# Supplementary Material to "Intrinsic Riemannian Functional Data Analysis for Sparse Longitudinal Observations"

Lingxuan Shao*, Zhenhua Lin†, and Fang Yao*

## S.1 Additional Background on Riemannian Manifold

Here we provide formal definitions related to Riemannian manifolds, starting with perhaps the most basic geometric space — the topological space.

**Definition S.1** (Topological space)**.** *A topological space is a set and a class of subsets of the set, called the open sets, such that the class contains the empty set and is closed under the formation of arbitrary unions and finite intersections. Such a class is called a topology.*

The most commonly seen topological space is $\mathbb{R}$ together with the standard (but often not explicitly mentioned) topology that contains all open intervals, as well as the $d$-dimensional Euclidean space $\mathbb{R}^d$ together with the standard topology that contains all open balls. In real analysis, continuous functions play an important role. The concept of continuity can be generalized to functions defined on and/or taking values in general topological spaces.

**Definition S.2** (Continuity and homeomorphism)**.** *A function $f : \mathcal{T}_1 \to \mathcal{T}_2$ between two topological spaces is continuous if to every open set $B$ of $\mathcal{T}_2$, the set $f^{-1}(B) := \{x \in \mathcal{T}_1 : f(x) \in B\}$ is an open set of $\mathcal{T}_1$. If $f$ is continuous and bijective and its inverse is also continuous, then we say $f$ is a homeomorphism between $\mathcal{T}_1$ and $\mathcal{T}_2$. Two topological spaces are homeomorphic to each other if there exists a homeomorphism between them.*

When both $\mathcal{T}_1$ and $\mathcal{T}_2$ are $\mathbb{R}$ with the standard topology, the continuity defined in the above coincides with the one via the $\epsilon$-$\delta$ definition. For instance, it is well know that $f(x) = x^3$ is a continuous function, and indeed, to every open interval $(a,b)$, $f^{-1}((a,b)) = (a^{1/3}, b^{1/3})$ is also an open interval, and this holds for all open sets of $\mathbb{R}$; see Example 1 of Section 18 in Munkres (2000). A property of a topological space is a topological property if it is preserved under a homeomorphism. Therefore, homeomorphic topological spaces share the same set of topological properties. A concrete example of topological properties is connectedness that is related to Assumption 4.4(a). A topological space $\mathcal{T}$ is **connected** if it is not the union of two disjoint non-empty open sets. For example, $\mathbb{R}$ is connected, but $\mathbb{R}\backslash\{0\}$ is not. If to every two points $p, q \in \mathcal{T}$ there is a continuous path connecting them, i.e., there exists a continuous function $\gamma : [0,1] \to \mathcal{T}$ such that $\gamma(0) = p$ and $\gamma(1) = q$, then we say $\mathcal{T}$ is **path-connected**. A topological space $\mathcal{T}$ is **simply connected** if and only if it is path-connected, and for any two continuous paths $\gamma_1$ and $\gamma_2$ with the same start and end points, $\gamma_1$ can be continuously deformed into $\gamma_2$, i.e., there exists a continuous function $\gamma : [0,1]^2 \to \mathcal{M}$ such that $\gamma(s,0) = \gamma_1(s)$ and $\gamma(s,1) = \gamma_2(s)$ for $s \in [0,1]$. Intuitively speaking, a topological space is simply

*School of Mathematical Sciences, Center for Statistical Science, Peking University
†Department of Statistics and Data Science, National University of Singapore

connected if there are no holes passing through the space. For instance, $\mathbb{R}^d$ and spheres are simply connected, while the torus is not. Here, connectedness, path-connectedness and simple connectedness are all topological properties, i.e., preserved under homeomorphisms.

There are topological spaces that locally resemble a Euclidean space $\mathbb{R}^d$, but globally may not be homeomorphic to $\mathbb{R}^d$. Such spaces are called topological manifolds. An example of topological manifolds is the surface of our earth that may be roughly parameterized by the two-dimensional sphere $\mathbb{S}_r^2 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = r^2\}$ of radius $r \approx 6371$km. Each small neighborhood of $\mathbb{S}_r^2$ looks like a (subset of) two-dimensional plane, and this is why our ancients perceived the flat earth model. However, there exists no homeomorphism between $\mathbb{S}_r^2$ and $\mathbb{R}^d$ for any $d$.

**Definition S.3** (Topological manifold)**.** *A topological space $\mathcal{T}$ is a topological manifold modeled on $\mathbb{R}^d$, if for each point $p \in \mathcal{T}$ there exists an open set that contains $p$ and is homeomorphic to $\mathbb{R}^d$. Here, $d$ is called the dimension of $\mathcal{T}$.*

In the above example of $\mathbb{S}_r^2$, let $U_1 = \mathbb{S}_r^2 \backslash \{(r, 0, 0)\}$ and $U_2 = \mathbb{S}_r^2 \backslash \{(-r, 0, 0)\}$. Then every point in $\mathbb{S}_r^2$ falls into one of these two open sets. Moreover, both $U_1$ and $U_2$ are homeomorphic to $\mathbb{R}^d$, with the following corresponding homeomorphisms

$$\phi_1(x_1, x_2, x_3) = \frac{r}{r - x_1}(x_2, x_3) \in \mathbb{R}^2$$
$$\phi_2(x_1, x_2, x_3) = \frac{r}{r + x_1}(x_2, x_3) \in \mathbb{R}^2.$$

This formally shows that $\mathbb{S}_r^2$ is a topological manifold of dimension 2.

Due to the local resemblance between $\mathbb{R}^d$ and a topological manifold, one might parameterize a local neighborhood of the topological manifold by using $\mathbb{R}^d$, as we did in the above for neighborhoods $U_1$ and $U_2$ of $\mathbb{S}_r^2$. Intuitively, the map $\phi_1$ (resp. $\phi_2$) assigns each point in $U_1$ (resp. $U_2$) a coordinate in $\mathbb{R}^2$. Since $U_1 \cup U_2 = \mathbb{S}_r^2$, each point gets a coordinate. However, for those points in $U_1 \cap U_2$, such as the points in the equator, are assigned two coordinates, one from $\phi_1$ and the other from $\phi_2$. In this case, one can obtain the coordinate under $\phi_1$ if we know its coordinate under $\phi_2$, and vice versa. For example, for $(x_1, x_2, x_3) \in U_1 \cap U_2$, if its coordinate under $\phi_1$ is $(y_1, z_1) \in \mathbb{R}^2$, then its coordinate under $\phi_2$ is $(y_2, z_2) = (\phi_2 \circ \phi_1^{-1})(y_1, z_1)$, where $\circ$ represents the composition of functions, i.e., $(f \circ g)(x) = f(g(x))$ for generic functions $f$, $g$ and argument $x$. Moreover, $\phi_2 \circ \phi_1^{-1}$, called a transition map, is a continuous function, since both $\phi_1$ and $\phi_2$ are homeomorphisms. Noting that $\phi_2 \circ \phi_1^{-1}$ is a function mapping $\mathbb{R}^2$ onto $\mathbb{R}^2$, we might impose regularity more than continuity on the transition maps, for instance, differentiability or smoothness. This consideration leads to the concept of differentiable and smooth manifolds.

**Definition S.4** (Differentiable and smooth manifolds)**.** *Let $\mathcal{M}$ be a $d$-dimensional topological manifold. For $k \geq 1$, a $C^k$-atlas on $\mathcal{M}$ is a collection of pairs $(U_\alpha, \phi_\alpha)$ that are indexed by an index set $J$ and satisfy the following axioms:*

- *Each $U_\alpha$ is an open subset of $\mathcal{M}$ and $\bigcup_{\alpha \in J} U_\alpha = \mathcal{M}$, i.e., the domains $U_\alpha$ together cover $\mathcal{M}$;*
- *Each $\phi_\alpha$ is a homeomorphism between $U_\alpha$ and the open set $\phi_\alpha(U_\alpha) = \{\phi_\alpha(x) \in \mathbb{R}^d : x \in U_\alpha\}$ of $\mathbb{R}^d$;*
- *For each pair of $\alpha, \beta \in J$, if $U_\alpha \cap U_\beta \neq \varnothing$, then the transition map $\phi_\alpha \circ \phi_\beta^{-1} : \phi_\beta(U_\alpha \cap U_\beta) \to \phi_\alpha(U_\alpha \cap U_\beta)$, illustrated in Figure S.1, is $k$ times differentiable; we say $\phi_\alpha$ and $\phi_\beta$ are compatible.*

*Two $C^k$-atlases are compatible if their union is again a $C^k$-atlas. An atlas is maximal if it contains any other atlas compatible with it. A $C^k$-manifold is a topological manifold together with a maximal $C^k$-atlas. A $C^\infty$-manifold is called a smooth manifold.*

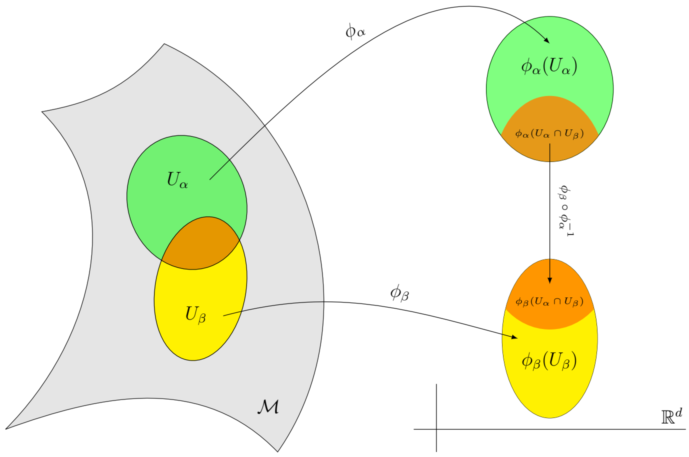


Figure S.1: Illustration of the chart and transition map.

The pair $(U_\alpha, \phi_\alpha)$ or sometimes $\phi_\alpha$ itself is called a **chart** (or **coordinate map**). One can check that compatibility defines an equivalence relation among atlases, and a maximal atlas is simply the union of atlases within the same equivalence class. Therefore, a differentiable manifold is essentially completely determined by an atlas, in the sense that any compatible atlas gives to the same differentiable manifold and incompatible atlases result in distinct differentiable manifolds. In light of this, in practice, we can describe a differentiable manifold simply by providing an atlas. For instance, in the example of $\mathbb{S}^2_r$, the collection $\mathscr{A} = \{(U_1, \phi_1), (U_2, \phi_2)\}$ forms a $C^\infty$-atlas that turns $\mathbb{S}^2_r$ into a smooth manifold.

Let $\mathcal{M}$ be a $d$-dimensional differentiable manifold in the sequel. One merit of the differentiable manifold is that we can discuss regularity of functions taking values in a differentiable manifold. For instance, for an interval $I \subset \mathbb{R}$, the function $\gamma : I \to \mathcal{M}$, which is also called a curve on $\mathcal{M}$, is differentiable at $t \in I$ if the function $\phi \circ \gamma : I \to \mathbb{R}^d$ is differentiable at $t$ for a chart $(U, \phi)$, and thus all charts, such that $\gamma(t) \in U$ in the maximal atlas associated with $\mathcal{M}$. The derivative, denoted by $\gamma'(t)$, measuring the velocity of the curve $\gamma$ at $t$, has different representations in different charts. However, once we know its representation in one chart, we can then obtain its representation in another chart via the transition maps. In this sense, the velocity of $\gamma$ at $t$ is essentially well defined, and is denoted by $\gamma'(t)$. Note that this sense of "well-definedness" applies generally to other concepts and quantities in differential geometry, that is, a manifold-related concept, such as the differentiability of a curve, is well defined if it holds for all relevant charts, and a manifold-related quantity, such the derivative of a differentiable curve at $t$, is well defined if its coordinate in one chart can be determined from the coordinate under another chart.

Consider all differentiable curves $\gamma : I \to \mathcal{M}$ with $\gamma(t) = p$. They may give rise to different derivatives $\gamma'(t)$. If we fix any chart $(U, \phi)$ with $p \in U$, then each $\gamma'(t)$ is represented by $(\phi \circ \gamma)'(t)$ which is a vector in $\mathbb{R}^d$ and all possible values of $(\phi \circ \gamma)'(t)$ form exactly the vector space $\mathbb{R}^d$. Therefore, with $\Gamma^t_p$ denoting the collection of differentiable curves $\gamma : I \to \mathcal{M}$ such that $\gamma(t) = p$, we can view the space $T_p\mathcal{M} := \{\gamma'(t) : \gamma \in \Gamma^t_p\}$ as a vector space with the vector addition $\gamma'_1(t) + \gamma'_2(t)$ defined to be $\gamma'_3(t) \in T_p\mathcal{M}$ such that $(\phi \circ \gamma_1)'(t) + (\phi \circ \gamma_2)'(t) = (\phi \circ \gamma_3)'(t)$, and the scalar multiplication $a\gamma'(t)$ defined to be $\eta'(t)$ such that $a(\phi \circ \gamma)'(t) = a(\phi \circ \eta)'(t)$, where $\gamma, \gamma_1, \gamma_2, \gamma_3, \eta \in \Gamma^t_p$ and $a \in \mathbb{R}$. Note that, the representation of both $\gamma'_1(t) + \gamma'_2(t)$ and $a\gamma'(t)$ in other charts is determined from its representation in $(U, \phi)$, and thus they are well defined manifold-related quantities. The space $T_p\mathcal{M}$ is called the **tangent space** at $p$ and its elements are called **tangent vectors** at $p$. Such (informal) definition also shows that $T_p\mathcal{M}$ depends on $p$ through its dependence on $\Gamma_p^t$. This agrees well with the physical meaning of the tangent vector $\gamma'(t)$ that represents the direction and amount to move if one wants to get to $\gamma(t+\Delta t)$ from $\gamma(t)$ within an infinitesimal amount time $\Delta t$, i.e., $\gamma'(t)$ encodes both the velocity and the base point $p$. In the chart $(U,\phi)$, $\gamma'(t)$ can be represented by $(\phi(p),(\phi\circ\gamma)'(t))$, $\gamma_1'(t)+\gamma_2'(t)$ by $(\phi(p),(\phi\circ\gamma_1)'(t)+(\phi\circ\gamma_2)'(t))$, and $a\gamma'(t)$ by $(\phi(p),a(\phi\circ\gamma)'(t))$, respectively; here note that in the above coordinate representation of the vector addition and scalar multiplication operations, the coordinate $\phi(t)$ for the base point $p$ does not participate in the operations.

The formal definition below of the tangent space and tangent vectors, with the same essence of the above informal discussion, eliminates the dependence on $I$ and $\gamma(t)=p$. We say two differentiable curves $\gamma$ and $\eta$ on a differentiable manifold $\mathcal{M}$ is equivalent at $p\in\mathcal{M}$, denoted by $\gamma\sim_p\eta$, if $\gamma(s)=\eta(t)=p$ and $(\phi\circ\gamma)'(s)=(\phi\circ\eta)'(t)$ for some $s,t\in\mathbb{R}$ and for all $(U,\phi)$ with $p\in U$. One can check that for each fixed $p\in\mathcal{M}$, $\sim_p$ defines an equivalence relation among differentiable curves passing $p$.

**Definition S.5** (Tangent space and tangent vector)**.** *Each class of equivalent differentiable curves at p is a tangent vector at p, and the collection of tangent vectors at p is the tangent space at p.*

The above formal definition seems abstract, but we can perceive tangent vectors and tangent spaces by using the aforementioned essentially equivalent informal definition, as well as the concrete coordinate representations. One thing we shall emphasize is that tangent vectors at distinct points are not directly comparable, since they have distinct base points. This is also evidenced by the abstract definition, as tangent vectors at distinct points are equivalence classes of distinct equivalence relation. For example, a tangent vector at $p$ is an equivalence class under the relation $\sim_p$, while a tangent vector at $q$ is an equivalence class under the relation $\sim_q$. In general, equivalence classes derived from different equivalence relations are not directly comparable.

For some manifolds, the tangent space might be visualized as a hyperplane that is tangent to the manifold. For instance, as depicted in Figure S.2, the tangent space at $p\in\mathbb{S}_r^2$ can be viewed as the two-dimensional affine plane that is tangent to $\mathbb{S}_r^2$ at $p$; such affine plane is regarded as a vector space with $p$ as the origin.

In elementary calculus, a function that maps a subset of a Euclidean space into potentially another Euclidean space is smooth if it is infinite times differentiable. Based on this elementary concept of smoothness, we can define smoothness for functions between smooth manifolds.

**Definition S.6** (Smooth manifold map)**.** *Let $\mathcal{M}$ be a $d_1$-dimensional smooth manifold and $\mathcal{N}$ a $d_2$-dimensional smooth manifold. A function $f:\mathcal{M}\to\mathcal{N}$ is smooth if to each chart $(U,\phi)$ of $\mathcal{M}$ and each chart $(V,\varphi)$ of $\mathcal{N}$ such that $V\cap f(U)\neq\varnothing$, the function $\varphi\circ f\circ\phi^{-1}$, that maps a subset of $\mathbb{R}^{d_1}$ into $\mathbb{R}^{d_2}$, is smooth.*

**Definition S.7** (Diffeomorphism)**.** *A bijective map between two smooth manifolds is a diffeomorphism if it and its inverse are smooth. Two smooth manifolds are diffeomorphic to each other if there exists a diffeomorphism between them.*

Let $\mathcal{M}$ be a $d$-dimensional smooth manifold in the sequel. A **vector field** $V$ on a set $A$ of $\mathcal{M}$ is a function that maps points in $A$ into $T\mathcal{M}:=\cup_{p\in\mathcal{M}}T_p\mathcal{M}$, such that $V(p)\in T_p\mathcal{M}$; here $T\mathcal{M}$ is called the **tangent bundle** of $\mathcal{M}$. Intuitively, at each point in $U$, the vector field $V$ assigns the point with a tangent vector at that point. Recall that for each tangent vector $u\in T_p\mathcal{M}$ and chart $(U,\phi)$ with $p\in U$, $u$ has a coordinate representation $(\phi(p),x_1,\dots,x_d)$, e.g., $(x_1,\dots,x_d)=(\phi\circ\gamma)'(t)$ for some differentiable curve $\gamma$ with $\gamma(t)=p$. This gives rise to a local coordinate representation of a smooth vector field, i.e., for a chart $(U,\phi)$ such that $A\cap U\neq\varnothing$, the coordinate of $V$ at $u$ is the coordinate of the tangent vector $V(u)\in T_p\mathcal{M}$. Let $\Phi_\phi(u)$ denote the coordinate of the tangent vector $u$ in the chart $(U,\phi)$. Then $\Phi\circ V$, restricted to the domain $A \cap U$, is a function mapping $U$ into $\mathbb{R}^{2d}$. We say $V$ is a **smooth vector field** if to each chart $(U, \phi)$, the function $\Phi \circ V$ is a smooth manifold map. An example of (smooth) vector fields is the vector field for the movement of air on Earth that represents the wind speed and direction at each location.

**Definition S.8** (Riemannian manifold)**.** *A Riemannian manifold is a smooth manifold $\mathcal{M}$ endowed with an inner product $\langle\cdot,\cdot\rangle_p$ on $T_p\mathcal{M}$ for each $p \in \mathcal{M}$ such that, for any smooth vector fields $V_1, V_2$ on $\mathcal{M}$, the function $p \mapsto \langle V_1(p), V_2(p)\rangle_p$ is a smooth manifold map defined on $\mathcal{M}$. The inner products $\langle\cdot,\cdot\rangle_p$ are collectively referred to as Riemannian metric or Riemannian metric tensor.*

For a Riemannian manifold, each of its tangent spaces is now an inner product space, and thus along a normed space with the induced norm $\|v\|_p = \langle v, v\rangle_p$ for $v \in T_p\mathcal{M}$. As a vector space, each tangent space is entitled to an independent basis. For Riemannian manifold, since each tangent space is an inner product space, it is also entitled to orthonormal basis. In this context, a **frame** refers to a map that assigns each point of the manifold an independent basis, and when all of such basis are orthonormal, we say the frame is an **orthonormal frame**.

The Riemannian metric also induces a (canonical) **distance** $d_{\mathcal{M}}$ on the Riemannian manifold, as follows. For a smooth curve $\gamma : I \to \mathcal{M}$, the restriction to an interval $[a, b] \subset I$ is referred to as a segment of $\gamma$ and is denoted by $\gamma([a, b])$. The length of the segment $\gamma([a, b])$ is defined by

$$\ell_\gamma(a, b) = \int_a^b \|\gamma'(t)\|_{\gamma(t)} \mathrm{d}t.$$

Such definition can be extended to regular curves that are formed by connecting a finite number of smooth segments, i.e., the length of a piecewisely smooth segment is defined to be the sum of the lengths of its smooth segments. For a connected Riemannian manifold $\mathcal{M}$, we can define the distance between two points $p, q$ by

$$d_{\mathcal{M}}(p, q) = \inf\{\ell_\gamma(a, b) : \gamma(a) = p, \gamma(b) = q, \text{ and } \gamma \text{ is a regular curve}\}.$$

The distance $d_{\mathcal{M}}$ then turns $\mathcal{M}$ also into a metric space. If such metric space is complete, then we say $\mathcal{M}$ is a complete Riemannian manifold. By Hopf–Rinow theorem, on a connected complete Riemannian manifold, two points can be connected by a minimizing geodesic of which the length is exactly the distance of the two points. By definition, a **geodesic** is a curve $\gamma : I \to \mathcal{M}$ such that for all sufficiently small $\epsilon > 0$ and all interior $t \in I$, $\gamma([t, t + \epsilon])$ is a shortest path that connects $\gamma(t)$ and $\gamma(t + \epsilon)$, and $\ell_\gamma(s, t) = a|s - t|$ for a constant $a$ and all $s, t \in I$. In words, geodesics are constant-speed curves that are locally shortest segments. Note that a geodesic may not be globally a shortest path connecting two points. For example, $\gamma(t) = (0, \cos(2\pi t), \sin(2\pi t))$ for $t \in [0, 1]$ is a geodesic, but it is not a shortest path from $\gamma(\epsilon)$ to $\gamma(1)$ for any $\epsilon \in [0, 1/2)$. A minimizing geodesic refers to those geodesics $\gamma : I \to \mathcal{M}$ such that $\gamma([s, t])$ is a shortest path connecting $\gamma(s)$ and $\gamma(t)$ for all $s, t \in I$.

**Remark S.1.** *In some textbooks, the term "curve" (and similarly, "geodesic") is sometimes used to denote the image of $\gamma : I \to \mathcal{M}$ on the manifold $\mathcal{M}$, rather than the map $\gamma$. In this paper, we do not adopt this practice. When we say a curve, we always refers to the map $\gamma$ itself. One shall note that, the length of a curve is invariant to parameterization, i.e., if $\eta : J \to \mathcal{M}$ is another curve such that $t = g(s)$ for a smooth function with $g'(s) \neq 0$ and $\eta(s) = \gamma(g(s))$ for all $s \in J$ and $I = g(J)$, then $\ell_\gamma(g(a), g(b)) = \ell_\eta(a, b)$ for all $a, b \in J$. This is because, $\ell_\gamma(g(a), g(b)) = \int_{g(a)}^{g(b)} \|\gamma'(t)\|_{\gamma(t)}\mathrm{d}t = \int_a^b \|\gamma'(g(s))\|_{\gamma(g(s))}\mathrm{d}g(s) = \int_a^b \|\eta'(s)\|_{\eta(s)}\mathrm{d}s = \ell_\eta(a, b)$.*

Below we assume $\mathcal{M}$ is complete and connected. The Riemannian metric also induces the Riemannian exponential map for each point. For a unit tangent vector $u \in T_p\mathcal{M}$, let $\gamma_u(t)$ be the geodesic such that

$\gamma_u(0) = p$ and $\gamma'_u(0) = u$. As a starting point specified by $p$ and an initial direction specified by the unit tangent vector $u$ together uniquely determine a geodesic, the celebrated Hopf–Rinow theorem asserts that the following map $\mathrm{Exp}_p$ is well defined at each $p \in \mathcal{M}$ on the entire tangent space $T_p\mathcal{M}$.

**Definition S.9** (Exponential map)**.** *The map* $\mathrm{Exp}_p(u) = \gamma_{\frac{u}{\|u\|_p}}(\|u\|_p)$ *for* $u \in T_p\mathcal{M}$ *is called the exponential map at* $p$*.*

**Remark S.2.** *Exponential maps might be defined also for incomplete Riemannian manifolds, but potentially only locally, i.e.,* $\mathrm{Exp}_p u$ *may be only defined for tangent vectors* $u \in T_p\mathcal{M}$ *in a neighborhood of the zero tangent vector* $0 \in T_p\mathcal{M}$*.*

For $p \in \mathcal{M}$ and $u \in T_p\mathcal{M}$, the curve $\gamma_u(t) = \mathrm{Exp}_p(tu)$ is a geodesic with speed $\|u\|_p$. The **cut time** $\mathfrak{c}(p,u)$ is defined to be $t \in \mathbb{R}_+$ such that $\gamma_u([0, t-\epsilon])$ is a minimizing geodesic for any $\epsilon > 0$ but $\gamma_u([0,t])$ is not, i.e., $\mathfrak{c}(p,u) = \sup\{t \in \mathbb{R}_+ : \gamma_u([0,t])$ is a minimizing geodesic with $\gamma_u(t) = \mathrm{Exp}_p(tu)\}$. Let $\mathcal{E}_p = \{tu : u \in T_p\mathcal{M}, \|u\|_p = 1, 0 \le t < \mathfrak{c}(p,u)\}$, which is a neighborhood of the zero tangent vector in $T_p\mathcal{M}$, and define $\mathcal{D}_p = \{\mathrm{Exp}_p u : u \in \mathcal{E}_p\}$. Then $\mathrm{Exp}_p$ is bijective between $\mathcal{E}_p$ and $\mathcal{D}_p$ and thus its inverse exists on $\mathcal{D}_p$.

**Definition S.10** (Riemannian logarithmic map)**.** *The inverse of* $\mathrm{Exp}_p$ *on* $\mathcal{D}_p$*, denoted by* $\mathrm{Log}_p$*, is called the Riemannian logarithmic map at* $p$*.*

**Remark S.3.** *Generally* $\mathrm{Log}_p$ *is defined in a neighborhood at* $p$*. When* $\mathfrak{c}(p,u) = \infty$ *for all* $u$*, then* $\mathcal{D}_p = \mathcal{M}$ *for a complete and connected manifold* $\mathcal{M}$*, and in this case,* $\mathrm{Log}_p$ *is defined on the entire* $\mathcal{M}$*.*

The Riemannian logarithmic map, illustrated in the left panel of Figure S.2, has important applications in statistical analysis of manifold-valued data. For example, a $\mathcal{M}$-valued random process $X(t)$ with the mean function $\mu(t)$ may be converted by the logarithmic map into $\mathrm{Log}_{\mu(t)}X(t)$, a process taking values in $T_{\mu(t)}\mathcal{M}$ for each $t$, for further analysis. As $t$ varies, $\mathrm{Log}_{\mu(t)}X(t)$ moves across different tangent spaces. This requires us to consider the space formed by the union of all tangent spaces. Such space is a special case of vector bundle in which each point of the manifold is associated with a vector space.

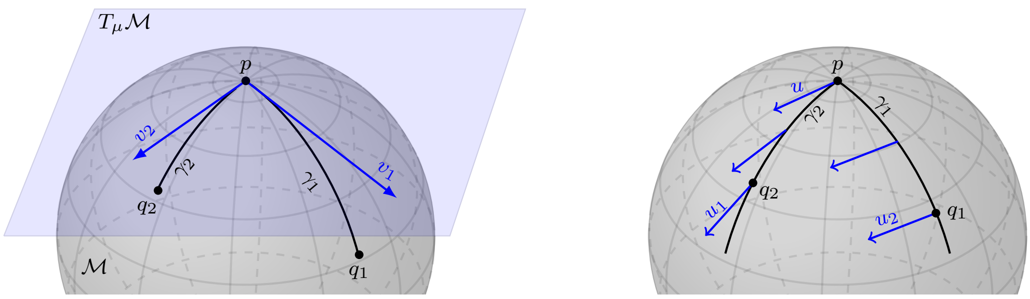


Figure S.2: Left: Illustration of the tangent space, geodesic, Riemannian exponential map and logarithmic map, where $\gamma_j(0) = p$, $\gamma_j(1) = q_j$, $q_j = \mathrm{Exp}_p v_j$ and $v_j = \mathrm{Log}_p q_j$, for $j = 1, 2$. Right: Illustration of parallel transport. The tangent vector $u$ at $p$ is parallelly transported along geodesics $\gamma_1$ and $\gamma_2$ to $q_1$ and $q_2$, resulting in tangent vectors $u_1 = \mathcal{P}_p^{q_1} u$ and $u_2 = \mathcal{P}_p^{q_2} u$, respectively.

**Definition S.11** (Vector bundle)**.** *A smooth vector bundle, denoted by* $\pi : \mathcal{E} \to \mathcal{M}$ *or simply* $\mathcal{E}$*, consists of a base smooth manifold* $\mathcal{M}$*, a smooth manifold* $\mathcal{E}$ *called total space, and a smooth bundle projection* $\pi$*, such that for every* $p \in \mathcal{M}$*, the fiber* $\pi^{-1}(p)$ *is a* $k$*-dimensional real vector space, and there is an open neighborhood* $U \subset \mathcal{M}$ *of* $p$ *and a diffeomorphism* $\Phi : \pi^{-1}(U) \to U \times \mathbb{R}^k$ *satisfying the property that for all* $z \in U$*,* $(\pi \circ \Phi^{-1})(z, v) = z$ *for all* $v \in \mathbb{R}^k$ *and the map* $v \mapsto \Phi^{-1}(z, v)$ *is a linear isomorphism between* $\mathbb{R}^k$ *and* $\pi^{-1}(z)$*.*

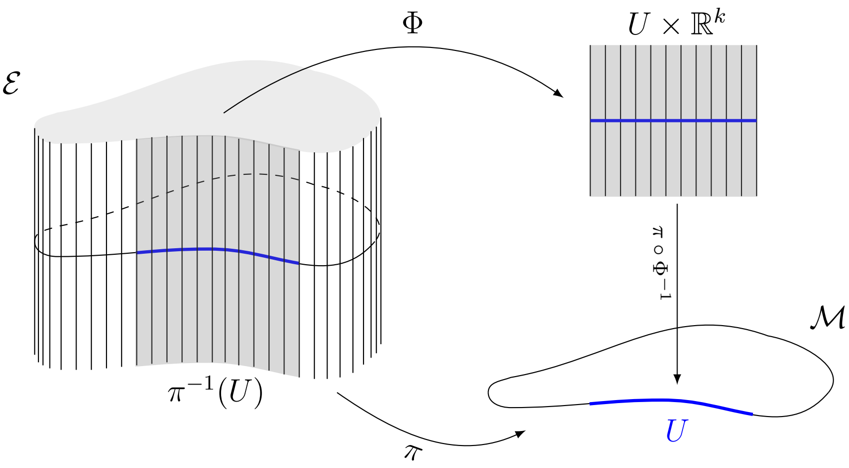


Figure S.3: Illustration of the vector bundle. The closed curve in the bottom represents the base manifold $\mathcal{M}$ and the figure on the left represents the total space $\mathcal{E}$, where each vertical line represents a fiber. The thickened segment $U \subset \mathcal{M}$ represents an open subset of the manifold $\mathcal{M}$, while $\Phi$ is a local trivialization defined on $\pi^{-1}(U)$ that is highlighted in gray in the total space.

The map $\Phi$ in the above definition is called a **local trivialization**. As graphically illustrated in Figure S.3, a vector bundle locally resembles the product space $U \times \mathbb{R}^k$ for some integer $k$. A function $V$ defined on $\mathcal{M}$ is called a **section** of the vector bundle if $V(p) \in \pi^{-1}(p)$ for all $p \in \mathcal{M}$. As previously mentioned, the union of all tangent spaces of a manifold, called the **tangent bundle** of the manifold, is a prominent example of vector bundle, where the tangent space at each point is a fiber. In particular, a section of a tangent bundle is also a vector field.

For a smooth function $f : \mathcal{M} \to \mathbb{R}$ and a tangent vector $v \in T_p\mathcal{M}$, the covariant derivative of $f$ at $p$ along the direction $v$, denoted by $\nabla_v f$, is defined by

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

where $\gamma : [-1, 1] \to \mathcal{M}$ is a differentiable curve such that $\gamma(0) = p$ and $\gamma'(0) = v$. For a smooth vector field $U$, $\nabla_{V(p)} f$ is a real-valued function of $p$. Let $C^\infty(\mathcal{M})$ denote the collection of smooth real-valued functions defined on $\mathcal{M}$. In addition, for $f \in C^\infty(\mathcal{M})$ and a smooth vector field $U$, $fU$ denotes a smooth vector field defined by $(fU)(p) = f(p)U(p)$ for all $p \in \mathcal{M}$. Let $\Gamma(\mathcal{E})$ be the collection of smooth sections.

For a smooth curve in a Euclidean space, it is meaningful to discuss its acceleration which is represented by the second derivative of the curve. Note that the definition of second derivative involves differentiating the first derivative. To generalize the concept of acceleration to manifold-valued curves, we then need to differentiate the velocity — represented by tangent vectors — of the curve, and this involves the concept of connection.

**Definition S.12** (Connection)**.** *A connection in a vector bundle $\mathcal{E}$ is a map $\nabla : \Gamma(\mathcal{E}) \times \Gamma(\mathcal{E}) \to \Gamma(\mathcal{E})$, with $(V, U) \mapsto \nabla_V U$, that satisfies the following properties:*

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

- $\nabla_V(fU) = f\nabla_V U + (\nabla_V f)U$ *for* $f \in C^\infty(\mathcal{M})$.

In the above, the value of $\nabla_V U$ at $p$ depends on $V$ only through its value at $p$ (Proposition 4.5, Lee, 2018). This observation leads to the definition of covariant derivative of a vector field at $p$ along $v \in T_p\mathcal{M}$. Consequently, the expression $\nabla_v U$ is sensible for $v \in T_p\mathcal{M}$, and is called the **covariant derivative** of $U$ at $p$ along the tangent vector $v$.

Let $f_{U,V}(p) = \langle U(p), V(p)\rangle_p$. For a connection $\nabla$ on the tangent bundle of $\mathcal{M}$, we say $\nabla$ is compatible with the metric on $\mathcal{M}$ if $\nabla_v f_{U,V} = \langle \nabla_v U, V\rangle_p + \langle U, \nabla_v V\rangle_p$ for all smooth vector fields $U$ and $V$, each $p \in \mathcal{M}$ and each tangent vector $v \in T_p\mathcal{M}$. For vector fields $U$ and $V$, we use $[U,V]$ to denote a new vector field such that $\nabla_{[U,V]} f = \nabla_U\nabla_V f - \nabla_V\nabla_U f$ for all $f \in C^\infty(\mathcal{M})$. Similarly, for $u, v \in T_p\mathcal{M}$, $[u,v]$ denotes the tangent at $p$ such that $\nabla_{[u,v]} f = \nabla_u\nabla_v f - \nabla_v\nabla_u f$ for all $f \in C^\infty(\mathcal{M})$. A connection is torsion-free if $\nabla_U V - \nabla_v U = [U,V]$ for all smooth vector fields $U, V$. For a Riemannian manifold, there exists a unique connection is both torsion-free and compatible with the Riemannian metric. Such connection is called the **Levi–Civita** connection and deemed the canonical connection in the tangent bundle.

To identify different fibers, one can introduce a parallel transport $\mathscr{P}$ on a vector bundle along a curve $\gamma$ on the base manifold. Such parallel transport must satisfy the following axioms: 1) $\mathscr{P}_p^p$ is the identity map on $\pi^{-1}(p)$ for all $p \in \mathcal{M}$, 2) $\mathscr{P}_{\gamma(u)}^{\gamma(t)} \circ \mathscr{P}_{\gamma(s)}^{\gamma(u)} = \mathscr{P}_{\gamma(s)}^{\gamma(t)}$, and 3) the dependence of $\mathscr{P}$ on $\gamma$, $s$ and $t$ are smooth. An example is the vector bundle and the parallel transport constructed in Section 2.4. For a tangent bundle, such parallel transport can be induced by a connection.

**Definition S.13** (Parallel transport)**.** *Let $\nabla$ be a connection in the tangent bundle of a Riemannian manifold $\mathcal{M}$. A smooth vector field $U$ is parallel along $\gamma : I \to \mathcal{M}$ (with respect to $\nabla$) if $\nabla_{\gamma'(t)} U = 0$ for all $t \in I$. The parallel transport of $v \in T_p\mathcal{M}$ along $\gamma$ with $p = \gamma(0)$ is $U(\gamma(t))$ for the unique smooth vector field $U$ along $\gamma$ such that $U$ is parallel along $\gamma$ and $U(0) = v$.*

Unlike Euclidean spaces, manifolds are often not flat and exhibit curvature that measures the degree of deviation from being flat. For smooth vector fields $U, V, W$, we define the map $R(U,V,W) = \nabla_U\nabla_V W - \nabla_V\nabla_U W - \nabla_{[U,V]} W$. It turns out that the value of $R(U,V,W)$ at $p$ depends only on the values of $U, V, W$ at $p$, and therefore it is sensible to write $R(u,v,w)$ for tangent vectors at the same point.

**Definition S.14** (Sectional curvature)**.** *The sectional curvature at $p$ is a real-valued function on $T_p\mathcal{M} \times T_p\mathcal{M}$ defined for $u, v \in T_p\mathcal{M}$ by $\mathfrak{K}(u,v) = \langle R(u,v,v), u\rangle_p / (\langle u,u\rangle_p \langle v,v\rangle_p - \langle u,v\rangle_p^2)$.*

Note that sectional curvature is invariant to the length of tangent vectors $u$ and $v$. We say the sectional curvature of a Riemannian manifold $\mathcal{M}$ is upper (lower, resp.) bounded by $\kappa$ if $\mathfrak{K}(u,v) \le \kappa$ ($\mathfrak{K}(u,v) \ge \kappa$, resp.) for all $p \in \mathcal{M}$ and $u, v \in T_p\mathcal{M}$.

## S.2 Asymptotic distribution of the covariance estimator

In this section, we provide a weak convergence result for the estimated covariance under the assumption $\hat{\mu} = \mu$ that is also adopted in Zhang and Wang (2016) for simplification. With the consistency property of $\hat{\mu}$, the derived asymptotic normality under the assumption $\hat{\mu} = \mu$ may approximate the reality well when sample size is sufficiently large. To drop this assumption, a detailed analysis on the asymptotic normality of $\hat{\mu}$ seems needed. However, this turns out to be very challenging in the context of Riemannian data due to the curvature effect. Since the focus of this paper is the construction of the covariance vector bundle framework rather than the mean estimation by local linear smoothing, we decide to leave it for future study.

For $v_s \in T_{\mu(s)}\mathcal{M}$ and $v_t \in T_{\mu(t)}\mathcal{M}$, let $\gamma_{v_s,v_t} = \langle \mathcal{C}v_s, v_t\rangle_{\mu(t)}$ and $\hat{\gamma}_{v_s,v_t} = \langle \mathscr{P}_{(\hat{\mu}(s),\hat{\mu}(t))}^{(\mu(s),\mu(t))} \hat{\mathcal{C}}(s,t) v_s, v_t\rangle_{\mu(t)}$. It is seen that each pair $(v_s, v_t)$ defines a linear functional on $\mathbb{L}(\mu(s),\mu(t))$. Then $\mathscr{P}_{(\hat{\mu}(s),\hat{\mu}(t))}^{(\mu(s),\mu(t))} \hat{\mathcal{C}}(s,t)$ is weakly convergent to $\mathcal{C}(s,t)$ if we can show that $\hat{\gamma}_{v_s,v_t}$ weakly converges to $\gamma_{v_s,v_t}$ for all $(v_s,v_t)$ according to the Cramér–Wold device. Below we demonstrate this for the random design; similar result can be proved for the hybrid design, and for deterministic design by utilizing the concept of design densities (Sacks and Ylvisaker, 1970).

To state the result, let $f$ be the probability density of $T_{11}$, and $\|K\|^2 = \int K^2(u)\mathrm{d}u$. Define

$$\begin{aligned}
V_1(s,t,v_s,v_t) =&\mathrm{Var}\{\langle \mathrm{Log}_{\mu(T_1)}X(T_1),v_s\rangle\langle \mathrm{Log}_{\mu(T_2)}X(T_2),v_t\rangle \mid T_1=s,T_2=t\},\\
V_2(s,t,v_s,v_t) =&\mathrm{Cov}\{\langle \mathrm{Log}_{\mu(T_1)}X(T_1),v_s\rangle\langle \mathrm{Log}_{\mu(T_2)}X(T_2),v_t\rangle,\langle \mathrm{Log}_{\mu(T_1)}X(T_1),v_s\rangle\langle \mathrm{Log}_{\mu(T_3)}X(T_3),v_t\rangle \mid\\
&T_1=s,T_2=t,T_3=t\},\\
V_3(s,t,v_s,v_t) =&\mathrm{Cov}\{\langle \mathrm{Log}_{\mu(T_1)}X(T_1),v_s\rangle\langle \mathrm{Log}_{\mu(T_2)}X(T_2),v_t\rangle,\langle \mathrm{Log}_{\mu(T_3)}X(T_3),v_s\rangle\langle \mathrm{Log}_{\mu(T_4)}X(T_4),v_t\rangle \mid\\
&T_1=s,T_2=t,T_3=s,T_4=t\}.
\end{aligned}$$

**Theorem S.2.1.** *Suppose that Assumptions 2.1, 2.2, 3.1, 4.1, 4.4, 4.5 and 4.6 hold. In addition, assume $h_{\mathcal{C}}\to 0$, $nm^2h_{\mathcal{C}}^2\to\infty$, $m^3h_{\mathcal{C}}^3\ll n$, and $mh_{\mathcal{C}}\to c$ for some constant $c\in[0,\infty]$. Then for $s,t\in\mathcal{T}$ and $v_s\in T_{\mu(s)}\mathcal{M}$ and $v_t\in T_{\mu(t)}\mathcal{M}$,*

$$\Sigma_{\mathcal{C}}^{-1/2}\Big(\hat{\gamma}_{v_s,v_t}-\gamma_{v_s,v_t}-b(h_{\mathcal{C}})+o_P(h_{\mathcal{C}}^2)\Big)\xrightarrow{D}N(0,1), \tag{S.1}$$

*where $b(h)=\frac{1}{2}h^2\big(\int u^2K(u)\mathrm{d}u\big)\Big(\frac{\partial^2\gamma}{\partial s^2}(s,t)+\frac{\partial^2\gamma}{\partial t^2}(s,t)\Big)$ and*

$$\Sigma_{\mathcal{C}} =\{1+1_{s=t}\}\left\{\tfrac{\|K\|^4}{nm(m-1)h_{\mathcal{C}}^2}\tfrac{V_1(s,t,v_s,v_t)}{f(s)f(t)}+\tfrac{\|K\|^2}{n(m-1)h_{\mathcal{C}}}\tfrac{f(s)V_2(t,s,v_t,v_s)+f(t)V_2(s,t,v_s,v_t)}{f(s)f(t)}\right\}+\tfrac{(m-2)(m-3)V_3(s,t,v_s,v_t)}{nm(m-1)}.$$

The proof for the above theorem follows from Zhang and Wang (2016) once one realizes that $\hat{\gamma}_{v_s,v_t}$ is the estimated covariance based on the raw covariance $\langle \mathrm{Log}_{\mu(T_{ij})}X(T_{ij}),v_s\rangle\langle \mathrm{Log}_{\mu(T_{ik})}X(T_{ik}),v_t\rangle$. From the theorem we then observe the same phase transition as that in Zhang and Wang (2016) in the following corollary.

**Corollary S.2.1.** *Assume the conditions of Theorem S.2.1.*

(a) *When $m\gg n^{1/4}$, with $h_{\mathcal{C}}\ll n^{-1/4}$ and $mh_{\mathcal{C}}\to\infty$, one has*

$$\sqrt{n}\Big(\hat{\gamma}_{v_s,v_t}-\gamma_{v_s,v_t}\Big)\xrightarrow{D}N\Big(0,V_1(s,t,v_s,v_t)\Big).$$

(b) *When $m/n^{1/4}\to c_\circ$, with $h_{\mathcal{C}}=c_*n^{-1/4}$, one has*

$$\sqrt{n}\Big(\hat{\gamma}_{v_s,v_t}-\gamma_{v_s,v_t}-b(h)\Big)\xrightarrow{D}N(0,\Sigma_*)$$

*where*

$$\Sigma_* =\{1+1_{s=t}\}\left\{\frac{\|K\|^4}{c_\circ^2c_*^2}\frac{V_1(s,t,v_s,v_t)}{f(s)f(t)}+\frac{\|K\|^2}{c_\circ c_*}\frac{f(s)V_2(t,s,v_t,v_s)+f(t)V_2(s,t,v_s,v_t)}{f(s)f(t)}\right\}+V_3(s,t,v_s,v_t).$$

(c) *When $m\ll n^{1/4}$, with $h_{\mathcal{C}}\asymp n^{-1/6}m^{-1/3}$, one has*

$$n^{1/3}m^{2/3}\Big(\hat{\gamma}_{v_s,v_t}-\gamma_{v_s,v_t}-b(h)\Big)\xrightarrow{D}N\Big(0,\{1+1_{s=t}\}\|K\|^4\tfrac{V_1(s,t,v_s,v_t)}{f(s)f(t)}\Big).$$

## S.3 Proofs of Main Results

*Proof of Lemma 2.1.* We prove this result for any fixed $t \in \mathcal{T}$. First, we show $\mathbf{E}\{\mathrm{Log}_{\mu(t)} X(t)\} = 0$. Suppose that $\gamma$ is a geodesic emanating from $\mu(t)$ with velocity $v \in T_{\mu(t)}\mathcal{M}$ and $\|v\|_{\mu(t)} = 1$. According to Proposition 2.10 of Oller and Corcuera (1995), one has

$$\begin{aligned}\frac{d}{ds}F(\gamma(s),t)\Big|_{s=0} =&\mathbf{E}\{2d_{\mathcal{M}}(X(t),\mu(t))\cos\langle v, \mathrm{Log}_{\mu(t)}X(t)\rangle\}\\ =&2\mathbf{E}\{\|\mathrm{Log}_{\mu(t)}X(t)\|\cos\langle v, \mathrm{Log}_{\mu(t)}X(t)\rangle\}.\end{aligned}$$

Since $F(p,t)$ reaches the minimum at $p = \mu(t)$, we have $\frac{d}{ds}F(\gamma(s),t)\big|_{s=0} = 0$ for any $v \in T_{\mu(t)}\mathcal{M}$. As $\|\mathrm{Log}_{\mu(t)}X(t)\|\cos\langle v, \mathrm{Log}_{\mu(t)}X(t)\rangle$ is the projection of $\mathrm{Log}_{\mu(t)}X(t)$ onto $v$, $\mathbf{E}\{\|\mathrm{Log}_{\mu(t)}X(t)\|\cos\langle v, \mathrm{Log}_{\mu(t)}X(t)\rangle\} = 0$ for all $v \in T_{\mu(t)}\mathcal{M}$ then implies $\mathbf{E}\{\mathrm{Log}_{\mu(t)}X(t)\} = 0$. According to the definition of $Y(t)$, it holds that $\mathbf{E}\{\mathrm{Log}_{\mu(t)}Y(t)\} = 0$. Similarly, it can be shown that the derivative of $F^*(\cdot,t) = \mathbf{E}\{d^2_{\mathcal{M}}(Y(t),\cdot)\}$ vanishes at $\mu(t)$. With Assumption 2.1, this implies that $\mu(t)$ is the unique minimum of $F^*(\cdot,t)$ and thus is the Fréchet mean of $Y(t)$. □

*Proof of Theorem 2.1.* The mean continuity and joint measurability ensure that $\mathrm{Log}_\mu X$ is a random element in $\mathscr{T}(\mu)$. According to the definition of $\mathcal{C}(s,t)$, for any $u, v \in \mathscr{T}(\mu)$,

$$\begin{aligned}\langle\!\langle \int_{\mathcal{T}}\mathcal{C}(s,\cdot)u(s)\mathrm{d}s, v\rangle\!\rangle_\mu =& \int_{\mathcal{T}}\int_{\mathcal{T}}\langle \mathcal{C}(s,t)u(s), v(t)\rangle_{\mu(t)}\mathrm{d}s\mathrm{d}t\\ =&\int_{\mathcal{T}}\int_{\mathcal{T}}\mathbf{E}\{\langle \mathrm{Log}_{\mu(s)}X(s), u(s)\rangle_{\mu(s)}\langle \mathrm{Log}_{\mu(t)}X(t), v(t)\rangle_{\mu(t)}\}\mathrm{d}s\mathrm{d}t\\ =&\mathbf{E}\left\{\int_{\mathcal{T}}\langle \mathrm{Log}_{\mu(s)}X(s), u(s)\rangle_{\mu(s)}\mathrm{d}s\int_{\mathcal{T}}\langle \mathrm{Log}_{\mu(t)}X(t), v(t)\rangle_{\mu}(t)\mathrm{d}t\right\}\\ =&\mathbf{E}(\langle\!\langle \mathrm{Log}_\mu X, u\rangle\!\rangle_\mu\langle\!\langle \mathrm{Log}_\mu X, v\rangle\!\rangle_\mu) = \langle\!\langle \mathbf{C}u, v\rangle\!\rangle_\mu,\end{aligned}$$

which implies that $(\mathbf{C}u)(t) = \int_{\mathcal{T}}\mathcal{C}(s,t)u(s)\mathrm{d}s$. □

*Proof of Theorem 2.2.* To see that $\{(\pi^{-1}(U_\alpha \times U_\beta), \varphi_{\alpha,\beta}) : (\alpha,\beta) \in J^2\}$ is a smooth atlas, it is sufficient to check the transition maps. Suppose that $(p, q, \sum_{j,k=1}^d v_{jk}B_{\alpha,j}(p)\otimes B_{\beta,k}(q)) \in \pi^{-1}(U_\alpha \times U_\beta)\cap\pi^{-1}(U_{\tilde\alpha}\times U_{\tilde\beta})$ is also represented by $(p, q, \sum_{j,k=1}^d \tilde v_{jk}B_{\tilde\alpha,j}(p)\otimes B_{\tilde\beta,k}(q))$. The transformation from the coefficient vector $v = (v_{11}, v_{12}, \ldots, v_{dd})$ to $\tilde v = (\tilde v_{11}, \tilde v_{12}, \ldots, \tilde v_{dd})$ is smooth, since $\tilde v = \{\mathcal{J}_\alpha^\top(p)\otimes\mathcal{J}_\beta^\top(q)\}v$ and $\mathcal{J}_\alpha^\top(p)$, $\mathcal{J}_\beta^\top(q)$ and their Kronecker product $\mathcal{J}_\alpha^\top(p)\otimes\mathcal{J}_\beta^\top(q)$ are respectively smooth in $p$, $q$ and $(p,q)$, where $\mathcal{J}_\alpha(\cdot)$ denotes the Jacobian matrix that transforms the basis $\{B_{\alpha,1}(\cdot),\ldots,B_{\alpha,d}(\cdot)\}$ into $\{B_{\tilde\alpha,1}(\cdot),\ldots,B_{\tilde\alpha,d}(\cdot)\}$.

According to the vector bundle construction lemma (Lemma 5.5, Lee, 2002), it is sufficient to check that when $U := (U_\alpha \times U_\beta)\cap(U_{\tilde\alpha}\times U_{\tilde\beta}) \neq \varnothing$ for some indices $\alpha, \beta, \tilde\alpha, \tilde\beta$, the composite map $\Phi_{\alpha,\beta}\circ\Phi^{-1}_{\tilde\alpha,\tilde\beta}$ from $U \times \mathbb{R}^{d^2}$ to itself has the form $\Phi_{\alpha,\beta}\circ\Phi^{-1}_{\tilde\alpha,\tilde\beta} = (p, q, \mathcal{J}(p,q)v)$ for a smooth map $\mathcal{J} : U \to \mathrm{GL}(d^2,\mathbb{R})$, where $\mathrm{GL}(d^2,\mathbb{R})$ is the collection of invertible real $d^2\times d^2$ matrices. From above discussion, we have $\mathcal{J}(p,q) = \mathcal{J}_\alpha^\top(p)\otimes\mathcal{J}_\beta^\top(q)$ is smooth in $(p,q)$. In addition, $\mathcal{J}(p,q) \in \mathrm{GL}(d^2,\mathbb{R})$ since both $\mathcal{J}_\alpha^\top(p)$ and $\mathcal{J}_\alpha^\top(q)$ are invertible and so is their Kronecker product. Note that the vector bundle construction lemma also asserts that any compatible atlas for $\mathcal{M}$ gives rise to the same smooth structure on $\mathbb{L}$. □

*Proof of Theorem 2.3.* One can show that the parallel transport defined in (5) is a genuine parallel transport satisfying the property of Definition A.54 of Rodrigues and Capelas de Oliveira (2007) on the vector bundle. Then the conclusion directly follows from Definitions A.55 and A.57 of Rodrigues and Capelas de Oliveira (2007) and the remarks right below them. □

*Proof of Theorem 2.4.* We first show that the definition (7) is invariant to the choice of orthonormal bases. To this end, fix an orthonormal basis in $T_q\mathcal{M}$, and suppose that $\{\tilde{e}_1,\dots,\tilde{e}_d\}$ is another orthonormal basis in $T_p\mathcal{M}$ and is related to $\{e_1,\dots,e_d\}$ by a $d\times d$ unitary matrix $\mathbf{O}$. Let $A_1$ and $A_2$ be the respective matrix
310 representation of $L_1$ and $L_2$ under the basis $\{e_1,\dots,e_d\}$. Then their matrix representation under the basis $\{\tilde{e}_1,\dots,\tilde{e}_d\}$ is $\tilde{A}_1=\mathbf{O}A_1$ and $\tilde{A}_2=\mathbf{O}A_2$, respectively. The inner product $G_{p,q}(L_1,L_2)$ is then calculated by $\mathrm{tr}(\tilde{A}_1^\top\tilde{A}_2)=\mathrm{tr}(A_1^\top\mathbf{O}^\top\mathbf{O}A_2)=\mathrm{tr}(A_1^\top A_2)$, which shows that $G_{p,q}(L_1,L_2)$ is invariant to the choice of bases in $T_p\mathcal{M}$. Its invariance to the choice of bases in $T_q\mathcal{M}$ can be proved in a similar fashion.

The smoothness of $G$ can be established by an argument similar to the one leading to Theorem 2.2 in conjunction with smoothness of the trace of matrices. To see that the parallel transport (5) preserves the bundle metric and thus defines isometries among fibers of $\mathbb{L}$, i.e., for any $L_1,L_2\in\mathbb{L}(p_1,q_1)$,

$$G_{(p_1,q_1)}(L_1,L_2)=G_{(p_2,q_2)}(\mathscr{P}_{(p_1,q_1)}^{(p_2,q_2)}L_1,\mathscr{P}_{(p_1,q_1)}^{(p_2,q_2)}L_2),$$

suppose that $\{e_1,\dots,e_d\}$ is an orthogonal basis of $T_{p_1}\mathcal{M}$. Then $\{\mathcal{P}_{p_1}^{p_2}e_1,\dots,\mathcal{P}_{p_1}^{p_2}e_d\}$ is an orthogonal basis of $T_{p_2}\mathcal{M}$. This further implies that

$$\begin{aligned}G_{(p_2,q_2)}(\mathscr{P}_{(p_1,q_1)}^{(p_2,q_2)}L_1,\mathscr{P}_{(p_1,q_1)}^{(p_2,q_2)}L_2)&=\sum_{k=1}^d\langle(\mathscr{P}_{(p_1,q_1)}^{(p_2,q_2)}L_1)(\mathcal{P}_{p_1}^{p_2}e_k),(\mathscr{P}_{(p_1,q_1)}^{(p_2,q_2)}L_2)(\mathcal{P}_{p_1}^{p_2}e_k)\rangle_{q_2}\\&=\sum_{k=1}^d\langle\mathcal{P}_{q_1}^{q_2}[L_1(e_k)],\mathcal{P}_{q_1}^{q_2}[L_2(e_k)]\rangle_{q_2}\\&=\sum_{k=1}^d\langle L_1(e_k),L_2(e_k)\rangle_{q_1}=G_{(p_1,q_1)}(L_1,L_2),\end{aligned}$$

which completes the proof. □

*Proof of Proposition 3.1.* Suppose that $(\tilde{B}_{ij,1},\dots,\tilde{B}_{ij,d})$ is another orthonormal basis for $T_{\hat{\mu}(T_{ij})}\mathcal{M}$, and $\mathbf{O}_{ij}$ is the unitary matrix relating $(B_{ij,1},\dots,B_{ij,d})$ to $(\tilde{B}_{ij,1},\dots,\tilde{B}_{ij,d})$. Then the coefficient vectors $\tilde{z}_{ij}$ and $\tilde{g}_{k,ij}$ of $\mathrm{Log}_{\hat{\mu}(T_{ij})}Y_{ij}$ and $\hat{\psi}_k(T_{ij})$ under the basis $(\tilde{B}_{ij,1},\dots,\tilde{B}_{ij,d})$ are linked to $z_{ij}$ and $g_{k,ij}$ by $\tilde{z}_{ij}=\mathbf{O}_{ij}z_{ij}$ and $\tilde{g}_{k,ij}=\mathbf{O}_{ij}g_{k,ij}$, respectively. Similarly, $\tilde{C}_{i,jl}$ is linked to $C_{i,jl}$ by $\tilde{C}_{i,jl}=\mathbf{O}_{ij}C_{i,jl}\mathbf{O}_{il}^\top$. More concisely, if we put

$$\mathbf{O}_i=\begin{pmatrix}\mathbf{O}_{i1}&&&\\&\mathbf{O}_{i2}&&\\&&\ddots&\\&&&\mathbf{O}_{im_i}\end{pmatrix},$$

315 then $\tilde{z}_i=\mathbf{O}_iz_i$, $\tilde{g}_{k,i}=\mathbf{O}_ig_{k,i}$ and $\tilde{\Sigma}_i=\mathbf{O}_i\Sigma_i\mathbf{O}_i^\top$, which are the counterpart of $z_i$, $g_{k,i}$ and $\Sigma_i$ under the bases $(\tilde{B}_{ij,1},\dots,\tilde{B}_{ij,d})$, respectively. Note that $\tilde{\Sigma}_i^{-1}=(\mathbf{O}_i\Sigma_i\mathbf{O}_i^\top)^{-1}=\mathbf{O}_i^{-\top}\Sigma_i^{-1}\mathbf{O}_i^{-1}=\mathbf{O}_i\Sigma_i^{-1}\mathbf{O}_i^\top$ since $\mathbf{O}_{ij}$ are unitary matrices and thus $\mathbf{O}_i^{-1}=\mathbf{O}_i^\top$. Now we see that $\tilde{g}_{k,i}^\top\tilde{\Sigma}_i^{-1}\tilde{z}_i=g_{k,i}^\top\mathbf{O}_i^\top\mathbf{O}_i\Sigma_i^{-1}\mathbf{O}_i^\top\mathbf{O}_iz_i=g_{k,i}^\top\Sigma_i^{-1}z_i$, which clearly implies that the scores $\hat{\xi}_{ik}$ calculated under the bases $(\tilde{B}_{ij,1},\dots,\tilde{B}_{ij,d})$ is identical to the one computed under the bases $(B_{ij,1},\dots,B_{ij,d})$. □

*Proof of Lemma 4.1.* Notice that

$$\begin{aligned}&\|\mathcal{P}_{q_1}^{p_1}\mathcal{P}_{q_2}^{q_1}\mathrm{Log}_{q_2}y-\mathcal{P}_{p_2}^{p_1}\mathrm{Log}_{p_2}y\|_{p_1}\\=&\|\mathcal{P}_{p_2}^{q_2}\mathcal{P}_{p_1}^{p_2}\mathcal{P}_{q_1}^{p_1}\mathcal{P}_{q_2}^{q_1}\mathrm{Log}_{q_2}y-\mathcal{P}_{p_2}^{q_2}\mathrm{Log}_{p_2}y\|_{q_2}\\\le&\|\mathcal{P}_{p_2}^{q_2}\mathcal{P}_{p_1}^{p_2}\mathcal{P}_{q_1}^{p_1}\mathcal{P}_{q_2}^{q_1}\mathrm{Log}_{q_2}y-\mathrm{Log}_{q_2}y\|_{q_2}+\|\mathrm{Log}_{q_2}y-\mathcal{P}_{p_2}^{q_2}\mathrm{Log}_{p_2}y\|_{q_2},\end{aligned}$$

where the equality follows from the fact that parallel transport preserves the inner product. Note that the operator $\mathcal{P}_{p_2}^{q_2}\mathcal{P}_{p_1}^{p_2}\mathcal{P}_{q_1}^{p_1}\mathcal{P}_{q_2}^{q_1}$ moves a tangent vector parallelly along a geodesic quadrilateral defined by the the points $p_1, p_2, q_1, q_2$. The holonomy theory (Eq (6), Nichols et al., 2016) and the compactness of $\mathcal{G}$ suggests that there exists a constant $c_1 > 0$ depending only on $\mathcal{G}$, such that for any $v \in T_{q_2}\mathcal{M}$ with $\|v\|_{q_2} \leq \mathrm{diam}(\mathcal{G})$,

$$\|\mathcal{P}_{p_2}^{q_2}\mathcal{P}_{p_1}^{p_2}\mathcal{P}_{q_2}^{p_1}v - v\|_{q_2} \leq c_1\|\mathrm{Log}_{q_2}p_2\|_{q_2} = c_1 d_{\mathcal{M}}(p_2, q_2),$$
$$\|\mathcal{P}_{p_1}^{q_2}\mathcal{P}_{q_1}^{p_1}\mathcal{P}_{q_2}^{q_1}v - v\|_{q_2} \leq c_1\|\mathrm{Log}_{q_1}p_1\|_{q_2} = c_1 d_{\mathcal{M}}(p_1, q_1),$$

which further imply that

$$\begin{aligned}
\|\mathcal{P}_{p_2}^{q_2}\mathcal{P}_{p_1}^{p_2}\mathcal{P}_{q_1}^{p_1}\mathcal{P}_{q_2}^{q_1}v - v\|_{q_2} &= \|(\mathcal{P}_{p_2}^{q_2}\mathcal{P}_{p_1}^{p_2}\mathcal{P}_{q_2}^{p_1})(\mathcal{P}_{p_1}^{q_2}\mathcal{P}_{q_1}^{p_1}\mathcal{P}_{q_2}^{q_1})v - v\|_{q_2} \\
&\leq \|(\mathcal{P}_{p_2}^{q_2}\mathcal{P}_{p_1}^{p_2}\mathcal{P}_{q_2}^{p_1})(\mathcal{P}_{p_1}^{q_2}\mathcal{P}_{q_1}^{p_1}\mathcal{P}_{q_2}^{q_1})v - (\mathcal{P}_{p_1}^{q_2}\mathcal{P}_{q_1}^{p_1}\mathcal{P}_{q_2}^{q_1})v\|_{q_2} + \|(\mathcal{P}_{p_1}^{q_2}\mathcal{P}_{q_1}^{p_1}\mathcal{P}_{q_2}^{q_1})v - v\|_{q_2} \\
&\leq c_1(d_{\mathcal{M}}(p_2, q_2) + d_{\mathcal{M}}(p_1, q_1)).
\end{aligned}$$

According to Theorem 3 in Pennec (2019), we have

$$\|\mathrm{Log}_{q_2}y - \mathcal{P}_{p_2}^{q_2}\mathrm{Log}_{p_2}y\|_{q_2} \leq c_2\|\mathrm{Log}_{q_2}p_2\|_{q_2} \leq c_2 d_{\mathcal{M}}(p_2, q_2)$$

for some constant $c_2 > 0$ depending only on $\mathcal{G}$. The proof is then completed by taking $c = c_1 + c_2$. □

*Proof of Propositions 4.1 and 4.2.* Simple computation shows that

$$\begin{aligned}
&\big(\hat{Q}_n(y,\tau) - F^*(y,\tau)\big) \\
=&\frac{\hat{u}_2(\tau)}{\hat{\sigma}_0^2(\tau)}\frac{1}{nm}\sum_{ij}K_{h_\mu}(T_{ij}-\tau)\Big(d_{\mathcal{M}}^2(Y_{ij},y) - F^*(y,\tau)\Big) \\
&-\frac{\hat{u}_1(\tau)}{\hat{\sigma}_0^2(\tau)}\frac{1}{nm}\sum_{ij}K_{h_\mu}(T_{ij}-\tau)(T_{ij}-\tau)\Big(d_{\mathcal{M}}^2(Y_{ij},y) - F^*(y,\tau)\Big) \\
=&\frac{\hat{u}_2(\tau)}{\hat{\sigma}_0^2(\tau)}\frac{1}{nm}\sum_{ij}K_{h_\mu}(T_{ij}-\tau)\Big(d_{\mathcal{M}}^2(Y_{ij},y) - F^*(y,\tau) - \partial_\tau F^*(y,\tau)(T_{ij}-\tau)\Big) \\
&-\frac{\hat{u}_1(\tau)}{\hat{\sigma}_0^2(\tau)}\frac{1}{nm}\sum_{ij}K_{h_\mu}(T_{ij}-\tau)(T_{ij}-\tau)\Big(d_{\mathcal{M}}^2(Y_{ij},y) - F^*(y,\tau) - \partial_\tau F^*(y,\tau)(T_{ij}-\tau)\Big).
\end{aligned}$$

Below we focus on the fist term, noting that the second term can be analyzed in a similar way.

Define

$$U := \frac{1}{nm}\sum_{ij}K_{h_\mu}(T_{ij}-\tau)\Big(d_{\mathcal{M}}^2(Y_{ij},y) - F^*(y,\tau) - \partial_\tau F^*(y,\tau)(T_{ij}-\tau)\Big).$$

Then, according to either Lemma S.4.1 or Lemma S.4.3, the rate of the first term depends on the rate of $U$. By Taylor expansion of $F^*(y, T_{ij})$ at $\tau$ and Assumption 4.5(a), we have

$$\begin{aligned}
\sup_{\tau\in B(t;h)}|\mathbf{E}U| &= \sup_{\tau\in B(t;h)}\left|\mathbf{E}\left(\frac{1}{nm}\sum_{ij}K_{h_\mu}(T_{ij}-\tau)\big(F^*(y,T_{ij}) - F^*(y,\tau) - \partial_\tau F^*(y,\tau)(T_{ij}-\tau)\big)\right)\right| \\
&= \sup_{\tau\in B(t;h)}\left|\mathbf{E}\left(\frac{1}{nm}\sum_{ij}K_{h_\mu}(T_{ij}-\tau)\times O(h_\mu^2)\right)\right| = O(h_\mu^2).
\end{aligned}$$

For the random and hybrid designs, define the envelop function

$$H := \frac{2\mathrm{diam}(\mathcal{K})^2}{m}\sum_{j=1}^{m}\sup_{\tau\in B(t;h)} K_{h_\mu}(T_{1j}-\tau).$$

According to Lemma S.4.1(a), we have $\mathbf{E}(H^2) = O(1+\frac{1}{mh_\mu})$ and thus

$$\sup_{\tau\in B(t;h)}|U-\mathbf{E}U| = O_p\left(\sqrt{\frac{1}{n}+\frac{1}{nmh_\mu}}\right)$$

according to Theorems 2.7.11 and 2.14.2 of van der Vaart and Wellner (1996). Lemma S.4.6 asserts that the last equation also holds for a deterministic design. With Lemma S.4.1 we deduce that

$$\sup_{\tau\in B(t;h)}\left|\hat{Q}_n(y,\tau)-F^*(y,\tau)\right| = O_p\left(h_\mu^2+\sqrt{\frac{1}{n}+\frac{1}{nmh_\mu}}\right).$$

A similar argument leads to

$$\sup_{\substack{d_{\mathcal{M}}(y_1,y_2)<\delta\\ \tau\in B(t;h)}}\left|\left(\hat{Q}_n(y_1,\tau)-\hat{Q}_n(y_2,\tau)\right)-\left(F^*(y_1,\tau)-F^*(y_2,\tau)\right)\right| = O_p\left(\delta h_\mu^2+\delta\sqrt{\frac{1}{n}+\frac{1}{nmh_\mu}}\right). \tag{S.2}$$

for any $y_1,y_2\in\mathcal{K}$ and $\delta>0$. Following from the argument in the proof of Lemma 2 in Petersen and Müller (2019), one can verify that for any $\kappa>0$

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\Pr\left\{\sup_{d_{\mathcal{M}}(y_1,y_2)<\delta,\tau\in B(t;h)}\left|\left(\hat{Q}_n(y_1,\tau)-\hat{Q}_n(y_2,\tau)\right)-\left(F^*(y_1,\tau)-F^*(y_2,\tau)\right)\right|>\kappa\right\}=0,$$

and further

$$\sup_{\tau\in B(t;h)} d_{\mathcal{M}}(\mu(\tau),\hat{\mu}(\tau)) = o_p(1) \tag{S.3}$$

given Assumption 4.5(b).

To derive the rate we apply (S.2) with $y_1=y$ and $y_2=\mu(\tau)$ to obtain

$$\sup_{\substack{d_{\mathcal{M}}(y,\mu(\tau))<\delta\\ \tau\in B(t;h)}}\left|\left(\hat{Q}_n(y,\tau)-\hat{Q}_n(\mu(\tau),\tau)\right)-\left(F^*(y,\tau)-F^*(\mu(\tau),\tau)\right)\right| = O_p\left(\delta h_\mu^2+\delta\sqrt{\frac{1}{n}+\frac{1}{nmh_\mu}}\right). \tag{S.4}$$

By (S.3), the the event $\{d_{\mathcal{M}}(\hat{\mu}(\tau),\mu(\tau))<\eta_1\}$ occurs with probability tending to one. On this event, according to Assumption 4.5(c), we have

$$F^*(\hat{\mu}(\tau),\tau)-F^*(\mu(\tau),\tau)-C_1 d_{\mathcal{M}}(\hat{\mu}(\tau),\mu(\tau))^2\geq 0.$$

Since $\hat{\mu}(\tau)$ is the minimizer of $\hat{Q}_n(y,\tau)$, we have $\hat{Q}_n(\mu(\tau),\tau)-\hat{Q}_n(\hat{\mu}(\tau),\tau)\geq 0$ and the following inequality on the event $\{d_{\mathcal{M}}(\hat{\mu}(\tau),\mu(\tau))<\eta_1\}$,

$$\left(F^*(\hat{\mu}(\tau),\tau)-F^*(\mu(\tau),\tau)\right)-\left(\hat{Q}_n(\hat{\mu}(\tau),\tau)-\hat{Q}_n(\mu(\tau),\tau)\right)\geq C_1 d_{\mathcal{M}}(\hat{\mu}(\tau),\mu(\tau))^2. \tag{S.5}$$

Let $a_n = h_\mu^2 + \sqrt{n^{-1} + (nmh_\mu)^{-1}}$. Below we fix an arbitrary $\epsilon > 0$, and find $M > 0$ accordingly to satisfy

$$\Pr\left\{\sup_{\tau\in B(t;h)} d_{\mathcal{M}}(\hat{\mu}(\tau), \mu(\tau)) > Ma_n\right\} \le \epsilon.$$

To this end, for $R > 0$ to be determined later, let

$$B_R(\delta) = \left\{\sup_{\substack{d_{\mathcal{M}}(y,\mu(\tau))<\delta \\ \tau\in B(t;h)}} \left|\left(\hat{Q}_n(y,\tau) - \hat{Q}_n(\mu(\tau),\tau)\right) - \left(F^*(y,\tau) - F^*(\mu(\tau),\tau)\right)\right| \le R\delta\left(h_\mu^2 + \sqrt{\frac{1}{n} + \frac{1}{nmh_\mu}}\right)\right\},$$

$$B_j = \{2^j Ma_n \le \sup_{\tau\in B(t;h)} d_{\mathcal{M}}(\hat{\mu}(\tau), \mu(\tau)) \le 2^{j+1}Ma_n\},$$

$$B_C = \{\sup_{\tau\in B(t;h)} d_{\mathcal{M}}(\hat{\mu}(\tau), \mu(\tau)) > \frac{1}{2}\eta_1\}.$$

Let $j_0 \ge 0$ be an integer satisfying $\frac{1}{2}\eta_1 < 2^{j_0+1}Ma_n \le \eta_1$. We then have

$$\begin{aligned}
&\Pr\left\{\sup_{\tau\in B(t;h)} d_{\mathcal{M}}(\hat{\mu}(\tau), \mu(\tau)) > Ma_n\right\} \le \sum_{j=0}^{j_0}\Pr\{B_j \cap B_R(2\eta_1)\} + \Pr\{B_C \cap B_R(2\eta_1)\} + \Pr\{\Omega\backslash B_R(2\eta_1)\}\\
&\le \sum_{j=0}^{j_0}\Pr\left(B_j \cap B_R(2\eta_1) \cap \left\{\sup_{\tau\in B(t;h)}\left|\left(F^*(\hat{\mu}(\tau),\tau) - F^*(\mu(\tau),\tau)\right) - \left(\hat{Q}_n(\hat{\mu}(\tau),\tau) - \hat{Q}_n(\mu(\tau),\tau)\right)\right| \ge C_1(2^jMa_n)^2\right\}\right)\\
&\quad + \Pr(B_C) + \Pr\{\Omega\backslash B_R(2\eta_1)\}\\
&\le \sum_{j=0}^{j_0}\Pr\left(1_{B_R(2^{j+1}Ma_n)} \sup_{\substack{d_{\mathcal{M}}(y,\mu(\tau))\le 2^{j+1}Ma_n \\ \tau\in B(t;h)}}\left|\left(F^*(y,\tau) - F^*(\mu(\tau),\tau)\right) - \left(\hat{Q}_n(y,\tau) - \hat{Q}_n(\mu(\tau),\tau)\right)\right| > C_1(2^jMa_n)^2\right)\\
&\quad + \Pr(B_C) + \Pr\{\Omega\backslash B_R(2\eta_1)\}\\
&\le R\sum_{j=0}^{j_0}\frac{2^{j+1}Ma_n^2}{C_1(2^jMa_n)^2} + \Pr(B_C) + \Pr\{\Omega\backslash B_R(2\eta_1)\} \le \frac{4R}{C_1M} + \Pr(B_C) + \Pr\{\Omega\backslash B_R(2\eta_1)\}.
\end{aligned}$$

Since $\lim_{n\to\infty}\Pr\{B_C\} = 0$ according to $\sup_{\tau\in B(t;h)} d_{\mathcal{M}}(\mu(\tau), \hat{\mu}(\tau)) = o_p(1)$ and $\lim_{R\to\infty}\liminf_{n\to\infty}\Pr\{\Omega\backslash B_R(2\eta_1)\} = 0$ according to (S.4), there exist $n_0$ and $R_0$ such that for any $n > n_0$,

$$\Pr\{\Omega\backslash B_{R_0}(2\eta_1)\} < \frac{\epsilon}{3} \quad \text{and} \quad \Pr\{B_C\} < \frac{\epsilon}{3}.$$

Taking $M = \frac{12R_0}{\epsilon C_1}$, we have

$$\Pr\left\{\sup_{\tau\in B(t;h)} d_{\mathcal{M}}(\hat{\mu}(\tau), \mu(\tau)) > Ma_n\right\} \le \frac{4R_0}{C_1M} + \Pr(B_C) + \Pr\{\Omega\backslash B_{R_0}(2\eta_1)\} \le \epsilon.$$

Therefore,

$$\lim_{M\to\infty}\limsup_{n\to\infty}\Pr\left\{\sup_{\tau\in B(t;h)} d_{\mathcal{M}}(\hat{\mu}(\tau), \mu(\tau)) > Ma_n\right\} = 0.$$

This yields the desired convergence rate of Proposition 4.2.

To establish the other proposition, by using the techniques in the proof of Lemma 5 of Zhang and Wang (2016) for the random and hybrid design, and Lemma S.4.6 combined with the above argument for the deterministic design, we can show that

$$\sup_{\tau\in\mathcal{T},y\in\mathcal{K}}\left|\hat{Q}_n(y,\tau)-F^*(y,\tau)\right| = O_p\left(h_\mu^2+\sqrt{\frac{1}{n}+\frac{1}{nmh_\mu}}\right)(\log n)^{1/2}.$$

Together with (S.5), this proves Proposition 4.1. □

*Proof of Theorems 4.1.* We divide the proof into three steps. In the first step, we identify two key terms that determine the convergence rate. In the second step and third step, we address the terms separately. Note that $v_1=\cdots=v_n=1/\{nm(m-1)\}$.

*Step 1.* Define $\gamma:=\hat{\mu}$ to simplify notation in the sequel. Since the parallel transport $\mathcal{P}^{(\mu(s),\mu(t))}_{(\gamma(s),\gamma(t))}$ preserves the fiber metric according to Theorem 2.4,

$$\mathscr{P}^{(\mu(s),\mu(t))}_{(\gamma(s),\gamma(t))}\hat{\mathcal{C}}(s,t)=\arg_{\beta_0}\min_{\beta_0,\beta_1,\beta_2\in\mathbb{L}(\mu(s),\mu(t))}\Bigg\{\sum_i\nu_i\sum_{j\neq k}\|\mathscr{P}^{(\mu(s),\mu(t))}_{(\gamma(s),\gamma(t))}\mathscr{P}^{(\gamma(s),\gamma(t))}_{(\gamma(T_{ij}),\gamma(T_{ik}))}\hat{\mathcal{C}}_{i,jk}$$
$$-\beta_0-\beta_1(T_{ij}-s)-\beta_2(T_{ik}-t)\|^2_{G_{(\mu(s),\mu(t))}}K_{h_\mathcal{C}}(s-T_{ij})K_{h_\mathcal{C}}(t-T_{ik})\Bigg\}.$$

Simple computation shows that

$$\begin{aligned}&\mathscr{P}^{(\mu(s),\mu(t))}_{(\gamma(s),\gamma(t))}\hat{\mathcal{C}}(s,t)-\mathcal{C}(s,t)\\&=\frac{(S_{20}S_{02}-S_{11}^2)R_{00}-(S_{10}S_{02}-S_{01}S_{11})R_{10}+(S_{10}S_{11}-S_{01}S_{20})R_{01}}{(S_{20}S_{02}-S_{11}^2)S_{00}-(S_{10}S_{02}-S_{01}S_{11})S_{10}+(S_{10}S_{11}-S_{01}S_{20})S_{01}},\end{aligned}\tag{S.6}$$

where $R_{ab}$ and $S_{ab}$ are defined as

$$\begin{aligned}R_{ab}&:=R_{ab}(s,t):=\sum_i\nu_i\sum_{j\neq k}\varpi(T_{ij},T_{ik})\left(\tfrac{T_{ij}-s}{h_\mathcal{C}}\right)^a\left(\tfrac{T_{ik}-t}{h_\mathcal{C}}\right)^b\left\{\mathscr{P}^{(\mu(s),\mu(t))}_{(\gamma(s),\gamma(t))}\mathscr{P}^{(\gamma(s),\gamma(t))}_{(\gamma(T_{ij}),\gamma(T_{ik}))}\hat{\mathcal{C}}_{i,jk}-\mathcal{C}(s,t)\right\},\\S_{ab}&:=S_{ab}(s,t):=\sum_i\nu_i\sum_{j\neq k}\varpi(T_{ij},T_{ik})\left(\tfrac{T_{ij}-s}{h_\mathcal{C}}\right)^a\left(\tfrac{T_{ik}-t}{h_\mathcal{C}}\right)^b,\end{aligned}\tag{S.7}$$

with $\varpi(s',t')=K_{h_\mathcal{C}}(s-s')K_{h_\mathcal{C}}(t-t')$ for $s',t'\in\mathcal{T}$ to further simplify notations.

Define $\tilde{\mathcal{C}}_{i,jk}:=\mathscr{P}^{(\mu(s),\mu(t))}_{(\mu(T_{ij}),\mu(T_{ik}))}(\mathrm{Log}_{\mu(T_{ij})}Y_{ij}\otimes\mathrm{Log}_{\mu(T_{ik})}Y_{ik})\in\mathbb{L}(\mu(s),\mu(t))$. We consider the decomposition

$$\begin{aligned}&\mathscr{P}^{(\mu(s),\mu(t))}_{(\gamma(s),\gamma(t))}\mathscr{P}^{(\gamma(s),\gamma(t))}_{(\gamma(T_{ij}),\gamma(T_{ik}))}\hat{\mathcal{C}}_{i,jk}-\mathcal{C}(s,t)\\&=\left\{\mathscr{P}^{(\mu(s),\mu(t))}_{(\gamma(s),\gamma(t))}\mathscr{P}^{(\gamma(s),\gamma(t))}_{(\gamma(T_{ij}),\gamma(T_{ik}))}\hat{\mathcal{C}}_{i,jk}-\tilde{\mathcal{C}}_{i,jk}\right\}+\left\{\tilde{\mathcal{C}}_{i,jk}-\mathcal{C}(s,t)\right\},\end{aligned}\tag{S.8}$$

according to which we split $R_{ab}$ into $R_{ab}=R_{ab,1}+R_{ab,2}$ with

$$\begin{aligned}R_{ab,1}&:=\sum_i\nu_i\sum_{j\neq k}\varpi(T_{ij},T_{ik})\left(\tfrac{T_{ij}-s}{h_\mathcal{C}}\right)^a\left(\tfrac{T_{ik}-t}{h_\mathcal{C}}\right)^b\left\{\mathscr{P}^{(\mu(s),\mu(t))}_{(\gamma(s),\gamma(t))}\mathscr{P}^{(\gamma(s),\gamma(t))}_{(\gamma(T_{ij}),\gamma(T_{ik}))}\hat{\mathcal{C}}_{i,jk}-\tilde{\mathcal{C}}_{i,jk}\right\},\\R_{ab,2}&:=\sum_i\nu_i\sum_{j\neq k}\varpi(T_{ij},T_{ik})\left(\tfrac{T_{ij}-s}{h_\mathcal{C}}\right)^a\left(\tfrac{T_{ik}-t}{h_\mathcal{C}}\right)^b\left\{\tilde{\mathcal{C}}_{i,jk}-\mathcal{C}(s,t)\right\}.\end{aligned}$$

Combining (S.6) and (S.8), we deduce that

$$\begin{aligned}
&\mathscr{P}_{(\gamma(s),\gamma(t))}^{(\mu(s),\mu(t))}\hat{\mathcal{C}}(s,t)-\mathcal{C}(s,t)\\
=&\frac{(S_{20}S_{02}-S_{11}^2)[R_{00,1}+R_{00,2}-\partial_s\mathcal{C}(s,t)h_{\mathcal{C}}S_{10}-\partial_t\mathcal{C}(s,t)h_{\mathcal{C}}S_{01}]}{(S_{20}S_{02}-S_{11}^2)S_{00}-(S_{10}S_{02}-S_{01}S_{11})S_{10}+(S_{10}S_{11}-S_{01}S_{20})S_{01}}\\
-&\frac{(S_{10}S_{02}-S_{01}S_{11})[R_{10,1}+R_{10,2}-\partial_s\mathcal{C}(s,t)h_{\mathcal{C}}S_{20}-\partial_t\mathcal{C}(s,t)h_{\mathcal{C}}S_{11}]}{(S_{20}S_{02}-S_{11}^2)S_{00}-(S_{10}S_{02}-S_{01}S_{11})S_{10}+(S_{10}S_{11}-S_{01}S_{20})S_{01}}\\
+&\frac{(S_{10}S_{11}-S_{01}S_{20})[R_{01,1}+R_{01,2}-\partial_s\mathcal{C}(s,t)h_{\mathcal{C}}S_{11}-\partial_t\mathcal{C}(s,t)h_{\mathcal{C}}S_{02}]}{(S_{20}S_{02}-S_{11}^2)S_{00}-(S_{10}S_{02}-S_{01}S_{11})S_{10}+(S_{10}S_{11}-S_{01}S_{20})S_{01}}.
\end{aligned}\tag{S.9}$$

In light of Lemmas S.4.2 and S.4.3, the convergence rate of (S.6) depends on $R_{ab,1}$ and $R_{ab,2}-\partial_s\mathcal{C}(s,t)h_{\mathcal{C}}S_{a+1,b}-\partial_t\mathcal{C}(s,t)h_{\mathcal{C}}S_{a,b+1}$.

*Step 2.* In this step we address $R_{ab,1}$. According to the definition of $\mathscr{P}$ in (5), the first part in Equation (S.8) is

$$\big(\mathcal{P}_{\gamma(s)}^{\mu(s)}\mathcal{P}_{\gamma(T_{ij})}^{\gamma(s)}\mathrm{Log}_{\gamma(T_{ij})}Y_{ij}\big)\otimes\big(\mathcal{P}_{\gamma(t)}^{\mu(t)}\mathcal{P}_{\gamma(T_{ik})}^{\gamma(t)}\mathrm{Log}_{\gamma(T_{ik})}Y_{ik}\big)-\big(\mathcal{P}_{\mu(T_{ij})}^{\mu(s)}\mathrm{Log}_{\mu(T_{ij})}Y_{ij}\big)\otimes\big(\mathcal{P}_{\mu(T_{ik})}^{\mu(t)}\mathrm{Log}_{\mu(T_{ik})}Y_{ik}\big).$$

Then according to Assumption 4.4(b) and Lemma 4.1, its rate is

$$\left\|\mathscr{P}_{(\gamma(s),\gamma(t))}^{(\mu(s),\mu(t))}\mathscr{P}_{(\gamma(T_{ij}),\gamma(T_{ik}))}^{(\gamma(s),\gamma(t))}\hat{\mathcal{C}}_{i,jk}-\tilde{\mathcal{C}}_{i,jk}\right\|_G=O\left(\sup_{\tau:|\tau-s|<h_{\mathcal{C}}\text{ or }|\tau-t|<h_{\mathcal{C}}}d_{\mathcal{M}}(\gamma(\tau),\mu(\tau))\right).$$

By Proposition 4.2, we conclude that

$$R_{ab,1}=O_p\left(h_\mu^2+\sqrt{\tfrac{1}{n}+\tfrac{1}{nmh_\mu}}\right).$$

*Step 3.* In this step, we first analyze the term $R_{00,2}-\partial_s\mathcal{C}(s,t)h_{\mathcal{C}}S_{10}-\partial_t\mathcal{C}(s,t)h_{\mathcal{C}}S_{01}$ in (S.9), which equals to

$$U:=\sum_i\nu_i\sum_{j\neq k}\varpi(T_{ij},T_{ik})\left\{\tilde{\mathcal{C}}_{i,jk}-\mathcal{C}(s,t)-\partial_s\mathcal{C}(s,t)(T_{ij}-s)-\partial_t\mathcal{C}(s,t)(T_{ik}-t)\right\}.$$

We start with bounding its mean. Let $\mathbb{T}=\{T_{ij}:i=1,\dots,n,\ j=1,\dots,m_i\}$ and observe that

$$\mathbf{E}\big(\tilde{\mathcal{C}}_{i,jk}\mid\mathbb{T}\big)=\mathscr{P}_{(\mu(T_{ij}),\mu(T_{ik}))}^{(\mu(s),\mu(t))}\mathcal{C}(T_{ij},T_{ik}).$$

340 In addition, since $\mathcal{C}$ is twice differentiable and the parallel transport $\mathscr{P}$ is depicted by a partial differential equation, we have the following Taylor expansion at $(s,t)$,

$$\mathscr{P}_{(\mu(T_{ij}),\mu(T_{ik}))}^{(\mu(s),\mu(t))}\mathcal{C}(T_{ij},T_{ik})=\mathcal{C}(s,t)+\partial_s\mathcal{C}(s,t)(T_{ij}-s)+\partial_t\mathcal{C}(s,t)(T_{ik}-t)+O(h_{\mathcal{C}}^2)\tag{S.10}$$

for all $T_{ij},T_{ik}$ such that $|T_{ij}-s|<h_{\mathcal{C}}$ and $|T_{ik}-t|<h_{\mathcal{C}}$, where $O(h_{\mathcal{C}}^2)$ is uniform over all $T_{ij}$ and $T_{ik}$ due to Assumption 4.6 and the compactness of $\mathcal{K}$. Then we further deduce that

$$\begin{aligned}
&\mathbf{E}(U)\\
=&\mathbf{E}\left\{\mathbf{E}\left[\sum_i\nu_i\sum_{j\neq k}\varpi(T_{ij},T_{ik})\left\{\tilde{\mathcal{C}}_{i,jk}-\mathcal{C}(s,t)-\partial_s\mathcal{C}(s,t)(T_{ij}-s)-\partial_t\mathcal{C}(s,t)(T_{ik}-t)\right\}\middle|\mathbb{T}\right]\right\}\\
=&\mathbf{E}\left\{\sum_i\nu_i\sum_{j\neq k}\varpi(T_{ij},T_{ik})\times O(h_{\mathcal{C}}^2)\right\}=O(h_{\mathcal{C}}^2).
\end{aligned}$$

For the random and hybrid designs, the i.i.d assumption on trajectories and Lemma S.4.2 imply that

$$\mathbf{E}\|U-\mathbf{E}U\|_G^2 \le \frac{1}{n^2}\sum_{i=1}^n \mathbf{E}\left\|\frac{1}{m(m-1)}\sum_{j\neq k}\varpi(T_{ij},T_{ik})\{\tilde{\mathcal{C}}_{i,jk}-\mathscr{P}^{(\mu(s),\mu(t))}_{(\mu(T_{ij}),\mu(T_{ik}))}\mathcal{C}(T_{ij},T_{ik})\}\right\|_G^2$$
$$\le \frac{4\mathrm{diam}(\mathcal{M})^4}{n}\mathbf{E}\left|\frac{1}{m(m-1)}\sum_{j\neq k}\varpi(T_{ij},T_{ik})\right|^2 = O\left(\frac{1}{n}+\frac{1}{nm^2h_{\mathcal{C}}^2}\right).$$

Lemma S.4.7 asserts that this also holds for the deterministic design. Combining this with $\mathbf{E}U=O(h_{\mathcal{C}}^2)$, we deduce that $\mathbf{E}\|U-\mathbf{E}U\|_G^2=O(n^{-1}+n^{-1}m^{-2}h_{\mathcal{C}}^{-2})$, and with Markov inequality, further conclude that $R_{00,2}-\partial_s\mathcal{C}(s,t)h_{\mathcal{C}}S_{10}-\partial_t\mathcal{C}(s,t)h_{\mathcal{C}}S_{01}=O_p(h_{\mathcal{C}}^2+n^{-1/2}+n^{-1/2}m^{-1}h_{\mathcal{C}}^{-1})$.

Similar arguments can show that the terms $R_{10,2}-\partial_s\mathcal{C}(s,t)h_{\mathcal{C}}S_{20}-\partial_t\mathcal{C}(s,t)h_{\mathcal{C}}S_{11}$ and $R_{01,2}-\partial_s\mathcal{C}(s,t)h_{\mathcal{C}}S_{11}-\partial_t\mathcal{C}(s,t)h_{\mathcal{C}}S_{02}$ in (S.9) are of the same order. The equation (10) is then obtained by inserting the results in Steps 2 and 3 into Step 1. □

*Proof of Theorem 4.2.* Similar to the proof of Theorem 4.1, we only need to consider the uniform rate of the term $R_{00,1}+R_{00,2}-\partial_s\mathcal{C}(s,t)h_{\mathcal{C}}S_{10}-\partial_t\mathcal{C}(s,t)h_{\mathcal{C}}S_{01}$ in (S.9).

Due to boundedness of $\mathcal{K}$ and Lemma 4.1, we have

$$\sup_{s,t}\|\mathcal{P}^{(\mu(s),\mu(t))}_{(\gamma(s),\gamma(t))}\mathcal{P}^{(\gamma(s),\gamma(t))}_{(\gamma(T_{ij}),\gamma(T_{ik}))}\hat{\mathcal{C}}_{i,jk}-\tilde{\mathcal{C}}_{i,jk}\|_G \le C\sup_\tau d_{\mathcal{M}}(\gamma(\tau),\mu(\tau)).$$

Therefore, according to Lemma S.4.2 and Proposition 4.1, we deduce that

$$\sup_{s,t}\|R_{00,1}\|_G \le c\left(\sup_{s,t}|S_{00}|\right)\left(\sup_\tau d_{\mathcal{M}}(\gamma(\tau),\mu(\tau)\right)=O_p\left(h_\mu^2+\sqrt{\frac{\log n}{nmh_\mu}+\frac{\log n}{n}}\right)$$

for a universal constant $c>0$ depending only on $\mathcal{K}$.

The uniform convergence rates of $R_{00,2}-\partial_s\mathcal{C}(s,t)h_{\mathcal{C}}S_{10}-\partial_t\mathcal{C}(s,t)h_{\mathcal{C}}S_{01}$ and other similar terms are obtained by arguments similar to those in Theorem 5.2 of Zhang and Wang (2016), except that no truncation argument is needed due to Assumption 4.4(b), and moments of some random quantities are calculated by using the techniques in Lemma S.4.2 for the hybrid design and by using Lemma S.4.7 for the deterministic design. □

## S.4 Technical Lemmas

The following lemma is used to establish the convergence rate of the mean estimator under the random or hybrid design.

**Lemma S.4.1** (mean, random)**.** *Suppose that Assumptions 2.1, 2.2, 3.1, 4.4, 4.5. Under either Assumption 4.1 or Assumption 4.3, if $h_\mu\to 0$ and $nmh_\mu\to\infty$, then for any $t$ and $h=O(h_\mu)$, we have*

(a) $\mathbf{E}\left|\frac{1}{m}\sum_j\sup_{\tau\in B(t;h)}K_{h_\mu}(T_{1j}-\tau)\right|^2=O\left(1+\frac{1}{mh_\mu}\right)$;

(b) $\sup_{\tau\in\mathcal{T}}|\hat{u}_k(\tau)|=O_p(h_\mu^k)$ *for* $k=0,1,2$;

(c) $\inf_{\tau\in B(t;h)}|\hat{\sigma}_0^2(\tau)|\asymp h_\mu^2(1+o_P(1))$.

*Proof of Lemma S.4.1.* Under either Assumption 4.1 or Assumption 4.3, $T_{i1},\dots,T_{im}$ are identically distributed (since they are exchangeable), and we deduce that

$$\sup_{\tau\in B(t;h)}|\mathbf{E}\hat{u}_k| = \sup_{\tau\in B(t;h)}\left|\frac{1}{nm}\sum_{ij}\mathbf{E}[K_{h_\mu}(T_{ij}-\tau)(T_{ij}-\tau)^k]\right| = \sup_{\tau\in B(t;h)}|\mathbf{E}K_{h_\mu}(T_{11}-\tau)(T_{11}-\tau)^k| = O(h_\mu^k).$$

Define an envelop function

$$H_k := \frac{1}{m}\sum_j \sup_{\tau\in B(t;h)}\left|K_{h_\mu}(T_{1j}-\tau)(T_{1j}-\tau)^k\right|$$

for $\hat{u}_k$. Under Assumption 4.1, the second moment of $H_k$ is

$$\begin{aligned}
\mathbf{E}(H_k^2) =& \frac{1}{m^2}\sum_{j_1,j_2}\left\{\mathbf{E}\sup_{\tau\in B(t;h)}\left|K_{h_\mu}(T_{1j_1}-\tau)(T_{1j_1}-\tau)^k\right|\times\sup_{\tau\in B(t;h)}\left|K_{h_\mu}(T_{1j_2}-\tau)(T_{1j_2}-\tau)^k\right|\right\}\\
=&\frac{1}{m}\mathbf{E}\sup_{\tau\in B(t;h)}|K_{h_\mu}(T_{1j_1}-\tau)(T_{1j_1}-\tau)^k|^2\\
&+\frac{m-1}{m}\mathbf{E}\Big\{\sup_{\tau\in B(t;h)}\left|K_{h_\mu}(T_{11}-\tau)(T_{11}-\tau)^k\right|\Big\}\times\mathbf{E}\Big\{\sup_{\tau\in B(t;h)}\left|K_{h_\mu}(T_{12}-\tau)(T_{12}-\tau)^k\right|\Big\}\\
=&O\left(h_\mu^{2k}(1+\frac{1}{mh_\mu})\right).
\end{aligned}$$

Under Assumption 4.3,

$$\begin{aligned}
\mathbf{E}(H_k^2) =& \frac{1}{m^2}\sum_{j_1,j_2}\left\{\mathbf{E}\sup_{\tau\in B(t;h)}\left|K_{h_\mu}(T_{1j_1}-\tau)(T_{1j_1}-\tau)^k\right|\times\sup_{\tau\in B(t;h)}\left|K_{h_\mu}(T_{1j_2}-\tau)(T_{1j_2}-\tau)^k\right|\right\}\\
=&\frac{1}{m}\mathbf{E}\sup_{\tau\in B(t;h)}|K_{h_\mu}(T_{1j_1}-\tau)(T_{1j_1}-\tau)^k|^2\\
&+\frac{m-1}{m}\mathbf{E}\Big\{\sup_{\tau\in B(t;h)}\left|K_{h_\mu}(T_{11}-\tau)(T_{11}-\tau)^k\right|\times\sup_{\tau\in B(t;h)}\left|K_{h_\mu}(T_{12}-\tau)(T_{12}-\tau)^k\right|\Big\}\\
\le&O\Big(\frac{h_\mu^{2k-1}}{m}\Big)+\mathbf{E}\Big[\mathbf{E}\Big\{\sup_{\tau\in B(t;h)}\left|K_{h_\mu}(T_{11}-\tau)(T_{11}-\tau)^k\right| \mid S_{11},S_{12}\Big\}\\
&\times\mathbf{E}\Big\{\sup_{\tau\in B(t;h)}\left|K_{h_\mu}(T_{12}-\tau)(T_{12}-\tau)^k\right|\mid S_{11},S_{12}\Big\}\Big]\\
\le&O\Big(\frac{h_\mu^{2k-1}}{m}\Big)+O(h_\mu^{2k-2})\mathbf{E}\Big[\mathbf{E}\Big\{1_{t-S_{11}-O(h_\mu)\le\zeta_{11}\le t-S_{11}+O(h_\mu)}\mid S_{11}\Big\}\mathbf{E}\Big\{1_{t-S_{12}-O(h_\mu)\le\zeta_{12}\le t-S_{12}+O(h_\mu)}\mid S_{12}\Big\}\Big].
\end{aligned}$$

When $h_\mu\lesssim L^{-1}$, $\mathbf{E}\big\{1_{t-S_{11}-O(h_\mu)\le\zeta_{11}\le t-S_{11}+O(h_\mu)}\mid S_{11}\big\}$ is of order $O(hL)$ when $|S_{11}-t|=O(L^{-1})$ and zero otherwise, an similar observation applies to $\mathbf{E}\big\{1_{t-S_{12}-O(h_\mu)\le\zeta_{12}\le t-S_{12}+O(h_\mu)}\mid S_{12}\big\}$. Together, they imply that

$$\begin{aligned}
&\mathbf{E}\Big[\mathbf{E}\Big\{1_{t-S_{11}-O(h_\mu)\le\zeta_{11}\le t-S_{11}+O(h_\mu)}\mid S_{11}\Big\}\mathbf{E}\Big\{1_{t-S_{11}-O(h_\mu)\le\zeta_{11}\le t-S_{11}+O(h_\mu)}\mid S_{12}\Big\}\Big]\\
&=O(h_\mu^2L^2)\mathbf{E}\{1_{|S_{11}-t|=O(L^{-1})}1_{|S_{12}-t|=O(L^{-1})}\}=O(h_\mu^2L^2)O(L^{-2})=O(h_\mu^2).
\end{aligned}$$

When $h_\mu\gtrsim L^{-1}$, $\mathbf{E}\big\{1_{t-S_{11}-O(h_\mu)\le\zeta_{11}\le t-S_{11}+O(h_\mu)}\mid S_{11}\big\}$ is of order $O(1)$ when $|S_{11}-t|=O(h_\mu)$ and zero otherwise, an similar observation applies to $\mathbf{E}\big\{1_{t-S_{12}-O(h_\mu)\le\zeta_{12}\le t-S_{12}+O(h_\mu)}\mid S_{12}\big\}$. Together, they imply that

$$\begin{aligned}
&\mathbf{E}\Big[\mathbf{E}\Big\{1_{t-S_{11}\le\zeta_{11}\le t-S_{11}+2h}\mid S_{11}\Big\}\mathbf{E}\Big\{1_{t-S_{12}\le\zeta_{12}\le t-S_{12}+2h}\mid S_{12}\Big\}\Big]\\
&=O(1)\mathbf{E}\{1_{|S_{11}-t|=O(h_\mu)}1_{|S_{12}-t|=O(L^{-1})}\}=O(1)O(h_\mu^2)=O(h_\mu^2).
\end{aligned}$$

In summary, we still have $\mathbf{E}H_k^2 = O\left(h_\mu^{2k}(1+\frac{1}{mh_\mu})\right)$ under Assumption 4.3. Part (a) is then verified by taking $k=0$ in the above.

Part (b) can be proved by an argument analogous to the proof for Lemma 4 of Zhang and Wang (2016). For part (c), it is seen that $\hat\sigma_0(\tau) \asymp \{\mathbf{E}\hat u_0 \mathbf{E}\hat u_2 - (\mathbf{E}\hat u_1)^2\}(1+o_P(1))$, where the $o_P(1)$ component is uniform over $\tau$. Define $V := K_{h_\mu}(T_{11}-\tau)$ and $W := \mathbf{E}V \asymp 1$. Simple calculation shows that

$$\mathbf{E}\hat u_0 \mathbf{E}\hat u_2 - (\mathbf{E}\hat u_1)^2 = W\mathbf{E}\big(V[(T_{11}-\tau) - W^{-1}\mathbf{E}\{V(T_{11}-\tau)\}]^2\big) \asymp h_\mu^2$$

uniformly over all $\tau \in \mathcal{T}$. □

The following lemma is used to establish Theorems 4.1 and 4.2 under the random or hybrid design. Its proof is similar to that for Lemma S.4.1 and thus is omitted.

**Lemma S.4.2** (covariance, random)**.** *Suppose that Assumptions 2.1, 2.2, 3.1, 4.4 and 4.5. Under either of additional Assumptions 4.1 and 4.3, if $h_\mathcal{C} \to 0$ and $nm^2h_\mathcal{C}^2 \to \infty$, we have*

$$\begin{aligned}
&\sup_{s,t\in\mathcal{T}} \mathbf{E}\{S_{ab}(s,t)\} = O(1),\\
&\sup_{s,t\in\mathcal{T}} |S_{ab}(s,t) - \mathbf{E}\{S_{ab}(s,t)\}| = o_P(1),\\
&\inf_{s,t\in\mathcal{T}} \{(S_{20}S_{02}-S_{11}^2)S_{00} - (S_{10}S_{02}-S_{01}S_{11})S_{10} + (S_{10}S_{11}-S_{01}S_{20})S_{01}\} \asymp 1 + o_P(1),\\
&\sup_{s,t\in\mathcal{T}} \mathbf{E}\left|\frac{1}{m(m-1)}\sum_{j\neq k} K_{h_\mathcal{C}}(s-T_{ij})K_{h_\mathcal{C}}(t-T_{ik})\right|^2 = O\left(1+\frac{1}{m^2h_\mathcal{C}^2}\right).
\end{aligned}$$

The next lemma is used to prove the convergence rates of the mean and covariance estimators under the deterministic design.

**Lemma S.4.3** (mean and covariance, deterministic)**.** *Suppose that Assumptions 4.4(c)(d) and 4.2 hold, and $K$ is decreasing on $[0,1]$. If $nmh_\mu \to \infty$ and $h \asymp h_\mu$, then*

(a) $\sup_{\tau\in\mathcal{T}} |\hat u_k(\tau)| = O(h_\mu^k)$ *for $k = 0,1,2$;*

(b) $\sup_{\tau\in\mathcal{T}} |\hat u_k(\tau)| \asymp h_\mu^k$ *for $k=0,2$;*

(c) $\inf_{\tau\in\mathcal{T}} |\hat\sigma_0^2(\tau)| \asymp h_\mu^2$.

*If $nm^2h_\mathcal{C}^2 \to \infty$, then*

(d) $\sup_{s,t\in\mathcal{T}} S_{ab}(s,t) = O(1)$;

(e) $\inf_{s,t\in\mathcal{T}}\{(S_{20}S_{02}-S_{11}^2)S_{00} - (S_{10}S_{02}-S_{01}S_{11})S_{10} + (S_{10}S_{11}-S_{01}S_{20})S_{01}\} \asymp 1$;

(f) $\sup_{s,t\in\mathcal{T}} \sum_{i,j\neq k} \nu_i K_{h_\mathcal{C}}(s-T_{ij})K_{h_\mathcal{C}}(t-T_{ik}) \left|\frac{T_{ij}-s}{h_\mathcal{C}}\right|^a \left|\frac{T_{ik}-t}{h_\mathcal{C}}\right|^b \asymp 1$,

*where $S_{ab}$ is defined in (S.7).*

*Proof.* Part (a) can be verified by simple calculation. For part (b), we fix $\tau \in \mathcal{T}$ and let $W = \sum_{ij} K\big(\frac{T_{ij}-\tau}{h_\mu}\big)$. The assumptions on the kernel function imply that $K(u) \geq c_0$ on $[-3/4, 3/4]$ for some constant $c_0 > 0$ depending only on $K$. In the sequel, we assume $n$ is sufficiently large so that $nmh_\mu \gg 1$. Assumption 4.2 implies that there are at least $c_1 nmh_\mu/2$ points within the interval $[\tau - 3h_\mu/4, \tau + 3h_\mu/4]$, from which we deduce that $W \geq c_0 c_1 nmh_\mu/2 > 0$ regardless of the location of $\tau$. Let $w_{ij} = K\left(\frac{T_{ij}-\tau}{h_\mu}\right)/W$, which is well defined. Observe that

$$\hat{u}_k(\tau) = \frac{W}{nmh_\mu}\sum_{ij} w_{ij}(T_{ij}-\tau)^k.$$

According the assumptions on the kernel function and Assumption 4.2, at least $c_1 nmh_\mu/4$ of the pairs $(i,j)$ satisfy $|T_{ij}-\tau| \geq h_\mu/8$ and $w_{ij} \geq c_0/W$. Thus,

$$\hat{u}_k(\tau) \geq \frac{W}{nmh_\mu}\frac{c_1 nmh_\mu}{4}\frac{c_0}{W}\frac{h_\mu^k}{8^k} \geq \frac{c_0 c_1}{2^{3k+2}}h_\mu^k$$

regardless of the value of $\tau$. Combining this with the first statement we prove the second statement. The last statement can be established in a similar fashion.

To establish part (c), let $E = \sum_{ij} w_{ij}T_{ij}$. As $w_{ij}$ is nonzero if and only if $T_{ij} \in (\tau - h_\mu, \tau + h_\mu)$ and $\sum_{ij} w_{ij} = 1$, we have $E \in [\tau - h_\mu, \tau + h_\mu]$. We then observe that

$$\begin{aligned}
\hat{\sigma}_0^2(\tau) &= \frac{W^2}{(nmh_\mu)^2}\left\{\sum_{ij} w_{ij}(T_{ij}-\tau)^2\right\} - \frac{W^2}{(nmh_\mu)^2}\left\{\sum_{ij} w_{ij}(T_{ij}-\tau)\right\}^2 \\
&= \frac{W^2}{(nmh_\mu)^2}\left[\sum_{ij} w_{ij}\left\{(T_{ij}-\tau) - \sum_{i'j'} w_{i'j'}(T_{i'j'}-\tau)\right\}^2\right] \\
&= \frac{W^2}{(nmh_\mu)^2}\left\{\sum_{ij} w_{ij}(T_{ij}-E)^2\right\}.
\end{aligned}$$

According to Assumption 4.2, there are at least $c_1 nmh_\mu/4$ of $T_{ij}$ such that $|T_{ij}-E| \geq h_\mu/8$ and $w_{ij} \geq c_0/W$. This implies that

$$\hat{\sigma}_0^2(\tau) \geq \frac{c_0 c_1 h_\mu^2}{256}\frac{W}{nmh_\mu} \geq \frac{c_0^2 c_1^2}{512}h_\mu^2$$

regardless of the value of $\tau$, where the last inequality is due to $W \geq c_0 c_1 nmh_\mu/2$ that we have deduced previously.

The other statements can be established by similar arguments. □

An $\epsilon$-cover of a subset $S$ of a pseudo-metric space $(\Omega, d)$ is a subset $A \subset S$ such that for each $p \in S$ there exists a $q \in A$ such that $d(p,q) \leq \epsilon$. We define $N(\epsilon, S, d) = \min\{|A| : A \text{ is an } \epsilon \text{ cover of } S\}$ to be the $\epsilon$-covering number of $S$, where $|A|$ denotes the cardinality of the set $A$. An $\epsilon$-packing of $S$ is a subset $A \subset S$ such that $d(p,q) > \epsilon$ for $p,q \in A$. The $\epsilon$-packing number of $S$ is defined by $M(\epsilon, S, d) = \max\{|A| : A \text{ is an } \epsilon \text{ packing of } S\}$. A standard relation between $\epsilon$-covering number and $\epsilon$-packing number is $M(2\epsilon, S, d) \leq N(\epsilon, S, d) \leq M(\epsilon, S, d)$ for all $\epsilon > 0$.

**Lemma S.4.4.** *Let $(S_1, d_1)$ and $(S_2, d_2)$ be two pseudo-metric spaces and $(S_1 \times S_2, d_1 \times d_2)$ the product pseudo-metric space with the pseudo-metric $(d_1 \times d_2)(p_1 \times p_2, q_1 \times q_2) = \{d_1^2(p_1,q_1) + d_2^2(p_2,q_2)\}^{1/2}$ for $p_1 \times p_2, q_1 \times q_2 \in S_1 \times S_2$. Then $N(\epsilon, S_1 \times S_2, d_1 \times d_2) \leq N(\epsilon/\sqrt{2}, S_1, d_1)N(\epsilon/\sqrt{2}, S_2, d_2)$.*

*Proof of Lemma S.4.4.* Let $A_1$ and $A_2$ be an $\epsilon/\sqrt{2}$-cover of $A_1$ and $A_2$, respectively. For each $k = 1, 2$, for every $p_k \in S_k$ there exists $p_k' \in A_k$ such that $d_k(p_k, p_k') \leq \epsilon/\sqrt{2}$. Then for each $p_1 \times p_2 \in S_1 \times S_2$, we have $(d_1 \times d_2)(p_1 \times p_2, p_1' \times p_2') = \{d_1^2(p_1, p_1') + d_2^2(p_2, p_2')\}^{1/2} \leq \epsilon$. This shows that $A = \{p_1' \times p_2' : p_1' \in A_1, p_2' \in A_2\}$ is an $\epsilon$-cover. The conclusion of the lemma then follows from the observation $|A| = N(\epsilon/\sqrt{2}, S_1, d_1)N(\epsilon/\sqrt{2}, S_2, d_2)$. □

**Lemma S.4.5.** *Let* $d_h(y\times t, z\times s) := \{h^{-2}|s-t|^2 + d^2_{\mathcal{M}}(y,z)\}^{1/2}$ *be a distance on the product space* $\mathcal{M}\times\mathcal{T}$*, and* $c>0$ *a constant. Then we have* $\sup_t \operatorname{diam}(\mathcal{K}\times B(t;ch)) \le \sqrt{4c^2+\operatorname{diam}^2(\mathcal{K})}$*. In addition, for all sufficiently small* $\epsilon>0$, $\sup_t N(\epsilon, \mathcal{K}\times B(t;ch), d_h\times d_{\mathcal{M}}) \le c_0\epsilon^{-d-1}$*, where* $d$ *is the dimension of* $\mathcal{M}$ *and* $c_0$ *is a constant depending on* $c$ *and* $\mathcal{K}$*.*

*Proof.* Given Lemma S.4.4, it is sufficient to show that $N(\epsilon,\mathcal{K},d_{\mathcal{M}}) \le c_1\epsilon^{-d}$ and $N(\epsilon, B(t;ch), d_h) \le c_2\epsilon^{-1}$ for some constants $c_1, c_2 > 0$. The compactness implies that $\mathcal{K}$ has bounded sectional curvature. Bishop–Günther inequality (Gray, 2012) implies that $M(\epsilon,\mathcal{K},d_{\mathcal{M}}) \le c_1\epsilon^{-d}$. Note that the space $(B(t;ch), d_B)$, with $d_B(s_1,s_2) = h^{-1}|s_1-s_2|$ for all $s_1,s_2\in B(t;ch)$, is isometric to the interval $[-c,c]$ endowed with the standard distance $d_E(s_1,s_2)| = |s_1-s_2|$, which implies that $N(\epsilon, B(t;ch), d_B) = N(\epsilon,[-c,c],d_E) \le c_2\epsilon^{-1}$. □

The following lemma is used to establish the convergence rate of the mean estimator under the deterministic design.

**Lemma S.4.6** (mean, deterministic)**.** *Suppose that Assumptions 2.1, 2.2, 3.1, 4.4, 4.5 and 4.2 hold. Let*

$$U(y,\tau) = \frac{1}{nm}\sum_{ij} K_{h_\mu}(T_{ij}-\tau)\left(d^2_{\mathcal{M}}(Y_{ij},y) - F^*(y,\tau) - \partial_\tau F^*(y,\tau)(T_{ij}-\tau)\right).$$

*Then, if* $h_\mu\to 0$ *and* $nmh_\mu\to\infty$*, then for any deterministic* $t\in\mathcal{T}$ *and* $h = O(h_\mu)$*, for all sufficiently small* $h_\mu$*,*

$$\mathbf{E}\left\{\sup_{\substack{y\in\mathcal{K}\\ \tau\in B(t;h)}}\left|U(y,\tau)-\mathbf{E}U(y,\tau)\right|\right\} = O\left(n^{-1/2} + (nmh_\mu)^{-1/2}\right), \tag{S.11}$$

$$\mathbf{E}\left\{\sup_{\substack{d(y_1,y_2)<\delta\\ \tau\in B(t;h)}}\left|\{U(y_1,\tau)-\mathbf{E}U(y_1,\tau)\} - \{U(y_2,\tau)-\mathbf{E}U(y_2,\tau)\}\right|\right\} = O\left(\delta n^{-1/2} + \delta(nmh_\mu)^{-1/2}\right), \tag{S.12}$$

*where* $\delta>0$ *is a constant. In addition, if* $h_\mu\to 0$, $nh_\mu \gtrsim 1$ *and* $nmh_\mu/\log n\to\infty$*, then*

$$\mathbf{E}\left\{\sup_{\substack{y\in\mathcal{K}\\ \tau\in\mathcal{T}}}\left|U(y,\tau)-\mathbf{E}U(y,\tau)\right|\right\} = O\left(n^{-1/2} + (nmh_\mu)^{-1/2}\right)(\log n)^{1/2}, \tag{S.13}$$

$$\mathbf{E}\left\{\sup_{\substack{d(y_1,y_2)<\delta\\ \tau\in\mathcal{T}}}\left|\{U(y_1,\tau)-\mathbf{E}U(y_1,\tau)\} - \{U(y_2,\tau)-\mathbf{E}U(y_2,\tau)\}\right|\right\} = O\left(\delta n^{-1/2} + \delta(nmh_\mu)^{-1/2}\right)(\log n)^{1/2}. \tag{S.14}$$

*Proof.* To simplify notation, the symbol $c$ below, which denotes a constant not depending on $n, m, h_\mu$, $\tau$, $t$ but maybe depending on other constants such as $\operatorname{diam}(\mathcal{K})$, $\sup_{u\in[-1,1]} K(u)$, Lipschitz constant of $K$, etc, will often be re-used potentially with different values at each occurrence. Below we prove (S.11) and (S.13); the proofs for (S.12) and (S.14) are similar and thus omitted.

We first consider the case $mh_\mu \gtrsim 1$. Let

$$V_i(y,\tau) = \frac{1}{mh_\mu}\sum_{j=1}^m K\left(\frac{T_{ij}-\tau}{h_\mu}\right)\left(d^2_{\mathcal{M}}(Y_{ij},y) - F^*(y,T_{ij})\right) \tag{S.15}$$

and

$$Z_n(y,\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i(y,\tau). \tag{S.16}$$

Then $\mathbf{E}V_i(y,\tau) = 0$ and $U(y,\tau) - \mathbf{E}U(y,\tau) = n^{-1/2} Z_n(y,\tau)$. Now we observe that

$$\begin{aligned}
|V_i(y,\tau_1) - V_i(z,\tau_2)| \le & \frac{1}{mh_\mu} \left| \sum_{j=1}^{m} \left\{ K\left(\tfrac{T_{ij}-\tau_1}{h_\mu}\right) - K\left(\tfrac{T_{ij}-\tau_2}{h_\mu}\right) \right\} \left( d_{\mathcal{M}}^2(Y_{ij},y) - F^*(y,T_{ij}) \right) \right| \\
& + \frac{1}{mh_\mu} \left| \sum_{j=1}^{m} K\left(\tfrac{T_{ij}-\tau_2}{h_\mu}\right) \left( d_{\mathcal{M}}^2(Y_{ij},y) - F^*(y,T_{ij}) - d_{\mathcal{M}}^2(Y_{ij},z) + F^*(z,T_{ij}) \right) \right| \\
\le & \frac{c}{mh_\mu} \left( \frac{|\tau_2-\tau_1|}{h_\mu} + d(y,z) \right) \sum_{j=1}^{m} (1_{\tau_1-h_\mu \le T_{ij} \le \tau_1+h_\mu} + 1_{\tau_2-h_\mu \le T_{ij} \le \tau_2+h_\mu}) \\
\le & c \frac{\max(c_2 mh_\mu, 1)}{mh_\mu} d_h(y\times\tau_1, z\times\tau_2) \\
\le & c d_h(y\times\tau_1, z\times\tau_2)
\end{aligned}$$

where $d_h(y\times\tau_1, z\times\tau_2) := \{h_\mu^{-2}|\tau_2-\tau_1|^2 + d_{\mathcal{M}}^2(y,z)\}^{1/2}$ defines a distance on the product space $\mathcal{K}\times\mathcal{T}$. With the entropy bound in Lemma S.4.5, by Theorem 3.3 of van de Geer (1990) we deduce that

$$\Pr\left\{ \sup_{y\in\mathcal{K},\tau\in B(t;h)} |Z_n(y,\tau)| \ge x \right\} \le \exp(-cx^2), \tag{S.17}$$

which directly implies that $\mathbf{E}\big\{\sup_{y\in\mathcal{K},\tau\in B(t;h)} |Z_n(y,\tau)|\big\} = O(1)$ and further $\mathbf{E}\big\{\sup_{y\in\mathcal{K},\tau\in B(t;h)} |U(y,\tau) - \mathbf{E}U(y,\tau)|\big\} = O(n^{-1/2})$.

Next we consider the case $mh_\mu \to 0$. Let

$$V_i(y,\tau) = \sum_{j=1}^{m} K\left(\frac{T_{ij}-\tau}{h_\mu}\right) \left( d_{\mathcal{M}}^2(Y_{ij},y) - F^*(y,T_{ij}) \right) \tag{S.18}$$

and

$$Z_n(y,\tau) = \frac{1}{\sqrt{nmh_\mu}} \sum_{i=1}^{n} V_i(y,\tau). \tag{S.19}$$

Then $\mathbf{E}V_i(y,\tau) = 0$ and $U(y,\tau) - \mathbf{E}U(y,\tau) = (nmh_\mu)^{-1/2} Z_n(y,\tau)$. Observe that

$$\begin{aligned}
|V_i(y,\tau_1) - V_i(z,\tau_2)| \le & c\left( \frac{|\tau_2-\tau_1|}{h_\mu} + d(y,z) \right) \sum_{j=1}^{m} (1_{\tau_1-h_\mu \le T_{ij} \le \tau_1+h_\mu} + 1_{\tau_2-h_\mu \le T_{ij} \le \tau_2+h_\mu}) \\
\le & c d_h(y\times\tau_1, z\times\tau_2),
\end{aligned}$$

where we use the fact that $\sum_{j=1}^{m} (1_{\tau_1-h_\mu \le T_{ij} \le \tau_1+h_\mu} + 1_{\tau_2-h_\mu \le T_{ij} \le \tau_2+h_\mu}) \le c$ due to the assumption $mh_\mu \to 0$ and Assumption 4.2. Note that for all sufficiently small $h_\mu$, there is at most one non-zero item in (S.18) and thus $Z_n(y,\tau)$ in (S.19) is sum of independent random variables. In addition, there are only at most $cnmh_\mu$ non-zero terms in (S.19). Based on Theorem 3.3 of van de Geer (1990) again we see that (S.17) holds, which implies that $\mathbf{E}\big\{\sup_{y\in\mathcal{K},\tau\in B(t;h)} |Z_n(y,\tau)|\big\} = O(1)$ and further $\mathbf{E}\big\{\sup_{y\in\mathcal{K},\tau\in B(t;h)} |U(y,\tau) - \mathbf{E}U(y,\tau)|\big\} = O\big(\frac{1}{\sqrt{nmh_\mu}}\big)$.

To establish (S.13), let $R = \lceil h_\mu^{-1}|\mathcal{T}|\rceil = O(h_\mu^{-1})$ and $A_1,\ldots,A_R$ a partition of $\mathcal{T}$ with $|A_r| \le h_\mu$. According to (S.17), we observe that, in either case of $mh_\mu \gtrsim 1$ and $mh_\mu \to 0$,

$$\begin{aligned}\Pr\left\{\sup_{y\in\mathcal{K},\tau\in\mathcal{T}} |Z_n(y,\tau)| \geq x\sqrt{\log n}\right\} &\leq \sum_{r=1}^{R} \Pr\left\{\sup_{y\in\mathcal{K},\tau\in A_r} |Z_n(y,\tau)| \geq x\sqrt{\log n}\right\} \\ &= O(h_\mu^{-1})\exp(-cx\log n) \leq O(n^{-1}h_\mu^{-1})n^{1-x} \\ &= O(1)n^{1-x},\end{aligned}$$

which then implies (S.13). □

The following lemma is used to establish Theorems 4.1 and 4.2 under the deterministic design. Its proof is similar to that of Lemma S.4.6 and thus is omitted.

**Lemma S.4.7** (covariance, deterministic). *Suppose that Assumptions 2.1, 2.2, 3.1, 4.4, 4.5, 4.6 and 4.2 hold. Let*

$$U(s,t) := \sum_i \nu_i \sum_{j\neq k} \varpi(T_{ij},T_{ik})\left\{\tilde{\mathcal{C}}_{i,jk} - \mathcal{C}(s,t) - \partial_s\mathcal{C}(s,t)(T_{ij}-s) - \partial_t\mathcal{C}(s,t)(T_{ik}-t)\right\},$$

*where $\varpi(s',t') = K_{h_\mathcal{C}}(s-s')K_{h_\mathcal{C}}(t-t')$ for $s',t'\in\mathcal{T}$. If $h_\mathcal{C}\to 0$ and $nm^2h_\mathcal{C}^2\to\infty$, then for all sufficient small $h_\mathcal{C}$,*

$$\sup_{s,t\in\mathcal{T}} \mathbf{E}\{|U(s,t)-\mathbf{E}U(s,t)|\} = O\left(n^{-1/2} + (nm^2h_\mathcal{C}^2)^{-1/2}\right). \tag{S.20}$$

*If $h_\mathcal{C}\to 0$, $nh_\mathcal{C}^2 \gtrsim 1$ and $nm^2h_\mathcal{C}^2/\log n\to\infty$, then*

$$\mathbf{E}\left\{\sup_{s,t\in\mathcal{T}} |U(s,t)-\mathbf{E}U(s,t)|\right\} = O\left(n^{-1/2} + (nm^2h_\mathcal{C}^2)^{-1/2}\right)(\log n)^{1/2}. \tag{S.21}$$

## S.5 Theoretical Results for Regular Design

In a regular design, each sample path is observed on a common set of time points $\{T_j\}_{1\leq j\leq m}$. From a theoretical perspective, this design is fundamentally different from the designs discussed in Section 4, as under such design, $m\to\infty$ is required for the estimators $\hat{\mu}$ and $\hat{\mathcal{C}}$ to be consistent. Below we consider both random and deterministic regular design which includes the often-encountered equally-spaced design as a special case.

**Assumption S.5.1** (Regular Random Design). *The design points $\{T_j\}_{1\leq j\leq m}$, independent of other random quantities, are i.i.d. sampled from a distribution on $\mathcal{T}$ with a probability density that is bounded away from zero and infinity.*

**Assumption S.5.2** (Regular Deterministic Design). *The design points $\{T_j\}_{1\leq j\leq m}$ are nonrandom, and there exist constants $c_2\geq c_1>0$, such that for any intervals $A,B\subset\mathcal{T}$,*

(a) $c_1m|A| - 1 \leq \sum_{j=1}^m 1_{T_j\in A} \leq \max\{c_2m|A|,1\}$,

(b) $c_1m^2|A||B| - 1 \leq \sum_{j,k} 1_{T_j\in A}1_{T_k\in B} \leq \max\{c_2m^2|A||B|,1\}$,

*where $|A|$ denotes the length of $A$.*

For any fixed $t\in\mathcal{T}$, under either of Assumptions S.5.1 or S.5.2, the number of distinct observed time points in the interval of length $h_\mu$ is $O(mh_\mu)$, and thus the condition of $mh_\mu\gtrsim 1$ is necessary for consistency of the mean and covariance estimators. Proposition S.5.1 presents the local and global uniform convergence rates for the mean estimation under regular design. The optimal bandwidth $h_\mu \asymp \frac{1}{m}$ leads to the same convergence rate as that from Cai and Yuan (2011) in the Euclidean case.

**Proposition S.5.1.** *Suppose that Assumptions 2.1, 2.2, 3.1, 4.4 and 4.5 hold. Under either of Assumptions S.5.1 or S.5.2, if $h_\mu \to 0$ and $mh_\mu > 1/c_1$, then*

$$\sup_{t\in\mathcal{T}} d^2_{\mathcal{M}}(\mu(t),\hat{\mu}(t)) = O_p\left(h_\mu^4 + \frac{\log n}{n}\right),$$

*and for any fixed $t \in \mathcal{T}$ and $h = O(h_\mu)$,*

$$\sup_{\tau:|\tau-t|\le h} d^2_{\mathcal{M}}(\mu(\tau),\hat{\mu}(\tau)) = O_p\left(h_\mu^4 + \frac{1}{n}\right).$$

In the above, the condition $mh_\mu > 1/c_1$ ensures that there is at least one observation of the time point for the interval $[t-h_\mu, t+h_\mu]$ for each $t$, according to Assumption S.5.2 for the regular deterministic design. The proof of Proposition S.5.1 is similar to that of Propositions 4.1 and 4.2, where Lemma S.4.1 is replaced with Lemma S.5.1 below for the regular random design and Lemmas S.4.3 and S.4.6 are replaced with Lemma S.5.2 for the regular deterministic design.

**Lemma S.5.1** (mean, regular random)**.** *Suppose that Assumptions 2.1, 2.2, 3.1, 4.4, 4.5. Define*

$$U := \frac{1}{nm}\sum_{ij} K_{h_\mu}(T_j-\tau)\left(d^2_{\mathcal{M}}(Y_{ij},y) - F^*(y,\tau) - \partial_\tau F^*(y,\tau)(T_j-\tau)\right).$$

*Under Assumption S.5.1, if $h_\mu \to 0$ and $mh_\mu \gtrsim 1$, then for any fixed $t\in\mathcal{T}$ and $h = O(h_\mu)$,*

(a) $\sup_{\tau\in B(t;h)} |U - \mathbf{E}U| = O_p\left(\sqrt{\frac{1}{n}}\right)$;

(b) $\sup_{\tau\in\mathcal{T}} |\hat{u}_k(\tau)| = O_p(h_\mu^k)$ *for* $k = 0,1,2$;

(c) $\inf_{\tau\in B(t;h)} |\hat{\sigma}_0^2(\tau)| \asymp h_\mu^2(1+o_P(1))$.

*Proof of Lemma S.5.1.* Define the envelop function

$$H := \frac{2\mathrm{diam}(\mathcal{K})^2}{m}\sum_{j=1}^m \sup_{\tau\in B(t;h)} K_{h_\mu}(T_j-\tau).$$

Since $mh_\mu \gtrsim 1$, simple computation leads to $\mathbf{E}(H^2) = O(1)$ and thus

$$\sup_{\tau\in B(t;h)} |U-\mathbf{E}U| = O_p\left(\sqrt{\frac{1}{n}}\right)$$

according to Theorems 2.7.11 and 2.14.2 of van der Vaart and Wellner (1996). Combining above results together, we deduce part (a). Similar technique leads to part (b). For part (c), it is seen that $\hat{\sigma}_0(\tau) \asymp \{\mathbf{E}\hat{u}_0\mathbf{E}\hat{u}_2 - (\mathbf{E}\hat{u}_1)^2\}(1+o_P(1))$, where the $o_P(1)$ component is uniform over $\tau$. Define $V := K_{h_\mu}(T_{11}-\tau)$ and $W := \mathbf{E}V \asymp 1$. Simple calculation shows that

$$\mathbf{E}\hat{u}_0\mathbf{E}\hat{u}_2 - (\mathbf{E}\hat{u}_1)^2 = W\mathbf{E}\left(V[(T_{11}-\tau) - W^{-1}\mathbf{E}\{V(T_{11}-\tau)\}]^2\right) \asymp h_\mu^2$$

uniformly over all $\tau\in\mathcal{T}$. □

**Lemma S.5.2** (mean, regular deterministic). *Suppose that Assumptions 2.1, 2.2, 3.1, 4.4, 4.5. Define*

$$U := \frac{1}{nm}\sum_{ij} K_{h_\mu}(T_j-\tau)\bigg(d^2_{\mathcal{M}}(Y_{ij},y) - F^*(y,\tau) - \partial_\tau F^*(y,\tau)(T_j-\tau)\bigg).$$

*Under Assumption S.5.2, if $h_\mu \to 0$ and $mh_\mu > 1/c_1$, then for any fixed $t\in\mathcal{T}$ and $h = O(h_\mu)$,*

(a) $\sup_{\tau\in B(t;h)} |U - \mathbf{E}U| = O_p\left(\sqrt{\frac{1}{n}}\right)$;

(b) $\sup_{\tau\in\mathcal{T}} |\hat{u}_k(\tau)| = O_p(h_\mu^k)$ *for* $k = 0,1,2$;

(c) $\inf_{\tau\in B(t;h)} |\hat\sigma_0^2(\tau)| \asymp h_\mu^2$.

*Proof of Lemma S.5.2.* Part (a) can be established by an argument similar to that of Lemma S.4.6 under the condition $mh_\mu \gtrsim 1$. Part (b) can be verified by simple calculation. For part (c), we fix $\tau\in\mathcal{T}$ and let $W := \sum_j K\big(\frac{T_j-\tau}{h_\mu}\big)$. The assumptions on the kernel function imply that $K(u)\ge c_0$ on $[-3/4,3/4]$ for some constant $c_0>0$ depending only on $K$. Assumption S.5.2 implies that there are at least $3c_1mh_\mu/2$ points within the interval $[\tau-3h_\mu/4, \tau+3h_\mu/4]$, from which we deduce that $W \ge 3c_0c_1mh_\mu/2 > 0$ regardless of the location of $\tau$. Let $w_j = K\left(\frac{T_j-\tau}{h_\mu}\right)/W$, which is well defined. Observe that $\hat u_k(\tau) = \frac{W}{mh_\mu}\sum_j w_j(T_j-\tau)^k$. According to the assumptions on the kernel function and Assumption S.5.2, at least $c_1mh_\mu/4$ of sampling points $T_j$ satisfy $|T_j-\tau|\ge h_\mu/8$ and $w_j\ge c_0/W$. Thus, for $k=0,2$,

$$\hat u_k(\tau) \ge \frac{W}{mh_\mu}\frac{c_1mh_\mu}{4}\frac{c_0}{W}\frac{h_\mu^k}{8^k} \ge \frac{c_0c_1}{2^{3k+2}}h_\mu^k$$

regardless of the value of $\tau$. Combining this with part (b) we prove part (c). □

The following theorem provides the pointwise and uniform convergence rates of the covariance estimator under a regular design, where the optimal bandwidth $h_\mu \asymp h_{\mathcal{C}} \asymp \frac{1}{m}$ leads to the same convergence rate as that from Cai and Yuan (2011) in the Euclidean case.

**Theorem S.5.1.** *Suppose that Assumptions 2.1, 2.2, 3.1, 4.4, 4.5 and 4.6 hold. Under either of Assumptions S.5.1 or S.5.2, if $h_\mu\to 0$, $h_{\mathcal{C}} = O(h_\mu)$, $mh_\mu > 1/c_1$, $m^2h_{\mathcal{C}}^2 > 1/c_1$ and $nh_{\mathcal{C}}\gtrsim 1$, then*

$$\sup_{(s,t)\in\mathcal{T}^2}\left\|\mathscr{P}^{(\mu(s),\mu(t))}_{(\hat\mu(s),\hat\mu(t))}\hat{\mathcal{C}}(s,t) - \mathcal{C}(s,t)\right\|^2_{G_{(\mu(s),\mu(t))}} = O_p\left(h_\mu^4 + h_{\mathcal{C}}^4 + \frac{\log n}{n}\right),$$

*and for any fixed $s,t\in\mathcal{T}$,*

$$\left\|\mathscr{P}^{(\mu(s),\mu(t))}_{(\hat\mu(s),\hat\mu(t))}\hat{\mathcal{C}}(s,t) - \mathcal{C}(s,t)\right\|^2_{G(\mu(s),\mu(t))} = O_p\left(h_\mu^4 + h_{\mathcal{C}}^4 + \frac{1}{n}\right).$$

The proof of Theorem S.5.1 is similar to that of Theorems 4.1 and 4.2, where Lemmas S.4.2, S.4.3 and S.4.7 are replaced by Lemma S.5.3 below to analyze the parts $S_{ab}$ and $U$. The proof of Lemma S.5.3 is similar to that of Lemmas S.5.1 and S.5.2 and thus omitted.

**Lemma S.5.3** (covariance, regular). *Suppose that Assumptions 2.1, 2.2, 3.1, 4.4, 4.5 and 4.6 hold. Define*

$$U(s,t) := \sum_i \nu_i \sum_{j\ne k} \varpi(T_{ij},T_{ik})\left\{\tilde{\mathcal{C}}_{i,jk} - \mathcal{C}(s,t) - \partial_s\mathcal{C}(s,t)(T_{ij}-s) - \partial_t\mathcal{C}(s,t)(T_{ik}-t)\right\},$$

*where $\varpi(s',t') = K_{h_{\mathcal{C}}}(s-s')K_{h_{\mathcal{C}}}(t-t')$ for $s',t'\in\mathcal{T}$. Under either of Assumptions S.5.1 or S.5.2, if $h_\mu\to 0$, $h_{\mathcal{C}} = O(h_\mu)$, $mh_\mu > 1/c_1$, $m^2h_{\mathcal{C}}^2 > 1/c_1$, and $nh_{\mathcal{C}}\gtrsim 1$, then*

(a) $\sup_{s,t\in\mathcal{T}} \mathbf{E}\{S_{ab}(s,t)\} = O(1)$;

(b) $\sup_{s,t\in\mathcal{T}} |S_{ab}(s,t) - \mathbf{E}\{S_{ab}(s,t)\}| = o_P(1)$;

(c) $\inf_{s,t\in\mathcal{T}}\{(S_{20}S_{02} - S_{11}^2)S_{00} - (S_{10}S_{02} - S_{01}S_{11})S_{10} + (S_{10}S_{11} - S_{01}S_{20})S_{01}\} \asymp 1 + o_P(1)$;

(d) $\sup_{s,t\in\mathcal{T}} |U(s,t) - \mathbf{E}U(s,t)| = O_p\left(\sqrt{\frac{1}{n}}\right)$.

## S.6 Additional Illustration of Invariance

The covariance function and its estimator proposed in our paper are invariant to the manifold parameterization, choice of frame and embedding. This important invariance property is a consequence of the intrinsic perspective we take, and below we demonstrate that it is not shared by non-intrinsic statistical methods.

**A method non-invariant to parameterization and frame selection.** An "obvious" estimator for $\mathcal{C}$ might be obtained by utilizing a frame along $\hat{\mu}(\cdot)$ and the coefficient process of Lin and Yao (2019). Specifically, fix a frame along $\hat{\mu}$ which determines an orthonormal basis of $T_{\hat{\mu}(t)}\mathcal{M}$ for each $t \in \mathcal{T}$. Then $\mathrm{Log}_{\hat{\mu}(T_{ij})}Y_{ij}$ can be represented by its coefficient vector $\hat{c}_{ij}$ with respect to the frame, and $\hat{\mathcal{C}}_{i,jk}$ is also represented by the observed coefficient matrix $\hat{c}_{ij}\hat{c}_{ik}^\top$. Local linear smoothing (Yao et al., 2005) or other smoothing methods can be applied on these matrices to yield an estimated coefficient matrix at any pair $(s,t)$ of time points, and the corresponding estimate $\hat{\mathcal{C}}(s,t)$ is recovered from the estimated coefficient matrix and the frame. However, this estimate is not invariant to the frame, i.e., different frames give rise to different estimates $\hat{\mathcal{C}}(s,t)$. As a simple example, consider two frames that coincide on all $T_{\hat{\mu}(T_{ij})}\mathcal{M}$ but not on $T_{\hat{\mu}(s)}\mathcal{M}$ and $T_{\hat{\mu}(t)}\mathcal{M}$, and assume that $s,t \notin \{T_{ij} : i = 1,\ldots,n,\ j = 1,\ldots,m_i\}$. Then the coefficient matrices $\hat{c}_{ij}\hat{c}_{ik}^\top$ with respect of the two frames are identical and thus this "obvious" estimator will produce identical estimated coefficient matrix at the pair $(s,t)$. However, since the two frames differ at $s$ and $t$, the estimates $\hat{\mathcal{C}}(s,t)$ recovered from the estimated coefficient matrix under the two frames are different. In addition, smoothing methods optimize certain objective function of the observations which are the frame-dependent coefficient matrices $\hat{c}_{ij}\hat{c}_{ik}^\top$ in this context, while most objective functions, like sum of squared errors, are not invariant to the frame, and consequently the corresponding estimate is frame-dependent.

We now numerically demonstrate that the above method based on Yao et al. (2005) is not invariant to parameterization and frame selection. For this purpose, we generate data from the two-dimensional sphere $\mathbb{S}^2 = \{(x,y,z) \in \mathbb{R}^3 : x^2+y^2+z^2 = 1\}$ with the same setting in Section 5 with sample size $n = 100$ and sampling rate $m = 10$. Consider the following three frames:

- The frame $(B_1(t) = \frac{\partial\phi}{\partial u}, B_2(t) = \frac{\partial\phi}{\partial v})$ derived from the polar parameterization in Equation (12);
- The frame $(B_1^{4\pi}(t), B_2^{4\pi}(t))$ constructed by

  $$B_1^{4\pi}(t) = \cos(4\pi t)B_1(t) + \sin(k\pi t)B_2(t), \quad B_2^{4\pi}(t) = \sin(4\pi t)B_1(t) + \cos(4\pi t)B_2(t),$$

  which is a rotated version of $(B_1(t), B_2(t))$;
- The frame $(\tilde{B}_1(t) = \frac{\partial\varphi}{\partial u}, \tilde{B}_2(t) = \frac{\partial\varphi}{\partial v})$ derived from the parameterization in Equation (15).

For each of these frames, we apply the method described above to estimate $\mathcal{C}$ under the identical conditions, e.g., with the same logarithmically equidistant grid of bandwidths $h_{\mathcal{C}} = 0.20, 0.28, 0.40, 0.56, 0.80$ and known mean function. If the method were invariant to frames, then we would expect to observe *identical* relative root mean integral square error (rRMISE) quantified by

$$\text{rRMISE} := \frac{\{\mathbf{E}\int_{\mathcal{T}^2} \|\hat{\mathcal{C}}(s,t) - \mathcal{C}(s,t)\|_G^2 \mathrm{d}s\mathrm{d}t\}^{1/2}}{\{\int_{\mathcal{T}^2} \|\mathcal{C}(s,t)\|_G^2 \mathrm{d}s\mathrm{d}t\}^{1/2}} \tag{S.22}$$

for any fixed bandwidth. The results, presented in Table S.1 and based on 100 independent Monte Carlo replicates, however, show that different frames lead to distinct rRMISE for a fixed bandwidth and distinct minimum rRMISE over a grid of bandwidths, and thus clearly show that the above method based on Yao et al. (2005) is not invariant to frames.

Table S.1: rRMISE under different frames and bandwidths

| rRMISE | $h_{\mathcal{C}} = 0.20$ | $h_{\mathcal{C}} = 0.28$ | $h_{\mathcal{C}} = 0.40$ | $h_{\mathcal{C}} = 0.56$ | $h_{\mathcal{C}} = 0.80$ |
|---|---|---|---|---|---|
| $(B_1(t), B_2(t))$ | 25.48% (20.20%) | 22.71% (19.38%) | 20.69% (18.86%) | **19.44% (18.29%)** | 19.94% (17.36%) |
| $(B_1^{4\pi}(t), B_2^{4\pi}(t))$ | 116.32% (34.57%) | 109.13% (30.08%) | 98.24% (23.31%) | **91.38% (24.19%)** | 93.07% (23.65%) |
| $(\tilde{B}_1(t), \tilde{B}_2(t))$ | 49.19% (21.77%) | 46.63% (20.59%) | 43.52% (19.38%) | 39.78% (18.04%) | **36.67% (16.83%)** |

**A method non-invariant to embedding.** We demonstrate that different embeddings for the method of Dai et al. (2020) yield distinct estimates of the covariance function. Consider a plane $\mathcal{M} = (0,1) \times [0,1]$ with the metric inherited from $\mathbb{R}^2$. The underlying population $X$ on $\mathcal{M}$ is $X(t) = (0.25 + 0.5t + Z_1, 0.5 + Z_2)$ where $Z_1, Z_2 \sim \text{Uniform}(-0.1, 0.1)$ and $\mu(t) = (0.25 + 0.5t, 0.5)$. We generate $n = 100$ paths from $X$ and for each path we randomly sample $\text{Poisson}(10) + 2$ observations, where $\text{Poisson}(10)$ is the Poisson distribution with mean parameter 10. Consider the following three isometric embeddings of $\mathcal{M}$ into $\mathbb{R}^3$:

$$\begin{aligned} &\iota_1 : (x,y) \to (x,y,0) && \text{plane;} \\ &\iota_2 : (x,y) \to (\frac{1}{\pi}\sin(\pi x), \frac{1}{\pi}\cos(\pi x), y) && \text{half cylindrical surface;} \\ &\iota_3 : (x,y) \to (\frac{1}{2\pi}\sin(2\pi x), \frac{1}{2\pi}\cos(2\pi x), y) && \text{cylindrical surface.} \end{aligned}$$

For each of these embeddings, we apply the method of Dai et al. (2020) to produce an estimate of $\mathcal{C}$ and calculate rRMISE (S.22) of these estimates, where for illustration, we consider logarithmically equidistant grid of bandwidths $h_{\mathcal{C}} = 0.10, 0.14, 0.22, 0.33, 0.50$ and the known true mean function $\mu(t)$. If the method of Dai et al. (2020) were invariant to embeddings, then we would expect to observe identical rRMISE for these embeddings for each bandwidth. Table S.2 with the rRMISE results based on 100 Monte Carlo simulation replicates, suggesting the opposite, clearly shows that the method of Dai et al. (2020) is not invariant to choices of the frame.

Table S.2: rRMISE under different embeddings and bandwidths

| rRMISE | $h_{\mathcal{C}} = 0.10$ | $h_{\mathcal{C}} = 0.14$ | $h_{\mathcal{C}} = 0.22$ | $h_{\mathcal{C}} = 0.33$ | $h_{\mathcal{C}} = 0.50$ |
|---|---|---|---|---|---|
| $\iota_1$ | 26.91% (17.21%) | 22.38% (15.72%) | 19.50% (14.76%) | **17.85% (14.49%)** | 19.25% (31.40%) |
| $\iota_2$ | 51.52% (21.16%) | 49.22% (19.54%) | 47.88% (18.48%) | **47.10% (17.92%)** | 54.57% (90.06%) |
| $\iota_3$ | 81.83% (29.78%) | 79.97% (28.08%) | 78.49% (26.78%) | **77.15% (25.40%)** | 90.01% (150.56%) |

**A real data example.** We now demonstrate that different choices of the frame for the extrinsic method based on Yao et al. (2005) lead to distinct statistical results for the real data analyzed in Section 6. Consider the frame $\{(B_k)\}_{1\le k\le 6}$ induced from the parameterization

$$\phi : (x_1, x_2, x_3, x_4, x_5, x_6) \in \mathbb{R}^6 \to \begin{pmatrix} e^{x_1} & 0 & 0 \\ x_4 & e^{x_2} & 0 \\ x_5 & x_6 & e^{x_3} \end{pmatrix} \begin{pmatrix} e^{x_1} & x_4 & x_5 \\ 0 & e^{x_2} & x_6 \\ 0 & 0 & e^{x_3} \end{pmatrix} \in \mathrm{Sym}^+_{LC}$$

and the frame $\{(\tilde{B}_k)\}_{1\le k\le 6}$ derived from a rotation of $\{(B_k)\}_{1\le k\le 6}$ on each tangent space $T_{\hat{\mu}(T_{ij})}\mathrm{Sym}^+_{LC}$ by

$$\{(\tilde{B}_k)\}_{1\le k\le 6} = \{(B_k)\}_{1\le k\le 6}\times$$
$$\mathrm{diag}\left\{\begin{pmatrix} \cos(4\pi T_{ij}) & -\sin(4\pi T_{ij}) \\ \sin(4\pi T_{ij}) & \cos(4\pi T_{ij}) \end{pmatrix}, \begin{pmatrix} \cos(4\pi T_{ij}) & -\sin(4\pi T_{ij}) \\ \sin(4\pi T_{ij}) & \cos(4\pi T_{ij}) \end{pmatrix}, \begin{pmatrix} \cos(4\pi T_{ij}) & -\sin(4\pi T_{ij}) \\ \sin(4\pi T_{ij}) & \cos(4\pi T_{ij}) \end{pmatrix}\right\},$$

where $\mathrm{diag}(M_1, M_2, M_3)$ for matrices $M_1, M_2, M_3$ denotes the block diagonal matrix formed by $M_1, M_2, M_3$. For each of these two frames, we apply the extrinsic method based on Yao et al. (2005) to estimate the covariance function and its eigenfunctions under identical conditions, e.g., with the same estimated mean function (with $h_\mu = 10$) and the same choice of bandwidth $h_{\mathcal{C}} = 20$. Figure S.4, depicting the first three functional principal components obtained from the two frames, clearly shows that the two frames yield distinct estimates. This demonstrates that the above method based on Yao et al. (2005) is not invariant to frames.

# References


Cai, T. and Yuan, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *The Annals of Statistics*, 39(5):2330–2355.

Dai, X., Lin, Z., and Müller, H.-G. (2020). Modeling sparse longitudinal data on riemannian manifolds. *Biometrics*, 77(4):1328–1341.

Gray, A. (2012). *Tubes*, volume 221. Springer Basel AG, Boston, second edition.

Lee, J. M. (2002). *Introduction to Smooth Manifolds*. Springer, New York.

Lee, J. M. (2018). *Introduction to Riemannian Manifolds*, volume 176. Springer, Cham, second edition.

Lin, Z. and Yao, F. (2019). Intrinsic Riemannian functional data analysis. *The Annals of Statistics*, 47(6):3533–3577.

Munkres, J. R. (2000). *Topology*. Prentice Hall, Upper Saddle River, second edition.

Nichols, David, A., Vines, and Justin (2016). Properties of an affine transport equation and its holonomy. *General Relativity and Gravitation*, 48(127).

Oller, J. M. and Corcuera, J. M. (1995). Intrinsic Analysis of Statistical Estimation. *The Annals of Statistics*, 23(5):1562–1581.

Pennec, X. (2019). Curvature effects on the empirical mean in riemannian and affine manifolds: a non-asymptotic high concentration expansion in the small-sample regime. *arxiv*.

Petersen, A. and Müller, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *The Annals of Statistics*, 47(2):691–719.

Rodrigues, W. A. and Capelas de Oliveira, E. (2007). *Many Faces of Maxwell, Dirac and Einstein Equations: A Clifford Bundle Approach*, volume 722. Springer, Heidelberg.

Sacks, J. and Ylvisaker, D. (1970). Designs for regression problems with correlated errors iii. *The Annals of Mathematical Statistics*, 41(6):2057–2074.

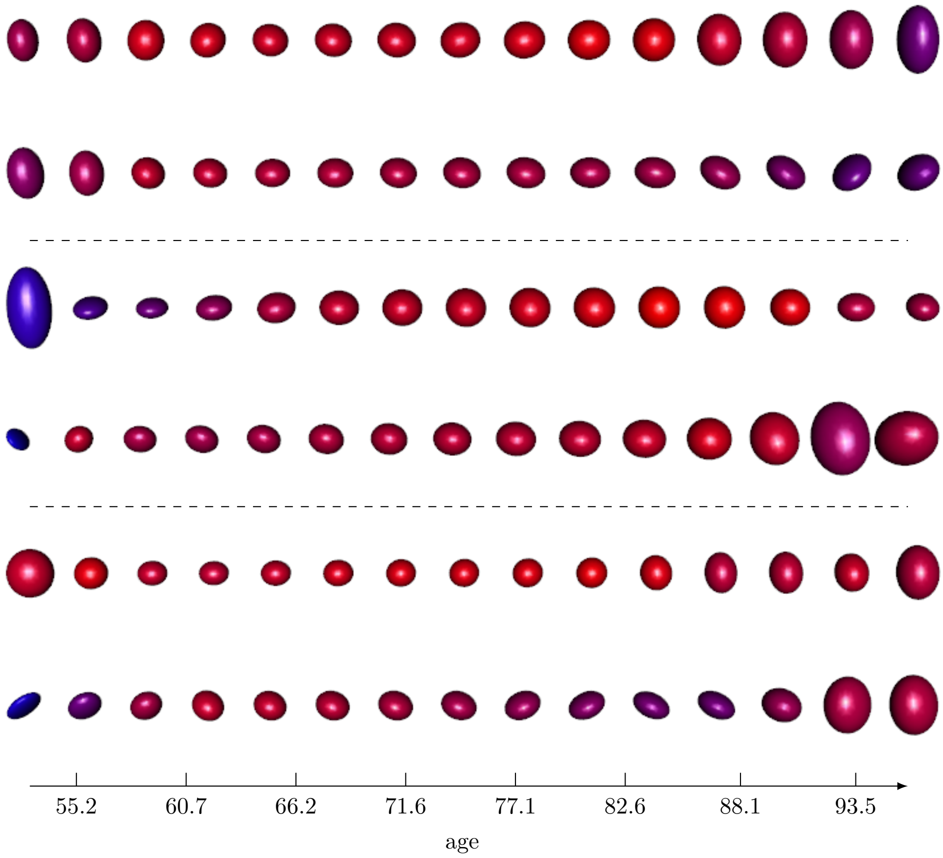


Figure S.4: The first principal component resulting from the frame $\{(B_k)\}_{1\le k\le 6}$ (Row 1) and the frame $\{(\tilde{B}_k)\}_{1\le k\le 6}$ (Row 2), the second principal component resulting from the frame $\{(B_k)\}_{1\le k\le 6}$ (Row 3) and the frame $\{(\tilde{B}_k)\}_{1\le k\le 6}$ (Row 4), and the third principal component resulting from the frame $\{(B_k)\}_{1\le k\le 6}$ (Row 5) and the frame $\{(\tilde{B}_k)\}_{1\le k\le 6}$ (Row 6). The color encodes fractional anisotropy.

van de Geer, S. (1990). Estimating a regression function. *Annals of Statistics*, 18(2):907–924.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590.

Zhang, X. and Wang, J. L. (2016). From sparse to dense functional data and beyond. *The Annals of Statistics*, 44:2281–2321.