The Annals of Statistics
2022, Vol. 50, No. 5, 2973–3008
https://doi.org/10.1214/22-AOS2218

NONPARAMETRIC REGRESSION ON LIE GROUPS WITH MEASUREMENT ERRORS

BY JEONG MIN JEON[1,a], BYEONG U. PARK[2,c] AND INGRID VAN KEILEGOM[1,b]

[1]*ORSTAT, KU Leuven,* [a]*jeongmin.jeon@kuleuven.be,* [b]*ingrid.vankeilegom@kuleuven.be*
[2]*Department of Statistics, Seoul National University,* [c]*bupark@stats.snu.ac.kr*

This paper develops a foundation of methodology and theory for nonparametric regression with Lie group-valued predictors contaminated by measurement errors. Our methodology and theory are based on harmonic analysis on Lie groups, which is largely unknown in statistics. We establish a novel deconvolution regression estimator, and study its rate of convergence and asymptotic distribution. We also provide asymptotic confidence intervals based on the asymptotic distribution of the estimator and on the empirical likelihood technique. Several theoretical properties are also studied for a deconvolution density estimator, which is necessary to construct our regression estimator. The case of unknown measurement error distribution is also covered. We present practical details on implementation as well as the results of simulation studies for several Lie groups. A real data example is also provided.

1. Introduction. Regression analysis is an important topic in statistics, which offers useful information on underlying relationships between variables of interest. However, when some variables are not precisely observed due to measurement errors, direct application of existing methods designed for error-free variables results in incorrect inference. To give an introductory account of the problem, let us consider a simple case where both the covariate (predictor) X and the response Y are real-valued. To estimate the regression function $m(x) = \mathrm{E}(Y|X = x)$ at a point x, one may apply "local smoothing" to Y_i around each point x. For example, the Nadaraya–Watson estimator of m is to take a weighted average of Y_i corresponding to X_i that fall in a neighborhood of each point x. This makes sense since Y_i corresponding to X_i near x have "correct" information about $m(x)$. Now, suppose that X_i are not available but $Z_i = X_i + U_i$ are, where U_i are unobserved measurement errors. In this case, the naive approach, simply taking a weighted average of Y_i corresponding to Z_i that fall in a neighborhood of the point x, should fail since X_i corresponding to such Z_i may locate far away from x, and thus the corresponding Y_i may not have correct information about m at x. To overcome the issue with measurement errors, appropriate correction methods have been proposed for the case where all variables are Euclidean. To list only a few, [72] introduced a deconvolution kernel density estimator, and [22] and [23] studied its rate of convergence and asymptotic distribution, respectively. Based on the deconvolution kernel, [24] and [16], respectively, investigated the rate of convergence and the construction of bootstrap confidence bands for a Nadaraya–Watson-type regression estimator, and [15] studied the asymptotic distribution of a local polynomial-type regression estimator. For an introduction to such errors-in-Euclidean-variables problems, we refer to [57] and [14], for example.

Analyzing non-Euclidean data is becoming an important task in modern statistics due to rapidly emerging non-Euclidean data in various fields. For a recent trend on non-Euclidean

Received July 2021; revised April 2022.
MSC2020 subject classifications. Primary 62G08; secondary 62G20.
Key words and phrases. Deconvolution, errors-in-variables, Lie groups, manifold-valued data, measurement errors.

data analysis, we refer to [56] and the references therein. In particular, functional data analysis has been widely studied in recent years ([53, 70, 78]). However, much fewer works exist for the case where a functional predictor in functional regression is contaminated by a genuinely functional measurement error ([5, 7, 39]). More recently, [54] considered a regression problem with a real-valued response and a functional predictor that takes values in a finite-dimensional submanifold of $L^2(D)$ for some compact set $D \subset \mathbb{R}$. In the latter paper, it is assumed that one observes noisy values of the functional predictor at random time points contaminated by real-valued noises. From the noisy observations, they constructed the estimated values of the functional predictor whose errors correspond to the measurement errors U_i in the aforementioned errors-in-variables problem. The estimation errors in their work are vanishing as the sample size increases to infinity, so that one does not need a deconvolution method to correct for the discrepancy between the true and observed values of the predictor. To the best of our knowledge, regression analysis with genuinely manifold-valued measurement errors has not been studied.

Manifold-valued data are also often subject to measurement errors as Euclidean and functional data are. For example, wind directions at particular times are difficult to measure exactly due to the fast speed of wind. Also, periodic time variables such as the time of a daily event and the date of an yearly event are prone to contain measurement errors unless one observes them all time. Those are examples of circular data. Another area where contaminated manifold-valued data arise is astronomy. For example, exact positioning of sunspots on the sun or measuring the oriented directions of astronomical objects to the earth is very hard since they move very fast from far away. Those are examples of spherical data. For the same reason, the orbits of comets or asteroids, which can be transformed to SO(3)-valued data ([68]), can also contain measurement errors, where $\mathrm{SO}(p)$ is the space of $p \times p$ orthogonal matrices having unit determinant. In addition, the reachable orientations of robot arms taking values in SO(3) are subject to measurement errors as noted in [55]. Another class of examples is hyperspherical data originated from Euclidean data sources. This is particularly the case when observed Euclidean vectors are normalized to have unit Euclidean norm to ensure that data analysis is affected only by the relative magnitudes of vector elements rather than their absolute magnitudes. If the original Euclidean data contain measurement errors, then the resulting hyperspherical data also contain measurement errors.

In general, analyzing manifold-valued data is challenging since there is no vector space structure on most manifolds, unlike Banach/Hilbert spaces where functional data take values in. For error-free manifold-valued data, many papers exist for density estimation (e.g., [4, 29, 32, 34, 64]) and regression analysis (e.g., [2, 6, 8, 10, 19, 30, 35, 40, 42, 49, 65, 80]). Among them, [6] considered the case where response and predictor are spherical variables and the response is symmetrically distributed around the product of an unknown orthogonal matrix and the predictor. Much fewer works exist, however, for manifold-valued variables that are contaminated by manifold-valued measurement errors, and they are restricted to density estimation. For example, [33, 48] and [46] studied deconvolution density estimation on the unit sphere $\mathbb{S}^2 = \{x \in \mathbb{R}^3 : \|x\| = 1\}$. A few others include [47] for special orthogonal groups, [50] for compact and connected Lie groups, [38] for the Poincaré upper half-plane and [55] for the 6-dimensional Euclidean motion group. All of these works on deconvolution density estimation studied only the rates of convergence of the estimators.

In this paper, we consider a regression setting where the response variable is real-valued and the predictor takes values in a compact and connected Lie group. In particular, we consider for the first time the case where the Lie group-valued predictor is contaminated by a Lie group-valued measurement error. We build up a theoretical foundation for such a new regression setup. Contrary to the aforementioned works on density estimation, the theory for nonparametric regression includes the analysis of the conditional distribution of the observed values of the contaminated predictor and response variable given unobservable true values of the predictor. Furthermore, in addition to the rate of convergence, we derive the asymptotic distribution for a novel deconvolution regression estimator and construct asymptotic confidence intervals. We develop two types of confidence intervals, one via the asymptotic distribution of the estimator and the other by empirical likelihood. We also provide several new properties of a deconvolution density estimator such as the uniform consistency, asymptotic distribution and asymptotic confidence intervals, which have not been considered before. In addition, we present full practical details of implementation and numerical studies, which have received less attention in the literature despite their importance. We emphasize that deriving such results are highly nontrivial and nonstandard, since they heavily rely on harmonic analysis on Lie groups and use various facts and tools that have not been known for non-Euclidean cases. Indeed, the derivation of the results is quite different from the ways in the Euclidean case.

Compact and connected Lie groups are important classes of non-Euclidean spaces having both manifold and algebraic group structures. Examples are toruses including the unit circle as a special case, special orthogonal groups, special unitary groups, unitary groups, compact symplectic groups, metaplectic groups and their product spaces. Hence, our setting covers various data types such as torus-valued data, which include circular data as a special case, special-orthogonal-matrix-valued data, some hyperspherical data and so on.

This paper is organized as follows. In Section 2, our regression setting and the proposed regression estimator are presented, after some terminologies are introduced, which are unfamiliar in statistics. The rates of convergence of the regression estimator (and also of the similarly defined density estimator) are provided in Section 3 and the asymptotic distributions of both estimators are given in Section 4. Construction of asymptotic confidence intervals for densities and regression functions is discussed in Section 5. Section 6 is devoted to the case where the measurement error distribution is unknown. In Section 7, the results of several simulation studies are discussed and an application to a real data set is presented. Full practical details on the implementation of our methods are collected in the Appendix for certain Lie groups. The Supplementary Material [41] contains the definitions of additional terminologies and all technical proofs.

2. Methodology.

2.1. *Preliminaries.* We start with giving the definitions of some unfamiliar notions in statistics required to introduce our methodology. We refer to standard books on harmonic analysis on Lie groups (e.g., [1, 25, 26, 66, 69]) for a comprehensive understanding of the notions. A set $\mathbb{G}$ equipped with an operation $\circ$ is called a *group* if (i) there exists an (identity) element $e \in \mathbb{G}$ such that $e \circ g = g \circ e = g$ for all $g \in \mathbb{G}$; (ii) for each $g \in \mathbb{G}$, there exists an (inverse) element $g^{-1} \in \mathbb{G}$ such that $g \circ g^{-1} = g^{-1} \circ g = e$; (iii) $g_1 \circ (g_2 \circ g_3) = (g_1 \circ g_2) \circ g_3$ for all $g_1, g_2, g_3 \in \mathbb{G}$. A group $\mathbb{G}$ equipped with a group operation $\circ$ is called a *Lie group* if it is a finite-dimensional smooth manifold such that the map $f : \mathbb{G} \times \mathbb{G} \to \mathbb{G}$ defined by $f(g_1, g_2) = g_1 \circ g_2^{-1}$ is smooth (C^∞). The definition of smooth manifold is given in the Supplementary Material S.1.

For a separable complex Hilbert space $\mathbb{H}$, we let $\mathrm{Inv}(\mathbb{H})$ denote the space of all bounded linear invertible operators from $\mathbb{H}$ to itself, and $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ denote an inner product of $\mathbb{H}$. For a Lie group $\mathbb{G}$ and a separable complex Hilbert space $\mathbb{H}$, a homomorphism $\sigma : \mathbb{G} \to \mathrm{Inv}(\mathbb{H})$ is called a *strongly continuous unitary representation* of $\mathbb{G}$ if (i) for each $h \in \mathbb{H}$ the map $g \mapsto \sigma(g)(h)$ is continuous and (ii) for each $g \in \mathbb{G}$ the map $\sigma(g) : \mathbb{H} \to \mathbb{H}$ is a unitary operator, that is, $\langle h_1, h_2 \rangle_{\mathbb{H}} = \langle \sigma(g)(h_1), \sigma(g)(h_2) \rangle_{\mathbb{H}}$ for all $h_1, h_2 \in \mathbb{H}$. Such $\mathbb{H}$ is called the *representation space* of σ and is denoted by $\mathbb{H}_\sigma$. Also, $\dim(\mathbb{H}_\sigma)$ is called the *dimension* of σ. Hereafter, we call strongly continuous unitary representation simply *representation.*

For a representation $\sigma : \mathbb{G} \to \mathrm{Inv}(\mathbb{H}_\sigma)$, a closed linear subspace $\mathbb{W}_\sigma$ of $\mathbb{H}_\sigma$ is called *invariant for* σ if $\sigma(g)(\mathbb{W}_\sigma) \subset \mathbb{W}_\sigma$ for all $g \in \mathbb{G}$. A representation $\sigma : \mathbb{G} \to \mathrm{Inv}(\mathbb{H}_\sigma)$ is called *irreducible* if $\mathbb{H}_\sigma \neq \{0_\sigma\}$ and the only invariant subspaces for σ are $\mathbb{H}_\sigma$ and $\{0_\sigma\}$, where 0_σ is the zero vector of $\mathbb{H}_\sigma$. It is known that (i) if $\mathbb{G}$ is compact, then every irreducible representation of $\mathbb{G}$ is of finite-dimension, and (ii) if $\mathbb{G}$ is Abelian, that is, $g_1 \circ g_2 = g_2 \circ g_1$ for all $g_1, g_2 \in \mathbb{G}$, then every irreducible representation of $\mathbb{G}$ is of one dimension. Two irreducible representations $\sigma : \mathbb{G} \to \mathrm{Inv}(\mathbb{H}_\sigma)$ and $\tau : \mathbb{G} \to \mathrm{Inv}(\mathbb{H}_\tau)$ are said *equivalent*, denoted by $\sigma \sim \tau$, if $\mathbb{H}_\sigma$ and $\mathbb{H}_\tau$ are isomorphic and there exists a bounded linear unitary operator $T : \mathbb{H}_\sigma \to \mathbb{H}_\tau$ satisfying $T(\sigma(g)(h)) = \tau(g)(T(h))$ for all $g \in \mathbb{G}$ and $h \in \mathbb{H}_\sigma$. The relation $\sim$ between irreducible representations of $\mathbb{G}$ is an equivalence relation. We take a representative for each equivalence class $[\sigma] = \{\tau : \tau$ is an irreducible representation of $\mathbb{G}$ such that $\tau \sim \sigma\}$ and write the space of all representatives by $\hat{\mathbb{G}}$. It is known that $\hat{\mathbb{G}}$ is countable if $\mathbb{G}$ is compact.

EXAMPLE 1. We illustrate $\hat{\mathbb{G}}$ and $\mathbb{H}_\sigma$ for several Lie groups $\mathbb{G}$ to give their tangible pictures in relation to the abstract notions for the representation of Lie groups. These would help be a better understanding of our proposed methodology in general, especially of the discussion on the exemplified Lie groups later in this paper. Some further ingredients needed in the practical implementation of the methodology for those Lie groups are discussed in the Appendix. Below, $D \geq 1$ is a given integer and $\dim(\mathbb{G})$ is the manifold dimension of $\mathbb{G}$.

(i) *Euclidean spaces*. For $\mathbb{G} = \mathbb{R}^D$ with $\circ$ being the usual addition operation $+$, we get $\dim(\mathbb{G}) = D$, $\hat{\mathbb{G}} = \{\sigma_t : t \in \mathbb{R}^D\}$, $\mathbb{H}_{\sigma_t} \equiv \mathbb{C}$ and $\dim(\mathbb{H}_{\sigma_t}) \equiv 1$, where $\sigma_t(g)(h) = h \cdot \exp(-\sqrt{-1} \cdot \langle t, g \rangle_{\mathbb{R}^D})$ for $g \in \mathbb{R}^D$ and $h \in \mathbb{C}$. We refer to Chapter 2.1 in [1] for details.

(ii) *Toruses*. Let $\mathbb{T}^D = \{g = (g_1, \dots, g_D) \in \mathbb{C}^D : |g_d| = 1$ for all $d = 1, \dots, D\}$. For $\mathbb{G} = \mathbb{T}^D$ with $\circ$ being the usual componentwise multiplication in $\mathbb{C}^D$, we have $\dim(\mathbb{G}) = D$, $\hat{\mathbb{G}} = \{\sigma_l : l = (l_1, \dots, l_D) \in \mathbb{Z}^D\}$, $\mathbb{H}_{\sigma_l} \equiv \mathbb{C}$ and $\dim(\mathbb{H}_{\sigma_l}) \equiv 1$, where $\sigma_l(g)(h) = h \cdot \prod_{d=1}^D (g_d)^{l_d}$ for $h \in \mathbb{C}$. We refer to Chapter 2.1 in [1] for details. In fact, dealing with $\mathbb{T}^D$ allows us to cover $\prod_{d=1}^D \mathbb{S}^1$, where $\mathbb{S}^1 = \{x \in \mathbb{R}^2 : \|x\| = 1\}$ is the unit circle; see Remark A.1 in the Appendix for details.

(iii) *Special unitary group of degree 2*. Consider $\mathbb{G} = \mathrm{SU}(2)$ with $\circ$ being the usual matrix multiplication, where

$$\mathrm{SU}(2) := \left\{ g = \begin{pmatrix} g_{11} & g_{12} \\ -\overline{g_{12}} & \overline{g_{11}} \end{pmatrix} : g_{11}, g_{12} \in \mathbb{C}, |g_{11}|^2 + |g_{12}|^2 = 1 \right\}$$

with $\bar{a}$ standing for the conjugate of a complex number $a \in \mathbb{C}$. In this case, $\dim(\mathbb{G}) = 3$, $\hat{\mathbb{G}} = \{\sigma_l : l \in \{0\} \cup \mathbb{N}\}$ and $\dim(\mathbb{H}_{\sigma_l}) = l + 1$. For concrete forms of σ_l and $\mathbb{H}_{\sigma_l}$, we refer to Chapter 7.5 in [25]. In fact, dealing with SU(2) allows us to cover the 3-dimensional unit hypersphere $\mathbb{S}^3 = \{x \in \mathbb{R}^4 : \|x\| = 1\}$; see Remark A.1 in the Appendix for details.

(iv) *Rotation group*. Let $O(3)$ be the space of all 3×3 real orthogonal matrices and define $\mathrm{SO}(3) = \{g \in O(3) : \det(g) = 1\}$. For $\mathbb{G} = \mathrm{SO}(3)$ with $\circ$ being the usual matrix multiplication, we obtain $\dim(\mathbb{G}) = 3$, $\hat{\mathbb{G}} = \{\sigma_l : l \in \{0\} \cup \mathbb{N}\}$ and $\dim(\mathbb{H}_{\sigma_l}) = 2l + 1$. For concrete forms of σ_l and $\mathbb{H}_{\sigma_l}$, we refer to Chapter 7.6 in [25].

(v) *Product of compact and connected Lie groups*. Let $\mathbb{G}_1$ and $\mathbb{G}_2$ be compact and connected Lie groups equipped with respective group operations $\circ_1$ and $\circ_2$. Then the product space $\mathbb{G} = \mathbb{G}_1 \times \mathbb{G}_2$ equipped with the group operation $\circ$, defined by $(g_1, g_2) \circ (g_1^*, g_2^*) = (g_1 \circ_1 g_1^*, g_2 \circ_2 g_2^*)$ for $g_1, g_1^* \in \mathbb{G}_1$ and $g_2, g_2^* \in \mathbb{G}_2$, forms a compact and connected Lie group. In this case, $\dim(\mathbb{G}) = \dim(\mathbb{G}_1) + \dim(\mathbb{G}_2)$, $\hat{\mathbb{G}} = \{\sigma_1 \otimes \sigma_2 : \sigma_1 \in \hat{\mathbb{G}}_1, \sigma_2 \in \hat{\mathbb{G}}_2\}$ and $d_{\sigma_1 \otimes \sigma_2} = d_{\sigma_1} \cdot d_{\sigma_2}$, where $\sigma_1 \otimes \sigma_2 : \mathbb{G} \to \mathrm{Inv}(\mathbb{H}_{\sigma_1} \otimes \mathbb{H}_{\sigma_2})$ is defined by $\sigma_1 \otimes \sigma_2(g_1, g_2)(h_1 \otimes h_2) = \sigma_1(g_1)(h_1) \otimes \sigma_2(g_2)(h_2)$, and $\mathbb{H}_{\sigma_1} \otimes \mathbb{H}_{\sigma_2} = \{h_1 \otimes h_2 : h_1 \in \mathbb{H}_{\sigma_1}, h_2 \in \mathbb{H}_{\sigma_2}\}$ is the tensor product of $\mathbb{H}_{\sigma_1}$ and $\mathbb{H}_{\sigma_2}$. We refer to Theorem 3.9 in [69] for more details.

A Borel measure μ on a Lie group $\mathbb{G}$ is called a *left Haar measure* if $\mu(g \circ A) = \mu(A)$ for all $g \in \mathbb{G}$ and Borel sets $A \subset \mathbb{G}$, where $g \circ A = \{g \circ a : a \in A\}$. In the case where $\mathbb{G} = \mathbb{R}^D$, the corresponding Lebesgue measure is a left Haar measure. It is known that, if $\mathbb{G}$ is compact, then there exists a unique left Haar measure μ on $\mathbb{G}$ such that $\mu(\mathbb{G}) = 1$, which is called the *normalized Haar measure* on $\mathbb{G}$.

2.2. *Model and estimation.* Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathbb{G}$ be a compact and connected Lie group equipped with a group operation $\circ$. Examples of such $\mathbb{G}$ include $\mathbb{T}^D$, special orthogonal groups, special unitary groups, unitary groups, compact symplectic groups, metaplectic groups and their product groups. It excludes the Euclidean spaces, however, since the latter spaces are not compact. We study the estimation of the nonparametric regression model

$$Y = m(X) + \epsilon, \tag{1}$$

where $Y : \Omega \to \mathbb{R}$ is a response variable with finite second moment, $X : \Omega \to \mathbb{G}$ is a predictor, $\epsilon : \Omega \to \mathbb{R}$ is an error term satisfying $\mathrm{E}(\epsilon|X) = 0$ and $m : \mathbb{G} \to \mathbb{R}$ is the regression function to estimate. We consider the case where X is contaminated by an unobservable measurement error $U : \Omega \to \mathbb{G}$ satisfying $U \perp (X, Y)$, so that we only observe $Z := U \circ X$. Here, $\perp$ stands for statistical independence. Let $(Y_1, Z_1), \dots, (Y_n, Z_n)$ be n i.i.d copies of (Y, Z) that we actually observe.

To describe our methodology of estimating the regression map m, we let f_X and f_U be the respective square integrable densities of X and U with respect to the normalized Haar measure μ on $\mathbb{G}$. Then one can easily check that the density f_Z of Z with respect to μ exists and is given by the following convolution formula:

$$f_Z(z) = (f_U * f_X)(z) := \int_{\mathbb{G}} f_U(u) f_X\big(u^{-1} \circ z\big)\, d\mu(u),$$

where u^{-1} is the inverse element of $u \in \mathbb{G}$. For a given $\sigma \in \hat{\mathbb{G}}$ and its dimension $d_\sigma \in \mathbb{N}$, let $\{e_i^\sigma : 1 \le i \le d_\sigma\}$ be an orthonormal basis of the representation space $\mathbb{H}_\sigma$ of σ. Also, for each $g \in \mathbb{G}$, let $\sigma^M(g)$ denote the $d_\sigma \times d_\sigma$ complex matrix whose (i, j)th element equals $\sigma_{ij}^M(g) := \langle \sigma(g)(e_j^\sigma), e_i^\sigma \rangle_{\mathbb{H}_\sigma} = \overline{\langle e_i^\sigma, \sigma(g)(e_j^\sigma) \rangle_{\mathbb{H}_\sigma}}$. We call $\sigma^M(g)$ the *matrix form of* $\sigma(g)$ with respect to $\{e_i^\sigma : 1 \le i \le d_\sigma\}$. Then, for V being any of X, U and Z, the Fourier transform $\phi^V(\sigma^M)$ of f_V at σ^M is given by the following matrix-valued integral:

$$\phi^V(\sigma^M) := \int_{\mathbb{G}} \sigma^M(g^{-1}) f_V(g)\, d\mu(g) = \mathrm{E}\big(\sigma^M(V^{-1})\big), \tag{2}$$

where "E" denotes expectation. Also, the following convolution identity holds (e.g., (5.1.7) in [26]):

$$\phi^Z(\sigma^M) = \phi^X(\sigma^M)\phi^U(\sigma^M), \quad \sigma \in \hat{\mathbb{G}}, \tag{3}$$

where the right-hand side is to be understood as the matrix multiplication.

REMARK 1. The definition of Fourier transform at (2) actually applies to a general Lie group. In the case of $\mathbb{R}^D$ for some $D \ge 1$, the Fourier transforms corresponding to different orthonormal bases of $\mathbb{C}$ are identical and the common transform coincides with the usual Euclidean Fourier transform. In the latter case, $d_{\sigma_t} \equiv 1$ so that σ_t^M is given by $\sigma_t^M(g^{-1}) = \sigma_t^M(-g) = \exp(\sqrt{-1} \cdot \langle t, g \rangle_{\mathbb{R}^D}) \in \mathbb{T}^1$; see Example 1(i). Also, μ equals the Lebesgue measure on $\mathbb{R}^D$.

Now, let $L^2((\mathbb{G},\mu),\mathbb{C})$ be the space of complex-valued functions $f:\mathbb{G}\to\mathbb{C}$ such that $\int_{\mathbb{G}} f(g)\overline{f(g)}\,d\mu(g)<\infty$, endowed with the inner product $\langle f_1, f_2\rangle_2=\int_{\mathbb{G}} f_1(g)\overline{f_2(g)}\,d\mu(g)$. The following lemma is a part of the so-called Peter–Weyl theorem (e.g., Theorem 5.12 in [26]).

LEMMA 1. *The collection* $\{d_\sigma^{1/2}\sigma_{ij}^M(\cdot):\sigma\in\hat{\mathbb{G}}, 1\le i,j\le d_\sigma\}$ *forms an orthonormal basis of* $L^2((\mathbb{G},\mu),\mathbb{C})$.

According to Lemma 1, f_X admits the following Fourier series expansion:

$$f_X(\cdot)=\sum_{\sigma\in\hat{\mathbb{G}}} d_\sigma \sum_{i,j=1}^{d_\sigma}\langle f_X(\cdot),\sigma_{ij}^M(\cdot)\rangle_2\sigma_{ij}^M(\cdot)=\sum_{\sigma\in\hat{\mathbb{G}}} d_\sigma \operatorname{Tr}(\phi^X(\sigma^M)\sigma^M(\cdot)), \tag{4}$$

where the equalities hold in $L^2((\mathbb{G},\mu),\mathbb{C})$ sense and $\operatorname{Tr}(A)$ for a square matrix A denotes the trace of A. In the next section, we give sufficient conditions under which (4) holds in pointwise sense. In the Euclidean measurement error case where the predictor takes values in $\mathbb{R}^D$ for some $D\ge 1$, it is usually assumed that the density of a measurement error is known and its Fourier transform is nonzero. In our non-Euclidean case, we also assume that f_U is known and $\phi^U(\sigma^M)$ is invertible for all $\sigma\in\hat{\mathbb{G}}$. This kind of assumptions is typical in the literature of non-Euclidean deconvolution density estimation (e.g., [33, 38, 46–48, 50, 55]). We discuss the case of unknown f_U in Section 6. From (3) and (4), we get the following Fourier series expansion:

$$f_X(\cdot)=\sum_{\sigma\in\hat{\mathbb{G}}} d_\sigma \operatorname{Tr}(\phi^Z(\sigma^M)\phi^U(\sigma^M)^{-1}\sigma^M(\cdot)). \tag{5}$$

We first introduce an estimator of the density f_X, which is needed to propose our estimator of the regression function m. For this, we use the representation (5). Since $\phi^Z(\sigma^M)=\mathrm{E}(\sigma^M(Z^{-1}))$ by definition, we estimate it by the empirical mean $n^{-1}\sum_{i=1}^n\sigma^M(Z_i^{-1})$, where Z^{-1} is the random element whose evaluation at $\omega\in\Omega$ is the inverse element of $Z(\omega)$ in $\mathbb{G}$. We plug the empirical mean in the place of $\phi^Z(\sigma^M)$ in (5), which gives an estimator of f_X. The estimator, however, is subject to a large variability since for each Z_i it involves the infinite sum $\sum_{\sigma\in\hat{\mathbb{G}}} d_\sigma \operatorname{Tr}(\sigma^M(Z_i^{-1})\phi^U(\sigma^M)^{-1}\sigma^M(\cdot))$, and the value of $\|\phi^U(\sigma^M)^{-1}\|_{\mathrm{op}}:=\sup\{\|\phi^U(\sigma^M)^{-1}v\|_{\mathbb{C}^{d_\sigma}}:v\in\mathbb{C}^{d_\sigma},\|v\|_{\mathbb{C}^{d_\sigma}}=1\}$ is unbounded as σ varies in $\hat{\mathbb{G}}$. For example, for the general Laplace and Gaussian distributions on $\mathbb{G}$ to be introduced in Section 3.1, we have $\|\phi^U(\sigma^M)^{-1}\|_{\mathrm{op}}=C(1+s\cdot f(\sigma))$ and $\|\phi^U(\sigma^M)^{-1}\|_{\mathrm{op}}=C\exp(s\cdot f(\sigma))$, respectively, for some constants $C,s>0$ and an unbounded nonnegative function f on $\hat{\mathbb{G}}$. This is analogous to the phenomenon in the Euclidean case that the reciprocal of a Euclidean Fourier transform tends to infinity in the tails. To avoid the difficulty, we take a finite sum over $\sigma\in\hat{\mathbb{G}}$ with a truncation point determined by the *Casimir spectrum* of $\mathbb{G}$ and a smoothing parameter $T_n>0$ such that $\lim_{n\to\infty}T_n=\infty$. The Casimir spectrum is an unbounded set $\{k_\sigma:\sigma\in\hat{\mathbb{G}}\}$ with $k_\sigma\ge 0$ such that $-k_\sigma$ for each σ is the eigenvalue of the Laplace–Beltrami operator Δ (twice differential operator acting on twice continuously differentiable functions mapping $\mathbb{G}$ to $\mathbb{C}$), corresponding to $\sigma_{ij}^M(\cdot)$ as eigenfunctions, that is, $\Delta(\sigma_{ij}^M)=-k_\sigma\cdot\sigma_{ij}^M$ for all $1\le i,j\le d_\sigma$. We refer to the Supplementary Material S.1 for the formal definition of Casimir spectrum.

We cut off the infinite sum $\sum_{\sigma\in\hat{\mathbb{G}}} d_\sigma \operatorname{Tr}(\sigma^M(Z_i^{-1})\phi^U(\sigma^M)^{-1}\sigma^M(\cdot))$ at σ where the Casimir spectrum k_σ exceeds T_n. In this way, we obtain a finite sum since the set $\{\sigma\in\hat{\mathbb{G}}:k_\sigma<T_n\}$ is finite for each n. The Casimir spectrum also has the property $\sum_{\sigma\in\hat{\mathbb{G}}:k_\sigma<T_n}d_\sigma^2\asymp$

$T_n^{\dim(\mathbb{G})/2}$ (e.g., Corollary in [58]), which is useful in controlling the strength of the variability of the resulting finite sum by adjusting T_n. The truncation gives a density estimator

$$\tilde{f}_X(x) = n^{-1}\sum_{i=1}^{n} K_{T_n}(x, Z_i),$$
$$K_{T_n}(x, z) = \sum_{\sigma\in\hat{\mathbb{G}}:k_\sigma<T_n} d_\sigma \operatorname{Tr}\bigl(\sigma^M(z^{-1})\phi^U(\sigma^M)^{-1}\sigma^M(x)\bigr), \quad x, z \in \mathbb{G}. \tag{6}$$

The above $K_{T_n}(\cdot,\cdot)$ plays the role of the Euclidean "deconvolution kernel" used in [72] and [24], which studied the cases of Euclidean measurement errors. Indeed, the L^2 error rates of $\tilde{f}_X$ were studied by [50] for $\mathbb{G}$-valued measurement errors. In this paper, we consider instead

$$\hat{f}_X(x) = n^{-1}\sum_{i=1}^{n}\operatorname{Re}\bigl(K_{T_n}(x, Z_i)\bigr), \tag{7}$$

where $\operatorname{Re}(a)$ stands for the real part of a complex number $a \in \mathbb{C}$. Taking the real part of $K_{T_n}(x, Z_i)$ for each x and Z_i as in (7) is natural since the target f_X is real-valued. It actually gives an estimator with a smaller error since $|\operatorname{Re}(a) - b| \le |a - b|$ for all $a \in \mathbb{C}$ and $b \in \mathbb{R}$. Including the imaginary part gives $\tilde{f}_X$ and an estimator of m whose asymptotic distributions are supported on $\mathbb{C}$, which would mystify the construction of the asymptotic confidence intervals on $\mathbb{R}$ for the targets $f_X(x)$ and $m(x)$. The asymptotic distribution of $\tilde{f}_X$ can be found in Proposition S.1 in the Supplementary Material S.20. The following new proposition gives that $\hat{f}_X$ integrates to one, which is a desired property for density estimators.

PROPOSITION 1. *$\int_{\mathbb{G}} K_{T_n}(x, z)\, d\mu(x) = 1$ for any $z \in \mathbb{G}$, so that $\int_{\mathbb{G}} \operatorname{Re}(K_{T_n}(x, z))\, d\mu(x) = 1$ for any $z \in \mathbb{G}$ and $\int_{\mathbb{G}} \hat{f}_X(x)\, d\mu(x) = 1$.*

Noting that $\operatorname{Re}(K_{T_n}(\cdot,\cdot))$ plays the role of the Euclidean deconvolution kernel, we define the following novel deconvolution regression estimator $\hat{m}$:

$$\hat{m}(x) = \frac{1}{\hat{f}_X(x)} \cdot n^{-1}\sum_{i=1}^{n}\operatorname{Re}\bigl(K_{T_n}(x, Z_i)\bigr)Y_i, \quad x \in \mathbb{G}. \tag{8}$$

The evaluation of $K_{T_n}(x, Z_i)$ in the estimators at (7) and (8) needs the knowledge of k_σ, σ^M and ϕ^U. In Example A.1 in the Appendix, we present the specific forms of these quantities for several popular examples of $\mathbb{G}$.

REMARK 2. The estimators $\hat{f}_X$ and $\hat{m}$ are independent of the choice of an orthonormal basis of $\mathbb{H}_\sigma$. Let $\{e_i^\sigma : 1 \le i \le d_\sigma\}$ and $\{v_i^\sigma : 1 \le i \le d_\sigma\}$ be two different orthonormal bases of $\mathbb{H}_\sigma$, and $\sigma^{M,e}(g)$ and $\sigma^{M,v}(g)$ denote the respective matrix forms of $\sigma(g)$. In Proposition S.2 in the Supplementary Material S.20, we show that $\sigma^{M,v}(g) = A\sigma^{M,e}(g)A^*$ for all $g \in \mathbb{G}$, where A is a $d_\sigma \times d_\sigma$ unitary matrix that depends only on the orthonormal bases $\{e_i^\sigma : 1 \le i \le d_\sigma\}$ and $\{v_i^\sigma : 1 \le i \le d_\sigma\}$, and A^* is the conjugate transpose of A. Since $\operatorname{Tr}(A_1A_2) = \operatorname{Tr}(A_2A_1)$ holds for all complex square matrices A_1 and A_2, we may see that the deconvolution kernels $K_{T_n}(\cdot,\cdot)$ at (6) resulting from $\{e_i^\sigma : 1 \le i \le d_\sigma\}$ and $\{v_i^\sigma : 1 \le i \le d_\sigma\}$ are identical.

REMARK 3. Note that the estimator in (8) looks like a Nadaraya–Watson-type estimator popularized for Euclidean data, but it is expressed so for the reader to connect the estimator for Lie group data to the Euclidean counterpart. The theory for the estimator in (8) and its practical implementation are far different from those for Euclidean Nadaraya–Watson-type estimators.

An important functionality that $\mathrm{Re}(K_{T_n}(\cdot,\cdot))$ needs to possess is to remove the effect of U that might be of a nonnegligible magnitude if

$$K^0_{T_n}(x,v)=\sum_{\sigma\in\hat{\mathbb{G}}:k_\sigma<T_n} d_\sigma \operatorname{Tr}(\sigma^M(v^{-1})\sigma^M(x)),\quad x,v\in\mathbb{G} \tag{9}$$

is used, instead of $K_{T_n}(\cdot,\cdot)$, with the noisy Z_i. In view of (4) and the fact $\mathrm{E}(\sigma^M(X^{-1}))=\phi^X(\sigma^M)$, one may use $K^0_{T_n}(\cdot,\cdot)$ to estimate f_X in case the *true predictor values* X_i are observed. For example, one may estimate $f_X(x)$ by $n^{-1}\sum_{i=1}^n K^0_{T_n}(x,X_i)$ or by $n^{-1}\sum_{i=1}^n \mathrm{Re}(K^0_{T_n}(x,X_i))$, where the former has been studied by [34] for the error-free case. The following new proposition gives a relationship between $K_{T_n}(\cdot,\cdot)$ and $K^0_{T_n}(\cdot,\cdot)$.

PROPOSITION 2. $\mathrm{E}(\sigma^M(Z^{-1})|X)=\sigma^M(X^{-1})\phi^U(\sigma^M)$, *so that* $\mathrm{E}(\mathrm{Re}(K_{T_n}(x,Z))|X)=\mathrm{Re}(K^0_{T_n}(x,X))$ *for any* $x\in\mathbb{G}$.

The conclusion asserted in the above proposition, termed as "unbiased scoring," implies that $\mathrm{E}(\sigma^M(Z^{-1})Y)\phi^U(\sigma^M)^{-1}=\phi^{m\cdot f_X}(\sigma^M)$, where $\phi^{m\cdot f_X}(\sigma^M)$ is the Fourier transform of $m\cdot f_X$ at σ^M given by $\int_{\mathbb{G}}\sigma^M(g^{-1})m(g)f_X(g)\,d\mu(g)$. Indeed, it holds that

$$\begin{aligned}\mathrm{E}(\sigma^M(Z^{-1})Y)\phi^U(\sigma^M)^{-1}&=\mathrm{E}(\mathrm{E}(\sigma^M(Z^{-1})Y|X))\phi^U(\sigma^M)^{-1}\\&=\mathrm{E}(\sigma^M(X^{-1})\mathrm{E}(\sigma^M(U^{-1})|X)m(X))\phi^U(\sigma^M)^{-1}\\&=\mathrm{E}(\sigma^M(X^{-1})m(X))\\&=\phi^{m\cdot f_X}(\sigma^M),\end{aligned}$$

where the second and third equalities follow from $U\perp(X,Y)$. The above result is useful in various places of our asymptotic analysis. In particular, the unbiased scoring property gives that

$$\mathrm{E}(\hat{f}_X(x)|X_1,\dots,X_n)=n^{-1}\sum_{i=1}^n \mathrm{Re}(K^0_{T_n}(x,X_i)),$$

$$\mathrm{E}(\hat{m}(x)\hat{f}_X(x)|X_1,\dots,X_n)=n^{-1}\sum_{i=1}^n \mathrm{Re}(K^0_{T_n}(x,X_i))m(X_i).$$

The above identities basically tell that the "cut-off spectral kernel" K_{T_n} defined at (6) deconvolutes efficiently the influence of measurement errors. They indicate that the biases of $\hat{f}_X$ and $\hat{m}_X$ based on the contaminated observations Z_i are the same as the respective biases of the corresponding estimators based on the unobservable original X_i.

REMARK 4. The density estimator $\tilde{f}_X$ defined at (6) can be also constructed by plugging $\int_{\mathbb{G}}\sigma^M(g^{-1})\hat{f}^0_Z(g)\,d\mu(g)$ in the place of the Fourier transform $\phi^Z(\sigma^M)$ in (5), where $\hat{f}^0_Z(g)=n^{-1}\sum_{i=1}^n K^0_{T_n}(g,Z_i)$ is the estimator of $f_Z(g)$ introduced by [34] for the case of no measurement error. In this construction, the infinite sum $\sum_{\sigma\in\hat{\mathbb{G}}}$ at (5) is automatically reduced to the finite sum $\sum_{\sigma\in\hat{\mathbb{G}}:k_\sigma<T_n}$ given in the expression of $\tilde{f}_X(x)$ due to Lemma 1. One may think of using a kernel density estimator of f_Z such as

$$\bar{f}_Z(g):=(nh)^{-1}\sum_{i=1}^n(\theta_{Z_i}(g))^{-1}K(d(g,Z_i)/h)$$

introduced by [64] instead of $\hat{f}^0_Z(g)$, where K is a baseline kernel function, h is the bandwidth, θ_{Z_i} is the volume density function at Z_i and d is the geodesic distance. The resulting density and regression estimators based on the use of $\bar{f}_Z$ may contain the infinite sum $\sum_{\sigma\in\hat{\mathbb{G}}}$ since the associated kernel weighting scheme for the estimators does not reduce automatically the infinite sum to a finite sum. Hence, the latter estimators are subject to a large variability. One can take an arbitrary finite sum, but this would introduce another tuning parameter in addition to the bandwidth h. Moreover, the estimators based on $\bar{f}_Z$ do not possess an unbiased scoring property that removes the influence of measurement errors in the bias parts.

3. Error rate analysis. In this section, we discuss the rates of convergence of our estimators defined at (7) and (8).

3.1. *Smoothness of measurement error distribution.* In the Euclidean case, two smoothness classes are usually considered for the distributions of measurement errors. They are ordinary-smoothness and supersmoothness scenarios; see, for example, [22]. A Euclidean example of ordinary-smoothness scenario is the Laplace distribution and the one with supersmoothness is the Gaussian distribution. We consider the extended notions of smoothness for the current non-Euclidean setting introduced by [50]. For a complex matrix A, define $\|A\|_{\mathrm{op}} = \sup\{\|Av\|_{\mathbb{C}^{d_\sigma}} : v \in \mathbb{C}^{d_\sigma}, \|v\|_{\mathbb{C}^{d_\sigma}} = 1\}$.

(S1) *Ordinary-smoothness class of order $\beta \geq 0$*: There exist constants $c_1, c_2 > 0$ such that, for all $\sigma \in \hat{\mathbb{G}}$ with $k_\sigma > 0$, (i) $\|\phi^U(\sigma^M)^{-1}\|_{\mathrm{op}} \leq c_1 k_\sigma^\beta$ and (ii) $\|\phi^U(\sigma^M)\|_{\mathrm{op}} \leq c_2 k_\sigma^{-\beta}$.

(S2) *Supersmoothness class of order $\beta > 0$*: There exist constants $c_1, c_2, \gamma > 0$ and $\alpha \in \mathbb{R}$ such that, for all $\sigma \in \hat{\mathbb{G}}$ with $k_\sigma > 0$, (i) $\|\phi^U(\sigma^M)^{-1}\|_{\mathrm{op}} \leq c_1 k_\sigma^\alpha \exp(\gamma \cdot k_\sigma^\beta)$ and (ii) $\|\phi^U(\sigma^M)\|_{\mathrm{op}} \leq c_2 k_\sigma^{-\alpha} \exp(-\gamma \cdot k_\sigma^\beta)$.

(S3) *Log-supersmoothness class of order $\beta > 0$*: There exist constants $c_1, c_2, \gamma > 0$ and $\alpha, \xi_1, \xi_2 \in \mathbb{R}$ such that, for all $\sigma \in \hat{\mathbb{G}}$ with $k_\sigma > 0$, (i) $\|\phi^U(\sigma^M)^{-1}\|_{\mathrm{op}} \leq c_1 k_\sigma^\alpha \exp(\gamma \cdot k_\sigma^\beta(\log k_\sigma - \xi_1))$ and (ii) $\|\phi^U(\sigma^M)\|_{\mathrm{op}} \leq c_2 k_\sigma^{-\alpha} \exp(-\gamma \cdot k_\sigma^\beta(\log k_\sigma - \xi_2))$.

For the rates of convergence, we only need the condition (i) in the respective scenarios (Sj) for $j \in \{1, 2, 3\}$. For the asymptotic distributions to be presented in Section 4.1, however, we need both conditions (i) and (ii) in (Sj). Below, we provide some examples of measurement error distributions belonging to the above scenarios. As a trivial case, consider the case of no measurement error where $P(U = e) = 1$ and $P(U = u) = 0$ for all $u \neq e$. Here and throughout this paper, e denotes the identity element of $\mathbb{G}$. We also let I_{d_σ} denote the $d_\sigma \times d_\sigma$ identity matrix. In the trivial case, $\phi^U(\sigma^M) = \sigma^M(e^{-1}) = I_{d_\sigma}$, so that its satisfies (S1) with $\beta = 0$. Another example of ordinary-smoothness is the general Laplace distribution on $\mathbb{G}$ whose Fourier transform is of the form $\phi^U(\sigma^M) = c(1 + s \cdot k_\sigma)^{-1} I_{d_\sigma}$, where $c \in \mathbb{C}$ is a nonzero constant and $s > 0$ is a real parameter. Thus, it satisfies (S1) with $\beta = 1$. An important example of supersmoothness is the general Gaussian distribution on $\mathbb{G}$ whose Fourier transform is of the form $\phi^U(\sigma^M) = c\exp(-s \cdot k_\sigma) I_{d_\sigma}$ for a nonzero complex constant $c \in \mathbb{C}$ and a real parameter $s > 0$. The distribution satisfies (S2) with $\beta = 1$, $\gamma = s$ and $\alpha = 0$. More specific examples are given in Examples A.2–A.4 in the Appendix for the Lie groups discussed in Example 1.

3.2. *Rates of convergence.* We start with giving the uniform consistency of the density estimator $\hat{f}_X$. This is new and has not been studied even for the complex-valued version of $\hat{f}_X$. We note that [50] obtained only the L^2 rates of convergence for the complex-valued $\tilde{f}_X$ defined at (6). The uniform consistency of $\hat{f}_X$ is required for the main results on our regression estimator $\hat{m}$. It is also important in its own right. For each smoothness scenario, we consider T_n growing to infinity as $n \to \infty$ at a speed as specified below:

(T1) *Ordinary-smoothness*: $n^{-1/2}T_n^{\beta+\dim(\mathbb{G})/2} = o(1)$.
(T2) *Supersmoothness*: $n^{-1/2}T_n^{\alpha+\dim(\mathbb{G})/2}\exp(\gamma\cdot T_n^{\beta}) = o(1)$.
(T3) *Log-supersmoothness*: $n^{-1/2}T_n^{\alpha+\dim(\mathbb{G})/2}\exp(\gamma\cdot T_n^{\beta}(\log T_n - \xi_1)) = o(1)$.

PROPOSITION 3. *Assume that the series in* (4) *converges uniformly to* f_X. *Then, under either of the conditions* (S1)-(i)+(T1), (S2)-(i)+(T2) *and* (S3)-(i)+(T3), *it holds that*

$$\sup_{x\in\mathbb{G}}\left|\hat{f}_X(x) - f_X(x)\right| = o_p(1).$$

A sufficient condition on f_X under which the assumption of Proposition 3 holds is that f_X is $2\times\lceil\dim(\mathbb{G})/4\rceil$-times continuously differentiable on $\mathbb{G}$, where $\lceil a\rceil = \min\{b\in\mathbb{N}: b > a\}$; see, for example, Theorem 3.3.1 in [1]. For some $\mathbb{G}$, it requires much weaker conditions than this. In the case where $\mathbb{G} = \mathbb{T}^1$, for example, the uniform convergence of the series in (4) holds if f_X is Hölder continuous with a positive exponent. In the case where $\mathbb{G} = \mathrm{SU}(2)$, the uniform convergence is implied by the Lipschitz continuity of f_X. We refer to [60] for details.

The error rate of our regression estimator $\hat{m}$ at (8) depends on the level of smoothness of f_X as well as that of m. Let $C^{\ell}(\mathbb{G})$ denote the space of ℓ-times continuously differentiable real-valued functions on $\mathbb{G}$.

(A1) For some $r > \dim(\mathbb{G})/4$ with $r\in\mathbb{N}$, (i) $f_X\in C^{2r}(\mathbb{G})$ and (ii) $m\in C^{2r}(\mathbb{G})$.
(A2) f_X is bounded away from zero on $\mathbb{G}$.

The reason we assume the same level of smoothness for f_X and m in (A1) is that the asymptotic analysis of $n^{-1}\sum_{i=1}^{n}\mathrm{Re}(K_{T_n}(x, Z_i))Y_i$ in $\hat{m}$ requires expressing the Fourier series of $m\cdot f_X$ in terms of $\Delta^r(m\cdot f_X)$, where Δ^r is the r-times composition of the Laplace–Beltrami operator Δ; see the proof of Theorem 1 in the Supplementary Material S.6 for more details. Recall that the operator Δ acts on twice continuously differentiable functions mapping $\mathbb{G}$ to $\mathbb{C}$. The condition (A2) is a regularity condition whose version for the Euclidean case is commonly assumed in nonparametric statistics.

We are ready to state our first theorem, which gives the rate of convergence of $\hat{m}$ for each smoothness class.

THEOREM 1. *Assume that the conditions* (A1) *and* (A2) *hold. Then the following results are valid*:

(a) *Under* (S1)(i) *and* (T1), *it holds that*

$$\int_{\mathbb{G}}\left|\hat{m}(x) - m(x)\right|^2 d\mu(x) = O_p\left(T_n^{-2r} + n^{-1}T_n^{2\beta+\dim(\mathbb{G})/2}\right).$$

(b) *Under* (S2)(i) *and* (T2), *it holds that*

$$\int_{\mathbb{G}}\left|\hat{m}(x) - m(x)\right|^2 d\mu(x) = O_p\left(T_n^{-2r} + n^{-1}T_n^{2\alpha+\dim(\mathbb{G})/2}\exp\left(2\gamma\cdot T_n^{\beta}\right)\right).$$

(c) *Under* (S3)(i) *and* (T3), *it holds that*

$$\int_{\mathbb{G}}\left|\hat{m}(x) - m(x)\right|^2 d\mu(x) = O_p\left(T_n^{-2r} + n^{-1}T_n^{2\alpha+\dim(\mathbb{G})/2}\exp\left(2\gamma\cdot T_n^{\beta}(\log T_n - \xi_1)\right)\right).$$

REMARK 5. We may show that $\int_{\mathbb{G}}|\hat{f}_X(x) - f_X(x)|^2\,d\mu(x)$ achieves the same rates given in Theorem 1 under (A1)(i). Proposition 3 and Theorem 1 also hold for the complex-valued versions of $\hat{f}_X$ and $\hat{m}$, respectively, under the same respective conditions.

We note that, for each smoothness class, the term of magnitude T_n^{-2r} in the L^2 error in Theorem 1 comes from the (conditional) bias of $\hat{m}$ and the other one is originated from a stochastic part contributing to the variance; see the proof of Theorem 1 in the Supplementary Material S.6. We may optimize the L^2 error rate for each smoothness class in Theorem 1 by taking a suitable speed of $T_n \to \infty$. We consider the following speeds:

(T1′) *Ordinary-smoothness*: $T_n \asymp n^{2/(4r+4\beta+\dim(\mathbb{G}))}$.
(T2′) *Supersmoothness*: $T_n = K \cdot (\log n)^{1/\beta}$ for $0 < K < (2\gamma)^{-1/\beta}$.
(T3′) *Log-supersmoothness*: $T_n = K \cdot (\log n/\log\log n)^{1/\beta}$ for $0 < K < (2\gamma/\beta)^{-1/\beta}$.

The speed of T_n in the ordinary-smoothness case actually balances the two terms of the magnitudes, T_n^{-2r} and $n^{-1}T_n^{2\beta+\dim(\mathbb{G})/2}$, so that it optimizes the L^2 error rate. In the cases of supersmoothness and log-supersmoothness, however, there exists no such thing that makes the corresponding two terms be of the same magnitude. This is because T_n also appears in the exponents of $\exp(2\gamma \cdot T_n^\beta)$ and $\exp(2\gamma \cdot T_n^\beta(\log T_n - \xi_1))$, respectively. The choices of T_n given in (T2′) and (T3′) have specific constant factors K with constraints. The upper bounds of K are actually the thresholds, beyond which $n^{-1}T_n^{2\alpha+\dim(\mathbb{G})/2}\exp(2\gamma \cdot T_n^\beta)$ and $n^{-1}T_n^{2\alpha+\dim(\mathbb{G})/2}\exp(2\gamma \cdot T_n^\beta(\log T_n - \xi_1))$, respectively, diverge to infinity, while they are dominated by T_n^{-2r} for K smaller than the thresholds. Thus, the choices of T_n are optimal in the sense that they minimize the L^2 error rates in the respective scenarios. We note that, in the Euclidean case, similar constraints are enforced on the constant factors of the bandwidths that take the role of T_n here; see, for example, Theorem 1 and Remark 1 in [24].

COROLLARY 1. *Assume that the conditions* (A1) *and* (A2) *hold. Then the following results are valid*:

(a) *Under* (S1)(i) *and* (T1′), *it holds that*

$$\int_{\mathbb{G}} |\hat{m}(x) - m(x)|^2 \, d\mu(x) = O_p\big(n^{-4r/(4r+4\beta+\dim(\mathbb{G}))}\big).$$

(b) *Under* (S2)(i) *and* (T2′), *it holds that*

$$\int_{\mathbb{G}} |\hat{m}(x) - m(x)|^2 \, d\mu(x) = O_p\big((\log n)^{-2r/\beta}\big).$$

(c) *Under* (S3)(i) *and* (T3′), *it holds that*

$$\int_{\mathbb{G}} |\hat{m}(x) - m(x)|^2 \, d\mu(x) = O_p\big((\log n/\log\log n)^{-2r/\beta}\big).$$

We note that similar rates were obtained by [24] for the Euclidean measurement error case under the Euclidean ordinary-smoothness and supersmoothness scenarios.

4. Asymptotic distributions.

4.1. *Asymptotic distributions for certain Lie groups.* In this section, we discuss the asymptotic distributions of our density and regression estimators. For non-Euclidean errors-in-variables problems, there has been no study on asymptotic distributions, even for density estimation, to the best of our knowledge. We provide the asymptotic distributions for the following Lie groups:

(G1) $\mathbb{G} = \mathbb{T}^D$ for some $D \geq 1$.
(G2) $\mathbb{G}$ is either SU(2) or SO(3), and $\phi^U(\sigma_l^M) = s_l \cdot I_{d_{\sigma_l}}$ for some $0 \neq s_l \in \mathbb{R}$ and for all $l \in \{0\} \cup \mathbb{N}$.

We note that all compact and connected Abelian Lie groups are algebraically and topologically isomorphic to $\mathbb{T}^D$ for some $D \geq 1$, that is, there exists a group isomorphism between any two compact and connected Abelian Lie groups of same dimension, which is also a homeomorphism (e.g., Chapter 4.4.2 in [66]). Since all properties of a Lie group are determined by its algebraic and topological structure, dealing with the case (G1) essentially covers all Abelian cases. As for the case (G2), the assumption $\phi^U(\sigma_l^M) = s_l \cdot I_{d_{\sigma_l}}$ is not too restrictive since it covers important distributions such as the families of Laplace and Gaussian distributions that we introduced in Section 3.1, as well as the case of no measurement error. Laplace (Gaussian) distributions on $\mathbb{G} = \mathrm{SU}(2)$ and $\mathbb{G} = \mathrm{SO}(3)$, respectively, are called the hyperspherical Laplace (Gaussian) distributions and the rotational Laplace (Gaussian) distributions, and they are respectively given in Examples A.3 and A.4. We now state the asymptotic distribution of $\hat{f}_X$.

THEOREM 2. *Assume that either* (G1) *or* (G2) *holds and that the series in* (4) *converges pointwise to* f_X. *Under either of the conditions* (S1)+(T1′), (S2)+(T2′) *and* (S3)+(T3′), *it holds that, for all* $x \in \mathbb{G}$,

$$\sqrt{n} \cdot \frac{\hat{f}_X(x) - \mathrm{E}(\mathrm{Re}(K_{T_n}(x, Z)))}{\sqrt{\mathrm{Var}(\mathrm{Re}(K_{T_n}(x, Z)))}} \xrightarrow{d} N(0, 1).$$

In the above theorem, we only assume the pointwise convergence for the series in (4), which is weaker than the uniform convergence assumed in Proposition 3. In the case where $\mathbb{G} = \mathbb{T}^D$, the Fourier series of any L^2 function converges almost everywhere to the function, according to Proposition 3.1.16 in [31]. In the case where $\mathbb{G} = \mathrm{SU}(2)$, a weak sufficient condition on the pointwise convergence of Fourier series is given by Theorem 3.15 in [60].

We now give the asymptotic distribution of $\hat{m}$ for the ordinary-smooth scenario. For this, we make a condition on the conditional moments of Y and ϵ.

(A3) $\mathrm{E}(|Y|^{2+\delta}|X = \cdot)$ is bounded on $\mathbb{G}$ for some $\delta > 0$ and $\mathrm{E}(\epsilon^2|X = \cdot)$ is bounded away from zero on $\mathbb{G}$.

The first condition in (A3) is an immediate extension of the standard regularity condition in nonparametric regression with Euclidean predictors. The second one in (A3) excludes the trivial case $\epsilon = 0$, which is of little practical importance since it is rare in real world data. In the latter case, the asymptotic distribution of $T_n^r(\hat{m}(x) - m(x))$ degenerates to zero for $T_n = n^p$ with $p < 1/(2\beta + 2r + \dim(\mathbb{G}))$, where β and r are the constants in (S1) and (A1), respectively. To state the theorem below, we recall that Δ denotes the Laplace–Beltrami operator associated with $\mathbb{G}$, and Δ^r the r-times composition of Δ. We note that, if a function $f : \mathbb{G} \to \mathbb{C}$ is $2r$-times continuously differentiable on $\mathbb{G}$, then the function $\Delta^r(f) : \mathbb{G} \to \mathbb{C}$ is well defined and is continuous on $\mathbb{G}$.

THEOREM 3. *Assume that either* (G1) *or* (G2) *holds*, *that the Fourier series of* $\Delta^r(f_X)$ *and of* $\Delta^r(m \cdot f_X)$ *converge absolutely on* $\mathbb{G}$ *and that the conditions* (S1), (A1), (A2), (A3) *and* (T1′) *hold. Then it holds that, for all* $x \in \mathbb{G}$,

$$\sqrt{n} \cdot \frac{\hat{m}(x) - m(x) - \mathrm{E}(\mathrm{Re}(K_{T_n}(x, Z))(Y - m(x)))/f_X(x)}{\sqrt{\mathrm{Var}(\mathrm{Re}(K_{T_n}(x, Z))(Y - m(x)))}/f_X(x)} \xrightarrow{d} N(0, 1).$$

Theorem 3 does not cover the supersmooth and log-supersmooth scenarios due to a technical reason. In fact, the proof of Theorem 3 involves proving

$$\sqrt{n} \cdot \frac{\mathrm{E}(\mathrm{Re}(K_{T_n}(x, Z))(Y - m(x))) \cdot (\hat{f}_X(x) - f_X(x))}{\sqrt{\mathrm{E}((\mathrm{Re}(K_{T_n}(x, Z)))^2)}} = o_p(1), \tag{10}$$

which does not hold in the supersmooth and log-supersmooth scenarios. A more detailed account is given in Remark S.1 after the proof of Theorem 3 in the Supplementary Material S.8.

For the absolute convergence of the Fourier series in Theorem 3, we note that, if a function $f : \mathbb{G} \to \mathbb{C}$ is $2 \times \lceil \dim(\mathbb{G})/4 \rceil$-times continuously differentiable, then its Fourier series is absolutely convergent (Theorem 3.1.3 in [1]). In certain cases, however, much weaker conditions suffice. For example, the Fourier series of a function on $\mathbb{T}^1$ is absolutely convergent if the function is Hölder continuous with exponent greater than 1/2; see [45]. Sufficient conditions for general toruses can be found in Theorem 3 of [76], for example. For SU(2), some weak conditions are given in [60].

4.2. *Extension to general Lie groups*. In this section, we extend the asymptotic distribution results in Section 4.1 to general $\mathbb{G}$ under some high-level conditions. To state the conditions, for any two positive sequences a_n and b_n, we let $a_n \gtrsim b_n$ (resp., $a_n \lesssim b_n$) mean that there exists a constant $c > 0$ such that, for all n, $a_n \geq c \cdot b_n$ (resp., $a_n \leq c \cdot b_n$). The constants α and γ below in the conditions are those appearing in the smoothness scenarios (S2) and (S3).

(B1) *Ordinary-smoothness*: There exists some constant $0 \leq q \leq \dim(\mathbb{G})/2$ such that, for each $x \in \mathbb{G}$, $\mathrm{E}((\mathrm{Re}(K_{T_n}(x, Z)))^2) \gtrsim T_n^{2\beta+q}$.

(B2) *Supersmoothness*: There exists some constant $0 \leq q \leq \dim(\mathbb{G})/2$ such that, for each $x \in \mathbb{G}$ and $\eta \in (0, 1)$, $\mathrm{E}((\mathrm{Re}(K_{T_n}(x, Z)))^2) \gtrsim T_n^{2\alpha+q} \exp(2\gamma(\eta \cdot T_n)^\beta)$.

(B3) *Log-supersmoothness*: There exist some constants $0 \leq q \leq \dim(\mathbb{G})/2$ and $\zeta \in \mathbb{R}$ such that, for each $x \in \mathbb{G}$ and $\eta \in (0, 1)$, $\mathrm{E}((\mathrm{Re}(K_{T_n}(x, Z)))^2) \gtrsim T_n^{2\alpha+q} \exp(2\gamma(\eta \cdot T_n)^\beta (\log T_n - \zeta))$.

We note that the lower bounds to $\mathrm{E}(\mathrm{Re}(K_{T_n}(x, Z))^2)$ in (B1)–(B3) are motivated by the upper bounds, which are of the magnitude

$$T_n^{2\beta+\dim(\mathbb{G})/2}, T_n^{2\alpha+\dim(\mathbb{G})/2} \exp\bigl(2\gamma \cdot T_n^\beta\bigr) \quad \text{or} \quad T_n^{2\alpha+\dim(\mathbb{G})/2} \exp\bigl(2\gamma \cdot T_n^\beta (\log T_n - \xi_1)\bigr),$$

depending on the smoothness scenarios. Note that $q = \dim(\mathbb{G})/2$ is maximal in view of the upper bounds. In fact, the cases (G1) and (G2) satisfy (B1)–(B3) with $q = \dim(\mathbb{G})/2$ as demonstrated in Lemma S.4 in the Supplementary Material S.2. We also consider general diverging speeds of T_n instead of the specific ones in (T1′)–(T3′) according to (B1)–(B3). We note that the following ranges cover the optimal speeds in (T1′)–(T3′):

(T1″) $T_n \asymp n^p$ for some $0 < p < 1/(\dim(\mathbb{G}) - q)$, where q is the constant in (B1).
(T2″) $T_n \lesssim (\log n)^{1/\beta}$ for β in (S2).
(T3″) $T_n \lesssim (\log n / \log\log n)^{1/\beta}$ for β in (S3).

We now state the asymptotic distribution of $\hat{f}_X$ for general $\mathbb{G}$ and T_n.

THEOREM 4. *Assume that the series in* (4) *converges pointwise to* f_X. *Under either of the conditions* (S1)-(i)+(T1″)+(B1), (S2)-(i)+(T2″)+(B2) *and* (S3)-(i)+(T3″)+(B3), *the asymptotic distribution given in Theorem* 2 *remains valid for all* $x \in \mathbb{G}$.

To state the version of the above theorem for $\hat{m}$ in the ordinary-smoothness scenario, we consider a new range of diverging T_n based on the constants β in (S1) and r in (A1).

(T1‴) $T_n \asymp n^p$ for some $1/(2\beta + 4r + \dim(\mathbb{G})/2) < p < 1/(2\beta + \dim(\mathbb{G}))$.

The above range of p is valid since r in (A1) is larger than $\dim(\mathbb{G})/4$. The upper bound $1/(2\beta + \dim(\mathbb{G}))$ is required for T_n to satisfy (T1). We also note that the range in (T1‴) covers the optimal speed in (T1′).

THEOREM 5. *Assume that the Fourier series of $\Delta^r(f_X)$ and of $\Delta^r(m \cdot f_X)$ converge absolutely on $\mathbb{G}$ and that the conditions* (S1)(i), (A1), (A2), (A3), (B1) *with* $q = \dim(\mathbb{G})/2$ *and* (T1$'''$) *hold. Then the asymptotic distribution given in Theorem* 3 *remains valid for all* $x \in \mathbb{G}$.

We may prove that the conclusion of Theorem 5 remains to hold for any $0 \le q \le \dim(\mathbb{G})/2$ with more complex versions of (A1) and (T1$'''$), although we state it with $q = \dim(\mathbb{G})/2$ for simplicity. The verification of the high-level conditions (B1)–(B3) is a challenging task since finding the exact size of $\mathrm{E}((\mathrm{Re}(K_{T_n}(x,Z)))^2)$ turns out to be difficult with the existing Taylor expansions on manifolds (e.g., [36, 52]). An easy but rough lower bound to $\mathrm{E}((\mathrm{Re}(K_{T_n}(x,Z)))^2)$ is $[\mathrm{E}(\mathrm{Re}(K_{T_n}(x,Z)))]^2$, but we only know that this lower bound is bounded away from zero. Below, we introduce one way of verifying (B1)–(B3) with $q = \dim(\mathbb{G})/2$, which gives the sharpest lower bounds. One can verify (B1)–(B3) similarly with $q = 0$ by arguing as in the case of $q = \dim(\mathbb{G})/2$ using Proposition S.3 in the Supplementary Material S.20.

We first investigate $\mathrm{E}(|K_{T_n}(x,Z)|^2)$ instead of $\mathrm{E}((\mathrm{Re}(K_{T_n}(x,Z)))^2)$ and show that it attains the lower bounds given in (B1)–(B3) with $q = \dim(\mathbb{G})/2$. For this, we introduce a condition. To state the condition, we let $\|\cdot\|_{\mathrm{HS}}$ denote the Hilbert–Schmidt norm on complex matrices. For a square complex matrix A, it is defined by $\|A\|_{\mathrm{HS}} = (\mathrm{Tr}(AA^*))^{1/2}$, where A^* denotes the conjugate transpose of A.

(B4) There exists a constant $c > 0$ such that, for all $\sigma \in \hat{\mathbb{G}}$, $c \cdot d_\sigma^{1/2} \|\phi^U(\sigma^M)^{-1}\|_{\mathrm{op}} \le \|\phi^U(\sigma^M)^{-1}\|_{\mathrm{HS}}$.

We note that $\|\phi^U(\sigma^M)^{-1}\|_{\mathrm{op}} \le \|\phi^U(\sigma^M)^{-1}\|_{\mathrm{HS}}$ always holds. Thus, if $\sup_{\sigma\in\hat{\mathbb{G}}} d_\sigma < \infty$, then the condition (B4) is satisfied with $c = (\sup_{\sigma\in\hat{\mathbb{G}}} d_\sigma)^{-1/2}$. We refer to Theorem 1 in [59] for an equivalent condition to $\sup_{\sigma\in\hat{\mathbb{G}}} d_\sigma < \infty$. This implies that (B4) is satisfied with $c = 1$ for any distribution of U on Abelian $\mathbb{G}$, since $d_\sigma \equiv 1$ for Abelian $\mathbb{G}$. Also, since $\|I_{d_\sigma}\|_{\mathrm{op}} = 1$ and $\|I_{d_\sigma}\|_{\mathrm{HS}} = d_\sigma^{1/2}$, (B4) with $c = 1$ is satisfied for the families of Laplace and Gaussian distributions introduced in Section 3.1 as well as the case of no measurement error. We note that the cases (G1) and (G2) also satisfy (B4) with $c = 1$.

LEMMA 2. *Assume that the conditions* (A2) *and* (B4) *hold. Then, for each* $j \in \{1,2,3\}$, *under* (S*j*)(ii), $\mathrm{E}(|K_{T_n}(x,Z)|^2)$ *attains the lower bound given in* (B*j*) *with* $q = \dim(\mathbb{G})/2$.

Now, we present a version of Lemma 2 for $\mathrm{E}((\mathrm{Re}(K_{T_n}(x,Z)))^2)$. We note that the latter is not direct from Lemma 2 since $\mathrm{E}(|K_{T_n}(x,Z)|^2) \ge \mathrm{E}((\mathrm{Re}(K_{T_n}(x,Z)))^2)$. Hence, we consider the case where

$$\int_{\mathbb{G}} |K_{T_n}(x,z)|^2 \, d\mu(z) \asymp \int_{\mathbb{G}} \mathrm{Re}\big(K_{T_n}(x,z)\big)^2 \, d\mu(z). \tag{11}$$

Since $\int_{\mathbb{G}} |K_{T_n}(x,z)|^2 \, d\mu(z)$ attains the same lower bounds as those to $\mathrm{E}(|K_{T_n}(x,Z)|^2)$ as demonstrated in the proof of Lemma 2, (11) with (A2) gives $\mathrm{E}(|K_{T_n}(x,Z)|^2) \asymp \mathrm{E}((\mathrm{Re}(K_{T_n}(x,Z)))^2)$, so that (B1)–(B3) follow with $q = \dim(\mathbb{G})/2$ from Lemma 2.

LEMMA 3. *Assume that the conditions* (A2) *and* (B4) *hold and that* (11) *holds. Then, for each* $j \in \{1,2,3\}$, *under* (S*j*)(ii), (B*j*) *holds with* $q = \dim(\mathbb{G})/2$.

The approximation (11) holds for the cases (G1) and (G2) as demonstrated in Lemma S.4 in the Supplementary Material S.2. We conjecture that it holds for general $\mathbb{G}$ and f_U although we could not prove it due to the computational complexity. We leave it as an open problem.

5. Asymptotic confidence intervals. In this section, we provide two types of asymptotic confidence intervals for both $f_X(x)$ and $m(x)$. One type is based on the asymptotic normal distributions as given by Theorems 2 and 3 in Section 4.1, and the other based on empirical likelihoods.

5.1. *Confidence intervals based on normal approximation.* Typically, the first step in deriving a confidence interval based on a normal approximation is to find the leading terms, with exact constant factors, of the bias and variance of the estimator under study. However, this does not seem to be feasible for $\hat{f}_X$ and $\hat{m}$ in our non-Euclidean setting, because the existing Taylor expansions on manifolds, such as those in [36] and [52], are not helpful. We estimate the biases and variances directly without quantifying their leading terms. The bias parts are $\mathrm{E}(\mathrm{Re}(K_{T_n}(x,Z))) - f_X(x)$ and $\mathrm{E}(\mathrm{Re}(K_{T_n}(x,Z))(Y - m(x)))$ for $\hat{f}_X$ and $\hat{m}$, respectively; see Theorems 2 and 3. We estimate them simply by zero. Indeed, plugging $\hat{f}_X(x) = n^{-1}\sum_{i=1}^n \mathrm{Re}(K_{T_n}(x,Z_i))$ into $\mathrm{E}(\mathrm{Re}(K_{T_n}(x,Z)))$ and $f_X(x)$, and $n^{-1}\sum_{i=1}^n \mathrm{Re}(K_{T_n}(x,Z_i))(Y_i - \hat{m}(x))$ into $\mathrm{E}(\mathrm{Re}(K_{T_n}(x,Z))(Y - m(x)))$, gives zero estimates. One might employ a sophisticated method of bias correction here, but we do not pursue it in this paper.

To justify the zero estimators of the biases in the construction of confidence intervals based on Theorems 2 and 3, it is essential to verify that the variances dominate the squared biases. In the case of $\hat{f}_X$, this amounts to showing

$$\sqrt{n}\cdot\frac{\mathrm{E}(\mathrm{Re}(K_{T_n}(x,Z))) - f_X(x)}{\sqrt{\mathrm{Var}((\mathrm{Re}(K_{T_n}(x,Z))))}} = o(1).$$

It turns out that the latter is implied by

$$\sqrt{n}\cdot\frac{\mathrm{E}(\mathrm{Re}(K_{T_n}(x,Z))) - f_X(x)}{\sqrt{\mathrm{E}((\mathrm{Re}(K_{T_n}(x,Z)))^2)}} = o(1). \tag{12}$$

However, $\mathrm{E}((\mathrm{Re}(K_{T_n}(x,Z)))^2)$ with the rates in (T2′) and (T3′) do not satisfy (12) in the scenarios (S2) and (S3). Hence, we construct a confidence interval for f_X only for the scenario (S1). As we mentioned above, we estimate $\mathrm{E}(\mathrm{Re}(K_{T_n}(x,Z))) - f_X(x)$ by zero. For the variance $V_{f_X}(x) := \mathrm{Var}(\mathrm{Re}(K_{T_n}(x,Z)))$, we estimate it by

$$\hat{V}_{f_X}(x) := n^{-1}\sum_{i=1}^n \big(\mathrm{Re}\big(K_{T_n}(x,Z_i)\big)\big)^2 - \big(\hat{f}_X(x)\big)^2.$$

By proving (12) and $\hat{V}_{f_X}(x)/V_{f_X}(x) \to 1$ in probability for all $x \in \mathbb{G}$, we get the following theorem.

THEOREM 6. *Assume that either* (G1) *or* (G2) *holds, that the Fourier series of* $\Delta^r(f_X)$ *converges absolutely on* $\mathbb{G}$ *and that the conditions* (S1), (A1)(i), (A2) *and* (T1′) *hold. Then, for all* $x \in \mathbb{G}$,

$$\sqrt{n}\cdot\frac{\hat{f}_X(x) - f_X(x)}{(\hat{V}_{f_X}(x))^{1/2}} \xrightarrow{d} N(0,1).$$

Therefore, a $(1-\alpha)\times 100\%$ *asymptotic confidence interval for* $f_X(x)$ *is given by*

$$\left(\hat{f}_X(x) - z_{\alpha/2}\cdot\frac{(\hat{V}_{f_X}(x))^{1/2}}{\sqrt{n}},\ \hat{f}_X(x) + z_{\alpha/2}\cdot\frac{(\hat{V}_{f_X}(x))^{1/2}}{\sqrt{n}}\right).$$

Similarly, we construct a confidence interval for m, now based on the asymptotic distribution in Theorem 3. Again, we estimate $\mathrm{E}(\mathrm{Re}(K_{T_n}(x,Z))(Y-m(x)))$ in the bias of $\hat{m}$ by zero. For the variance $V_m(x) := \mathrm{Var}(\mathrm{Re}(K_{T_n}(x,Z))(Y-m(x)))$, we estimate it by

$$\hat{V}_m(x) := n^{-1}\sum_{i=1}^{n}\big(\mathrm{Re}\big(K_{T_n}(x,Z_i)\big)\big(Y_i-\hat{m}(x)\big)\big)^2.$$

THEOREM 7. *Assume that either* (G1) *or* (G2) *holds, that the Fourier series of* $\Delta^r(f_X)$ *and of* $\Delta^r(m\cdot f_X)$ *converge absolutely on* $\mathbb{G}$ *and that the conditions* (S1), (A1), (A2), (A3) *and* (T1′) *hold. Then, for all* $x\in\mathbb{G}$,

$$\sqrt{n}\cdot\frac{\hat{m}(x)-m(x)}{(\hat{V}_m(x))^{1/2}/\hat{f}_X(x)} \xrightarrow{d} N(0,1).$$

Therefore, a $(1-\alpha)\times 100\%$ *asymptotic confidence interval for* $m(x)$ *is given by*

$$\left(\hat{m}(x)-z_{\alpha/2}\cdot\frac{(\hat{V}_m(x))^{1/2}/\hat{f}_X(x)}{\sqrt{n}},\hat{m}(x)+z_{\alpha/2}\cdot\frac{(\hat{V}_m(x))^{1/2}/\hat{f}_X(x)}{\sqrt{n}}\right).$$

5.2. *Confidence regions based on empirical likelihood.* Asymptotic confidence regions based on empirical likelihoods, called empirical likelihood confidence regions, are useful alternatives to those based on asymptotic distributions. The empirical likelihood technique has many advantages in constructing confidence regions. It allows the data to determine the shape of the confidence region, respects the boundaries of the area where possible values of the parameter of interest belong, produces transformation-invariant confidence regions and often does not require the estimation of the variance. We refer to [63] for an introduction to empirical likelihood methods. A broad review and a general theory for empirical likelihood methods can be found in [9] and [37], respectively. Some examples of applying the technique to the case of measurement errors include Euclidean density estimation ([71]) and Euclidean partially linear regression ([79]).

To construct the empirical likelihood confidence regions for f_X and m, let $F_{f_X}(z,\theta;\cdot) = \mathrm{Re}(K_{T_n}(\cdot,z))-\theta$ and $F_m(z,y,\theta;\cdot)=\mathrm{Re}(K_{T_n}(\cdot,z))(y-\theta)$ for $\theta\in\mathbb{R}$, both as functions defined on $\mathbb{G}$. Define the corresponding empirical likelihood ratio functions on $\mathbb{G}$ by

$$\mathrm{EL}_{f_X}(\theta;\cdot)=\max\left\{\prod_{i=1}^{n}(nw_i):w_1,\dots,w_n>0,\sum_{i=1}^{n}w_i=1,\sum_{i=1}^{n}w_iF_{f_X}(Z_i,\theta;\cdot)=0\right\},$$

$$\mathrm{EL}_{m}(\theta;\cdot)=\max\left\{\prod_{i=1}^{n}(nw_i):w_1,\dots,w_n>0,\sum_{i=1}^{n}w_i=1,\sum_{i=1}^{n}w_iF_{m}(Z_i,Y_i,\theta;\cdot)=0\right\}.$$

Here, we set the maximum of an empty set to be zero. Then we define the respective empirical likelihood confidence regions for $f_X(x)$ and $m(x)$, respectively, as $\{\theta\in\mathbb{R}:\mathrm{EL}_{f_X}(\theta;x)\ge c_{f_X}\}$ and $\{\theta\in\mathbb{R}:\mathrm{EL}_m(\theta;x)\ge c_m\}$ for suitable positive constants c_{f_X} and c_m. To determine the constants, we derive the asymptotic distributions of the empirical likelihood ratio functions. For this, we make the following assumptions:

(E1) $P(\mathrm{EL}_{f_X}(f_X(x);x)>0)\to 1$ for each $x\in\mathbb{G}$.
(E2) $P(\mathrm{EL}_m(m(x);x)>0)\to 1$ for each $x\in\mathbb{G}$.

The conditions (E1) and (E2) are basic in empirical likelihood methods. The inequality $\mathrm{EL}_{f_X}(f_X(x);x)>0$ is satisfied as long as there are at least two data points Z_i and Z_j such that $F_{f_X}(Z_i,f_X(x);x)>0$ and $F_{f_X}(Z_j,f_X(x);x)<0$. Likewise, $\mathrm{EL}_m(m(x);x)>0$ is satisfied as long as there are at least two data points (Z_i,Y_i) and (Z_j,Y_j) such that

$F_m(Z_i, Y_i, m(x); x) > 0$ and $F_m(Z_j, Y_j, m(x); x) < 0$. In the statements of the theorems below, $\chi^2_\alpha(1)$ denotes the $(1-\alpha)$ quantile of the chi-square distribution with degree of freedom 1.

THEOREM 8. *Assume that either* (G1) *or* (G2) *holds, that the Fourier series of* $\Delta^r(f_X)$ *converges absolutely on* $\mathbb{G}$ *and that the conditions* (S1), (A1)(i), (A2), (T1′) *and* (E1) *hold. Then, for all* $x \in \mathbb{G}$,

$$-2\log \mathrm{EL}_{f_X}(f_X(x); x) \xrightarrow{d} \chi^2(1).$$

Therefore, a $(1-\alpha) \times 100\%$ *asymptotic confidence region for* $f_X(x)$ *is given by* $\{\theta \in \mathbb{R} : -2\log \mathrm{EL}_{f_X}(\theta; x) \le \chi^2_\alpha(1)\} = \{\theta \in \mathbb{R} : \mathrm{EL}_{f_X}(\theta; x) \ge \exp(-\chi^2_\alpha(1)/2)\}$.

THEOREM 9. *Assume that either* (G1) *or* (G2) *holds, that the Fourier series of* $\Delta^r(f_X)$ *and of* $\Delta^r(m \cdot f_X)$ *converge absolutely on* $\mathbb{G}$ *and that the conditions* (S1), (A1), (A2), (A3), (T1′) *and* (E2) *hold. Then, for all* $x \in \mathbb{G}$,

$$-2\log \mathrm{EL}_m(m(x); x) \xrightarrow{d} \chi^2(1).$$

Therefore, a $(1-\alpha) \times 100\%$ *asymptotic confidence region for* $m(x)$ *is given by* $\{\theta \in \mathbb{R} : -2\log \mathrm{EL}_m(\theta; x) \le \chi^2_\alpha(1)\} = \{\theta \in \mathbb{R} : \mathrm{EL}_m(\theta; x) \ge \exp(-\chi^2_\alpha(1)/2)\}$.

The asymptotic confidence regions in Theorems 8 and 9 are in fact intervals. This is because $t\theta_1 + (1-t)\theta_2$ for $0 < t < 1$ belongs to the confidence regions whenever θ_1 and θ_2 belong to those regions. In the practical implementation of the confidence regions, we need to compute $\mathrm{EL}_{f_X}(\theta; x)$ and $\mathrm{EL}_m(\theta; x)$. A Lagrange multiplier technique shows that the unique maximizing weights w_i are $1/(n(1+\lambda_{f_X} F_{f_X}(Z_i, \theta; x)))$ for $\mathrm{EL}_{f_X}(\theta; x)$, and are $1/(n(1+\lambda_m F_m(Z_i, Y_i, \theta; x)))$ for $\mathrm{EL}_m(\theta; x)$, where $\lambda_{f_X} \in \mathbb{R}$ and $\lambda_m \in \mathbb{R}$ are the solutions of

$$\sum_{i=1}^n \frac{F_{f_X}(Z_i, \theta; x)}{1+\lambda_{f_X} F_{f_X}(Z_i, \theta; x)} = 0 \quad \text{and} \quad \sum_{i=1}^n \frac{F_m(Z_i, Y_i, \theta; x)}{1+\lambda_m F_m(Z_i, Y_i, \theta; x)} = 0, \tag{13}$$

respectively.

REMARK 6. Although the asymptotic confidence intervals are obtained only for the cases (G1) and (G2), versions of Theorems 6, 7, 8 and 9 can be readily obtained for other $\mathbb{G}$ and f_U if they satisfy (B1)–(B3) with $q = \dim(\mathbb{G})/2$. We note that, however, the cases (G1) and (G2) still cover a large important class of $\mathbb{G}$ and f_U.

6. Cases of unknown measurement error distribution. In this section, we extend the results in Section 3.2 to the case where the measurement error distribution f_U is unknown. We refer to [18] for a review on Euclidean measurement error problems with unknown measurement error distribution.

6.1. *General results.* We first study the effect of estimating $\phi^U(\sigma^M)$ with an arbitrary estimator $\hat{\phi}^U(\sigma^M)$ such that $\hat{\phi}^U(\sigma^M)^{-1}$ exists. We write $\nu = m \cdot f_X$ and let $\phi^\nu(\sigma^M)$ denote its Fourier transform at σ^M. Let $\hat{\phi}^Z(\sigma^M)$ be an arbitrary estimator of $\phi^Z(\sigma^M) = \phi^X(\sigma^M)\phi^U(\sigma^M)$ and $\hat{\phi}^{\nu,U}(\sigma^M)$ an arbitrary estimator of $\phi^{\nu,U}(\sigma^M) := \phi^\nu(\sigma^M)\phi^U(\sigma^M)$. Natural examples of these are $\hat{\phi}^Z(\sigma^M) = n^{-1}\sum_{i=1}^n \sigma^M(Z_i^{-1})$ and $\hat{\phi}^{\nu,U}(\sigma^M) =$

$n^{-1}\sum_{i=1}^n \sigma^M(Z_i^{-1})Y_i$. We note that $\mathrm{E}(\sigma^M(Z^{-1})Y) = \phi^{\nu,U}(\sigma^M)$. We define new density and regression estimators,

$$\hat{f}_X^*(x) = \mathrm{Re}\Bigg(\sum_{\sigma\in\hat{\mathbb{G}}:k_\sigma<T_n} d_\sigma \operatorname{Tr}\big(\hat{\phi}^Z(\sigma^M)\hat{\phi}^U(\sigma^M)^{-1}\sigma^M(x)\big)I_n(\sigma^M)\Bigg),$$

$$\hat{m}^*(x) = \frac{1}{\hat{f}_X^*(x)}\mathrm{Re}\Bigg(\sum_{\sigma\in\hat{\mathbb{G}}:k_\sigma<T_n} d_\sigma \operatorname{Tr}\big(\hat{\phi}^{\nu,U}(\sigma^M)\hat{\phi}^U(\sigma^M)^{-1}\sigma^M(x)\big)I_n(\sigma^M)\Bigg),$$

where $I_n(\sigma^M) = I(\|\hat{\phi}^U(\sigma^M)\|_{\mathrm{op}} \ge a_n)$ and a_n is a positive sequence converging to zero. Introducing the indicator term $I_n(\sigma^M)$ in the above definitions is necessary in practice to prevent a large variability induced by possibly small values of $\|\hat{\phi}^U(\sigma^M)\|_{\mathrm{op}}$. This treatment is also important in our theoretical development. Such a truncation is common as well in Euclidean measurement error problems with unknown measurement error distribution (e.g., [13, 17, 61]).

We start with the uniform error rate for $\hat{f}_X^*$, which is required to derive the L^2 error rate for $\hat{m}^*$. For this, we let C_σ denote the condition number of $\phi^U(\sigma^M)$, that is, $C_\sigma = \|\phi^U(\sigma^M)\|_{\mathrm{op}}\|\phi^U(\sigma^M)^{-1}\|_{\mathrm{op}}$. Likewise, let $\hat{C}_\sigma$ denote the condition number of $\hat{\phi}^U(\sigma^M)$. We choose a_n such that

$$0 < a_n < \begin{cases} c_1^{-1}T_n^{-\beta} & \text{for (S1)},\\ c_1^{-1}T_n^{-\alpha}\exp(-\gamma\cdot T_n^\beta) & \text{for (S2)},\\ c_1^{-1}T_n^{-\alpha}\exp\big(-\gamma\cdot T_n^\beta(\log T_n-\xi_1)\big) & \text{for (S3)},\end{cases}$$

for sufficiently large n, where c_1 is the constant in (Sj) for each $j \in \{1,2,3\}$. We note that the above upper bounds for a_n are the reciprocals of the respective upper bounds for $\|\phi^U(\sigma^M)^{-1}\|_{\mathrm{op}}$ in (Sj)(i) with k_σ being replaced by T_n.

THEOREM 10. *Assume that $\phi^U(\sigma^M)\hat{\phi}^U(\sigma^M) = \hat{\phi}^U(\sigma^M)\phi^U(\sigma^M)$ for all $\sigma \in \hat{\mathbb{G}}$ and that $\|\phi^U(\sigma^M)\|_{\mathrm{op}} \ge \|\phi^U(\tau^M)\|_{\mathrm{op}}$ for all $\sigma,\tau \in \hat{\mathbb{G}}$ with $k_\sigma \le k_\tau$. Also, assume that the series in* (4) *converges uniformly to f_X. Then, for each $j \in \{1,2,3\}$, under* (Sj)(i), *it holds that*

$$\begin{aligned}&\sup_{x\in\mathbb{G}}\big|\hat{f}_X^*(x) - f_X(x)\big|\\ &\quad= o_p(1) + O_p\Bigg(a_n^{-1}\sum_{\sigma\in\hat{\mathbb{G}}:k_\sigma<T_n} d_\sigma^{3/2}\mathrm{E}\big(\hat{C}_\sigma\cdot\|\hat{\phi}^Z(\sigma^M)-\phi^Z(\sigma^M)\|_{\mathrm{HS}}\cdot I_n(\sigma^M)\big)\\ &\qquad+ a_n^{-1}\sum_{\sigma\in\hat{\mathbb{G}}:k_\sigma<T_n} d_\sigma^{3/2}\|\phi^X(\sigma^M)\|_{\mathrm{op}}C_\sigma\mathrm{E}\big(\hat{C}_\sigma\cdot\|\hat{\phi}^U(\sigma^M)-\phi^U(\sigma^M)\|_{\mathrm{HS}}\cdot I_n(\sigma^M)\big)\\ &\qquad+ (b_n-a_n)^{-2}\sum_{\sigma\in\hat{\mathbb{G}}:k_\sigma<T_n} d_\sigma^{3/2}\|\phi^X(\sigma^M)\|_{\mathrm{HS}}\mathrm{E}\big(\|\hat{\phi}^U(\sigma^M)-\phi^U(\sigma^M)\|_{\mathrm{op}}^2\big)\Bigg),\end{aligned}$$

where $b_n = c_1^{-1}T_n^{-\beta}$ for (S1), $c_1^{-1}T_n^{-\alpha}\exp(-\gamma\cdot T_n^\beta)$ *for* (S2) *and* $c_1^{-1}T_n^{-\alpha}\exp(-\gamma\cdot T_n^\beta(\log T_n-\xi_1))$ *for* (S3).

We put the conditions on $\phi^U(\sigma^M)$ and $\hat{\phi}^U(\sigma^M)$ in Theorem 10 for technical reasons. The first condition in Theorem 10 clearly holds when $\phi^U(\sigma^M) = c_\sigma I_{d_\sigma}$ for some $0 \ne c_\sigma \in \mathbb{C}$. We note that every distribution on Abelian $\mathbb{G}$ has $\phi^U(\sigma^M)$ of this form since $d_\sigma \equiv 1$ in that case. Also, the families of Laplace and Gaussian distributions on $\mathbb{G}$ that we introduced in Section 3.1 have $\phi^U(\sigma^M)$ of this form. More generally, this condition holds when $\phi^U(\sigma^M)$ and $\hat{\phi}^U(\sigma^M)$ are symmetric matrices such that either $\phi^U(\sigma^M)\hat{\phi}^U(\sigma^M)$ or $\hat{\phi}^U(\sigma^M)\phi^U(\sigma^M)$ is symmetric. The second condition in Theorem 10 also holds for most distributions. For example, the distributions in Examples A.2–A.4 satisfy the condition. We note that C_σ is uniformly bounded over σ by 1 if $\phi^U(\sigma^M)$ is of the form $\phi^U(\sigma^M) = c_\sigma I_{d_\sigma}$, and by $\max\{1, c_1 c_2\}$ if one of (S1)–(S3) holds, where c_1 and c_2 are the constants in (Sj). In the next section, we show that $\hat{f}_X^*(x)$ achieves the uniform consistency with certain choices of $\hat{\phi}^U(\sigma^M)$, $\hat{\phi}^Z(\sigma^M)$, T_n and a_n.

We now provide the L^2 error rates of $\hat{f}_X^*$ and $\hat{m}^*$. For this, we define the following term for an integrable function f on $\mathbb{G}$:

$$R_n(f) = a_n^{-2} \sum_{\sigma\in\hat{\mathbb{G}}:k_\sigma<T_n} d_\sigma \|\phi^f(\sigma^M)\|_{\mathrm{op}}^2 C_\sigma^2 \mathrm{E}(\hat{C}_\sigma^2 \cdot \|\hat{\phi}^U(\sigma^M) - \phi^U(\sigma^M)\|_{\mathrm{HS}}^2 \cdot I_n(\sigma^M))$$
$$+ (b_n - a_n)^{-2} \sum_{\sigma\in\hat{\mathbb{G}}:k_\sigma<T_n} d_\sigma \|\phi^f(\sigma^M)\|_{\mathrm{HS}}^2 \mathrm{E}(\|\hat{\phi}^U(\sigma^M) - \phi^U(\sigma^M)\|_{\mathrm{op}}^2),$$

where b_n is the sequence in Theorem 10.

THEOREM 11. *Assume the first two conditions in Theorem 10 and the condition* (A1)(i). *Then, for each $j \in \{1, 2, 3\}$, under (Sj)*(i), *it holds that*

$$\int_{\mathbb{G}} |\hat{f}_X^*(x) - f_X(x)|^2 \, d\mu(x)$$
$$= O_p\Big(T_n^{-2r} + R_n(f_X) + a_n^{-2} \sum_{\sigma\in\hat{\mathbb{G}}:k_\sigma<T_n} d_\sigma \cdot \mathrm{E}\big[\hat{C}_\sigma^2 \cdot \|\hat{\phi}^Z(\sigma^M) - \phi^Z(\sigma^M)\|_{\mathrm{HS}}^2 \cdot I_n(\sigma^M)\big]\Big).$$

If we further assume the conditions (A1)(ii) *and* (A2), *and that* $\sup_{x\in\mathbb{G}} |\hat{f}_X^*(x) - f_X(x)| = o_p(1)$, *then for each $j \in \{1, 2, 3\}$, under (Sj)*(i), *it holds that*

$$\int_{\mathbb{G}} |\hat{m}^*(x) - m(x)|^2 \, d\mu(x)$$
$$= O_p\Big(T_n^{-2r} + R_n(f_X) + R_n(\nu) + a_n^{-2} \sum_{\sigma\in\hat{\mathbb{G}}:k_\sigma<T_n} d_\sigma$$
$$\cdot \mathrm{E}\big[\hat{C}_\sigma^2 \cdot (\|\hat{\phi}^Z(\sigma^M) - \phi^Z(\sigma^M)\|_{\mathrm{HS}}^2 + \|\hat{\phi}^{\nu,U}(\sigma^M) - \phi^\nu(\sigma^M)\phi^U(\sigma^M)\|_{\mathrm{HS}}^2) \cdot I_n(\sigma^M)\big]\Big).$$

In the next section, we show that the above L^2 error rates achieve the same or similar L^2 error rates given in Corollary 1 for certain choices of $\hat{\phi}^U(\sigma^M)$, $\hat{\phi}^Z(\sigma^M)$, $\hat{\phi}^{\nu,U}(\sigma^M)$, T_n and a_n.

6.2. *Specific results.* In Euclidean measurement error problems with unknown measurement error distribution, the following two situations are usually considered for the estimation of f_U or ϕ^U: (i) there is a sample from the measurement error distribution (e.g., [13, 20, 43, 44, 61]); (ii) there are repeated measurements of covariates for each subject (e.g., [11, 17, 44, 62]). In fact, such situations are frequent in the real world as noted in the literature. For simplicity in both situations, we focus on the case where $\phi^U(\sigma^M) = c_\sigma I_{d_\sigma}$ for some $0 \neq c_\sigma \in \mathbb{C}$ satisfying $|c_\sigma| \geq |c_\tau|$ for $k_\sigma \leq k_\tau$. In this case, the conditions on $\phi^U(\sigma^M)$ and $\hat{\phi}^U(\sigma^M)$ in Theorem 10 hold. As we noted in Section 6.1, the case covers the most important distributions. One may be able to treat cases with more general $\phi^U(\sigma^M)$ along the lines of the theoretical development to be presented below, at a cost of more complexity under slightly different conditions.

We first consider the case where we have a random sample $\{\tilde{U}_j : 1 \le j \le N\}$ from f_U. In this case, we take $\hat{\phi}^U(\sigma^M) = \hat{c}_\sigma I_{d_\sigma}$ with $\hat{c}_\sigma = d_\sigma^{-1} \mathrm{Tr}(N^{-1} \sum_{j=1}^N \sigma^M(\tilde{U}_j^{-1}))$. We note that $\hat{\phi}^U(\sigma^M)$ reduces to $N^{-1} \sum_{j=1}^N \sigma^M(\tilde{U}_j^{-1})$ for Abelian $\mathbb{G}$. We also choose $\hat{\phi}^Z(\sigma^M) = n^{-1} \sum_{i=1}^n \sigma^M(Z_i^{-1})$ and $\hat{\phi}^{\nu,U}(\sigma^M) = n^{-1} \sum_{i=1}^n \sigma^M(Z_i^{-1}) Y_i$. For T_n and a_n, we consider the following choices:

$$T_n = \begin{cases} L_1 \cdot n^{2/(4r+4\beta+\dim(\mathbb{G}))} & \text{for (S1)}, \\ L_2 \cdot (\log n)^{1/\beta} & \text{for (S2)}, \\ L_3 \cdot (\log n / \log\log n)^{1/\beta} & \text{for (S3)}, \end{cases}$$
$$(14) \qquad a_n = \begin{cases} \tilde{L} \cdot n^{-2\beta/(4r+4\beta+\dim(\mathbb{G}))} & \text{for (S1)}, \\ n^{-1/(2+\delta_2)} & \text{for (S2)}, \\ n^{-1/(2+\delta_3)} & \text{for (S3)}, \end{cases}$$

where $\delta_2, \delta_3 > 0$. In the above specifications of T_n and a_n, L_j are the constants satisfying $L_1 > 0$, $0 < L_2 < ((2+\delta_2)\gamma)^{-1/\beta}$ and $0 < L_3 < ((2+\delta_3)\gamma/\beta)^{-1/\beta}$. Also, $0 < \tilde{L} < (c_1 L_1^\beta)^{-1}$. We note that the speeds of T_n at (14) satisfy (T1′)–(T3′). We choose them so that $\hat{f}_X^*$ and $\hat{m}^*$ can achieve optimal L^2 rates for suitable a_n such as those at (14). The above choice of a_n is an example and other options may be also possible. In practice, we may consider a_n as a tuning parameter. However, in the supersmooth and log-supersmooth scenarios, we can use the above a_n with a small $\delta_j > 0$ without tuning. In the following theorem, we assume that the additional sample size N from f_U satisfies $N \gtrsim n$, that is, $n = O(N)$; see the first paragraph of Section 4.2 for the definition of $\gtrsim$.

THEOREM 12. *Suppose we choose T_n and a_n as specified at* (14). *Assume that $N \gtrsim n$ and that the series in* (4) *converges uniformly to f_X. Then, it holds that $\sup_{x\in\mathbb{G}} |\hat{f}_X^*(x) - f_X(x)| = o_p(1)$. If we further assume the condition* (A1)(i), *then $\int_{\mathbb{G}} |\hat{f}_X^*(x) - f_X(x)|^2 \, d\mu(x)$ achieves the same rates as $\int_{\mathbb{G}} |\hat{m}(x) - m(x)|^2 \, d\mu(x)$ given in Corollary 1. If we further assume the conditions* (A1)(ii) *and* (A2), *then $\int_{\mathbb{G}} |\hat{m}^*(x) - m(x)|^2 \, d\mu(x)$ achieves the same rates as given in Corollary 1.*

We now consider the case where we have repeated measurements

$$(15) \qquad Z_{ij} = U_{ij} \circ X_i, \quad 1 \le j \le R_i, 1 \le i \le n,$$

where $R_i \ge 1$ are the numbers of replicates bounded by a constant $R_{\max}$ and $U_{ij} \sim f_U$ are independent across all i and j. The assumption that $\max_{1\le i\le n} R_i \le R_{\max}$ is natural and also adopted in the Euclidean case since we usually do not have too many replicates due to time, financial or other constraints. In this repeated measurements case, we assume that c_σ in $\phi^U(\sigma^M) = c_\sigma I_{d_\sigma}$ is real and positive. Then it is easy to check that f_U is symmetric, that is, $f_U(u) = f_U(u^{-1})$ for all $u \in \mathbb{G}$. Considering symmetric measurement error distributions is standard in Euclidean measurement error problems with repeated measurements. We note that most of the distributions in Examples A.2–A.4 satisfy the assumption on c_σ. For the estimation of $\phi^U(\sigma^M)$, we note that $Z_{ij} \circ Z_{ik}^{-1} = U_{ij} \circ U_{ik}^{-1}$ holds for $j \ne k$ due to (15) and that $U_{ij} \circ U_{ik}^{-1} \sim f_U * f_U$ holds for $j \ne k$ due to the symmetry of f_U. Hence, $\mathrm{E}(\sigma^M((Z_{ij} \circ Z_{ik}^{-1})^{-1})) = \mathrm{E}(\sigma^M((U_{ij} \circ U_{ik}^{-1})^{-1})) = \phi^U(\sigma^M)^2 =: \phi^{U,2}(\sigma^M)$ for $j \ne k$. Let $R = \sum_{i=1}^n R_i(R_i - 1)/2$. We consider $\hat{\phi}^{U,2}(\sigma^M) = R^{-1} \sum_{i=1}^n \sum_{1\le j<k\le R_i} \sigma^M(Z_{ik} \circ Z_{ij}^{-1})$ as an estimator of $\phi^{U,2}(\sigma^M)$, and take $\hat{\phi}^U(\sigma^M) = \hat{c}_\sigma I_{d_\sigma}$ with $\hat{c}_\sigma = |\mathrm{Re}(d_\sigma^{-1} \mathrm{Tr}(\hat{\phi}^{U,2}(\sigma^M)))|^{1/2}$. We choose $\hat{\phi}^Z(\sigma^M) = (\sum_{i=1}^n R_i)^{-1} \sum_{i=1}^n \sum_{j=1}^{R_i} \sigma^M(Z_{ij}^{-1})$ and $\hat{\phi}^{\nu,U}(\sigma^M) = (\sum_{i=1}^n R_i)^{-1} \sum_{i=1}^n \sum_{j=1}^{R_i} \sigma^M(Z_{ij}^{-1}) Y_i$.

THEOREM 13. *Assume the conditions in Theorem 12 with* $4r + 4\beta + \dim(\mathbb{G})$ *in* T_n *and* a_n *at* (14) *for* (S1) *being replaced by* $4r + 8\beta + \dim(\mathbb{G})$, *the condition* $\delta_j > 0$ *being replaced by* $\delta_j = 2$ *for each* $j \in \{2, 3\}$ *and the condition* $N \gtrsim n$ *being replaced by* $R \gtrsim n$. *Then the conclusions of Theorem 12 for* (S2)(i) *and* (S3)(i) *are valid. For* (S1)(i), *the* L^2 *errors of* $\hat{f}_X^*$ *and* $\hat{m}^*$ *are* $O_p(n^{-4r/(4r+8\beta+\dim(\mathbb{G}))})$.

The L^2 error rate $n^{-4r/(4r+8\beta+\dim(\mathbb{G}))}$ for the ordinary-smooth scenario in the above theorem is suboptimal to the rate $n^{-4r/(4r+4\beta+\dim(\mathbb{G}))}$, which is achieved by the estimators based on a random sample from f_U as asserted in Theorem 12. This is because $\hat{\phi}^U(\sigma^M)$ based on the repeated measurements is subject to a large variability; see the proof of Theorem 13 in the Supplementary Material S.19. The large variability of $\hat{\phi}^U(\sigma^M)$ does not influence on the L^2 rates in the supersmooth and log-supersmooth scenarios. We recall that, in those scenarios, there is a thresholding size of T_n around which the variance of $\hat{m}$ either diverges to infinity or is dominated by the bias. With T_n at the threshold, the variance and bias are not traded-off but apart far away in their sizes (polynomial versus logarithmic); see the discussion in the paragraph immediately above Corollary 1. The size of T_n at (14) differs from the aforementioned threshold only by a constant factor. For such T_n, the large variability of $\hat{\phi}^U$ does not push up the polynomial rate of the variance of $\hat{m}$ to a logarithmic rate. Thus, the bias still dominates the variance with the estimated ϕ^U, which lets $\hat{m}^*$ still achieve the optimal L^2 error rates. For the same reason $\hat{f}_X^*$ also has the optimal rates in the supersmooth and log-supersmooth scenarios. This phenomena was also observed in the Euclidean case; see [17], for example.

The condition $R \gtrsim n$ is very weak. It is satisfied with a few replicates. For example, it holds when $R_i = 2$ for all i or even in the case where some $R_i = 1$ (no replication for some subjects) with some others greater than 1. The choices a_n with $\delta_2 = \delta_3 = 2$ in the super-smooth and log-supersmooth scenarios are made to obtain the optimal L^2 rates. The sub-optimal rate for the ordinary-smooth scenario gets closer to the optimal rate as β is smaller. We note that most of important ordinary-smooth measurement error distributions, such as those in Examples A.2–A.4, have small values of β.

7. Finite sample performance.

7.1. *Simulation study.* In this section, we present the results of two simulation studies. For both simulation studies, we considered the cases of $\mathbb{G} = \mathbb{T}^1$ and $\mathbb{G} = \mathrm{SO}(3)$. We did not cover the case where $\mathbb{G} = \mathrm{SU}(2)$ since the lesson would be similar. The goal of the first simulation study is to compare our regression estimators $\hat{m}$ and $\hat{m}^*$ with some competitors in terms of estimation accuracy. For $\hat{m}^*$, we considered the situation where there exists a random sample from f_U. We did not include the case of repeated measurements in the comparison since in the latter case $\hat{m}^*$ would use more observations of Z so that a fair comparison with $\hat{m}$ and other competitors could not be made. Instead, the case was treated in the real data example to be presented in Section 7.2.

Since there is no other regression estimator designed for the non-Euclidean errors-in-variables problem under study, we chose, as a competitor to $\hat{m}$ and $\hat{m}^*$, the estimator $\hat{m}^0$ defined in the same way as $\hat{m}$ with $K_{T_n}(x, Z_i)$ being replaced by $K_{T_n}^0(x, Z_i)$ at (9). We note that $\hat{m}^0$ is an estimator that one can use when there is no measurement error, that is, $Z_i = X_i$. In the case where $\mathbb{G} = \mathbb{T}^1$, we also considered two Nadaraya–Watson-type estimators, $\hat{m}^{\mathrm{M}}$ of [19] and $\hat{m}^{\mathrm{P}}$ of [65], the local M-estimator $\hat{m}^{\mathrm{H}}$ of [35] constructed by using the kernel weights of [65], and the Euclidean deconvolution regression estimator $\hat{m}^{\mathrm{E}}$ of [24]. The estimators $\hat{m}^{\mathrm{M}}$, $\hat{m}^{\mathrm{P}}$ and $\hat{m}^{\mathrm{H}}$ neglect measurement errors in covariate observation, while the estimator $\hat{m}^{\mathrm{E}}$ ignores the geometric structure of $\mathbb{T}^1$. To construct the estimator $\hat{m}^{\mathrm{E}}$, we treated

θ_U and θ_Z, respectively, as the Euclidean measurement error and contaminated variable, and thus used the Euclidean Fourier transform of θ_U for deconvolution. Here, $\theta_U \in [0, 2\pi)$ and $\theta_Z \in [0, 2\pi)$ are the angles corresponding to $U = \exp(\sqrt{-1} \cdot \theta_U)$ and $Z = \exp(\sqrt{-1} \cdot \theta_Z)$, respectively. In the case where $\mathbb{G} = \mathrm{SO}(3)$, we considered as another competitor of our estimators a multivariate version $\hat{m}^{\mathrm{E}*}$ of [24] treating the nine entries of SO(3) as a 9-dimensional Euclidean variable. For the latter estimator, we estimated the Euclidean Fourier transform of the vectorized version of U by the empirical Euclidean Fourier transform defined by $\hat{\phi}_U^{\mathrm{E}}(t) = N^{-1} \sum_{j=1}^N \exp(\sqrt{-1} \cdot \langle t, \tilde{U}_j^{\mathrm{vec}} \rangle_{\mathbb{R}^9})$ for $t \in \mathbb{R}^9$, where $\{\tilde{U}_j^{\mathrm{vec}} : 1 \le j \le N\}$ is the vectorized version of a random sample $\{\tilde{U}_j : 1 \le j \le N\}$ from f_U. As suggested by [61], we also added the indicator term $I(|\hat{\phi}_U^{\mathrm{E}}(t/h)| \ge N^{-1/2})$ to the integrand of the corresponding Euclidean deconvolution kernel, where h is the bandwidth. The role of the indicator term is similar to that of the indicator term $I(\|\hat{\phi}^U(\sigma^M)\|_{\mathrm{op}} \ge a_n)$ for our estimator $\hat{m}^*$. We selected T_n and other tuning parameters for our methods and their competitors by a five-fold cross-validation.

For the case of $\mathbb{G} = \mathbb{T}^1$, we generated X from the wrapped Gaussian distribution with circular variance $\mathrm{Var}(X) = 5/6$. For the distribution of U in this case, we took the wrapped Laplace distribution for (S1), the wrapped Gaussian distribution for (S2) and the von Mises distribution for (S3). The definitions of these distributions are given in Example A.2. We set the noise-to-signal ratio (NSR), defined as the ratio $\mathrm{Var}(U)/\mathrm{Var}(X)$, to 0.2 and 0.4. For the case of $\mathbb{G} = \mathrm{SO}(3)$, we generated X from the rotational von Mises–Fisher (vMF) distribution with $\lambda = 0.1$ and $A = I_3$. As for the distribution of U, we took the rotational Laplace distribution for (S1), the rotational Gaussian distribution for (S2) and the rotational vMF distribution with $A = I_3$ for (S3). The definitions of these distributions are given in Example A.4. For each distribution of U, we took two values of λ, one producing a relatively low NSR and one producing a relatively high NSR, but we did not calculate the NSR values since there is no closed form of the variance for each distribution on SO(3). Specifically, we took $\lambda = \sqrt{0.2}$ and $\sqrt{0.25}$ for the rotational Laplace and Gaussian f_U, and $\lambda = 2.5$ and 2 for the rotational vMF f_U.

For the case of $\mathbb{G} = \mathbb{T}^1$, we generated Y from the model $Y = \log(\cos\theta_X + 2) + \epsilon$, where $\theta_X \in [0, 2\pi)$ is the angle corresponding to $X = \exp(\sqrt{-1} \cdot \theta_X)$, and ϵ is the Gaussian random variable with mean zero and standard deviation 0.1. We note that this model can also be regarded as a regression model on the unit circle $\mathbb{S}^1$. We generated $R = 500$ Monte Carlo samples $\{(Y_i^{(r)}, U_i^{(r)} \circ X_i^{(r)}) : 1 \le i \le n\}$ of sizes $n = 125, 250$ and 500 for $r = 1, \dots, R$. Next, for the case of $\mathbb{G} = \mathrm{SO}(3)$, we considered $Y = \det(X + 0.1 \cdot I_3) + \epsilon$, where ϵ is again the Gaussian random variable with mean zero and standard deviation 0.1. We generated $R = 500$ Monte Carlo samples of sizes $n = 400$ and 800.

For $\hat{m}^*$ in both cases, $\mathbb{G} = \mathbb{T}^1$ and $\mathbb{G} = \mathrm{SO}(3)$, we generated an additional random sample $\{\tilde{U}_j : 1 \le j \le N\}$ of size $N = n$ from f_U, corresponding to each pseudo sample $\{(Y_i^{(r)}, U_i^{(r)} \circ X_i^{(r)}) : 1 \le i \le n\}$ for $1 \le r \le R$. For simplicity, we took $a_n = n^{-2/5}$ for all the smoothness scenarios, where a_n is the threshold used in the construction of $\hat{f}_X^*$ and $\hat{m}^*$. In the rotational vMF measurement error case, $\phi^U(\sigma^M)$ does not take the form of $c_\sigma I_{d_\sigma}$, which our estimator $\hat{\phi}^U(\sigma^M)$ introduced in Section 6.2 is based on. In this case, we used the empirical mean defined by $\hat{\phi}^U(\sigma^M) = N^{-1} \sum_{j=1}^N \sigma^M(\tilde{U}_j^{-1})$.

We compared the integrated squared bias (ISB), integrated variance (IV) and integrated mean squared error (IMSE) defined by

$$\mathrm{ISB} = \int_{\mathbb{G}} \left(R^{-1} \sum_{r=1}^{R} \check{m}^{(r)}(x) - m(x) \right)^2 d\mu(x),$$

$$\mathrm{IV} = R^{-1} \sum_{r=1}^{R} \int_{\mathbb{G}} \left(R^{-1} \sum_{s=1}^{R} \breve{m}^{(s)}(x) - \breve{m}^{(r)}(x) \right)^2 d\mu(x),$$

$$\mathrm{IMSE} = \mathrm{ISB}+\mathrm{IV} = R^{-1} \sum_{r=1}^{R} \int_{\mathbb{G}} \left(\breve{m}^{(r)}(x) - m(x)\right)^2 d\mu(x),$$

where $\breve{m}^{(r)}(x)$ is an estimator of $m(x)$ obtained from $\{(Y_i^{(r)}, U_i^{(r)} \circ X_i^{(r)}) : 1 \le i \le n\}$, the rth Monte Carlo sample. Tables 1 and 2 give the results of the first simulation. The ISB values for $\hat{m}$ in Tables 1 and 2 are always much smaller than the corresponding values for $\hat{m}^0$. This is well explained by the unbiased scoring property of $\hat{m}$ as demonstrated in Proposition 2. The estimator $\hat{m}^*$ also performs similarly as $\hat{m}$, which demonstrates the validity of using $\hat{\phi}^U$ in the estimation of m. The IMSE values in Tables 1 and 2 show that $\hat{m}$ and $\hat{m}^*$ behave better than their competitors in all scenarios except some cases with smaller sample sizes. In the

TABLE 1

Integrated squared bias (ISB), integrated variance (IV) and integrated mean squared error (IMSE), multiplied by 10^3, of the proposed estimators and competitors for $\mathbb{G} = \mathbb{T}^1$ and scenarios (S1)–(S3), based on $R = 500$ Monte Carlo samples

	NSR	0.2							0.4						
n	Criterion	$\hat{m}$	$\hat{m}^*$	$\hat{m}^0$	$\hat{m}^{\mathrm{E}}$	$\hat{m}^{\mathrm{H}}$	$\hat{m}^{\mathrm{M}}$	$\hat{m}^{\mathrm{P}}$	$\hat{m}$	$\hat{m}^*$	$\hat{m}^0$	$\hat{m}^{\mathrm{E}}$	$\hat{m}^{\mathrm{H}}$	$\hat{m}^{\mathrm{M}}$	$\hat{m}^{\mathrm{P}}$
	f_U								Laplace						
125	ISB	0.12	0.11	5.58	8.11	9.85	10.28	10.79	13.40	16.45	21.36	27.31	28.85	31.12	31.64
	IV	9.39	10.27	4.47	3.03	3.13	3.35	3.05	38.53	38.09	5.07	3.43	4.17	4.21	3.84
	IMSE	9.51	10.38	10.05	11.14	12.98	13.63	13.84	51.93	54.54	26.43	30.74	33.02	35.33	35.48
250	ISB	0.07	0.07	5.91	7.14	8.63	9.11	9.49	0.67	1.31	21.86	24.38	26.79	28.76	29.32
	IV	3.80	4.07	1.96	1.51	1.60	1.78	1.61	15.68	18.41	2.43	1.69	2.03	2.21	2.00
	IMSE	3.87	4.14	7.87	8.65	10.23	10.89	11.10	16.35	19.72	24.29	26.07	28.82	30.97	31.32
500	ISB	0.08	0.09	5.91	6.86	8.03	8.47	8.76	0.07	0.10	22.30	23.59	25.40	27.39	27.83
	IV	2.04	2.16	1.05	0.74	0.81	0.93	0.85	4.81	6.40	1.13	0.79	1.10	1.22	1.13
	IMSE	2.12	2.25	6.96	7.60	8.84	9.40	9.61	4.88	6.50	23.43	24.38	26.50	28.61	28.96
	f_U								Gaussian						
125	ISB	0.27	0.20	5.83	9.07	10.15	10.26	10.66	14.00	16.32	21.30	27.21	29.21	30.70	30.79
	IV	8.28	8.97	4.03	3.18	3.31	3.38	3.13	38.76	39.75	4.77	3.43	4.33	4.22	3.94
	IMSE	8.55	9.17	9.86	12.25	13.46	13.64	13.79	52.76	56.07	26.07	30.64	33.54	34.92	34.73
250	ISB	0.07	0.09	5.81	7.69	8.87	9.07	9.31	0.80	1.98	21.53	23.56	27.25	27.96	28.71
	IV	3.92	4.27	2.11	1.65	1.70	1.72	1.62	17.78	21.38	2.56	2.20	2.25	2.32	2.11
	IMSE	3.99	4.38	7.92	9.34	10.57	10.79	10.93	18.58	23.36	24.09	25.76	29.50	30.28	30.82
500	ISB	0.07	0.07	5.90	6.94	8.29	8.31	8.70	0.14	0.15	21.91	21.39	26.12	26.68	27.46
	IV	2.21	2.31	1.06	0.82	0.91	0.95	0.88	5.95	6.89	1.24	1.12	1.24	1.28	1.19
	IMSE	2.28	2.38	6.96	7.76	9.20	9.26	9.58	6.09	7.04	23.15	22.51	27.36	27.96	28.65
	f_U								von Mises						
125	ISB	0.18	0.13	5.58	8.72	9.84	10.32	10.47	9.48	11.82	21.08	26.24	29.11	30.70	31.27
	IV	9.11	9.72	4.22	3.19	3.29	3.40	3.20	36.44	39.38	5.02	3.92	4.20	4.54	3.91
	IMSE	9.29	9.85	9.80	11.91	13.23	13.72	13.67	45.92	51.20	26.10	30.16	33.31	35.24	35.18
250	ISB	0.07	0.07	5.73	7.34	8.95	8.99	9.42	1.35	1.87	21.90	23.95	27.17	28.41	29.29
	IV	3.99	4.31	2.00	1.61	1.60	1.70	1.58	19.33	21.12	2.38	1.80	2.19	2.17	2.07
	IMSE	4.06	4.38	7.73	8.95	10.55	10.69	11.00	20.68	22.99	24.28	25.75	29.36	30.58	31.36
500	ISB	0.08	0.07	6.00	6.88	8.26	8.37	8.83	0.09	0.10	22.07	22.21	25.75	26.99	27.59
	IV	2.35	2.49	1.01	0.80	0.87	0.97	0.86	5.17	5.53	1.20	0.90	1.17	1.24	1.13
	IMSE	2.43	2.56	7.01	7.68	9.13	9.34	9.69	5.26	5.63	23.27	23.11	26.92	28.23	28.72

TABLE 2
Integrated squared bias (ISB), integrated variance (IV) and integrated mean squared error (IMSE), multiplied by 10^3, of the proposed estimators and competitors for $\mathbb{G} = \mathrm{SO}(3)$ and scenarios (S1)–(S3), based on $R = 500$ Monte Carlo samples

	NSR	Low				High			
n	Criterion	$\hat{m}$	$\hat{m}^*$	$\hat{m}^0$	$\hat{m}^{E*}$	$\hat{m}$	$\hat{m}^*$	$\hat{m}^0$	$\hat{m}^{E*}$
	f_U				Laplace				
400	ISB	0.16	0.15	0.82	12.12	0.22	0.24	1.14	12.04
	IV	1.83	1.88	0.61	0.17	3.18	3.22	0.61	0.91
	IMSE	1.99	2.03	1.43	12.29	3.40	3.46	1.75	12.95
800	ISB	0.14	0.14	0.86	12.10	0.14	0.14	1.19	12.07
	IV	0.69	0.69	0.29	0.20	0.85	0.86	0.30	0.45
	IMSE	0.83	0.83	1.15	12.30	0.99	1.00	1.49	12.52
	f_U				Gaussian				
400	ISB	0.15	0.16	0.34	12.15	0.16	0.16	0.50	12.15
	IV	1.03	1.07	0.59	0.06	1.20	1.26	0.59	0.06
	IMSE	1.18	1.23	0.93	12.21	1.36	1.42	1.09	12.21
800	ISB	0.13	0.13	0.36	12.12	0.13	0.13	0.52	12.13
	IV	0.50	0.49	0.30	0.03	0.56	0.56	0.30	0.03
	IMSE	0.63	0.62	0.66	12.15	0.69	0.69	0.82	12.16
	f_U				von Mises-Fisher				
400	ISB	0.16	0.16	0.47	12.13	0.18	0.16	0.78	12.12
	IV	1.22	1.22	0.60	0.06	1.62	1.66	0.61	0.12
	IMSE	1.38	1.38	1.07	12.19	1.80	1.82	1.39	12.24
800	ISB	0.13	0.13	0.50	12.12	0.15	0.14	0.82	12.13
	IV	0.54	0.56	0.29	0.03	0.68	0.70	0.30	0.03
	IMSE	0.67	0.69	0.79	12.15	0.83	0.84	1.12	12.16

latter cases, the small biases of $\hat{m}$ and $\hat{m}^*$ fail to offset the large variability produced in the estimation of $\phi^Z(\sigma^M)$, which is further amplified by $\phi^U(\sigma^M)^{-1}$ or $\hat{\phi}^U(\sigma^M)^{-1}$. However, $\hat{m}$ and $\hat{m}^*$ overtake the competitors quickly in the IMSE performance as the sample size increases. Note that, unlike the proposed estimators, the performance of the competitors does not get better much for larger sample sizes, due to the intrinsic bias caused by ignoring measurement errors or the geometry of $\mathbb{G}$. As expected, the performance of all the methods becomes worse as the NSR increases. We note that the proposed estimators resist better than the competitors as the NSR increases when the sample size is relatively large.

Now, we move to the discussion of the second simulation study. It is to compare the two types of asymptotic confidence intervals for $m(x)$ that appear in Theorems 7 and 9, namely the confidence interval based on the asymptotic normality (AN) and the one based on the empirical likelihood (EL). We note that there is no other method of constructing confidence intervals that is currently available for this problem. Recall that we discussed the construction of confidence intervals only for the scenario (S1). The distributions of X and U were the same as in the first simulation study. We considered a set $\mathcal{G}$ of dense grid points of $\mathbb{G}$, and for each $x \in \mathcal{G}$ we computed the coverage rate $C_{1-\alpha}(x)$ and the average length $L_{1-\alpha}(x)$ from the corresponding $R = 500$ confidence intervals of level $(1-\alpha) \times 100\%$. We then compared the average values $|\mathcal{G}|^{-1} \sum_{x \in \mathcal{G}} C_{1-\alpha}(x)$ and $|\mathcal{G}|^{-1} \sum_{x \in \mathcal{G}} L_{1-\alpha}(x)$ for $\alpha = 0.9$ and 0.95, where $|\mathcal{G}|$ denotes the cardinality of $\mathcal{G}$.

Tables 3 and 4 contain the results of the second simulation. Table 3 demonstrates that the two methods are comparable in terms of coverage probability although the EL-based

TABLE 3
Average coverage rate $|\mathcal{G}|^{-1}\sum_{x\in\mathcal{G}} C_{1-\alpha}(x)$ (Cov) and average length $|\mathcal{G}|^{-1}\sum_{x\in\mathcal{G}} L_{1-\alpha}(x)$ (Len) of $(1-\alpha)\times 100\%$ confidence intervals of m for $\mathbb{G}=\mathbb{T}^1$ and for the wrapped Laplace density f_U, based on $R=500$ Monte Carlo samples

		NSR = 0.2				NSR = 0.4			
		AN		EL		AN		EL	
$1-\alpha$	n	Cov	Len	Cov	Len	Cov	Len	Cov	Len
0.9	125	0.898	0.248	0.896	0.269	0.674	0.259	0.675	0.285
	250	0.898	0.196	0.900	0.208	0.838	0.226	0.837	0.240
	500	0.894	0.146	0.898	0.150	0.888	0.189	0.888	0.200
0.95	125	0.940	0.295	0.945	0.334	0.705	0.309	0.708	0.357
	250	0.942	0.233	0.946	0.255	0.878	0.269	0.878	0.295
	500	0.942	0.173	0.945	0.182	0.937	0.225	0.938	0.245

method is slightly better, but that the EL-based confidence intervals are wider than the AN-based ones. On the other hand, Table 4 reveals that the AN-based method outperforms the EL-based in terms of both coverage probability and length. These results suggest that the AN-based method is generally a better option than the EL-based, although the latter method is also an attractive alternative. We note that these results are somewhat different from the existing results for Euclidean data such as those in [71] and [79], which demonstrate the superiority of EL-based methods against other competitors in terms of both coverage and length. As the sample size increases, both AN- and EL-based methods tend to produce confidence intervals with more accurate coverage rate and shorter length. As the NSR increases, the confidence intervals get lengthier, which is well expected, and their coverage probabilities become worse especially for smaller sample sizes.

7.2. *Real data analysis*. Particulate matter whose size is less than 2.5 micrometer, called PM2.5, is an air pollutant mainly produced by a combination of other air pollution sources. Due to its tiny size, it can penetrate lungs and blood vessels, which causes various health problems. In Korea, the city "Seosan" is known to have the worst air quality among all Korean cities and it is reported that there are many patients suffering from respiratory diseases. In fact, its subregion "Daesan," located in the north and northwest parts of the city made the largest increase of air pollutants in the world as noted in [21]. It has been believed that a number of large petrochemical and mechanical plants that have been built in the subregion are the main cause of air pollution. Nevertheless, there have not been enough investigations

TABLE 4
Average coverage rate $|\mathcal{G}|^{-1}\sum_{x\in\mathcal{G}} C_{1-\alpha}(x)$ (Cov) and average length $|\mathcal{G}|^{-1}\sum_{x\in\mathcal{G}} L_{1-\alpha}(x)$ (Len) of $(1-\alpha)\times 100\%$ confidence intervals of m for $\mathbb{G}=\mathrm{SO}(3)$ and for the rotational Laplace density f_U, based on $R=500$ Monte Carlo samples

		Low NSR				High NSR			
		AN		EL		AN		EL	
$1-\alpha$	n	Cov	Len	Cov	Len	Cov	Len	Cov	Len
0.9	400	0.888	0.123	0.866	0.135	0.776	0.119	0.755	0.133
	800	0.882	0.087	0.864	0.090	0.885	0.095	0.863	0.099
0.95	400	0.931	0.147	0.918	0.169	0.812	0.141	0.799	0.169
	800	0.940	0.104	0.924	0.110	0.942	0.113	0.924	0.121

on this belief, in particular, on the impact of the local plants on the bad air quality in the city center. As an investigation on this issue, we studied how the PM2.5 level in the city center is related to the wind direction oriented from Daesan.

For this, we used a data set for these variables collected during the year 2020, available from the Korea air pollution information system https://www.airkorea.or.kr/web/last_amb_hour_data?pMENU_NO=123 and the Korea meteorological administration https://data.kma.go.kr/data/grnd/selectAsosRltmList.do?pgmNo=36&tabNo=1. In the databases, wind direction is recorded in minute-level while PM2.5 level is hourly. We took the hourly measured PM2.5 level as the response Y. Considering the time for air pollutants produced by the plants to arrive at the city center, we took the wind direction one hour prior to the measuring time of Y as the predictor X. As noted in [28] and [27] among many others, it is known that wind direction is hard to measure precisely due to the inaccuracy of measuring devices. Hence, it is natural to assume that recorded wind direction, say Z_1, contains a measurement error added to X. To calibrate such measurement errors, we took an additional observation of Z_2 that is the wind direction measured one minute after the measurement of Z_1. After deleting some missing observations, we obtained a data set $\{(Z_{i1}, Z_{i2}, Y_i) : 1 \le i \le n\}$ with $n = 6215$, where Z_{i1} and Z_{i2} are considered as the repeated measurements of X_i with errors. Among them, 5938 observations contain both Z_{i1} and Z_{i2} and the remaining observations contain either Z_{i1} or Z_{i2}. We then applied the proposed method for repeated measurements presented in Section 6.2. For comparison, we also applied the naive approach based on the assumption that there is no measurement error. The latter estimator $\hat{m}^{0*}$ takes the form of $\hat{m}^*$ with all $\hat{\phi}^U(\sigma^M) \in \mathbb{C}$ being replaced by 1. We note that $\phi^U(\sigma^M) \equiv 1$ for the circular case in the absence of measurement error. We also applied the methods of [19, 65] and [35] by taking the circular mean of Z_{i1} and Z_{i2} as the predictor corresponding to Y_i for each i. The latter three methods also ignore measurement errors. We selected tuning parameters by a five-fold cross-validation. The selected values of tuning parameters were $T_n = 4$ for $\hat{m}^*$, $T_n = 16$ for $\hat{m}^{0*}$, 0.27 for the estimator of [19] and 1.5 for the estimators of [65] and of [35].

Figure 1 shows the estimated regression functions for the five approaches. The top panel clearly shows higher PM2.5 level for wind blown from the directions of the petrochemical and mechanical plants. On the contrary, the middle and bottom panels do not reveal this pattern but depict almost uniform functions on the unit circle. Hence, the proposed method finds an evidence that the plants pollute the air of the city center, which might not be detected by methods ignoring measurement errors.

8. Further discussion. One may be interested in the case where the predictor takes values in a general k-dimensional manifold, say $\mathcal{M}$. Suppose that we cannot observe $X(\omega)$ of a covariate $X : \Omega \to \mathcal{M}$, but observe only its contaminated value $Z(\omega) \in \mathcal{M}$. If there is no group structure and no other alternative on $\mathcal{M}$, then we cannot define a measurement error variable on $\mathcal{M}$. A possible approach to this is to assume that $\phi_Z(Z) = \phi_X(X) + U^*$ for some U^* taking values in $\mathbb{R}^k$, where $\phi_p : V_p \to \phi_p(V_p) \subset \mathbb{R}^k$ is a homeomorphism on an open neighborhood V_p of $p \in \mathcal{M}$. Suppose that ϕ_p are known for all p although it is unrealistic for many manifolds. Then, by applying Euclidean deconvolution techniques to the observations $\{\phi_{Z_i(\omega)}(Z_i(\omega)) : 1 \le i \le n\} \subset \mathbb{R}^k$, we may be able to estimate the density $f_{\phi_X(X)}$ of $\phi_X(X)$. However, since the map ϕ_p depends on $p \in \mathcal{M}$, it is difficult to construct an estimator of f_X from an estimator of $f_{\phi_X(X)}$. In regression function estimation, in addition to overcoming the hurdle, we need to establish an unbiased scoring property, which is a key to the success of a deconvolution technique. The latter problem is also very hard to solve for manifolds without group or other relevant structure. The Lie group structure and the associated harmonic analysis on which our approach relies facilitate all these components of the work.

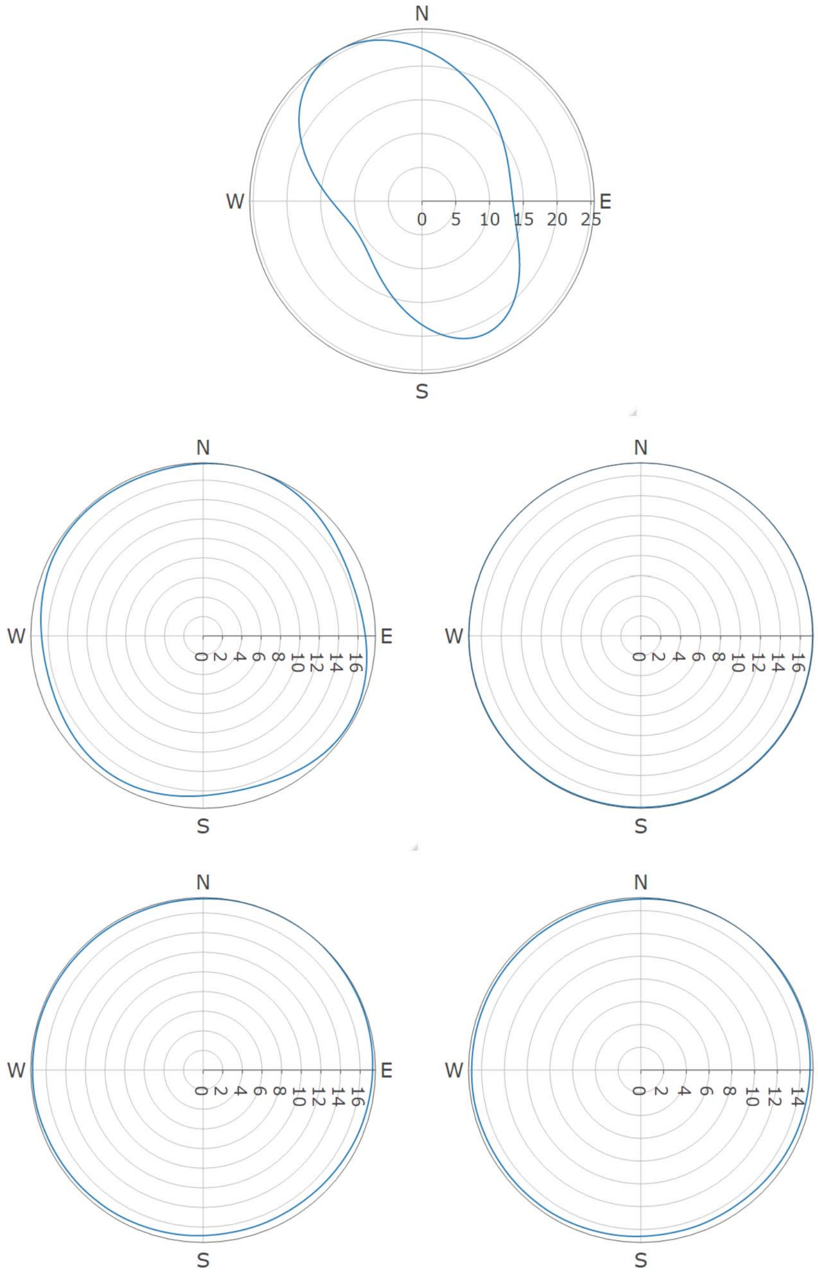

FIG. 1. *Estimated regression functions for the proposed* $\hat{m}^*$ (*top*), *for* $\hat{m}^{0*}$ (*middle left*) *and for the methods of* [19] (*middle right*), *of* [65] (*bottom left*) *and of* [35] (*bottom right*). *In the panels, the directional position of a point on the solid curve represents the value of the circular predictor corresponding to the value of the regression function depicted as the distance from the point to the origin.*

One may be also interested in the case where $\mathbb{G}$ is not compact. We summarize, in the Supplementary Material S.21, some basic notions and results in harmonic analysis on general Lie groups. Dealing with general Lie groups have many obstacles. For examples, some irreducible representations on noncompact Lie groups can be infinite-dimensional and the Fourier transforms at such irreducible representations become bounded linear operators on infinite-dimensional Hilbert spaces. Treating infinite-dimensional irreducible representations and operator-valued Fourier transforms are challenging both in theory and practice. To mention first a couple of theoretical difficulties among many others, we may not have a property such as $\sum_{\sigma\in\hat{\mathbb{G}}:k_\sigma<T_n} d_\sigma^2 \asymp T_n^{\dim(\mathbb{G})/2}$ that appears in the discussion above (6), neither can we assume the condition (B4) since $d_\sigma = \infty$ for some $\sigma \in \hat{\mathbb{G}}$. In practical aspects, implementing infinite-dimensional irreducible representations and operator-valued Fourier transforms are very difficult, and there are possibly uncountably many irreducible representations, which makes the implementation even harder. Another fundamental obstacle to treating general Lie groups is that the L^2 Fourier inversion formula for the general case is very abstract, unlike the one for the compact case given at (4), and its specific form is not discovered for many noncompact Lie groups in mathematics. Furthermore, its pointwise, uniform-type and absolute-type versions are not well studied in mathematics, to the best of our knowledge.

Another interesting case is that $X : \Omega \to \mathcal{M}$, where $\mathcal{M}$ is a Riemannian symmetric space. A Riemmanian symmetric space can be viewed as a quotient space $\mathcal{G}/\mathcal{H}$ for some Lie group $\mathcal{G}$ and a subgroup $\mathcal{H}$ of $\mathcal{G}$. For the definition of Riemmanian symmetric space and also for regression analysis with $\mathcal{M}$-valued responses, we refer to [12]. For such an $\mathcal{M}$, there exists a smooth map $\odot : \mathcal{G} \times \mathcal{M} \to \mathcal{M}$ such that, for any $x, z \in \mathcal{M}$, there exists an element $u \in \mathcal{G}$ satisfying $u \odot x = z$. Hence, for a true value $X(\omega) \in \mathcal{M}$ and its contaminated observation $Z(\omega) \in \mathcal{M}$, we may consider a measurement error structure $Z = U \odot X$ for U taking values in $\mathcal{G}$. However, studying on the deconvolution theory for a general class of Riemannian symmetric spaces also has the same obstacles as the case for general Lie groups we described above. We refer to [74] and [73] for harmonic analysis on Riemannian symmetric spaces. Among Riemannian symmetric spaces that are not a Lie group, the unit sphere $\mathbb{S}^2$ is a statistically most important example. In the case of $\mathbb{S}^2$, for any $x, z \in \mathbb{S}^2$, there exists $u \in \mathrm{SO}(3)$ such that $u \odot x = z$, where $u \odot x$ is the usual matrix multiplication between the matrix u and the vector x. In fact, the case of $\mathbb{S}^2$ is well studied in deconvolution density estimation (e.g., [33, 46, 48]) based on well-studied spherical harmonic analysis in mathematics. We believe that it is possible to extend it to regression analysis and work on asymptotic confidence intervals as well as asymptotic distributions, for which our novel idea and techniques we have developed in this paper would be very useful.

APPENDIX: SOME DETAILS FOR PRACTICAL IMPLEMENTATION

We present full practical details on the implementation of $K_{T_n}(x, Z_i)$ for certain Lie groups. They are largely unknown in statistics.

EXAMPLE A.1. (i) *Toruses*. Suppose that $\mathbb{G} = \mathbb{T}^D$. Recall from Example 1(ii) that $\dim(\mathbb{G}) = D$, $\hat{\mathbb{G}} = \{\sigma_l : l = (l_1, \ldots, l_D) \in \mathbb{Z}^D\}$ and $d_{\sigma_l} \equiv 1$. In this case, $k_{\sigma_l} = \sum_{d=1}^D l_d^2$ and the matrix form $\sigma_l^M(g)$ of $\sigma_l(g)$ is given by $\sigma_l^M(g) = \prod_{d=1}^D (g_d)^{l_d} \in \mathbb{T}^1$ regardless of the choice of an orthonormal basis of $\mathbb{C}$; see [25]. We note that each $g_d \in \mathbb{T}^1$ can be written as $\exp(\sqrt{-1} \cdot \theta_d)$ for some $\theta_d \in [0, 2\pi)$. It holds that

$$\begin{aligned}\phi^U(\sigma_l^M) &= \int_{\mathbb{T}^D} f_U(g) \cdot \prod_{d=1}^D (g_d^{-1})^{l_d}\, d\mu(g)\\ &= \frac{1}{(2\pi)^D} \int_{[0,2\pi]^D} f_U\big(\exp(\sqrt{-1}\cdot\theta_1), \ldots, \exp(\sqrt{-1}\cdot\theta_D)\big)\\ &\quad \cdot \exp\left(-\sqrt{-1} \cdot \sum_{d=1}^D l_d\theta_d\right) d\theta_1 \cdots d\theta_D,\end{aligned}$$

where the latter integral is the usual Lebesgue integral; see [25].

(ii) *Special unitary group of degree 2*. Suppose that $\mathbb{G} = \mathrm{SU}(2)$. Recall from Example 1(iii) that $\dim(\mathbb{G}) = 3$, $\hat{\mathbb{G}} = \{\sigma_l : l \in \{0\} \cup \mathbb{N}\}$ and $d_{\sigma_l} = l + 1$. In this case, $k_{\sigma_l} = l(l + 2)$ and a matrix form $\sigma_l^M(g)$ of $\sigma_l(g)$ is given by a $(l+1)\times(l+1)$ complex matrix whose (i,j)th element equals

$$\begin{aligned}\left(\frac{(i-1)!(l+1-i)!}{(j-1)!(l+1-j)!}\right)^{1/2} &\int_0^1 \bigl(\overline{g_{11}}\cdot\exp(2\pi t\cdot\sqrt{-1})-g_{12}\bigr)^{j-1}\\ &\cdot\bigl(\overline{g_{12}}\cdot\exp(2\pi t\cdot\sqrt{-1})+g_{11}\bigr)^{l+1-j}\cdot\exp\bigl(-2\pi t(i-1)\sqrt{-1}\bigr)\,dt;\end{aligned}\tag{16}$$

see [25] and [26]. We also note that each $g\in \mathrm{SU}(2)$ can be written as

$$\begin{aligned}&R(\varphi,\theta,\psi)\\ &=\begin{pmatrix}\cos\varphi+\sqrt{-1}\cdot\sin\varphi\cos\theta & \sin\varphi\sin\theta\cos\psi+\sqrt{-1}\cdot\sin\varphi\sin\theta\sin\psi\\ -\sin\varphi\sin\theta\cos\psi+\sqrt{-1}\cdot\sin\varphi\sin\theta\sin\psi & \cos\varphi-\sqrt{-1}\cdot\sin\varphi\cos\theta\end{pmatrix}\end{aligned}$$

for some $\varphi,\theta\in[0,\pi)$ and $\psi\in[0,2\pi)$. It holds that

$$\begin{aligned}\phi^U(\sigma_l^M)&=\int_{\mathrm{SU}(2)}\sigma_l^M(g^{-1})f_U(g)\,d\mu(g)\\ &=\frac{1}{2\pi^2}\int_0^{2\pi}\int_0^{\pi}\int_0^{\pi}\sigma_l^M\bigl(R(-\varphi,\theta,\psi)\bigr)f_U\bigl(R(\varphi,\theta,\psi)\bigr)\sin^2(\varphi)\sin(\theta)\,d\varphi\,d\theta\,d\psi;\end{aligned}$$

see [25]. The integration on the right-hand side of the second equation is the usual matrix-valued Lebesgue integral, so that $\phi^U(\sigma_l^M)$ can be readily computed from the specific forms of σ_l^M and $R(\varphi,\theta,\psi)$.

(iii) *Rotation group*. Suppose that $\mathbb{G}=\mathrm{SO}(3)$. Recall from Example 1(iv) that $\dim(\mathbb{G})=3$, $\hat{\mathbb{G}}=\{\sigma_l : l\in\{0\}\cup\mathbb{N}\}$ and $d_{\sigma_l}=2l+1$. In this case, $k_{\sigma_l}=l(l+1)$; see [75]. To give an explicit form of σ_l^M, we note that each element $g\in\mathrm{SO}(3)$ can be written as $g=R(\varphi)S(\theta)R(\psi)$ for some Euler angles $\varphi,\psi\in[0,2\pi)$ and $\theta\in[0,\pi)$, where

$$R(\vartheta)=\begin{pmatrix}\cos\vartheta & -\sin\vartheta & 0\\ \sin\vartheta & \cos\vartheta & 0\\ 0 & 0 & 1\end{pmatrix},\qquad S(\vartheta)=\begin{pmatrix}\cos\vartheta & 0 & \sin\vartheta\\ 0 & 1 & 0\\ -\sin\vartheta & 0 & \cos\vartheta\end{pmatrix}.$$

Let $c_{ij}^l=((2l+1-i)!(i-1)!(2l+1-j)!(j-1)!)^{1/2}$ and $d_{ij}^l(\theta)$ be the (i,j)th element of the Wigner's d-matrix such that

$$d_{ij}^l(\theta)=c_{ij}^l\cdot\sum_{k=\max\{0,j-i\}}^{\min\{2l+1-i,j-1\}}(-1)^{k+i-j}\frac{(\cos(\theta/2))^{2l-2k+j-i}(\sin(\theta/2))^{2k+i-j}}{(2l+1-i-k)!(j-1-k)!(k+i-j)!k!}.$$

Then a matrix form $\sigma_l^M(R(\varphi)S(\theta)R(\psi))$ of $\sigma_l(R(\varphi)S(\theta)R(\psi))$ is given by a $(2l+1)\times(2l+1)$ complex matrix whose (i,j)th element equals

$$\exp\bigl(-\sqrt{-1}(i-l-1)\varphi\bigr)\cdot d_{ij}^l(\theta)\cdot\exp\bigl(-\sqrt{-1}(j-l-1)\psi\bigr);$$

see [75] and [67]. It holds that

$$\begin{aligned}\phi^U(\sigma_l^M)&=\int_{\mathrm{SO}(3)}\sigma_l^M(g^{-1})f_U(g)\,d\mu(g)\\ &=\frac{1}{8\pi^2}\int_0^{2\pi}\int_0^{\pi}\int_0^{2\pi}\sigma_l^M\bigl(R(-\psi)S(-\theta)R(-\varphi)\bigr)f_U\bigl(R(\varphi)S(\theta)R(\psi)\bigr)\\ &\quad\cdot\sin(\theta)\,d\varphi\,d\theta\,d\psi,\end{aligned}$$

where again the integration on the right-hand side of the second equation is the usual matrix-valued Lebesgue integral; see [75].

(iv) *Product of compact and connected Lie groups*. Suppose that $\mathbb{G} = \mathbb{G}_1 \times \mathbb{G}_2$ for compact and connected Lie groups $\mathbb{G}_1$ and $\mathbb{G}_2$ equipped with respective group operations $\circ_1$ and $\circ_2$. Recall from Example 1(v) that $\dim(\mathbb{G}) = \dim(\mathbb{G}_1) + \dim(\mathbb{G}_2)$, $\hat{\mathbb{G}} = \{\sigma_1 \otimes \sigma_2 : \sigma_1 \in \hat{\mathbb{G}}_1, \sigma_2 \in \hat{\mathbb{G}}_2\}$ and $d_{\sigma_1 \otimes \sigma_2} = d_{\sigma_1} \cdot d_{\sigma_2}$. In this case, $k_{\sigma_1 \otimes \sigma_2} = k_{\sigma_1} + k_{\sigma_2}$; see [3]. For orthonormal bases $\{e_k^{\sigma_1} : 1 \le k \le d_{\sigma_1}\}$ of $\mathbb{H}_{\sigma_1}$ and $\{e_l^{\sigma_2} : 1 \le l \le d_{\sigma_2}\}$ of $\mathbb{H}_{\sigma_2}$, $\{e_i^{\sigma_1 \otimes \sigma_2} : 1 \le i \le d_{\sigma_1 \otimes \sigma_2}\} := \{e_k^{\sigma_1} \otimes e_l^{\sigma_2} : 1 \le k \le d_{\sigma_1}, 1 \le l \le d_{\sigma_2}\}$ forms an orthonormal basis of $\mathbb{H}_{\sigma_1} \otimes \mathbb{H}_{\sigma_2}$. For $e_i^{\sigma_1 \otimes \sigma_2} = e_{k_i}^{\sigma_1} \otimes e_{l_i}^{\sigma_2}$ and $e_j^{\sigma_1 \otimes \sigma_2} = e_{k_j}^{\sigma_1} \otimes e_{l_j}^{\sigma_2}$ with $1 \le k_i, k_j \le d_{\sigma_1}$ and $1 \le l_i, l_j \le d_{\sigma_2}$, it holds that

$$\begin{aligned}
(\sigma_1 \otimes \sigma_2)_{ij}^M(g_1, g_2) &= \langle \sigma_1 \otimes \sigma_2(g_1, g_2)(e_j^{\sigma_1 \otimes \sigma_2}), e_i^{\sigma_1 \otimes \sigma_2} \rangle_{\mathbb{H}_{\sigma_1} \otimes \mathbb{H}_{\sigma_2}} \\
&= \langle \sigma_1(g_1)(e_{k_j}^{\sigma_1}) \otimes \sigma_2(g_2)(e_{l_j}^{\sigma_2}), e_{k_i}^{\sigma_1} \otimes e_{l_i}^{\sigma_2} \rangle_{\mathbb{H}_{\sigma_1} \otimes \mathbb{H}_{\sigma_2}} \\
&= \langle \sigma_1(g_1)(e_{k_j}^{\sigma_1}), e_{k_i}^{\sigma_1} \rangle_{\mathbb{H}_{\sigma_1}} \langle \sigma_2(g_2)(e_{l_j}^{\sigma_2}), e_{l_i}^{\sigma_2} \rangle_{\mathbb{H}_{\sigma_2}} \\
&= (\sigma_1)_{k_i k_j}^M(g_1)(\sigma_2)_{l_i l_j}^M(g_2).
\end{aligned}$$

Also,

$$\phi^U\big((\sigma_1 \otimes \sigma_2)^M\big) = \int_{\mathbb{G}_1 \times \mathbb{G}_2} (\sigma_1 \otimes \sigma_2)^M\big(g_1^{-1}, g_2^{-1}\big) f_U(g_1, g_2)\, d\mu_1 \otimes \mu_2(g_1, g_2),$$

where μ_1 and μ_2 are the respective normalized Haar measures on $\mathbb{G}_1$ and $\mathbb{G}_2$, and $\mu_1 \otimes \mu_2$ is the product measure of μ_1 and μ_2. We note that $\mu_1 \otimes \mu_2$ is the normalized Haar measure on $\mathbb{G}_1 \times \mathbb{G}_2$; see [77] for more details.

REMARK A.1. If implementation details such as those in Example A.1 are available for some Lie group $\mathbb{G}$ equipped with a group operation $\circ$, then one may use them for another Lie group $\mathcal{M}$ equipped with a group operation $\circledast$ for which there exists a group isomorphism $\Phi : \mathcal{M} \to \mathbb{G}$, which is also a homeomorphism. Such $\mathcal{M}$ and $\mathbb{G}$ are said to be algebraically and topologically isomorphic or simply *isomorphic as Lie group*. We note that they are also diffeomorphic since every homeomorphism between two Lie groups is a diffeomorphism. In that case, suppose that we are interested in estimating $\mathfrak{m} : \mathcal{M} \to \mathbb{R}$ in the model $Y_i = \mathfrak{m}(\tilde{X}_i) + \epsilon_i$ based on $\{(\tilde{Z}_i, Y_i), 1 \le i \le n\}$, where $\tilde{Z}_i = \tilde{U}_i \circledast \tilde{X}_i$ for some measurement errors $\tilde{U}_i$ that are independent of $(\tilde{X}_i, Y_i)$. Define $m : \mathbb{G} \to \mathbb{R}$ by $m = \mathfrak{m}(\Phi^{-1}(\cdot))$. Let $X_i = \Phi(\tilde{X}_i)$, $U_i = \Phi(\tilde{U}_i)$ and $Z_i = \Phi(\tilde{Z}_i)$. Then we have $Y_i = m(X_i) + \epsilon_i$ and $Z_i = \Phi(\tilde{Z}_i) = \Phi(\tilde{U}_i \circledast \tilde{X}_i) = \Phi(\tilde{U}_i) \circ \Phi(\tilde{X}_i) = U_i \circ X_i$. We note that $U_i \perp (X_i, Y)$ since $\tilde{U}_i \perp (\tilde{X}_i, Y)$. If we obtain an estimator $\hat{m}$ of m based on $\{(Z_i, Y_i) : 1 \le i \le n\}$, then we may estimate $\mathfrak{m}$ by $\hat{\mathfrak{m}} = \hat{m}(\Phi(\cdot))$. The same method applies to density estimation.

As an example, consider $\mathcal{M} = \prod_{d=1}^D \mathbb{S}^1$ for $D \ge 1$, where $\mathbb{S}^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ is the unit circle. In this case, a point on $\mathcal{M}$ is identified to a point on $\mathbb{G} = \mathbb{T}^D$ by the homeomorphism

$$\Phi : \big((x_1, y_1), \dots, (x_D, y_D)\big) \mapsto \big((x_1 + \sqrt{-1} \cdot y_1), \dots, (x_D + \sqrt{-1} \cdot y_D)\big).$$

Then $\mathcal{M}$ equipped with $\circledast$, defined by $p \circledast q = \Phi^{-1}(\Phi(p) \circ \Phi(q))$ for $p, q \in \mathcal{M}$, is isomorphic to $\mathbb{G}$ as Lie groups. As another example, consider the 3-dimensional unit hypersphere $\mathcal{M} = \mathbb{S}^3 := \{(x, y, z, w) \in \mathbb{R}^4 : x^2 + y^2 + z^2 + w^2 = 1\}$. In this case, $\mathcal{M}$ is homeomorphic to $\mathbb{G} = \mathrm{SU}(2)$ by the map

$$\Phi : (x, y, z, w) \mapsto \begin{pmatrix} x + \sqrt{-1} \cdot y & z + \sqrt{-1} \cdot w \\ -z + \sqrt{-1} \cdot w & x - \sqrt{-1} \cdot y. \end{pmatrix}.$$

Then, for $\circledast$ defined in the same way as in the first example, $\mathcal{M}$ and $\mathbb{G}$ are isomorphic as Lie groups. These group operations $\circledast$ on $\prod_{d=1}^D \mathbb{S}^1$ and $\mathbb{S}^3$ are commonly used for the respective Lie groups.

We note that all compact and connected Abelian Lie groups of the same dimension $D > 1$ are isomorphic as Lie group to each other, indeed isomorphic to $\mathbb{T}^D$ (e.g., Chapter 4.4.2, [66]). Hence, compact and connected Abelian Lie groups are essentially unique. We also note that, in case $\mathcal{M}$ and $\mathbb{G}$ are compact and connected semisimple Lie groups, then a group isomorphism between them is automatically a homeomorphism ([51]). The definition of semisimple Lie group can be found in the Supplementary Material S.1. We note that SU(2) and SO(3) are examples of compact and connected semisimple Lie group. Some examples of Lie groups isomorphic to SU(2) other than $\mathbb{S}^3$ and those isomorphic to SO(3) as Lie group can be found in Chapter 1.2 of [69], for example.

We now present some special distributions for the Lie groups in Example A.1 that can be commonly adopted as a measurement error distribution in practice.

EXAMPLE A.2. Consider the case $\mathbb{G} = \mathbb{T}^1$. The following examples of distributions on $\mathbb{T}^1$ have not been considered in measurement error problems despite their practical importance. Below, in the expressions of the densities f_U at $u \in \mathbb{T}^1$, we use the representation $u = \exp(\sqrt{-1} \cdot \theta_u)$, where $\theta_u \in [0, 2\pi)$ is the angle corresponding to u. In all examples, $\lambda > 0$ in the densities $f_U(\cdot, \lambda)$ are real parameters and σ_l^M is the matrix form given in Example A.1(i).

(i) *Wrapped Laplace distribution*. The density of a wrapped Laplace distribution on $\mathbb{T}^1$ is given by

$$f_U(u, \lambda) = \frac{\pi}{\lambda} \cdot \left[\frac{\exp(-\theta_u/\lambda)}{1 - \exp(-2\pi/\lambda)} + \frac{\exp(\theta_u/\lambda)}{\exp(2\pi/\lambda) - 1}\right].$$

It holds that $\phi^U(\sigma_l^M) = (1 + \lambda^2 l^2)^{-1}$ for $l \in \mathbb{Z}$, and this Fourier transform satisfies (S1) with $\beta = 1$.

(ii) *Wrapped exponential distribution*. The density of a wrapped exponential distribution on $\mathbb{T}^1$ is given by

$$f_U(u, \lambda) = \frac{2\pi \exp(-\theta_u/\lambda)}{\lambda(1 - \exp(-2\pi/\lambda))}.$$

It holds that $\phi^U(\sigma_l^M) = (1 + \sqrt{-1} \cdot \lambda \cdot l)^{-1}$ for $l \in \mathbb{Z}$, and this Fourier transform satisfies (S1) with $\beta = 1/2$.

(iii) *Wrapped Lindley distribution*. The density of a wrapped Lindley distribution on $\mathbb{T}^1$ is given by

$$f_U(u, \lambda) = 2\pi\lambda^2(1 + \lambda)^{-1} \exp(-\lambda\theta_u)\left[\frac{1 + \theta_u}{1 - \exp(-2\pi\lambda)} + \frac{2\pi \exp(-2\pi\lambda)}{(1 - \exp(-2\pi\lambda))^2}\right].$$

It holds that $\phi^U(\sigma_l^M) = \lambda^2(1 + \lambda)^{-1}(\lambda + \sqrt{-1} \cdot l)^{-2}(1 + \lambda + \sqrt{-1} \cdot l)$ for $l \in \mathbb{Z}$, and this Fourier transform satisfies (S1) with $\beta = 1/2$.

(iv) *Wrapped Gaussian distribution*. The density of a wrapped Gaussian distribution on $\mathbb{T}^1$ is given by

$$f_U(u, \lambda) = \frac{\sqrt{2\pi}}{\lambda} \cdot \sum_{q \in \mathbb{Z}} \exp\bigl(-(\theta_u + 2\pi q)^2/(2\lambda^2)\bigr).$$

It holds that $\phi^U(\sigma_l^M) = \exp(-\lambda^2 l^2/2)$ for $l \in \mathbb{Z}$, and this Fourier transform satisfies (S2) with $\beta = 1$, $\alpha = 0$ and $\gamma = \lambda^2/2$.

(v) *Wrapped Cauchy distribution*. The density of a wrapped Cauchy distribution on $\mathbb{T}^1$ is given by

$$f_U(u,\lambda) = \frac{\sinh(\lambda)}{\cosh(\lambda) - \cos(\theta_u)},$$

where sinh and cosh denote the hyperbolic sine and cosine functions, respectively. It holds that $\phi^U(\sigma_l^M) = \exp(-\lambda|l|)$ for $l \in \mathbb{Z}$, and this Fourier transform satisfies (S2) with $\beta = 1/2$, $\alpha = 0$ and $\gamma = \lambda$.

(vi) *Wrapped Lévy distribution*. The density of a wrapped Lévy distribution on $\mathbb{T}^1$ is given by

$$f_U(u,\lambda) = \sqrt{2\pi\lambda} \cdot \sum_{q \in \mathbb{Z}} \frac{\exp(-\lambda(\theta_u + 2\pi q)/2)}{(\theta_u + 2\pi q)^{3/2}}.$$

It holds that $\phi^U(\sigma_l^M) = \exp(-\sqrt{\lambda|l|} \cdot (1 + \sqrt{-1} \cdot \mathrm{sgn}(l)))$ for $l \in \mathbb{Z}$, and this Fourier transform satisfies (S2) with $\beta = 1/4$, $\alpha = 0$ and $\gamma = \sqrt{\lambda}$.

(vii) *von Mises distribution*. The density of a von Mises distribution on $\mathbb{T}^1$ is given by

$$f_U(u,\lambda) = \exp(\lambda \cos\theta_u)/B_0(\lambda),$$

where B_k is the modified Bessel function of the first kind of order k. It holds that $\phi^U(\sigma_l^M) = B_{|l|}(\lambda)/B_0(\lambda)$ for $l \in \mathbb{Z}$. Using the result of Section 5.3 in [50], we may prove that it satisfies (S3) with $\beta = 1/2$, $\alpha = 0$, $\gamma = 1/2$, $\xi_1 = 2(1 + \log\lambda)$ and $\xi_2 = 2(1 + \log(2\lambda))$.

EXAMPLE A.3. Next, consider the case $\mathbb{G} = \mathrm{SU}(2)$. Again, $\lambda > 0$ in the densities $f_U(\cdot,\lambda)$ below are real parameters and σ_l^M is the matrix form given in Example A.1(ii).

(i) *Hyperspherical Laplace distribution*. The density of a hyperspherical Laplace distribution is given by

$$f_U(u,\lambda) = \sum_{l=0}^{\infty} (l+1)\big(1 + \lambda^2 \cdot l(l+2)\big)^{-1} \operatorname{Tr}\big(\sigma_l^M(u)\big), \quad u \in \mathrm{SU}(2).$$

For this distribution, $\phi^U(\sigma_l^M) = (1 + \lambda^2 \cdot l(l+2))^{-1} I_{l+1}$ so that it belongs to the ordinary-smoothness scenario (S1) with $\beta = 1$.

(ii) *Hyperspherical Gaussian distribution*. The density of a hyperspherical Gaussian distribution is given by

$$f_U(u,\lambda) = \sum_{l=0}^{\infty} (l+1) \exp\big(-\lambda^2 \cdot l(l+2)/2\big) \operatorname{Tr}\big(\sigma_l^M(u)\big), \quad u \in \mathrm{SU}(2).$$

For this distribution, $\phi^U(\sigma_l^M) = \exp(-\lambda^2 \cdot l(l+2)/2) I_{l+1}$ so that it belongs to the super-smoothness scenario (S2) with $\beta = 1$, $\alpha = 0$ and $\gamma = \lambda^2/2$.

EXAMPLE A.4. Now, consider the case $\mathbb{G} = \mathrm{SO}(3)$. Below, $\lambda > 0$ in the densities $f_U(\cdot,\lambda)$ are real parameters and σ_l^M is the matrix form given in Example A.1(iii).

(i) *Rotational Laplace distribution*. Let $r_u = \arccos((\operatorname{Tr}(u) - 1)/2) \in [0,\pi]$ for $u \in \mathrm{SO}(3)$ and $a_\lambda = \sqrt{1/4 - \lambda^{-2}} \in \mathbb{C}$. The density of a rotational Laplace distribution is given by

$$f_U(u,\lambda) = \lambda^{-2}\pi \cdot \frac{\cos(a_\lambda(\pi - r_u))}{\cos(a_\lambda \pi)\sin(r_u/2)} \cdot I(r_u > 0), \quad u \in \mathrm{SO}(3).$$

In this case, $\phi^U(\sigma_l^M) = (1 + \lambda^2 \cdot l(l+1))^{-1} I_{2l+1}$ (Theorem 3.5 in [33]). Thus, it satisfies (S1) with $\beta = 1$.

(ii) *Rotational Gaussian distribution*. The density of a rotational Gaussian distribution is given by

$$f_U(u, \lambda) = \sum_{l=0}^{\infty} (2l+1) \exp\bigl(-\lambda^2 \cdot l(l+1)/2\bigr) \operatorname{Tr}\bigl(\sigma_l^M(u)\bigr), \quad u \in \mathrm{SO}(3).$$

For this distribution introduced in [48], $\phi^U(\sigma_l^M) = \exp(-\lambda^2 \cdot l(l+1)/2) I_{2l+1}$ so that it satisfies (S2) with $\beta = 1$, $\alpha = 0$ and $\gamma = \lambda^2/2$.

(iii) *Rotational von Mises–Fisher distribution*. The density of a rotational von Mises–Fisher distribution is given by

$$f_U\bigl(u, (\lambda, A)\bigr) = c(\lambda, A)^{-1} \exp\bigl(\lambda \cdot \operatorname{Tr}\bigl(A^{-1}u\bigr)\bigr), \quad u \in \mathrm{SO}(3),$$

where $A \in \mathrm{SO}(3)$ is an additional parameter that indicates the mean direction and $c(\lambda, A)$ is the normalizing constant. This belongs to the log-supersmoothness scenario (S3) with $\beta = 1/2$, $\gamma = 1/2$, $\alpha = 2$, $\xi_1 = 2(1 + \log \lambda)$ and $\xi_2 = 2(1 + \log(3\lambda))$ ([50]).

Acknowledgments. The authors would like to thank an Associate Editor and four referees for helpful and constructive comments, which led to significant improvement of our paper.

Funding. The first author acknowledges financial support from the European Research Council (2016–2021, Horizon 2020/ERC grant agreement No. 694409) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2020R1A6A3A03037314). Research of the second author was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2019R1A2C3007355). The third author acknowledges financial support from the European Research Council (2016–2021, Horizon 2020/ERC grant agreement No. 694409).

SUPPLEMENTARY MATERIAL

Supplement to "Nonparametric regression on Lie groups with measurement errors" (DOI: 10.1214/22-AOS2218SUPP; .pdf). The Supplementary Material contains additional propositions and all technical proofs. It starts with a brief introduction to some notions on manifolds and to harmonic analysis on general Lie groups.

REFERENCES

[1] APPLEBAUM, D. (2014). *Probability on Compact Lie Groups*. *Probability Theory and Stochastic Modelling* **70**. Springer, Cham. MR3243650 https://doi.org/10.1007/978-3-319-07842-7

[2] ASWANI, A., BICKEL, P. and TOMLIN, C. (2011). Regression on manifolds: Estimation of the exterior derivative. *Ann. Statist.* **39** 48–81. MR2797840 https://doi.org/10.1214/10-AOS823

[3] BERESTOVSKIĬ, V. N. and SVIRKIN, V. M. (2010). The Laplace operator on normal homogeneous Riemannian manifolds. *Siberian Adv. Math.* **20** 231–255.

[4] BHATTACHARYA, A. and DUNSON, D. B. (2010). Nonparametric Bayesian density estimation on manifolds with applications to planar shapes. *Biometrika* **97** 851–865. MR2746156 https://doi.org/10.1093/biomet/asq044

[5] CHAKRABORTY, A. and PANARETOS, V. M. (2017). Regression with genuinely functional errors-in-covariates. arXiv preprint, arXiv:1712.04290.

[6] CHANG, T. (1989). Spherical regression with errors in variables. *Ann. Statist.* **17** 293–306. MR0981451 https://doi.org/10.1214/aos/1176347017

[7] CHEN, C., GUO, S. and QIAO, X. (2022). Functional linear regression: Dependence and error contamination. *J. Bus. Econom. Statist.* **40** 444–457. MR4356585 https://doi.org/10.1080/07350015.2020.1832503

[8] CHEN, M., JIANG, H., LIAO, W. and ZHAO, T. (2022). Nonparametric regression on low-dimensional manifolds using deep ReLU networks: Function approximation and statistical recovery. arXiv preprint. Available at arXiv:1908.01842v5.

[9] CHEN, S. X. and VAN KEILEGOM, I. (2009). A review on empirical likelihood methods for regression. *TEST* **18** 415–447. MR2566404 https://doi.org/10.1007/s11749-009-0159-5

[10] CHENG, M.-Y. and WU, H.-T. (2013). Local linear regression on manifolds and its geometric interpretation. *J. Amer. Statist. Assoc.* **108** 1421–1434. MR3174718 https://doi.org/10.1080/01621459.2013.827984

[11] COMTE, F., SAMSON, A. and STIRNEMANN, J. J. (2014). Deconvolution estimation of onset of pregnancy with replicate observations. *Scand. J. Stat.* **41** 325–345. MR3207174 https://doi.org/10.1111/sjos.12029

[12] CORNEA, E., ZHU, H., KIM, P. and IBRAHIM, J. G. (2017). Regression models on Riemannian symmetric spaces. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 463–482. MR3611755 https://doi.org/10.1111/rssb.12169

[13] DATTNER, I., REISS, M. and TRABS, M. (2016). Adaptive quantile estimation in deconvolution with unknown error distribution. *Bernoulli* **22** 143–192. MR3449779 https://doi.org/10.3150/14-BEJ626

[14] DELAIGLE, A. (2014). Nonparametric kernel methods with errors-in-variables: Constructing estimators, computing them, and avoiding common mistakes. *Aust. N. Z. J. Stat.* **56** 105–124. MR3226432 https://doi.org/10.1111/anzs.12066

[15] DELAIGLE, A., FAN, J. and CARROLL, R. J. (2009). A design-adaptive local polynomial estimator for the errors-in-variables problem. *J. Amer. Statist. Assoc.* **104** 348–359. MR2504382 https://doi.org/10.1198/jasa.2009.0114

[16] DELAIGLE, A., HALL, P. and JAMSHIDI, F. (2015). Confidence bands in non-parametric errors-in-variables regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 149–169. MR3299403 https://doi.org/10.1111/rssb.12067

[17] DELAIGLE, A., HALL, P. and MEISTER, A. (2008). On deconvolution with repeated measurements. *Ann. Statist.* **36** 665–685. MR2396811 https://doi.org/10.1214/009053607000000884

[18] DELAIGLE, A. and VAN KEILEGOM, I. (2021). Deconvolution with unknown error distribution. In *Handbook on Measurement Error Models* (G. Yi, A. Delaigle and P. Gustafson, eds.) CRC Press/CRC, Boca Raton.

[19] DI MARZIO, M., PANZERA, A. and TAYLOR, C. C. (2009). Local polynomial regression for circular predictors. *Statist. Probab. Lett.* **79** 2066–2075. MR2571770 https://doi.org/10.1016/j.spl.2009.06.014

[20] DIGGLE, P. J. and HALL, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *J. Roy. Statist. Soc. Ser. B* **55** 523–531. MR1224414

[21] DUNCAN, B. N., LAMSAL, L. N., THOMPSON, A. M., YOSHIDA, Y., LU, Z., STREETS, D. G., HURWITZ, M. M. and PICKERING, K. E. (2016). A space-based, high-resolution view of notable changes in urban NOx pollution around the world (2005–2014). *J. Geophys. Res., Atmos.* **121** 976–996.

[22] FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272. MR1126324 https://doi.org/10.1214/aos/1176348248

[23] FAN, J. (1991). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā Ser. A* **53** 97–110. MR1177770

[24] FAN, J. and TRUONG, Y. K. (1993). Nonparametric regression with errors in variables. *Ann. Statist.* **21** 1900–1925. MR1245773 https://doi.org/10.1214/aos/1176349402

[25] FARAUT, J. (2008). *Analysis on Lie Groups. Cambridge Studies in Advanced Mathematics* **110**. Cambridge Univ. Press, Cambridge. MR2426516 https://doi.org/10.1017/CBO9780511755170

[26] FOLLAND, G. B. (2016). *A Course in Abstract Harmonic Analysis*. CRC Press/CRC, Boca Raton, FL.

[27] GAO, F., HUANG, X.-Y., JACOBS, N. A. and WANG, H. (2015). Assimilation of wind speed and direction observations: Results from real observation experiments. *Tellus, Ser. A Dyn. Meteorol. Oceanogr.* **67** 27132.

[28] GARCÍA-PORTUGUÉS, E., CRUJEIRAS, R. M. and GONZÁLEZ-MANTEIGA, W. (2013). Exploring wind direction and SO_2 concentration by circular-linaer density estimation. *Stoch. Environ. Res. Risk Assess.* **27** 1055–1067.

[29] GARCÍA-PORTUGUÉS, E., CRUJEIRAS, R. M. and GONZÁLEZ-MANTEIGA, W. (2013). Kernel density estimation for directional-linear data. *J. Multivariate Anal.* **121** 152–175. MR3090475 https://doi.org/10.1016/j.jmva.2013.06.009

[30] GONZALEZ-MANTEIGA, W., HENRY, G. and RODRIGUEZ, D. (2012). Partly linear models on Riemannian manifolds. *J. Appl. Stat.* **39** 1797–1809. MR2935559 https://doi.org/10.1080/02664763.2012.683169

[31] GRAFAKOS, L. (2008). *Classical Fourier Analysis*, 2nd ed. *Graduate Texts in Mathematics* **249**. Springer, New York. MR2445437

[32] HALL, P., WATSON, G. S. and CABRERA, J. (1987). Kernel density estimation with spherical data. *Biometrika* **74** 751–762. MR0919843 https://doi.org/10.1093/biomet/74.4.751

[33] HEALY, D. M. JR., HENDRIKS, H. and KIM, P. T. (1998). Spherical deconvolution. *J. Multivariate Anal.* **67** 1–22. MR1659108 https://doi.org/10.1006/jmva.1998.1757

[34] HENDRIKS, H. (1990). Nonparametric estimation of a probability density on a Riemannian manifold using Fourier expansions. *Ann. Statist.* **18** 832–849. MR1056339 https://doi.org/10.1214/aos/1176347628

[35] HENRY, G. and RODRIGUEZ, D. (2009). Robust nonparametric regression on Riemannian manifolds. *J. Nonparametr. Stat.* **21** 611–628. MR2543576 https://doi.org/10.1080/10485250902846439

[36] HEYER, H. (1977). *Probability Measures on Locally Compact Groups*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete*, *Band* 94. Springer, Berlin. MR0501241

[37] HJORT, N. L., MCKEAGUE, I. W. and VAN KEILEGOM, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37** 1079–1111. MR2509068 https://doi.org/10.1214/07-AOS555

[38] HUCKEMANN, S. F., KIM, P. T., KOO, J.-Y. and MUNK, A. (2010). Möbius deconvolution on the hyperbolic plane with application to impedance density estimation. *Ann. Statist.* **38** 2465–2498. MR2676895 https://doi.org/10.1214/09-AOS783

[39] JADHAV, S. and MA, S. (2020). Functional measurement error in functional regression. *Canad. J. Statist.* **48** 238–258. MR4095262 https://doi.org/10.1002/cjs

[40] JEON, J. M., PARK, B. U. and VAN KEILEGOM, I. (2021). Additive regression for non-Euclidean responses and predictors. *Ann. Statist.* **49** 2611–2641. MR4338377 https://doi.org/10.1214/21-aos2048

[41] JEON, J. M, PARK, B. U and VAN KEILEGOM, I. (2022). Supplement to "Nonparametric regression on Lie groups with measurement errors." https://doi.org/10.1214/22-AOS2218SUPP

[42] JIAO, Y., SHEN, G., LIN, Y. and HUANG, J. (2022). Deep nonparametric regression on approximately low-dimensional manifolds. arXiv preprint. Available at arXiv:2104.06708v4.

[43] JOHANNES, J. (2009). Deconvolution with unknown error distribution. *Ann. Statist.* **37** 2301–2323. MR2543693 https://doi.org/10.1214/08-AOS652

[44] KAPPUS, J. and MABON, G. (2014). Adaptive density estimation in deconvolution problems with unknown error distribution. *Electron. J. Stat.* **8** 2879–2904. MR3299125 https://doi.org/10.1214/14-EJS976

[45] KATZNELSON, Y. (2004). *An Introduction to Harmonic Analysis*, 3rd ed. *Cambridge Mathematical Library*. Cambridge Univ. Press, Cambridge. MR2039503 https://doi.org/10.1017/CBO9781139165372

[46] KERKYACHARIAN, G., PHAM NGOC, T. M. and PICARD, D. (2011). Localized spherical deconvolution. *Ann. Statist.* **39** 1042–1068. MR2816347 https://doi.org/10.1214/10-AOS858

[47] KIM, P. T. (1998). Deconvolution density estimation on SO(N). *Ann. Statist.* **26** 1083–1102. MR1635446 https://doi.org/10.1214/aos/1024691089

[48] KIM, P. T. and KOO, J.-Y. (2002). Optimal spherical deconvolution. *J. Multivariate Anal.* **80** 21–42. MR1889831 https://doi.org/10.1006/jmva.2000.1968

[49] KIM, P. T. and KOO, J.-Y. (2005). Statistical inverse problems on manifolds. *J. Fourier Anal. Appl.* **11** 639–653. MR2190676 https://doi.org/10.1007/s00041-005-3041-1

[50] KIM, P. T. and RICHARDS, D. S. P. (2001). Deconvolution density estimation on compact Lie groups. In *Algebraic Methods in Statistics and Probability* (*Notre Dame*, *IN*, 2000). *Contemp. Math.* **287** 155–171. Amer. Math. Soc., Providence, RI. MR1873674 https://doi.org/10.1090/conm/287/04784

[51] KRAMER, L. (2011). The topology of a semisimple Lie group is essentially unique. *Adv. Math.* **228** 2623–2633. MR2838051 https://doi.org/10.1016/j.aim.2011.07.019

[52] LANG, S. (1999). *Fundamentals of Differential Geometry*. *Graduate Texts in Mathematics* **191**. Springer, New York. MR1666820 https://doi.org/10.1007/978-1-4612-0541-8

[53] LIN, Z. and YAO, F. (2019). Intrinsic Riemannian functional data analysis. *Ann. Statist.* **47** 3533–3577. MR4025751 https://doi.org/10.1214/18-AOS1787

[54] LIN, Z. and YAO, F. (2021). Functional regression on the manifold with contamination. *Biometrika* **108** 167–181. MR4226196 https://doi.org/10.1093/biomet/asaa041

[55] LUO, Z. M., KIM, P. T., KIM, T. Y. and KOO, J. Y. (2011). Deconvolution on the Euclidean motion group $\mathbb{SE}(3)$. *Inverse Probl.* **27** 035014. MR2772533 https://doi.org/10.1088/0266-5611/27/3/035014

[56] MARRON, J. S. and ALONSO, A. M. (2014). Overview of object oriented data analysis. *Biom. J.* **56** 732–753. MR3258083 https://doi.org/10.1002/bimj.201300072

[57] MEISTER, A. (2009). *Deconvolution Problems in Nonparametric Statistics*. *Lecture Notes in Statistics* **193**. Springer, Berlin. MR2768576 https://doi.org/10.1007/978-3-540-87557-4

[58] MINAKSHISUNDARAM, S. and PLEIJEL, Å. (1949). Some properties of the eigenfunctions of the Laplace-operator on Riemannian manifolds. *Canad. J. Math.* **1** 242–256. MR0031145 https://doi.org/10.4153/cjm-1949-021-5

[59] MOORE, C. C. (1972). Groups with finite dimensional irreducible representations. *Trans. Amer. Math. Soc.* **166** 401–410. MR0302817 https://doi.org/10.2307/1996058

[60] MYERS, D. F. (2016). *Pointwise and Uniform Convergence of Fourier Series on SU*(2). ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—Missouri University of Science and Technology. MR3597695

[61] NEUMANN, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Stat.* **7** 307–330. MR1460203 https://doi.org/10.1080/10485259708832708

[62] NEUMANN, M. H. (2007). Deconvolution from panel data with unknown error distribution. *J. Multivariate Anal.* **98** 1955–1968. MR2396948 https://doi.org/10.1016/j.jmva.2006.09.012

[63] OWEN, A. (2001). *Empirical Likelihood*. CRC Press/CRC, London.

[64] PELLETIER, B. (2005). Kernel density estimation on Riemannian manifolds. *Statist. Probab. Lett.* **73** 297–304. MR2179289 https://doi.org/10.1016/j.spl.2005.04.004

[65] PELLETIER, B. (2006). Non-parametric regression estimation on closed Riemannian manifolds. *J. Nonparametr. Stat.* **18** 57–67. MR2214065 https://doi.org/10.1080/10485250500504828

[66] PROCESI, C. (2007). *Lie Groups: An Approach Through Invariants and Representations. Universitext.* Springer, New York. MR2265844

[67] SAKURAI, J. J. and NAPOLITANO, J. (2017). *Modern Quantum Mechanics*. Cambridge Univ. Press, Cambridge.

[68] SEI, T., SHIBATA, H., TAKEMURA, A., OHARA, K. and TAKAYAMA, N. (2013). Properties and applications of Fisher distribution on the rotation group. *J. Multivariate Anal.* **116** 440–455. MR3049915 https://doi.org/10.1016/j.jmva.2013.01.010

[69] SEPANSKI, M. R. (2007). *Compact Lie Groups. Graduate Texts in Mathematics* **235**. Springer, New York. MR2279709 https://doi.org/10.1007/978-0-387-49158-5

[70] SHAO, L., LIN, Z. and YAO, F. (2022). Intrinsic Riemannian functional data analysis for sparse longitudinal observations. *Ann. Statist.* **50** 1696–1721. MR4441137 https://doi.org/10.1214/22-aos2172

[71] SONG, W. (2011). Empirical likelihood confidence intervals for density function in errors-in-variables model. *J. Statist. Res.* **45** 95–110. MR2934364

[72] STEFANSKI, L. and CARROLL, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **21** 169–184. MR1054861 https://doi.org/10.1080/02331889008802238

[73] TERRAS, A. (2016). *Harmonic Analysis on Symmetric Spaces—Higher Rank Spaces, Positive Definite Matrix Space and Generalizations*, 2nd ed. Springer, New York. MR3496932 https://doi.org/10.1007/978-1-4939-3408-9

[74] VAN DEN BAN, E., FLENSTED-JENSEN, M. and SCHLICHTKRULL, H. (1994). Basic harmonic analysis on pseudo-Riemannian symmetric spaces. In *Noncompact Lie Groups and Some of Their Applications* (E. A. Tanner and R. Wilson, eds.) Springer, Dordrecht.

[75] VOLLRATH, A. (2010). The nonequispaced fast SO(3) Fourier transform, generalisations and applications Ph.D. thesis, Universität Lübeck.

[76] WALKER, P. L. (1969). Lipschitz classes on finite dimensional groups. *Proc. Camb. Philos. Soc.* **66** 31–38. MR0240565 https://doi.org/10.1017/s0305004100044686

[77] WALTERS, P. (1982). *An Introduction to Ergodic Theory. Graduate Texts in Mathematics* **79**. Springer, New York-Berlin. MR0648108

[78] WANG, J.-L., CHIOU, J.-M. and MÜLLER, H.-G. (2016). Functional data analysis. *Annu. Rev. Stat. Appl.* **3** 257–295.

[79] YAN, L. and CHEN, X. (2014). Empirical likelihood for partly linear models with errors in all variables. *J. Multivariate Anal.* **130** 275–288. MR3229538 https://doi.org/10.1016/j.jmva.2014.06.007

[80] YUAN, Y., ZHU, H., LIN, W. and MARRON, J. S. (2012). Local polynomial regression for symmetric positive definite matrices. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 697–719. MR2965956 https://doi.org/10.1111/j.1467-9868.2011.01022.x