j'})\pi(h_{j'}), \right\|_{HS}^2 \tag{7}$$

and our M-estimator is given by $\hat{h}_\epsilon = \arg\min_{h\in\mathcal{A}_0} M_\epsilon(h)$.

3 Consistency and asymptotic normality of the M-estimator

Theorem 1. *Assume that $f^* \in \mathbb{L}^2(G)$ is such that there does not exist a closed subgroup H (except $H = \{e\}$ or $H = G$) such that $f(gh) = f(g)$ for all $g \in G$ and $h \in H$. Moreover, suppose that for all $\pi \in \hat{G}$ such that $\pi(f^*)$ is not identically null, then $\pi(f^*)$ is invertible. Then, M has a unique minimum at h^*, and if $\lim_{\epsilon\to 0}\epsilon^2\sum_{\pi\in\hat{G}_\epsilon} d_\pi^2 = 0$, then $\hat{h}_\epsilon$ converges in probability to $h^* = (h_1^*, \ldots, h_J^*)$.*

This Theorem shows the consistency of the estimators of the parameters h^* when the noise level goes to zero (recall that in the discretized model, this is equivalent to the fact that the number of observations increase). We stress that this result is obtained by minimizing a quadratic contrast without prior knowledge of the main pattern, thanks to the empirical criterion (7) which enables to get rid of f^*. Asymptotic normality of these estimators and thus rates of convergence will provided in Theorem 2. The first condition of Theorem 1 ensures identifiability of the model since a function which would not satisfy this condition would be shift invariant, preventing any estimation. The second condition can be viewed as a smoothness condition which ensures unicity of the minimization scheme. In the following, we will denote by $\Re(z)$ the real part of any complex number z.

As the estimator $\hat{h}_\epsilon$ takes its values in a Lie group, it is not that obvious to define a notion of asymptotic normality as the space $\mathcal{A}_0$ is typically not a linear space if the group G is not Abelian. To overcome this, a classical approach is to use the exponential map to "project" $\hat{h}_\epsilon$ into the Lie algebra $\mathcal{G}$ of G. If we can write $\hat{h}_\epsilon = \exp(\hat{u}_\epsilon)$, then one can study the asymptotic normality of $\hat{u}_\epsilon$ which belongs to the linear space $\mathcal{G}$ supposed to be of finite dimension p. For this, we first re-express the criterion M_ϵ defined on G^J as a function $\tilde{M}_\epsilon$ defined on $\mathcal{G}^J$. If G is a compact group, then the exponential map is surjective, but it is not necessarily injective. Hence, define $\mathcal{U}$ to be a compact neighborhood of $0 \in \mathcal{G}^J$ onto which the exponential map is a smooth diffeomorphism, and let $\mathcal{A} = \exp(\mathcal{U})$. For $u \in \mathcal{U}^J$, we re-express our criterion as $\tilde{M}(u_1,\ldots,u_J) = M(\exp(u_1),\ldots,\exp(u_J))$, and $\tilde{M}_\epsilon(u_1,\ldots,u_J) = M_\epsilon(\exp(u_1),\ldots,\exp(u_J))$. Then, define

$$\hat{u}_\epsilon = (\hat{u}_1,\ldots,\hat{u}_J) = \arg\min_{u\in\mathcal{U}_1} \tilde{M}_\epsilon(u_1,\ldots,u_J),$$

where $\mathcal{U}_1 = \left\{(u_1,\ldots,u_J)\in\mathcal{U}^J,\ u_1=0\right\}$. Suppose that $h_1^* = \exp(u_1^*),\ldots,h_J^* = \exp(u_J^*)$, with $u^* = (u_1,\ldots,u_J)\in\mathcal{U}^J$, then the following result holds:

Theorem 2. *Assume that the conditions of Theorem 1 hold. Moreover, assume that for all $j = 2,\ldots,J$ and $k = 1,\ldots,p$*

$$\lim_{\epsilon\to 0}\epsilon\sum_{\pi\in\hat{G}_\epsilon} d_\pi^2 \|d_e\pi\left(d_{u_j^*}\exp(x^k)\right)\|_{HS}^2 = 0, \tag{8}$$

and that

$$\lim_{\epsilon\to 0}\sup_{u_1\in\mathcal{U}}\left\{\sum_{\pi\in\hat{G}\setminus\hat{G}_\epsilon} d_\pi \|\pi(f^*)\|_{HS}^2 \left\|d_e\pi\left(d_{u_1}\exp(x^{k_1})\right)\right\|_{HS}^2\right\} = 0 \tag{9}$$

$$\lim_{\epsilon\to 0}\sup_{u_1\in\mathcal{U}}\left\{\sum_{\pi\in\hat{G}\setminus\hat{G}_\epsilon} d_\pi \|\pi(f^*)\|_{HS}^2 \left\|\nabla_{u_1}^{x^{k_2}} d_e\pi\left(d_{u_1}\exp(x^{k_1})\right)\right\|_{HS}\right\} = 0 \tag{10}$$

$$\lim_{\epsilon\to 0}\epsilon^2\sup_{u_1\in\mathcal{U}}\left\{\sum_{\pi\in\hat{G}_\epsilon} d_\pi^2 \left\|d_e\pi\left(d_{u_1}\exp(x^{k_1})\right)\right\|_{HS}^2\right\} = 0 \tag{11}$$

$$\lim_{\epsilon\to 0}\epsilon^2\sup_{u_1\in\mathcal{U}}\left\{\sum_{\pi\in\hat{G}_\epsilon} d_\pi^2 \left\|\nabla_{u_1}^{x^{k_2}} d_e\pi\left(d_{u_1}\exp(x^{k_1})\right)\right\|_{HS}\right\} = 0, \tag{12}$$

where $x^1,\ldots,x^p$ is an arbitrary basis of $\mathcal{G}$. Then,

$$\epsilon^{-1}(\hat{u}_\epsilon - u^*) \to N(0, H^{-1}\Sigma H^{-1}),\ \text{as } \epsilon\to 0,$$

where Σ is a positive definite $(J-1)p\times(J-1)p$ matrix whose entries for $2\le j_1,j_2\le J$ and $1\le k_1,k_2\le p$ are given by

$$\Sigma_{(j_1,k_1)\times(j_1,k_1)} = \frac{4}{J^2}\sum_{\pi\in\hat{G}} d_\pi\left(1-\frac{1}{J}\right)\|\pi(f^*)d_e\pi\left(d_{u_{j_1}^*}\exp(x^{k_1})\right)\|_{HS}^2$$

$$\Sigma_{(j_1,k_1)\times(j_1,k_2)} = \frac{4}{J^2}\sum_{\pi\in\hat{G}} d_\pi\left(1-\frac{1}{J}\right)\Re\left\langle\pi(f^*)d_e\pi\left(d_{u_{j_1}^*}\exp(x^{k_1})\right),\pi(f^*)d_e\pi\left(d_{u_{j_1}^*}\exp(x^{k_2})\right)\right\rangle_{HS},$$

and for $j_1\ne j_2$ by

$$\Sigma_{(j_1,k_1)\times(j_2,k_2)} = -\frac{4}{J^2}\sum_{\pi\in\hat{G}} d_\pi\frac{1}{J}\Re\left\langle\pi(f^*)d_e\pi\left(d_{u_{j_1}^*}\exp(x^{k_1})\right),\pi(f^*)d_e\pi\left(d_{u_{j_2}^*}\exp(x^{k_2})\right)\right\rangle_{HS},$$

and where H is a positive definite $(J-1)p \times (J-1)p$ matrix whose entries for $2 \leq j_1, j_2 \leq J$ and $1 \leq k_1, k_2 \leq p$ are given by

$$H_{(j_1,k_1)\times(j_1,k_2)} = -\frac{2}{J}\sum_{\pi\in\hat{G}} d_\pi \Re\Big\{\Big\langle \pi(f^*)d_e\pi\left(d_{u^*_{j_1}}\exp(x^{k_2})\right)d_e\pi\left(d_{u^*_{j_1}}\exp(x^{k_1})\right) + \pi(f^*)\nabla^{x^{k_2}}_{u^*_{j_1}} d_e\pi\left(d_{u^*_{j_1}}\exp(x^{k_1})\right), \pi(f^*)\Big\rangle_{HS} + \frac{1}{J}\left\langle \pi(f^*)d_e\pi\left(d_{u^*_{j_1}}\exp(x^{k_1})\right), \pi(f^*)d_e\pi\left(d_{u^*_{j_1}}\exp(x^{k_2})\right)\right\rangle_{HS}\Big\},$$

and for $j_1 \neq j_2$ by

$$H_{(j_1,k_1)\times(j_2,k_2)} = -\frac{2}{J^2}\sum_{\pi\in\hat{G}} d_\pi \Re\left\langle \pi(f^*)d_e\pi\left(d_{u^*_{j_1}}\exp(x^{k_1})\right), \pi(f^*)d_e\pi\left(d_{u^*_{j_2}}\exp(x^{k_2})\right)\right\rangle_{HS}.$$

The convergence result in the above theorem must be understood for the vector $\hat{u}_\epsilon = (\hat{u}_2, \ldots, \hat{u}_J) \in \mathcal{G}^{J-1}$ since the first component is fixed to $\hat{u}_1 = 0$. Moreover, the notation $d_e\pi$ stands for the differential of π at e, while $d_{u_1}\exp(x^{k_1})$ corresponds to the differential of the exponential at u_1 in the direction x^{k_1}. We point out that the estimator converges at the parametric rate of convergence, and thus optimal rate of convergence, which would not have been the case if we had considered a preliminar estimate of f^*. This is one of the main achievements of the semiparametric type methodology proposed in this paper. Proving the optimality up to the constants imply studying the semiparametric efficiency of the estimators and falls beyond the scope of this paper. Some intuitions about such a result is provided in Section 6.

3.1 Abelian groups : the special case of the torus

The assumptions of Theorem 2 are rather technical and difficult to state in the very general case. However, for Abelian groups their statement is much simpler, which is due to the fact that the mapping $d\exp_u : \mathcal{G} \to \mathcal{G}$ reduces to the identity on $\mathcal{G}$ i.e. $d_u \exp(v) = v$ for all $u \in G$ and $v \in \mathcal{G}$. The assumptions can be rewritten as

$$\lim_{\epsilon\to 0}\epsilon^2\#\{\hat{G}_\epsilon\} = 0, \quad \lim_{\epsilon\to 0}\epsilon\sum_{\pi\in\hat{G}_\epsilon}|d_e\pi(x^k)|^2 = 0, \quad \lim_{\epsilon\to 0}\sum_{\pi\in\hat{G}\setminus\hat{G}_\epsilon}|\pi(f^*)|^2\left|d_e\pi(x^k)\right|^2 = 0, \tag{13}$$

where $x^1, \ldots, x^p$ is an arbitrary basis of $\mathcal{G}$. Thus, Condition (**??**) states that the common shape is differentiable and its derivatives are square integrable on $\mathcal{G}$. Conditions (13) give some assumptions on the choice of $\hat{G}_\epsilon$.

As an illustrative example, let us consider the case where $G = (\mathbb{R}/\mathbb{Z})^p$ which corresponds to the classical multi-dimensional Fourier decomposition of a function $f \in \mathbb{L}^2([0,1]^p)$

$$f(x) = \sum_{\ell\in\mathbb{Z}^p} c_\ell(f)e_\ell(x), \text{ for } x = (x_1, \ldots, x_p) \in [0,1]^d \text{ and } \ell = x = (\ell_1, \ldots, \ell_p) \in \mathbb{Z}^d,$$

where $e_\ell(x) = \pi(x) = e^{-i2\pi(\sum_{k=1}^{p} \ell_k x_k)}$ and $c_\ell(x) = \pi(f) = \int_{[0,1]^d} f(x)e_\ell(x)dx$. Note that also that $d_e\pi(x^k) = -i2\pi\ell_k$. Now, take $\hat{G}_\epsilon = \{(\ell_1, \ldots, \ell_p) \in \mathbb{Z}^p,\ |\ell_k| \leq \ell_\epsilon$ for all $k = 1, \ldots, p\}$, for some positive integer ℓ_ϵ. Then, the following corollary holds:

Corollary 1. *Let $G = (\mathbb{R}/\mathbb{Z})^p$ and $f^* \in \mathbb{L}^2([0,1]^p)$ be a periodic function satisfying the conditions of Theorem 1. Assume that $h^* \in G^J$ or equivalently that $u^* \in ([0,1]^p)^J$. If*

$$\epsilon\ell_\epsilon^{p+2} = o(1) \text{ and } \sum_{(\ell_1,\ldots,\ell_p)\in\mathbb{Z}^p} \left(|\ell_1|^2 + \ldots + |\ell_p|^2\right) |c_\ell(f^*)|^2 < \infty,$$

then $\epsilon^{-1}(\hat{u}_\epsilon - u^*) \to N(0, \Gamma^{-1})$, as $\epsilon \to 0$, where the matrix Γ is given by

$$\Gamma_{(j_1,k_1)\times(j_1,k_2)} = \sum_{\ell\in\mathbb{Z}} \left(1 - \frac{1}{J}\right) |c_\ell(f^*)|^2 (2\pi)^2 \ell_{k_1}\ell_{k_2},$$
$$\Gamma_{(j_1,k_1)\times(j_2,k_2)} = -\frac{1}{J} \sum_{\ell\in\mathbb{Z}} |c_\ell(f^*)|^2 (2\pi)^2 \ell_{k_1}\ell_{k_2} \text{ for } j_1 \neq j_2,$$

4 Numerical simulations and some illustrative examples

4.1 A general gradient descent algorithm

To compute the estimator $\hat{h}_\epsilon$ one has to minimize the function $M_\epsilon(h)$. As this criterion is defined on a Lie group, a direct numerical optimization is generally not feasible. Finding minima of functions defined on a Lie group is generally done by reformulating the problem as an optimization problem on its Lie algebra. Since the expression of the gradient of $\tilde{M}_\epsilon(u)$ is available in a closed form, the following gradient descent algorithm with an adaptive step can be easily implemented:

Initialization : let $u^0 = 0 \in \mathcal{G}^J$, $\gamma_0 = \frac{1}{\|\nabla_{u^0}\tilde{M}_\epsilon\|}$, $M(0) = \tilde{M}_\epsilon(u^0)$, and set $m = 0$.
Step 2 : let $u^{new} = u^m - \gamma_m \nabla_{u^m}\tilde{M}_\epsilon$ and $M(m+1) = \tilde{M}_\epsilon(u^{new})$
  While $M(m+1) > M(m)$ **do**
  $\gamma_m = \gamma_m/\kappa$, and $u^{new} = u^m - \gamma_m \nabla_{u^m}\tilde{M}_\epsilon$, and $M(m+1) = \tilde{M}_\epsilon(u^{new})$
  End while
Then, take $u^{m+1} = u^{new}$ and set $m = m+1$.
Step 3 : if $M(m) - M(m+1) \geq \rho(M(1) - M(m+1))$ then return to Step 2, else stop the iterations, and take $\hat{h}_\epsilon = \exp(u^{m+1})$.

In the above algorithm, $\rho > 0$ is a small stopping parameter and $\kappa > 1$ is a parameter to control the choice of the adaptive step γ_m. Note that to satisfy the identifiability constraints the first p components of u^m are held fixed to zero at each iteration m.

4.2 Registration of translated 2D images

As an illustrative example, we consider the registration of translated 2D images (see [9] for a related work in a similar setting for 2D images). The chosen template f^* is the Shepp-Logan phantom image [1] of size $N \times N$ with $N = 100$ (see Figure 1). Noisy images can be generated by translating this image and adding Gaussian noise to each pixel value:

$$Y^j(i_1, i_2) = f^*\left(\frac{i_1}{N} - h_j^1, \frac{i_2}{N} - h_j^2\right) + \sigma z_j(i_1, i_2),\ 1 \leq i_1, i_2 \leq N,\ j = 1, \ldots, J \quad (14)$$

where i_1, i_2 denotes a pixel position in the image, $z_j(i_1, i_2) \sim_{i.i.d.} N(0,1)$, σ is the level of noise, and $h_j^1, h_j^2 \in [0,1]$ are the unknown translation parameters to estimate. One could argue that the sampled data model (14) does not truly correspond to the white noise model (3). However, as previously explained there exists a correspondence between these two models in the sense that they are asymptotically equivalent if $\epsilon = \frac{\sigma}{N}$ (see [5]).

We have repeated $M = 100$ simulations with $J = 6$ noisy images simulated from the model (14). The various values taken for the translation parameters are the bold numbers given in Table 1. A typical example of a simulation run is shown in Figure 1. Here, $G = [0,1] \times [0,1]$ and its Lie algebra is $\mathcal{G} = \mathbb{R}^2$. The criterion $\tilde{M}_\epsilon(u)$ can be easily implemented via the use of the fast Fourier transform for 2D images:

$$\tilde{M}_\epsilon(u) = \frac{1}{J} \sum_{j=1}^{J} \sum_{|\ell_1| \leq \ell_\epsilon} \sum_{|\ell_2| \leq \ell_\epsilon} \left| y^j_{\ell_1,\ell_2} e^{i2\pi(\ell_1 u_1^j + \ell_2 u_2^j)} - \frac{1}{J} \sum_{j'=1}^{J} y^{j'}_{\ell_1,\ell_2} e^{i2\pi(\ell_1 u_1^{j'} + \ell_2 u_2^{j'})} \right|^2$$

for $u = (u_1^1, u_1^2, \ldots, u_J^1, u_J^2)$, and where the $y^j_{\ell_1,\ell_2}$'s are the empirical Fourier coefficients of the image Y^j. Moreover, if $(x_1^1, x_1^2, \ldots, x_J^1, x_J^2)$ denotes the canonical basis of the product space $(\mathbb{R}^2)^J$, then the components of the gradient of $\tilde{M}_\epsilon(u)$ are given by

$$\frac{\partial}{\partial x_j^k} \tilde{M}_\epsilon(u) = -\frac{2}{J} \sum_{|\ell_1| \leq \ell_\epsilon} \sum_{|\ell_2| \leq \ell_\epsilon} \Re \left((i2\pi\ell_k) y^j_{\ell_1,\ell_2} e^{i2\pi(\ell_1 u_1^j + \ell_2 u_2^j)} \overline{\left(\frac{1}{J} \sum_{j'=1}^{J} y^{j'}_{\ell_1,\ell_2} e^{i2\pi(\ell_1 u_1^{j'} + \ell_2 u_2^{j'})}\right)} \right).$$

According to Corollary 1, the smoothing parameter ℓ_ϵ should be chosen such that $\epsilon \ell_\epsilon^4 = o(1)$. Since the models (14) and (3) are asymptotically equivalent if $\epsilon = \frac{\sigma}{N}$, this condition becomes $\ell_\epsilon = \ell_N = o(N^{1/4})$. With $N = 100$, this suggests to take $\ell_N \leq 100^{1/4} \approx 3.16$. In Table 1 we give the empirical average of the estimated parameters over the $M = 100$ simulations, for the choice $\ell_N = 3$, together with their standard deviation. The results are extremely satisfactory as averages are very close to the true values and standard deviations are very small.

Fig. 1. A typical simulation run for $J = 6$ images generated from the model (14).

Table 1. Average and standard deviation (in brackets) of the estimators $\hat{h}_j = (\hat{h}_j^1, \hat{h}_j^2)$ over $M = 100$ simulations. The bold numbers represent the true values of the parameters (h_j^1, h_j^2).

	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$
h_j^1	**0.07**	**0.1**	**0.05**	**-0.05**	**-0.08**
$\hat{h}_j^1$	0.0704 (0.0031)	0.0997 (0.0031)	0.0494 (0.0028)	-0.0502 (0.0031)	-0.0801 (0.0032)
h_j^2	**0.02**	**0.08**	**-0.10**	**-0.05**	**0.06**
$\hat{h}_j^2$	0.0201 (0.0031)	0.0803 (0.0031)	-0.1002 (0.0030)	-0.0493 (0.0029)	0.0604 (0.0032)

5 Registration of spherical images

In many applications, data can be organized as spherical images i.e. as functions defined on the unit sphere $\mathbb{S}^2$. For instance, spherical images are widely used in robotics since the sphere is a domain where perspective projection can be mapped, and an important question is the estimation of the camera rotation from such images (see [11]). Obviously such data do not correspond exactly to the shift model on group (3) as spherical images are defined on $\mathbb{S}^2$ while the shifts parameters belong the special orthogonal group SO(3). However this setting corresponds to the general model (2) with $\mathcal{X} = \mathbb{S}^2$ and $G = \mathrm{SO}(3)$, and a matching criterion similar to the one defined in equation (6) can still be defined by combining the spherical harmonics on $\mathbb{S}^2$ with the irreducible representations of SO(3).

Indeed, let $x \in \mathbb{S}^2$ be a point on the unit sphere parameterized with spherical coordinates $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi[$. Then any $f \in \mathbb{L}^2(\mathbb{S}^2)$ can be decomposed as (see [6])

$$f(x) = \sum_{\ell=0}^{+\infty} \sum_{m=-\ell}^{\ell} c_\ell^m(f) \psi_\ell^m(x), \text{ with } c_\ell^m(f) = \int_0^\pi \int_0^{2\pi} f(\theta, \phi) \overline{\psi_\ell^m(\theta, \phi)} d\phi \sin(\theta) d\theta,$$

and where the functions $(\psi_\ell^m, \ell \in \mathbb{N}, m = -\ell, \ldots, \ell)$ are the usual spherical harmonics which form an orthonormal basis of $\mathbb{L}^2(\mathbb{S}^2)$. Since these functions form a basis for the irreducible representations $(\pi_\ell)_{\ell \in \mathbb{N}}$ of SO(3) which are matrices of size $(2\ell + 1) \times (2\ell + 1)$, it follows that the action of a rotation $h \in \mathrm{SO}(3)$ on a function $f \in \mathbb{L}^2(\mathbb{S}^2)$ is given by (see [6])

$$f(h^{-1}x) = \sum_{\ell=0}^{+\infty} c_\ell(f)^T \pi_\ell(h^{-1}) \psi_\ell(x) \quad \text{for all } x \in \mathbb{S}^2, \tag{15}$$

where $c_\ell(f) = (c_\ell^m(f))_{m=-\ell,\ldots,\ell}$ and $\psi_\ell(x) = (\psi_\ell^m(x))_{m=-\ell,\ldots,\ell}$ denotes vectors in $\mathbb{C}^{2\ell+1}$.

Now, suppose that one observes a set of noisy spherical images f_j that satisfy the following shift model: for $j = 1, \dots, J$ and $x \in \mathbb{S}^2$

$$dZ_j(x) = f_j(x)dx + \epsilon dW_j(x) \text{ with } dx = d\phi \sin(\theta) d\theta, \tag{16}$$

where $f_j(x) = f^*(h_j^{*-1}x)$, and $h_j^* \in \mathrm{SO}(3), j = 1, \dots, J$ are rotation parameters to estimate. For $h = (h_1, \dots, h_J) \in G^J$, the shift property (16) and the orthonormality of the spherical harmonics imply that the following matching criterion

$$N(h) = \frac{1}{J} \sum_{j=1}^{J} \left\| f_j \circ T_{h_j} - \frac{1}{J} \sum_{j'=1}^{J} f_{j'} \circ T_{h_{j'}} \right\|_{\mathbb{L}^2(\mathbb{S}^2)}^2, \tag{17}$$

where $T_{h_j} : x \in \mathbb{S}^2 \to h_j x \in \mathbb{S}^2$, can be written as

$$N(h) = \frac{1}{J} \sum_{j=1}^{J} \sum_{\ell=0}^{+\infty} \left\| c_\ell(f_j)^T \pi_\ell(h_j) - \frac{1}{J} \sum_{j'=1}^{J} c_\ell(f_{j'})^T \pi_\ell(h_{j'}) \right\|_{\mathbb{C}^{2\ell+1}}^2, \tag{18}$$

where $c_\ell(f_j)^T = c_\ell(f^*)^T \pi_\ell(h_j^{*-1})$ and $\|\cdot\|_{\mathbb{C}^{2\ell+1}}^2$ denotes the usual euclidean norm in $\mathbb{C}^{2\ell+1}$. Then, remark that the spherical harmonic coefficients of the noisy images Z_j are given by (in vector form) by $c_\ell(Z_j) = \int_{\mathbb{S}^2} \psi_\ell(x) dZ_j(x) = c_\ell(f_j) + \epsilon c_\ell(W_j), \quad j = 1 \dots J$, where $c_\ell(W_j) = \int_{\mathbb{S}^2} \psi_\ell(x) dW_j(x)$ is a complex random vector whose components are independent and identically distributed Gaussian variables $\mathcal{N}_{\mathbb{C}}(0, 1)$. Now, let ℓ_ϵ be an appropriate frequency cut-off parameter. The following empirical criterion can thus be proposed for registering spherical images:

$$N_\epsilon(h_1, \dots, h_J) = \frac{1}{J} \sum_{j=1}^{J} \sum_{\ell=0}^{\ell_\epsilon} \left\| c_\ell(Z_j)^T \pi_\ell(h_j) - \frac{1}{J} \sum_{j'=1}^{J} c_\ell(Z_{j'})^T \pi_\ell(h_{j'}) \right\|_{\mathbb{C}^{2\ell+1}}^2, \tag{19}$$

and an M-estimator of the rotation parameters is given by $\hat{h}_\epsilon = \arg\min_{h \in \mathcal{A}_0} N_\epsilon(h)$.

The criterion N_ϵ is very similar to the criterion M_ϵ, and the study of the consistency and the asymptotic normality of $\hat{h}_\epsilon$ can be done by following exactly the arguments as those developed in the previous sections.

6 Some perspectives and future work

The results on the asymptotic normality of the estimators show that there exists a significant difference between semi-parametric estimation on a linear Euclidean space and semi-parametric estimation on a nonlinear manifold. If the group G is non-commutative, then the asymptotic covariance matrix of the estimator $\hat{u}_\epsilon$ depends on the point u^* (and thus on h^*). Hence, this matrix can be interpreted as a Riemanian metric on G. This is a classical result in parametric statistics for random variables whose law is indexed by parameters belonging to a finite-dimensional manifold. In such setting, the Fisher information matrix is a Riemanian metric and lower bounds analogue to the classical Cramer-Rao bound for parameters in an Euclidean space can be derived (see e.g. [17]). If G is supposed to be an Abelian group, then the asymptotic covariance matrix of the estimator does not depend on the point h^* since the parameter space for the shifts is a flat manifold.

An important issue is then the question of optimality of our estimators. We are currently studying analogs of the Cramer-Rao bound for the semi-parametric model (3), and in particular we are investigating if the covariance matrix given in Theorem 2 corresponds to the Fisher information matrix of this model. This result would provide a proof of the optimality of our reconstruction, even in a non asymptotic framework.

Another important question is the implementation of our approach for non-commutative groups. The numerical computation of our method for the registration of spherical images is more involved that the one used for the alignment of 2D images. Indeed, one has to deal with both the problem of computing the Fourier transform for images defined on a sphere, and with the problem of computing the irreducible representations of the group SO(3) from its Lie algebra. We are currently working on the development of an efficient and fast numerical scheme to minimize the criterion N_ϵ, and we believe that this approach could yield good results for the registration of spherical images.

Finally, it should be mentioned that this work is rather preliminary and practical applications are restricted to estimation of shifts in 2D images and rotation on a sphere for the registration of spherical images. Another application that would be of great interest is the analysis of images of the fundus of the eye as described in [7]. However, many other interesting applications in medical images involve the study of more sophisticated Lie groups of transformations that are generally non-compact. We believe that there is a good chance to obtain satisfactory results for the estimation of deformations on non-commutative and compact groups. However, an important challenge is to investigate the extension of this work to the case of non-compact groups and at least to identify the difficulties introduced by such a generalization.

Acknowledgments

We are very much indebted to the three anonymous referees for their constructive criticism, comments and remarks that resulted in a substantial revision of the original manuscript.

References

1. Anil, K. J. (1989) Fundamentals of digital image processing. Prentice-Hall, Inc..
2. Bigot, J., Loubes, J.-M. & Vimond, M. (2008) Semiparametric Estimation of Shifts on Lie Groups for Image Registration. submitted paper.

3. Beg, M. F., Miller, M. I., Trouvé, A. & Younes, L. (2005) Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int. J. Comput. Vis., **61**, 139157.
4. P.J. Bickel, Klaassen C.A.J., Ritov Y. and Wellner J.A. (1998) Efficient and Adaptive Estimation for Semiparametric Models. Springer Verlag.
5. Brown, L.D. & Low, M.G. (1996). Asymptotic equivalence of nonparametric regression and white noise. Ann. Statist., **24**, 2384–2398.
6. Chirikjian, G. S. & Kyatkin, A. B. (2001). Engineering Applications of Noncommutative Harmonic Analysis. CRC Press.
7. Deguchi, K., Noami, J. and Hontani, H. (2000). 3D fundus pattern reconstruction and display from multiple fundusimages. Proceedings 15th International Conference on Pattern Recognition, Volume 4, 94 - 97.
8. Gamboa, F., Loubes, J.-M. and Maza, E. (2007) Semi-parametric estimation of shifts. Electronic Journal of Statistics, 1, 616–640.
9. Glasbey, C. A. and Mardia, K. V. (2001) A penalized likelihood approach to image warping. Journal of The Royal Statistical Society Series B, 3, 465-492.
10. Klassen, E., Srivastava, A., Mio,W. & Joshi, S. (2004) Analysis of planar shapes using geodesic paths on shape spaces. IEEE Trans. Pattn Anal. Mach. Intell., 26, 372383.
11. Makadia, A. and Daniilidis, K. (2006) Rotation Recovery from Spherical Images without Correspondences. IEEE Trans. Pattern Anal. Mach. Intell., 28(7), 1170–1175.
12. Marsland, S. and Twining, C. (2004) Constructing diffeomorphic representations for the groupwise analysis of non-rigid registrations of medical images. IEEE Trans. Med. Imgng, 23.
13. Michor, P. and Mumford, D. (2006) Riemannian geometries on spaces of planar curves. J. Eur. Math. Soc., 8, 1-48.
14. Pennec, X. and Fillard, P. (2008) Statistical computing on non-linear spaces for computational anatomy. Biomedical Image Analysis: Methodologies and Applications, Springer, Editors: Paragios, N. and Duncan, J. and Ayache, N.
15. Rudin, W. (1990). Fourier analysis on groups. Wiley Classics Library, Reprint of the 1962 original edition, New York.
16. Sepanski, M.R. (2007). Compact Lie groups. Graduate Texts in Mathematics, Springer, New York.
17. Smith, S.T. (2005) Covariance, subspace, and intrinsic Cramér-Rao bounds. IEEE Transactions on Signal Processing, **53**, 1610-1630.
18. Trouvé, A. and Younes, L. (2005) Metamorphoses through lie group action. Found. Computnl Math., 5, 173198.
19. van der Vaart, A. W. (1998). Asymptotic statistics. Cambridge University Press.
20. Vimond, M. (2008) Efficient estimation in homothetic shifted in regression models. To be published in the Annals of Statistics.