Biometrika (xxxx), **xxx**, x, *pp*. 1–19
Printed in Great Britain

Additive Models for Symmetric Positive-Definite Matrices and Lie Groups

BY Z. LIN

Department of Statistics and Applied Probability, National University of Singapore
21 Lower Kent Ridge Road, 119077, Singapore
linz@nus.edu.sg

H.-G. MÜLLER

Department of Statistics, University of California, Davis
One Shields Avenue, Davis, CA 95616, U.S.A.
hgmueller@ucdavis.edu

AND B. U. PARK

Department of Statistics, Seoul National University
1 Gwanak-ro, Gwanak-gu, Seoul, Republic of Korea
bupark@stats.snu.ac.kr

FOR THE ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE*

SUMMARY

We propose and investigate an additive regression model for symmetric positive-definite matrix valued responses and multiple scalar predictors. The model exploits the abelian group structure inherited from either the Log-Cholesky metric or the Log-Euclidean framework that turns the space of symmetric positive-definite matrices into a Riemannian manifold and further a bi-invariant Lie group. The additive model for responses in the space of symmetric positive-definite matrices with either of these metrics is shown to connect to an additive model on a tangent space. This connection not only entails an efficient algorithm to estimate the component functions but also allows to generalize the proposed additive model to general Riemannian manifolds. Beyond symmetric positive-definite matrix valued responses the proposed additive model also covers more general Lie groups. Optimal asymptotic convergence rates and normality of the estimated component functions are established and numerical studies show that the proposed model enjoys good numerical performance and is not subject to the curse of dimensionality when there are multiple predictors. The practical merits of the proposed model are demonstrated through an analysis of brain diffusion tensor imaging data.

Some key words: Additive regression; Asymptotic normality; Diffusion tensor; Log-Euclidean metric; Log-Cholesky metric; Riemannian manifold; Toroidal data.

* Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`adni.loni.usc.edu`). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: `http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`

1. INTRODUCTION

Data in the form of symmetric positive-definite matrices arise in many areas, including computer vision (Caseiro et al., 2012; Rathi et al., 2007), signal processing (Arnaudon et al., 2013; Hua et al., 2017), medical imaging (Dryden et al., 2009; Fillard et al., 2007) and neuroscience (Friston, 2011), among other fields and applications. For instance, they are used to model brain functional connectivity that is often characterized by covariance matrices of blood-oxygen-level dependent signals (Huettel et al., 2008). In diffusion tensor imaging analysis (Le Bihan, 1991), a 3×3 symmetric positive matrix that is computed for each voxel describes the dominant shape of local diffusion of water molecules. While the analysis of symmetric positive-definite matrices as responses in a regression model is our primary emphasis in this paper, our results more generally extend to responses with a Lie group structure that include data on the torus.

The space $\mathcal{S}^+$ of symmetric positive-definite matrices is a nonlinear metric space and, depending on the metric, forms a Riemannian manifold. Various metrics have been studied (Pigoli et al., 2014); an important criterion for choosing a metric is to avoid the swelling effect. This refers to the phenomenon that the determinant of the Fréchet mean of a set of symmetric positive-definite matrices may be substantially larger than that of any of the constituent matrices. The swelling effect becomes evident in the geodesics connecting two elements of $\mathcal{S}^+$ (Arsigny et al., 2007), negatively affects specifically the Frobenius metric and various other metrics, and is problematic for many applications. These include diffusion tensor imaging, where a diffusion tensor is represented by a symmetric positive-definite matrix and the determinant quantifies diffusion of the tensor. The swelling effect will then lead to distortions when quantifying diffusion.

The abundance of $\mathcal{S}^+$-valued data in many areas stands in contrast with the relative sparsity of work on their statistical analysis, in particular regarding regression with $\mathcal{S}^+$-valued responses. Existing work includes Riemannian frameworks to analyze diffusion tensor images with a focus on averages and modes of variation (Fletcher & Joshib, 2007; Pennec et al., 2006) and various versions of nonparametric regression such as spline regression (Barmpoutis et al., 2007), local constant regression (Davis et al., 2010), intrinsic local linear regression (Yuan et al., 2012), wavelet regression (Chau & von Sachs, 2019) and Fréchet regression (Petersen et al., 2019). Various metric, manifold and Lie group structures have been proposed, for example, the trace metric (Lang, 1999), affine-invariant metric (Moakher, 2005; Pennec et al., 2006; Fletcher & Joshib, 2007), Log-Euclidean metric (Arsigny et al., 2007), Log-Cholesky metric (Lin, 2019), scaling-rotation distance (Jung et al., 2015) and Procrustes distance (Dryden et al., 2009). As $\mathcal{S}^+$is a Riemannian manifold and more generally a metric space, regression techniques developed for general Riemannian manifolds (e.g. Pelletier, 2006; Shi et al., 2009; Steinke et al., 2010; Davis et al., 2010; Fletcher, 2013; Hinkle et al., 2014; Cornea et al., 2017, among many others) and metric spaces (Hein, 2009; Chen & Müller, 2022; Lin & Müller, 2021) also apply to $\mathcal{S}^+$.

Additive regression originated with Stone (1985) and is known to be an efficient way of avoiding the well known curse of dimensionality problem that one faces in nonparametric regression when the dimension of the covariate vector increases, but so far has been by and large limited to the case of real-valued and functional responses. Examples for additive regression approaches for real-valued responses include the original work on smooth backfitting (Mammen et al., 1999) and extensions to generalized additive models (Yu et al., 2008), additive quantile models (Lee et al., 2010), generalized varying coefficient models (Lee et al., 2012), errors-in-variables (Han & Park, 2018) and functional or distributional responses (Scheipl et al., 2015; Park et al., 2018; Han et al., 2020; Jeon & Park, 2020).

This paper contains three major contributions. First, to the best of our knowledge, this is the first paper to study additive regression for $\mathcal{S}^+$-valued responses, counteracting the curse

of dimensionality while maintaining a high degree of flexibility. Previous studies for modeling $\mathcal{S}^+$-valued responses focused on unstructured nonparametric regression such as local constant/polynomial regression that are well known to be subject to the curse of dimensionality when there are many predictors.

Second, by focusing on abelian Lie group-valued responses that include data on the torus (Eltzner et al., 2018) as well as $\mathcal{S}^+$-valued responses, we propose a novel intrinsic group additive regression model that directly exploits the abelian group structure of the responses. This sets our work apart, as previously only the general manifold structure of responses was considered in the few existing non-additive regression approaches. To the best of our knowledge, this is the first work to connect additive regression with Lie groups. This connection extends the reach of additive models in a very natural way.

Third, we show that this group additive model can be transformed into an additive model on tangent spaces by utilizing the Riemannian logarithmic map, which paves the way for extending the additive model to the case of responses that lie on more general manifolds. While throughout we showcase the proposed approaches for the space $\mathcal{S}_m^+$ of symmetric positive-definite matrices with suitable metrics, our results are by no means limited to this type of responses and are applicable to additive regression modeling for a much larger class of manifold-valued responses.

2. Preliminaries on Differential Geometry

We compile here some basic notions for Riemannian manifolds and Lie groups, referring readers to Section S.1 of Shao et al. (2022+) for a self-contained note on basic concepts of Riemannian geometry and to the text by Lee (2018) for a comprehensive treatment. In the following, $\mathcal{L}$ denotes a simply connected smooth manifold situated in a D-dimensional Euclidean space.

Tangent vectors, tangent spaces and Riemannian metrics. The tangent space at $y \in \mathcal{L}$, denoted by $T_y\mathcal{L}$, is a linear space consisting of velocity vectors $\alpha'(0)$ where $\alpha : (-1, 1) \to \mathcal{L}$ represents a differentiable curve passing through y, i.e. $\alpha(0) = y$. Each tangent space $T_y\mathcal{L}$ is endowed with an inner product g_y that varies smoothly with y and is a D-dimensional Hilbert space with the induced norm denoted by $\|\cdot\|_y$. The inner products $\{g_y : y \in \mathcal{L}\}$ are collectively denoted by g, referred to as the Riemannian metric of $\mathcal{L}$ that also defines a distance d on $\mathcal{L}$.

Geodesics and exponential maps. A geodesic γ is a constant-speed curve defined on $[0, \infty)$ such that for each $t \in [0, \infty)$, the segment $\gamma([t, t + \epsilon])$ is the shortest path connecting $\gamma(t)$ and $\gamma(t + \epsilon)$ for all sufficiently small $\epsilon > 0$. The Riemannian exponential map Exp_y at $y \in \mathcal{L}$ is a function mapping $T_y\mathcal{L}$ into $\mathcal{L}$ and defined by $\mathrm{Exp}_y(u) = \gamma(1)$ with $\gamma(0) = y$ and $\gamma'(0) = u \in T_y\mathcal{L}$. Conversely, $\gamma_{y,u}(t) = \mathrm{Exp}_y(tu)$ is a geodesic starting at y and with direction u.

Cut time and logarithmic maps. For a tangent vector $u \in T_y\mathcal{L}$, the cut time c_u is the positive number such that $\gamma_{y,u}([0, c_u])$ is a shortest path connecting $\gamma_{y,u}(0)$ and $\gamma_{y,u}(c_u)$, but $\gamma_{y,u}([0, c_u + \epsilon])$ is not a shortest path for any $\epsilon > 0$. It turns out that the exponential map $\mathrm{Exp}_y(tu)$ for $u \in T_y\mathcal{L}$ is invertible before the cut time c_u. Formally, the inverse of Exp_y, denoted by Log_y and called the Riemannian logarithmic map at y, can be defined by $\mathrm{Log}_y z = tu$ for $z \in \mathcal{E}_y := \{\mathrm{Exp}_y(tu) : u \in T_y\mathcal{L}, \|u\|_y = 1, 0 \le t < c_u\}$ such that $\mathrm{Exp}_y(tu) = z$.

Covariant derivative of smooth functions. Let $C^\infty(\mathcal{L})$ denote the collection of smooth real-valued functions defined on $\mathcal{L}$. For a smooth function $f \in C^\infty(\mathcal{L})$ and a tangent vector $v \in T_y\mathcal{L}$, the covariant derivative of f at y along the direction v, denoted by $\nabla_v f$, is defined by

$$\nabla_v f := (f \circ \gamma)'(0) = \lim_{t \to 0} \frac{f\{\gamma(t)\} - f(y)}{t},$$

where $\gamma : [-1, 1] \to \mathcal{L}$ is a differentiable curve such that $\gamma(0) = y$ and $\gamma'(0) = v$.

Covariant derivative of smooth vector fields. A smooth vector field U is a smooth function defined on $\mathcal{L}$ such that $U(y) \in T_y\mathcal{L}$ for all $y \in \mathcal{L}$. The covariant derivative measures how fast a map changes along a direction ands is defined for smooth vector fields as follows. Let $\Gamma(\mathcal{L})$ denote the collection of smooth vector fields on $\mathcal{L}$. A *connection* is a map $\nabla : \Gamma(\mathcal{L}) \times \Gamma(\mathcal{L}) \to \Gamma(\mathcal{L})$, with $(V, U) \mapsto \nabla_V U$, that satisfies the following properties:

- $\nabla_V U$ is linear over $C^\infty(\mathcal{L})$ in V, i.e. $\nabla_{fV_1+gV_2}U = f\nabla_{V_1}U + g\nabla_{V_2}U$ for $f, g \in C^\infty(\mathcal{L})$ and $V_1, V_2 \in \Gamma(\mathcal{L})$;
- $\nabla_V U$ is linear over $\mathbb{R}$ in U, i.e. $\nabla_V(a_1U_1 + a_2U_2) = a_1\nabla_V U_1 + a_2\nabla_V U_2$ for $a_1, a_2 \in \mathbb{R}$ and $U_1, U_2 \in \Gamma(\mathcal{L})$;
- $\nabla_V(fU) = f\nabla_V U + (\nabla_V f)U$ for $f \in C^\infty(\mathcal{L})$,

where, for $f \in C^\infty(\mathcal{L})$ and a smooth vector field U, fU denotes a smooth vector field defined by $(fU)(y) = f(y)U(y)$ for all $y \in \mathcal{L}$, and $\nabla_V f$ is a smooth real-valued function defined by $(\nabla_V f)(y) = \nabla_{V(y)} f$ for $y \in \mathcal{L}$. The quantity $\nabla_V U$ is called the covariant derivative of U in the direction V. Note that the value of $\nabla_V U$ at y depends on V only through its value at y (Proposition 4.5, Lee, 2018), which makes the expression $\nabla_v U$ sensible for $v \in T_y\mathcal{L}$; $\nabla_v U$ is called the covariant derivative of U at y in the direction v.

Levi–Civita covariant derivative. For $U, V \in \Gamma(\mathcal{L})$, the function $f_{U,V}$ defined by $f_{U,V}(y) = g_y(U(y), V(y))$ is in $C^\infty(\mathcal{L})$. A connection ∇ on $\mathcal{L}$ is compatible with the metric g on $\mathcal{L}$ if $\nabla_v f_{U,V} = g_y(\nabla_v U, V(y)) + g_y(U(y), \nabla_v V)$ for all $U, V \in \Gamma(\mathcal{L})$, each $y \in \mathcal{L}$ and each tangent vector $v \in T_y\mathcal{L}$. For $U, V \in \Gamma(\mathcal{L})$, $[U, V]$ denotes a new vector field satisfying $\nabla_{[U,V]} f = \nabla_U\nabla_V f - \nabla_V\nabla_U f$ for all $f \in C^\infty(\mathcal{L})$. If $\nabla_U V - \nabla_V U = [U, V]$ for all $U, V \in \Gamma(\mathcal{L})$, then we say the connection ∇ is torsion-free. For a Riemannian manifold, there exists a unique connection that is both torsion-free and compatible with the Riemannian metric. It is called the Levi–Civita connection with Levi–Civita covariant derivative as its induced covariant derivative.

Sectional Curvature. Curvature quantifies the degree of deviation from being flat. Define the map $R(U, V, W) = \nabla_U\nabla_V W - \nabla_V\nabla_U W - \nabla_{[U,V]}W$ for $U, V, W \in \Gamma(\mathcal{L})$. The value of $R(U, V, W)$ at y depends only on the values of U, V, W at y, and therefore we can write $R(u, v, w)$ for tangent vectors u, v, w at the same point. The *sectional curvature* at $y \in \mathcal{L}$ is a real-valued function on $T_y\mathcal{L} \times T_y\mathcal{L}$ defined for $u, v \in T_y\mathcal{L}$ by

$$\mathfrak{K}(u, v) = \frac{g_y\{R(u, v, v), u\}}{g_y(u, u)g_y(v, v) - g_y(u, v)^2}.$$

A *Hadamard manifold* is a complete and simply connected Riemannian manifold that has everywhere non-positive sectional curvature and thus is a Hadamard space.

Parallel transport. Given a curve $\gamma(t)$ on $\mathcal{L}$, $t \in I$ for a real interval I, a vector field U along γ is a smooth map defined on I such that $U(t) \in T_{\gamma(t)}\mathcal{L}$. We say U is parallel along γ if $\nabla_{\gamma'(t)}U = 0$ for all $t \in I$. In this paper, we primarily focus on parallel vector fields along geodesics. Let $\gamma : [0, 1] \to \mathcal{L}$ be a geodesic connecting y and z, and U a parallel vector field along γ such that $U(0) = u$ and $U(1) = v$. Then we say v is the parallel transport of u along γ, denoted by $\tau_{y,z}u = v$. Parallel transport can be used as an intrinsic mechanism to compare tangent vectors residing at different points, e.g. via parallelly transporting the tangent vectors to the tangent space at a fixed point on $\mathcal{L}$, where they can be easily compared.

Lie algebra. A Lie algebra is a vector space $\mathfrak{g}$ endowed with an alternating binary operation $[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ satisfying the following axioms:

- (Bilinearity) $[au + bv, w] = a[u, w] + b[v, w]$ and $[w, au + bv] = a[w, u] + b[w, v]$ for all $a, b \in \mathbb{R}$ and $u, v, w \in \mathfrak{g}$;
- (Alternativity) $[u, u] = 0$ for all $u \in \mathfrak{g}$;
- (Jacobi identity) $[u, [v, w]] + [v, [w, u]] + [w, [u, v]] = 0$ for all $u, v, w \in \mathfrak{g}$.

Lie group. When $\mathcal{L}$ is a group such that the group operation $\oplus$ and inverse $\iota : y \mapsto y^{-1}$ are smooth, $(\mathcal{L}, \oplus)$ is a Lie group $\mathcal{L}$. We say a vector field U on a Lie group $(\mathcal{L}, \oplus)$ is *left-invariant* if $U(y \oplus z) = (D_z L_y)(U(z))$ for all $y, z \in \mathcal{L}$, where $L_y : z \mapsto y \oplus z$ is the left translation induced by y and $D_z L_y$ is the differential of L_y at z. Right-invariant vector fields are defined in a similar fashion via right translations. One can show that any left-invariant vector field U is fully specified by its value $U(e)$ at the group identity element e. Consequently, the collection of left-invariant vector fields, denoted by $\mathfrak{h}$, can be identified with the tangent space $T_e\mathcal{L}$. In addition, the space $\mathfrak{h}$ gives rise to a Lie algebra, as follows. For two smooth vector fields U, V on the Lie group $\mathcal{L}$, the Lie bracket $[U, V]$ is the vector field determined by $[U, V](f) = \nabla_U \nabla_V f - \nabla_V \nabla_U f$ for $f \in C^\infty(\mathcal{L})$. It can be shown that the vector space $\mathfrak{h}$ endowed with the Lie bracket is a Lie algebra. If $\mathcal{L}$ is abelian, then $[U, V] = 0$ for any $U, V \in \mathfrak{h}$.

Fréchet function and Fréchet means. For random elements $Y \in \mathcal{L}$, for a Lie group $\mathcal{L}$, the Fréchet function is $F(y) = \mathbb{E}d^2(y, Y)$, where d is the Riemannian distance function induced by the metric on $\mathcal{L}$. If $\mathcal{L}$ is a Hadamard manifold and $F(y) < \infty$ for some $y \in \mathcal{L}$, and hence $F(y) < \infty$ for all $y \in \mathcal{L}$ according to the triangle inequality,the minimizer of $F(y)$ exists and is unique (Sturm, 2003). It is known as the Fréchet mean and is denoted by $\mathbb{E}_o plus Y$.

Bi-invariant metric. A Riemannian metric g on a Lie group is *left-invariant* if $g_z(u, v) = g_{y \oplus z}\{(D_z L_y)u, (D_z L_y)v\}$ for all $y, z \in \mathcal{L}$ and $u, v \in T_z\mathcal{L}$. Right-invariant metrics can be defined in a similar fashion. A metric is *bi-invariant* if it is both left-invariant and right-invariant.

Fréchet function and Fréchet mean. For random elements $Y \in \mathcal{L}$, for a Lie group $\mathcal{L}$, the Fréchet function is $F(y) = \mathbb{E}d^2(y, Y)$, where d is the Riemannian distance function induced by the metric on $\mathcal{L}$. If $\mathcal{L}$ is a Hadamard manifold and $F(y) < \infty$ for some $y \in \mathcal{L}$, and hence $F(y) < \infty$ for all $y \in \mathcal{L}$ according to the triangle inequality, the minimizer of $F(y)$ exists and is unique (Sturm, 2003). It is known as the Fréchet mean and is denoted here by $\mathbb{E}_\oplus Y$.

Lie exponential and logarithmic maps. The Lie exponential map, denoted by $\mathfrak{exp}$ that maps $\mathfrak{g}$ into $\mathcal{L}$, is defined by $\mathfrak{exp}(u) = \gamma(1)$ where $\gamma : \mathbb{R} \to \mathcal{L}$ is the unique one-parameter subgroup such that $\gamma'(0) = u \in \mathfrak{g}$. Its inverse, if it exists, is called the Lie logarithmic map and denoted by $\mathfrak{log}$. When g is bi-invariant, $\mathfrak{exp} = \mathrm{Exp}_e$, i.e. the Riemannian exponential map at the identity element coincides with the Lie exponential map.

3. Additive models for Lie groups

While we develop additive regression models for Lie group-valued responses, the main motivating examples are symmetric positive matrices due to their ubiquity and practical relevance. Accordingly, we first provide details about this motivating example and then consider the general Lie group framework, of which this is a special case. The space of $m \times m$ symmetric positive-definite matrices $\mathcal{S}_m^+$ is a smooth submanifold of $\mathbb{R}^{m \times m}$. Its tangent spaces are $\mathcal{S}_m$, the collection of $m \times m$ symmetric matrices. Upon endowing these tangent spaces with a Riemannian metric g, $\mathcal{S}_m^+$ becomes a Riemannian manifold. Here we focus on the Log-Euclidean (Arsigny et al., 2007) and Log-Cholesky (Lin, 2019) metrics, which are designed to eliminate the swelling effect; extensions to other metrics and general Riemannian manifolds will be discussed in the next section. Each of these metrics is associated with a group operation $\oplus$ that turns $\mathcal{S}_m^+$ into an abelian Lie group in which the metric is bi-invariant. In our experience, the Log-Cholesky metric

is computationally more efficient but it is not invariant to permutation, while the Log-Euclidean metric is invariant to permutation but is not affine-invariant (Arsigny et al., 2007).

Example 1 (*Log-Cholesky metric*). The Log-Cholesky metric utilizes the Cholesky decomposition to transfer a Lie group structure on the space of lower triangular matrices of positive diagonals to the space of symmetric positive-definitive matrices. Specifically, let $\mathrm{LT}(m)$ be the space of $m \times m$ lower triangular matrices and $\mathrm{LT}_+(m) \subset \mathrm{LT}(m)$ the subspace such that $L \in \mathrm{LT}_+(m)$ if all diagonal elements of L are positive. One can show that $\mathrm{LT}_+(m)$ is a smooth submanifold of $\mathrm{LT}(m)$ and its tangent spaces are identified with $\mathrm{LT}(m)$. For a fixed $L \in \mathrm{LT}_+(m)$, we define a Riemannian metric $\tilde{g}$ on $\mathrm{LT}_+(m)$ by $\tilde{g}_L(A, B) = \sum_{1 \le j < i \le m} A_{ij} B_{ij} + \sum_{j=1}^{m} A_{jj} B_{jj} L_{jj}^{-2}$, where A_{ij} denotes the element of A in the ith row and jth column. It is an abelian Lie group with the operation $\circledcirc$ defined by $L_1 \circledcirc L_2 = \mathfrak{L}(L_1) + \mathfrak{L}(L_2) + \mathfrak{D}(L_1)\mathfrak{D}(L_2)$, where $\mathfrak{L}(L)$ is the strict lower triangular part of L, that is, $(\mathfrak{L}(L))_{ij} = L_{ij}$ if $j < i$ and $(\mathfrak{L}(L))_{ij} = 0$ otherwise, and $\mathfrak{D}(L)$ is the diagonal part of L, that is, a diagonal matrix whose diagonals are equal to the respective diagonals of L. One can show that $\tilde{g}$ is a bi-invariant metric for the Lie group $\mathrm{LT}_+(m)$ with the group operation $\circledcirc$. It is well known that a symmetric positive-definite matrix P is associated with a unique matrix L in $\mathrm{LT}_+(m)$ such that $LL^\top = P$. We refer to L as the Cholesky factor of P. For $U, V \in T_P\mathcal{S}_m^+ = \mathcal{S}_m$, we define the metric $g_P(U, V) = \tilde{g}_L(L(L^{-1}UL^{-\top})_{\frac{1}{2}}, L(L^{-1}VL^{-\top})_{\frac{1}{2}})$, where $(S)_{\frac{1}{2}} = \mathfrak{L}(S) + \mathfrak{D}(S)/2$ for a matrix S. We also turn $\mathcal{S}_m^+$ into an abelian Lie group with the operator $\oplus$ such that $P_1 \oplus P_2 = (L_1 \circledcirc L_2)(L_1 \circledcirc L_2)^\top$, where L_1 and L_2 are the Cholesky factors of P_1 and P_2, respectively. The metric g is a bi-invariant metric of the Lie group $(\mathcal{S}_m^+, \oplus)$ and turns $\mathcal{S}_m^+$ into a Hadamard manifold.

Example 2 (*Log-Euclidean metric*). The matrix logarithm is a smooth bijective map between $\mathcal{S}_m^+$ and $\mathcal{S}_m$ and thus can be used to transfer the canonical Euclidean structure of $\mathcal{S}_m$ to $\mathcal{S}_m^+$. Recall that for a symmetric matrix S, $\exp(S) = I_m + \sum_{j=1}^{\infty} \frac{1}{j!} S^j$ is a symmetric positive-definite matrix. The inverse of $\exp$, denoted by $\log$, exists and is called the matrix logarithmic map. Both $\exp$ and $\log$ are smooth maps between $\mathcal{S}_m^+$ and $\mathcal{S}_m$. The operation $\oplus$ defined as $P_1 \oplus P_2 = \exp(\log(P_1) + \log(P_2))$ for $P_1, P_2 \in \mathcal{S}_m^+$ turns $\mathcal{S}_m^+$ into an abelian group. The canonical Riemannian metric on $\mathcal{S}_m$ is $\mathrm{trace}(S_1 S_2)$ for $S_1, S_2 \in \mathcal{S}_m$ and can be transferred to a Riemannian metric on $\mathcal{S}_m^+$ given by $g_P(U, V) = \mathrm{trace}\left[\{(D_P \log)U\}\{(D_P \log)V\}\right]$ for $U, V \in \mathcal{S}_m$, where $D_P \log$ denotes the differential of the $\log$ at P. It turns out that g is a bi-invariant metric on $(\mathcal{S}_m^+, \oplus)$ that is isomorphic to the group of $\mathcal{S}_m$ with the usual matrix addition as the group operation (Proposition 3.4, Arsigny et al., 2007). This metric turns $\mathcal{S}_m^+$ into a Hadamard manifold.

Another example for a relevant Lie group that is not related to the space $\mathcal{S}_m^+$ and shows that our approach is not limited to symmetric positive matrices is as follows.

Example 3 (*Tori*). Let $\mathbb{S} = \{a + b\sqrt{-1} : a, b \in \mathbb{R}, a^2 + b^2 = 1\}$ be the unit circle. It is an abelian Lie group when the group operation is the standard multiplication of complex numbers. In addition, $\mathbb{S}$ is a Riemannian submanifold of $\mathbb{R}^2$. An m-torus $\mathbb{T}^m$ is the direct product of m copies of $\mathbb{S}$, and therefore, is also an abelian Lie group. Its Lie algebra is $\mathfrak{g} = \sqrt{-1}\mathbb{R}^m = \{(a_1\sqrt{-1}, \ldots, a_m\sqrt{-1}) : a_1, \ldots, a_m \in \mathbb{R}\}$ with the trivial Lie bracket $[u, v] = 0$ for all $u, v \in \mathfrak{g}$. The Riemannian metric on $\mathbb{T}^m$ is the product Riemannian metric of $\mathbb{S}$, and this metric is bi-invariant. Data on tori occur in the statistical analysis of wind directions (Hundrieser et al., 2021) and RNA data (Eltzner et al., 2018).

We note that it follows from (...) that Fréchet means are unique if $\mathcal{L} = \mathcal{S}_m^+$, equipped with the Log-Cholesky or Log-Euclidean metric. Given scalar variables $X_1 \in \mathcal{X}_1, \ldots, X_q \in \mathcal{X}_q$, which are predictors that are paired with a Lie group valued response Y and where $\mathcal{X}_j \subset \mathbb{R}$, $j = 1, \ldots q$, are compact domains, the proposed Lie group additive model is as follows, where we make use of the Lie group operation $\oplus$,

$$Y = \mu \oplus w_1(X_1) \oplus \cdots \oplus w_q(X_q) \oplus \zeta. \tag{1}$$

Here $\mu = \mathbb{E}_\oplus Y$ is the Fréchet mean of Y, each w_k is a function that maps X_k into $\mathcal{L}$, and ζ is random noise with $\mathbb{E}_\oplus \zeta = e$, the group identity element e. For identifiability, we require $\mathbb{E}_\oplus\{w_k(X_k)\} = e$, $k = 1, \cdots, q$. This model generalizes the additive model for Euclidean responses to $\mathcal{L}$-valued responses. It includes the effect of of the additive component functions on the mean response and of noise in the responses, neither of which can be additively modeled in the absence of a linear structure of the responses. The task at hand is to estimate the unknown parameter μ and the component functions $w_1, \ldots, w_q$, given a sample of independently and identically distributed (i.i.d.) observations of size n. To overcome the challenge of the absence of a linear structure in $\mathcal{L}$, the following result proves essential.

PROPOSITION 1. *If $(\mathcal{L}, \oplus)$ is an abelian Lie group endowed with a bi-invariant metric g that turns $\mathcal{L}$ into a Hadamard manifold, then (1) is equivalent to*

$$\mathrm{Log}_\mu Y = \sum_{k=1}^{q} \tau_{e,\mu} \mathfrak{log} w_k(X_k) + \tau_{e,\mu} \mathfrak{log} \zeta. \tag{2}$$

In (2), we transform the left-hand side of (1) by the Riemannian log map Log_μ, so that $\mathbb{E}\mathrm{Log}_\mu Y = 0$, and the right-hand side of (1) by the Lie log map $\mathfrak{log}$, whereupon $w_1(X_1) \oplus \cdots \oplus w_q(X_q)$ is decomposed into additive components in the vector space $T_\mu\mathcal{L}$, so that the standard smooth backfitting algorithm can be adopted to estimate $w_1, \ldots, w_q$. Specifically, let $f_k(X_k) = \tau_{e,\mu} \mathfrak{log} w_k(X_k)$ and $\varepsilon = \tau_{e,\mu} \mathfrak{log} \zeta$. Then according to Proposition 1, one may rewrite the model (1) as $\mathrm{Log}_\mu Y = \sum_{k=1}^q f_k(X_k) + \varepsilon$, where $\mathbb{E}\varepsilon = \mathbb{E}\tau_{e,\mu} \mathfrak{log} \zeta = \tau_{e,\mu} \mathbb{E} \mathfrak{log} \zeta = 0$ since $\mathbb{E}_\oplus \zeta = e$. Note that $\mathbb{E}\big(\sum_{k=1}^q f_k(X_k)\big) = 0$ since $\mathbb{E}\mathrm{Log}_\mu Y = 0$. The identifiability of the individual component functions f_k follows from $\mathbb{E} f_k(X_k) = 0$ for all $k = 1, \ldots, q$, which is a consequence of $\mathbb{E}_\oplus\{w_k(X_k)\} = e$ in (1). These considerations motivate to estimate the component functions w_k through estimation of the f_k, as follows.

Step 1: Compute the sample Fréchet mean $\hat{\mu}$. Closed-form expressions of $\hat{\mu}$ are available for special cases of $\mathcal{L}$, including when $\mathcal{L} = \mathcal{S}_m^+$ with the Log-Cholesky or Log-Euclidean metric; see Section S1 of the Supplementary Material for details. Alternative numerical algorithms (e.g. Yang, 2007) are also available.

Step 2: Compute $\mathrm{Log}_{\hat{\mu}} Y_i$. There are closed-form expressions available for $\mathcal{L} = \mathcal{S}_m^+$ with the Log-Cholesky or Log-Euclidean metric; see Section S1 of the Supplementary Material. Numerical methods as described in Section 5.3 of Brun (2007) are also available.

Step 3: Solve the system of integral equations

$$\hat{f}_k(x_k) = \hat{m}_k(x_k) - n^{-1} \sum_{i=1}^{n} \mathrm{Log}_{\hat{\mu}} Y_i - \sum_{j:j\neq k} \int_{\mathcal{X}_j} \hat{f}_j(x_j) \frac{\hat{p}_{kj}(x_k, x_j)}{\hat{p}_k(x_k)} \mathrm{d}x_j, \quad 1 \le k \le q, \tag{3}$$

subject to the constraints $\int_{\mathcal{X}_j} \hat{f}_k(x_k)\hat{p}_k(x_k)\mathrm{d}x_k = 0$ for $1 \le k \le q$. Here, $\hat{p}_k(x_k) = n^{-1}\sum_{i=1}^n K_{h_k}(x_k, X_{ik})$, $\hat{p}_{kj}(x_k, x_j) = n^{-1}\sum_{i=1}^n K_{h_k}(x_k, X_{ik})K_{h_j}(x_j, X_{ij})$, and

$$\hat{m}_k(x_k) = n^{-1}\hat{p}_k(x_k)^{-1}\sum_{i=1}^n K_{h_k}(x_k, X_{ik})\mathrm{Log}_{\hat{\mu}}Y_i, \tag{4}$$

where K_{h_j} is a kernel function with $\int_{\mathcal{X}_j} K_{h_j}(u, v)\,du = 1$ for all $v \in \mathcal{X}_j$, see Jeon & Park (2020). Note that $n^{-1}\sum_{i=1}^n \mathrm{Log}_{\hat{\mu}}Y_i = 0$ since $\hat{\mu}$ is the sample Fréchet mean.

Step 4: Finally, estimate $w_k(x_k)$ by $\hat{w}_k(x_k) = \mathfrak{exp}\{\tau_{\hat{\mu},e}\hat{f}_k(x_k)\}$.

Step 3 is a multivariate version of the standard Smooth Backfitting (SBF) system of equations (Mammen et al., 1999). Since the tangent space $T_{\hat{\mu}}\mathcal{S}_m^+$ is also a Hilbert space, the above SBF system of equations can be interpreted from a Bochner integral perspective, see Jeon & Park (2020), where also the empirical selection of bandwidths h_k is discussed.

Remark 1. According to Theorem 3.6 of Bröcker & tom Dieck (1985), any abelian Lie group is isomorphic to $\mathbb{T}^k \times \mathbb{R}^s$, where $\mathbb{T}^k$ denotes a k-torus defined in Example 3. However, our model, estimation and theory do not rest on this isomorphism, as it does not lead to a natural extension to general Riemannian manifolds. Instead, in the above development we connect the additive model (1) on Lie groups to an additive model (2) on the tangent space at the Fréchet mean via Proposition 1. This allows an immediate extension to general Riemannian manifolds. Since neither our estimation method nor the theory presented in Section 5 rely on the isomorphism to $\mathbb{T}^k \times \mathbb{R}^s$, our methods and theory cover both Lie groups and general Riemannian manifolds.

4. Extension to Riemannian manifolds

When $\mathcal{S}_m^+$ is endowed with the affine-invariant metric (Moakher, 2005; Pennec et al., 2006; Fletcher & Joshib, 2007), it generally is not an abelian Lie group with a bi-invariant metric and then model (1) depends on the order of operations and thus is not additive. Specifically Proposition 1 ceases to hold. However, model (2) remains additive in all cases, suggesting a natural extension to accommodate other metrics and general Riemannian manifolds. If $\mathcal{M}$ is a Riemannian manifold that is not necessarily an abelian Lie group, consider

$$\mathrm{Log}_{\mu}Y = \sum_{k=1}^q f_k(X_k) + \varepsilon, \tag{5}$$

where $\mu = \mathbb{E}_{\oplus}Y$, $\varepsilon \in T_\mu\mathcal{M}$ is centered and of finite variance, and $f_1, \ldots, f_q : \mathbb{R} \to T_\mu\mathcal{M}$ are unknown functions to be estimated. Model (5) includes (1) as a special case by setting $f_k(x) = \tau_{e,\mu}\mathfrak{log}w_k(x)$ and $\varepsilon = \tau_{e,\mu}\mathfrak{log}\zeta$ according to Proposition 1, transforms the response into the tangent space $T_\mu\mathcal{M}$ and thus is applicable to general Riemannian manifolds, at the expense of interpretability in the original space, as per the group operation in model (1).

For general Riemannian manifolds that might feature positive sectional curvature, the Fréchet mean may not exist and therefore additional conditions are required for model (5). Specifically, if $(\mathcal{M}, g)$ denotes a general Riemannian manifold and Y a random element on $\mathcal{M}$, we assume

(A1) The minimizer of the Fréchet function F exists and is unique.

This is automatically satisfied when $\mathcal{M}$ is a Hadamard manifold. For other manifolds, we refer to Bhattacharya & Patrangenaru (2003) and Afsari (2011) for conditions that imply (A1).

For a nonempty subset $A \subset \mathcal{L}$, let $d(y, A) = \inf\{d(y, z) : z \in A\}$ denote the distance between y and the set A. For a positive real number ϵ, set $A^{\epsilon} = \{y : d(y, A) < \epsilon\}$ and $A^{-\epsilon} = \mathcal{M}\backslash(\mathcal{M}\backslash A)^{\epsilon}$. When $A = \emptyset$, set $A^{\epsilon} = \emptyset$. The following assumption (A2) is only needed for the case where $\mathcal{M}$ is not a Hadamard manifold:

(A2) $\Pr\{Y \in \mathcal{E}_{\mu}^{-\epsilon_0}\} = 1$ for some $\epsilon_0 > 0$, where $\mathcal{E}_{\mu}$ is defined in Section 2.

If (A1) and (A2) are satisfied, the proposed manifold additive model (5) remains well defined, and the first three steps of the estimation method described in the previous section are still valid and can be employed to estimate $f_1, \ldots, f_q$, with $\mathcal{S}_m^+$ replaced by $\mathcal{L}$.

5. THEORY

We first establish convergence rates and asymptotic normality of the estimators for the mean and the component functions for general manifolds in the manifold additive model (5) and then provide additional details for the space $\mathcal{S}_m^+$ endowed with either the Log-Cholesky metric or the Log-Euclidean metric. We consider a manifold $\mathcal{M}$ that satisfies at least one of

(M1) $\mathcal{M}$ is a finite-dimensional Hadamard manifold that has sectional curvature bounded from below by $\mathfrak{c}_0 \leq 0$.

(M2) $\mathcal{M}$ is a complete compact Riemannian manifold.

The space $\mathcal{S}_m^+$ with the Log-Cholesky metric, Log-Euclidean metric or affine-invariant metric is a manifold that satisfies (M1), while the unit sphere that is used to model compositional data (Dai & Müller, 2018) serves as an example of a manifold that satisfies (M2).

To establish the convergence rate of $\hat{\mu}$, we also make the following assumptions.

(A3) The manifold $\mathcal{M}$ satisfies at least one of the conditions (M1) and (M2).

(A4) For some constant $\mathfrak{c}_2 > 0$, $F(y) - F(\mu) \geq \mathfrak{c}_2 d^2(y, \mu)$ when $d(y, \mu)$ is sufficiently small.

(A5) For some constant $\mathfrak{c}_3 > 0$, for all $y, z \in \mathcal{M}$, the linear operator $H_{y,z} : T_z\mathcal{M} \to T_z\mathcal{M}$, defined by $g_z(H_{y,z}u, v) = g_z(\nabla_u \mathrm{Log}_z y, v)$ for $u, v \in T_z\mathcal{M}$, has an operator norm that is bounded by $\mathfrak{c}_3\{1 + d(z, y)\}$.

Condition (A4) is satisfied for Hadamard manifolds with $\mathfrak{c}_2 = 1$ according to Lemma S.7 of Lin & Müller (2021) and the $\mathrm{CAT}(0)$ inequality that holds for Hadamard manifolds (Chapter II.1, Bridson & Häfliger, 1999). The condition also holds for some manifolds of positive curvature when data concentrate on a small region; see Example 4 of Lin & Müller (2021). The operator $H_{y,z}$ in the technical condition (A5) is the Hessian of the squared distance function d; see also equation (5.4) of Kendall & Le (2011). Assumption (A5) is superfluous if the manifold $\mathcal{M}$ is compact and is satisfied by manifolds of zero curvature. It can also be replaced by a uniform moment condition on the operator norm of $H_{z,Y}$ over all z in a small local neighborhood of μ. We then obtain a parametric convergence rate for the estimates $\hat{\mu}$ of the Fréchet mean μ.

PROPOSITION 2. *Assume that (A1), (A3) and (A4) hold and Y is of the second order. Then* $d(\hat{\mu}, \mu) = O_P(n^{-1/2})$.

To obtain convergence rates of the estimated component functions, we require some additional conditions that are standard in the literature on additive regression.

(B1) The kernel function K is positive, symmetric, Lipschitz continuous and supported on $[-1, 1]$ with $\int K(x)\, dx = 1$.

(B2) The bandwidths h_k satisfy $n^{1/5} h_k \to \alpha_k > 0$.

(B3) The joint density p of $X_1, \ldots, X_q$ is bounded away from zero and infinity on $\mathcal{X} \equiv \mathcal{X}_1 \times \cdots \times \mathcal{X}_q$. The densities p_{kj} are continuously differentiable for $1 \le j \ne k \le q$.

(B4) The additive functions f_k are twice continuously (Fréchet) differentiable.

Without loss of generality, assume $\mathcal{X}_k = [0,1]$ for all k and let $\mathcal{I}_k = [2h_k, 1-2h_k]$. The moment condition on ε in the following theorem is required to control the effect of the error of $\hat{\mu}$ as an estimator of μ on the discrepancies of $\mathrm{Log}_{\hat{\mu}} Y_i$ from $\mathrm{Log}_{\mu} Y_i$ after parallel transportation; see Lemma S3 of the Supplementary Material. It is a mild requirement and is satisfied for example when the manifold is compact or $\|\varepsilon\|_\mu$ follows a sub-exponential distribution.

THEOREM 1. *Under (A1)–(A5) and (B1)–(B4), if* $\mathbb{E}\|\varepsilon\|_\mu^\alpha < \infty$ *for some* $\alpha \ge 10$ *and* $\mathbb{E}(\|\varepsilon\|_\mu^2 \mid X_j = \cdot)$ *are bounded on* $\mathcal{X}_j$*, respectively, for* $1 \le j \le q$*, it holds that*

$$\max_{1\le k\le q} \int_{\mathcal{I}_k} \|\tau_{\hat{\mu},\mu} \hat{f}_k(x_k) - f_k(x_k)\|_\mu^2 \, p_k(x_k) \mathrm{d}x_k = O_P(n^{-4/5}),$$
$$\max_{1\le k\le q} \int_{\mathcal{X}_k} \|\tau_{\hat{\mu},\mu} \hat{f}_k(x_k) - f_k(x_k)\|_\mu^2 \, p_k(x_k) \mathrm{d}x_k = O_P(n^{-3/5}),$$

where $\tau_{\hat{\mu},\mu}$ *is the parallel transport operator along geodesics.*

The following corollary is an immediate consequence.

COROLLARY 1. *Under the conditions of Theorem* 1*, if* $\mathcal{L}$ *is* $\mathcal{S}_m^+$ *endowed with either the Log-Cholesky metric or the Log-Euclidean metric,*

$$\max_{1\le k\le q} \int_{\mathcal{I}_k} \|\mathfrak{log}\hat{w}_k(x_k) - \mathfrak{log} w_k(x_k)\|_e^2 \, p_k(x_k) \mathrm{d}x_k = O_P(n^{-4/5}),$$
$$\max_{1\le k\le q} \int_{\mathcal{X}_k} \|\mathfrak{log}\hat{w}_k(x_k) - \mathfrak{log} w_k(x_k)\|_e^2 \, p_k(x_k) \mathrm{d}x_k = O_P(n^{-3/5}).$$

To derive the asymptotic distribution of $\hat{f}_k$, we define $\mathscr{C}_k(x) = \mathbb{E}\{\varepsilon \otimes \varepsilon \mid X_k = x\}$, where $u \otimes v : T_\mu\mathcal{M} \to T_\mu\mathcal{M}$ is a tensor product operator such that $(u \otimes v)z = g_\mu(u, z)v$. Define

$$\Sigma_k(x) = \alpha_k^{-1} p_k(x)^{-1} \int K(u)^2 \mathrm{d}u \cdot \mathscr{C}_k(x), \tag{6}$$

$$\delta_k(x) = \frac{p_k'(x)}{p_k(x)} \int u^2 K(u) \, \mathrm{d}u \cdot f_k'(x), \tag{7}$$

$$\delta_{jk}(x, v) = \frac{\partial p_{jk}(x, v)}{\partial v} \frac{1}{p_{jk}(x, v)} \int u^2 K(u) \, \mathrm{d}u \cdot f_k'(v), \tag{8}$$

$$\tilde{\Delta}_k(x) = \alpha_k^2 \cdot \delta_k(x) + \sum_{j:j\ne k} \alpha_j^2 \int_{\mathcal{X}_j} \frac{p_{kj}(x, u)}{p_k(x)} \cdot \delta_{kj}(x, u) \, \mathrm{d}u, \tag{9}$$

where α_k are the constants in the condition (B2). Let $(\Delta_1, \ldots, \Delta_q)$ be a solution of the system of equations

$$\Delta_k(x) = \tilde{\Delta}_k(x) - \sum_{j:j\ne k} \int_{\mathcal{X}_j} \frac{p_{kj}(x, u)}{p_k(x)} \cdot \Delta_j(u) \mathrm{d}u, \quad 1 \le k \le q, \tag{10}$$

satisfying the constraints

$$\int_{\mathcal{X}_k} p_k(x) \cdot \Delta_k(x) \mathrm{d}x = \alpha_k^2 \cdot \int_{\mathcal{X}_k} p_k(x) \cdot \delta_k(x) \mathrm{d}x, \quad 1 \le k \le q. \tag{11}$$

Finally, define $c_k(x) = \frac{1}{2}\int u^2 K(u)\,\mathrm{d}u \cdot f_k''(x)$ and $\theta_k(x) = \alpha_k^2 \cdot c_k(x) + \Delta_k(x)$.

We assume that

(B5) $\mathbb{E}\{\varepsilon \otimes \varepsilon \mid X_k = \cdot\}$ are continuous operators on $\mathcal{X}_k$ for all $1 \le k \le q$ and operators $\mathbb{E}\{\varepsilon \otimes \varepsilon \mid X_j = \cdot, X_k = \cdot\}$ are bounded on $\mathcal{X}_j \times \mathcal{X}_k$ for all $1 \le j \ne k \le q$.

(B6) $\partial p/\partial x_k$, $k = 1, \ldots, q$, exist and are bounded on $\mathcal{X} = \prod_{k=1}^q \mathcal{X}_k$.

Note that condition (B5) is superfluous if the random noise ε is independent of the predictors $X_1, \ldots, X_q$. Let $N_\mu(\boldsymbol{x})$ be the product measure $N(\theta_1(x_1), \Sigma_1(x_1)) \times \cdots \times N(\theta_q(x_q), \Sigma_q(x_q))$ on $(T_\mu\mathcal{L})^q$, where $N(\theta, \Sigma)$ denotes a Gaussian measure on $T_\mu\mathcal{L}$ with the mean vector θ and covariance operator Σ. For a set A, let $\mathrm{Int}(A)$ denote the interior of A.

THEOREM 2. *Assume that conditions (A1)–(A5) and (B1)–(B6) hold, that* $\mathbb{E}\|\varepsilon\|_\mu^\alpha < \infty$ *for some* $\alpha > 10$ *and that there exists* $\alpha' > 5/2$ *such that* $\mathbb{E}(\|\varepsilon\|_\mu^{\alpha'} \mid X_k = \cdot)$ *are bounded on* $\mathcal{X}_k$ *for all* $1 \le k \le q$. *Then, for* $\boldsymbol{x} = (x_1, \ldots, x_q) \in \mathrm{Int}(\mathcal{X})$, *it holds that* $\left[n^{2/5}\{\tau_{\hat{\mu},\mu}\hat{f}_k(x_k) - f_k(x_k)\} : 1 \le k \le q\right] \to N_\mu(\boldsymbol{x})$ *in distribution. In addition,* $n^{2/5}\left\{\sum_{k=1}^q \tau_{\hat{\mu},\mu}\hat{f}_k(x_k) - \sum_{k=1}^q f_k(x_k)\right\}$ *converges to* $N_\mu(\theta(\boldsymbol{x}), \Sigma(\boldsymbol{x}))$, *where* $\theta(\boldsymbol{x}) = \sum_{k=1}^q \theta_k(x_k)$ *and* $\Sigma(\boldsymbol{x}) = \Sigma_1(x_1) + \cdots + \Sigma_q(x_q)$.

When $\mathcal{L}$ is $\mathcal{S}_m^+$ equipped with either the Log-Cholesky metric or the Log-Euclidean metric, the above asymptotic normality can be formulated on the Lie algebra $\mathfrak{g}$. To this end, assume that $\Sigma_1^{\mathrm{SPD}}, \ldots, \Sigma_q^{\mathrm{SPD}}$ and $\Delta_1^{\mathrm{SPD}}, \ldots, \Delta_q^{\mathrm{SPD}}$ are defined by equations (6)–(11) with $\mathscr{C}_k(x)$ and f_k replaced by $\mathbb{E}\{\mathfrak{log}\zeta \otimes \mathfrak{log}\zeta \mid X_k = x\}$ and $\psi_k := \mathfrak{log} w_k$, respectively. Also, let $c_k^{\mathrm{SPD}} = \frac{1}{2}\int u^2 K(u)\,\mathrm{d}u \cdot \psi_k''(x)$ and $\theta_k^{\mathrm{SPD}}(x) = \alpha_k^2 \cdot c_k^{\mathrm{SPD}}(x) + \Delta_k^{\mathrm{SPD}}(x)$, for $k = 1, \ldots, q$. The following corollary is an immediate consequence of Theorem 2, by noting that the manifold $\mathcal{S}_m^+$ when equipped with the Log-Cholesky metric or the Log-Euclidean metric satisfies the conditions (A1)–(A4) when the second moment of the random noise ζ is finite.

COROLLARY 2. *Assume that the conditions (B1)–(B6) hold and that* $\mathbb{E}\|\mathfrak{log}\zeta\|_\mu^\alpha < \infty$ *for some* $\alpha > 10$. *Furthermore, assume that there exists* $\alpha' > 5/2$ *such that* $\mathbb{E}(\|\mathfrak{log}\zeta\|_e^{\alpha'} \mid X_k = \cdot)$ *are bounded on* $\mathcal{X}_k$ *for all* $1 \le k \le q$. *For* $\mathcal{S}_m^+$ *endowed with either the Log-Cholesky metric or the Log-Euclidean metric, for* $\boldsymbol{x} = (x_1, \ldots, x_q) \in \mathrm{Int}(\mathcal{X})$, *it holds that* $\left[n^{2/5}\{\mathfrak{log}\hat{w}_k(x_k) - \mathfrak{log} w_k(x_k)\} : 1 \le k \le q\right] \to N_{I_m}(\boldsymbol{x})$ *in distribution. In addition,* $n^{2/5}\left\{\sum_{k=1}^q \mathfrak{log}\hat{w}_k(x_k) - \sum_{k=1}^q \mathfrak{log} w_k(x_k)\right\}$ *converges to* $N_{I_m}\{\theta(\boldsymbol{x}), \Sigma(\boldsymbol{x})\}$, *where* I_m *is the* $m \times m$ *identity matrix,* $\theta(\boldsymbol{x}) = \sum_{k=1}^q \theta_k^{\mathrm{SPD}}(x_k)$ *and* $\Sigma(\boldsymbol{x}) = \Sigma_1^{\mathrm{SPD}}(x_q) + \cdots + \Sigma_q^{\mathrm{SPD}}(x_q)$.

These results generalize the work of Jeon & Park (2020) One of the major technical challenges that is addressed in Lemmas S1 and S2 in the Supplementary Material is to uniformly quantify the discrepancy between $\mathrm{Log}_{\hat{\mu}} Y_i$ and $\mathrm{Log}_\mu Y_i$ due to $\hat{\mu} \ne \mu$, as the asymptotic behavior of this discrepancy plays an important role in the theoretical analysis.

6. SIMULATIONS

To illustrate the numerical performance of the proposed manifold additive model estimators, we conducted simulations for $\mathcal{L} = \mathcal{S}_m^+$ endowed with the Log-Cholesky and Log-Euclidean metrics, respectively. We consider two matrix dimensions, $m = 3$ and $m = 10$. We set $\mathcal{X}_k = [0, 1]$ for $k = 1, \ldots, q$. The predictors $X_1, \ldots, X_q$ are set to $X_1 = B_{2/3,2/3}(T_1)$ and $X_k = \Phi(T_k)$ for $k = 2, \ldots, q$, where $B_{2/3,2/3}$ denotes the cumulative distribution function of the beta distribution with both shape parameters equal to $2/3$, Φ denotes the cumulative distribution function of

Table 1. *Prediction RMSE and its Monte Carlo standard error with $m = 3$ (Log-Cholesky)*

Setting	q	n	MAM		ILPR		CHOL	
			SNR=2	SNR=4	SNR=2	SNR=4	SNR=2	SNR=4
I	3	50	1.046 (0.093)	0.967 (0.092)	1.604 (0.105)	1.546 (0.233)	1.376 (0.112)	1.329 (0.120)
		100	0.695 (0.049)	0.624 (0.040)	1.528 (0.127)	1.452 (0.218)	1.248 (0.108)	1.171 (0.123)
		200	0.541 (0.040)	0.412 (0.038)	1.392 (0.110)	1.320 (0.151)	1.200 (0.109)	1.187 (0.117)
	4	50	1.665 (0.082)	1.458 (0.154)	1.865 (0.065)	1.871 (0.170)	1.931 (0.096)	1.890 (0.074)
		100	1.087 (0.049)	0.965 (0.065)	1.786 (0.033)	1.759 (0.028)	1.829 (0.134)	1.678 (0.045)
		200	0.754 (0.073)	0.608 (0.050)	1.743 (0.029)	1.726 (0.033)	1.683 (0.087)	1.616 (0.050)
II	3	50	1.121 (0.045)	1.073 (0.063)	1.399 (0.161)	1.195 (0.128)	1.186 (0.064)	1.156 (0.052)
		100	0.910 (0.045)	0.822 (0.037)	1.191 (0.131)	1.125 (0.103)	1.016 (0.085)	0.915 (0.055)
		200	0.774 (0.042)	0.711 (0.025)	1.151 (0.122)	1.093 (0.126)	0.861 (0.043)	0.782 (0.026)
	4	50	1.495 (0.037)	1.463 (0.049)	1.601 (0.075)	1.644 (0.117)	1.654 (0.076)	1.582 (0.070)
		100	1.203 (0.064)	1.117 (0.099)	1.521 (0.025)	1.507 (0.021)	1.569 (0.094)	1.451 (0.062)
		200	0.916 (0.051)	0.817 (0.043)	1.481 (0.022)	1.469 (0.020)	1.396 (0.103)	1.350 (0.103)
III	3	50	0.640 (0.054)	0.629 (0.055)	0.713 (0.107)	0.696 (0.101)	0.593 (0.053)	0.609 (0.050)
		100	0.574 (0.041)	0.555 (0.050)	0.678 (0.051)	0.616 (0.077)	0.559 (0.042)	0.547 (0.063)
		200	0.530 (0.043)	0.515 (0.045)	0.637 (0.084)	0.589 (0.067)	0.535 (0.037)	0.526 (0.051)
	4	50	0.603 (0.059)	0.551 (0.036)	0.647 (0.068)	0.624 (0.097)	0.602 (0.055)	0.597 (0.036)
		100	0.584 (0.048)	0.549 (0.046)	0.616 (0.092)	0.617 (0.054)	0.573 (0.041)	0.576 (0.045)
		200	0.540 (0.063)	0.521 (0.039)	0.586 (0.058)	0.583 (0.037)	0.584 (0.061)	0.566 (0.038)

the standard Gaussian distribution, and $(T_1, \ldots, T_q)$ is sampled from the centered q-dimensional Gaussian distribution with the covariance matrix whose (j,k)-entry is 1 if $j = k$ and $1/5$ if $j \neq k$. Consequently, $X_1, \ldots, X_q$ are correlated and X_1 has a non-uniform distribution on $[0,1]$. The mean μ is an $m \times m$ matrix whose (j,k)-entry is $2^{-|j-k|}$. We then generate the response variable Y by $Y = \mu \oplus w(X_1, \ldots, X_q) \oplus \zeta$, where $w(X_1, \ldots, X_q) = \mathfrak{exp}\tau_{\mu,e} f(X_1, \ldots, X_q)$ with three settings for f:

I. $f(x_1, \ldots, x_q) = \sum_{k=1}^{q} f_k(x_k)$ with $f_k(x_k)$ being an $m \times m$ matrix whose (j,l)-entry is $g(x_k; j, l, q) = \exp(-|j-l|/q)\sin(2q\pi(x_k - (j+l)/q))$;
II. $f(x_1, \ldots, x_q) = f_{12}(x_1, x_2) + \sum_{k=3}^{q} f_k(x_k)$, where f_k is defined as in setting I, while $f_{12}(x_1, x_2)$ is an $m \times m$ matrix whose (j,l)-entry is $g(x_1; j, l, q)g(x_2, j, l, q)$;
III. $f(x_1, \ldots, x_q) = f_{12}(x_1, x_2) \prod_{k=3}^{q} f_k(x_k)$, where $f_{12}(x_1, x_2)$ is an $m \times m$ matrix whose (j,l)-entry is $\exp\{-(j+l)(x_1+x_2)\}/3$, and $f_k(x_k)$ is an $m \times m$ matrix whose (j,l)-entry is $\sin(2\pi x_k)$.

The random noise ζ is generated according to $\mathfrak{log}\zeta = \sum_{j=1}^{p} Z_j v_j$, where $p = m(m+1)/2$ is the dimension of $T_e\mathcal{S}_m^+$, $Z_1, \ldots, Z_p$ are independently sampled from $N(0, \sigma^2)$, and $v_1, \ldots, v_p$ form an orthonormal basis of the tangent space $T_e\mathcal{S}_m^+$. The signal-to-ratio (SNR) is measured by $\text{SNR} = \mathbb{E}\|\mathfrak{log}w(X_1, \ldots, X_q)\|_e^2 / \mathbb{E}\|\mathfrak{log}\zeta\|_e^2$. We tweak the value of the parameter σ^2 to cover two settings for the SNR, namely, $\text{SNR} = 2$ and $\text{SNR} = 4$. We note that the model for f in I is an additive model, while models II and III are not additive. In particular, model III has no additive components and thus represents the most challenging scenario for the proposed additive regression. We consider $q = 3$ and $q = 4$ to probe the effect of the dimensionality of the predictor vector and study sample sizes $n = 50, 100, 200$.

The quality of the estimation is measured by the root mean squared error

$$\text{RMSE} = \left[\int_{[0,1]^q} d^2\{\hat{\mu} \oplus \hat{w}_1(x_1) \oplus \cdots \oplus \hat{w}_q(x_q), f(x_1, \ldots, x_q)\} \mathrm{d}x_1 \cdots \mathrm{d}x_q \right]^{1/2}.$$

Table 2. *Prediction RMSE and its Monte Carlo standard error with $m = 3$ (Log-Euclidean)*

Setting	q	n	MAM		ILPR		CHOL	
			SNR=2	SNR=4	SNR=2	SNR=4	SNR=2	SNR=4
I	3	50	2.452 (0.258)	2.122 (0.218)	3.450 (0.396)	3.121 (0.351)	3.441 (0.323)	3.129 (0.301)
		100	1.576 (0.094)	1.305 (0.127)	3.135 (0.240)	3.120 (0.217)	3.034 (0.218)	2.890 (0.260)
		200	1.217 (0.067)	0.914 (0.060)	3.025 (0.109)	2.945 (0.190)	2.922 (0.172)	2.823 (0.223)
	4	50	3.573 (0.182)	3.355 (0.350)	3.949 (0.205)	3.879 (0.132)	4.478 (0.299)	4.388 (0.223)
		100	2.436 (0.201)	2.079 (0.163)	3.712 (0.196)	3.680 (0.184)	4.173 (0.224)	4.059 (0.276)
		200	1.802 (0.176)	1.333 (0.120)	3.621 (0.152)	3.603 (0.137)	4.015 (0.231)	3.701 (0.274)
II	3	50	2.418 (0.107)	2.239 (0.120)	2.736 (0.259)	2.517 (0.126)	2.756 (0.175)	2.467 (0.112)
		100	1.960 (0.149)	1.750 (0.077)	2.493 (0.290)	2.480 (0.301)	2.491 (0.183)	2.140 (0.155)
		200	1.665 (0.083)	1.563 (0.051)	2.440 (0.311)	2.411 (0.223)	1.954 (0.112)	1.782 (0.062)
	4	50	3.292 (0.341)	3.150 (0.254)	3.396 (0.266)	3.412 (0.276)	3.693 (0.245)	3.654 (0.289)
		100	2.769 (0.286)	2.442 (0.108)	3.215 (0.286)	3.201 (0.251)	3.475 (0.250)	3.424 (0.189)
		200	2.151 (0.102)	1.872 (0.073)	3.149 (0.230)	3.130 (0.235)	3.289 (0.279)	3.207 (0.308)
III	3	50	0.850 (0.030)	0.937 (0.138)	1.093 (0.080)	0.922 (0.063)	0.875 (0.042)	0.833 (0.059)
		100	0.801 (0.050)	0.762 (0.042)	0.895 (0.055)	0.813 (0.074)	0.875 (0.165)	0.823 (0.060)
		200	0.706 (0.048)	0.674 (0.063)	0.829 (0.033)	0.801 (0.037)	0.773 (0.067)	0.753 (0.074)
	4	50	0.853 (0.066)	0.832 (0.058)	1.002 (0.074)	0.902 (0.062)	0.889 (0.077)	0.872 (0.096)
		100	0.837 (0.044)	0.832 (0.039)	0.890 (0.066)	0.845 (0.071)	0.871 (0.058)	0.855 (0.057)
		200	0.815 (0.059)	0.802 (0.032)	0.844 (0.056)	0.810 (0.052)	0.860 (0.088)	0.839 (0.043)

As a comparison method for the proposed manifold additive model (MAM), we also implemented the intrinsic local polynomial regression (ILPR) proposed in Yuan et al. (2012), which is a fully nonparametric approach. In addition, we implemented the following baseline method proposed by an anonymous reviewer, which provides a simple and straightforward approach based on the Cholesky decomposition (CHOL): Each symmetric positive-definite matrix is represented by its Cholesky factor and then a standard multivariate additive model is applied for the Cholesky factor. Note that this simple method may be subject to the swelling effect as it does not use a swelling-free geometry on $\mathcal{S}_m^+$; see Example 1 of Lin (2019) for an illustration. Each simulation setting was repeated 100 times. The Monte Carlo prediction RMSE and its standard error are shown in Table 1 for the Log-Cholesky metric and in Table 2 for the Log-Euclidean metric for $m = 3$; the results for $m = 10$ are similar and can be found in Tables S1 and S2 of the Supplementary Material, where we also graphically illustrate the estimation quality of the proposed method for SNR $= 4$, $q = 3$, $m = 3$ and the Log-Euclidean metric in Figure S1.

When the model is correctly specified as in Setting I, the proposed model outperforms ILPR and CHOL by a significant margin. When the underlying model is not fully additive but contains some additive components, such as the model in Setting II, the MAM approach still outperforms the other two methods. When the true model has no additive components, such as the model in Setting III, all methods have comparable performance when data are sampled from the Log-Cholesky metric, while MAM and CHOL exhibit advantages for the Log-Euclidean metric. Also, MAM enjoys smaller prediction RMSE, compared to the baseline method CHOL in Settings I and II, while both have similar prediction RMSE in Setting III. As the true model is unknown in reality and MAM is competitive even when there are no additive components, MAM is overall preferrable in applications. In terms of computational efficiency, we found that the Log-Cholesky metric is computationally faster than the Log-Euclidean metric. In addition, the proposed method for the Log-Cholesky metric and the baseline method have comparable computational efficiency, while ILPR is fastest in computation; see Table S3 of the Supplementary Material.

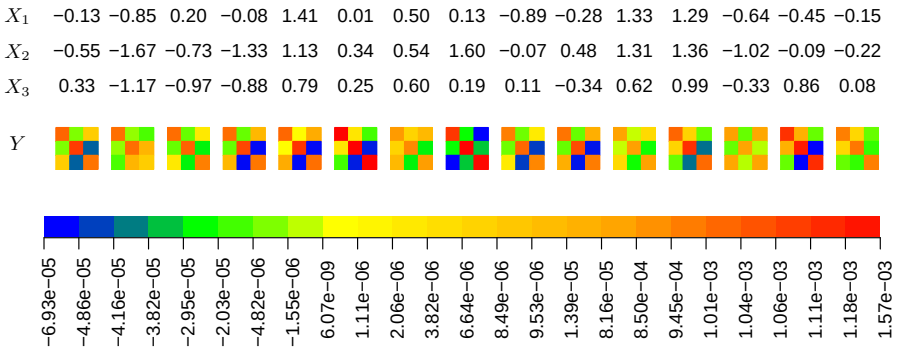

Fig. 1. A random sample of 15 observations from the data.

7. APPLICATION TO DIFFUSION TENSOR IMAGING

We applied the proposed additive model for diffusion tensors obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) at `http://adni.loni.usc.edu/` and `www.adni-info.org`. Diffusion tensors are 3×3 symmetric positive-definite matrices that characterize diffusion of water molecules in tissues and convey rich information about brain tissues with important applications in tractography. They are utilized to to aid in the diagnosis of brain related diseases. In statistical modeling, diffusion tensors are typically considered to be random elements in the space $\mathcal{S}_3^+$ (Fillard et al., 2005; Arsigny et al., 2006; Lenglet et al., 2006; Pennec, 2006; Zhou et al., 2016; Fletcher & Joshib, 2007; Dryden et al., 2009; Zhu et al., 2009; Pennec, 2020). A traditional Euclidean framework for diffusion tensors suffers from significant swelling effects that undesirably inflate the diffusion tensors (Arsigny et al., 2007) and impede their interpretation. In our analysis we use diffusion tensors as responses under the Log-Euclidean metric, which is designed to eliminate the swelling effect and relate these responses to several covariates.

We focus on the hippocampus, which plays a central role in Alzheimer's disease (Lindberg et al., 2012). In the ADNI study, brain images and assessment of memory, executive functioning and language ability were obtained for participating subjects. For each raw diffusion tensor image, a standard preprocessing protocol including denoising, eddy current and motion correction, skull stripping, bias correction and normalization was applied and then diffusion tensors for each hippocampal voxel were extracted, followed by computing their Log-Euclidean mean. For each raw image this resulted in an average diffusion tensor representing the typical hippocampal diffusion of the corresponding subject at the time of visit. To study the relation between the average hippocampal diffusion tensor and memory, executive functioning and language ability of the subject, we utilized the neuropsychological summary scores available from ADNI (Gibbons et al., 2012). We only included data from the first visit of subjects who were diagnosed as having either early mild cognitive impairment, mild cognitive impairment, late mild cognitive impairment or Alzheimer's disease and excluded records with missing values. This resulted in 220 data tuples of the form (Y, X_1, X_2, X_3), where Y is the average diffusion tensor, which serves as response, while the predictors X_1, X_2, X_3 are standardized scores for memory, executive functioning and language ability, respectively. A subset of the data is presented in Figure 1.

The estimated component functions $\hat{w}_1(x_1), \hat{w}_2(x_2), \hat{w}_3(x_3)$ and their individual effect on the diffusion tensors are depicted in Figure 2. For interpretation, we denote the coordinate system of $\mathbb{R}^3$ that was adopted to record the diffusion tensors by $\{e_1, e_2, e_3\}$, so that the matrices pre-

sented in Figures 1 and 2 indicate the coefficients of the corresponding diffusion tensors in this coordinate system. We find that the component functions have distinct effects on the outcome.

Considering for example how $\hat{w}_3$, which encodes the fit for language ability as predictor and acts on the average diffusion tensor $\hat{\mu}$ by the group operation $\oplus$, affects the diffusion along the directions e_1 and e_2 we find that they become increasingly negatively correlated as the standardized language ability drops, while the opposite happens for e_2 and e_3, indicating nonlinear changes in the diffusion that are associated with changing language ability. As a second example consider $\hat{w}_1$, which encodes the effect of memory. A negative correlation between e_2 and e_3 at high memory levels changes into a positive correlation as memory levels drop in a nonlinear fashion, while the correlation between e_1 and e_2 oscillates between mildly positive and mildly negative, suggesting distinctive relationships between diffusion patterns and performance scores.

This leads to the question whether the covariates X_k, for $k = 1, 2, 3$, are significantly related to the spatially averaged hippocampal diffusion tensor and motivates testing the global null hypothesis

$$H_0 : w_k(x_k) = I_3 \qquad \text{for all } x_k, \tag{12}$$

where I_3 denotes the 3×3 identity matrix that is also the group identity of the Lie group $(\mathcal{S}_m^+, \oplus)$ in the Log-Euclidean framework, that is, $P \oplus I_3 = P$ for all $P \in \mathcal{S}_3^+$. Corollary 2 can be employed for testing this hypothesis: For a set $\mathbb{H}_k$ of values of X_1, a test based on the asymptotic normality of Corollary 2 can be implemented to obtain the p-value for testing the local null hypothesis $H_0 : w_k(x_k) = I_3$ for each $x_k \in \mathbb{H}_k$, followed by adjustment for multiple comparisons, e.g. by the Benjamini–Hochberg method. Here, a natural choice of $\mathbb{H}_k$ is the set of the observed values for X_k in the data. The global null hypothesis (12) is then rejected if at least one adjusted p-value is less the nominal level α. Implementing this apporach and applying it to the ADNI data leads to rejecting the null hypothesis (12) at the level $\alpha = 0.05$ for all $k = 1, 2, 3$, with the minimal corrected p-values 1.028×10^{-9}, 1.80×10^{-5} and $< 10^{-10}$, respectively. This suggests that there are indeed associations between the spatially averaged hippocampal diffusion tensor and memory, executive functioning and language ability.

8. Discussion

There are at least three potential extensions of the proposed methods and theory. First, in the data application we consider the spatially averaged hippocampal diffusion tensor. Although significantly reducing the data noise level, the average may conceal some spatial structure of interest within the hippocampus. One way to address this problem is to view all hippocampal diffusion tensors derived from an image as a $\mathcal{S}_3^+$-valued function $Y : s \mapsto Y(s) \in \mathcal{S}_3^+$ with s ranging over all hippocampal voxels. This functional perspective enables one to borrow information from neighboring voxels to counter the high-level data noise. However, it is rather challenging to develop an additive model for $\mathcal{S}_3^+$-valued functions (see Dubey & Müller, 2020).

Second, we focus only on the first visit of each subject, and thus do not utilize all available data. To analyze the data of repeated visits which by default are correlated, the theory needs to be extended to account for such correlation, which seems nontrivial, especially when the number of visits may grow with the sample size. Third, when testing hypothesis (12), we perform multiple local tests by using the pointwise asymptotic normality of Corollary 2 and then make corrections for multiple comparisons. This approach, although sufficient for our data application, will be too conservative in general. An alternative method is to develop an asymptotic normality result for the random process $\{\hat{w}_k(x) : x \in \mathcal{X}_k\}$, which may lead to more powerful one-step tests. These extensions are left for future studies.

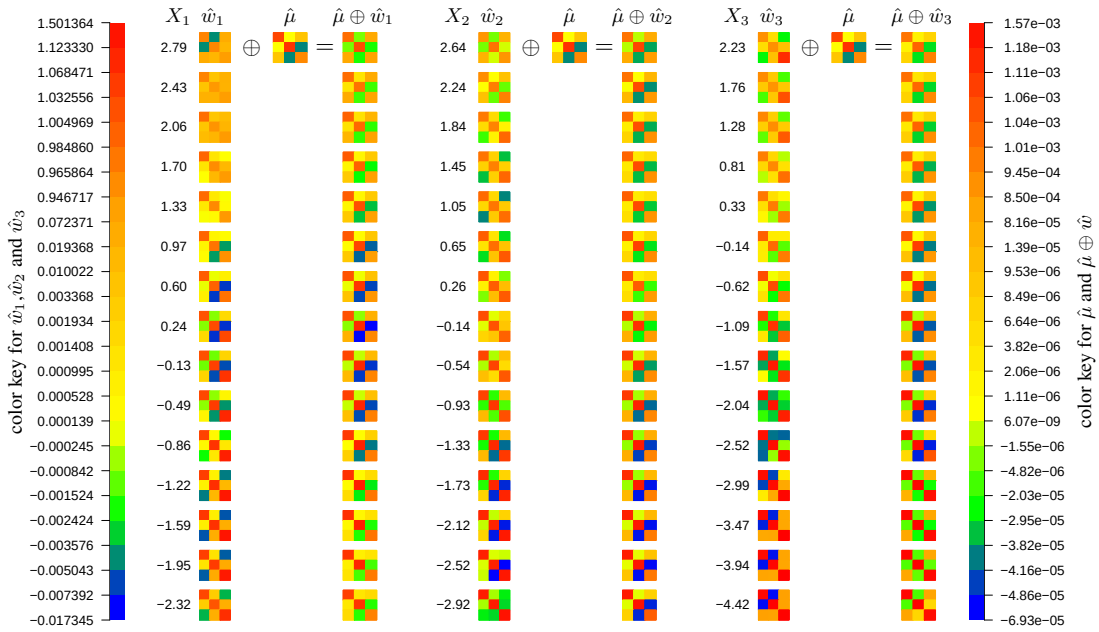

Fig. 2. Estimated additive component functions $\hat{w}_1, \hat{w}_2, \hat{w}_3$ and their effect on the response when other component functions are fixed at identity.

ACKNOWLEDGEMENTS

Zhenhua Lin's Research was partially supported by a NUS startup grant. Hans-Georg Müller's research was supported by a NSF grant. Byeong U. Park's research was supported by Samsung Science and Technology Foundation. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI). A detailed list of funding for ADNI can be found at `https://adni.loni.usc.edu/`.

SUPPLEMENTARY MATERIAL

The Supplementary Material contains additional computational details, auxiliary simulation results and proofs. An `R` implementation of the proposed method is available in the github repository `https://www.github.com/linulysses/mam`.

REFERENCES

AFSARI, B. (2011). Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society* **139**, 655–673.

ARNAUDON, M., BARBARESCO, F. & YANG, L. (2013). Riemannian medians and means with applications to radar signal processing. *IEEE Journal of Selected Topics in Signal Processing* **7**, 595–604.

ARSIGNY, V., FILLARD, P., PENNEC, X. & AYACHE, N. (2006). Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine* **56**, 411–421.

ARSIGNY, V., FILLARD, P., PENNEC, X. & AYACHE, N. (2007). Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal of Matrix Analysis and Applications* **29**, 328–347.

BARMPOUTIS, A., VEMURI, B. C., SHEPHERD, T. M. & FORDER, J. R. (2007). Tensor splines for interpolation and approximation of DT-MRI with applications to segmentation of isolated rat hippocampi. *IEEE transactions on medical imaging* **26**, 1537–1546.

BHATTACHARYA, R. & PATRANGENARU, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *The Annals of Statistics* **31**, 1–29.

BRIDSON, M. R. & HÄFLIGER, A. (1999). *Metric Spaces of Non-Positive Curvature*. Springer-Verlag.

BRÖCKER, T. & TOM DIECK, T. (1985). *Representations of Compact Lie Groups*. Springer.

BRUN, A. (2007). *Manifolds in Image Science and Visualization*. Ph.D. thesis, Linköping University.

CASEIRO, R., HENRIQUES, J. F., MARTINS, P. & BATISTA, J. (2012). A nonparametric Riemannian framework on tensor field with application to foreground segmentation. *Pattern Recognition* **45**, 3997–4017.

CHAU, J. & VON SACHS, R. (2019). Intrinsic wavelet regression for surfaces of Hermitian positive definite matrices. *arXiv:1808.08764 [stat]* ArXiv: 1808.08764.

CHEN, Y. & MÜLLER, H.-G. (2022). Uniform convergence of local Fréchet regression, with applications to locating extrema and time warping for metric-space valued trajectories. *Annals of Statistics* **50**, xxx–xxx.

CORNEA, E., ZHU, H., KIM, P. & IBRAHIM, J. G. (2017). Regression models on Riemannian symmetric spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79**, 463–482.

DAI, X. & MÜLLER, H.-G. (2018). Principal component analysis for functional data on Riemannian manifolds and spheres. *The Annals of Statistics* **46**, 3334–3361.

DAVIS, B. C., FLETCHER, P. T., BULLITT, E. & JOSHI, S. (2010). Population shape regression from random design data. *International Journal of Computer Vision* **90**, 255–266.

DRYDEN, I. L., KOLOYDENKO, A. & ZHOU, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *The Annals of Applied Statistics* **3**, 1102–1123.

DUBEY, P. & MÜLLER, H.-G. (2020). Functional models for time-varying random objects. *Journal of the Royal Statistical Society B (with discussion)* **82**, 275–327.

ELTZNER, B., HUCKEMANN, S. & MARDIA, K. V. (2018). Torus principal component analysis with applications to rna structure. *The Annals of Applied Statistics* **12**, 1332–1359.

FILLARD, P., ARSIGNY, V., AYACHE, N. & PENNEC, X. (2005). A Riemannian framework for the processing of tensor-valued images. In *International Workshop on Deep Structure, Singularities, and Computer Vision*.

FILLARD, P., ARSIGNY, V., PENNEC, X., M.HAYASHI, K., M.THOMPSON, P. & AYACHE, N. (2007). Measuring brain variability by extrapolating sparse tensor fields measured on sulcal lines. *NeuroImage* **34**, 639–650.

FLETCHER, P. T. (2013). Geodesic regression and the theory of least squares on Riemannian manifolds. *International Journal of Computer Vision* **105**, 171–185.

FLETCHER, P. T. & JOSHIB, S. (2007). Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing* **87**, 250–262.

FRISTON, K. J. (2011). Functional and effective connectivity: a review. *Brain Connectivity* **1**, 13–36.

GIBBONS, L. E., CARLE, A. C., MACKIN, R. S., HARVEY, D., MUKHERJEE, S., INSEL, P., CURTIS, S. M., MUNGAS, D. & CRANE, P. K. (2012). A composite score for executive functioning, validated in Alzheimer's Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. *Brain Imaging and Behavior* **6**, 517–527.

HAN, K., MÜLLER, H.-G. & PARK, B. U. (2020). Additive functional regression for densities as responses. *Journal of the American Statistical Association* **115**, 997–1010.

HAN, K. & PARK, B. U. (2018). Smooth backfitting for errors-in-variables additive models. *The Annals of Statistics* **46**, 216–2250.

HEIN, M. (2009). Robust nonparametric regression with metric-space valued output. In *Advances in Neural Information Processing Systems*.

HINKLE, J., FLETCHER, P. T. & JOSHI, S. (2014). Intrinsic polynomials for regression on Riemannian manifolds. *Journal of Mathematical Imaging and Vision* **50**, 32–52.

HUA, X., CHENG, Y., WANG, H., QIN, Y., LI, Y. & ZHANG, W. (2017). Matrix CFAR detectors based on symmetrized Kullback-Leibler and total Kullback-Leibler divergences. *Digital Signal Processing* **69**, 106–116.

HUETTEL, S. A., SONG, A. W. & MCCARTHY, G. (2008). *Functional Magnetic Resonance Imaging*. Sinauer Associates, 2nd ed.

HUNDRIESER, S., ELTZNER, B. & HUCKEMANN, S. F. (2021). Finite sample smeariness of fréchet means and application to climate. *arxiv* .

JEON, J. M. & PARK, B. U. (2020). Additive regression with Hilbertian responses. *The Annals of Statistics* **48**, 2671–2697.

JUNG, S., SCHWARTZMAN, A. & GROISSER, D. (2015). Scaling-rotation distance and interpolation of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications* **36**, 1180–1201.

KENDALL, W. S. & LE, H. (2011). Limit theorems for empirical Fréchet means of independent and non-identically distributed manifold-valued random variables. *Brazilian Journal of Probability and Statistics* **25**, 323–352.

LANG, S. (1999). *Fundamentals of Differential Geometry*. New York: Springer.

LE BIHAN, D. (1991). Molecular diffusion nuclear magnetic resonance imaging. *Magnetic Resonance Quarterly* **7**, 1–30.

LEE, J. M. (2018). *Introduction to Riemannian Manifolds*, vol. 176. Cham: Springer, 2nd ed.

LEE, Y. K., MAMMEN, E. & PARK, B. U. (2010). Backfitting and smooth backfitting for additive quantile models. *The Annals of Statistics* **38**, 2857–2883.

LEE, Y. K., MAMMEN, E. & PARK, B. U. (2012). Flexible generalized varying coefficient regression models. *The Annals of Statistics* **40**, 1906–1933.

LENGLET, C., ROUSSON, M., DERICHE, R. & FAUGERAS, O. (2006). Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing. *Journal of Mathematical Imaging and Vision* **25**, 423–444.

LIN, Z. (2019). Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition. *SIAM Journal on Matrix Analysis and Applications* **40**, 1353–1370.

LIN, Z. & MÜLLER, H.-G. (2021). Total variation regularized Fréchet regression for metric-space valued data. *The Annals of Statistics.* **49**, 3510–3533.

LINDBERG, O., WALTERFANG, M., LOOI, J. C., MALYKHIN, N., ÖSTBERG, P., ZANDBELT, B., STYNER, M., VELAKOULIS, D., ÖRNDAHL, E., CAVALLIN, L. & WAHLUND, L.-O. (2012). Shape analysis of the hippocampus in alzheimer's disease and subtypes of frontotemporal lobar degeneration. *Journal of Alzheimer's Disease* **30**, 355–365.

MAMMEN, E., LINTON, O. & NIELSEN, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *The Annals of Statistics* **27**, 1443–1490.

MOAKHER, M. (2005). A differential geometry approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications* **26**, 735–747.

PARK, B. U., CHEN, C.-J., TAO, W. & MÜLLER, H.-G. (2018). Singular additive models for function to function regression. *Statistica Sinica* **28**, 2497–2520.

PELLETIER, B. (2006). Non-parametric regression estimation on closed Riemannian manifolds. *Journal of Nonparametric Statistics* **18**, 57–67.

PENNEC, X. (2006). Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements,. *Journal of Mathematical Imaging and Vision* **25**, 127–154.

PENNEC, X. (2020). Manifold-valued image processing with SPD matrices. In *Riemannian Geometric Statistics in Medical Image Analysis*. Elsevier, pp. 75–134.

PENNEC, X., FILLARD, P. & AYACHE, N. (2006). A Riemannian framework for tensor computing. *International Journal of Computer Vision* **66**, 41–66.

PETERSEN, A., DEONI, S. & MÜLLER, H.-G. (2019). Fréchet estimation of time-varying covariance matrices from sparse data, with application to the regional co-evolution of myelination in the developing brain. *The Annals of Applied Statistics* **13**, 393–419.

PIGOLI, D., ASTON, J. A., DRYDEN, I. L. & SECCHI, P. (2014). Distances and inference for covariance operators. *Biometrika* **101**, 409–422.

RATHI, Y., TANNENBAUM, A. & MICHAILOVICH, O. (2007). Segmenting images on the tensor manifold. In *Proocedings of Computer Vision and Pattern Recognition*.

SCHEIPL, F., STAICU, A.-M. & GREVEN, S. (2015). Functional additive mixed models. *Journal of Computational and Graphical Statistics* **24**, 477–501.

SHAO, L., LIN, Z. & YAO, F. (2022+). Intrinsic riemannian functional data analysis for sparse longitudinal observations. *The Annals of Statistics* , To appear.

SHI, X., STYNER, M., LIEBERMAN, J., IBRAHIM, J. G., LIN, W. & ZHU, H. (2009). Intrinsic regression models for manifold-valued data. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, vol. 12.

STEINKE, F., HEIN, M. & SCHÖLKOPF, B. (2010). Nonparametric regression between general Riemannian manifolds. *SIAM Journal on Imaging Sciences* **3**, 527–563.

STONE, C. J. (1985). Additive regression and other nonparametric models. *The Annals of Statistics* **13**, 689–705.

STURM, K.-T. (2003). Probability measures on metric spaces of nonpositive curvature. In *Heat kernels and analysis on manifolds, graphs, and metric spaces (Paris, 2002), vol. 338 of Contemporary Mathematics*. Providence, RI: American Mathematical Society, pp. 357–390.

YANG, Y. (2007). Globally convergent optimization algorithms on riemannian manifolds: Uniform framework for unconstrained and constrained optimization. *Journal of Optimization Theory and Applications* **132**, 245–265.

YU, K., PARK, B. U. & MAMMEN, E. (2008). Smooth backfitting in generalized additive models. *The Annals of Statistics* **36**, 228–260.

YUAN, Y., ZHU, H., LIN, W. & MARRON, J. S. (2012). Local polynomial regression for symmetric positive definite matrices. *Journal of Royal Statistical Society: Series B (Statistical Methodology)* **74**, 697–719.

ZHOU, D., DRYDEN, I. L., KOLOYDENKO, A. A., AUDENAERT, K. M. & BAI, L. (2016). Regularisation, interpolation and visualisation of diffusion tensor images using non-Euclidean statistics. *Journal of Applied Statistics* **43**, 943–978.

ZHU, H., CHEN, Y., IBRAHIM, J. G., LI, Y., HALL, C. & LIN, W. (2009). Intrinsic regression models for positive-definite matrices with applications to diffusion tensor imaging. *Journal of the American Statistical Association* **104**, 1203–1212.

[*Received on* x xxx xxxx. *Editorial decision on* x xxx xxxx]