Probab. Theory Relat. Fields (2012) 152:425–473
DOI 10.1007/s00440-010-0327-2

Semiparametric estimation of shifts on compact Lie groups for image registration

Jérémie Bigot · Jean-Michel Loubes · Myriam Vimond

Received: 9 September 2008 / Revised: 29 August 2010 / Published online: 18 November 2010

Abstract In this paper we focus on estimating the deformations that may exist between similar images in the presence of additive noise when a reference template is unknown. The deformations are modeled as parameters lying in a finite dimensional compact Lie group. A general matching criterion based on the Fourier transform and its well known shift property on compact Lie groups is introduced. M-estimation and semiparametric theory are then used to study the consistency and asymptotic normality of the resulting estimators. As Lie groups are typically nonlinear spaces, our tools rely on statistical estimation for parameters lying in a manifold and take into account the geometrical aspects of the problem. Some simulations are used to illustrate the usefulness of our approach and applications to various areas in image processing are discussed.

Keywords Image registration · Lie group · Semi-parametric estimation · M-estimation · Fourier transform · Statistical inference on manifolds · White noise model

Mathematics Subject Classification (2000) Primary 62F12 · Secondary 65Hxx

1 Introduction

In order to extract any information from a set of images, it is common sense that one has to be able to compare the images one together. However, such a comparison is a difficult task due to the lack of convexity of the space of images, and even giving

J. Bigot · J.-M. Loubes (✉)
Institut de Mathématiques de Toulouse, Université de Toulouse, 31062 Toulouse Cedex 9, France
e-mail: jean-michel.loubes@math.univ-toulouse.fr

M. Vimond
CREST-ENSAI, ENSAI, Ker-Lann, 35172 Bruz, France

a sense to the notion of a mean image is not an easy matter. Hence, one of the most commonly used model is to consider that the data are obtained through the deformation of the same image often called template or reference image. In Grenander's theory of shapes [15], images are considered as points in an infinite dimensional manifold and the variations of the images are modeled by the action of Lie groups on the manifold. Finite dimensional Lie groups can be used to model rigid displacement such as translation or rotation, while infinite dimensional groups such as spaces of diffeomorphisms can model local and non-rigid deformations of an image and thus provide much more flexibility than finite dimensional groups. In the last decade, there has been a growing interest in transformation Lie groups to model the variability of natural images, and the study of the properties and intrinsic geometries of such deformation groups is now an active field of research (see e.g. [4,26,30,38] and references therein).

An important problem in this setting is the estimation of the mean pattern, achieved through the estimation of the deformations between similar images in the presence of additive noise when a reference template is unknown. This is the so-called registration or warping problem of images (see [14] and the discussion therein for a detailed overview of image registration in a statistical setting). The main goal of this paper is to build such estimates and study their statistical properties when the deformation parameters are modeled by finite dimensional Lie groups. Statistical estimation of parameters lying in a smooth Riemannian manifold has been originally studied by [1]. A general overview and extensive references on the geometrical aspect of statistical inference on manifold can be found in [22]. The difficulty of statistical analysis on manifolds comes from the fact that the parameter space is generally not linear which makes the definition of simple notions such as mean or covariance a difficult task. Yet, various statistical problems in this context have been studied such as mean estimation from a sample of random variables on a manifold [6,7], nonparametric estimation of location and dispersion parameters in a Riemannian manifold [5] or statistical estimation and nonparametric inference in group models for manifold valued variables [10,23–25]. However, to the best of our knowledge, the literature on statistical estimation on Lie groups for warping problems is scarce.

Consider the following general model for the registration of images: let $\mathcal{X}$ be a subset of $\mathbb{R}^d$ (with $d = 2, 3$ in our applications) and G be a connected Lie group acting on $\mathcal{X}$. For $x \in \mathcal{X}$ and $h \in G$, the action of h onto x will de denoted by hx. To model a set of J images (with $J > 1$), let us consider the following general deformation white noise model

$$dY_j(x) = f_j(x)dx + \epsilon dW_j(x) \quad \text{for } x \in \mathcal{X}, \quad j = 1 \dots J \tag{1.1}$$

where

$$f_j(x) = f^*(h_j{}^{*-1}x).$$

The function $f^* : \mathcal{X} \to \mathbb{R}$ is the unknown common shape of the observed images Y_j.

The $h_j^* \in G$, $j = 1, \dots, J$ are the unknown deformation parameters that we wish to estimate, W_j, $j = 1, \dots, J$ are independent standard Brownian sheets on the topological space $\mathcal{X}$ with reference measure dx, and ϵ is an unknown noise level parameter.

Note that the white noise model (1.1) is a continuous model which is a very useful tool for the theoretical study of statistical problem in image analysis. In practice, images are typically discretely sampled on a regular grid, and thus the model (1.1) may seem inappropriate at a first glance. However, asymptotic results obtained in the white noise model can be shown to lead to comparable asymptotic theory in a sampled data model provided the regression function satisfies appropriate smoothness conditions, see [3] for further details. Moreover a continuous model avoids the problem of controlling the bias introduced by any discretization scheme, and allows one to rather focus on the statistical properties of the estimators.

A typical example of the above model is the registration of translated two-dimensional (2D) images for which $\mathcal{X} = [0, 1]^2$, $G = \mathbb{R}^2/\mathbb{Z}^2$ (the torus in dimension two), and which founds its applications in biomedical imaging or satellite remote sensing (see [14,27]). Another example is a rotation model for spherical images for which $\mathcal{X} = \mathbb{S}^2$ (the unit sphere in dimension 3), and $G = \mathrm{SO}(3)$ (the special orthogonal group). Indeed in many applications, data can be organized as functions defined on a sphere. 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 e.g. [29]). A Bayesian approach in such model has been proposed also in [16] for automated target recognition of a deformable template under the action of rotations and translations in dimension 3.

Within the model (1.1), the problem of optimal recovery of the shift parameters h_j^* involves semiparametric techniques. Indeed, semiparametric modeling is concerned with statistical problems where the parameters of interest are both finite and infinite-dimensional. Here, the finite-dimensional parameters are the Lie group elements, and the infinite-dimensional parameter is the unknown template which is typically a 2D or 3D image (see [34] for a detail presentation of semiparametric statistics), which blurs the parametric estimation issue. The main idea in semiparametric statistics is to find an efficient tool which separates the effect of the parameters from the influence of the blurring infinite-dimensional parameter.

A matching criterion has been proposed in [13,42] for the mere problem of recovering shifts between noisy one-dimensional curves observed on an interval i.e. when the model (1.1) can be written as

$$dY_j(x) = f^*(x - h_j^*)dx + \epsilon dW_j(x) \quad \text{for } x \in [0, 1] \quad \text{and} \quad h_j \in [0, 1],\ j = 1, \ldots, J. \tag{1.2}$$

This criterion is based on the Fourier transform of the data and on its well-known shift property for one-dimensional translations. Indeed, let $e_\ell(x) = e^{-i2\pi\ell x}$, $\ell \in \mathbb{Z}$ denotes the standard Fourier basis. Then by taking the Fourier coefficients $d_{j,\ell} = \int_{[0,1]} e_\ell(x) dY_j(x)$ of the observed curves the model (1.2) becomes

$$d_{j,\ell} = e^{-i2\pi\ell h_j^*} c_\ell + \epsilon z_{j,\ell}, \quad \text{with} \quad c_\ell = \int\limits_{[0,1]} f^*(x) e_\ell(x) dx \quad \text{and} \quad z_{j,\ell} \sim_{i.i.d.} N(0, 1).$$

Hence, for a set of parameters $(h_1, \ldots, h_J) \in [0, 1]^J$, the following contrast function is defined in [13,42]

$$M(h_1, \ldots, h_J) = \sum_{|\ell| \leq \ell_\epsilon} \sum_{j=1}^{J} \left| e^{i2\pi \ell h_j} d_{j,\ell} - \frac{1}{J} \sum_{j'=1}^{J} e^{i2\pi \ell h_{j'}} d_{j',\ell} \right|^2 \tag{1.3}$$

where ℓ_ϵ is a frequency cut-off parameter. Under appropriate conditions, minimization of M over $[0, 1]^J$ is shown to yield consistent estimators. Note that the above criterion is closely related to Procrustean analysis which is classically used for the statistical analysis of shapes (see e.g. [28]) and the registration of a set of curves onto a common target function.

In this paper, we extend this approach to a multi-dimensional setting and to the general case where the shift parameters belong to a compact Lie group. First, as Lie groups are typically not linear 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 [17,18], and we refer to [36] for a detailed discussion and review. However, in the more general situation of semi-parametric models, there is few work dealing with the estimation of parameters lying in a Lie group. Then, in order to use the same kind of matching criterion, we need to use the extension of the the standard one-dimensional Fourier transform to functions defined on a compact Lie group. It is achieved via the theory of representations (see e.g. [35]). Thanks to a general shift property of the Fourier transform on arbitrary compact Lie group, a similar matching criterion based on the Fourier transform of the data can still be defined, and enables to investigate the statistical properties of the resulting estimators. Note that M-estimation for parameters in groups models has been considered in [10], but applying M-estimation theory in the context of image warping to compact Lie groups has not been proposed before.

The main contributions of this paper are the following: we provide a general framework for the registration problem of noisy images without a reference template. We build a general matching criterion for recovering the deformations that may exist between similar images, and we also study consistency and asymptotic normality for parameters lying in a Lie group. Although the model (1.1) looks as a toy model, our results already provides some insights into the estimation of deformations over Lie group. In particular, an important and new result is the study of the asymptotic covariance matrix of estimators belonging to non-commutative groups within a semiparametric framework. Indeed, our results on the asymptotic normality of the estimators show that there exists a significant difference between semiparametric estimation on a linear Euclidean space and semi-parametric estimation on a nonlinear manifold. Finally, our general matching criterion provides a feasible method to estimate the parameters h_j^*, which induces an estimator of the common shape f^* using the inversion theorem of the Fourier Transform. Then, the convergence of this estimator of the common shape is also studied.

The rest of the paper is organized as follows. In Sect. 2, some properties on the Fourier transform are briefly recalled, and a simple model for shifts on Lie groups is introduced. In Sect. 3, the shift property of the Fourier transform is used to define a general matching criterion on compact Lie groups, and the consistency of the estimator is established. The problem of studying and defining a notion of asymptotic normality of estimators belonging to a Lie group is studied in Sect. 4, which also includes a study of an estimator of the common shape. The efficiency of the resulting estimators for the shifts is discussed in Sect. 5. A general gradient descent algorithm, to minimize the matching criterion, is described in Sect. 6 and some numerical simulations are presented to illustrate the usefulness of this approach. Finally in Sect. 7 some extensions of our simple shift model are described and applied to the problem of registering spherical images. The main proofs are gathered in a technical Appendix.

2 A shift model on Lie groups

2.1 The Fourier transform on compact Lie groups

In what follows, some aspects of the theory of the Fourier transform on compact Lie groups are briefly summarized. For more details, we refer to the books of [8,11] and [35]. 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 complex 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 to the automorphism group of a vector space. So let V be a finite-dimensional vector space, we defined a *representation* of G in V 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.

A representation π on V is irreducible if the only invariant subspaces by the set of homomorphism $\pi(g)$, $g \in G$, are $\{0\}$ and V. 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. Two representations will be call equivalent if they are the same up to, basically a change of basis. Denote the set of equivalence classes of irreducible representations of G by $\hat{G}$. For simplicity, the same notation is used for π and its equivalence class in $\hat{G}$.

The function $g \mapsto \mathrm{Tr}\, \pi(g)$ is called the *character* of π, and 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 $\hat{G}$. 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

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

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

$$f^*(g) = \sum_{\pi \in \hat{G}} d_\pi \operatorname{Tr}\left(\pi(g)\pi(f^*)\right) \quad \text{and} \quad \|f^*\|^2_{\mathbb{L}^2(G)} = \sum_{\pi \in \hat{G}} d_\pi \operatorname{Tr}\left(\pi(f^*)\overline{\pi(f^*)}^T\right) \tag{2.1}$$

In the sequel, we will also denote by $\langle A, B\rangle_{HS} = \operatorname{Tr}(\overline{A}^T B)$ the Hilbert-Schmidt inner product between two finite dimensional $d_\pi \times d_\pi$ matrices A and B. Note that if G equals the circle $\mathbb{R}/\mathbb{Z}$, then $\hat{G} = \mathbb{Z}$, the representation are the trigonometric polynomials, the "matrices" $\pi(f^*)$ are one-dimensional and equal the standard Fourier coefficients, and one finally retrieves the classical Fourier decomposition of a periodic function in $L^2[0, 1]$.

2.2 A simple shift model

To focus on the geometrical aspects of the statistical procedure and to simplify the presentation, the simplest model for which $\mathcal{X} = G$ is studied to give the main ideas of our approach. A discussion in the case where $\mathcal{X} \neq G$ is deferred to Sect. 7 to show that the methodology can be extended to more complex situations. In this case the general model (2.2) becomes

$$dY_j(g) = f_j(g)dg + \epsilon dW_j(g), \quad j = 1 \ldots J \tag{2.2}$$

where $f_j(g) = f^*(h_j^{*-1}g)$, and W_j are independent standard Brownian sheets on the Lie group G. Surveys on the constructions of Brownian motions indexed by a Lie group can be found in [2,12,20] and references therein. Obviously, without any further restriction on the set of possible shifts, the model (2.2) is not identifiable. Indeed, if s is an element of G with $s \neq e$, then one can replace the h_j^*'s in Eq. (2.2) by $\tilde{h}_j = h_j^* s$ and f^* by $\tilde{f}(g) = f^*(sg)$ without changing the formulation of the model.

Let $\mathcal{A}$ denote the space G^J. To ensure identification, we further assume that the set of parameters $\mathcal{A}$ is reduced to the subset $\mathcal{A}_0 \subset \mathcal{A}$ such that

$$\mathcal{A}_0 = \{(h_1, \ldots, h_J) \in \mathcal{A}, h_1 = e\}. \tag{2.3}$$

The above assumption will also guarantee the convergence of our estimators (see Theorem 3.1). Since $\overline{\pi(g)}^T = \pi(g^{-1})$, one has that for all $j = 1, \ldots, J$

$$\pi(f_j) = \int_G f^*(h_j^{*-1}g)\pi(g^{-1})dg = \int_G f^*(g)\pi((h_j^* g)^{-1})dg = \pi(f^*)\pi(h_j^{*-1}).$$

The above formula is classically referred to as the *shift property* of the Fourier transform. Indeed, it is well known that for the standard Fourier transform on $\mathbb{R}$, then shifting a function only amounts to a phase correction of its Fourier coefficients. This property is at the heart of our estimation procedure to exhibit the shift parameters h_j^*.

2.3 Regularity assumption on the common shape

Since we use the Fourier transform to build our estimation method, it will be natural to suppose that the common shape f^* satisfy the following assumption:

$$f^* \in \mathbb{L}^2(G) \subset \mathbb{L}^1(G),$$

where $\mathbb{L}^1(G)$ denotes the set of integrable function on G with respect to dg.

Now remark that the function f^* should satisfy some geometric conditions to make the estimation of the shift parameters feasible. Indeed, think of a spherical image that would be symmetric with respect to some axis through the origin. Such an image is thus rotation invariant and a proper estimation of the shifts is therefore impossible. Now, let us study the general case. Assume that there exists a closed subgroup H of G (not reduced to e) such that $f^*(gh) = f^*(g)$ for all $g \in G$ and $h \in H$. Then, there is a unique manifold structure on the quotient group $K = G/H$ so that the projection map $P_H : G \to K$ is smooth. Let π be an irreducible representation of K on the vector space V. Then, π can be used to define an irreducible representation of G by $\pi \circ P_H$. Furthermore, from the Parseval formula (2.1) we have that:

$$\|f^*\|^2_{\mathbb{L}^2(G)} = \sum_{\pi \in \hat{G}} d_\pi \operatorname{Tr}\left(\pi(f^*)\overline{\pi(f^*)}^T\right) = \sum_{\pi \in \hat{K}} d_\pi \operatorname{Tr}\left(\pi(f^*)\overline{\pi(f^*)}^T\right).$$

For any two (generic) sets A and B, the set $A \setminus B$ denotes the set A minus B. Then we deduce that for all irreducible representation $\pi \in \hat{G} \setminus \hat{K}$ the linear mapping $\pi(f^*) : V \to V$ is identically null and thus the shift property of the Fourier transform can not be used to recover the shifts for such π's, and of course the set $\hat{G} \setminus \hat{K}$ is clearly unknown in practice. Thus, the following definition is introduced.

Definition 2.1 A function $f^* \in \mathbb{L}^2(G)$ is said to be *not shift-invariant* if 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 $h \in H$.

Finally, we also impose some smoothness assumptions on the function to recover f^* which are given by the following definition. This assumption is used to guarantee the unicity of the minimum of the M-criterion defined in the next section.

Definition 2.2 A function $f^* \in \mathbb{L}^2(G)$ is said to be regular if for all $\pi \in \hat{G}$ such that $\pi(f^*)$ is not identically null, then $\pi(f^*)$ is invertible.

3 The *M*-estimation criterion

3.1 A matching criterion based on the Fourier transform

For $h=(h_1,\ldots,h_J)\in\mathcal{A}_0$, we propose to minimize the following criterion inspired by recent results of [13] and [42] for the estimation of shifts between curves:

$$M(h_1,\ldots,h_J)=\frac{1}{J}\sum_{j=1}^{J}\left\| f_j\circ L_{h_j}-\frac{1}{J}\sum_{j'=1}^{J} f_{j'}\circ L_{h_{j'}}\right\|^2_{\mathbb{L}^2(G)}, \tag{3.1}$$

where $L_h : g\in G\to hg\in G$ and $f_j : g\in G\to f^*(h_j^{*-1}g)\in\mathbb{R}$. Using the Parseval–Plancherel formula, the criterion may be rewritten in the Fourier domain as:

$$M(h)=M(h_1,\ldots,h_J)=\frac{1}{J}\sum_{j=1}^{J}\sum_{[\pi]\in\hat{G}} d_\pi\left\|\pi(f_j)\pi(h_j)-\frac{1}{J}\sum_{j'=1}^{J}\pi(f_{j'})\pi(h_{j'})\right\|^2_{HS}, \tag{3.2}$$

for $h=(h_1,\ldots,h_J)\in\mathcal{A}_0$. Given that $\pi(f_j)=\pi(f^*)\pi(h_j^{*-1})$, the criterion M has a minimum at $h^*=(h_1^*,\ldots,h_J^*)$ such that $M(h^*)=0$. If f^* is assumed to be not shift-invariant and regular (see Definitions 2.1 and 2.2), then this minimum is unique (see the proof of Theorem 3.1).

3.2 The empirical criterion

Our estimation method is then based on the Fourier Transform of the noisy data given by model (2.2). 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 (2.2):

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

where

$$\pi(W_j)=\int_G \pi(g^{-1})dW_j(g),\quad j=1\ldots J.$$

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 $\mathcal{N}_{\mathbb{C}}(0, d_\pi^{-1})$. Notice that if π is an irreducible representation, then the conjugate representation $\overline{\pi} : g \in G \to \overline{\pi(g)}$ is irreducible too.

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

$$\cup_{\epsilon>0}\hat{G}_\epsilon = \hat{G}.$$

Moreover, we assume that if $\pi \in \hat{G}_\epsilon$, then $\overline{\pi} \in \hat{G}_\epsilon$. Practical choices for the set $\hat{G}_\epsilon$ will be discussed later on in the paper for the case of Abelian groups and the non-commutative group SO(3). Then, we consider the following criterion:

$$M_\epsilon(h_1, \ldots, h_J) = \frac{1}{J}\sum_{j=1}^{J}\sum_{\pi\in\hat{G}_\epsilon} d_\pi \left\| \pi(Y_j)\pi(h_j) - \frac{1}{J}\sum_{j'=1}^{J}\pi(Y_{j'})\pi(h_{j'}), \right\|_{HS}^2 \tag{3.3}$$

and the M-estimator given by

$$\hat{h}_\epsilon = \arg\min_{h\in\mathcal{A}_0} M_\epsilon(h).$$

The following theorem provides the consistency of $\hat{h}_\epsilon$. Note that a Lie group is a topological space which can be equipped with a metric, and thus the convergence in probability of $\hat{h}_\epsilon$ is defined with respect to this metric.

Theorem 3.1 *Assume that f^* is not shift-invariant and regular. Moreover suppose that*

$$\lim_{\epsilon\to 0}\epsilon^2\sum_{\pi\in\hat{G}_\epsilon} d_\pi^2 = 0$$

then $\hat{h}_\epsilon$ converges in probability to $h^* = (h_1^*, \ldots, h_J^*)$.

The condition that $\lim_{\epsilon\to 0}\epsilon^2\sum_{\pi\in\hat{G}_\epsilon} d_\pi^2 = 0$ in Theorem 3.1 restricts the choice of the subset $\hat{G}_\epsilon$. Such a condition leads to the choice of a subset $\hat{G}_\epsilon$ with a small number of irreducible representations of low dimensions d_π. Some examples in the case of an Abelian group and for $G = \mathrm{SO}(3)$ are given in the next section.

4 Asymptotic normality of the estimator

By their nature, groups are usually nonlinear objects. Thus, it is not obvious to define a notion of asymptotic normality for an M-estimator such as $\hat{h}_\epsilon$ which takes its values in a group. Indeed asymptotic normality of M-estimators is classically derived using the differentiability over a vector space of the criterion to minimize. To overcome this, we assume that G is a Lie group. Then, a way to linearize a Lie group is to look at its Lie algebra via the exponential map. The Lie algebra is the vector space of left invariant vectors fields equipped with the Lie bracket of vector fields. The exponential map is a bridge between the structure of a Lie group and the structure of its Lie algebra which is a vector space. At the neighborhood of some point $g \in G$, a Lie group is very similar to its tangent space at g which is a vector space that can be identified to the Lie algebra of G. In this setting, the asymptotic normality of an estimator can be established by examining the behavior of the estimator in the immediate neighborhood of the parameter h^* to estimate. Thus, all we need is to define an appropriate system of coordinates to parametrize the group G in the vicinity of the point h^*. Such an approach has been proposed for instance in [7] to study the asymptotic properties of the intrinsic mean of a sample of random variables taking their values in a manifold. In the next sections, we provide some background on Lie groups and fix the notations. Then, we study the asymptotic normality of the estimator $\hat{h}_\epsilon$.

4.1 Lie group, Lie algebra and the exponential map

Let us first introduce some definitions. A Lie group $(G, \cdot)$ is a group which has also the structure of a manifold such that the group product and the inversion are smooth mapping with respect to the differential structure for the manifold. Let F be a smooth manifold of finite dimension. For each point $p \in F$, recall that the tangent space $T_p F$ is the vector space of all point-derivations of the algebra of smooth germs defined in the vicinity of p. The tangent bundle of F is the disjoint union of all tangent space spaces of F,

$$TF = \bigcup_{p \in F} \{p\} \times T_p F.$$

The derivative of a function $m : G \to F$ at a point h in the direction $v \in T_h G$ will be written as $d_h m(v) \in T_{m(h)} F$ and $d_h m : T_h G \to T_{m(h)} F$ is a linear map. Then, it defines a linear operator $dm : TG \to TF$, called the tangent map of m, such that:

$$\forall (g, X_g) \in TG, \quad dm(g, X_g) = d_g m(X_g).$$

Then, consider the left multiplication which is defined for any $h \in G$ as the mapping $L_h : G \to G$ such that $L_h(g) = hg$. The left multiplication L_h is diffeomorphism. The derivative of $L_{h^{-1}}$ at point h is known to determine an isomorphism between $T_e G$ and $T_h G$. Therefore any element of $T_h G$ can be identified with an element of $T_e G$ via the relation $T_e G = dL_{h^{-1}}(h, T_h G)$.

Now, recall that a vector field X on G is a smooth section of the tangent bundle TG

$$\begin{aligned} X : G &\to TG \\ g &\to (g, X_g), \end{aligned}$$

where $X_g \in T_gG$ for all $g \in G$. A vector field X is left invariant if $d_gL_hX_g = X_{hg}$ for all $(g,h) \in G \times G$, and let us denote by $\Gamma^L(G)$ the space of left invariant vector fields. Hence a left invariant vector field X is completely determined by its value at the identity e since $X_g = d_eL_gX_e$. Moreover one can associate to any $u \in T_eG$, a left invariant vector field X^u given by $X^u_g = d_eL_g(u)$, and one can check that the mapping

$$\begin{cases} \psi : T_eG \to \Gamma^L(G) \\ \qquad\quad u \mapsto X^u \end{cases}$$

is invertible by simply associating to any left-invariant vector field $X \in \Gamma^L(G)$ its value at the identify e namely $\psi^{-1}(X) = X_e$. Clearly ψ is an isomorphism and therefore, the tangent space T_eG of G at the identity e is in bijection with the space $\Gamma^L(G)$ which will be written as $T_eG \simeq \Gamma^L(G)$. Then, let us now define the Lie algebra of G:

Definition 4.1 The Lie algebra $\mathcal{G}$ of G is the tangent space at the identity e i.e. $\mathcal{G} = T_eG \simeq \Gamma^L(G)$.

The dimension of G as a Lie group will always be assumed to be finite. The Lie algebra $\mathcal{G}$ is thus a vector space of finite dimension $p \geq 1$. Since a Lie group is a topological space, the notion of convergence is defined with respect to this topology. However, for studying convergence results on $\mathcal{G}$, it will be useful to equipped the Lie algebra with a Banach space structure. We therefore suppose that there exists a norm $\|\cdot\|$ on $\mathcal{G}$ which induces a complete metric. For a detailed presentation of compact Lie groups and Lie algebra we refer to [35] and [39].

Each left invariant vector field X defines a differential equation governed by a flow denoted by $\phi_X(t,g) \in G$ for all $(t,g) \in \mathbb{R} \times G$ such that

$$\begin{cases} \frac{\partial}{\partial t}\phi_X(t,g) = X_{\phi_X(t,g)} \\ \phi_X(0,g) \;\; = g \end{cases}$$

Since X is left invariant, one has that $\phi_X(t,g) = L_g\phi_X(t,e)$ and thus the flow is completely determined by the initial condition $\phi_X(0,e) = e$. Then, we arrive at the following definition:

Definition 4.2 The exponential map exp is the mapping from $\mathcal{G} \simeq \Gamma^L(G) \to G$ defined by $\exp(X) = \phi_X(1,e)$ for $X \in \Gamma^L(G)$.

Using that $T_eG \simeq \Gamma^L(G)$, the exponential map can also be seen as a mapping from $\mathcal{G} = T_eG$ to G given by

$$\exp(u) = \phi_{X^u}(1,e) \quad \text{for} \quad u \in \mathcal{G} = T_eG.$$

This application maps $0 \in \mathcal{G}$ to the identity $e \in G$. Moreover, the exponential map can be shown (see e.g. [39]) to be an analytical diffeomorphism from an open neighborhood $\mathcal{V}(0)$ of $0 \in \mathcal{G}$ to a neighborhood $V(e) = \exp(\mathcal{V}(0))$ of $e \in G$. The differential

Table 1 Examples of Lie groups with their associated Lie algebra and exponential map, where $M(n,\mathbb{R})$ denotes the set of all $n \times n$ matrices with real entries and $S(n,\mathbb{R}) = \{u \in M(n,\mathbb{R})$ such that $u^T + u = 0$ and $Tr(u) = 0\}$

The Lie group G	The Lie Algebra	The exponential map
$(\mathbb{R}/\mathbb{Z},+)$	$\mathbb{R}$	$\exp(u) = u \bmod 1$, for $u \in \mathbb{R}$
$(GL(n,\mathbb{R}),\cdot)$	$M(n,\mathbb{R})$	$\exp(u) = \sum_{k=0}^{+\infty} \frac{u^k}{k!}$, for $u \in M(n,\mathbb{R})$
$(\mathrm{SO}(n,\mathbb{R}),\cdot)$	$S(n,\mathbb{R})$	$\exp(u) = \sum_{k=0}^{+\infty} \frac{u^k}{k!}$, for $u \in S(n,\mathbb{R})$

$d_u \exp$ of exp at $u \in \mathcal{G}$ is a linear function given by

$$\begin{cases} d_u \exp : \mathcal{G} \to & T_{\exp(u)} G \\ \qquad\qquad v \mapsto & \frac{d}{dt}|_{t=0} \exp(u + tv) \end{cases}$$

Moreover one can check that $d_{\exp(u)} L_{\exp(-u)} \circ d_u \exp$ is an endomorphism of $\mathcal{G}$ and by a slight abuse of notations, $d_{\exp(u)} L_{\exp(-u)} \circ d_u \exp$ is also denoted by $d_u \exp$. In what follows the application $d_u \exp$ is thus considered as a map $d_u \exp : \mathcal{G} \to \mathcal{G}$.

Since the exponential map is an application from $\mathcal{G}$ to G it will play a fundamental role to define the asymptotic normality of the estimator $\hat{h}_\epsilon$. Indeed, using that exp is a diffeomorphism from $\mathcal{V}(0)$ to $V(e)$, one has that if $\hat{h}_\epsilon \in V(e)$, then there exists a unique $\hat{u}_\epsilon \in \mathcal{V}(0)$ such that $\hat{h}_\epsilon = \exp(\hat{u}_\epsilon)$. Then, we finally arrive at the following definition:

Definition 4.3 The operation $\hat{h}_\epsilon \mapsto \exp^{-1}(\hat{h}_\epsilon)$ is defined as the projection of the estimator $\hat{h}_\epsilon$ onto the vector space $\mathcal{G}$.

In Table 1, a few illustrative examples are given to better explain how the estimates are mapped to the Lie Algebra in the case of the circle group $G = (\mathbb{R}/\mathbb{Z},+)$ of dimension $p = 1$, the group of all invertible $n \times n$ matrices with real entries $G = GL(n,\mathbb{R})$ of dimension $p = n^2$, and the special orthogonal group $G = \mathrm{SO}(n,\mathbb{R})$ of dimension $p = (n^2 - n)/2$.

One can see that, in the case of the circle group, then $d_u \exp(v) = v$ for all $u, v \in \mathbb{R}$. In contrast, in the case where $G = GL(p,\mathbb{R})$, the differential of exp at $u \in \mathcal{G} = M(n,\mathbb{R})$ is (with our slight abuse of notations for $d_u \exp$),

$$d_u \exp(v) = \sum_{k \geq 0} \frac{(-1)^k}{(k+1)!} (\mathrm{ad}_u)^k (v) \text{ for } v \in \mathcal{G}, \text{ where } \begin{cases} \mathrm{ad}_u : \mathcal{G} \to & \mathcal{G} \\ \qquad v \mapsto & uv - vu. \end{cases} \tag{4.1}$$

and $(\mathrm{ad}_u)^k (v)$ is defined recursively by $(\mathrm{ad}_u)^k (v) = \mathrm{ad}_u \left((\mathrm{ad}_u)^{(k-1)} (v) \right)$ and $(\mathrm{ad}_u)^0 (v) = v$. More generally, in the case of Abelian groups, the mapping $d_u \exp$ reduces to the identity on $\mathcal{G}$ and is therefore independent of u, whereas in the case of non-commutative groups, this differential depends on u. This will make a fundamental difference between Abelian and non-commutative groups for the interpretation of the asymptotic covariance matrix of $h^* = (h_1^*, \dots, h_J^*)$ in the model (2.2), see the following section for a precise definition.

4.2 Projection of the estimator

Our main idea is to re-express the criterion M_ϵ defined on G^J as a function $\tilde{M}_\epsilon$ defined on $\mathcal{G}^J$ using the exponential map. If G is a compact group, then the exponential map is surjective (see [39]), which means that for any $g \in G$, there exists $u \in \mathcal{G}$ such that $g = \exp(u)$. However, this map is not necessarily injective (think of the circle group for instance). To overcome this, we will use the fact that the exponential chart $\exp : \mathcal{V}(0) \to V(e)$ is a diffeomorphism, where $\mathcal{V}(0)$ is an open neighborhood of $0 \in \mathcal{G}$, and $V(e) = \exp(\mathcal{V}(0))$ is a neighborhood of $e \in G$.

Assumption 4.1 Let $\tilde{h} = (\tilde{h}_1 \dots, \tilde{h}_J)$ be in $\mathcal{A}_0$ with $\tilde{h}_1 = e$, such that the true parameters $(h_1^* \dots, h_J^*)$ belong to the neighborhood of $(\tilde{h}_1, \dots, \tilde{h}_J)$,

$$(h_1^*, \dots, h_J^*) \in V(\tilde{h}) = \{(h_1, \dots, h_J) \in \mathcal{A}_0, h_j \in \tilde{h}_j V(e)\}.$$

Then we can re-express our criteria on the vicinity of $\tilde{h}$ as functions on the vector space $\mathcal{G}^J$,

$$\tilde{M}(u_1, \dots, u_J) = M(\tilde{h}_1 \exp(u_1), \dots, \tilde{h}_J \exp(u_J)),$$

and

$$\tilde{M}_\epsilon(u_1, \dots, u_J) = M_\epsilon(\tilde{h}_1 \exp(u_1), \dots, \tilde{h}_J \exp(u_J)).$$

Both functions $\tilde{M}$ and $\tilde{M}_\epsilon$ are thus defined on the vector space $\mathcal{G}^J$ of dimension $J \times p$. Using the exponential chart, there exist $u^* = (u_1^*, \dots, u_J^*) \in \mathcal{V}(0)^J$ such that $h_1^* = \tilde{h}_1 \exp(u_1^*)$, $h_2^* = \tilde{h}_2 \exp(u_2^*), \dots, h_J^* = \tilde{h}_J \exp(u_J^*)$. Let $\mathcal{U}$ be a compact neighborhood of $0 \in \mathcal{G}$, with $\mathcal{U} \subset \mathcal{V}(0)$ and such that $u^* \in \mathcal{U}^J$. Note that Assumption 4.1 imply that the true parameters $u^* = (u_1^*, \dots, u_J^*)$ belong to the compact set

$$\mathcal{U}_0 = \left\{(u_1, \dots, u_J) \in \mathcal{U}^J,\ u_1 = 0\right\}.$$

Note that under Assumption 4.1, it follows that $h_1^* = e$ is fixed which corresponds to the identifiability condition (2.3). Then, we define,

$$\hat{u}_\epsilon = (\hat{u}_1, \dots, \hat{u}_J) = \arg\min_{u \in \mathcal{U}_0} \tilde{M}_\epsilon(u_1, \dots, u_J),$$

Arguing as in the proof of Theorem 3.1, we immediately have the following proposition:

Proposition 4.1 *Suppose that Assumption 4.1 holds. Assume that f^* is not shift-invariant and regular. Moreover, suppose that*

$$\lim_{\epsilon\to 0} \epsilon^2 \sum_{\pi \in \hat{G}_\epsilon} d_\pi^2 = 0$$

then $\hat{u}_\epsilon$ converges in probability to $u^* = (u_1^*, \ldots, u_J^*)$ as $\epsilon \to 0$.

As $\hat{u}_\epsilon$ belongs to a linear space, the problem of studying the asymptotic normality of $\hat{h}_\epsilon$ amounts to studying the asymptotic normality of $\hat{u}_\epsilon$ with the local exponential chart centered in $\tilde{h}$. In the case where G is a non commutative group, we will see that the asymptotic covariance matrix of $\hat{u}_\epsilon$ can be interpreted as a Riemannian metric i.e. as an inner product on the tangent space $T_{h^*}G^J$ that depends on the point $h^* \in G^J$ and the chosen coordinate chart (and thus on $\tilde{h}$). This is the standard fact for statistical models indexed by parameters belonging to a manifold and we will comment more on this in the next section, see e.g. [36] and the references therein.

In practice, we recommend the choice $\tilde{h}_j = e$ for all $j = 1, \ldots, J$ which corresponds a local parametrization of the group G around the identity e. Note that this local parametrization is also used to calculate $\hat{h}_\epsilon = \arg\min_{h\in\mathcal{A}_0} M_\epsilon(h)$ since once $\hat{u}_\epsilon$ has been computed then the choice $\tilde{h}_j = e$ for all $j = 1, \ldots, J$ automatically yields an expression for the value of $\hat{h}_\epsilon$. Such an choice is equivalent to suppose that the true parameter h^* belong to $V(e)$. Somehow, it restricts the study of asymptotic normality to the choice of the chart at the origin. Being able to do the estimation without assuming that the true parameter lies in the domain of a specific chart is an interesting topic for future work. One possibility would be to use the estimator $\hat{h}_\epsilon$ to define a random chart depending on this point. However, we believe that studying the asymptotic normality of the estimator on such a random chart is a difficult task that is beyond the scope of the paper.

4.3 Asymptotic normality of $\hat{u}_\epsilon$

Let us first introduce and recall some notations. Let F be a smooth manifold of finite dimension. Recall that the derivative of a function $m : G \to F$ at a point h in the direction $v \in T_hG$ will be written as $d_hm(v) \in T_{m(h)}F$, $d_hm : T_hG \to T_{m(h)}F$ is a linear map. The second derivative of a function $m : G \to F$ at a point h in the direction $v \in T_hG$ and $w \in T_hG$ will be written as $d_h^2m(v, w) \in T_{m(h)}F$, $d_h^2m : T_hG \times T_hG \to T_{m(h)}F$ is a bilinear map. By abusing notations, $d_{\exp(u)}L_{\exp(-u)}(d_u^2 \exp(v, w))$ is denoted by $d_u^2 \exp(v, w)$ which is bilinear map from $\mathcal{G} \times \mathcal{G} \to \mathcal{G}$.

Finally, it will be convenient to express our results for a given basis of $\mathcal{G}^J$. Let $e_1, \ldots, e_J$ be the canonical basis of $\mathbb{R}^J$, and let $x^1, \ldots, x^p$ be a basis of $\mathcal{G}$. Then, $\mathcal{G}^J$ can be viewed as the tensor product space of $\mathbb{R}^J$ and $\mathcal{G}$: $\mathcal{G}^J = \mathbb{R}^J \otimes \mathcal{G}$. For example, let v be in $\mathcal{G}$ and $j \in \{1 \ldots J\}$, $e_j \otimes v$ is the element $(0, \ldots, 0, v, 0, \ldots, 0)$ of $\mathcal{G}^J$ where v is the j^{th} coordinate. Then, a basis of $\mathcal{G}^J$ is $(e_j \otimes x^k)_{1\le j\le J, 1\le k\le p}$. With a such basis, we can identify the differential of order 1 at point $u \in \mathcal{G}^J$ of a function

$m : \mathcal{G}^J \to \mathbb{R}$ as a element $\nabla_u m$ of $\mathbb{R}^{Jd}$. Likewise, the differential of order 2 at point $u \in \mathcal{G}^J$ of a function $m : \mathcal{G}^J \to \mathbb{R}$ can be identified as as an element $\nabla_u^2 m$ of the space of $Jd \times Jd$ real matrices.

Note that the consistency of $\epsilon^{-1}(\hat{u}_\epsilon - u^*)$ should be actually understood for the vector $\hat{u}_\epsilon = (\hat{u}_2, \dots, \hat{u}_J) \in \mathcal{G}^{J-1}$ since the first component is fixed to $\hat{u}_1 = 0$ for identifiability reasons. By definition of $\hat{u}_\epsilon$, one has that

$$\nabla_{\hat{u}_\epsilon} \tilde{M}_\epsilon = 0.$$

Thus, the Taylor theorem with integral remainder states that,

$$0 = \epsilon^{-1} \nabla_{u^*} \tilde{M}_\epsilon + \int_0^1 \nabla^2_{\bar{u}_\epsilon(t)} \tilde{M}_\epsilon \epsilon^{-1}(\hat{u}_\epsilon - u^*) dt, \tag{4.2}$$

where for $t \in [0, 1]$

$$\bar{u}_\epsilon(t) = u^* + t(\hat{u}_\epsilon - u^*) \in \mathcal{U}_\epsilon = \{u \in \mathcal{U}_0, \|u - u^*\| \leq \|\hat{u}_\epsilon - u^*\|\}. \tag{4.3}$$

Then, let us introduce the following matrix norm:

Definition 4.4 For any matrix A of size $q \times q$ with complex entries $A_{k,\ell}$, $\|A\|$ denotes the norm

$$\|A\| = \sum_{k,\ell=1}^{q} |A_{k,\ell}|.$$

Under appropriate conditions, it will be shown that $\epsilon^{-1} \nabla_{u^*} \tilde{M}_\epsilon$ converges to a centered Gaussian variable $N(0, 4\Sigma/J^2)$ (see Proposition 4.2 for the expression of Σ), and that $\sup_{u \in \mathcal{U}_\epsilon} \|\nabla_u^2 \tilde{M}_\epsilon - 2\Sigma/J\|$ converges in probability to 0. Then, using Slutsky's lemma (see e.g. [41]), it will follow that $\epsilon^{-1}(\hat{u}_\epsilon - u^*)$ converges to $N(0, \Sigma^{-1})$ which is the main result of this section (see Theorem 4.1).

Obviously, to compute $\nabla_u \tilde{M}_\epsilon$ and $\nabla_u^2 \tilde{M}_\epsilon$, it will be necessary to compute the gradient and the Hessian of the function

$$\begin{cases} \tilde{\pi} : \mathcal{G} \to \mathrm{GL}(V) \\ u \mapsto \pi(\exp(u)), \end{cases}$$

where π is a finite dimensional representation of G in the vector space V. First remark that the differential of π at the identity e can be computed as

$$d_e \pi(u) = \lim_{t \to 0} \frac{1}{t} \left(\pi(\exp(tu)) - \pi(e)\right)$$

for $u \in \mathcal{G}$. Therefore, its differential at point $h \in G$ is given by

$$
\begin{aligned}
d_h \pi(u) &= \lim_{t \to 0} \frac{1}{t} \left(\pi(h \cdot \exp(tu)) - \pi(h) \right) \\
&= \pi(h) \lim_{t \to 0} \frac{1}{t} \left(\pi(\exp(tu)) - \pi(e) \right) \\
&= \pi(h) d_e \pi(u). \qquad (4.4)
\end{aligned}
$$

The above Eq. (4.4) shows that differentiating π just amounts to right multiplication by $d_e\pi(u)$. This fact is of fundamental importance to prove the ϵ^{-1} consistency of our estimators. Finally, by applying the chain rule of differentiation and the above results one has that for $u, v \in \mathcal{G}$

$$
d_u \tilde{\pi}(v) = \pi(\exp(u)) d_e \pi \left(d_u \exp(v) \right). \qquad (4.5)
$$

Then, the following results hold (proofs are deferred to the Appendix):

Proposition 4.2 *Assume that the conditions of Proposition 4.1 hold. Moreover, assume that for all $j = 2, \ldots, J$ and $k = 1, \ldots, p$*

$$
\lim_{\epsilon \to 0} \epsilon \sum_{\pi \in \hat{G}_\epsilon} d_\pi^2 \left\| d_e \pi \left(d_{u_j^*} \exp(x^k) \right) \right\|_{HS}^2 = 0, \qquad (4.6)
$$

$$
\sum_{\pi \in \hat{G}} d_\pi \left\| \pi(f^*) d_e \pi \left(d_{u_j^*} \exp(x^k) \right) \right\|_{HS}^2 < \infty, \qquad (4.7)
$$

where $x^1, \ldots, x^p$ is an arbitrary basis of $\mathcal{G}$. Then, as $\epsilon \to 0$

$$
\epsilon^{-1} \nabla_{u^*} \tilde{M}_\epsilon \to N \left(0, \frac{4}{J^2} \Sigma \right),
$$

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

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

and for $j_1 \neq j_2$ by

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

Proposition 4.3 *Assume that the conditions of Proposition 4.1 hold. Moreover, assume that for all* $1 \leq k_1, k_2 \leq p$

$$\lim_{\epsilon \to 0} \sup_{u \in \mathcal{U}} \left\{ \sum_{\pi \in \hat{G} \setminus \hat{G}_\epsilon} d_\pi \left\| \pi(f^*) \right\|_{HS}^2 \left\| d_e \pi \left(d_u \exp(x^{k_1}) \right) \right\|_{HS}^2 \right\} = 0 \quad (4.8)$$

$$\lim_{\epsilon \to 0} \sup_{u \in \mathcal{U}} \left\{ \sum_{\pi \in \hat{G} \setminus \hat{G}_\epsilon} d_\pi \left\| \pi(f^*) \right\|_{HS}^2 \left\| d_e \pi \left(d_u^2 \exp(x^{k_1}, x^{k_2}) \right) \right\|_{HS} \right\} = 0 \quad (4.9)$$

$$\lim_{\epsilon \to 0} \epsilon^2 \sup_{u \in \mathcal{U}} \left\{ \sum_{\pi \in \hat{G}_\epsilon} d_\pi^2 \left\| d_e \pi \left(d_u \exp(x^{k_1}) \right) \right\|_{HS}^2 \right\} = 0 \quad (4.10)$$

$$\lim_{\epsilon \to 0} \epsilon^2 \sup_{u \in \mathcal{U}} \left\{ \sum_{\pi \in \hat{G}_\epsilon} d_\pi^2 \left\| d_e \pi \left(d_u^2 \exp(x^{k_1}, x^{k_2}) \right) \right\|_{HS} \right\} = 0 \quad (4.11)$$

where $x^1, \ldots, x^p$ *is an arbitrary basis of* $\mathcal{G}$. *Then, as* $\epsilon \to 0$

$$\sup_{u \in \mathcal{U}_1} \| \nabla_u^2 \tilde{M}_\epsilon - \nabla_u^2 \tilde{M} \| \to 0 \text{ in probability} .$$

Finally, combining the above propositions we arrive at the following result

Theorem 4.1 *Under the assumptions of Propositions 4.1, 4.2 and 4.3,*

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

Intuitively, the conditions of Propositions 4.2 and 4.3 impose smoothness constraints of the reference template f^* and also give an idea of how choosing the set $\hat{G}_\epsilon$ with respect to the level of noise ϵ. The interpretation of these various conditions is easier in the case of Abelian groups, in particular when G is the multi-dimensional torus, and this will be discussed in the following sections.

Let us define $I(u^*) = \Sigma$. The matrix $I^{-1}(u^*)$ is the asymptotic covariance matrix of the estimator $\hat{u}_\epsilon$. As it depends on the point u^* (and thus of h^*), this matrix can be interpreted as a Riemannian metric on G. This is a classical result in mathematical statistics for random variables whose law is indexed by parameters belonging to a finite-dimensional manifold. In such settings, the Fisher information matrix is a Riemannian metric and lower bounds analogue to the classical Cramer–Rao bound for parameters in an Euclidean space can be derived (see e.g. [36] for a detailed review and discussion of this notion). When the exponential chart is centered at the point h^* ($\tilde{h} = h^*$), the covariance matrix can be rewritten as a tensor product of matrices:

$$\Sigma = \left(I_{J-1} - \frac{1}{J} \mathbb{I}_{J-1} \right) \otimes \text{Gramm}(\nabla f^*),$$

where I_{J-1} is the identity matrix of $\mathbb{R}^{J-1}$, $\mathbb{I}_{J-1}$ is the $(J-1)\times(J-1)$ matrix whose the elements are equal to 1, and $\text{Gramm}(\nabla f^*)$ is the $p \times p$ matrix defined as, for $1 \leq k_1, k_2 \leq p$,

$$\text{Gramm}(\nabla f^*)_{k_1,k_2} = \sum_{\pi\in\hat{G}} d_\pi \Re \times \left\langle \pi(f^*) d_e\pi \left(d_0 \exp(x^{k_1})\right), \pi(f^*) d_e\pi \left(d_0 \exp(x^{k_2})\right)\right\rangle_{HS}. \quad (4.12)$$

4.4 The special case of Abelian groups

In the particular case where G is an Abelian group, the conditions of Theorem 4.1 are much simpler and easier to interpret, which is due to the fact that the mapping $d_u \exp$ reduces to the identity on $\mathcal{G}$ i.e. $d_u \exp(v) = v$ for all u. Moreover, recall that in this case $d_\pi = 1$. Let $\tilde{f}^* : u \in \mathcal{G} \to f^*(gh \exp(u))$ be the function defined at the neighborhood of $gh \in G$. From our notations, the gradient of $\tilde{f}^*$ at point u is the following vector of $\mathbb{R}^p$,

$$\nabla_u \tilde{f}^* = \left(d_{gh\exp(u)} f^*(d_u \exp(x^k))\right)_{1\leq k\leq p}.$$

In the Abelian case one has that $d_u \exp(x^k) = x^k$, and thus $\nabla_u \tilde{f}^* = \left(d_{gh\exp(u)} f^*(x^k)\right)_{1\leq k\leq p}$ can be seen as a function of $gh\exp(u)$. By abusing notations, we denote by $d_{gh\exp(u)} f^*$ that function. Using the Fourier inverse formula, we get that

$$\tilde{f}^*(u) = \sum_\pi \pi(f^*)\pi(gh\exp(u)).$$

Then by differentiation, the derivative of the function $\tilde{f}^*$ at point u in the direction $v \in \mathcal{G}$ is,

$$d_{gh\exp(u)} f^*(v) = \sum_\pi \pi(f^*) d_e\pi(v)\pi(gh\exp(u)).$$

Consequently, the Parseval formula and the expression (4.12) of the matrix Gramm (∇f^*) imply that in the case of Abelian group the matrix Gramm (∇f^*) is given by

$$\text{Gramm}\left(\nabla f^*\right)_{k_1,k_2} = \int_G d_g f^*(x^{k_1})\overline{d_g f^*(x^{k_2})} dg.$$

If we suppose that $\hat{G}_\epsilon$ is a finite subset of $\hat{G}$, then one has the following result which is an immediate consequence of Theorem 4.1.

Proposition 4.4 *Let G be an Abelian group. Assume that Assumption 4.1 holds and that $f^* \in \mathbb{L}^2(G)$ is regular and not shift-invariant. Moreover suppose that for $k = 1, \ldots, p$*

$$\lim_{\epsilon \to 0} \epsilon^2 \#\{\hat{G}_\epsilon\} = 0 \tag{4.13}$$

$$\lim_{\epsilon \to 0} \epsilon \sum_{\pi \in \hat{G}_\epsilon} |d_e \pi(x^k)|^2 = 0 \tag{4.14}$$

$$\lim_{\epsilon \to 0} \sum_{\pi \in \hat{G} \backslash \hat{G}_\epsilon} \left|\pi(f^*)\right|^2 \left|d_e \pi(x^k)\right|^2 = 0, \tag{4.15}$$

where $x^1, \ldots, x^p$ is an arbitrary basis of $\mathcal{G}$. Then, as $\epsilon \to 0$

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

for $2 \leq j_1, j_2 \leq J$ and $1 \leq k_1, k_2 \leq p$ are given by

$$\Sigma_{(j_1,k_1)\times(j_1,k_2)} = \sum_{\pi \in \hat{G}} \left(1 - \frac{1}{J}\right) \left|\pi(f^*)\right|^2 \Re\left(d_e \pi \left(x^{k_1}\right) \overline{d_e \pi \left(x^{k_2}\right)}\right),$$

$$\Sigma_{(j_1,k_1)\times(j_2,k_2)} = \sum_{\pi \in \hat{G}} \frac{1}{J} \left|\pi(f^*)\right|^2 \Re\left(d_e \pi \left(x^{k_1}\right) \overline{d_e \pi \left(x^{k_2}\right)}\right) \quad for\ j_1 \neq j_2,$$

or,

$$\Sigma = \left(I_{J-1} - \frac{1}{J}\mathbb{I}_{J-1}\right) \otimes Gramm(\nabla f^*),$$

with $Gramm(\nabla f^*)_{k_1,k_2} = \sum_{\pi \in \hat{G}} |\pi(f^*)|^2 \Re\left(d_e \pi \left(x^{k_1}\right) \overline{d_e \pi \left(x^{k_2}\right)}\right) = \int_G d_g f^*(x^{k_1}) \overline{d_g f^*(x^{k_2})} dg$.

Thus, condition (4.15) (or (4.7)) state that the common shape is differentiable and its derivatives are square integrable on $\mathcal{G}$. Conditions (4.13)–(4.14) (or (4.6)–(4.10)–(4.11)) give some sufficient assumptions on the choice of $\hat{G}_\epsilon$ to guarantee the asymptotic normality of the estimators.

Note that the asymptotic covariance matrix Σ^{-1} does not depend on the point h^* since the parameter space for the shifts is a flat subset of $\mathbb{R}^p$ in the case of Abelian groups. The matrix Gramm(∇f^*) can be viewed as the scalar product of the partial derivatives $g \mapsto df^*(x^{k_1})$ and $g \mapsto df^*(x^{k_2})$. Then, the covariance matrix Σ is invertible if, and only if the matrix Gramm(∇f^*) is invertible. This means that the partial derivatives of the common shape have to be linearly independent in $\mathbb{L}^2(G)$. Moreover if the partial derivatives of f^* are orthogonal, the covariance matrix may be rewritten as a block diagonal matrix: for a fixed j, the estimators of the components $u^*_{j,1}, \ldots, u^*_{j,p}$ of the vector $u^*_j \in \mathbb{R}^p$ are asymptotically independent.

Then, let us study Proposition 4.4 in the case where $G = (\mathbb{R}/\mathbb{Z})^p$ which corresponds to the torus in dimension p. This case corresponds to the case of periodic functions defined on $[0, 1]^p$ for which $\hat{G} = \mathbb{Z}^p$ and $\mathcal{G} = \mathbb{R}^p$. Therefore, one retrieves the classical multi-dimensional Fourier decomposition of a function $f \in \mathbb{L}^2([0, 1]^p)$

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

where $e_\ell(x) = \pi(x) = e^{-i2\pi(\sum_{k=1}^p \ell^k x^k)}$ and $c_\ell(f) = \pi(f) = \int_{[0,1]^d} f(x) e_\ell(x) dx$. Note also that $d_e \pi(x^k) = -i2\pi \ell_k$. The condition that the function f^* is not shift invariant means that f^* cannot be rewritten as a function $m : (\mathbb{R}/\mathbb{Z})^{p-1} \to \mathbb{R}$, and $\mathbb{Z}^p$ is the minimal network of periodicity. This assumption implies that the partial derivatives are linearly independent. Moreover, one can check that f^* is not shift invariant if one of these two conditions holds:

1. there exist $\ell_1, \ldots, \ell_p \in \mathbb{Z}^p$ such that for all $r = 1 \ldots p, c_{\ell_r}(f^*) \neq 0$, and $\det\left(\{l_r^k\}_{1 \le r,k \le p}\right) = 1$,
2. there exist $\ell_{11}, \ldots, \ell_{1p} \in \mathbb{Z}^p$ and $\ell_{21}, \ldots, \ell_{2p} \in \mathbb{Z}^p$ such that for all $r = 1 \ldots p, i = 1, 2, c_{\ell_{ir}}(f^*) \neq 0$, and $\det\left(\{\ell_{1r}^k\}_{1 \le r,k \le p}\right)$ and $\det\left(\{\ell_{2r}^k\}_{1 \le r,k \le p}\right)$ are relatively prime.

Now, take

$$\hat{G}_\epsilon = \{(\ell^1, \ldots, \ell^p) \in \mathbb{Z}^p, \ |\ell^k| \le \ell_\epsilon \text{ for all } k = 1, \ldots, p\},$$

for some positive integer ℓ_ϵ. Then, the three conditions of Proposition 4.4 are satisfied if

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

The last above condition implies that the template function f^* should be at least differentiable. Also note that in this case, the two criterion $M_\epsilon(h)$ and $\tilde{M}_\epsilon(u)$ coincide for $h \in G^J$ and $u \in ([0, 1]^p)^J$. Since the condition $\epsilon \ell_\epsilon^{p+2} = o(1)$ implies that $\epsilon^2 \ell_\epsilon^p = o(1)$ if $\ell_\epsilon \to +\infty$ as $\epsilon \to 0$, we arrive at the following proposition:

Proposition 4.5 *Let $G = (\mathbb{R}/\mathbb{Z})^p$ and $f^* \in \mathbb{L}^2([0, 1]^p)$ be a periodic function. Assume that $h^* \in G^J$ or equivalently that $u^* \in ([0, 1]^p)^J$. Moreover, assume that f^* is regular and not shift-invariant, and suppose that*

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

then, as $\epsilon \to 0$

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

where the matrix Σ *simplifies to*

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

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

Proposition 4.5 shows that we retrieve the results in [13,42] obtained in related nonparametric regression models for one-dimensional shifted curves ($p = 1$) with sampled design points. However, with sampled design points, we have to assume the following stronger regularity on the common shape in order to estimate the Fourier coefficient (see [42])

$$\sum_{|\ell|\geq m} |c_l(f^*)| = o\left(m^{-p/2}\right).$$

4.5 The case of the special orthogonal group SO(3)

Now let us consider the case where $G = \mathrm{SO}(3) = \mathrm{SO}(3,\mathbb{R})$ (the special orthogonal group) to illustrate the influence of the geometry of non-commutative groups on the estimation of u^*. The group SO(3) is the space of 3×3 orthogonal matrices with determinant equal to one, and thus a Lie group of dimension $p = 3$.

First, let us describe the irreducible representations of this group. Let (e_1, e_2, e_3) be the canonical basis of $\mathbb{R}^3$. We define the rotation matrices $r_i(\alpha)$ $(i = 1, 2, 3)$ as the counter-clockwise rotation by an angle α about the e_i axes:

$$r_1(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{pmatrix}, \quad r_2(\alpha) = \begin{pmatrix} \cos(\alpha) & 0 & \sin(\alpha) \\ 0 & 1 & 0 \\ -\sin(\alpha) & 0 & \cos(\alpha) \end{pmatrix},$$

and

$$r_3(\alpha) = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It is a classical fact that element of SO(3) are parameterized by three Euler angles α, β, γ : for every $g \in \mathrm{SO}(3)$ there exist angles $\alpha, \gamma \in [0, 2\pi), \beta \in [0, \pi]$, such that $g = g(\alpha, \beta, \gamma) = r_3(\alpha) r_2(\beta) r_3(\gamma)$. This parameterization is not everywhere injective: for $\beta = 0$, the two parameters α and γ are only fixed up their sum. Using the Euler angles, the Haar measure of SO(3) is $dg = \frac{1}{8\pi^2}\sin(\beta) d\alpha d\beta d\gamma$, where $d\beta$ (resp. α, γ) is the Lebesgue measure on $[0, \pi]$ (resp. $[0, 2\pi)$).

Now let us define the representations of SO(3). For $m \in \mathbb{N}$, $\mathcal{H}_m$ denotes the space of all harmonic homogeneous polynomials on $\mathbb{R}^3$. The space $\mathcal{H}_m$ is a complex vector space of dimension $d_\pi = 2m + 1$. We may define a representation π_m of SO(3) in $\mathcal{H}_m$ as the linear endomorphism of $\mathcal{H}_m$ such that for all $g \in$ SO(3):

$$\begin{cases} \pi_m(g) : \mathcal{H}_m \to \mathcal{H}_m \\ \qquad\qquad h(x) \mapsto h(g^{-1}x), \quad \text{for} \quad x \in \mathbb{R}^3 \text{ and } h \in \mathcal{H}_m. \end{cases}$$

Hence, evaluating the matrix element of π_m for $g = g(\alpha, \beta, \gamma)$, we find that (see e.g. [9])

$$\pi^m_{k,l}(g) = \pi^m_{k,l}(g(\alpha, \beta, \gamma)) = e^{-ik\alpha} P^m_{k,l}(\cos(\beta)) e^{-il\gamma}, \quad -m \le k, l \le m,$$

where the functions $P^m_{k,l}(\cos(\beta))$ are generalizations of the associated Legendre functions and we refer to [9, Chapter 9, p. 295] for their exact expression. The representations $\pi_m, m \in \mathbb{N}$, are all irreducible unitary representation of SO(3).

Then any $f^* \in \mathbb{L}^2(\text{SO}(3))$ can be decomposed as (see e.g. [9])

$$\begin{aligned} f^*(g) = \sum_{m=0}^{+\infty} (2m+1) \sum_{k=-m}^{m} \sum_{\ell=-m}^{m} \pi^m_{k,l}(f^*) \pi^m_{k,l}(g), \text{ with } \pi^m_{k,l}(f^*) \\ = \int_{\text{SO}(3)} f^*(g) \overline{\pi^m_{l,k}(g)} dg. \end{aligned}$$

In this case, a possible choice for the set $\hat{G}_\epsilon$ defined in Theorem 3.1 is

$$\hat{G}_\epsilon = \{m = -m_\epsilon, \ldots, m_\epsilon\},$$

where m_ϵ is an appropriate cut-off parameter whose choice is given by the condition $\lim_{\epsilon \to 0} \epsilon^2 \sum_{\pi \in \hat{G}_\epsilon} d^2_\pi = \lim_{\epsilon \to 0} \epsilon^2 \sum_{|m| \le m_\epsilon} (2m + 1) = 0$ which is satisfied as soon as $m_\epsilon = o(\epsilon)$ as $\epsilon \to 0$.

Note that SO(3) is a Lie group of dimension 3, and that a vectorial basis of its associate Lie algebra is:

$$x^1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad x^2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad x^3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Alternatively, if one parametrizes an element of $g \in$ SO(3) using the exponential map and the Lie algebra as

$$g = g(\theta_1, \theta_2, \theta_3) = \exp \begin{pmatrix} 0 & -\theta_3 & \theta_2 \\ \theta_3 & 0 & -\theta_1 \\ -\theta_2 & \theta_1 & 0 \end{pmatrix}, \quad \text{for } (\theta_1, \theta_2, \theta_3) \in \mathbb{R}^3,$$

then the matrix elements $\pi^m_{k,l}(g)$ are given by the following formula

$$\pi^m_{k,l}(g(\theta_1,\theta_2,\theta_3)) = (-1)^{2m+k+l}\left[\frac{(m-k)!}{(m+k)!(m-l)!(m+l)!}\right] \times (\sin(\theta/2))^{k-l}\left(\frac{-\theta_1+i\theta_2}{\theta}\right)^{k-l}\left(\cos(\theta/2)-i\frac{\theta_3}{\theta}\sin(\theta/2)\right)^{k+l} \times P^{(k-l,k+l)}_{m-k}\left((1-\theta_3^2/\theta^2)\cos(\theta)+\theta_3^2/\theta^2\right),$$

where $\theta = \sqrt{\theta_1^2+\theta_2^2+\theta_3^2}$ and $P^{q,q'}_n(\cdot)$, for $(n,q,q')\in\mathbb{Z}^3$, are the Jacobi polynomials.

From Proposition 4.2 it follows that the entry $(j_1,k_1)\times(j_2,k_2)$ of Σ (the inverse of the asymptotic covariance matrix of $\hat{u}_\epsilon$) depends on $d_e\pi\left(d_{u^*_{j_1}}\exp(x^{k_1})\right)$ and $d_e\pi\left(d_{u^*_{j_2}}\exp(x^{k_2})\right)$. Hence, $\Sigma_{(j_1,k_1)\times(j_2,k_2)}$ depends on the parameter $u^*\in G^J$ to estimate only through the differential of the exponential map at the points $u^*_{j_1}$ and $u^*_{j_2}$ and in the directions x^{k_1} and x^{k_2} for $j_1,j_2=2,\ldots,J$ and $k_1,k_2=1,2,3$. From Eq. (4.1) it follows that for

$$u=\begin{pmatrix}0 & -\theta_3 & \theta_2\\ \theta_3 & 0 & -\theta_1\\ -\theta_2 & \theta_1 & 0\end{pmatrix}, \text{ with } (\theta_1,\theta_2,\theta_3)\in\mathbb{R}^3,$$

then

$$d_u\exp(x^\ell)=\sum_{k\geq 0}\frac{(-1)^k}{(k+1)!}(\mathrm{ad}_u)^k(x^\ell) \quad \text{for } \ell=1,2,3$$

Note that $\mathrm{ad}_u(x^1)=\begin{pmatrix}0 & -\theta_3 & \theta_2\\ \theta_3 & 0 & 0\\ -\theta_2 & 0 & 0\end{pmatrix}$, $\mathrm{ad}_u(x^2)=\begin{pmatrix}0 & -\theta_3 & 0\\ \theta_3 & 0 & -\theta_1\\ 0 & \theta_1 & 0\end{pmatrix}$, $\mathrm{ad}_u(x^3)=\begin{pmatrix}0 & 0 & \theta_2\\ 0 & 0 & -\theta_1\\ -\theta_2 & \theta_1 & 0\end{pmatrix}$, and that clearly $(\mathrm{ad}_u)^k(x^\ell)$ depends on $(\theta_1,\theta_2,\theta_3)$, for $k>1$ and $\ell=1,2,3$.

Hence, $d_u\exp(x^\ell)$ generally depends on the point u. Note that for some special values of u this may not be the case. For instance suppose that $\theta_2=\theta_3=0$ then $\mathrm{ad}_u(x^1)=0$, which implies that $d_u\exp(x^1)=0$ for any value of θ_1, but clearly $d_u\exp(x^2)$ and $d_u\exp(x^3)$ depend on θ_1. Hence, contrary to the case of Abelian groups (see Proposition 4.4), the matrix Σ (and thus the asymptotic covariance of $\hat{u}_\epsilon$) depends on the parameter u^* to be estimated. This example illustrates the influence of the geometry of non-commutative groups on the expression of the asymptotic covariance of u^* through the differential of the exponential map.

4.6 The difference between non-Abelian and Abelian groups

These results on the asymptotic normality of the estimators show that there exists a significant difference between semiparametric estimation on a linear Euclidean space and semiparametric 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^* (and also on the point $\tilde{h}$ used to define an appropriate projection of the estimator $\hat{h}_\epsilon$ on a vector space). Hence, this matrix can be interpreted as a Riemannian metric on G which depends on the point h^*. 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 Riemannian metric and lower bounds analogue to the classical Cramer–Rao bound for parameters in an Euclidean space can be derived (see e.g. [36]). If G is supposed to be an Abelian group, then the asymptotic covariance matrix of the estimator is still a Riemannian metric but its expression does not depend on the point h^* since the parameter space G for the shifts is a flat manifold.

4.7 The estimation of the common shape

The estimation of the parameter $h_1^*, \ldots, h_J^*$ allows us to align the signals. Therefore, it is desirable to be able to define an estimator of the common shape f^*. Our estimation method suggests to use the following estimators of the coefficients $\pi(f^*), \pi \in \hat{G}$:

$$\hat{\pi}_\epsilon(Y) = \frac{1}{J}\sum_{j=1}^{J} \pi(Y_j)\pi(\hat{h}_{j,\epsilon}).$$

Using the Peter–Weyl Theorem (2.1), one can then take the following estimator of the common shape,

$$\hat{f}_\epsilon(g) = \sum_{\pi \in \hat{G}_\epsilon} d_\pi \operatorname{Tr}\left(\pi(g)\hat{\pi}_\epsilon(Y)\right).$$

In order to simplify the study of this estimator, we restrict it to the case where $G = (\mathbb{R}/\mathbb{Z})^p$, the multidimensional torus. In this case $\hat{G} = \mathbb{Z}^p$, and a possible choice for $\hat{G}_\epsilon$ is to take

$$\hat{G}_\epsilon = \{\ell \in \mathbb{Z}^p,\ |\ell|_\infty \leq \ell_\epsilon\}$$

for some frequency cutoff parameter $\ell_\epsilon > 0$, where $|\ell|_\infty = \max(|\ell^k|,\ 1 \leq k \leq p)$ for $\ell \in \mathbb{Z}^p$. To study the convergence of the estimator $\hat{f}_\epsilon$, let us introduce the following smoothness class

$$\mathcal{F}(s, M) = \left\{ f : (\mathbb{R}/\mathbb{Z})^p \to \mathbb{R},\ \sum_{\ell \in \mathbb{Z}^p} (1 + |\ell|)^{2s} |c_l(f^*)|^2 < M \right\}$$

for some $s \geq p/2$ and some constant $M > 0$, where $|\ell|^2 = |\ell^1|^2 + \cdots + |\ell^p|^2$. The parameter s can be thought as a parameter which controls the smoothness of the functions in the above ellipsoid. In the case $p = 1$, it is well known that such ellipsoids can be identified with periodic Sobolev classes (see e.g. [32]), and the problem of estimating functions lying in such sets has been widely studied in nonparametric regression (see e.g. [37]).

Proposition 4.6 *Assume that the conditions of Proposition 4.5 hold. Moreover, assume that the common shape f^* belongs to $\mathcal{F}(s, M)$ for $s \geq p/2$ and some constant $M > 0$. Then, as $\epsilon \to 0$*

$$MISE_{f^*}(\hat{f}_\epsilon) = \mathbb{E}\int_G \left(\hat{f}_\epsilon(g) - f^*(g)\right)^2 dg = \mathcal{O}\left(\frac{1}{\ell_\epsilon^{2s}} + \epsilon^2 \ell_\epsilon^p\right).$$

Moreover, if $\ell_\epsilon \sim \epsilon^{-2/(2s+p)}$, we have $MISE_{f^*}(\hat{f}_n) = \mathcal{O}\left(\epsilon^{4s/(2s+p)}\right)$.

The above theorem shows that aligning the noisy images Y_j using the estimated deformations $\hat{h}_{j,\epsilon}$ yields a consistent estimate of the common shape f^*. Note that if $\ell_\epsilon \sim \epsilon^{-2/(2s+p)}$, then one retrieves the optimal rate of convergence in the minimax sense for standard nonparametric regression problems [37]. Since $h_1^* = e$, one could simply denoise the first image to estimate the common shape f^* with the same asymptotic rate of convergence. Nevertheless using the above estimate by aligning all the images reduces the variance from ϵ^2 to ϵ^2/J which yields important improvement in practice. Moreover aligning images is a fundamental task in image registration that is commonly done to estimate a common shape.

However, this estimator is not adaptive in the sense that the choice of ℓ_ϵ depends on the unknown smoothness s of f^*. An interesting extension of this work would be to investigate data-based choices of ℓ_ϵ to estimate f^* in an optimal way, but we leave this problem open for future work. Also, we have only investigated the case $G = (\mathbb{R}/\mathbb{Z})^p$. However, at the price of additional technicalities, it is also possible to define a notion of Sobolev ellipsoid for functions defined on other compact groups (see e.g. [23] for an example with $G = \mathrm{SO}(3)$) and then to obtain analog results.

5 Efficiency of the estimators

In this section, we discuss the optimality of the covariance matrix Σ^{-1} of the estimators given in Theorem 4.1 from the point of view of asymptotic efficiency in locally asymptotic normal (LAN) semi-parametric models, see [31] for a detailed exposition of this concept. We mainly discuss the efficiency of the estimators for Abelian groups and more particularly for the special case of the torus in dimension p. To the best of our knowledge, asymptotic efficiency in LAN models has been mainly developed for the estimation of parameters belonging to a linear space (see [31] and references therein). Extending the notion of LAN models and efficiency for parameters lying in a non-commutative Lie group remains a challenge that is beyond the scope of this paper. Nevertheless, in what follows, we have tried to keep general notations (using the exponential map and its derivative) to better highlight how the results could be generalized to non-commutative Lie groups.

In linear spaces, the concept of efficiency is based on what is commonly referred to as the convolution theorem (see [31]). The two key hypothesis of this theorem are the LAN property of the model and the differentiability of the parameter of interest. Hereafter, we restate this approach into the framework of Abelian Lie groups.

5.1 The LAN property on compact Abelian Lie groups

In the rest of this section, G is supposed to be Abelian. Let $\mathbb{P}_\theta^{(\epsilon)}$ denote the distribution of the model (2.2) for the parameter $\theta = (h, f) \in \mathcal{A}_0 \times \mathcal{F}$, where $\mathcal{F}$ is the set of real-valued continuous function on G. The family $\left(\mathbb{P}_\theta^{(\epsilon)}\right)$ is LAN at point θ indexed by a linear space $\mathcal{T}$ endowed with an inner product $< \cdot, \cdot >_{\mathcal{T}}$ and a norm $\| \cdot \|_{\mathcal{T}}$ such that for every $t \in \mathcal{T}$ there exists a sequence $\left(\mathbb{P}_{\theta_\epsilon(t)}^{(\epsilon)}\right)$ of probability measures with $\theta_\epsilon(t) \in \mathcal{A}_0 \times \mathcal{F}$ such that the log-likelihood ratio Λ_ϵ for θ and $\theta_\epsilon(t)$ admits the following representation:

$$\Lambda_\epsilon\left(\theta_\epsilon(t), \theta\right) = \log \frac{d\mathbb{P}_{\theta_\epsilon(t)}^{(\epsilon)}}{d\mathbb{P}_\theta^{(\epsilon)}} = \Delta_\epsilon(t) - \frac{1}{2}\|t\|_{\mathcal{T}}^2 + o_{\mathbb{P}_\theta^{(\epsilon)}}(1),$$

where the process $\Delta_\epsilon(t)$ is linear in t and converges in $\mathbb{P}_\theta^{(\epsilon)}$-distribution to $\mathcal{N}\big(0, \|t(\theta)\|_{\mathcal{T}}^2\big)$, for some $t(\theta) \in \mathcal{T}$ depending on θ. The tangent space $\mathcal{T}$ is used to parameterize the neighborhoods of the point $\theta^* = (h^*, f^*)$. In the context of Lie group, the LAN property must be true for every local map in the neighborhood of h^*. Using Assumption 4.1, a convenient choice in this setting is

$$\theta_\epsilon^*(t) = \Big((\tilde{h}_2 \exp(u_2^* + \epsilon u_2), \ldots, \tilde{h}_J \exp(u_J^* + \epsilon u_J)),\, f^* + \epsilon f\Big)$$

where $t = (u_2, \ldots, u_J, f) \in \mathcal{T} = \mathcal{G}^{J-1} \times \mathcal{F}$. Thanks to the Girsanov formula (see [19, Appendix 2]), the log-likelihood ratio Λ_ϵ is (under Assumption 4.1)

$$\begin{aligned}\Lambda_\epsilon\left(\theta_\epsilon^*(t), \theta^*\right) = &\int_G f(g)dW_1(g) + \frac{1}{\epsilon}\sum_{j=2}^{J}\int_G (f^* + \epsilon f)(g\tilde{h}_j^{-1}\exp(-u_j^* - \epsilon u_j)) \\ &- f^*(g\tilde{h}_j^{-1}\exp(-u_j^*))dW_j(g) \\ &- \frac{1}{2}\int_G f(g)^2 dg - \frac{1}{2\epsilon^2}\sum_{j=2} J \int_G \Big\{(f^* + \epsilon f)(g\tilde{h}_j^{-1}\exp(-u_j^* - \epsilon u_j)) \\ &- f^*(g\tilde{h}_j^{-1}\exp(-u_j^*)\Big\}^2 dg.\end{aligned}$$

Note that the Girsanov's formula in [19]] is not stated over a Lie group. We only apply it in the case of an Abelian group which can considered as an Euclidean space by some abuse of notations. Thus, if the common shape f^* is continuously differentiable and using the uniform continuity of the functions f^* and f on the compact group G, the following proposition holds (the proof is omitted as it is a direct consequence of the above formula for the log-likelihood ratio):

Proposition 5.1 *Assume that G is a compact Abelian Lie group. Let $\langle \cdot, \cdot \rangle_{\mathbb{L}^2(G)}$ be the standard inner-product in $\mathbb{L}^2(G)$. Assume that the function f^* is not shift-invariant and differentiable with a continuous tangent map function df^* such that the matrix,*

$$\mathrm{Gramm}(\nabla f^*) = \left(\left\langle df^*(x^{k_1}), df^*(x^{k_2}) \right\rangle_{\mathbb{L}^2(G)} \right)_{1 \le k_1, k_2 \le p},$$

is invertible. Then the model $\left(\mathbb{P}_\theta^{(\epsilon)}\right)$ is LAN at point $\theta^* = (h^*, f^*)$ indexed by the tangent space $\mathcal{T} = \mathcal{G}^{J-1} \times \mathcal{F}$ with,

$$\Delta_\epsilon(t) = \int_G f(g) dW_1(g) + \sum_{j=2}^{J} \int_G f(g\tilde{h}_j^{-1} \exp(-u_j^*)) - d_{g\tilde{h}_j^{-1} \exp(-u_j^*)} f^*(u_j) dW_j(g).$$

Moreover, the tangent space $\mathcal{T} = \mathcal{G}^{J-1} \times \mathcal{F}$ is a vector space when endowed with the following inner-product:

$$\langle t, t' \rangle_{\mathcal{T}} = \left\langle f, f' \right\rangle_{\mathbb{L}^2(G)} + \sum_{j=2}^{J} \left\langle f - d_{g \exp(0)} f^*(u_j), f' - d_{g \exp(0)} f^*(u'_j) \right\rangle_{\mathbb{L}^2(G)},$$

and the closure of $\mathcal{T} = \mathcal{G}^{J-1} \times \mathbb{L}^2(G)$ is a Hilbert space.

Let T_ϵ be any estimator of h^* and let us denote by $\hat{u}_\epsilon \in \mathcal{G}^{J-1}$ its local coordinates in the Lie Algebra $\mathcal{G}$ via the exponential map. Using the centered process $\Delta_\epsilon(t)$, we can characterize the class of asymptotically linear estimators in the sense of the following definition:

Definition 5.1 An estimator T_ϵ of h^* is said to be asymptotically linear if and only if there exists $t \in \mathcal{T}$ such that

$$\epsilon^{-1}(\hat{u}_\epsilon - u^*) = \Delta_\epsilon(t) + o_{\mathbb{P}_{\theta^*}^{(\epsilon)}}(1),$$

where $\Delta_\epsilon(t) = \left(\Delta_\epsilon(t_{1,2}), \ldots, \Delta_\epsilon(t_{p,2}), \ldots, \Delta_\epsilon(t_{p,J})\right)$ is a multi-variate centered process linear in $t = (t_{1,2}, \ldots, t_{p,J}) \in \mathcal{T}^{p(J-1)}$.

5.2 The differentiability of the estimation parameter

Let us now consider the special case where $G = (\mathbb{R}/\mathbb{Z})^p$ is the torus in dimension p. Denote the parameter to estimate relative to the distribution $\mathbb{P}_\theta^{(\epsilon)}$ by

$$\nu_\epsilon \left(\mathbb{P}_\theta^{(\epsilon)}\right) := (v_2, \ldots, v_J) \in \mathcal{G}^{J-1},$$

where $\theta = ((\tilde{h}_2 \exp(v_2), \ldots, \tilde{h}_J \exp(v_J)), f) \in \mathcal{A}_0 \times \mathcal{F}$. This parameter is differentiable relative to the tangent space $\mathcal{T}$ in the sense that

$$\lim_{\epsilon \to 0} \epsilon^{-1} \left\{ \nu_\epsilon \left(\mathbb{P}_{\theta_\epsilon(t)}^{(\epsilon)}\right) - \nu_\epsilon \left(\mathbb{P}_\theta^{(\epsilon)}\right) \right\} = (u_2, \ldots, u_J), \quad t \in \mathcal{T},$$

and thus there exists a continuous linear map $\dot{\nu}$ from $\mathcal{T}^{p(J-1)}$ to $\mathcal{G}^{p(J-1)}$. According to the Riesz representation theorem, there exist $p(J-1)$ vectors $(\dot{\nu}_{k,j})_{1 \le k \le p, 2 \le j \le J}$ of the closure of $\mathcal{T}$ such that:

$$\forall t \in \mathcal{T}, \quad \langle \dot{\nu}_{k,j}, t \rangle_{\mathcal{T}} = u_j^k,$$

where $t = (u, f) \in \mathcal{T}$ and $u = (u_2^1, \ldots, u_2^p, \ldots u_J^p) \in \mathcal{G}^{J-1}$.

Then, the tangents vectors $\dot{\nu}_{k,j} = (\dot{u}_{k,j}, \dot{f}_{k,j}) \in \mathcal{G}^{J-1} \times \mathcal{F}$ are such that,

$$\dot{f}_{k,j} = \frac{1}{J} \sum_{j'=2}^{J} df^* \left((\dot{u}_{k,j})_{j'}\right), \tag{5.1}$$

and the vectors $(\dot{u}_{k,j})_{k,j}$ are the solutions of the equation,

$$\left(I_{J-1} - \frac{1}{J}\mathbb{I}_{J-1}\right) \otimes \text{Gramm}(\nabla f^*)(\dot{u}_{1,2}, \ldots, \dot{u}_{p,2}, \ldots, \dot{u}_{p,J}) = I_{p(J-1)}. \tag{5.2}$$

Since the process $\Delta_\epsilon(t)$ is linear with t, a consequence of the Proposition 5.3 of [31] allows us to link the notions of asymptotic linearity and asymptotic efficiency.

Proposition 5.2 *Let T_ϵ be an asymptotic linear estimator of h^* with associate centered process*

$$\Delta_\epsilon(t) = \left(\Delta_\epsilon(t_{1,2}), \ldots, \Delta_\epsilon(t_{p,2}), \ldots, \Delta_\epsilon(t_{p,J})\right).$$

T_ϵ is asymptotically regular and efficient if and only if,

$$\forall j = 2 \ldots J, \quad \forall k = 1 \ldots p, \qquad t_{p,j} = \dot{\nu}_{k,j}.$$

Let $\xi_j \in \mathbb{R}^d$ be the centered random vectors defined as,

$$\xi_j = \int_G \nabla_{g\tilde{h}_j^{-1}\exp(-u_j^*)} f^* dW_j(g\exp(u)), \quad j = 1, \dots J.$$

Let T_ϵ be an asymptotic linear estimator of h^* and denote by $\hat{u}_\epsilon \in \mathcal{G}^{J-1}$ its local coordinates in the Lie Algebra $\mathcal{G}$ via the exponential map. Using Propositions 5.1 and 5.2, Eqs. (5.1) and (5.2), it follows that the estimator T_ϵ is asymptotically efficient if and only if,

$$\epsilon^{-1}(\hat{u}_\epsilon - u^*) = \left(I_{J-1} - \frac{1}{J}\mathbb{I}_{J-1}\right)^{-1} \otimes \text{Gramm}(\nabla f^*)^{-1} \left(-\sum_{j=2}^{J} e_j \otimes \xi_j + \frac{1}{J}\sum_{j=1}^{J} 1_{J-1} \otimes \xi_j\right) + o_{\mathbb{P}_{\theta^*}^{(\epsilon)}}(1), \tag{5.3}$$

where $1_{J-1} = (1, \dots, 1)^T \in \mathbb{R}^{J-1}$. The following proposition (whose proof is deferred to the Appendix) finally shows that in the case of $G = (\mathbb{R}/\mathbb{Z})^p$ (the torus in dimension p) then the estimator is asymptotically efficient.

Proposition 5.3 *Suppose that $G = (\mathbb{R}/\mathbb{Z})^p$. Assume that the function f^* is not shift-invariant and differentiable with a continuous tangent map function df^* such that the matrix,*

$$\text{Gramm}(\nabla f^*) = \left(\left\langle df^*(x^{k_1}), df^*(x^{k_2})\right\rangle_{\mathbb{L}^2(G)}\right)_{1\le k_1,k_2\le p},$$

is invertible. Then, the estimator $\hat{u}_\epsilon$ is asymptotically efficient.

6 Numerical simulations and some illustrative examples

6.1 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 if G is not a linear space, as in this case one may compute estimates which do not belong to the search space (take for instance the problem of optimizing a function over the space of positive definite matrices). Finding minima of functions defined on a Lie group is generally done by reformulating the problem as an optimization problem on the Lie algebra of G. Such an approach has been for instance proposed in [33] to formulate a general Newton optimization method over Lie groups. Here, we propose to find a minima $\hat{u}_\epsilon$ of $\tilde{M}_\epsilon(u)$ for $u \in \mathcal{G}^J$, and then to take $\hat{h}_\epsilon = \exp(\hat{u}_\epsilon)$. Since the expression of the gradient of $\tilde{M}_\epsilon(u)$ is available in a closed form, a gradient ! descent algorithm

Fig. 1 Shepp–Logan phantom image of size 100 × 100 used as the template function f^*

with an adaptive step can be easily implemented. More precisely the algorithm is composed of the following steps:

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**

$$\gamma_m = \gamma_m/\kappa, \quad \text{and } u^{new} = u^m - \gamma_m \nabla_{u^m} \tilde{M}_\epsilon, \quad \text{and } M(m+1) = \tilde{M}_\epsilon(u^{new})$$

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 the choice of a basis for the product space $\mathcal{G}^J$ can be arbitrary and is left to the statistician. Moreover, to satisfy the identifiability constraints the first p components of u^m are held fixed to zero at each iteration m.

6.2 Registration of translated 2D images

As an illustrative example, the above described algorithm has been implemented for the registration of translated 2D images. All simulations have been carried out with Matlab, and the chosen template f^* is the Shepp–Logan phantom image (see [21]) of size $N \times N$ with $N = 100$ displayed in Fig. 1. Data can be generated by translating this image and adding Gaussian noise to each pixel value:

$$Y^j(i_1, i_2) = f\left(\frac{i_1}{N} - h_j^1, \frac{i_2}{N} - h_j^2\right) + \sigma z_j(i_1, i_2), \quad 1 \leq i_1, i_2 \leq N, j = 1, \ldots, J \tag{6.1}$$

Table 2 Average and standard deviation (in brackets) of the estimators $\hat{h}_j = (\hat{h}_j^1, \hat{h}_j^2)$ over $M = 100$ simulations

	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$
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)

The bold numbers represent the true values of the parameters (h_j^1, h_j^2)

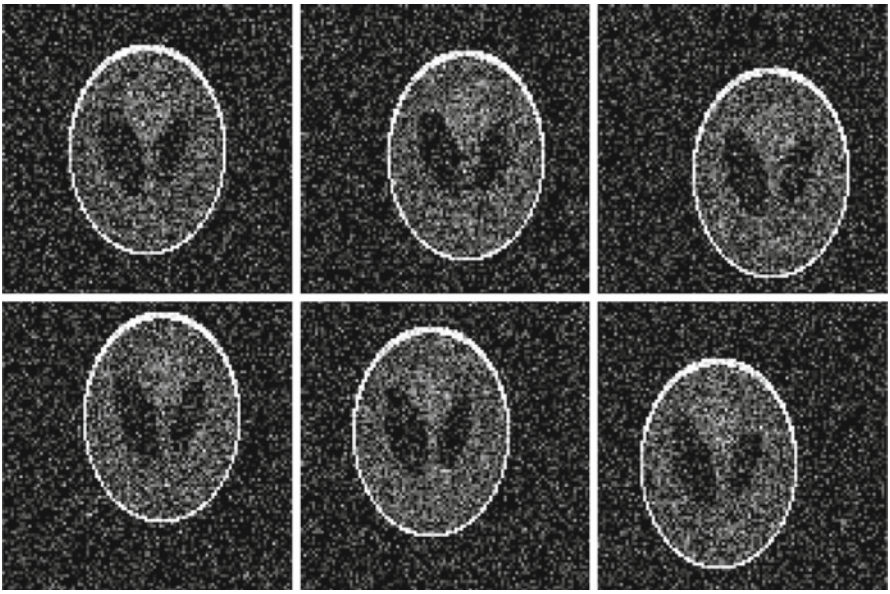

Fig. 2 A typical simulation run for $J = 6$ images generated from the model (6.1)

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. Note that in the above model, the image f is considered to be periodic function on the square $[0, 1]^2$ so that it is also defined outside the range of pixels $1, \ldots, N \times 1, \ldots, N$. One could argue that the sampled data model (6.1) does not truly correspond to the white noise model (2.2). However, as previously explained the white noise model is a useful theoretical tool to study the properties of statistical procedures in image analysis. Moreover, there exists a correspondence between these two models in the sense that they are asymptotically equivalent if $\epsilon = \frac{\sigma}{N}$ (see [3]).

We have repeated $M = 100$ simulations with $J = 6$ noisy images simulated from the model (6.1). The various values taken for the translation parameters are the bold numbers given in Table 2. A typical example of a simulation run is shown in Fig. 2 (note that the signal-to-noise ratio is quite low).

Here, $G = [0, 1] \times [0, 1]$ and the 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:

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

for $u = (u_1^1, u_1^2, \dots, u_J^1, u_J^2)$, and where the $y^j_{\ell_1,\ell_2}$'s are the empirical Fourier coefficients of the image Y^j. Moreover, if $(x_1^1, x_1^2, \dots, x_J^1, x_J^2)$ 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

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

As discussed in Sect. 4.4 and according to Proposition 4.5, the smoothing parameter ℓ_ϵ should be chosen such that $\epsilon\ell_\epsilon^4 = o(1)$. Because of the equivalence between models (6.1) and (2.2) given by $\epsilon = \frac{\sigma}{N}$, this condition becomes

$$\ell_\epsilon = \ell_N = o(N^{1/4}).$$

Hence, since $N = 100$, the above condition suggests to take $\ell_N \le 100^{1/4} \approx 3.16$. However the choice of ℓ_ϵ is a delicate model selection problem. The condition $\ell_\epsilon = o(N^{1/4})$ is a purely asymptotic result but the choice $\ell_\epsilon \le 100^{1/4}$ is rather arbitrary since for many functions it can be important to select harmonics above the third one with much care. In, Table 2, 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 quite satisfactory as averages are close to the true values and standard deviations are small.

We have also conducted the same simulations but with a slightly different model. Indeed in real applications, images have a similar shape but are typically not the deformation of *exactly* the same image since some portions of an image may not be common to all the other images. In Fig. 2 an example of such a data set without noise is displayed. These images have created by taking only a portion of the previous images generated by translations of the Shepp–Logan image. Then, $M = 100$ images are generated by adding Gaussian noise to these $J = 6$ images. The values of the "best translation parameters" to align the images in Fig. 2 are the same as in the previous simulation, and results are reported in Table 3. Again, the estimations are very satisfactory and they demonstrate somehow the robustness of this approach with respect to some deviation from the ideal model (6.1).

Table 3 Average and standard deviation (in brackets) of the estimators $\hat{h}_j = (\hat{h}_j^1, \hat{h}_j^2)$ over $M = 100$ simulations generated by adding noise to the images shown in Fig. 3

	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$
h_j^1	**0.07**	**0.1**	**0.05**	**−0.05**	**−0.08**
$\hat{h}_j^1$	0.0692 (0.0016)	0.1011 (0.0017)	0.0503 (0.0019)	−0.0473 (0.0017)	−0.0758 (0.0018)
h_j^2	**0.02**	**0.08**	**−0.10**	**−0.05**	**0.06**
$\hat{h}_j^2$	0.0251 (0.0032)	0.0830 (0.0032)	−0.0900 (0.0033)	−0.0434 (0.0034)	0.0684 (0.0037)

The bold numbers represent the true values of the parameters (h_j^1, h_j^2)

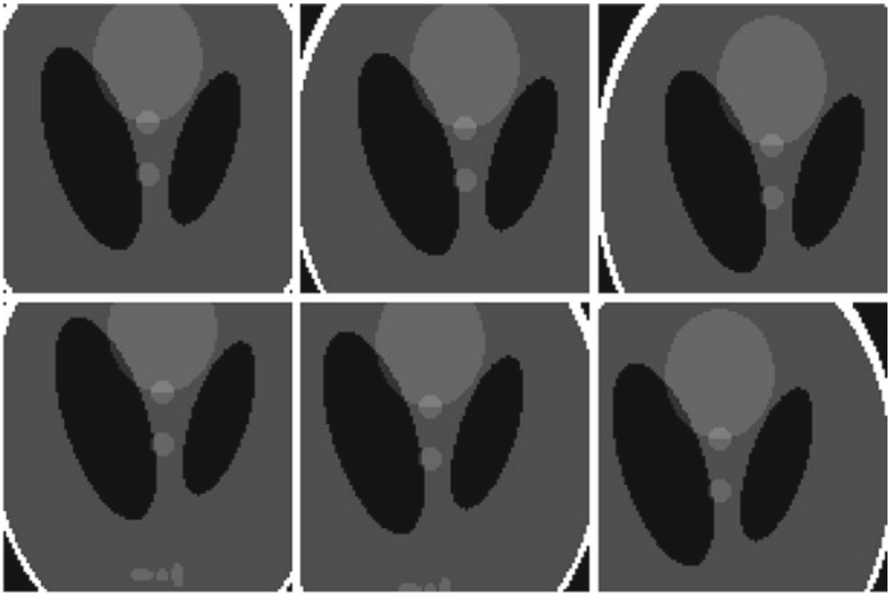

Fig. 3 A more realistic situation: the images look similar, but they are not exactly translated versions of the same image. Noise is then added to these $J = 6$ images to create a second data set

7 Some extensions of the simple shift model

7.1 The general case

Return now to the general case where the space $\mathcal{X}$ is not necessarily equal to G. A possible extension of the empirical matching criterion (3.3) is to take

$$M_\epsilon(h_1, \ldots, h_J) = \frac{1}{J} \sum_{j=1}^{J} \left\| \hat{f}_j \circ L_{h_j} - \frac{1}{J} \sum_{j'=1}^{J} \hat{f}_{j'} \circ L_{h_{j'}} \right\|_{\mathbb{L}^2(\mathcal{X})}^2 \qquad (7.1)$$

where $L_h : x \in \mathcal{X} \to hx \in \mathcal{X}$ is the action of $h \in G$ on $\mathcal{X}$, and $\hat{f}_1, \ldots, \hat{f}_J$ represent some estimators of the functions $f_1, \ldots, f_J \in \mathbb{L}^2(\mathcal{X})$ obtained by smoothing the noisy images $Y_1, \ldots, Y_J$. In the case $\mathcal{X} = G$ these estimators have been obtained via low-pass filtering in the Fourier domain, while the deformations of the functions $\hat{f}_j$ by the transformations L_{h_j} are easily implemented via a simple multiplication of their Fourier coefficients. In the next section, we show that for some particular choices of $\mathcal{X}$ and G, a similar analysis based on Fourier transforms can still be investigated. Note that in a future work, we also plan to study the criterion (7.1) in a more general setting using other smoothing and deformation methods than those based on Fourier analysis.

7.2 Registration of spherical images

Consider the problem of estimating three-dimensional rotations between images defined on the three-dimensional unit sphere $\mathbb{S}^2 = \{x \in \mathbb{R}^3, \ \|x\| = 1\}$. In many applications, data can be organized as spherical images. 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 displacement from such images (see [29]). Data collected on the sphere can also be found in other applications such as molecular biology or crystallography (see [29,40] and the references therein).

Obviously such data do not correspond exactly to the simple shift model on group (2.2) as spherical images are defined on $\mathcal{X} = \mathbb{S}^2$ while the shifts parameters belong the special orthogonal group $G = SO(3)$. However, a matching criterion similar to the one defined in Eq. (3.2) 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 $\phi \in [0, 2\pi[$. For $x = x(\theta, \phi)$ let us denote by dx the measure $dx = d\phi \sin(\theta) d\theta$, where $d\phi$ and $d\theta$ are the Lebesgue measures on $[0, 2\pi]$ and $[0, \pi]$. Then any $f \in \mathbb{L}^2(\mathbb{S}^2)$ (the space of square integrable functions on $\mathbb{S}^2$ with respect to dx) can be decomposed as (see e.g. [9])

$$f(x) = \sum_{\ell=0}^{+\infty} \sum_{m=-\ell}^{\ell} c_\ell^m(f) Y_\ell^m(x),$$

with $c_\ell^m(f) = \int_{\mathbb{S}^2} f(x) \overline{Y_\ell^m(x)} dx = \int_0^\pi \int_0^{2\pi} f(\theta, \phi) \overline{Y_\ell^m(\theta, \phi)} d\phi \sin(\theta) d\theta$, and where the functions $(Y_\ell^m, \ell \in \mathbb{N}, m = -\ell, \ldots, \ell)$ are the usual spherical harmonics which form an orthonormal basis of $(\mathbb{L}^2(\mathbb{S}^2), dx)$, and are given by

$$Y_\ell^m(\theta, \phi) = \Gamma_{\ell,m} P_\ell^m(\cos(\theta)) e^{im\phi},$$

where the P_ℓ^m's are the associated Legendre functions and $\Gamma_{\ell,m}$ are normalizing constants to satisfy the orthonormality conditions. For further details on spherical harmonics we refer to [9].

Now, to each $g \in G = \mathrm{SO}(3)$, one can associate a linear mapping $\pi(g)$ which acts on $\mathbb{L}^2(\mathbb{S}^2)$ by $\pi(g)f(x) = f(g^{-1}x)$. This defines a left regular representation of SO(3) on the vector space $\mathbb{L}^2(\mathbb{S}^2)$. Moreover, one has that $\hat{G} = \mathbb{N}$, and the invariant subspaces are the vector spaces $\{V_\ell, \ell \in \mathbb{N}\}$ defined as the set of functions spanned by the spherical harmonics at frequency ℓ, i.e.

$$V_\ell = \mathbf{Vect}\{Y_\ell^m, m = -\ell, \ldots, \ell\}.$$

Then, using the decomposition of a representation into a direct sum of irreducible representations, and if we identify the irreducible representations of SO(3) as $(2\ell + 1) \times (2\ell + 1)$ matrices π_ℓ (with respect to the above basis for V_ℓ), 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

$$f(h^{-1}x) = \sum_{\ell=0}^{+\infty} c_\ell(f)^T \pi_\ell(h) Y_\ell(x) \quad \text{for all } x \in \mathbb{S}^2, \tag{7.2}$$

where $c_\ell(f) = (c_\ell^m(f))_{m=-\ell,\ldots,\ell}$ denotes the vector in $\mathbb{C}^{2\ell+1}$ of spherical coefficients of f, and $Y_\ell(x) = (Y_\ell^m(x))_{m=-\ell,\ldots,\ell}$ is the vector in $\mathbb{C}^{2\ell+1}$ of spherical harmonics at frequency ℓ. Depending on the chosen parametrization for SO(3) (e.g. by Euler angles), various formulas are available to express the coefficients of the matrices π_ℓ and we refer to [9] for further details.

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

$$\begin{aligned} dZ_j(x) &= f_j(x)dx + \epsilon dW_j(x), \\ \text{where} \quad f_j(x) &= f^*(h_j^{*-1}x), \end{aligned} \tag{7.3}$$

where W_j, $j = 1, \ldots, J$ are standard Brownian sheets on the topological space $\mathbb{S}^2$ with measure dx, ϵ is an unknown noise level parameter, $f^* : \mathbb{S}^2 \to \mathbb{R}$ is an unknown template, and h_j^*, $j = 1, \ldots, J$ are rotation parameters in $G = \mathrm{SO}(3)$ to estimate. For $h = (h_1, \ldots, h_J) \in \mathcal{A}_0$, where $\mathcal{A}_0$ is the subset of G^J defined in Eq. (2.3), the shift property (7.3) and the orthonormality of the spherical harmonics imply that the following matching criterion

$$N(h) = \frac{1}{J} \sum_{j=1}^{J} \left\| f_j \circ T_{h_j} - \frac{1}{J} \sum_{j'=1}^{J} f_{j'} \circ T_{h_{j'}} \right\|_{\mathbb{L}^2(\mathbb{S}^2)}^2, \tag{7.4}$$

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

$$N(h) = \frac{1}{J} \sum_{j=1}^{J} \sum_{\ell=0}^{+\infty} \left\| c_\ell(f_j)^T \pi_\ell(h_j^{-1}) - \frac{1}{J} \sum_{j'=1}^{J} c_\ell(f_{j'})^T \pi_\ell(h_{j'}^{-1}) \right\|_{\mathbb{C}^{2\ell+1}}^2, \tag{7.5}$$

where $\|\cdot\|^2_{\mathbb{C}^{2\ell+1}}$ 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)

$$c_\ell(Z_j) = \int_{\mathbb{S}^2} Y_\ell(x) dZ_j(x) = c_\ell(f_j) + \epsilon c_\ell(W_j), \quad j = 1 \dots J,$$

where $c_\ell(W_j) = \int_{\mathbb{S}^2} Y_\ell(x) dW_j(x)$ 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 to be chosen later, the following empirical criterion can thus be proposed for registering spherical images:

$$N_\epsilon(h_1, \dots, h_J) = \frac{1}{J} \sum_{j=1}^{J} \sum_{\ell=0}^{\ell_\epsilon} \left\| c_\ell(Z_j)\pi(h_j^{-1}) - \frac{1}{J} \sum_{j'=1}^{J} c_\ell(Z_{j'})\pi(h_{j'}^{-1}) \right\|^2_{\mathbb{C}^{2\ell+1}}, \quad (7.6)$$

and an M-estimator of the rotation parameters is thus 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. Indeed, its formulation is equivalent to that of M_ϵ is one replaces, in the expression (3.3), the matrix $\pi(Y_j)$ by the vector $c_\ell(Z_j)$, the norm $\|\cdot\|^2_{HS}$ by $\|\cdot\|^2_{\mathbb{C}^{2\ell+1}}$, and the summation $\sum_{\pi \in \hat{G}_\epsilon}$ by $\sum_{\ell=0}^{\ell_\epsilon}$. Note that the weighting by the dimension $d_\pi = (2\ell + 1)$ disappears in the formulation of N_ϵ due the chosen normalization for the spherical harmonics. Therefore, the study of the consistency and the asymptotic normality of $\hat{h}_\epsilon$ can be done by following exactly the arguments developed in Sects. 2.2 and 4. For this, let us introduce the following definitions:

Definition 7.1 A function $f \in \mathbb{L}^2(\mathbb{S}^2)$ is said to be not shift-invariant if there does not exist a closed subgroup H of SO(3) (except $H = \{e\}$ or $H = \text{SO}(3)$) such that $f(hx) = f(x)$ for all $x \in \mathbb{S}^2$ and $h \in H$.

Definition 7.2 A function $f \in \mathbb{L}^2(\mathbb{S}^2)$ is said to be regular if for all $\ell \in \mathbb{N}$ such that $c_\ell(f)$ is not identically null, then the linear mapping $A \mapsto c_\ell(f)^T A$ is injective, for A belonging to the set of $(2\ell + 1) \times (2\ell + 1)$ matrices with complex entries.

Then, the following proposition holds (its proof is omitted since it follows from a simple adaptation of the proof of Theorem 3.1)

Proposition 7.1 *Assume that $f^* \in \mathbb{L}^2(\mathbb{S}^2)$ is not shift-invariant and regular. Suppose that*

$$\lim_{\epsilon \to 0} \epsilon^2 \sum_{\ell=0}^{\ell_\epsilon} (2\ell + 1) = 0,$$

then $\hat{h}_\epsilon$ converges in probability to $h^* = (h_1^*, \dots, h_J^*)$.

The asymptotic normality of $\hat{h}_\epsilon$ can also be studied by reformulating the criterion $N_\epsilon(h)$ as a function $\tilde{N}_\epsilon(u)$ defined on $\mathcal{G}^J$, and by taking $\hat{h}_\epsilon = \exp(\hat{u}_\epsilon)$. The Lie algebra of SO(3) is the space $\mathfrak{so}(3)$ of 3×3 skew symmetric matrices which is a linear space of dimension $p = 3$ generated by the basis (see [9])

$$x_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}, \ x_2 = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \ x_3 = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Adapting the conditions of Theorem 4.1 to the formulation of the criterion $\tilde{N}_\epsilon(u)$, replacing d_π by $\sqrt{2\ell+1}$, $\pi(f^*)$ by $c_\ell(f^*)$, and the Hilbert-Schmidt inner product and norm for $d_\pi \times d_\pi$ matrices by the euclidean inner product and norm in $\mathbb{C}^{2\ell+1}$, it is also possible to derive the asymptotic normality of $\hat{u}_\epsilon$.

Simulations: the numerical implementation of the above method for the registration of spherical images is more involved that 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 computation of the irreducible representation of the group SO(3) from its Lie algebra. Then, a numerical method to find a minimum of $N_\epsilon(h)$ could be developed by following the ideas of the general gradient descent algorithm described previously. Due to the large size of spherical data, it is essential to develop an efficient and fast numerical scheme. However, we believe that it is far beyond the scope of this paper to develop such a fast numerical method, so we prefer to leave this for a future work, but encouraged by the good numerical results shown in the previous section, we think that this approach could certainly yield satisfactory results for the registration of! spherical images.

Acknowledgments We would like to thank Dominique Bakry for fruitful discussions about the Fourier transform on compact groups, and his suggestions of potential applications to registration problems in image processing. We are very much indebted to the referees and the Associate Editor for their constructive comments and remarks that resulted in a major revision of the original manuscript. The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-JCJC-SIMI1 DEMOS.

Appendix

Proof of Theorem 3.1 to derive the result, it is enough to prove that $M(\cdot)$ has a unique minimum at $(h_1, \ldots, h_J) = (h_1^*, \ldots, h_J^*)$, and that M_ϵ converges uniformly in probability to M i.e.

$$\sup_{h \in \mathcal{A}_0} |M_\epsilon(h) - M(h)| \to 0 \text{ in probability as } \epsilon \to 0.$$

Then, following e.g. the proof of Theorem 5.7 in [41], these two conditions ensure that $\hat{h}_\epsilon = (\hat{h}_{1,\epsilon}, \ldots, \hat{h}_{J,\epsilon})$ converges in probability to h^* as $\epsilon \to 0$.

Unicity of the minimum of $M(\cdot)$: From the definition of M, M is a positive function such that: $M(h_1^*, \ldots, h_J^*) = 0$. This means that $h^* = (h_1^*, \ldots, h_J^*)$ is a minimum of M.

Let h be a minimum of M such that: $M(h) = 0$. We shall prove that $h = h^*$. Indeed, if $M(h) = 0$ then for all $\pi \in \hat{G}$ such that $\pi(f)$ is not identically null we have that:

$$\left\| \pi(f^*) \left((\pi(h_j^{*-1} h_j) - \frac{1}{J} \sum_{j'=1}^{J} \pi(h_{j'}^{*}{}^{-1} h_{j'}) \right) \right\|_{HS}^2 = 0, \quad \forall j = 1, \ldots J$$

i.e.

$$\text{Im} \left((\pi(h_j^{*-1} h_j) - \frac{1}{J} \sum_{j'=1}^{J} \pi(h_{j'}^{*}{}^{-1} h_{j'}) \right) \subseteq \text{Ker}(\pi(f^*)), \quad \forall j = 1, \ldots J.$$

Using the assumption (**A2**), and the identifiability constraint $h_1 = e$, this means that, for all $\pi \in \hat{G}$ such that $\pi(f^*)$ is not identically null,

$$\pi(h_j^{*-1} h_j) = I_{V_\pi}, \quad \forall j = 1, \ldots J,$$

i.e.

$$h_j^{*-1} h_j \in H = \cap \text{Ker}(\pi, \ \pi \in \hat{G} \text{ and } \pi(f) \neq 0\}, \quad \forall j = 1, \ldots J.$$

But from the assumption (**A1**), the normal subgroup H is $\{e\}$. The result follows.

Uniform convergence of M_ϵ: Remark that $M_\epsilon(h)$ is the sum of three terms:

$$M_\epsilon(h) = D_\epsilon(h) + \epsilon L_\epsilon(h) + \epsilon^2 Q_\epsilon(h), \tag{7.7}$$

where

$$D_\epsilon(h) = \frac{1}{J} \sum_{j=1}^{J} \sum_{\pi \in \hat{G}_\epsilon} d_\pi \left\| \pi(f_j)\pi(h_j) - \frac{1}{J} \sum_{j'=1}^{J} \pi(f_{j'})\pi(h_{j'}) \right\|_{HS}^2$$

$$L_\epsilon(h) = \frac{2}{J} \sum_{j=1}^{J} \sum_{\pi \in \hat{G}_\epsilon} d_\pi \Re \left\langle \pi(f_j)\pi(h_j) - \frac{1}{J} \sum_{j'=1}^{J} \pi(f_{j'})\pi(h_{j'}), \pi(W_j)\pi(h_j) \right.$$
$$\left. - \frac{1}{J} \sum_{j'=1}^{J} \pi(W_{j'})\pi(h_{j'}) \right\rangle_{HS}$$

$$Q_\epsilon(h) = \frac{1}{J} \sum_{j=1}^{J} \sum_{\pi \in \hat{G}_\epsilon} d_\pi \left\| \pi(W_j)\pi(h_j) - \frac{1}{J} \sum_{j'=1}^{J} \pi(W_{j'})\pi(h_{j'}) \right\|_{HS}^2,$$

where $\Re(x)$ denotes the real part of a complex number x. Let us notice that by applying Cauchy–Schwarz inequality, we get that:

$$|\epsilon L_\epsilon(h)| \leq 2 \left\{ \sup_{h \in \mathcal{A}_0} |D_\epsilon(h) - M(h)| + \sup_{h \in \mathcal{A}_0} M(h) \right\}^{1/2} \left\{ \sup_{h \in \mathcal{A}_0} \epsilon^2 Q_\epsilon(h) \right\}^{1/2}$$

Since the function M is continuous on the compact set $\mathcal{A}_0$, we have just to consider the uniform convergence of D_ϵ to M and the uniform convergence in probability of $\epsilon^2 Q_\epsilon$ to zero.

First, we study the uniform convergence of $D_\epsilon(h)$ to $M(h)$. Remark that:

$$\begin{aligned} \left\| \pi(f_j)\pi(h_j) \right\|_{HS}^2 &= \mathrm{Tr} \left(\pi(h_j^{-1}) \overline{\pi(f_j)}^T \pi(f_j)\pi(h_j) \right) \\ &= \left\| \pi(f_j) \right\|_{HS}^2 = \left\| \pi(f^*) \right\|_{HS}^2, \end{aligned} \tag{7.8}$$

and that:

$$\left\| \frac{1}{J} \sum_{j=1}^{J} \pi(f_j)\pi(h_j) \right\|_{HS}^2 = \left| \frac{1}{J^2} \sum_{j=1}^{J} \sum_{j'=1}^{J} \left\langle \pi(f_j)\pi(h_j), \pi(f_{j'})\pi(h_{j'}) \right\rangle_{HS} \right| \tag{7.9}$$

$$\leq \frac{1}{J^2} \sum_{j=1}^{J} \sum_{j'=1}^{J} \left\| \pi(f_j)\pi(h_j) \right\|_{HS} \left\| \pi(f_{j'})\pi(h_{j'}) \right\|_{HS} \tag{7.10}$$

$$\leq \left\| \pi(f^*) \right\|_{HS}^2. \tag{7.11}$$

Then for all $h \in \mathcal{A}_0$, we have that

$$|M(h) - D_\epsilon(h)| \leq 2 \sum_{\pi \in \hat{G} \setminus \hat{G}_\epsilon} d_\pi \left\| \pi(f^*) \right\|_{HS}^2.$$

Thus D_ϵ converges uniformly to M, because $f^* \in \mathbb{L}^2(G)$ and $\lim_{\epsilon \to 0} \hat{G}_\epsilon = G$.

We show now that $\epsilon^2 Q_\epsilon$ converges uniformly in probability to 0. Using the equality (7.8), $\epsilon^2 Q_\epsilon$ may be rewritten as the sum of two terms:

$$\epsilon^2 Q_\epsilon(h) = \frac{\epsilon^2}{J} \sum_{j=1}^{J} \sum_{\pi \in \hat{G}_\epsilon} d_\pi \left\| \pi(W_j) \right\|_{HS}^2 - \epsilon^2 \sum_{\pi \in \hat{G}_\epsilon} d_\pi \left\| \frac{1}{J} \sum_{j'=1}^{J} \pi(W_{j'})\pi(h_{j'}) \right\|_{HS}^2,$$

where $\left\|\pi(W_j)\right\|_{HS}^2 = \sum_{k=1}^{d_\pi}\sum_{l=1}^{d_\pi}\left|\pi_{k,j}(W_j)\right|^2$. Then, the first term of the sum converges uniformly in probability to 0, because

$$\lim_{\epsilon\to 0}\mathbb{E}\left(\frac{1}{J}\sum_{j=1}^{J}\sum_{\pi\in\hat{G}_\epsilon} d_\pi \left\|\pi(W_j)\right\|_{HS}^2\right) = \lim_{\epsilon\to 0}\epsilon^2\sum_{\pi\in\hat{G}_\epsilon} d_\pi^2 = 0.$$

Similarly to the inequality (7.9), the second term may be uniformly bounded by:

$$0\leq\epsilon^2\sum_{\pi\in\hat{G}_\epsilon} d_\pi\left\|\frac{1}{J}\sum_{j'=1}^{J}\pi(W_{j'})\pi(h_{j'})\right\|_{HS}^2 \leq\epsilon^2\sum_{\pi\in\hat{G}_\epsilon} d_\pi\frac{1}{J^2}\sum_{j,j'=1}^{J}\left\|\pi(W_j)\right\|_{HS}\left\|\pi(W_{j'})\right\|_{HS}.$$

Using the fact that,

$$\mathbb{E}\left(\left\|\pi(W_j)\right\|_{HS}\left\|\pi(W_{j'})\right\|_{HS}\right)\leq\left\{\mathbb{E}\left(\left\|\pi(W_j)\right\|_{HS}^2\right)\mathbb{E}\left(\left\|\pi(W_{j'})\right\|_{HS}^2\right)\right\}^{1/2} = d_\pi.$$

we deduce that the second term converges uniformly to 0, which completes the proof of Theorem 3.1. □

Proof of Proposition 4.2 from the decomposition (7.7), $\tilde{M}_\epsilon(u)$ can be written as the sum of three terms:

$$\epsilon^{-1}\tilde{M}_\epsilon(u) = \epsilon^{-1}\tilde{D}_\epsilon(u) + \tilde{L}_\epsilon(u) + \epsilon\tilde{Q}_\epsilon(u), \tag{7.12}$$

where $\tilde{D}_\epsilon(u) = D_\epsilon(\exp(u))$, $\tilde{L}_\epsilon(u) = L_\epsilon(\exp(u))$ and $\tilde{Q}_\epsilon(u) = Q_\epsilon(\exp(u))$. In what follows, we study the convergence of the three terms in the right part of equality (7.12).

Convergence of $\epsilon^{-1}\nabla_{u^*}\tilde{D}_\epsilon$: one can easily check that u^* is a minimum of $\tilde{D}_\epsilon(u)$ and thus $\nabla_{u^*}\tilde{D}_\epsilon = 0$ for any ϵ.

Convergence of $\epsilon\nabla_{u^*}\tilde{Q}_\epsilon$: remark that $\tilde{Q}_\epsilon$ can be written as

$$\tilde{Q}_\epsilon(u) = \sum_{\pi\in\hat{G}_\epsilon} d_\pi\times\left(\frac{1}{J}\sum_{j=1}^{J}\left\|\pi(W_j)\pi(\tilde{h}_j)\tilde{\pi}(u_j)\right\|_{HS}^2 - \left\|\frac{1}{J}\sum_{j'=1}^{J}\pi(W_{j'})\pi(\tilde{h}_{j'})\tilde{\pi}(u_{j'})\right\|_{HS}^2\right)$$

Let $2\leq j\leq J$ and $1\leq k\leq p$. Since $\left\|\pi(W_j)\pi(\tilde{h}_j)\tilde{\pi}(u_j)\right\|_{HS}^2 = \left\|\pi(W_j)\right\|_{HS}^2$ is independent of u_j, we obtain that for $u\in\mathcal{G}^J$

$$d_u\tilde{Q}_\epsilon(e_j\otimes x^k) = -\frac{2}{J}\sum_{\pi\in\hat{G}_\epsilon} d_\pi\Re$$
$$\times\left\langle\pi(W_j)\pi(\tilde{h}_j e^{u_j})d_e\pi(d_{u_j}\exp(x^k)), \frac{1}{J}\sum_{j'=1}^{J}\pi(W_{j'})\pi(\tilde{h}_{j'}e^{u_{j'}})\right\rangle_{HS}.$$

Thus, using Cauchy–Schwartz inequalities, one obtains that

$$\mathbb{E}|d_u \tilde{Q}_\epsilon(v)| \leq \frac{2}{J} \sum_{\pi \in \hat{G}_\epsilon} d_\pi \sqrt{\mathbb{E}\, \|A\|_{HS}^2} \sqrt{\mathbb{E}\, \|B\|_{HS}^2},$$

where

$$A = \pi(W_j)\pi(\tilde{h}_j e^{u_j}) d_e \pi(d_{u_j} \exp(x^k)) \quad \text{and} \quad B = \frac{1}{J} \sum_{j'=1}^{J} \pi(W_{j'})\pi(\tilde{h}_{j'} e^{u_{j'}}).$$

Arguing as in the proof of Theorem 3.1, one can easily prove that $\mathbb{E}\, \|B\|_{HS}^2 \leq d_\pi$. Then, remark that

$$\begin{aligned} \mathbb{E}\, \|A\|_{HS}^2 &\leq \left\| d_e \pi(d_{u_j} \exp(v)) \right\|_{HS}^2 \mathbb{E} \left\| \pi(W_j)\pi(\tilde{h}_j e^{u_j}) \right\|_{HS}^2 \\ &\leq d_\pi \left\| d_e \pi \left(d_{u_j} \exp(x^k) \right) \right\|_{HS}^2 \end{aligned}$$

Therefore, under condition (4.6) it follows that $\epsilon \mathbb{E}|\nabla_{u^*}^{x_j^k} \tilde{Q}_\epsilon|$ converges to 0 as $\epsilon \to 0$, and we conclude via Markov inequality.

Convergence of $\nabla_{u^*} \tilde{L}_\epsilon$: let $2 \leq j \leq J$ and $1 \leq k \leq p$. Then,

$$\begin{aligned} d_{u^*} \tilde{L}_\epsilon(e_j \otimes x^k) &= \frac{2}{J} \sum_{\pi \in \hat{G}_\epsilon} d_\pi \Re \\ &\times \left\langle \pi(f_j)\pi(h_j^*) d_e \pi(d_{u_j^*} \exp(x^k)), \pi(W_j)\pi(h_j^*) - \frac{1}{J} \sum_{j'=1}^{J} \pi(W_{j'})\pi(h_{j'}^*) \right\rangle_{HS} \end{aligned}$$

Let us introduce the following quantities

$$V_{j,k}^\pi = \pi(f^*) d_e \pi(d_{u_j^*} \exp(x^k)) \quad \text{and} \quad Z_j^\pi = \pi(W_j)\pi(h_j^*) \sim \mathcal{N}_{\mathbb{C}}(0, \quad d_\pi^{-1} \pi(h_j^*) \overline{\pi(h_j^*)}^T).$$

Hence, using that $\pi(f_j) = \pi(f^*)\pi(h_j^{*-1})$,

$$\begin{aligned} d_{u^*} \tilde{L}_\epsilon(e_j \otimes x^k) &= \frac{2}{J} \sum_{\pi \in \hat{G}_\epsilon} d_\pi \Re \\ &\times \left(\left\langle V_{j,k}^\pi, \left(1 - \frac{1}{J}\right) Z_j^\pi \right\rangle_{HS} + \left\langle V_{j,k}^\pi, -\frac{1}{J} \sum_{j'=1, j' \neq j}^{J} Z_{j'}^\pi \right\rangle_{HS} \right) \end{aligned}$$

Since $\mathrm{Var}\left(\Re\left\langle V_{j,k}^{\pi}, Z_j^{\pi}\right\rangle_{HS}\right) = \|V_{j,k}^{\pi}\|_{HS}^2/(2d_\pi)$, $d_{u^*}\tilde{L}_\epsilon(e_j \otimes x^k)$ is a Gaussian variable with zero mean, and variance:

$$\mathrm{Var}\left(d_{u^*}\tilde{L}_\epsilon(e_j \otimes x^k)\right) = = \frac{4}{J^2}\sum_{\pi\in\hat{G}_\epsilon} d_\pi \left(1-\frac{1}{J}\right)\left\|\pi(f^*)d_e\pi\left(d_{u_j^*}\exp(x^k)\right)\right\|_{HS}^2 \tag{7.13}$$

where we have used the fact that the Z_j^{π}'s are independent variables for $\pi \neq \pi'$ (except for $\pi' = \overline{\pi}$). Using similar calculations, one obtains that

$$\begin{aligned}
&\mathbb{E}\left(d_{u^*}\tilde{L}_\epsilon(e_j \otimes x^{k_1})\overline{d_{u^*}\tilde{L}_\epsilon(e_j \otimes x^{k_2})}\right)\\
&= \frac{4}{J^2}\sum_{\pi\in\hat{G}_\epsilon} d_\pi\left(1-\frac{1}{J}\right)\Re\left\langle \pi(f^*)d_e\pi\left(d_{u_j^*}\exp(x^{k_1})\right),\right.\\
&\quad \left.\times\, \pi(f^*)d_e\pi\left(d_{u_j^*}\exp(x^{k_2})\right)\right\rangle_{HS} \qquad (7.14)\\
&\mathbb{E}\left(d_{u^*}\tilde{L}_\epsilon(e_{j_1} \otimes x^{k_1})\overline{d_{u^*}\tilde{L}_\epsilon(e_{j_2} \otimes x^{k_2})}\right)\\
&= -\frac{4}{J^2}\sum_{\pi\in\hat{G}_\epsilon} d_\pi\frac{1}{J}\Re\left\langle \pi(f^*)d_e\pi\left(d_{u_{j_1}^*}\exp(x^{k_1})\right),\right.\\
&\quad \left.\times\, \pi(f^*)d_e\pi\left(d_{u_{j_2}^*}\exp(x^{k_2})\right)\right\rangle_{HS}, \qquad (7.15)
\end{aligned}$$

for $j_1 \neq j_2$ and $1 \leq k_1, k_2 \leq p$. Finally, for any vector $v \in \mathcal{G}^{J-1}$, by using Eqs. (7.13), (7.14) and (7.15), one has that as $\epsilon \to 0$

$$d_{u^*}\tilde{L}_\epsilon(v) \to N\left(0, \frac{4}{J^2}v^T\Sigma v\right),$$

which completes the proof of Proposition 4.2. □

Proof of Proposition 4.3 Let $(u_1, \ldots, u_J)$ be in $\mathcal{U}_0$. Let us denote by $(h_1, \ldots, h_J) \in G^J$ the corresponding element such that $h_j = \tilde{h}_j\exp(u_j)$, $j = 1\ldots J$. Let $2 \leq j_1, j_2 \leq J$ and $1 \leq k_1, k_2 \leq p$. Then, from the decomposition (7.7), one has that for any $u \in \mathcal{U}_1$:

$$d_u^2\tilde{M}_\epsilon = d_u^2\tilde{D}_\epsilon + \epsilon d_u^2\tilde{L}_\epsilon + \epsilon^2 d_u^2\tilde{Q}_\epsilon, \tag{7.16}$$

In what follows, we study the uniform convergence in probability over $\mathcal{U}_0$ of the three above terms.

Convergence of $d_u^2 \tilde{D}_\epsilon$: from the definition of $\tilde{D}_\epsilon(u)$ one has that for $j_1 \neq j_2$

$$d_u^2 \tilde{D}_\epsilon(e_{j_1} \otimes x^{k_1}, e_{j_2} \otimes x^{k_2}) = \frac{-2}{J^2} \sum_{\pi \in \hat{G}_\epsilon} d_\pi \Re \Big\langle \pi(f_{j_1})\pi(h_{j_1}) d_e\pi(d_{u_{j_1}} \times \exp(x^{k_1})), \pi(f_{j_2})\pi(h_{j_2}) d_e\pi(d_{u_{j_2}} \exp(x^{k_2})) \Big\rangle_{HS}.$$

Hence, using Cauchy–Schwartz inequality and the fact that $\|\pi(f_{j_1})\|_{HS} = \|\pi(f^*) \|_{HS}$ yields

$$\begin{aligned}
&\left| d_u^2 \tilde{M}(e_{j_1} \otimes x^{k_1}, e_{j_2} \otimes x^{k_2}) - d_u^2 \tilde{D}_\epsilon(e_{j_1} \otimes x^{k_1}, e_{j_2} \otimes x^{k_2}) \right| \\
&\leq \frac{2}{J^2} \sum_{\pi \in \hat{G} \backslash \hat{G}_\epsilon} d_\pi \left\| \pi(f_{j_1})\pi(h_{j_1}) d_e\pi(d_{u_{j_1}} \exp(x^{k_1})) \right\|_{HS} \\
&\quad \times \left\| \pi(f_{j_2})\pi(h_{j_2}) d_e\pi(d_{u_{j_2}} \exp(x^{k_2})) \right\|_{HS} \\
&\leq \frac{2}{J^2} \sum_{\pi \in \hat{G} \backslash \hat{G}_\epsilon} d_\pi \left\| \pi(f^*) \right\|_{HS}^2 \left\| d_e\pi(d_{u_{j_1}} \exp(x^{k_1})) \right\|_{HS} \left\| d_e\pi(d_{u_{j_2}} \exp(x^{k_2}))\tilde{\pi} \right\|_{HS} \\
&\leq \frac{2}{J^2} \left\{ \sum_{\pi \in \hat{G} \backslash \hat{G}_\epsilon} d_\pi \left\| \pi(f^*) \right\|_{HS}^2 \left\| d_e\pi \left(d_{u_{j_1}} \exp(x^{k_1}) \right) \right\|_{HS}^2 \right\}^{1/2} \\
&\quad \times \left\{ \sum_{\pi \in \hat{G} \backslash \hat{G}_\epsilon} d_\pi \left\| \pi(f^*) \right\|_{HS}^2 \left\| d_e\pi \left(d_{u_{j_2}} \exp(x^{k_2}) \right) \right\|_{HS}^2 \right\}^{1/2}.
\end{aligned}$$

Therefore, under Assumption (4.8), one has that $d_u^2 \tilde{D}_\epsilon(e_{j_1} \otimes x^{k_1}, e_{j_2} \otimes x^{k_2})$ converges uniformly to $d_u^2 \tilde{M}_\epsilon(e_{j_1} \otimes x^{k_1}, e_{j_2} \otimes x^{k_2})$ over $\mathcal{U}_0$.

Now, if $j_1 = j_2$, one has that

$$\begin{aligned}
&d_u^2 \tilde{D}_\epsilon(e_{j_1} \otimes x^{k_1}, e_{j_1} \otimes x^{k_2}) \\
&= \sum_{\pi \in \hat{G}_\epsilon} d_\pi \Re \Big\langle \frac{2}{J} \pi(f_{j_1})\pi(h_{j_1}) \Big[d_e\pi(d_{u_{j_1}} \exp(x^{k_2})) d_e\pi(d_{u_{j_1}} \exp(x^{k_1}))
\end{aligned}$$

$$+ d_e\pi(d^2_{u_{j_1}} \exp(x^{k_1}, x^{k_2}))\Big], \pi(f_{j_1})\pi(h_{j_1}) - \frac{1}{J}\sum_{j=1}^{J}\pi(f_j)\pi(h_j)\Big\rangle_{HS} + \frac{2(1-1/J)}{J}$$
$$\times \sum_{\pi\in\hat{G}_\epsilon} d_\pi \Re\left\langle \pi(f_{j_1})\pi(h_{j_1})d_e\pi(d_{u_{j_1}} \exp(x^{k_1})), \pi(f_{j_1})\pi(h_{j_1})d_e\pi(d_{u_{j_1}} \exp(x^{k_2}))\right\rangle_{HS}$$

By proceeding as previously using Cauchy–Schwartz inequality, Assumption (4.8) and Assumption (4.9) ensure that $d_u^2\tilde{D}_\epsilon(e_{j_1}\otimes x^{k_1}, e_{j_1}\otimes x^{k_2})$ converges uniformly to $d_u^2\tilde{M}_\epsilon(e_{j_1}\otimes x^{k_1}, e_{j_1}\otimes x^{k_2})$ over $\mathcal{U}_0$.

Convergence of $\epsilon^2 d_u^2\tilde{Q}_\epsilon$: from the definition of $\tilde{Q}_\epsilon(u)$ one has that for $j_1 \neq j_2$:

$$d_u^2\tilde{Q}_\epsilon(e_{j_1}\otimes x^{k_1}, e_{j_2}\otimes x^{k_2}) = -\frac{2}{J^2}\sum_{\pi\in\hat{G}_\epsilon} d_\pi \Re\big\langle \pi(W_{j_1})\pi(h_{j_1})d_e$$
$$\pi(d_{u_{j_1}} \exp(x^{k_1})), \pi(W_{j_2})\pi(h_{j_2})d_e\pi(d_{u_{j_2}} \exp(x^{k_2}))\big\rangle_{HS}$$

Thus, using Cauchy–Schwartz inequality,

$$\mathbb{E}|d_u^2\tilde{Q}_\epsilon(e_{j_1}\otimes x^{k_1}, e_{j_2}\otimes x^{k_2})| \leq \frac{2}{J^2}\sum_{\pi\in\hat{G}_\epsilon} d_\pi\sqrt{\mathbb{E}\,\|A\|^2_{HS}}\sqrt{\mathbb{E}\,\|B\|^2_{HS}},$$

where

$$A = \pi(W_{j_1})\pi(h_{j_1})d_e\pi(d_{u_{j_1}} \exp(x^{k_1})) \text{ and } B = \pi(W_{j_2})\pi(h_{j_2})d_e\pi(d_{u_{j_2}} \exp(x^{k_2})).$$

Now, remark that $\|A\|^2_{HS} \leq \left\|\pi(W_{j_1})\right\|^2_{HS}\left\|d_e\pi\left(d_{u_{j_1}} \exp(x^{k_1})\right)\right\|^2_{HS}$, which implies that (using again Cauchy–Schwartz inequality)

$$\begin{aligned}\mathbb{E}|d_u^2\tilde{Q}_\epsilon(e_{j_1}\otimes x^{k_1}, e_{j_2}\otimes x^{k_2})| &\leq \frac{2}{J^2}\sum_{\pi\in\hat{G}_\epsilon} d_\pi^2\left\|d_e\pi\left(d_{u_{j_1}} \exp(x^{k_1})\right)\right\|_{HS} \\ &\quad\times\left\|d_e\pi\left(d_{u_{j_2}} \exp(x^{k_2})\right)\right\|_{HS} \\ &\leq \frac{2}{J^2}\left\{\sum_{\pi\in\hat{G}_\epsilon} d_\pi^2\left\|d_e\pi\left(d_{u_{j_1}} \exp(x^{k_1})\right)\right\|^2_{HS}\right\}^{1/2} \\ &\quad\times\left\{\sum_{\pi\in\hat{G}_\epsilon} d_\pi^2\left\|d_e\pi\left(d_{u_{j_2}} \exp(x^{k_2})\right)\right\|^2_{HS}\right\}^{1/2}\end{aligned}$$

and therefore under Assumption (4.10), one has that $\epsilon^2\mathbb{E}|d_u^2\tilde{Q}_\epsilon(e_{j_1}\otimes x^{k_1}, e_{j_2}\otimes x^{k_2})|$ converges uniformly to zero over $\mathcal{U}_1$, and the uniform convergence in probability of $\epsilon^2 d_u^2\tilde{Q}_\epsilon(e_{j_1}\otimes x^{k_1}, e_{j_2}\otimes x^{k_2})$ to zero follows by Markov inequality.

Now for $j_1 = j_2$,

$$d_u^2 \tilde{Q}_\epsilon(e_{j_1} \otimes x^{k_1}, e_{j_1} \otimes x^{k_2}) = -\frac{2}{J^2} \sum_{\pi \in \hat{G}_\epsilon}$$

$$\times d_\pi \Re \left\langle \pi(W_{j_1})\pi(h_{j_1}) d_e\pi(d_{u_{j_1}} \exp(x^{k_1})), \pi(W_{j_1})\pi(h_{j_1}) d_e\pi(d_{u_{j_1}} \exp(x^{k_2})) \right\rangle_{HS}$$

$$-\frac{2}{J^2} \sum_{\pi \in \hat{G}_\epsilon} d_\pi \Re \left\langle \pi(W_{j_1})\pi(h_{j_1}) \left[d_e\pi(d_{u_{j_1}} \exp(x^{k_2})) d_e\pi(d_{u_{j_1}} \exp(x^{k_1})) \right.\right.$$

$$\left.\left. + \; d_e\pi(d^2_{u_{j_1}} \exp(x^{k_1}, x^{k_2})) \right], \sum_{j=1}^{J} \pi(W_j)\pi(h_j) \right\rangle_{HS},$$

By proceeding as previously using Cauchy–Schwartz inequality and Markov inequality, Assumption (4.10) and Assumption (4.11) ensure that $\epsilon^2 d_u^2 \tilde{M}_\epsilon(e_{j_1} \otimes x^{k_1}, e_{j_1} \otimes x^{k_2})$ converges uniformly in probability to zero over $\mathcal{U}_0$.

Convergence of $\epsilon d_u^2 \tilde{L}_\epsilon$: from the definition of $\tilde{L}_\epsilon(u)$ one has that for $j_1 \neq j_2$:

$$d_u^2 \tilde{L}_\epsilon(e_{j_1} \otimes x^{k_1}, e_{j_2} \otimes x^{k_2}) = \frac{-2}{J^2} \sum_{\pi \in \hat{G}_\epsilon} d_\pi \Re \left\{ \left\langle \pi(f_{j_1})\pi(h_{j_1}) d_e\pi(d_{u_{j_1}} \exp(x^{k_1})), \right.\right.$$

$$\left. \pi(W_{j_2})\pi(h_{j_2}) d_e\pi(d_{u_{j_2}} \exp(x^{k_2})) \right\rangle_{HS}$$

$$+ \left\langle \pi(f_{j_2})\pi(h_{j_2}) d_e\pi(d_{u_{j_2}} \exp(x^{k_2})), \right.$$

$$\left.\left. \pi(W_{j_1})\pi(h_{j_1}) d_e\pi(d_{u_{j_1}} \exp(x^{k_1})) \right\rangle_{HS} \right\}$$

Then, remarking that by Cauchy–Schwartz inequality

$$\mathbb{E} \left| d_u^2 \tilde{L}_\epsilon(e_{j_1} \otimes x^{k_1}, e_{j_2} \otimes x^{k_2}) \right|$$

$$\leq \frac{2}{J^2} \left\{ \sum_{\pi \in \hat{G}_\epsilon} d_\pi \left\| \pi(f_{j_1})\pi(h_{j_1}) d_e\pi(d_{u_{j_1}} \exp(x^{k_1})) \right\|_{HS}^2 \right\}^{1/2}$$

$$\times \left\{ \sum_{\pi \in \hat{G}_\epsilon} d_\pi \mathbb{E} \left\| \pi(W_{j_2})\pi(h_{j_2}) d_e\pi(d_{u_{j_2}} \exp(x^{k_2})) \right\|_{HS}^2 \right\}^{1/2}$$

$$+ \frac{2}{J^2} \left\{ \sum_{\pi \in \hat{G}_\epsilon} d_\pi \left\| \pi(f_{j_2})\pi(h_{j_2}) d_e\pi(d_{u_{j_2}} \exp(x^{k_2})) \right\|_{HS}^2 \right\}^{1/2}$$

$$\times \left\{ \sum_{\pi \in \hat{G}_\epsilon} d_\pi \mathbb{E} \left\| \pi(W_{j_1})\pi(h_{j_1}) d_e\pi(d_{u_{j_1}} \exp(x^{k_1})) \right\|_{HS}^2 \right\}^{1/2},$$

and arguing as above for the convergence of $\nabla_u^2 \tilde{Q}_\epsilon$, it follows from our assumptions that $\epsilon |d_u^2 \tilde{L}_\epsilon(e_{j_1} \otimes x^{k_1}, e_{j_2} \otimes x^{k_2})|$ converges uniformly to zero in probability.

Using similar arguments and our assumptions, one can also prove the uniform convergence in probability to zero of $\epsilon |d_u^2 \tilde{L}_\epsilon(e_{j_1} \otimes x^{k_1}, e_{j_1} \otimes x^{k_2})|$, which completes the proof of Proposition 4.3. □

Proof of Theorem 4.1 Recall that,

$$\mathcal{U}_\epsilon = \{u \in \mathcal{U}_0, \quad \|u - u^*\| \leq \|\hat{u}_\epsilon - u^*\|\}.$$

Let $\gamma > 0$, and remark that

$$\mathbb{P}\left(\sup_{u \in \mathcal{U}_\epsilon} \left\| \nabla_u^2 \tilde{M}_\epsilon - \frac{2}{J}\Sigma \right\| > 2\gamma \right) \leq \mathbb{P}\left(\sup_{u \in \mathcal{U}_\epsilon} \|\nabla_u^2 \tilde{M}_\epsilon - \nabla_u^2 \tilde{M}\| > \gamma \right) + \mathbb{P}\left(\sup_{u \in \mathcal{U}_\epsilon} \|\nabla_u^2 \tilde{M} - \nabla_{u^*}^2 \tilde{M}\| > \gamma \right),$$

where $\nabla_{u^*}^2 \tilde{M} = \frac{2}{J}\Sigma$. From Proposition 4.3, the first term in the above equation converges to zero as $\epsilon \to 0$. For the second term, one can remark that $u \mapsto \nabla_u^2 \tilde{M}$ is a uniformly continuous function the compact set $\mathcal{U}_0$, therefore there exists $\delta > 0$ such that (by inclusion of events)

$$\mathbb{P}\left(\sup_{u \in \mathcal{U}_\epsilon} \|\nabla_u^2 \tilde{M} - \nabla_{u^*}^2 \tilde{M}\| > \gamma \right) \leq \mathbb{P}\left(\|\hat{u}_\epsilon - u^*\| > \delta\right).$$

By Proposition 4.1 it follows that the right term in the above equation converges to zero which finally proves that $\sup_{u \in \mathcal{U}_\epsilon} \|\nabla_u^2 \tilde{M}_\epsilon - \frac{2}{J}\Sigma\|$ converges in probability to 0. Using a Taylor expansion of $\nabla_u \tilde{M}_\epsilon$ with an integral remainder, we get that

$$\left[\frac{2}{J}\Sigma + \int_0^1 \left(\nabla_{\bar{u}_\epsilon(t)}^2 \tilde{M}_\epsilon - \frac{2}{J}\Sigma\right)\right] \epsilon^{-1}(\hat{u}_\epsilon - u^*) = -\epsilon^{-1} \nabla_{u^*} \tilde{M}_\epsilon,$$

where $\bar{u}_\epsilon(t) = u^* + t(\hat{u}_\epsilon - u^*) \in \mathcal{U}_\epsilon$. From Proposition 4.2, one has that $\epsilon^{-1}\nabla_{u^*}\tilde{M}_\epsilon$ converges to the Gaussian variable $N(0, \frac{4}{J^2}\Sigma)$. Since $\sup_{u \in \mathcal{U}_\epsilon} \|\nabla_u^2 \tilde{M}_\epsilon - \frac{2}{J}\Sigma\|$ converges in probability to 0, the proof of Theorem 4.1 is completed. □

Proof of Proposition 4.6 first recall that $G = (\mathbb{R}/\mathbb{Z})^p$ and $\hat{G} = \mathbb{Z}^p$. Let $\ell \in \mathbb{Z}^d$ and for $j = 1 \ldots J$, let us denote by $c_\ell(W_j)$ the Gaussian variables defined as,

$$c_\ell(W_j) = \int_G \overline{e_\ell(x)} dW_j(x).$$

Then, we may write for $g \in G$,

$$f^*(g) - \hat{f}_\epsilon(g) = \sum_{|\ell|_\infty > \ell_\epsilon} c_\ell(f^*) e_\ell(g) + \sum_{1 \le |\ell|_\infty \le \ell_\epsilon} c_\ell(f^*) \left\{ 1 - \frac{1}{J} \sum_{j=1}^{J} e_\ell(\hat{h}_{j,\epsilon} - h_j^*) \right\} e_\ell(g) - \epsilon \mathcal{S}_\epsilon(g), \tag{7.17}$$

where

$$\mathcal{S}_\epsilon(g) = \sum_{1 \le |l|_\infty \le \ell_\epsilon} \left\{ \frac{1}{J} \sum_{j=1}^{J} c_\ell(W_j) e_\ell(\hat{h}_{j,\epsilon}) \right\} e_\ell(g).$$

From Proposition 4.5, $\epsilon(\hat{h}_\epsilon - h^*)$ converges weakly in distribution. Then, the delta method (see e.g. [41]) implies that

$$\mathbb{E} \left| 1 - \frac{1}{J} \sum_{j=1}^{J} e_\ell(\hat{h}_{j,\epsilon} - h_j^*) \right|^2 = (|\ell|^2) \mathcal{O}(\epsilon^2). \tag{7.18}$$

Furthermore, the stochastic term $\mathcal{S}_\epsilon$ is such that

$$\int_G \mathbb{E}(|\mathcal{S}_\epsilon(g)|^2) dg \le \frac{1}{J} \sum_{1 \le |\ell|_\infty \le \ell_\epsilon} \sum_{j=1}^{J} \mathbb{E}(|c_\ell(W_j)|^2) = \mathcal{O}(\ell_\epsilon^p). \tag{7.19}$$

Then, inserting (7.18) and (7.19) into (7.17) yields

$$\mathbb{E} \int_G |f^*(g) - \hat{f}_\epsilon(g)|^2 dg = \sum_{|\ell|_\infty > \ell_\epsilon} |c_\ell(f^*)|^2 + \mathcal{O}(\epsilon^2 + \ell_\epsilon^p \epsilon^2).$$

and using our assumptions on the smoothness of f completes the proof of Proposition 4.6. □

Proof of Proposition 5.3 from the proof of Proposition 4.2, the gradient of $\tilde{M}_\epsilon$ at u^* is a squared integrable random variable and may be rewritten as,

$$\epsilon^{-1} \nabla_{u^*} \tilde{M}_\epsilon = \nabla_{u^*} \tilde{L}_\epsilon + o_{\mathbb{P}_\theta^{(\epsilon)}},$$

where, for $j = 2 \ldots J$ and $k = 1 \ldots, p$,

$$\frac{J}{2} d_{u^*} \tilde{L}_\epsilon(e_j \otimes x^k) = \Re \left\{ \sum_{\pi \in \hat{G}} \pi(f^*) d_e \pi(x^k) \pi(\tilde{h}_j^{-1} \exp(-u_j^*)) \overline{\pi(W_j)} \right\} + o_{\mathbb{P}_{\theta^*}^{(\epsilon)}}.$$

Furthermore using the Fourier inversion formula (2.1), we have that for $j = 2, \dots, J$ and $k = 1, \dots, p$,

$$\int_G d_{-u_j^*} \tilde{f}(x^k) dW_j(g) = \int_G d_{g\tilde{h}_j^{-1}\exp(-u_j^*)} f(x^k) dW_j(g)$$
$$= \sum_{\pi \in \hat{G}} \pi(f^*) d_\pi \pi(x^k) \pi(\tilde{h}_j^{-1} \exp(-u_j^*)) \int_G \pi(g) dW_j(g)$$
$$= \sum_{\pi \in \hat{G}} \pi(f^*) d_\pi \pi(x^k) \pi(\tilde{h}_j^{-1} \exp(-u_j^*)) \overline{\pi(W_j)}.$$

Consequently, we may rewrite $\epsilon^{-1} \nabla_{u^*} \tilde{M}_\epsilon$ as,

$$\epsilon^{-1} \nabla_{u^*} \tilde{M}_\epsilon = \frac{2}{J} \left\{ \sum_{j=2}^{J} e_j \otimes \xi_j - \frac{1}{J} \sum_{j=1}^{J} 1_{J-1} \otimes \xi_j \right\} + o_{\mathbb{P}_{\theta^*}^{(\epsilon)}}(1).$$

Then, using a Taylor expansion, we have established in the proof of Theorem 4.1 that:

$$\epsilon^{-1}(\hat{u}_\epsilon - u^*) = -\frac{J}{2} \left(I_{J-1} - \frac{1}{J} \mathbb{I}_{J-1} \right)^{-1} \otimes \mathrm{Gramm}(\nabla f^*)^{-1} \epsilon^{-1} \nabla_{u^*} \tilde{M}_\epsilon + o_{\mathbb{P}_{\theta^*}^{(\epsilon)}}(1),$$

and thus the result of Proposition 5.3 follows from Eq. (5.3) and the arguments given in Sect. 5. □

References

1. Amari, S.-I.: Differential-geometrical methods in statistics. In: Lecture Notes in Statistics. Springer, Berlin (1985)
2. Applebaum, D.: Lévy processes in stochastic differential geometry. In: Levy Processes: Theory and Applications, pp. 111–137. Birkhäuser, Basel (2001)
3. Brown, R.L.D., Low, M.G.: Asymptotic equivalence of nonparametric regression and white noise. Ann. Stat. **24**, 2384–2398 (1996)
4. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int. J. Comput. Vis. **61**(2), 139–157 (2005)
5. Bhattacharya, R., Patrangenaru, V.: Nonparametric estimation of location and dispersion on Riemannian manifolds. J. Stat. Plan. Inference **108**(1–2):23–35 (2002). C. R. Rao 80th birthday felicitation volume, Part II
6. Bhattacharya, R., Patrangenaru, V.: Large sample theory of intrinsic and extrinsic sample means on manifolds—I. Ann. Stat. **31**(1), 1–29 (2003)
7. Bhattacharya, R., Patrangenaru, V.: Large sample theory of intrinsic and extrinsic sample means on manifolds—II. Ann. Stat. **33**, 1225–1259 (2005)
8. Bump, D.: Lie groups. In: Graduate Texts in Mathematics, vol. 225. Springer-Verlag, New York (2004)
9. Chirikjian, G.S., Kyatkin, A.B.: Engineering Applications of Noncommutative Harmonic Analysis. CRC Press, Boca Raton (2001)
10. Chang, T., Rivest, L.-P.: M-estimation for location and regression parameters in group models: a case study using Stiefel manifolds. Ann. Stat. **29**(3), 784–814 (2001)
11. Duistermaat, J.J., Kolk, J.A.C.: Lie Groups. Universitext. Springer-Verlag, Berlin (2000)

12. Elworthy, D.: Geometric aspects of diffusions on manifolds. In: École d'Été de Probabilités de Saint-Flour XV–XVII, Lecture Notes in Mathematics, pp. 277–425. Springer, Berlin (1988)
13. Gamboa, F., Loubes, J.-M., Maza, E.: Semi-parametric estimation of shifts. Electron. J. Stat. **1**, 616–640 (2007)
14. Glasbey, C.A., Mardia, K.V.: A penalized likelihood approach to image warping. J. R. Stat. Soc. Ser. B **63**(3), 465–492 (2001)
15. Grenander, U.: General Pattern Theory: A Mathematical Study of Regular Structures. Oxford University Press, New York (1993)
16. Grenander, U., Miller, M., Srivastava, A.: Hilbert-Schmidt lower bounds for estimators on matrix Lie groups. IEEE Trans. Pattern Anal. Mach. Intell. **20**(8), 790–801 (1998)
17. Hendriks, H.: A Cramér-Rao type lower bound for estimators with values in a manifold. J. Multivar. Anal. **38**(2), 245–261 (1991)
18. Hendriks, H.W., Janssen, J.H., Ruymgaart, F.H.: A Cramér-Rao type inequality for random variables in Euclidean manifolds. Sankhya A **54**(3), 387–401 (1992)
19. Ibragimov, I.A., Hasminskii, R.Z.: Statistical Estimation. Springer-Verlag, New York (1981)
20. Itô, K.: Brownian motions in a Lie group. Proc. Jpn. Acad. **26**(8), 4–10 (1950)
21. Jain, A.K.: Fundamentals of Digital Image Processing. Prentice-Hall, Inc., Upper Saddle River (1989)
22. Kass, R.E., Vos, P.W.: Geometrical Foundations of Asymptotic Inference. Wiley-Interscience (1997)
23. Kim, P.T.: Deconvolution density estimation on $\mathfrak{so}(n)$. Ann. Stat. **26**, 1083–1102 (1998)
24. Kim, P.T., Koo, J.Y.: Asymptotic minimax bounds for stochastic deconvolution over groups. IEEE Trans. Inf. Theory **54**, 289–298 (2008)
25. Kim, P.T., Koo, J.Y.: Statistical inverse problems on manifolds. J. Fourier Anal. Appl. **11**, 639–653 (2005)
26. Klassen, E., Srivastava, A., Mio, W., Joshi, S.H.: Analysis of planar shapes using geodesic paths on shape spaces. IEEE Trans. Pattern Anal. Mach. Intell. **26**(3), 372–383 (2004)
27. Li, H., Manjunath, B.S., Mitra, S.K.: A contour-based approach to multisensor image registration. IEEE Trans. Image Process. **4**(3), 320–334 (1995)
28. Mardia, K.V., Dryden, I.L.: Efficient and Adaptive Estimation for Semiparametric Models. Springer-Verlag, Berlin (1998)
29. Makadia, A., Daniilidis, K.: Rotation recovery from spherical images without correspondences. IEEE Trans. Pattern Anal. Mach. Intell. **28**(7), 1170–1175 (2006)
30. Marsland, S., Twining, C.J.: Constructing diffeomorphic representations for the groupwise analysis of nonrigid registrations of medical images. IEEE Trans. Med. Imaging **23**(8), 1006–1020 (2004)
31. McNeney, B., Wellner, J.A.: Application of convolution theorems in semiparametric models with non-i.i.d. data. J. Stat. Plan. Inference **91**(2), 441–480 (1998)
32. Nussbaum, M.: Spline smoothing in regression models and asymptotic efficiency in $l2$. Ann. Stat. **13**, 984–992 (1985)
33. Owren, B., Welfert, B.: The Newton iteration on Lie groups. BIT Numer. Math. **40**(1), 121–145 (2000)
34. Ritov, Y., Bickel, P.J., Klaassen, C.A.J., Wellner, J.A.: Efficient and Adaptive Estimation for Semiparametric Models. Springer-Verlag, New York (1998)
35. Sepanski, M.R.: Compact Lie groups. In: Graduate Texts in Mathematics, vol. 235. Springer, New York (2007)
36. Smith, S.T.: Covariance, subspace, and intrinsic cramér-rao bounds. IEEE Trans. Signal Process. **53**(5), 1610–1630 (2005)
37. Stone, C.J.: Optimal global rates of convergence for nonparametric regression. Ann. Stat. **10**(4), 1040–1053 (1982)
38. Trouvé, A., Younes, L.: Metamorphoses through lie group action. Found. Comput. Math. **5**(2), 173–198 (2005)
39. Varadarajan, V.S.: Lie Groups, Lie Algebras, and Their Representations. GTM 102. Springer-Verlag, New York (1984)
40. van den Boogaart, K.G., Hielscher, R., Prestin, J., Schaeben, H.: Kernel-based methods for inversion of the radon transform on so(3) and their applications to texture analysis. J. Comput. Appl. Math. **199**(1), 122–140 (2007)
41. Van der Vaart, A.W.: Asymptotic statistics. In: Cambridge Series in Statistical and Probabilistic Mathematics, vol. 3. Cambridge University Press, Cambridge (1998)
42. Vimond, M.: Efficient estimation in homothetic shifted in regression models. Ann. Stat. **38**(3), 1885–1912 (2010)