



University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): WILFRID S. KENDALL AND HUILING LE

Article Title: Limit theorems for empirical Fréchet means of independent and non-identically distributed manifold-valued random variables

Year of publication: 2011

Link to published article: <http://dx.doi.org/10.1214/11-BJPS14>

Publisher statement: Journal homepage: <http://www.imstat.org/bjps/>

Limit theorems for empirical Fréchet means of independent and non-identically distributed manifold-valued random variables

Wilfrid S. Kendall^a and Huiling Le^b

^a*University of Warwick*

^b*University of Nottingham*

Abstract. We prove weak laws of large numbers and central limit theorems of Lindeberg type for empirical centres of mass (empirical Fréchet means) of independent nonidentically distributed random variables taking values in Riemannian manifolds. In order to prove these theorems we describe and prove a simple kind of Lindeberg–Feller central approximation theorem for vector-valued random variables, which may be of independent interest and is therefore the subject of a self-contained section. This vector-valued result allows us to clarify the number of conditions required for the central limit theorem for empirical Fréchet means, while extending its scope.

1 Introduction

Fréchet means, or Riemannian centres of mass, were introduced at a relatively early stage of probability by Fréchet (1948). The idea is simple enough: generalize the mean-square characterization of the mean $\mathbb{E}[X]$ as the minimizer of the “energy function” $x \mapsto \frac{1}{2}\mathbb{E}[(X - x)^2]$. If X takes values in a metric space \mathcal{X} this can be achieved as follows: replace $(X - x)^2$ by the square of the distance function $\text{dist}(X, x)^2$.

Of course the theory of Fréchet means is subject to geometric complications. Uniqueness becomes the exception rather than the rule, though existence is guaranteed if the metric space satisfies some kind of local compactness condition. Ziezold (1977, 1989, 1994) established some basic results in this broad context, as well as developing some significant applications in applied statistics. If the metric space \mathcal{X} is specialized to a Riemannian manifold \mathbb{M} then it is possible to produce useful calculations and estimates using curvature; Karcher (1977) provides a good account of this as well as surveying substantial applications of Fréchet means in geometry.

Probabilistic interest in Fréchet means was initially spurred on by considerations of how to generate theories of martingales taking values in manifolds, and

Key words and phrases. Central approximation theorem, central limit theorem, curvature, empirical Fréchet mean, exponential map, Fréchet mean, gradient, Hessian, Kähler manifold, Lindeberg condition, Newton’s method, Riemannian centre of mass, weak law of large numbers.

Received February 2011.

in particular how then to extend the mathematical application of martingale theory beyond the theory of linear elliptic differential equations to the theory of harmonic maps [Kendall (1990); Picard (1994)]. In particular this led to strong connections with convexity theory for Riemannian manifolds, simply expressed in Kendall (1991b) and further developed in Kendall (1991a, 1992a, 1992b) and Corcuera and Kendall (1999); more recently see Afsari (2011). Ziezold (1989)'s application of Fréchet means to statistical shape theory has been taken up by several workers [see, e.g., Le (2001, 2004); also the recent survey by Kendall and Le (2010)]. In particular Bhattacharya and Patrangenaru (2003, 2005) and Bhattacharya and Bhattacharya (2008) have developed important statistical theory for *empirical* Fréchet means on Riemannian manifolds, including (but not limited to) laws of large numbers and central limit theory for independent and identically distributed manifold-valued random variables.

The present paper is inspired by these results of Bhattacharya and coworkers, and addresses the challenge of extending their theory to the nonidentically distributed case. After Section 2, which establishes basic definitions and notation, in Section 3 we develop a weak law of large numbers for empirical Fréchet means in a metric space context (Theorem 2) which is based on the most general possible weak law of large numbers for independent nonnegative random variables (stated here as Theorem 1). In particular, we pay attention to the question of when one can assert existence of *local* empirical Fréchet means lying close to a local minimizer of the aggregated energy function which is obtained by summing the individual energy functions of the random variables concerned.

It is a natural step from this theory to consider central limit theorems of Lindeberg type for empirical Fréchet means, since the conditions for the weak law of large numbers (Theorem 2) involve conditions of Lindeberg type. To do this one needs to specialize to the more specific case of Riemannian manifolds, since this allows one to use the Riemannian Exponential map to refer the manifold to an Euclidean approximation. It is therefore apparent that a central limit theorem for the Riemannian manifold case must depend on a central limit theorem for the random tangent vectors corresponding to the manifold-valued random variables via this Exponential map, and Section 4 considers the relevant theory.

In fact there is a substantial literature on central limit theorems and normal approximations for vector-valued random variables; see Bhattacharya and Rao (1976) for an exposition in book form, and more recently Chatterjee (2008) and Röllin (2011) (both of whom describe approaches which apply Stein's method). However, as we sought to generalize to a Lindeberg central limit theorem for empirical Fréchet means so it became clear that we needed a subtly different result: a theorem which would describe when a sequence of normalized random sums may be approximated by a second sequence of matching multivariate normal random variables, when there is no guarantee of weak convergence, and *when the normalization uses not individual coordinate variances but the trace of the variance-covariance matrix of the sum*. These requirements mean, for example, that one cannot simply apply the Cramér-Wold device. The closest general result we can find

in the published literature is that of [Bhattacharya and Rao \(1976, Corollary 18.2\)](#) [also see [Barbour and Gnedin \(2009\)](#) for specific cases arising in study of infinite occupancy schemes]; however this uses normalization in a matrix-valued sense, using the inverse of the symmetric square-root of the variance–covariance matrix (which is required to be nonsingular), whereas we need an approach which uses scalar normalization and which can work even when the variance–covariance matrix degenerates.

It turns out, as we describe in [Section 4](#), that it is possible to formulate such a result, a multidimensional Lindeberg central *approximation* theorem, which we state and prove as [Theorem 3](#) (and also [Corollary 2](#) for the Feller converse). Proofs vary little from the classic approach of, say, [Feller \(1966\)](#). However it is necessary to take account of the vector-valued context and to allow for a crucial intervention of the Wasserstein metric for the truncated Euclidean distance; therefore we give the proofs in full for the sake of completeness of exposition, since the application is unfamiliar.

These results allow us to prove a Lindeberg central approximation theorem for empirical Fréchet means, which forms [Theorem 4](#) in [Section 5](#). The basic idea uses Newton’s root-finding algorithm, and owes much to the work of [Bhattacharya](#) and coworkers; however while extending to the nonidentically distributed case we are also able to clarify the set of conditions required for the result, by exploiting the idea of central approximation rather than central limits, and we can derive a rather explicit form for the variance–covariance matrices of the approximating multivariate normal random variables. The paper concludes with a small number of illustrative examples, demonstrating how the results simplify in the case of independent and identically distributed random variables, and also in the case when the Riemannian manifold is of constant sectional curvature, or carries a Kähler structure with constant holomorphic sectional curvature.

2 Basic theory and notation

Consider the *energy function* of a random variable X taking values in a metric space \mathcal{X} :

$$\phi(x) = \mathbb{E} \left[\frac{1}{2} \text{dist}(X, x)^2 \right].$$

Observe that if ϕ is finite at one point of \mathcal{X} then it is finite everywhere, by an argument using the triangle inequality. Given independent X_1, \dots, X_n , the *aggregate energy function* is simply the sum

$$\phi_n(x) = \sum_{m=1}^n \mathbb{E} \left[\frac{1}{2} \text{dist}(X_m, x)^2 \right].$$

A *Fréchet mean* is a global minimizer of ϕ . Note that there can be more than one Fréchet mean: we then consider *the set of Fréchet means*

$$\arg \min_x \mathbb{E} \left[\frac{1}{2} \text{dist}(X, x)^2 \right].$$

An *empirical Fréchet mean* is a global minimizer of the energy function based on the empirical probability measure defined by a sample X_1, \dots, X_n : thus *the set of empirical Fréchet means* is

$$\arg \min_x \frac{1}{n} \sum_{i=1}^n \frac{1}{2} \text{dist}(X_i, x)^2.$$

[In case of local compactness, the existence of global minimizers of both kinds follows immediately from $\text{dist}(X, y) + \text{dist}(X, x) \geq \text{dist}(x, y)$.]

Some of our results hold for local minimizers; we use the term *local Fréchet mean* to describe a local minimizer of ϕ , while a *local empirical Fréchet mean* denotes a local minimizer of the energy function based on the empirical probability measure defined by a sample X_1, \dots, X_n of points from the metric space \mathcal{X} .

We shall use the operator-theoretic notation $\mathbb{E}[H]$ to denote the expectation of a random variable H . In particular we shall write $\mathbb{E}[H; A] = \mathbb{E}[H\mathbb{I}[A]]$, where $\mathbb{I}[A]$ is the indicator random variable for an event A .

3 Weak law of large numbers for empirical Fréchet means

[Ziezold \(1977\)](#) established a strong law of large numbers for sequences of independent identically distributed random variables X_1, X_2, \dots taking values in a separable metric space \mathcal{X} (actually [Ziezold](#) covered the more general case of a separable *finite quasi-metric space*). Imposing the condition that the energy function $\mathbb{E}[\frac{1}{2} \text{dist}(X_i, x)^2]$ be finite for some (and thus all) x , [Ziezold](#) was then able to show that almost surely the limit of the closure of the sup of the set of empirical Fréchet means is a subset of the set of Fréchet means (up to an event of zero probability measure):

$$\bigcap_{k=1}^{\infty} \overline{\bigcup_{n=k}^{\infty} \arg \min_x \frac{1}{n} \sum_{i=1}^n \frac{1}{2} \text{dist}(X_i, x)^2} \subseteq \arg \min_x \mathbb{E} \left[\frac{1}{2} \text{dist}(X_1, x)^2 \right]. \quad (3.1)$$

Here of course the $\arg \min$ are treated as random closed sets.

If \mathcal{X} is not compact then it is possible for a sequence of empirical Fréchet means to diverge to infinity even when (3.1) holds. Given uniqueness of the Fréchet mean, [Bhattacharya and Patrangenaru \(2003, Theorem 2.3\)](#) have shown that a strong law of large numbers follows from imposition of the additional condition that every

closed bounded subset of \mathcal{X} is compact; in that case every sequence of measurable choices from the sets

$$\arg \min_x \frac{1}{n} \sum_{i=1}^n \frac{1}{2} \text{dist}(X_i, x)^2$$

of empirical Fréchet means will almost surely converge to the unique Fréchet mean.

In this section we derive a *weak* law of large numbers in the more general case of nonidentically distributed independent random variables X_1, X_2, \dots , taking values in a separable metric space \mathcal{X} possessing the bounded compactness property of [Bhattacharya and Patrangenaru](#), and such that the individual energy functions $\mathbb{E}[\frac{1}{2} \text{dist}(X_n, x)^2]$ are finite for some (and therefore for all) $x \in \mathcal{X}$. Evidently we need to impose extra conditions to compensate for the lack of identical distribution; we will require that the aggregate energy function $\phi_n(x) = \sum_{i=1}^n \mathbb{E}[\frac{1}{2} \text{dist}(X_i, x)^2]$ has a strict local minimum near a fixed reference point $\mathbf{o} \in \mathcal{X}$, and we will require that this holds uniformly as $n \rightarrow \infty$ [in a particular sense captured in the displayed equation (3.2) below]. In recompense for this restriction, our results describe the behaviour of *local* empirical Fréchet means lying in a geodesic ball $\text{ball}(\mathbf{o}, \rho_1) \subseteq \mathcal{X}$. The particular uniformity requirement is that for each positive $\rho_0 \leq \rho_1$ there is positive $\kappa = \kappa(\rho_0, \rho_1)$ such that, for all n ,

$$(1 + \kappa)\phi_n(\mathbf{o}) < \inf\{\phi_n(y) : \rho_0 \leq \text{dist}(y, \mathbf{o}) \leq \rho_1\}. \tag{3.2}$$

Bearing in mind that the ultimate aim of this paper is to prove a central limit theorem, convergence in probability is a more natural objective than almost sure convergence. Therefore it is reasonable to restrict attention to the weaker notion of convergence in probability. Moreover even in the scalar case the law-of-large-numbers conditions for convergence in probability are clearer and more easily stated than for convergence almost surely. The key theorem for our treatment is the weak law of large numbers for nonidentically distributed nonnegative real random variables. We state a special case of this result:

Theorem 1. *Suppose that Z_1, Z_2, \dots are independent nonnegative real random variables, not necessarily of the same distribution. Suppose further that*

$$\frac{1}{\sum_{r=1}^n \mathbb{E}[Z_r]} \sum_{m=1}^n \mathbb{E} \left[Z_m; Z_m \geq \varepsilon \sum_{r=1}^n \mathbb{E}[Z_r] \right] \longrightarrow 0 \quad \text{for each } \varepsilon > 0. \tag{3.3}$$

Then it is the case that as $n \rightarrow \infty$ so

$$\frac{\sum_{r=1}^n Z_r}{\sum_{r=1}^n \mathbb{E}[Z_r]} \longrightarrow 1 \quad \text{in probability.} \tag{3.4}$$

This theorem follows directly from [Chow and Teicher \(2003, Chapter 10, Theorem 1, Corollary 2\)](#).

Remark 1. The condition (3.3) can be viewed as an equation of Lindeberg type. Indeed, if W_1, W_2, \dots are independent real random variables with $\mathbb{E}[W_m] = 0$ and such that $W_m^2 = Z_m$, then (3.3) corresponds exactly to the usual Lindeberg condition for the sequence $\{W_m : m \geq 1\}$. Thus [Chow and Teicher \(2003, Chapter 10, Theorem 1, Corollary 2\)](#) signals the close connection between weak laws of large numbers and the central limit theorem.

Our strategy for proving a weak law of large numbers for nonidentically distributed \mathcal{X} -valued random variables is as follows: consider the condition (3.3) applied to the case $Z_m^{(x)} = \frac{1}{2} \text{dist}(X_m, x)^2$, and then apply the corresponding weak laws of large numbers (3.4). Under suitable additional conditions the aggregate empirical energy functions $\sum_{m=1}^n Z_m^{(x)} = \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, x)^2$ can be made to approximate the aggregate energy functions $\phi_n(x)$ closely enough to ensure that the uniform local minimum property forces convergence to 1 of the probability of there being local empirical Fréchet means close to \mathbf{o} .

For a useful result it is preferable to require that the Lindeberg-type condition apply only at the chosen reference point \mathbf{o} . For a general metric space \mathcal{X} we should not expect the Lindeberg-type condition for the $Z_m^{(\mathbf{o})}$ to imply the corresponding conditions obtained when \mathbf{o} is replaced by a general $x \in \mathcal{X}$. However we can prove a partial result in this direction, which will be sufficient for our purposes:

Lemma 1. *Suppose as above that \mathcal{X} is a separable metric space. Let X_1, X_2, \dots be independent \mathcal{X} -valued random variables with finite energy functions. The following conditions of Lindeberg-type are equivalent:*

- *Firstly, a local Lindeberg condition:*

$$\frac{1}{\phi_n(x)} \sum_{m=1}^n \mathbb{E} \left[\frac{1}{2} \text{dist}(X_m, x)^2; \frac{1}{2} \text{dist}(X_m, x)^2 > \varepsilon \phi_n(x) \right] \rightarrow 0$$

(3.5)

as $n \rightarrow \infty$ for each $\varepsilon > 0$.

- *Secondly, a semi-global Lindeberg condition:*

$$\frac{1}{n\phi_n(x)} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E} \left[\frac{1}{2} \text{dist}(X_i, X_j)^2; \frac{1}{2} \text{dist}(X_i, X_j)^2 > \varepsilon \phi_n(x) \right] \rightarrow 0$$

(3.6)

as $n \rightarrow \infty$ for each $\varepsilon > 0$.

Remark 2. Note that the presence of $\phi_n(x)$ in (3.6) means that this semiglobal condition is *not* truly global, since $\phi_n(x) = \sum_{m=1}^n \mathbb{E}[\frac{1}{2} \text{dist}(X_m, x)^2]$ depends implicitly on the choice of $x \in \mathcal{X}$.

Proof of Lemma 1. First suppose that the local condition (3.5) holds. We shall use this to produce an upper bound on the quantity on the left-hand side of (3.6). Indeed

$$\begin{aligned} & \frac{1}{n\phi_n(x)} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E} \left[\frac{1}{2} \text{dist}(X_i, X_j)^2; \frac{1}{2} \text{dist}(X_i, X_j)^2 > \varepsilon\phi_n(x) \right] \\ & \leq \frac{4}{\phi_n(x)} \sum_{i=1}^n \mathbb{E} \left[\frac{1}{2} \text{dist}(X_i, x)^2; \frac{1}{2} \text{dist}(X_i, x)^2 > \frac{\varepsilon}{4}\phi_n(x) \right] \\ & \quad + \frac{4}{n} \sum_{i=1}^n \mathbb{P} \left[\frac{1}{2} \text{dist}(X_i, x)^2 > \frac{\varepsilon}{4}\phi_n(x) \right]. \end{aligned}$$

Here we make direct use of the triangle inequality via

$$\text{dist}(X_i, X_j)^2 \leq 2 \text{dist}(X_i, x)^2 + 2 \text{dist}(X_j, x)^2;$$

in particular the condition that $\text{dist}(X_i, X_j) > \sqrt{2\varepsilon\phi_n(x)}$ implies that at least one of $\text{dist}(X_i, x) > \frac{1}{2}\sqrt{2\varepsilon\phi_n(x)}$ or $\text{dist}(X_j, x) > \frac{1}{2}\sqrt{2\varepsilon\phi_n(x)}$ must hold.

The Markov inequality implies that

$$\begin{aligned} & \frac{4}{n} \sum_{i=1}^n \mathbb{P} \left[\frac{1}{2} \text{dist}(X_i, x)^2 > \frac{\varepsilon}{4}\phi_n(x) \right] \\ & \leq \frac{16}{\varepsilon n\phi_n(x)} \sum_{i=1}^n \mathbb{E} \left[\frac{1}{2} \text{dist}(X_i, x)^2; \frac{1}{2} \text{dist}(X_i, x)^2 > \frac{\varepsilon}{4}\phi_n(x) \right] \end{aligned}$$

and therefore we obtain

$$\begin{aligned} & \frac{1}{n\phi_n(x)} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E} \left[\frac{1}{2} \text{dist}(X_i, X_j)^2; \frac{1}{2} \text{dist}(X_i, X_j)^2 > \varepsilon\phi_n(x) \right] \\ & \leq 4 \left(1 + \frac{4}{\varepsilon n} \right) \frac{1}{\phi_n(x)} \sum_{i=1}^n \mathbb{E} \left[\frac{1}{2} \text{dist}(X_i, x)^2; \frac{1}{2} \text{dist}(X_i, x)^2 > \frac{\varepsilon}{4}\phi_n(x) \right]. \end{aligned}$$

For any $\varepsilon > 0$ this upper bound tends to zero as $n \rightarrow \infty$, by (3.5), and therefore we obtain (3.6).

Now suppose on the other hand that the semiglobal condition (3.6) holds. If $\text{dist}(X_i, x) > \sqrt{2\varepsilon\phi_n(x)}$ and $\text{dist}(X_j, x) \leq \frac{1}{2}\sqrt{2\varepsilon\phi_n(x)}$ then it follows that $\text{dist}(X_i, X_j) > \frac{1}{2}\sqrt{2\varepsilon\phi_n(x)}$. We deduce that

$$\begin{aligned} & \mathbb{E} \left[\frac{1}{2} \text{dist}(X_i, X_j)^2; \frac{1}{2} \text{dist}(X_i, X_j)^2 > \frac{\varepsilon}{4}\phi_n(x) \right] \\ & \geq \mathbb{E} \left[\frac{1}{2} \text{dist}(X_i, X_j)^2; \text{dist}(X_i, x) > \sqrt{2\varepsilon\phi_n(x)}, \text{dist}(X_j, x) \leq \frac{1}{2}\sqrt{2\varepsilon\phi_n(x)} \right]. \end{aligned}$$

If $\text{dist}(X_i, x) > \sqrt{2\varepsilon\phi_n(x)}$ and $\text{dist}(X_j, x) \leq \frac{1}{2}\sqrt{2\varepsilon\phi_n(x)}$ then $\text{dist}(X_i, X_j) \geq \text{dist}(X_i, x) - \text{dist}(X_j, x) \geq \frac{1}{2}\sqrt{2\varepsilon\phi_n(x)} \geq \frac{1}{2}\text{dist}(X_i, x)$, and so

$$\begin{aligned} & \frac{1}{n\phi_n(x)} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E} \left[\frac{1}{2} \text{dist}(X_i, X_j)^2; \frac{1}{2} \text{dist}(X_i, X_j)^2 > \frac{\varepsilon}{4} \phi_n(x) \right] \\ & \geq \frac{1}{4\phi_n(x)} \sum_{i=1}^n \left(\mathbb{E} \left[\frac{1}{2} \text{dist}(X_i, x)^2; \frac{1}{2} \text{dist}(X_i, x)^2 > \varepsilon\phi_n(x) \right] \right. \\ & \quad \left. \times \frac{1}{n} \sum_{j=1, j \neq i}^n \mathbb{P} \left[\frac{1}{2} \text{dist}(X_j, x)^2 \leq \frac{\varepsilon}{4} \phi_n(x) \right] \right). \end{aligned}$$

Finally we take complements and use Markov’s inequality to deduce

$$\begin{aligned} & \frac{1}{n} \sum_{j=1, j \neq i}^n \mathbb{P} \left[\frac{1}{2} \text{dist}(X_j, x)^2 \leq \frac{\varepsilon}{4} \phi_n(x) \right] \\ & \geq 1 - \frac{1}{n} - \frac{4}{n\varepsilon\phi_n(x)} \sum_{j=1, j \neq i}^n \mathbb{E} \left[\frac{1}{2} \text{dist}(X_j, x)^2; \frac{1}{2} \text{dist}(X_j, x)^2 > \frac{\varepsilon}{4} \phi_n(x) \right] \\ & \geq 1 - \frac{1}{n} - \frac{4}{n\varepsilon} \geq \frac{1}{2} \quad \text{once } n \geq 2 \left(1 + \frac{4}{\varepsilon} \right). \end{aligned}$$

Taking $n \geq 2(1 + \frac{4}{\varepsilon})$, we deduce that (3.6) implies (3.5) by arguing that

$$\begin{aligned} & \frac{1}{n\phi_n(x)} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E} \left[\frac{1}{2} \text{dist}(X_i, X_j)^2; \frac{1}{2} \text{dist}(X_i, X_j)^2 > \frac{\varepsilon}{4} \phi_n(x) \right] \\ & \geq \frac{1}{8} \frac{1}{\phi_n(x)} \sum_{i=1}^n \mathbb{E} \left[\frac{1}{2} \text{dist}(X_i, x)^2; \frac{1}{2} \text{dist}(X_i, x)^2 > \varepsilon\phi_n(x) \right]. \end{aligned}$$

This establishes the equivalence of local and semiglobal conditions. □

Effective use of the semiglobal Lindeberg condition depends on a lower bound on the growth of the energy function $\phi_n(y)$ as $\text{dist}(\mathbf{o}, y)$ increases.

Lemma 2. *Suppose as above that \mathcal{X} is a separable metric space. Let X_1, X_2, \dots be \mathcal{X} -valued random variables with finite energy functions. Suppose that the aggregate energy function $\phi_n(y) = \sum_{m=1}^n \mathbb{E}[\frac{1}{2} \text{dist}(X_m, y)^2]$ attains its minimum over \mathcal{X} at $y = \mathbf{o}$:*

$$\phi_n(\mathbf{o}) \leq \phi_n(y). \tag{3.7}$$

Then the aggregate energy function grows at least linearly at any $y \neq \mathbf{o}$:

$$\phi_n(y) \geq \frac{\text{dist}(y, \mathbf{o})^2}{16} n. \tag{3.8}$$

Proof. For convenience, set $\rho = \text{dist}(y, \mathbf{o})$. If $\phi_n(\mathbf{o}) \geq \rho^2 n/16$ then (3.8) follows from Inequality (3.7). So we can suppose that $\phi_n(\mathbf{o}) < \rho^2 n/16$.

For additional convenience let M be a random integer chosen uniformly from $\{1, 2, \dots, n\}$ (independently of X_1, \dots, X_n). Then

$$\begin{aligned} \frac{1}{n} \phi_n(y) &= \mathbb{E} \left[\frac{1}{2} \text{dist}(X_M, y)^2 \right] \geq \frac{\rho^2}{8} \mathbb{P} \left[\frac{1}{2} \text{dist}(X_M, y)^2 \geq \frac{\rho^2}{8} \right] \\ &\quad \text{(Markov inequality)} \\ &= \frac{\rho^2}{8} \mathbb{P} \left[\text{dist}(X_M, y) \geq \frac{\rho}{2} \right] \geq \frac{\rho^2}{8} \mathbb{P} \left[\text{dist}(X_M, \mathbf{o}) < \frac{\rho}{2} \right] \\ &\quad \text{(triangle inequality)} \\ &= \frac{\rho^2}{8} \left(1 - \mathbb{P} \left[\frac{1}{2} \text{dist}(X_M, \mathbf{o})^2 \geq \frac{\rho^2}{8} \right] \right) \geq \frac{\rho^2}{8} \left(1 - \frac{8}{\rho^2} \mathbb{E} \left[\frac{1}{2} \text{dist}(X_M, \mathbf{o})^2 \right] \right) \\ &\quad \text{(Markov inequality again)} \\ &\geq \frac{\rho^2}{8} \left(1 - \frac{8}{\rho^2} \frac{\rho^2}{16} \right) \geq \frac{\rho^2}{16}. \end{aligned}$$

So (3.8) follows in this case also. □

We are now in a position to state and prove the main result of this section. We follow [Bhattacharya and Patrangenaru \(2003\)](#) by imposing the compactness of bounded closed sets, and also impose the uniform local minimum property described above by Inequality (3.2).

Theorem 2. *Suppose \mathcal{X} is a separable metric space for which all bounded closed sets are compact. Let X_1, X_2, \dots be independent nonidentically distributed \mathcal{X} -valued random variables such that $\mathbb{E}[\frac{1}{2} \text{dist}(X_m, \mathbf{o})^2] < \infty$ for a given reference point $\mathbf{o} \in \mathcal{X}$ (hence for all points in \mathcal{X}), for each m . Suppose also that the uniform local minimum property obtains: there is fixed finite $\rho_1 > 0$ such that Inequality (3.2) holds for each positive $\rho_0 \leq \rho_1$. Thus there is $\kappa = \kappa(\rho_0, \rho_1)$ such that $(1 + \kappa)\phi_n(\mathbf{o})$ (for the aggregate energy function ϕ_n specified above) is a strict lower bound for the values of ϕ_n on the annulus centred at \mathbf{o} and defined by radii ρ_0, ρ_1 . Finally, suppose that the X_m satisfy a local condition of Lindeberg type at \mathbf{o} : for each $\varepsilon > 0$, as $n \rightarrow \infty$ so*

$$\frac{1}{\phi_n(\mathbf{o})} \sum_{m=1}^n \mathbb{E}[\text{dist}(X_m, \mathbf{o})^2; \text{dist}(X_m, \mathbf{o})^2 > \varepsilon \phi_n(\mathbf{o})] \longrightarrow 0. \tag{3.9}$$

Consider any measurable choice of a sequence of local minimizers

$$\mathcal{E}(X_1, \dots, X_n) = \arg \inf_{x \in \text{ball}(\mathbf{o}, \rho_1)} \left\{ \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, x)^2 \right\}.$$

There exists at least one such sequence such that

$$\mathbb{P}[\mathcal{E}(X_1, \dots, X_n) \in \text{ball}(\mathbf{o}, \rho_0)] \rightarrow 1,$$

and for any such sequence $\mathcal{E}(X_1, \dots, X_n) \rightarrow \mathbf{o}$ in probability.

Proof. First note that global (and hence also local) minimizers of the aggregate empirical energy function always exist and are confined to an almost surely bounded region: indeed global minimizers for the sample X_1, \dots, X_n are simply conventional Fréchet means of the n -point empirical distribution, and the argument of [Bhattacharya and Patrangenaru \(2003, Theorem 2.1\)](#) applies (this theorem is stated for Riemannian manifolds, but the portion relating to existence within a bounded region is a purely metric space argument, using the compactness of bounded sets).

Evidently it suffices to show that Inequality (3.2) has high probability of being replicated at the empirical level: it is enough to show that the following probability converges to 1 as $n \rightarrow \infty$ for each positive $\rho_0 < \rho_1$:

$$\mathbb{P} \left[\sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, \mathbf{o})^2 < \inf \left\{ \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, y)^2; \rho_0 \leq \text{dist}(y, \mathbf{o}) \leq \rho_1 \right\} \right]. \quad (3.10)$$

For then it follows immediately that any sequence of local minimizers of the aggregate empirical energy function *restricted to* $\text{ball}(\mathbf{o}, \rho_1)$

$$\mathcal{E}(X_1, \dots, X_n) = \arg \inf_{x \in \text{ball}(\mathbf{o}, \rho_1)} \left\{ \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, x)^2 \right\},$$

must (as $n \rightarrow \infty$) eventually have arbitrarily high probability of lying in $\text{ball}(\mathbf{o}, \rho_0)$, and must in this event be a local minimizer of the unrestricted aggregate empirical energy function. Since (3.10) holds for each positive $\rho_0 < \rho_1$, we may deduce that $\text{dist}(\mathcal{E}(X_1, \dots, X_n), \mathbf{o}) \rightarrow 0$ in probability.

To begin the proof, first note that the result follows trivially if $\phi_n(\mathbf{o}) = 0$ for all n , for then $X_m = \mathbf{o}$ almost surely for all n . Otherwise by Theorem 1

$$\frac{1}{\phi_n(\mathbf{o})} \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, \mathbf{o})^2 \rightarrow 1 \quad \text{in probability.} \quad (3.11)$$

Furthermore Lemma 1 and (3.9) show that, for each $\varepsilon > 0$, as $n \rightarrow \infty$ so

$$\frac{1}{n\phi_n(\mathbf{o})} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E} \left[\frac{1}{2} \text{dist}(X_i, X_j)^2; \frac{1}{2} \text{dist}(X_i, X_j)^2 > \varepsilon \phi_n(\mathbf{o}) \right] \rightarrow 0.$$

Moreover (3.2) implies that if $\rho_0 \leq \text{dist}(y, \mathbf{o}) \leq \rho_1$ then also

$$\frac{1}{n\phi_n(y)} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E} \left[\frac{1}{2} \text{dist}(X_i, X_j)^2; \frac{1}{2} \text{dist}(X_i, X_j)^2 > \varepsilon \phi_n(y) \right] \rightarrow 0.$$

A further application of Lemma 1 then shows that, for each $\varepsilon > 0$, as $n \rightarrow \infty$

$$\frac{1}{\phi_n(y)} \sum_{m=1}^n \mathbb{E} \left[\frac{1}{2} \text{dist}(X_m, y)^2; \frac{1}{2} \text{dist}(X_m, y)^2 > \varepsilon \phi_n(y) \right] \rightarrow 0.$$

Consequently we may also deduce that if $\rho_0 \leq \text{dist}(y, \mathbf{o}) \leq \rho_1$ then

$$\frac{1}{\phi_n(y)} \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, y)^2 \rightarrow 1 \quad \text{in probability.}$$

Now we have established suitable convergence in probability for the energy functions, but only holding pointwise not uniformly. Were we able to uniformize this over the whole of the annulus $A(\rho_0, \rho_1) = \{y : \rho_0 \leq \text{dist}(y, \mathbf{o}) \leq \rho_1\}$, and were we able to overcome the distinction between $\phi_n(\mathbf{o})$ and $\phi_n(y)$ for $y \in A(\rho_0, \rho_1)$, then we would achieve the required convergence for (3.10) via Inequality (3.2). Following [Bhattacharya and Patrangenaru \(2003\)](#), we do this by selecting y_1, \dots, y_k from $A(\rho_0, \rho_1)$ to form a finite δ -net for $A(\rho_0, \rho_1)$, for suitably small $\delta > 0$. Consider two points $y, z \in A(\rho_0, \rho_1)$ with $\text{dist}(y, z) < \delta$. Then we can use $\text{dist}(X_m, y) \leq 1 + \text{dist}(X_m, y)^2$ to deduce

$$\text{dist}(X_m, z)^2 \leq (\text{dist}(X_m, y) + \delta)^2 \leq (1 + 2\delta) \text{dist}(X_m, y)^2 + (2 + \delta)\delta,$$

likewise

$$\text{dist}(X_m, y)^2 \leq (1 + 2\delta) \text{dist}(X_m, z)^2 + (2 + \delta)\delta.$$

Applying this to whichever is the larger of $\text{dist}(X_m, y)^2$, $\text{dist}(X_m, z)^2$, and then using $\text{dist}(X_m, z)^2 \leq (\text{dist}(X_m, y) + \delta)^2 \leq 2 \text{dist}(X_m, y)^2 + 2\delta^2$,

$$\begin{aligned} & \left| \text{dist}(X_m, z)^2 - \text{dist}(X_m, y)^2 \right| \\ & \leq 2\delta \max\{\text{dist}(X_m, z)^2, \text{dist}(X_m, y)^2\} + (2 + \delta)\delta \\ & \leq 4\delta \text{dist}(X_m, y)^2 + 4\delta^3 + (2 + \delta)\delta = (4 \text{dist}(X_m, y)^2 + 4\delta^2 + \delta + 2)\delta. \end{aligned}$$

For $z \in A(\rho_0, \rho_1)$, choose $p(z)$ to be an element of the δ -net which is closest to z . Then the above implies that

$$\begin{aligned} & \sup \left\{ \left| 1 - \frac{1}{\phi_n(p(z))} \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, z)^2 \right|; z \in A(\rho_0, \rho_1) \right\} \\ & \leq \max_{i=1, \dots, k} \left\{ \left| 1 - \frac{1}{\phi_n(y_i)} \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, y_i)^2 \right| \right. \\ & \quad \left. + \left(\frac{4}{\phi_n(y_i)} \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, y_i)^2 + \frac{4\delta^2 + \delta + 2}{2} \frac{n}{\phi_n(y_i)} \right) \delta \right\}. \end{aligned}$$

Thus we establish useful limiting bounds holding in probability as $n \rightarrow \infty$ so long as we can show that if $y \in A(\rho_0, \rho_1)$ then

$$\liminf_n \frac{\phi_n(y)}{n} > 0.$$

But this follows (with an explicit lower bound) from Lemma 2: hence

$$\begin{aligned} & \sup \left\{ \left| 1 - \frac{1}{\phi_n(p(z))} \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, z)^2 \right|; z \in A(\rho_0, \rho_1) \right\} \\ & \leq \max_{i=1, \dots, k} \left\{ \left| 1 - \frac{1}{\phi_n(y_i)} \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, y_i)^2 \right| \right. \\ & \quad \left. + \left(\frac{4}{\phi_n(y_i)} \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, y_i)^2 + \frac{4\delta^2 + \delta + 2}{2} \frac{16}{\rho_0^2} \right) \delta \right\}. \end{aligned}$$

Consequently, once ρ_0 is fixed, for any $\varepsilon > 0$ we can choose δ small enough so that with probability tending to 1 as $n \rightarrow \infty$

$$\sup \left\{ \left| 1 - \frac{1}{\phi_n(p(z))} \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, z)^2 \right|; z \in A(\rho_0, \rho_1) \right\} \leq \frac{\varepsilon}{2}.$$

We now use (3.2) to deduce that with probability tending to 1 as $n \rightarrow \infty$

$$\begin{aligned} & \inf \left\{ \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, z)^2; z \in A(\rho_0, \rho_1) \right\} \\ & \geq (1 + \kappa) \phi_n(\mathbf{o}) \inf \left\{ \frac{1}{\phi_n(p(z))} \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, z)^2; z \in A(\rho_0, \rho_1) \right\} \\ & \geq \left(1 - \frac{\varepsilon}{2} \right) (1 + \kappa) \phi_n(\mathbf{o}) \geq (1 - \varepsilon) (1 + \kappa) \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, \mathbf{o})^2, \end{aligned}$$

where the last step uses the convergence in probability noted in (3.11). This establishes that the quantity in (3.10) must converge to 1; this completes the proof of the theorem. \square

We have therefore shown that sequences of local empirical Fréchet means must converge in probability to a reference point \mathbf{o} when this reference point is uniformly a strict local minimum of the aggregate energy function so long as a condition of Lindeberg-type is satisfied at \mathbf{o} . Under the additional condition of a linear bound on the growth of $\phi_n(\mathbf{o})$ it is possible also to control the behaviour of global minimizers and derive a result for global empirical Fréchet means.

Corollary 1. *In the situation of Theorem 2, suppose that condition (3.2) holds for all positive ρ_1 (thus in particular \mathbf{o} is the unique global Fréchet mean), and suppose in addition that there is a positive constant C such that*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \phi_n(\mathbf{o}) = \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n \mathbb{E} \left[\frac{1}{2} \text{dist}(X_m, \mathbf{o})^2 \right] \leq C^2. \tag{3.12}$$

Then any measurably selected sequence of local empirical Fréchet means converges to \mathbf{o} in probability.

Proof. Following the proof of Theorem 2, it would suffice to show that, for sufficiently large ρ_1 ,

$$\mathbb{P} \left[\frac{1}{n} \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, \mathbf{o})^2 + 1 \leq \inf \left\{ \frac{1}{n} \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, y)^2 : \text{dist}(y, \mathbf{o}) > \rho_1 \right\} \right]$$

converges to 1 as $n \rightarrow \infty$. To establish this, we once again adapt methods from the proof of [Bhattacharya and Patrangenaru \(2003, Theorem 2.3\)](#). First observe that we can apply the Cauchy–Schwartz inequality to show that

$$\begin{aligned} \frac{1}{n} \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, y)^2 &\geq \frac{1}{n} \sum_{m=1}^n \frac{1}{2} (\text{dist}(X_m, \mathbf{o}) - \text{dist}(y, \mathbf{o}))^2 \\ &\geq \frac{1}{2} \text{dist}(y, \mathbf{o})^2 + \frac{1}{n} \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, \mathbf{o})^2 \\ &\quad - \sqrt{2} \text{dist}(y, \mathbf{o}) \sqrt{\frac{1}{n} \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, \mathbf{o})^2}. \end{aligned}$$

As before, if $\phi_n(\mathbf{o}) = 0$ for all n then the corollary follows immediately. Otherwise from Theorem 1 and the local Lindeberg condition we know that

$$\frac{1}{\phi_n(\mathbf{o})} \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, \mathbf{o})^2 \rightarrow 1 \quad \text{in probability,}$$

and hence the growth condition (3.12) shows that as $n \rightarrow \infty$ so (for example)

$$\mathbb{P} \left[\frac{1}{n} \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, \mathbf{o})^2 \leq 2C^2 \right] \rightarrow 1.$$

This can be applied as follows; if we choose ρ_1 to exceed $2C + \sqrt{2 + 4C^2}$ then, with probability increasing to 1, as $n \rightarrow \infty$

$$\frac{1}{2} \text{dist}(y, \mathbf{o})^2 - \sqrt{2} \text{dist}(y, \mathbf{o}) \sqrt{\frac{1}{n} \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, \mathbf{o})^2} - 1 \geq 0$$

once $\text{dist}(y, \mathbf{o}) > \rho_1$. Consequently as $n \rightarrow \infty$ so

$$\mathbb{P} \left[\frac{1}{n} \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, \mathbf{o})^2 + 1 \leq \inf \left\{ \frac{1}{n} \sum_{m=1}^n \frac{1}{2} \text{dist}(X_m, y)^2 : \text{dist}(y, \mathbf{o}) > \rho_1 \right\} \right]$$

converges to 1 as required. □

4 Euclidean interlude

Before we turn to the central limit theorem on Riemannian manifolds, it is helpful to prove a modest variant on the usual central limit theorem for independent Euclidean (vector-valued) random variables, which may be of independent interest, and which could be argued to capture more precisely the conventional statistical use of the idea of a central limit theorem. The reader will see that the arguments in this section are almost entirely classical [see, e.g., Feller (1966)] and the main issue is simply to formulate the result. However we give complete proofs since we have not been able to trace general forms of these results in the literature, and also because the classical proofs must be adapted to the vector-valued nature of the summands.

A natural condition for central limit approximation for normalized partial sums of d -dimensional mean-zero finite-variance independent random vectors Y_1, \dots, Y_n, \dots is that they should satisfy a variant of Lindeberg’s condition: for each $\varepsilon > 0$, as $n \rightarrow \infty$ so

$$\frac{1}{\phi_n} \sum_{m=1}^n \mathbb{E}[\|Y_m\|^2; \|Y_m\|^2 > \varepsilon \phi_n] \rightarrow 0. \tag{4.1}$$

Here we abbreviate $\phi_n = \sum_{m=1}^n \mathbb{E}[\frac{1}{2}\|Y_m\|^2]$; this parallels the $\phi_n(\mathbf{o})$ used in Sections 3 and 5 and leads us to consider the normalized sums $(X_1 + \dots + X_n)/\sqrt{2\phi_n}$. (The factor $\frac{1}{2}$ is awkward in the Euclidean context, but eases details of calculations later in the geometric context of Section 5.) Note that (4.1) corresponds exactly to the local condition of Lindeberg type (3.5) for X_1, X_2, \dots . However it should be clear that (4.1) cannot be sufficient to establish weak convergence to normality of $(Y_1 + \dots + Y_n)/\sqrt{2\phi_n}$; consider two-dimensional examples in which the sequence Y_1, Y_2, \dots alternates between longer and longer stretches of $\mathcal{L}(Y_k) = (N(0, 1), 0)$ versus longer and longer stretches of $\mathcal{L}(Y_k) = (0, N(0, 1))$. So we cannot hope for a central *limit* theorem (thus the Cramér-Wold device is inapplicable); however it is the case that in fact (4.1) implies a central *approximation* theorem.

In order to describe the result we first recall that the topology of weak convergence of probability measures can be metrized using a truncated Wasserstein distance

$$\widetilde{W}_1(\mu, \nu) = \inf\{\mathbb{E}[1 \wedge \|U - V\|] : \mathcal{L}(U) = \mu, \mathcal{L}(V) = \nu\} \tag{4.2}$$

[see, e.g., Villani (2003, Chapter 7)]. Moreover by Kantorovich–Rubinstein representation [Villani (2003, Remark 7.5(i))] we may write

$$\widetilde{W}_1(\mu, \nu) = \sup \left\{ \int f \, d(\mu - \nu) : f \text{ is Lip}(1) \text{ for distance } 1 \wedge \|x - y\| \right\}. \quad (4.3)$$

We now consider when the law of $(Y_1 + \dots + Y_n)/\sqrt{2\phi_n}$ draws ever closer to the matching (but varying) multivariate normal distribution as $n \rightarrow \infty$.

Theorem 3 (Lindeberg central approximation theorem for vector-valued random variables). *Suppose that Y_1, \dots, Y_n, \dots are independent zero-mean random d -dimensional vectors with finite variance–covariance matrices and that the above variant of Lindeberg’s condition (4.1) is satisfied. Then*

$$\widetilde{W}_1 \left(\frac{Y_1 + \dots + Y_n}{\sqrt{2\phi_n}}, Z_n \right) \rightarrow 0,$$

where $\phi_n = \sum_{m=1}^n \mathbb{E}[\frac{1}{2}\|Y_m\|^2]$ and Z_n has the multivariate d -dimensional normal distribution of zero mean and variance–covariance matrix V_n , with

$$u^\top V_n u = \frac{1}{\phi_n} \sum_{m=1}^n \mathbb{E} \left[\frac{1}{2} \langle u, Y_m \rangle^2 \right] \quad \text{for all vectors } u. \quad (4.4)$$

Remark 3. Note that the variance–covariance matrix V_n has unit trace.

Proof of Theorem 3. The proof is based heavily on the classic proof of the Feller–Lindeberg central limit theorem using characteristic functions. First of all, observe that it is a consequence of the variant Lindeberg condition that

$$\sup_{m=1, \dots, n} \frac{1}{\phi_n} \mathbb{E}[\|Y_m\|^2] \rightarrow 0.$$

For otherwise we can find a subsequence $\{n_r\}$ and m_r in $1, \dots, n_r$ such that for some positive $c > 0$ we have $\mathbb{E}[\|Y_{m_r}\|^2] \geq c\phi_{n_r}$ for all r , and if we choose $\varepsilon < c$ then this implies that

$$\sum_{m=1}^{n_r} \mathbb{E}[\|Y_m\|^2; \|Y_m\|^2 > \varepsilon\phi_{n_r}] \geq \mathbb{E}[\|Y_{m_r}\|^2; \|Y_{m_r}\|^2 > \varepsilon\phi_{n_r}] \geq (c - \varepsilon)\phi_{n_r}.$$

Choosing $\varepsilon < c$, this contradicts the variant Lindeberg condition (4.1). Thus we can choose $N = N(u)$ large enough that $\mathbb{E}[\frac{1}{2}\langle u, Y_m \rangle^2] < \phi_n$ for all $m = 1, \dots, n$ and all $n \geq N(u)$.

Using independence, set

$$\Psi_n(u) = \mathbb{E} \left[\exp \left(i \frac{\langle u, Y_1 + \dots + Y_n \rangle}{\sqrt{2\phi_n}} \right) \right] = \prod_{m=1}^n \mathbb{E} \left[\exp \left(i \frac{\langle u, Y_m \rangle}{\sqrt{2\phi_n}} \right) \right].$$

By estimates based on Taylor expansion [Billingsley (1986, Section 27)],

$$\begin{aligned} & \left| 1 + i \frac{\langle u, Y_m \rangle}{\sqrt{2\phi_n}} - \frac{1}{2} \frac{\langle u, Y_m \rangle^2}{2\phi_n} - \exp\left(i \frac{\langle u, Y_m \rangle}{\sqrt{2\phi_n}}\right) \right| \\ & \leq \left(\frac{\langle u, Y_m \rangle^2}{2\phi_n} \right) \wedge \left(\frac{|\langle u, Y_m \rangle|^3}{(2\phi_n)^{3/2}} \right) \\ & \leq \max\{1, \|u\|^3\} \left\{ \frac{\|Y_m\|^2}{2\phi_n} \mathbb{I}[\|Y_m\|^2 > \varepsilon\phi_n] + \frac{\|Y_m\|^3}{(2\phi_n)^{3/2}} \mathbb{I}[\|Y_m\|^2 \leq \varepsilon\phi_n] \right\}. \end{aligned}$$

Hence for $n \geq N(u)$,

$$\begin{aligned} & \left| \Psi_n(u) - \prod_{m=1}^n \left(1 - \frac{1}{2} \frac{\mathbb{E}[\langle u, Y_m \rangle^2]}{2\phi_n} \right) \right| \\ & \leq \max\{1, \|u\|^3\} \left\{ \frac{1}{2\phi_n} \sum_{m=1}^n \mathbb{E}[\|Y_m\|^2; \|Y_m\|^2 > \varepsilon\phi_n] + \sqrt{\frac{\varepsilon}{2}} \sum_{m=1}^n \frac{\mathbb{E}[\|Y_m\|^2]}{2\phi_n} \right\} \\ & = \max\{1, \|u\|^3\} \left\{ \frac{1}{2\phi_n} \sum_{m=1}^n \mathbb{E}[\|Y_m\|^2; \|Y_m\|^2 > \varepsilon\phi_n] + \sqrt{\frac{\varepsilon}{2}} \right\} \end{aligned}$$

(recalling the definition of ϕ_n for the last step, and noting that for $n \geq N(u)$ we know that every $\frac{1}{2} \frac{\mathbb{E}[\langle u, Y_m \rangle^2]}{2\phi_n}$ is bounded above by $\frac{1}{2}$).

Now invoke the inequality

$$e^{-p/(1-p)} \leq 1 - p \leq e^{-p},$$

valid for $0 \leq p < 1$. Since $\frac{1}{2} \frac{\mathbb{E}[\langle u, Y_m \rangle^2]}{2\phi_n} < \frac{1}{2}$ when $n \geq N(u)$,

$$\begin{aligned} 0 & \leq \log \prod_{m=1}^n \exp\left(-\frac{1}{2} \frac{\mathbb{E}[\langle u, Y_m \rangle^2]}{2\phi_n}\right) - \log \prod_{m=1}^n \left(1 - \frac{1}{2} \frac{\mathbb{E}[\langle u, Y_m \rangle^2]}{2\phi_n} \right) \\ & \leq \sum_{m=1}^n \frac{1}{1 - (1/2)\mathbb{E}[\langle u, Y_m \rangle^2]/(2\phi_n)} \left(\frac{1}{2} \frac{\mathbb{E}[\langle u, Y_m \rangle^2]}{2\phi_n} \right)^2 \leq \frac{1}{2} \max_{m=1, \dots, n} \frac{\mathbb{E}[\langle u, Y_m \rangle^2]}{2\phi_n} \\ & \leq \frac{\|u\|^2}{2} \max_{m=1, \dots, n} \frac{\mathbb{E}[\|Y_m\|^2]}{2\phi_n} \rightarrow 0. \end{aligned}$$

Accordingly we may use (4.4) to deduce that if $n \geq N(u)$ then

$$\left| \Psi_n(u) - \exp\left(-\frac{1}{2} u^\top V_n u\right) \right| \leq \max\{1, \|u\|^3\} \times A_n, \tag{4.5}$$

$$A_n = \sqrt{\frac{\varepsilon}{2}} + \frac{1}{2} \max_{m=1, \dots, n} \frac{\mathbb{E}[\|Y_m\|^2]}{2\phi_n} + \frac{1}{2\phi_n} \sum_{m=1}^n \mathbb{E}[\|Y_m\|^2; \|Y_m\|^2 > \varepsilon\phi_n]. \tag{4.6}$$

Since ε can be chosen to be arbitrarily small, and the variant Lindeberg condition (4.1) implies the other quantities converge to 0, it follows that $|\Psi_n(u) - \exp(-\frac{1}{2}u^\top V_n u)|$ converges to 0 for each fixed u .

We now convert this relationship between characteristic functions into an inequality for the truncated Wasserstein distance between the corresponding distributions. To this end we use a Parseval equality [Feller (1966, XV.3)]:

$$e^{-i\langle u, t \rangle} \Psi_n(u) = \mathbb{E} \left[\exp \left(i \left\langle u, \frac{Y_1 + \dots + Y_n}{\sqrt{2\phi_n}} - t \right\rangle \right) \right].$$

We can multiply by the symmetric d -dimensional normal density of variance σ^{-2} , integrate with respect to u , and rearrange to obtain

$$\begin{aligned} & \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle u, t \rangle} \Psi_n(u) e^{-\sigma^2|u|^2/2} du \\ &= \frac{1}{(2\pi\sigma^2)^{d/2}} \mathbb{E} \left[\exp \left(-\frac{1}{2\sigma^2} \left\| t - \frac{Y_1 + \dots + Y_n}{\sqrt{2\phi_n}} \right\|^2 \right) \right]. \end{aligned} \tag{4.7}$$

The right-hand side (viewed as a function of t) is the density of $\frac{Y_1 + \dots + Y_n}{\sqrt{2\phi_n}} + Z'$, where Z' has a d -dimensional multivariate normal distribution of variance–covariance matrix $\sigma^2 \mathbb{I}_d$, independent of $\frac{Y_1 + \dots + Y_n}{\sqrt{2\phi_n}}$. By the definition (4.2) of Wasserstein distance the truncated Wasserstein distance between the distribution of $\frac{Y_1 + \dots + Y_n}{\sqrt{2\phi_n}}$ and the distribution of $\frac{Y_1 + \dots + Y_n}{\sqrt{2\phi_n}} + Z'$ is bounded by

$$\mathbb{E}[\|Z'\|] \leq \text{constant} \times \sigma.$$

Given any $\eta > 0$, we can choose σ to make this smaller than $\eta/5$.

Choose Z_n to be of d -dimensional multivariate normal distribution with variance–covariance matrix V_n , independent of Z' . The truncated Wasserstein distance between the distributions Z_n and $Z_n + Z'$ satisfies the same bound of $\eta/5$. So consider bounds on the truncated Wasserstein distance between the distributions of (a) $\frac{Y_1 + \dots + Y_n}{\sqrt{2\phi_n}} + Z'$, with density given by (4.7), and (b) $Z_n + Z'$ whose density satisfies a similar formula but with the normal characteristic function $\exp(-\frac{1}{2}u^\top V_n u)$ replacing $\Psi_n(u)$. By the Kantorovich–Rubinstein representation (4.3) of the truncated Wasserstein distance we may consider

$$\left| \mathbb{E} \left[f \left(\frac{Y_1 + \dots + Y_n}{\sqrt{2\phi_n}} + Z' \right) \right] - \mathbb{E} [f(Z_n + Z')] \right|,$$

where f is Lip(1) with respect to the *truncated* distance function $1 \wedge \|x - y\|$ [see (4.3)]. Without loss of generality we take $f(\mathbf{0}) = 0$; the Lipschitz condition then implies that $|f| \leq 1$ (since the truncated distance $1 \wedge \|x - y\|$ is always bounded above by 1). Now both $\frac{Y_1 + \dots + Y_n}{\sqrt{2\phi_n}} + Z'$ and $Z_n + Z'$ have variance–covariance matrices with traces bounded above by $1 + \sigma^2 d$; therefore once σ is fixed we may

choose a large radius R and deduce by Chebyshev that the distributions of both $\frac{Y_1 + \dots + Y_n}{\sqrt{2\phi_n}} + Z'$ and $Z_n + Z'$ place probability mass of at most $\eta/5$ outside the ball centred on \mathbf{o} and of radius R , so that

$$\begin{aligned} \left| \mathbb{E} \left[f \left(\frac{Y_1 + \dots + Y_n}{\sqrt{2\phi_n}} + Z' \right); \left\| \frac{Y_1 + \dots + Y_n}{\sqrt{2\phi_n}} + Z' \right\| > R \right] \right| &\leq \eta/5, \\ \left| \mathbb{E} [f(Z_n + Z'); \|Z_n + Z'\| > R] \right| &\leq \eta/5. \end{aligned}$$

Finally

$$\begin{aligned} &\left| \mathbb{E} \left[f \left(\frac{Y_1 + \dots + Y_n}{\sqrt{2\phi_n}} + Z' \right); \left\| \frac{Y_1 + \dots + Y_n}{\sqrt{2\phi_n}} + Z' \right\| \leq R \right] \right. \\ &\quad \left. - \mathbb{E} [f(Z_n + Z'); \|Z_n + Z'\| \leq R] \right| \\ &\leq \int_{\text{ball}(\mathbf{o}, R)} \left| \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle u, t \rangle} \Psi_n(u) e^{-\sigma^2 |u|^2/2} \, du \right. \\ &\quad \left. - \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle u, t \rangle} e^{-(1/2)u^\top V_n u} e^{-\sigma^2 |u|^2/2} \, du \right| dt \\ &\leq \frac{1}{(2\pi)^d} \int_{\text{ball}(\mathbf{o}, R)} \int_{\mathbb{R}^d} |\Psi_n(u) - e^{-(1/2)u^\top V_n u}| e^{-\sigma^2 |u|^2/2} \, du \, dt. \end{aligned}$$

Given σ and R , the dominated convergence theorem allows us to choose N (not depending on u) to make this arbitrarily small for all $n \geq N$, hence

$$\begin{aligned} &\left| \mathbb{E} \left[f \left(\frac{Y_1 + \dots + Y_n}{\sqrt{2\phi_n}} + Z' \right); \left\| \frac{Y_1 + \dots + Y_n}{\sqrt{2\phi_n}} + Z' \right\| \leq R \right] \right. \\ &\quad \left. - \mathbb{E} [f(Z_n + Z'); \|Z_n + Z'\| \leq R] \right| \leq \eta/5 \end{aligned}$$

for all $n \geq N$. It therefore follows that for $n \geq N$ we obtain

$$\tilde{W}_1 \left(\mathcal{L} \left(\frac{Y_1 + \dots + Y_n}{\sqrt{2\phi_n}} \right), Z_n \right) \leq \eta,$$

and since $\eta > 0$ was arbitrary the theorem follows. □

The following converse to this result mirrors Feller’s converse to Lindeberg’s theorem.

Corollary 2 (Feller converse to Lindeberg central approximation theorem).
In the situation of Theorem 3, suppose that in place of the above variant of the Lindeberg condition (4.1) it is the case that

$$\frac{1}{\phi_n} \mathbb{E} [\|Y_n\|^2] \rightarrow 0 \quad \text{and} \quad \phi_n \rightarrow \infty \tag{4.8}$$

and that

$$\widetilde{W}_1\left(\frac{Y_1 + \cdots + Y_n}{\sqrt{2\phi_n}}, Z_n\right) \rightarrow 0. \tag{4.9}$$

Then the Lindeberg condition (4.1) must be satisfied.

Proof. As a consequence of (4.8) and the fact that ϕ_n increases with n ,

$$\lim_{n \rightarrow \infty} \max_{1 \leq m \leq n} \frac{\mathbb{E}[\|Y_m\|^2]}{\phi_n} \leq \lim_{n \rightarrow \infty} \max_{1 \leq m \leq k} \frac{\mathbb{E}[\|Y_m\|^2]}{\phi_n} + \lim_{n \rightarrow \infty} \max_{k < m} \frac{\mathbb{E}[\|Y_m\|^2]}{\phi_m}$$

also tends to zero. For fixed $u \in \mathbb{R}^d$, the bounded Lipschitz nature of $\exp(i\langle u, x \rangle)$ as a function of x , applied to (4.9) and the Kantorovich–Rubinstein characterization (4.3) together imply that

$$\mathbb{E}\left[\exp\left(i\frac{\langle u, Y_1 + \cdots + Y_m \rangle}{\sqrt{2\phi_n}}\right)\right] - \exp\left(-\frac{1}{2}\langle u, V_n u \rangle\right) \rightarrow 0.$$

Since V_n has unit trace, we can multiply through by $\exp(\frac{1}{2}\langle u, V_n u \rangle)$, take logs and use independence to see that

$$\frac{1}{2}\langle u, V_n u \rangle + \sum_{m=1}^n \log \mathbb{E}\left[\exp\left(i\frac{\langle u, Y_m \rangle}{\sqrt{2\phi_n}}\right)\right] \rightarrow 0.$$

Standard estimates using Taylor expansion show that

$$\begin{aligned} & \left| \log \mathbb{E}\left[\exp\left(i\frac{\langle u, Y_m \rangle}{\sqrt{2\phi_n}}\right)\right] - \left(\mathbb{E}\left[\exp\left(i\frac{\langle u, Y_m \rangle}{\sqrt{2\phi_n}}\right)\right] - 1\right) \right| \\ & \leq \left| \mathbb{E}\left[\exp\left(i\frac{\langle u, Y_m \rangle}{\sqrt{2\phi_n}}\right)\right] - 1 \right|^2, \end{aligned}$$

while

$$\begin{aligned} & \sum_{m=1}^n \left| \mathbb{E}\left[\exp\left(i\frac{\langle u, Y_m \rangle}{\sqrt{2\phi_n}}\right)\right] - 1 \right|^2 \\ & \leq \left(\max_{1 \leq m \leq n} \left| \mathbb{E}\left[\exp\left(i\frac{\langle u, Y_m \rangle}{\sqrt{2\phi_n}}\right)\right] - 1 \right| \right) \times \sum_{m=1}^n \left| \mathbb{E}\left[\exp\left(i\frac{\langle u, Y_m \rangle}{\sqrt{2\phi_n}}\right)\right] - 1 \right| \\ & \leq \max_{1 \leq m \leq n} \frac{\|u\|^2}{2} \frac{\mathbb{E}[\|Y_m\|^2]}{2\phi_n} \times \sum_{m=1}^n \frac{\|u\|^2}{2} \frac{\mathbb{E}[\|Y_m\|^2]}{2\phi_n} = \max_{1 \leq m \leq n} \frac{\|u\|^4}{8} \frac{\mathbb{E}[\|Y_m\|^2]}{\phi_n} \\ & \rightarrow 0. \end{aligned}$$

Thus for fixed u

$$\frac{1}{2}\langle u, V_n u \rangle - \sum_{m=1}^n \mathbb{E}\left[1 - \exp\left(i\frac{\langle u, Y_m \rangle}{\sqrt{2\phi_n}}\right)\right] \rightarrow 0.$$

Taking real parts and splitting the expectation at $\|Y_m\|^2 = \varepsilon\phi_n$,

$$\begin{aligned} \frac{1}{2}\langle u, V_n u \rangle - \sum_{m=1}^n \mathbb{E} \left[1 - \cos \left(\frac{\langle u, Y_m \rangle}{\sqrt{2\phi_n}} \right); \|Y_m\|^2 \leq \varepsilon\phi_n \right] \\ = \sum_{m=1}^n \mathbb{E} \left[1 - \cos \left(\frac{\langle u, Y_m \rangle}{\sqrt{2\phi_n}} \right); \|Y_m\|^2 > \varepsilon\phi_n \right] + o(1), \end{aligned}$$

where we must bear in mind that the $o(1)$ term depends on u . The right-hand side is bounded above by

$$\sum_{m=1}^n \mathbb{E} \left[2 \times \frac{\|Y_m\|^2}{\varepsilon\phi_n}; \|Y_m\|^2 > \varepsilon\phi_n \right] + o(1) \leq \frac{4}{\varepsilon} + o(1);$$

while the left-hand side is bounded below by

$$\begin{aligned} \frac{1}{2}\langle u, V_n u \rangle - \sum_{m=1}^n \mathbb{E} \left[\frac{1}{2} \left(\frac{\langle u, Y_m \rangle}{\sqrt{2\phi_n}} \right)^2; \|Y_m\|^2 \leq \varepsilon\phi_n \right] \\ = \sum_{m=1}^n \mathbb{E} \left[\frac{1}{2} \left(\frac{\langle u, Y_m \rangle}{\sqrt{2\phi_n}} \right)^2; \|Y_m\|^2 > \varepsilon\phi_n \right] \end{aligned}$$

by (4.4); thus

$$\frac{1}{\phi_n} \sum_{m=1}^n \mathbb{E} \left[\left\langle \frac{u}{\|u\|}, Y_m \right\rangle^2; \|Y_m\|^2 > \varepsilon\phi_n \right] \leq \frac{1}{\|u\|^2} \left(\frac{16}{\varepsilon} + o(1) \right).$$

The variant Lindeberg condition (4.1) now follows by summing over vectors $u/\|u\|$ forming an orthonormal basis, and choosing suitably large $\|u\|$. \square

5 Central limit theory for empirical Fréchet means

In order to discuss the second-order theory of empirical Fréchet means, namely central limit theorems, we augment the metric space structure of \mathcal{X} by moving to the context of a complete and connected Riemannian manifold \mathbb{M} of dimension d . Let $\text{dist}(x, y)$ be the Riemannian distance between points $x, y \in \mathbb{M}$. For any $x \in \mathbb{M}$, let \mathcal{C}_x denote the cut locus of x . Let $\text{Exp}_x : T_x\mathbb{M} \rightarrow \mathbb{M}$ be the Exponential map from the tangent space $T_x\mathbb{M}$ to \mathbb{M} ; observe that $\text{Exp}_x^{-1}(y)$ can be defined uniquely for $y \notin \mathcal{C}_x$ by $\text{Exp}_x^{-1}(y) = \gamma'(1)$, where $\gamma : [0, 1] \rightarrow \mathbb{M}$ is the unique minimal geodesic running from x to y . Now let $\Pi_{x,y} : T_x\mathbb{M} \rightarrow T_y\mathbb{M}$ be the parallel transport map along the geodesic γ , and note that $\Pi_{x,y}^{-1} = \Pi_{y,x}$, both being defined when $x \notin \mathcal{C}_y$ equivalently $y \notin \mathcal{C}_x$. Finally, denote the covariant derivative by ∇ : if U is a smooth vectorfield and γ is a geodesic then the covariant derivative of U at $\gamma(0)$ in the direction $\gamma'(0)$ is given by

$$\nabla_{\gamma'(0)} U = \lim_{s \downarrow 0} \frac{\Pi_{\gamma(s), \gamma(0)} U(\gamma(s)) - U(\gamma(0))}{s}.$$

Moreover $\nabla_{\gamma'(0)}$ depends only on the tangent vector $\gamma'(0)$, rather than the actual curve γ .

Our discussion concerns a sequence of independent (but *not* identically distributed) random variables X_1, X_2, \dots , taking values in \mathbb{M} , for which each $\mathbb{E}[\text{dist}(x, X_i)^2]$ is finite for some (and therefore for all) x , and which share a common Fréchet mean $\mathbf{o} \in \mathbb{M}$. Furthermore we suppose that

$$\mathbb{P}[X_n \in \mathcal{C}_{\mathbf{o}}] = 0 \quad \text{for } n \geq 1. \tag{5.1}$$

For each n we choose $\mathcal{E}(X_1, \dots, X_n)$ to be a measurably selected empirical local Fréchet mean of X_1, \dots, X_n , and we suppose it possible to make these choices so that $\mathcal{E}(X_1, \dots, X_n)$ converges to \mathbf{o} in probability. (Theorem 2 delineates a large class of cases in which this can be done.)

For each $i \geq 1$ we can define a random vectorfield Y_i on $\mathbb{M} \setminus \mathcal{C}_{X_i}$ by

$$Y_i(x) = \text{Exp}_x^{-1}(X_i). \tag{5.2}$$

Here we use the definition of Exp_x^{-1} on $\mathbb{M} \setminus \mathcal{C}_x$; in the cases when $\text{Exp}_x^{-1}(X_i)$ is not defined we choose Y_i measurably but otherwise arbitrarily from the preimage of X_i under Exp_x . In fact it can be shown that (5.2) defines $Y_i(x)$ uniquely for almost all x with probability 1; moreover the cut locus condition (5.1) ensures that $Y_i(\mathbf{o})$ in particular is almost surely well defined.

Since \mathbf{o} is a Fréchet mean of each X_i , it follows that $\mathbb{E}[Y_i(\mathbf{o})] = 0$; moreover the finiteness of $\mathbb{E}[\text{dist}(\mathbf{o}, X_i)^2]$ implies the finiteness of $\mathbb{E}[\|Y_i\|^2]$, which is the trace of the variance–covariance matrix of the random vector Y_i . Moreover the calculus of manifolds shows that

$$Y_i(x) = -\text{dist}(x, X_i) \text{grad}_x \text{dist}(x, X_i) = \text{grad}_x \left(-\frac{1}{2} \text{dist}(x, X_i)^2 \right). \tag{5.3}$$

Indeed, if $x \in \mathbb{M} \setminus \mathcal{C}_{X_i}$ then covariant differentiation defines a symmetric $(d \times d)$ tensor $H_i(x) = -(\nabla_U Y_i)(x)$, acting on vectorfields U, V by

$$\langle H_i U, V \rangle(x) = \langle -\nabla_U Y_i, V \rangle(x) = \text{Hess}_x \left(\frac{1}{2} \text{dist}(x, X_i)^2 \right)(U, V). \tag{5.4}$$

[The sign of H_i is chosen so that if $X_i = \mathbf{o}$ then $H_i(\mathbf{o})$ is the identity tensor.]

As noted above, the assumption that \mathbf{o} is a Fréchet mean of X_i for all $i \geq 1$ implies that $Y_i(\mathbf{o}) = \text{Exp}_{\mathbf{o}}^{-1}(X_i)$ determines a sequence of independent random variables with zero mean on $T_{\mathbf{o}}(\mathbb{M})$. Then Theorem 3 and Corollary 2 capture the conditions under which the normalized sum $(Y_1(\mathbf{o}) + \dots + Y_n(\mathbf{o})) / \sqrt{2\phi_n(\mathbf{o})}$ is asymptotically multivariate normal (where ϕ_n is the aggregate energy function as defined in Section 2). Moreover a first-order Taylor expansion argument suggests that (under further regularity conditions) the Exponential map of a suitable transformation of this normalized sum should approximate the local empirical Fréchet mean $\mathcal{E}(X_1, \dots, X_n)$; this corresponds to an application of Newton’s root-finding method. This is indeed the case, and forms the main result of this section. However

before we turn to this we must first prove a preliminary geometric result, required in order to control the effects of the approximation.

We begin by constructing a certain orthonormal frame field e_1, \dots, e_d over $\mathbb{M} \setminus \mathcal{C}_\mathbf{o}$. Pick $e_1(\mathbf{o}), \dots, e_d(\mathbf{o})$ to be an orthonormal basis for $T_\mathbf{o}(\mathbb{M})$, and extend by parallel transport along minimal geodesics from \mathbf{o} over all of $\mathbb{M} \setminus \mathcal{C}_\mathbf{o}$: $e_r(x) = \Pi_{\mathbf{o},x} e_r(\mathbf{o})$, for $x \in \mathbb{M} \setminus \mathcal{C}_\mathbf{o}$. By the properties of geodesic normal coordinates, the vectorfields $\nabla_{e_r} e_s$ all vanish at \mathbf{o} .

Lemma 3. *For given $\varepsilon > 0$, choose $\rho > 0$ such that $\text{ball}(\mathbf{o}, \rho) \subseteq \mathbb{M} \setminus \mathcal{C}_\mathbf{o}$ and $\|\nabla_{e_r} e_s\| < \varepsilon/d$ within $\text{ball}(\mathbf{o}, \rho)$, for $r, s = 1, \dots, d$. Set $Z_{r,i} = \langle Y_i, e_r \rangle e_r$, for some Y_i . Then [viewing $\nabla Z_{r,i}$ as a symmetric $(d \times d)$ tensor] for $x \in \text{ball}(\mathbf{o}, \rho)$ we have*

$$\begin{aligned} \|\Pi_{x,\mathbf{o}} \nabla Z_{r,i}(x) - \nabla Z_{r,i}(\mathbf{o})\| &\leq (1 + 2\varepsilon\rho) \sup_{x' \in \text{ball}(\mathbf{o}, \rho)} \|\Pi_{x',\mathbf{o}} \nabla Y_i(x') - \nabla Y_i(\mathbf{o})\| \\ &\quad + 2\varepsilon(\|Y_i(\mathbf{o})\| + \|\nabla Y_i(\mathbf{o})\|\rho). \end{aligned} \quad (5.5)$$

Proof. We suppress the dependence on the suffix i for the sake of convenience of exposition, and write $Z_r = Z_{r,i}$, $Y = Y_i$. First consider $\nabla_V Z_r$ for a general smooth vectorfield V . By the calculus of covariant differentiation

$$\nabla_V Z_r = \nabla_V(\langle Y, e_r \rangle e_r) = \langle \nabla_V Y, e_r \rangle e_r + \langle Y, \nabla_V e_r \rangle e_r + \langle Y, e_r \rangle \nabla_V e_r.$$

Because $\nabla_V e_r$ vanishes at \mathbf{o} ,

$$\begin{aligned} \Pi_{x,\mathbf{o}} \nabla_V Z_r(x) - \nabla_V Z_r(\mathbf{o}) &= (\langle \nabla_V Y, e_r \rangle(x) - \langle \nabla_V Y, e_r \rangle(\mathbf{o})) e_r(\mathbf{o}) \\ &\quad + \langle Y, \nabla_V e_r \rangle(x) e_r(\mathbf{o}) + \langle Y, e_r \rangle(x) \Pi_{x,\mathbf{o}} \nabla_V e_r(x). \end{aligned}$$

The coefficient of $e_r(\mathbf{o})$ in the first term on the right-hand side can be rewritten as the evaluation of $\langle \Pi_{x,\mathbf{o}} \nabla_V Y - \nabla_V Y, e_r \rangle e_r$ at \mathbf{o} ; the other two terms can be expanded to achieve

$$\begin{aligned} \Pi_{x,\mathbf{o}} \nabla_V Z_r(x) - \nabla_V Z_r(\mathbf{o}) &= \langle \Pi_{x,\mathbf{o}} \nabla_V Y - \nabla_V Y, e_r \rangle(\mathbf{o}) e_r(\mathbf{o}) \\ &\quad + \langle Y, \Pi_{x,\mathbf{o}} \nabla_V e_r \rangle(\mathbf{o}) e_r(\mathbf{o}) + \langle \Pi_{x,\mathbf{o}} Y - Y, \Pi_{x,\mathbf{o}} \nabla_V e_r \rangle(\mathbf{o}) e_r(\mathbf{o}) \\ &\quad + \langle Y, e_r \rangle(\mathbf{o}) (\Pi_{x,\mathbf{o}} \nabla_V e_r)(\mathbf{o}) + \langle \Pi_{x,\mathbf{o}} Y - Y, e_r \rangle(\mathbf{o}) (\Pi_{x,\mathbf{o}} \nabla_V e_r)(\mathbf{o}). \end{aligned}$$

To control the size of the matrix $M = \Pi_{x,\mathbf{o}} \nabla_V Z_r - \nabla_V Z_r$ at \mathbf{o} we shall use the Frobenius norm $\|M\| = \sqrt{M^\top M} = \sqrt{\sum_{i,j} M_{i,j}^2}$. Now V is an arbitrary vectorfield, hence (evaluating tensor and vectorfields at \mathbf{o} throughout) we may deduce that

$$\begin{aligned} \|\Pi_{x,\mathbf{o}} \nabla_V Z_r(x) - \nabla_V Z_r(\mathbf{o})\| &\leq \|\Pi_{x,\mathbf{o}} \nabla Y(x) - \nabla Y(\mathbf{o})\| \\ &\quad + 2(\|\Pi_{x,\mathbf{o}} Y(x) - Y(\mathbf{o})\| + \|Y(\mathbf{o})\|) \|\nabla e_r(x)\| \\ &\leq \|\Pi_{x,\mathbf{o}} \nabla Y(x) - \nabla Y(\mathbf{o})\| + 2\varepsilon(\|\Pi_{x,\mathbf{o}} Y(x) - Y(\mathbf{o})\| + \|Y(\mathbf{o})\|) \end{aligned} \quad (5.6)$$

so long as $x \in \text{ball}(\mathbf{o}, \rho)$.

We now apply the Mean Value Theorem to observe that

$$\begin{aligned} & \|\Pi_{x,\mathbf{o}}Y(x) - Y(\mathbf{o})\| \\ & \leq \text{dist}(x, \mathbf{o}) \sup_{x' \in \text{ball}(\mathbf{o}, \rho)} \|\nabla Y(x')\| \\ & \leq \text{dist}(x, \mathbf{o}) \left(\|\nabla Y(\mathbf{o})\| + \sup_{x' \in \text{ball}(\mathbf{o}, \rho)} \|\Pi_{x',\mathbf{o}}\nabla Y(x') - \nabla Y(\mathbf{o})\| \right) \end{aligned} \tag{5.7}$$

and thus (restoring the dependence on the suffix i) we can apply (5.7) to (5.6) and combine with $\text{dist}(\mathbf{o}, x) \leq \rho$ to deduce the required inequality. \square

The above lemma allows us to control the errors arising from the approximation implicit in the Newton method described above. We can now state and prove the main theorem of this section.

Theorem 4. *Suppose that X_1, X_2, \dots are independent nonidentically distributed random variables taking values in \mathbb{M} , such that for all n and all $x \in \mathbb{M}$ the aggregate energy function $\phi_n(x) = \sum_{i=1}^n \frac{1}{2} \mathbb{E}[\text{dist}(x, X_i)^2]$ is finite. Suppose that \mathbf{o} is a local Fréchet mean of each of the X_i and moreover suppose that $\mathbb{P}[X_i \in \mathcal{C}_{\mathbf{o}}] = 0$ for each i . Let $Y_i = \text{Exp}_{\mathbf{o}}^{-1}(X_i)$. Let $x_n = \mathcal{E}(X_1, \dots, X_n)$ be a measurable choice of local empirical Fréchet means such that $x_n \rightarrow \mathbf{o}$ in probability. Suppose that the following conditions hold:*

1. $\phi_n(\mathbf{o})$ is of at least linear growth, so $\liminf_{n \rightarrow \infty} \frac{\phi_n(\mathbf{o})}{n} = C_1 > 0$ for a finite positive constant $C_1 > 0$;
2. For each sufficiently small $\rho > 0$, as $n \rightarrow \infty$ so

$$\frac{1}{\phi_n(\mathbf{o})} \sum_{i=1}^n \sup_{x' \in \text{ball}(\mathbf{o}, \rho)} \mathbb{E}[\|\Pi_{x',\mathbf{o}}H_i(x') - H_i(\mathbf{o})\|] \rightarrow 0,$$

where H_i is as given in (5.4);

3. There is a finite constant C_2 such that

$$\limsup_{n \rightarrow \infty} \frac{1}{\phi_n(\mathbf{o})} \sum_{i=1}^n \mathbb{E}[\|H_i(\mathbf{o})\|^2] \leq C_2;$$

4. Let \tilde{H}_n be the coordinate-wise expectation

$$\tilde{H}_n = \frac{\mathbb{E}[H_1(\mathbf{o}) + \dots + H_n(\mathbf{o})]}{2\phi_n(\mathbf{o})}.$$

Then the symmetric matrix \tilde{H}_n is asymptotically nonsingular; there is a positive constant $C_3 > 0$ with $\limsup_{n \rightarrow \infty} \|\tilde{H}_n^{-1}\| \leq C_3$;

5. Finally we require a condition of Lindeberg type: for each $\varepsilon > 0$, as $n \rightarrow \infty$ so

$$\frac{1}{\phi_n(\mathbf{o})} \sum_{i=1}^n \mathbb{E}[\text{dist}(\mathbf{o}, X_i)^2; \text{dist}(\mathbf{o}, X_i)^2 > \varepsilon \phi_n(\mathbf{o})] \rightarrow 0.$$

Let \tilde{Z}_n have the multivariate normal distribution with zero mean and variance-covariance matrix $\tilde{H}_n^{-1} V_n \tilde{H}_n^{-1}$, where V_n is the variance-covariance matrix of $\frac{Y_1 + \dots + Y_n}{\sqrt{2\phi_n(\mathbf{o})}}$. Then as $n \rightarrow \infty$ so

$$\tilde{W}_1(\sqrt{2\phi_n(\mathbf{o})} \text{Exp}_{\mathbf{o}}^{-1}(x_n), \tilde{Z}_n) \rightarrow 0.$$

Proof. Begin by representing $\sum_{i=1}^n Y_i(x) = \sum_{i=1}^n \text{grad}_x(-\frac{1}{2} \text{dist}(x, X_i)^2)$ by a first-order Taylor series expansion about \mathbf{o} : if γ_x is a minimal geodesic begun at \mathbf{o} and ending at $x \in \mathbb{M} \setminus \mathcal{C}_{\mathbf{o}}$ at unit time then

$$\begin{aligned} \Pi_{x,\mathbf{o}} \sum_{i=1}^n Y_i(x) &= \sum_{i=1}^n Y_i(\mathbf{o}) + \sum_{i=1}^n \nabla_{\gamma'_x(0)} Y_i(\mathbf{o}) + \Delta_n(x) \gamma'_x(0) \\ &= \sum_{i=1}^n Y_i(\mathbf{o}) - \sum_{i=1}^n H_i(\mathbf{o}) \gamma'_x(0) + \Delta_n(x) \gamma'_x(0), \end{aligned}$$

where the Mean Value Theorem can be applied to show that the matrix correction term $\Delta_n(x)$ can be written as

$$\Delta_n(x)U = \sum_{r=1}^d \sum_{i=1}^n (\Pi_{\gamma_x(\theta_r), \mathbf{o}} \nabla_U Z_{r,i}(\gamma_x(\theta_r)) - \nabla_U Z_{r,i}(\mathbf{o}))$$

for $Z_{r,i} = \langle Y_i, e_r \rangle e_r$ as defined in Lemma 3, and for suitable $0 \leq \theta_1, \dots, \theta_d \leq 1$. Choosing $\rho > 0$ given ε as in Lemma 3, if $x \in \text{ball}(\mathbf{o}, \rho)$ then

$$\begin{aligned} \|\Delta_n(x)\| &\leq \sum_{r=1}^d \sum_{i=1}^n \left((1 + 2\varepsilon\rho) \sup_{x' \in \text{ball}(\mathbf{o}, \rho)} \|\Pi_{x', \mathbf{o}} \nabla Y_i(x') - \nabla Y_i(\mathbf{o})\| \right. \\ &\quad \left. + 2\varepsilon(\|Y_i(\mathbf{o})\| + \rho \|\nabla Y_i(\mathbf{o})\|) \right). \end{aligned} \tag{5.8}$$

Now choose $x = x_n = \mathcal{E}(X_1, \dots, X_n)$. Since $\gamma'_x(0) = \text{Exp}_{\mathbf{o}}^{-1}(x)$, it follows that $\Pi_{x_n, \mathbf{o}} \sum_{i=1}^n Y_i(x_n) = 0$. If $x_n = \mathcal{E}(X_1, \dots, X_n) \in \text{ball}(\mathbf{o}, \rho)$ then

$$0 = \sum_{i=1}^n Y_i(\mathbf{o}) - \left(\sum_{i=1}^n H_i(\mathbf{o}) - \Delta_n(x_n) \right) \text{Exp}_{\mathbf{o}}^{-1}(x_n).$$

Consequently, so long as $\sum_{i=1}^n H_i(\mathbf{o}) - \Delta_n(x_n)$ is invertible, we may write

$$x_n = \mathcal{E}(X_1, \dots, X_n) = \text{Exp}_{\mathbf{o}} \left(\left(\sum_{i=1}^n H_i(\mathbf{o}) - \Delta_n(x_n) \right)^{-1} \sum_{i=1}^n Y_i(\mathbf{o}) \right). \tag{5.9}$$

Use the aggregate energy function $\phi_n(x) = \sum_{i=1}^n \mathbb{E}[\frac{1}{2} \text{dist}(X_i, x)^2]$ (defined in Section 3) to adjust the above equation into a form hinting at a central limit approximation for $\mathcal{E}(X_1, \dots, X_n)$:

$$\begin{aligned} & \sqrt{2\phi_n(\mathbf{o})} \times \text{Exp}_{\mathbf{o}}^{-1}(\mathcal{E}(X_1, \dots, X_n)) \\ &= \left(\sum_{i=1}^n \frac{H_i(\mathbf{o})}{2\phi_n(\mathbf{o})} - \frac{\Delta_n(x_n)}{2\phi_n(\mathbf{o})} \right)^{-1} \frac{\sum_{i=1}^n Y_i(\mathbf{o})}{\sqrt{2\phi_n(\mathbf{o})}}. \end{aligned} \tag{5.10}$$

Using our estimates on the Frobenius norm $\|\Delta_n(x_n)\|$,

$$\begin{aligned} \frac{1}{d} \left\| \frac{\Delta_n(x_n)}{2\phi_n(\mathbf{o})} \right\| &\leq (1 + 2\varepsilon\rho) \frac{\sum_{i=1}^n \sup_{x' \in \text{ball}(\mathbf{o}, \rho)} \|\Pi_{x', \mathbf{o}} \nabla Y_i(x') - \nabla Y_i(\mathbf{o})\|}{2\phi_n(\mathbf{o})} \\ &+ \varepsilon \frac{\sum_{i=1}^n \|Y_i(\mathbf{o})\|}{\phi_n(\mathbf{o})} + \varepsilon\rho \frac{\sum_{i=1}^n \|\nabla Y_i(\mathbf{o})\|}{\phi_n(\mathbf{o})}. \end{aligned}$$

We are given that $x_n \rightarrow \mathbf{o}$ in probability, so with probability tending to 1 we may apply condition 2 of the theorem to the first of these summands, together with the Markov inequality, and deduce that

$$(1 + 2\varepsilon\rho) \frac{\sum_{i=1}^n \sup_{x' \in \text{ball}(\mathbf{o}, \rho)} \|\Pi_{x', \mathbf{o}} \nabla Y_i(x') - \nabla Y_i(\mathbf{o})\|}{2\phi_n(\mathbf{o})} \rightarrow 0 \quad \text{in probability.}$$

Application of the Cauchy–Schwartz inequality to the second summand, together with the definition of the aggregate energy function, the fact that $\|Y_i(\mathbf{o})\| = \text{dist}(\mathbf{o}, X_i)$, and condition 1 of the theorem, shows that

$$\frac{\mathbb{E}[\sum_{i=1}^n \|Y_i(\mathbf{o})\|]}{\phi_n(\mathbf{o})} \leq \sqrt{\frac{\mathbb{E}[\sum_{i=1}^n \|Y_i(\mathbf{o})\|^2]}{\phi_n(\mathbf{o})}} \sqrt{\frac{n}{\phi_n(\mathbf{o})}} = \sqrt{\frac{2n}{\phi_n(\mathbf{o})}} \leq \frac{2}{\sqrt{C_1}}.$$

A similar argument, but using condition 3 of the theorem as well as condition 1, allows us to deduce that

$$\frac{\mathbb{E}[\sum_{i=1}^n \|\nabla Y_i(\mathbf{o})\|]}{\phi_n(\mathbf{o})} = \frac{\mathbb{E}[\sum_{i=1}^n \|H_i(\mathbf{o})\|]}{\phi_n(\mathbf{o})} \leq \sqrt{\frac{\mathbb{E}[\sum_{i=1}^n \|H_i(\mathbf{o})\|^2]}{\phi_n(\mathbf{o})}} \sqrt{\frac{n}{\phi_n(\mathbf{o})}} \leq \sqrt{\frac{C_2}{C_1}}.$$

Once again we may use the assumption that $x_n \rightarrow \mathbf{o}$ in probability; it follows from this and the Markov inequality that we may choose $\varepsilon = \varepsilon_n$ to decrease to zero in such a manner that

$$\varepsilon_n \frac{\sum_{i=1}^n \|Y_i(\mathbf{o})\|}{\phi_n(\mathbf{o})} + \varepsilon_n\rho \frac{\sum_{i=1}^n \|\nabla Y_i(\mathbf{o})\|}{\phi_n(\mathbf{o})} \rightarrow 0 \quad \text{in probability.}$$

Accordingly it follows that the matrix error term is negligible:

$$\frac{\Delta_n(x_n)}{2\phi_n(\mathbf{o})} \rightarrow 0 \quad \text{in probability.} \tag{5.11}$$

Now consider the behaviour of $\sum_{i=1}^n \frac{H_i(\mathbf{o})}{2\phi_n(\mathbf{o})}$. We can control the sum of the variances of the components of this matrix: by independence, and the fact that variance is always bounded above by second moment, we deduce that the sum of variances is bounded above by

$$\sum_{i=1}^n \frac{\mathbb{E}[\|H_i(\mathbf{o})\|^2]}{4\phi_n(\mathbf{o})^2}$$

which converges to zero by conditions 1 and 3 of the theorem. Accordingly

$$\sum_{i=1}^n \frac{H_i(\mathbf{o})}{2\phi_n(\mathbf{o})} - \tilde{H}_n \rightarrow 0 \quad \text{in probability.}$$

Condition 4 of the theorem, together with the negligibility of $\frac{\Delta_n(x_n)}{\phi_n(\mathbf{o})}$ established above in (5.11), implies that the probability of the following being invertible converges to 1:

$$\left(\sum_{i=1}^n \frac{H_i(\mathbf{o})}{2\phi_n(\mathbf{o})} - \frac{\Delta_n(x_n)}{2\phi_n(\mathbf{o})} \right).$$

Moreover we may deduce that

$$\left(\sum_{i=1}^n \frac{H_i(\mathbf{o})}{2\phi_n(\mathbf{o})} - \frac{\Delta_n(x_n)}{2\phi_n(\mathbf{o})} \right)^{-1} - \tilde{H}_n^{-1} \rightarrow 0 \quad \text{in probability.} \tag{5.12}$$

Finally we consider the asymptotic distributional behaviour of

$$\frac{\sum_{i=1}^n Y_i(\mathbf{o})}{\sqrt{2\phi_n(\mathbf{o})}}.$$

The Lindeberg condition 5 of the theorem translates directly into a condition of Lindeberg type on the Y_i : since $\|Y_i(\mathbf{o})\| = \text{dist}(\mathbf{o}, X_i)$, and since $\mathbb{E}[Y_i(\mathbf{o})] = 0$ as a consequence of \mathbf{o} being a local Fréchet mean of X_i , as $n \rightarrow \infty$ so

$$\frac{1}{\phi(\mathbf{o})} \sum_{i=1}^n \mathbb{E}[\|Y_i(\mathbf{o})\|^2; \|Y_i(\mathbf{o})\|^2 > \varepsilon\phi_n(\mathbf{o})] \rightarrow 0.$$

Now Theorem 3 shows that

$$\tilde{W}_1 \left(\frac{Y_1(\mathbf{o}) + \dots + Y_n(\mathbf{o})}{\sqrt{2\phi_n}}, Z_n \right) \rightarrow 0,$$

where Z_n has the multivariate d -dimensional normal distribution of zero mean and variance-covariance matrix V_n .

The proof of the theorem is now completed by using observation (5.12), since properties of the Wasserstein distance allow us to deduce

$$\tilde{W}_1 \left(\left(\sum_{i=1}^n \frac{H_i(\mathbf{o})}{2\phi_n(\mathbf{o})} - \frac{\Delta_n(x_n)}{2\phi_n(\mathbf{o})} \right)^{-1} \frac{Y_1(\mathbf{o}) + \dots + Y_n(\mathbf{o})}{\sqrt{2\phi_n}}, \tilde{H}_n^{-1} Z_n \right) \rightarrow 0$$

from the convergence in probability specified in (5.12), together with the upper bound supplied by condition 4 of the theorem. \square

We finish by looking at a few special cases. First, if we assume that the X_n are actually identically distributed, then $\phi_n(\mathbf{o}) = \frac{n}{2}\mathbb{E}[\text{dist}(\mathbf{o}, X_1)^2]$. Accordingly, if $0 < \mathbb{E}[\text{dist}(\mathbf{o}, X_1)^2] < \infty$ then the conditions 1 and 5 of Theorem 4 hold trivially. Moreover,

$$\tilde{H}_n = \tilde{H} = \mathbb{E}[H_1(\mathbf{o})]/\mathbb{E}[\text{dist}(\mathbf{o}, X_1)^2]$$

and

$$V_n = V = \mathbb{E}[\Phi_{\mathbf{o}, X_1}]/\mathbb{E}[\text{dist}(\mathbf{o}, X_1)^2],$$

where $\Phi_{x,y}$ is the self-adjoint linear operator on $T_x(\mathbb{M})$ defined by

$$\Phi_{x,y} : v \mapsto \langle \text{Exp}_x^{-1}(y), v \rangle \text{Exp}_x^{-1}(y). \tag{5.13}$$

Hence, the following is a direct consequence of Theorem 4.

Corollary 3. *Suppose that X_1, X_2, \dots is a sequence of independent and identically distributed random variables on \mathbb{M} with finite $\mathbb{E}[\text{dist}(x, X_1)^2]$. Suppose that \mathbf{o} is the local Fréchet mean of X_1 and that $\mathbb{P}[X_1 \in C_{\mathbf{o}}] = 0$. Let $x_n = \mathcal{E}(X_1, \dots, X_n)$ be a measurable choice of local empirical Fréchet means such that $x_n \rightarrow \mathbf{o}$ in probability. Assume that*

(i)

$$\lim_{\rho \rightarrow 0} \mathbb{E} \left[\sup_{x \in \text{ball}(\mathbf{o}, \rho)} \|\Pi_{x', \mathbf{o}} H_1(x') - H_1(\mathbf{o})\| \right] = 0;$$

(ii) $\mathbb{E}[\|H_1(\mathbf{o})\|^2] < \infty$;

(iii) $\mathbb{E}[H_1(\mathbf{o})]^{-1}$ exists.

Then we have the following weak convergence as $n \rightarrow \infty$:

$$\sqrt{n} \text{Exp}_{\mathbf{o}}^{-1}(x_n) \xrightarrow{d} \text{MVN}(0, \tilde{H}^{-1} \mathbb{E}[\Phi_{\mathbf{o}, X_1}] \tilde{H}^{-1}),$$

where the limit is the multivariate normal distribution with zero mean and variance-covariance matrix $\tilde{H}^{-1} \mathbb{E}[\Phi_{\mathbf{o}, X_1}] \tilde{H}^{-1}$.

If there exists a local coordinate chart $\psi(x) = (x_1(x), \dots, x_d(x))$ with a domain which contains the support of the distribution of X_1 , then let us write (ξ_1, \dots, ξ_d) and $(\zeta_1^n, \dots, \zeta_d^n)$ respectively for the coordinates of $\text{Exp}_{\mathbf{o}}^{-1}(X_1)$ and $\text{Exp}_{\mathbf{o}}^{-1}(x_n)$ with respect to the basis $(\partial_{x_1}, \dots, \partial_{x_d})$ in $T_{\mathbf{o}}(\mathbb{M})$. The following result is the version of Corollary 3 in terms of these coordinates.

Corollary 4. Write $\zeta^n = (\zeta_1^n, \dots, \zeta_d^n)^\top$. In the case of Corollary 3,

$$\sqrt{n}\zeta^n \xrightarrow{d} \text{MVN}(0, (\mathbb{E}[H^\psi])^{-1}GV_\psi G(\mathbb{E}[H^\psi])^{-1}) \quad \text{as } n \rightarrow \infty,$$

where $G = (\langle \partial_{x_j}, \partial_{x_k} \rangle)$, $V_\psi = (\mathbb{E}[\xi_j \xi_k])$ and H^ψ is the matrix of the linear operator $H_1(\mathbf{o})$ under the coordinate chart ψ with

$$H_{\ell k}^\psi = -\frac{\partial \xi_\ell}{\partial x_k} - \sum_{j=1}^d \Gamma_{kj}^\ell \xi_j$$

and with Γ_{ij}^k being the Christoffel symbols for the chosen coordinate chart.

If the coordinates ψ are normal coordinates centred at \mathbf{o} corresponding to an orthonormal basis (e_1, \dots, e_d) of $T_{\mathbf{o}}(\mathbb{M})$, then $G = I$, and (ξ_1, \dots, ξ_d) and $(\zeta_1^n, \dots, \zeta_d^n)$ become the normal coordinates, centred at \mathbf{o} , of X_1 and x_n , respectively. Moreover, under a normal coordinate system, all the Christoffel symbols disappear at the centre \mathbf{o} and $\xi_k = -\frac{1}{2}\nabla_{e_k} \text{dist}(x, X_1)^2|_{x=\mathbf{o}}$, where ∇_{e_k} acts on the first variable of dist^2 under normal coordinates, Corollary 4 recovers the result of [Bhattacharya and Patrangenaru \(2005\)](#) at the Fréchet mean \mathbf{o} .

Finally, if \mathbb{M} either has constant sectional curvature κ or is a Kähler manifold of constant holomorphic sectional curvature κ then the operator $H_i(x)$ defined by (5.4) can be expressed explicitly. In the former case

$$H_i(x) : v \mapsto \frac{1 - f_\kappa(\text{dist}(x, X_i))}{\text{dist}(x, X_i)^2} \Phi_{x, X_i}(v) + f_\kappa(\text{dist}(x, X_i))v$$

and, in the latter,

$$\begin{aligned} H_i(x) : v \mapsto & f_\kappa(\text{dist}(x, X_i))v + \frac{1 - f_{\kappa/4}(\text{dist}(x, X_i))}{\text{dist}(x, X_i)^2} \Phi_{x, X_i}(v) \\ & + \frac{f_\kappa(\text{dist}(x, X_i)) - f_{\kappa/4}(\text{dist}(x, X_i))}{\text{dist}(x, X_i)^2} \Phi_{x, X_i}^J(v), \end{aligned}$$

where

$$f_\kappa(s) = \begin{cases} \sqrt{|\kappa|}s \frac{C_\kappa(s)}{S_\kappa(s)}, & \kappa \neq 0, \\ 1, & \kappa = 0, \end{cases} \quad S_\kappa(s) = \begin{cases} \sin(\sqrt{\kappa}s), & \kappa > 0, \\ s, & \kappa = 0, \\ \sinh(\sqrt{-\kappa}s), & \kappa < 0, \end{cases}$$

$C_\kappa(s) = S'_\kappa(s)/\sqrt{|\kappa|}$, and where J is the tensor field of isometries J_x of the tangent spaces $T_x(\mathbb{M})$ such that $J_x^2 = -\text{id}$, $\Phi_{x,y}$ is defined by (5.13) and $\Phi_{x,y}^J$ is also defined by (5.13) but with Exp_x^{-1} there replaced by $J_x \circ \text{Exp}_x^{-1}$. Note that the consequent expression for the operator $\mathbb{E}[H_i(\mathbf{o})]$ was obtained in [Bhattacharya and Bhattacharya \(2008\)](#) when \mathbb{M} has constant curvature and an upper bound has also been given in the same paper for general \mathbb{M} in term of the bound of its curvature.

Acknowledgments

We gratefully acknowledge the helpful advice of A. Barbour, N. H. Bingham, C. M. Goldie, P. Hall and R. Bhattacharya.

References

- Afsari, B. (2011). Riemannian L^p center of mass: Existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* **139**, 655–655. [MR2736346](#)
- Barbour, A. D. and Gneden, A. V. (2009). Small counts in the infinite occupancy scheme. *Electron. J. Probab.* **14**, 365–384. [MR2480545](#)
- Bhattacharya, A. and Bhattacharya, R. (2008). Statistics on Riemannian manifolds: Asymptotic distribution and curvature. *Proc. Amer. Math. Soc.* **136**, 2959–2967. [MR2399064](#)
- Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds—I. *Ann. Statist.* **31**, 1–29. [MR1962498](#)
- Bhattacharya, R. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds—II. *Ann. Statist.* **33**, 1225–1259. [MR2195634](#)
- Bhattacharya, R. and Rao, R. R. (1976). *Normal Approximation and Asymptotic Expansions*. New York–London–Sydney: Wiley. [MR0436272](#)
- Billingsley, P. (1986). *Probability and Measure*. New York: Wiley. [MR0830424](#)
- Chatterjee, S. (2008). A new method of normal approximation. *Ann. Probab.* **36**, 1584–1610. [MR2435859](#)
- Chow, Y. S. and Teicher, H. (2003). *Probability Theory: Independence, Interchangeability, Martingales*. New York–Heidelberg–Berlin: Springer-Verlag.
- Corcuera, J.-M. and Kendall, W. S. (1999). Riemannian barycentres and geodesic convexity. *Math. Proc. Cambridge Philos. Soc.* **127**, 253–269. [MR1705458](#)
- Feller, W. (1966). *An Introduction to Probability Theory and Its Applications, Volume 2*. New York: Wiley. [MR0210154](#)
- Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. Henri Poincaré* **10**, 215–310. [MR0027464](#)
- Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.* **30**, 509–541. [MR0442975](#)
- Kendall, W. S. (1990). Probability, convexity, and harmonic maps with small image I: Uniqueness and fine existence. *Proc. Lond. Math. Soc. (3)* **61**, 371–406. [MR1063050](#)
- Kendall, W. S. (1991a). Convex geometry and nonconfluent Γ -martingales I: Tightness and strict convexity. In *Stochastic Analysis, Proceedings, LMS Durham Symposium, 11th–21st July 1990* (M. T. Barlow and N. H. Bingham, eds.) 163–178. Cambridge: Cambridge Univ. Press. [MR1166410](#)
- Kendall, W. S. (1991b). Convexity and the hemisphere. *J. Lond. Math. Soc. (2)* **43**, 567–576. [MR1113394](#)
- Kendall, W. S. (1992a). Convex geometry and nonconfluent Γ -martingales II: Well-posedness and Γ -martingale convergence. *Stochastics and Stochastic Reports* **38**, 135–147. [MR1274899](#)
- Kendall, W. S. (1992b). The Propeller: A counterexample to a conjectured criterion for the existence of certain convex functions. *J. Lond. Math. Soc. (2)* **46**, 364–374. [MR1182490](#)
- Kendall, W. S. and Le, H. (2010). Statistical shape theory. In *New Perspectives in Stochastic Geometry* (W. S. Kendall and I. S. Molchanov, eds.) **10** 348–373. Oxford: Clarendon Press, Oxford Univ. Press. [MR2654683](#)
- Le, H. (2001). Locating Fréchet means with application to shape spaces. *Adv. Appl. Probab.* **33**, 324–338. [MR1842295](#)

- Le, H. (2004). Estimation of Riemannian barycentres. *LMS J. Comput. Math* **7**, 193–200. [MR2085875](#)
- Picard, J. (1994). Barycentres et martingales sur une variété. *Ann. Inst. H. Poincaré Probab. Stat.* **30**, 647–702. [MR1302764](#)
- Röllin, A. (2011). Stein's method in high dimensions with applications. Available at [arXiv:1101.4454](#).
- Villani, C. (2003). *Topics in optimal transportation. Graduate Studies in Mathematics* **58**. Providence, RI: Amer. Math. Soc. [MR1964483](#)
- Ziezold, H. (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. *Conf. Inf. Theory, Statist. Decision Functions, Random Processes* **A**, 591–602. [MR0501230](#)
- Ziezold, H. (1989). On expected figures in the plane. In *Geobild '89 (Georgenthal, 1989)*. *Math. Res.* **51**, 105–110. Berlin: Akademie-Verlag. [MR1003326](#)
- Ziezold, H. (1994). Mean figures and mean shapes applied to biological figure and shape distributions in the plane. *Biom. J.* **36**, 491–510. [MR1292314](#)

Statistics Department
University of Warwick
Coventry CV4 7AL
UK
E-mail: w.s.kendall@warwick.ac.uk
URL: <http://go.warwick.ac.uk/wsk>

School of Mathematical Sciences
University of Nottingham
University Park
Nottingham NG7 2RD
UK
E-mail: Huilong.Le@nottingham.ac.uk