GEOMETRY OF SAMPLE SPACES

PHILIPP HARMS, PETER W. MICHOR, XAVIER PENNEC, AND STEFAN SOMMER

Abstract. In statistics, independent, identically distributed random samples do not carry a natural ordering, and their statistics are typically invariant with respect to permutations of their order. Thus, an n-sample in a space M can be considered as an element of the quotient space of M^n modulo the permutation group. The present paper takes this definition of sample space and the related concept of orbit types as a starting point for developing a geometric perspective on statistics. We aim at deriving a general mathematical setting for studying the behavior of empirical and population means in spaces ranging from smooth Riemannian manifolds to general stratified spaces.

We fully describe the orbifold and path-metric structure of the sample space when M is a manifold or path-metric space, respectively. These results are non-trivial even when M is Euclidean. We show that the infinite sample space exists in a Gromov–Hausdorff type sense and coincides with the Wasserstein space of probability distributions on M. We exhibit Fréchet means and k-means as metric projections onto 1-skeleta or k-skeleta in Wasserstein space, and we define a new and more general notion of polymeans. This geometric characterization via metric projections applies equally to sample and population means, and we use it to establish asymptotic properties of polymeans such as consistency and asymptotic normality.

1. Introduction

Following the pioneering developments of directional statistics [33] and shape statistics [35, 36, 19], there is a growing need in many application domains for the statistical analysis of populations of objects in complicated non-Euclidean spaces. One can cite for instance tree-spaces in biology [8], Riemannian manifolds and Lie groups, including diffeomorphism groups, in medical image analysis and computer vision [45, 47, 49], or more generally stratified spaces [42]. With the choice of a relevant distance, a natural generalization of the central values of a population of objects in these spaces is the Fréchet p-mean, that is the set of minima of the mean distance to the power p [25]. While the choice of $p = 2$ is often used because it corresponds to the usual arithmetic, lower values of p up to $p = 1$ defining the median ("valeur equiprobable" in Fréchet's words) are also often useful for robust statistics.

This paper develops a general mathematical setting to study the behavior of empirical and population Fréchet p-means in spaces ranging from smooth Riemannian manifolds to general stratified spaces. We start from the key observation that independent, identically distributed (i.i.d.) random samples do not carry a natural ordering, and their statistics are typically invariant with respect to permutations of their order. Thus, an n-sample in a space M can naturally be considered as an

element of the quotient space M^n/S_n of n-tuples modulo the permutation group S_n. This space shall accordingly be called *sample space*. The paper takes this definition as a starting point for developing a geometric perspective on statistics, guided by the notion of orbit type. This way, we provide a theoretical basis for further investigations on unordered samples in non-Euclidean spaces.

1.1. **Background.** For non-positively curved spaces in the sense of Alexandrov, the 2-mean is always unique when it exists [48]. For positively curved Riemannian manifolds, an important effort has been spent in determining the convexity conditions on the distribution that ensure uniqueness [34, 11, 1]. However, many very useful distributions such as wrapped or truncated Gaussian distributions on the tangent spaces do not fulfill these conditions even if they have a unique Fréchet mean.

Asymptotic properties of the sample mean for distributions on Riemannian manifolds with a unique population Fréchet mean were studied by Bhattacharya and Patrangenaru [5, 6, 7]. In particular, they showed the consistency of the sample Fréchet mean $\bar{x}_n$ of n i.i.d. samples of a random variable x for large sample sizes (law of large numbers), building on a strong consistency result of [55]. Under the Karcher and Kendall convexity conditions for the uniqueness of the population mean $\bar{x}$, the Bhattacharya-Patrangenaru central limit theorem (CLT) further states that the random variables $u_n = \sqrt{n}\log_{\bar{x}}(\bar{x}_n)$ converge in distribution to the Gaussian $\mathcal{N}(0, \bar{H}^{-1}\mathrm{Cov}(x)\bar{H}^{-1})$ in the tangent space at $\bar{x}$ whenever the expected Hessian $\bar{H}$ of half the Riemannian squared distance at the population mean $\bar{x}$ is invertible. This type of CLT based on the delta method was further generalised in [37] to non-i.i.d. variables and in [29] to summary statistics other than the mean, such as principal geodesics.

In non-manifold stratified spaces of negative curvature, an intriguing phenomenon was discovered 10 years ago: the Fréchet mean may be sticky on singular strata [28]. A regular random variable (that is a not fully concentrated on singular strata) whose Fréchet mean is located on a singular stratum is said to have a sticky mean if a sufficiently small variation of that random variable continues to have its Fréchet mean on the singular stratum. In other words, the singular strata are attractive. It is surprising that a CLT can still be derived under these conditions [28]. This suggests that some regularity can be used for deriving CLTs in more general settings. Stickiness does not seem to happen in positive curvature. For instance, Kendall shape spaces in three or higher dimensions are stratified, but the Fréchet mean of regular random variables was shown to belong to the top regular stratum (manifold-stability) [30]. In other words, singular strata of that kind are repulsive.

More recently, an apparently opposite unusual behavior of the CLT was discovered with smeary means, where the empirical Fréchet means converge at an asymptotic rate lower than $\sqrt{n}$; see [21] e.g. Other results show that intermediate repulsive or attractive behaviours can happen on Riemannian manifolds, controlled either by the curvature [46, 20] or by the topology [31]. Thus, classical tests based on asymptotic results for Euclidean spaces might be biased, which is a critical problem for many applications. This highlights the need for a new mathematical framework to study the distribution of the empirical Fréchet mean, either in the small sample regime or asymptotically.

While considering n-samples disregarding ordering is not new, the literature is sparse in linking geometric properties of the quotient space to sample statistics. In

the Euclidean case, de Finetti's theorem [23, 16] and the theory of Hewitt and Savage [27] on exchangeability and presentability characterized distributions invariant to finite permutations leading to central limit theorems based on exchangeability instead of independence [13, 9, 38]. We here develop a similar theory using additional geometric structures.

1.2. **Overview and results.** The convenient level of generality that we adopt is that of path-metric spaces [26, 10], see A.1, where the distance is given by the infimum of the length of curves joining the two points; for complete path-metric spaces the infimum is a minimum, see A.2.

We first describe in Section 2 the orbifold (resp. path-metric) structure of the sample space M^n/S_n when M is a manifold (resp. a path-metric space). These results are non-trivial even when M is Euclidean but well known in the realm of reflection groups and Weyl chambers. The sample space M^n/S_n can be stratified by the number of pairwise distinct points. The regular part $(M^n/S_n)_{\mathrm{reg}}$ contains the unordered configurations where the n points are distinct. The lower dimensional strata are called the q-skeleta, see 2.2, and comprise unordered configurations with exactly $q < n$ distinct points. A finer stratification classifying orbit types is based on the partition $(\mathbf{k}) := (k_1 \geq \cdots \geq k_q)$ of n describing the number of identical points; see 2.5 and 2.6. Sub-partitioning (distinguishing some of the points that were previously identified) gives a half-ordering on partitions which are thus organized in a geometric lattice structure. The orbit-type stratum $(M^n/S_n)_{(n)}$ with the smallest partition (n) is the *diagonal* $\{x : x_1 = \cdots = x_n\} \simeq M$ where all points coincide. This is the 1-skeleton, which can be identified with the base manifold M. At the other end of the lattice, the regular orbit stratum $(M^n/S_n)_{\mathrm{reg}} = (M^n/S_n)_{(1 \geq 1 \geq \cdots \geq 1)}$ is the open, dense, connected, and locally connected subset of all unordered configurations with n distinct points. The closure of $(M^n/S_n)_{(\mathbf{k})}$ in M^n/S_n is the disjoint union of all $(M^n/S_n)_{(\mathbf{k}')}$ with $(\mathbf{k}') \leq (\mathbf{k})$; see 2.10. The q-skeleton of M^n/S_n is the the union of all orbit strata $(M^n/S_n)_{(\mathbf{k})}$ corresponding to all partitions $(\mathbf{k}) = (k_1 \geq \cdots \geq k_p)$ with length $p \leq q \leq n$. The projection to q-skeleta and orbit strata will be used in Section 5 to characterize the Fréchet p-mean and to define a generalization called polymeans.

Section 3 investigates the metric properties of the sample spaces when we assume that M is a complete path-metric space. The L_p metric $d_p(x, y) = \left(\frac{1}{n}\sum_{i=1}^n d(x_i, y_i)^p\right)^{1/p}$ with $p \in [1, \infty)$ on M^n induces a canonical quotient metric on the sample space $(M^n/S_n, \bar{d}_p)$, which is then a complete path-metric space; see 3.2. Moreover, orbit-type strata have convex closures, and a minimizing geodesic in the sample space $(M^n/S_n, \bar{d}_p)$ is the projection of a minimizing geodesic in the configuration space (M^n, d_p). When M is Riemannian and $p = 2$, one can show that geodesics are more regular at interior points than at their end-points, 3.7. However, this assertion is generally wrong for non-Riemannian complete path-metric spaces, like for instance for the 3-spider, 3.8. This lack of regularity could be linked to stickiness.

In order to investigate sub-samples (bootstrap) and infinite samples together in the same space, we show in 4.7 that the sample space $(M^n/S_n, \bar{d}_p)$ is isometric to the space of mixtures of n-atomic measures (the empirical law of the samples) endowed with the p-Wasserstein metric. Moreover, the infinite sample space $\lim_{n \to \infty} M^n/S_n$ exists in a weakened Gromov–Hausdorff type sense and coincides with the p-Wasserstein space $(\mathcal{P}^p(M), \bar{d}_p)$ of p-integrable probability distributions on M; see 4.8. The extension of skeleta and orbit-type strata to infinite sample

spaces can then be done easily: the q-*skeleton* in the infinite-sample space $\mathcal{P}^p(M)$ is the subset $\mathcal{P}(M)_q$ of all probability distributions with at most q support points; see 4.11. Similarly, for any partition $(\mathbf{k}) \coloneqq (w_1 \geq \cdots \geq w_q)$ consisting of non-negative weights w_i summing up to 1, the $(\mathbf{k})$-*stratum* in the infinite-sample space $\mathcal{P}^p(M)$ is the subset of mixtures $P = \sum_{i=1}^q w_i \delta_{x_i} \in \mathcal{P}(M)_q$ with q distinct points x_i. It is interesting to note that such a mixture of q Diracs is realized in a finite sample space for some n if the weights are all rational, but irrational weights can only be achieved in the infinite-sample limit.

With this setting, we are in position to exhibit in Section 5 empirical and population Fréchet means as metric projections onto the 1-skeletum in sample space or Wasserstein space, and we define a new and more general notion of empirical and population polymeans by the projection on the q-skeleta $(M^n/S_n)_q$ or on the $(\mathbf{k})$-strata $(M^n/S_n)_{(\mathbf{k})}$. These polymeans can be interpreted as the clusters of the well known k-means clustering algorithm: the k distinct points are the cluster centroids (we also call them the unweighted polymeans) and the weights w_i are the relative masses of the clusters. As everything is defined for p-integrable distributions ($p \geq 1$), our definitions are actually valid for general Fréchet p-means and p-power k-means. Since q-skeleta and $(\mathbf{k})$-strata are closed in all sample spaces, as well as in the p-Wasserstein space, the existence of empirical and population polymeans is ensured. The uniqueness is a much harder problem. In the Riemannian case with $p = 2$, recent results on the regularity of the singular set of the distance to a sufficiently regular set show that empirical polymeans of i.i.d. samples with an absolutely continuous law are almost surely unique. This partly extends the previous result of [4] on the uniqueness of the empirical Fréchet p-mean.

We turn in Section 6 to probability distributions on sample spaces. It turns out that the correct space of infinite samples is not the quotient space $M^{\mathbb{N}}/S_{(\mathbb{N})}$ but the space $\mathcal{P}(M)$ of probability distributions on M. Indeed, using this definition one obtains as in the theory of Hewitt and Savage [27] that probability distributions on infinite sample spaces correspond exactly to symmetric probability distributions on configuration spaces, which in turn correspond exactly to mixtures of product distributions. This definition is also in line with the infinite-sample limit 4.8. The analogous statement for random variables instead of probability distributions is that random samples correspond exactly (possibly after passing to an extended probability space) to conditionally i.i.d. random configurations; see 6.6.

This setting allows us to establish in Section 7 asymptotic properties of polymeans. We first show that the empirical q-means are strongly consistent estimators for the population q-means, in the sense that any accumulation point of the sets of empirical q-means is a population q-mean. Thus, when the population q-mean is unique, any measurable selection of empirical q-means converges in probability to the population q-mean, and we may inquire about the rate of convergence. We derive in 7.4 an upper bound on the convergence rate of empirical q-means to the population q-mean. The bound depends first on the convergence rate in Wasserstein space of empirical distributions—a well studied subject—and second on the subspace geometry of the q-skeleton within Wasserstein space—a purely geometric question. It remains an open problem if the bound is sharp and if q-means are asymptotically normal after a suitable normalization. However, when M is a Riemannian manifold, we establish in 7.6 the asymptotic normality of unweighted q-means for any $p \geq 1$ under mild conditions (null measure of the union of the cut

loci of the centroids and of their "medial axis" and non-degenerate expected Hessian of the power p distance to the closest centroid). We further refine this central limit theorem in 7.7 from i.i.d. to exchangeable sequences under some additional conditional independence assumptions.

In the appendix we collect some tools from path-metric geometry.

1.3. **Open problems and future work.** Our framework opens the door to many further investigations by linking two traditionally distinct strands of literature, namely, statistics on manifolds and orbifold or path-metric geometry. Tools from these fields can be fruitfully combined. The setup is fully general and applies to curved spaces and more general stratified spaces, as needed in the previously cited applications. It also encompasses Fréchet p-means and not only the classical 2-mean, which opens the way to many useful asymptotic results for robust statistics.

Our results also suggest that the non-standard convergence rates in the CLT are not only due to the geometry of M but also the subspace geometry of the k-skeleta within the sample spaces. For instance, considering the Fréchet mean as a projection on the 1-skeleton casts a new geometric light on the uniqueness problem: in a Riemannian manifold, it is unique whenever there is no mass on the singular set of the distance function to the 1-skeleton. Thus, one can conjecture that the geometry of the "medial axis" of the q-skeleton in p-Wasserstein space controls the uniqueness of the polymeans and that advances on the sub-space geometry of this set within Wasserstein space would extend this uniqueness theorem to more general settings.

Likewise, the rate of convergence of the empirical 2-mean towards the population 2-mean is controlled by the eigenvalues of the expected Hessian of the squared distance (Corollary 7.7). The convergence rate towards the limiting distribution in the direction of an eigenvector falls below $\sqrt{n}$ whenever the corresponding eigenvalue vanishes. Conversely, stickiness could be induced by eigenvalues going to infinity. This last behavior cannot happen in smooth Riemannian manifolds, but it can be approached by concentrating the curvature at singular points. This could be a way to study stickiness on smoothable manifolds. For i.i.d. samples with distribution P, we conjecture that these condition could be linked to the convexity or concavity of the geodesic distance in Wasserstein space from P to the polymean in the k-skeleton, and thus that it can be controlled by some kind of Ma–Trudinger–Wang (MTW) condition [22].

2. Orbit type stratification of sample spaces

Let M be a topological space. For any natural number $n \in \mathbb{N}_{>0}$, the permutation group S_n of n symbols acts on the n-fold product M^n by permutation of the components. In symbols, we shall write $x_\sigma := x \circ \sigma$ for the action of $\sigma \in S_n$ on $x \in M^n$.

Definition 2.1 (Configurations and samples). An *n-point configuration* or *ordered n-sample* is an element of M^n, and this space is called *(ordered) configuration space*. An *n-sample* is an element of the quotient space M^n/S_n, and this space is called *sample space* or *unordered configuration space*. The *projection* is denoted by $\pi \colon M^n \to M^n/S_n$.

Note that this definition of configuration spaces differs from the one commonly used in topology, where the points are required to be pairwise distinct. The set of

pairwise distinct points is an open subset of M^n, and its fundamental group in the case $M = \mathbb{R}^2$ is the *braid group*. In contrast, we also consider the case where only $q < n$ points are mutually distinct:

Definition 2.2 (Skeleta). A configuration $(x_1, \ldots, x_n)$ is said to belong to the *q-skeleton* if it consists of at most $q \in \mathbb{N}$ distinct points x_i. As the number of distinct points is S_n-invariant, there is a corresponding notion of q-skeleta of samples.

The name skeleton is taken from the theory of simplicial complexes and cell complexes. The filtration of sample space into skeleta is rather coarse, and finer stratifications are needed to fully describe the local geometry of sample space. This is done next.

Definition 2.3 (Orbifolds [52]). A Hausdorff topological space $\mathcal{O}$ is an orbifold, if the following data are given:

- An open cover (U_i) of $\mathcal{O}$ which is closed under forming finite intersections.
- For each i there is an open subset $V_i \subset \mathbb{R}^N$ which is invariant under a faithful linear action of a finite group G_i on $\mathbb{R}^N$ and a G_i-invariant quotient map $\pi_i \colon V_i \to V_i/G_i \cong U_i$.
- If $U_i \subset U_j$ then there is an injective group homomorphism $\varphi_{ij} \colon G_i \to G_j$ and a gluing map ψ_{ij} from V_i to an open subset of V_j which is G_i-equivariant in the sense that $\psi_{i,j}(g.x) = \varphi_{ij}(g).\psi_{ij}(x)$ for all $x \in V_i$ and such that $\pi_j \circ \psi_{ij} = \pi_i$.

In this situation (V_i, π_i, G_i) is then called an orbifold chart.

Lemma 2.4 (Orbifold structure of sample space)**.** *If M is a manifold, then the sample space M^n/S_n is an orbifold.*

Proof. For any $x \in M^n$, choose a chart $(U_i, u_i \colon U_i \to \mathbb{R}^m)$ such that whenever $x_i = x_j$ we have $(U_i, u_i) = (U_j, u_j)$. Then $u_1(U_1) \times \cdots \times u_n(U_n) \subseteq (\mathbb{R}^m)^n$ is invariant under the isotropy group $(S_n)_x$ and $\pi \circ (u_1^{-1} \times \cdots \times u_n^{-1}) \colon u_1(U_1) \times \cdots \times u_n(U_n) \to \pi(U_1 \times \cdots \times U_n) \subset M^n/S_n$ is the required orbifold chart. $\qquad\square$

The proof of 2.4 shows more generally that the quotient space of a smooth manifold with respect to a properly discontinuous action of a group is an orbifold; in this case it is sometimes called a *developable* or (by Thurston) a *good* orbifold. To understand the orbifold structure of sample space, one has to describe the different *orbit types*.

Definition 2.5 (Orbit types). The orbit type of an ordered sample $x \in M^n$ is defined as the conjugacy class of its isotropy group $(S_n)_x := \{\sigma \in S_n : x_\sigma = x\}$. As the orbit type is S_n-invariant, there is a corresponding notion of orbit types of samples in M^n/S_n.

The following theorem classifies the orbit types of sample space. It turns out that there are many different orbit types, one for each partition of the integer n. This highlights the complicated geometry of sample space.

Theorem 2.6 (Classification of orbit types)**.** *The orbit types in the configuration space M^n are exactly given by the integer partitions of n of the form*

$$n = k_1 + k_2 + \cdots + k_q, \quad k_1 \geq k_2 \geq \cdots \geq k_q \geq 1.$$

We write $(\mathbf{k}) := (k_1 \geq \cdots \geq k_q)$ for such a partition.

Proof. This follows from the fact that a point $x \in M^n$ is fixed by a permutation

$$\sigma = (\sigma_1 \sigma_2 \ldots \sigma_{k_1})(\sigma_{k_1+1} \ldots \sigma_{k_1+k_2}) \ldots (\sigma_{k_1+\ldots k_{q-1}+1} \ldots \sigma_{k_1+\cdots+k_p}) \in S_n$$

if and only if

$$x_{\sigma_1} = x_{\sigma_2} = \cdots = x_{\sigma_{k_1}}, x_{\sigma_{k_1+1}} = \cdots = x_{\sigma_{k_1+k_2}}, \quad \cdots$$
$$\cdots \quad x_{\sigma_{k_1+\ldots k_{q-1}+1}} = \cdots = x_{\sigma_{k_1+\cdots+k_q}},$$

and all other x_i being distinct. Here $(k_1 \geq k_2 \geq \cdots \geq k_p)$ with $k_1 + \cdots + k_p \leq n$ is the *cycle type* of the permutation σ. For our purpose we enlarge the cycle type to $(k_1 \geq \cdots \geq k_p \geq \cdots \geq k_q) := (k_1 \geq \cdots \geq k_p \geq 1 \cdots \geq 1)$ until it becomes a *partition* of n, denoted by

$$(\mathbf{k}) = (k_1 \geq \cdots \geq k_q) \text{ with } k_1 + \cdots + k_q = n.$$

The conjugate by $\tau \in S_n$ of the k_1-cycle $\sigma' = (\sigma_1 \sigma_2 \ldots \sigma_{k_1})$ is the k_1-cycle $\tau\sigma'\tau^{-1} = (\tau(\sigma_1) \tau(\sigma_2) \ldots \tau(\sigma_{k_1}))$, and similarly for the other cycles in σ. Thus, the isotropy group of any x as above is conjugated to the subgroup $S_{k_1} \times S_{k_2} \times \ldots \times S_{k_p}$. Its conjugacy class is described by the cycle type $(k_1, \ldots, k_p)$ with $k_1, \ldots, k_q \in \mathbb{N}_{>0}$, and equivalently by its enlargement to a partition of n. $\square$

The configuration space M^n and the sample space M^n/S_n are *stratified* by orbit type.

Definition 2.7 (Orbit-type strata). Let (H) denote the conjugacy class of any subgroup H of S_n corresponding to a partition $(\mathbf{k})$. We write $(M^n)_{(H)}$ and $(M^n)_{(\mathbf{k})}$ for the *stratum* of all points in M^n of orbit type (H) and $(\mathbf{k})$, respectively. Similarly, we write $(M^n/S_n)_{(H)}$ and $(M^n/S_n)_{(\mathbf{k})}$ for the corresponding stratum in M^n/S_n.

Lemma 2.8 (Orbit-type strata). *The stratum $(M^n)_{(\mathbf{k})}$ of orbit type*

$$(\mathbf{k}) := (k_1 \geq \cdots \geq k_q)$$

consists of all $x = (x_1, \ldots, x_n)$ such that k_1 of the the x_i are equal to $y_1 \in M$, k_2 of the remaining x_i are equal to $y_2 \neq y_1$ in M, and so on, until the remaining k_q of the x_i are equal to $y_q \in M$, and all y_i are distinct. Thus, $(M^n)_{(\mathbf{k})}$ is the disjoint union of its connected components, which are all homeomorphic to the open subset of pairwise distinct points in M^q.

Proof. This follows from the description of orbit types in the proof of 2.6. $\square$

Definition 2.9 (Half-ordering of orbit types). For two conjugacy classes (H) and (H') of subgroups H and H' in S_n, we write $(H) \leq (H')$ if H is conjugated in S_n to a subgroup of H'. Correspondingly, for two partitions $(\mathbf{k}) = (k_1 \geq \cdots \geq k_q)$ and $(\mathbf{k}') = (k'_1 \geq \cdots \geq k'_{q'})$, we write $(\mathbf{k}) \geq (\mathbf{k}')$ if $(\mathbf{k})$ sub-partitions $(\mathbf{k}')$.

Note that the half-order between partitions is the inverse of the half-order between the corresponding conjugacy classes. The *diagonal* $\{x : x_1 = \cdots = x_n\}$ is the stratum with the largest conjugacy class (S_n) and the smallest partition (n). The projection onto the corresponding stratum in M^n/S_n is a homeomorphism. The *regular stratum* is the open and dense subset of all configurations x with mutually distinct components x_i. It has as orbit type the smallest conjugacy class $(\{\mathrm{Id}\})$ and the largest partition $(1 \geq \cdots \geq 1)$. The regular orbit stratum $M^n_{(\{\mathrm{Id}\})} = M^n_{(1 \geq 1 \geq \cdots \geq 1)}$ in M^n/S_n is open, dense, connected, and locally connected; it will also be denoted by M^n_{reg}. Likewise for $(M^n/S_n)_{\mathrm{reg}} = (M^n/S_n)_{(1 \geq 1 \geq \cdots \geq 1)}$.

Note that for $q \leq n$, the q-skeleton of M^n/S_n is the the union of all orbit strata $(M^n/S_n)_{(\mathbf{k})}$ corresponding to all partitions $(\mathbf{k}) = (k_1 \geq \cdots \geq k_p)$ with length $p \leq q$.

Lemma 2.10 (Closure of orbit-type strata). *The stratum $(M^n)_{(\mathbf{k}')}$ is contained in the closure of the stratum $(M^n)_{(\mathbf{k})}$ if and only if $(\mathbf{k}') \leq (\mathbf{k})$ if and only if $(S_{k_1} \times \ldots \times S_{k_q}) \leq (S_{k_1'} \times \ldots \times S_{k_{q'}'})$. Moreover, the closure of $(M^n)_{(\mathbf{k})}$ in M^n is the disjoint union of all $(M^n)_{(\mathbf{k}')}$ with $(\mathbf{k}') \leq (\mathbf{k})$. A similar statement holds with M^n replaced by M^n/S_n.*

Proof. This follows from the description of the orbit-type strata given in 2.8, since at the boundary some distinct x_i might become equal. $\qquad\square$

Lemma 2.11 (Bundle structure of orbit-type strata). *Let $(\mathbf{k}) \coloneqq (k_1 \geq \cdots \geq k_q)$ be a partition describing the orbit type (H) with $H \coloneqq S_{k_1} \times \ldots \times S_{k_q}$. Then the projection $(M^n)_{(\mathbf{k})} \to S_n/N_{S_n}(H)$ defines a topological fiber bundle, where $N_{S_n}(H)$ is the normalizer of H in S_n, and where for any $\sigma \in S_n$, the fiber over $\sigma.N_{S_n}(H)$ is the fixed-point set $(M^n)^{\sigma^{-1}H\sigma} \cap (M^n)_{(\mathbf{k})}$.*

Proof. The proof in [43, 29.22], although given for smooth manifolds, is purely topological and applies here without change. $\qquad\square$

3. Path metrics on sample spaces

The category of *path-metric spaces* is ideally suited for the description of sample spaces because it is well-behaved under quotients. We refer to the appendix for the definition of path metrics and some of their properties, and to the book of Gromov [26] and also [10] or [3] for further details. Throughout this section, d is a complete path metric on the separable topological space M, $n \in \mathbb{N}_{>0}$, and $p \in [1, \infty)$.

There are many choices of metrics on the configuration space M^n which are consistent with the product topology. The following lemma describes some of them.

Lemma 3.1 (Path metrics on configuration spaces). *The following is a complete path metric on the configuration space M^n:*

$$d_p(x, y) \coloneqq \left(\frac{1}{n} \sum_{i=1}^{n} d(x_i, y_i)^p\right)^{1/p}, \qquad x, y \in M^n.$$

The identity on M^n is Lipschitz continuous between any of the metrics d_p, $p \in [1, \infty)$.

Note that $d_p(x, y) = \|d(x, y)\|_{L^p}$, where $\|\cdot\|_{L^p}$ denotes the L^p norm of functions on the space $\{1, \ldots, n\}$ with the uniform probability distribution. The choice of normalizing constant $\frac{1}{n}$ is motivated by this probabilistic interpretation, as well as the large-sample limits in 4.3 and 4.8.

Proof. Completeness of (M^n, d_p) follows from completeness of (M, d). As M is a path-metric space, there exists by A.3 for any $r > 1/2$ and any $a, b \in M$ a point $c = c(a, b) \in M$ such that

$$\max\{d(a, c), d(c, b)\} \leq r d(a, b).$$

Then obtains for the configuration $z \coloneqq c(x, y)$ by applying the L^p norm that

$$\max\{d_p(x, z), d_p(z, y)\} \leq r d_p(x, y).$$

This implies by A.3 that d_p is a path metric on M^n. The identity $M^n \to M^n$ is Lipschitz continuous under any of the metrics d_p because

$$n^{-1/p} \max_i d(x_i, y_i) \leq d_p(x, y) \leq \max_i d(x_i, y_i), \qquad x, y \in M^n. \qquad \square$$

The complete path metric d_p on the ordered sample space M^n induces a canonical *quotient metric* on the sample space M^n/S_n. As the permutation group S_n acts isometrically on (M^n, d_p), this quotient metric is again complete and admits a particularly simple description, as shown next.

Lemma 3.2 (Quotient metrics on sample spaces)**.** *The following quotient metric is a complete path metric on the sample space M^n/S_n:*

$$\bar{d}_p(\bar{x}, \bar{y}) = \min_{\pi(x)=\bar{x}, \pi(y)=\bar{y}} d_p(x, y) = \min_{\sigma \in S_n} d_p(x, y_\sigma),$$

where $\bar{x}, \bar{y} \in M^n/S_n$ and $x, y \in M^n$ with $\pi(x) = \bar{x}$, $\pi(y) = \bar{y}$.

Proof. The fibers of the projection are the orbits of the permutation group S_n, which acts isometrically on (M^n, d_p). Therefore, the metric $\bar{d}_p$ is a path metric [10, Lemma 3.3.6]. Moreover, this metric is complete: Given a Cauchy sequence in M^n/S_n, take a subsequence such that the distances between subsequent points are summable. Lift the sequence to M^n such that distances between subsequent points are preserved. Then the lift is a Cauchy sequence, which converges thanks to the completeness of M^n. $\qquad \square$

Recall that a subset of a metric space is called *convex* if the restriction of the metric to this subset is a finite complete path metric [10, Definition 3.6.5]. If the surrounding space carries a complete path metric, then this is equivalent to the subset being *totally geodesic*, i.e., any two points in the subset can be connected by a minimizing geodesic in the subset.

Lemma 3.3 (Convexity of orbit-type strata)**.** *Connected components of orbit-type strata in the configuration space (M^n, d_p) have convex closures. Moreover, orbit-type strata in the sample space $(M^n/S_n, \bar{d}_p)$ have convex closures.*

Proof. If $(\mathbf{k}) := (k_1 \geq \cdots \geq k_q)$ is a partition of n, then by 2.8 each connected component K of $(M^n)_{(\mathbf{k})}$ is homeomorphic to the open subset of all pairwise distinct points in M^q. This homeomorphism is even an isometry (up to a normalizing constant) under the d_p metrics on M^n and M^q, respectively. Thus, the closure $\bar{K}$ is homeomorphic to M^q. As (M^q, d_p) is a complete path-metric space by 3.1, it follows that $\bar{K}$ is a convex subset of (M^n, d_p). The projection $\pi \colon M^n \to M^n/S_n$ restricts to an isometry $\pi \colon K \to (M^n/S_n)_{(\mathbf{k})}$. It follows that $\bar{d}_p$ restricts to a complete path metric on the closure of the stratum $(M^n/S_n)_{(\mathbf{k})}$. Therefore, by definition, the closure of the stratum $(M^n/S_n)_{(\mathbf{k})}$ is a convex subset of $(M^n/S_n, \bar{d}_p)$. $\qquad \square$

Example 3.4 (Lack of strict convexity)**.** *The closure of a connected component of an orbit stratum in M^n need not be strictly convex in the sense that each minimal geodesic connecting two points in this stratum lies also in the stratum.*

Proof. Let c_1 and c_2 be two distinct meridian geodesics in the 2-sphere $M = S^2$, which connect the north pole N to the south pole S. Then $c = (c_1, c_2)$ is a minimizing geodesic between the points (N, N) and (S, S) in M^2. These points belong to the closed and connected orbit stratum $(M^2)_{(2)}$, but the geodesic c does not lie in $(M^2)_{(2)}$. $\qquad \square$

Theorem 3.5 (Geodesics between configurations). *A continuous curve $c\colon [0,1] \to M^n$ is a constant-speed minimizing geodesic in (M^n, d_p) with $p \in (1, \infty)$ if and only if its component curves $c_1, \ldots, c_n \colon [0,1] \to M$ are constant-speed minimal geodesics in (M, d). For $p = 1$ a similar statement holds without the requirement of constant speed.*

Proof. For $p > 1$, we associate Lagrangian energy–action pair (E, A) and (E_p, A_p) to (M, d) and (M^n, d_p), respectively, as in A.5:

$$E^{s,t}(x_i, y_i) := \frac{d(x_i, y_i)^p}{|s-t|^{p-1}}, \quad A^{s,t}(c_i) := \sup_{\substack{n \in \mathbb{N} \\ s=u_0 \le \cdots \le u_n = t}} \sum_{m=0}^{n-1} \frac{d(c_i(u_m), c_i(u_{m+1}))^p}{|u_m - u_{m+1}|^{p-1}},$$

$$E_p^{s,t}(x, y) := \frac{d_p(x, y)^p}{|s-t|^{p-1}}, \quad A_p^{s,t}(c) := \sup_{\substack{n \in \mathbb{N} \\ s=u_0 \le \cdots \le u_n = t}} \sum_{m=0}^{n-1} \frac{d_p(c(u_m), c(u_{m+1}))^p}{|u_m - u_{m+1}|^{p-1}},$$

for any $i \in \{1, \ldots, n\}$, $0 \le s \le t \le 1$, $x, y \in M^n$, and continuous curve $c\colon [0,1] \to M^n$. By A.5, the given curve c is a length-minimizing constant-speed geodesic in (M^n, d_p) if and only if it satisfies for all $u \le v \le w$ in $[0,1]$ that

$$E_p^{u,v}(c(u), c(v)) + E_p^{v,w}(c(v), c(w) - E_p^{u,w}(c(u), c(w)) = 0.$$

Equivalently, by the definitions of E and E_p,

$$\frac{1}{n} \sum_{i=1}^n \Big(E^{u,v}\big(c_i(u), c_i(v)\big) + E^{v,w}\big(c_i(v), c_i(w)\big) - E^{u,w}\big(c_i(u), c_i(w)\big) \Big) = 0.$$

As all summands are non-negative by the triangle inequality, they vanish. Equivalently, by Lemma A.5, all components $c_i\colon [0,1] \to M$ are constant-speed minimizing geodesics.

For $p = 1$, one uses a similar argument for the energy-action pairs (d, ℓ) and (d_1, ℓ_1), where ℓ is the length functional in (M, d), and ℓ_1 is the length functional in (M^n, d_1). However, in this case, a curve is minimizing for these energy-action pairs if and only if it is a geodesic, regardless of whether it has constant speed or not. $\square$

Theorem 3.6 (Geodesics between samples). *Let M be a connected complete locally compact path-metric space. Then any minimizing geodesic in the sample space $(M^n/S_n, \bar{d}_p)$ is the projection of a minimizing geodesic in the configuration space (M^n, d_p), which we call its horizontal lift.*

Proof. Let $\bar{c} \in C([0,1], M^n/S_n)$ be a constant-speed minimizing geodesic, and let $x \in \pi^{-1}(\bar{c}(0))$. For each $m \in \mathbb{N}$ we construct a curve $c_m \in C([0,1], M^n)$ as follows: Set $c_m(0) := x$, and then inductively, for each $k \in \{0, \ldots, m-1\}$, choose $c_m|[(k/m, (k+1)/m]$ as a constant-speed minimizing geodesic from $c_m(k/m)$ to the orbit $\pi^{-1}(\bar{c}((k+1)/m))$, until c_m reaches the orbit $\pi^{-1}(\bar{c}(1))$. The family $\{c_m : m \in \mathbb{N}\}$ is equicontinuous because the curves c_m have constant speed. Moreover, all curves c_m take values in the compact ball of radius $\bar{d}_p(\bar{c}(0), \bar{c}(1))$ around x, which is compact by the Hopf–Rinov theorem A.2. Thus, by the Arzelà–Ascoli theorem [53, Theorem 43.15], the set $\{c_m : m \in \mathbb{N}\}$ is pre-compact in the topology of uniform convergence and therefore has a cluster point $c \in C([0,1], M^n)$. The cluster point satisfies $\pi \circ c = \bar{c}$ because the curves c_m satisfy $\pi(c_m(k/m)) = \bar{c}(k/m)$ for all $0 \le k \le m$. By construction, c is a minimizing geodesic. $\square$

We next consider the special case where M is a finite-dimensional manifold with Riemannian metric g and complete geodesic distance d. Then $\frac{1}{n}(g \oplus \cdots \oplus g)$ is an S_n-invariant Riemannian metric on M^n, whose geodesic distance is the metric d_2 on M^n. The quotient space M^n/S_n carries a rich differential-geometric structure, which is described in detail in [43, Sections 29 and 30]. In particular, one obtains by differential-geometric arguments that a minimal geodesic segment is more regular at interior points than at the end points. This is formalized in the following theorem.

Theorem 3.7 (Interior regularity of Riemannian geodesics [2, 3.5 and 3.4]). *Let M be a finite-dimensional manifold with complete Riemannian metric g, and let M^n be the product manifold with the product Riemannian metric $\frac{1}{n}(g \oplus \cdots \oplus g)$. Then, for any lift $c \colon [0,1] \to M^n$ of a minimal geodesic segment in M^n/S_n, the isotropy group $(S_n)_{c(t)}$ of an interior point of c is contained in the isotropy groups $(S_n)_{c(0)}, (S_n)_{c(1)}$ of the end points.*

Thus, for any subgroup $H \leq S_n$, the set $(M^n/S_n)_{\leq(H)}$ of orbits with orbit type smaller or equal to (H) is a strictly convex subset of M^n/S_n. This means that any minimal geodesic segment between two points in $(M^n/S_n)_{\leq(H)}$ lies entirely in $(M^n/S_n)_{\leq(H)}$. In particular, the regular orbit-type stratum in M^n/S_n is a strictly convex open dense subset. Recall for comparison that $(M^n/S_n)_{\geq(H)}$ is convex by 3.3 but may not be strictly convex by 3.4.

Example 3.8 (Lack of interior regularity). *The assertion of 3.7 is wrong for non-Riemannian complete path-metric spaces.*

Proof. Let (M, d) be an open book space, for example the 3-spider, one of the simplest tree spaces [8].
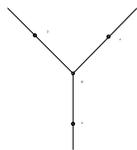
We choose 3 points x, y, z on the 3 lines with the same distance from the center 0. Let $c \colon [0,2] \to M^2$ be the minimal geodesic from $c(0) = (x,y)$ via $c(1) = (0,0)$ to $c(2) = (z,z)$. Then the isotropy group S_2 of $c(1)$ and $c(2)$ is not contained in the trivial isotropy group of $c(0) = (x,y)$.

See the related discussion in [51, Chapter 8]. Note that the 'curvature' of the spider at 0 is $-\infty$. $\qquad \square$

4. Infinite configuration and sample spaces

This section exhibits configuration spaces as spaces of random variables and sample spaces as spaces of probability distributions. Moreover, it identifies large-sample limits of these spaces. Throughout this section, (M, d) is a separable connected complete path-metric space, and $p \in [1, \infty)$.

Definition 4.1 (Random variables). For any complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we write $L^p(\Omega, M)$ for the space of all measurable functions $x \colon \Omega \to M$ which satisfy for one (or equivalently, all) $o \in M$ that $\|d(x, o)\|_{L^p(\Omega)} < \infty$. We endow the space $L^p(\Omega, M)$ with the metric

$$d_p(x, y) := \|d(x, y)\|_{L^p(\Omega)}, \qquad x, y \in L^p(\Omega, M).$$

Lemma 4.2 (Configurations as random variables). *For any $n \in \mathbb{N}$, the configuration space (M^n, d_p) is isometric to $(L^p(\{1, \ldots, n\}, M), d_p)$, where $\{1, \ldots, n\}$ is seen as a probability space with the uniform distribution.*

Proof. A configuration $x \in M^n$ is precisely a function $x \colon \{1, \ldots, n\} \to M$, and the metrics d_p defined on M^n and $L^p(\{1, \ldots, n\}, M)$ coincide. $\qquad\square$

The description of configurations as random variables allows one to pass to a *large-sample limit.* Similar results are shown in [39]. In the following lemma, $(0, 1)$ denotes the unit interval with the Lebesgue measure and could, for all purposes, be replaced by any standard probability space.

Lemma 4.3 (Infinite configurations)**.** *The configuration spaces (M^n, d_p) are isometrically embedded in the complete path-metric space $(L^p((0, 1), M), d_p)$ and converge to it in the following sense: for any compact $K \subset L^p((0, 1), M)$,*

$$\lim_{n \to \infty} \sup_{x \in K} \inf_{y \in M^n} d_p(x, y) = 0.$$

The lemma would imply pointed Gromov–Hausdorff convergence of (M^n, d_p) to the space $L^p((0, 1), M)$ if the uniform convergence on compacts could be strengthened to uniform convergence on bounded sets. However, this is not the case, as one easily verifies by considering functions x of the form $n^{1/p} \mathbb{1}_{[0, 1/n]}$ for large n.

Proof. The isometric immersion of $(M^n, d_p) \cong L^p(\{1, \ldots, n\}, M)$ into $L^p((0, 1), M)$ is given by the identification of n-tuples with piece-wise constant functions on $(0, 1)$. It remains to prove the convergence. Let $\epsilon > 0$. By the compactness of K, there are $m \in \mathbb{N}$ and $x_1, \ldots, x_m \in L^p((0, 1), M)$ such that the open d_p-balls $B_{\epsilon/3}(x_i)$ cover K. Let $o \in M$. By the dominated convergence theorem, there is $r > 0$ such that the configurations $y_i \in L^p((0, 1), M)$ defined by

$$y_i := \begin{cases} x_i, & d(x_i, o) \leq r, \\ o, & d(x_i, o) > r, \end{cases}$$

satisfy $d_p(x_i, y_i) \leq \epsilon/3$ for all $i \in \{1, \ldots, m\}$. Let F be the Banach space of continuous bounded functions on $B_r(o)$ with the uniform norm. Then $(B_r(o), d)$ embeds isometrically into F via the map $B_r(o) \ni a \mapsto d(a, \cdot) \in F$. Thus, $B_r(o)$ may be seen as a subset of F. Moreover, F is separable because $B_r(o)$ is separable. For any $n \in \mathbb{N}$, let $\mathbb{E}_n \colon L^p((0, 1), F) \to L^p((0, 1), F)$ be the conditional expectation with respect to the sigma-algebra generated by the intervals $[\frac{j-1}{n}, \frac{j}{n})$, $j \in \{1, \ldots, n\}$. Then, for sufficiently large n, the configurations $z_i := \mathbb{E}_n(y_i)$ satisfy $d_p(y_i, z_i) \leq \epsilon/3$ for all $i \in \{1, \ldots, m\}$. Let $A \colon F \to M$ be the metric projection from $f \in F$ to the nearest point $A(f) \in M$, and let $A_* \colon L^p((0, 1), F) \to L^p((0, 1), M)$ be the push-forward along A. Then the configurations $w_i := A_* z_i$ satisfy for all $i \in \{1, \ldots, m\}$ that

$$d_p(z_i, w_i) = d_p(z_i, A_* z_i) \leq d_p(z_i, y_i) \leq \epsilon/3.$$

It follows that every $x \in K$ is ϵ-close to some $w_i \in L^p(\{1, \ldots, n\}, M)$. $\qquad\square$

Recall that any continuous curve $c \colon [0, 1] \to L^p((0, 1), M)$ has a jointly measurable version $c \colon [0, 1] \times (0, 1) \to M$; see e.g. [15, Proposition 3.2]. Then the *sample paths* of c are the measurable functions $c(\cdot, \omega) \colon [0, 1] \to M$, $\omega \in (0, 1)$.

Lemma 4.4 (Geodesics between infinite configurations)**.** *$(L^p((0, 1), M), d_p)$ is a complete path-metric space. For $p > 1$, a continuous curve $c \colon [0, 1] \to L^p((0, 1), M)$ is a constant-speed minimizing geodesic in $(L^p((0, 1), M), d_p)$ if and only if almost all of its sample paths are constant-speed minimizing geodesics in M.*

Proof. To show that $L^p(\Omega, M)$ is a complete path-metric space, we proceed as in the proof of 3.1, noting that the point $c = c(a, b)$ can be chosen as a measurable function of a, b. Indeed, this follows from a measurable selection theorem [17] because the set

$$\Gamma \coloneqq \big\{(a, c, b) \in M^3 : \max\{d(a, c), d(c, b)\} \leq \alpha d(a, b)\big\}$$

is Polish, the projection $\Gamma \ni (a, c, b) \to (a, b) \in M^2$ is continuous, and the inverse image of any $(a, b) \in M^2$ under this projection is compact. To prove the statement about geodesics, we proceed as in the proof of 3.5 and associate Lagrangian energy–action pairs (E, A) and (E_p, A_p) to (M, d) and $(L^p((0, 1), M), d^p)$, respectively. By A.3 a continuous curve $c \colon [0, 1] \to L^p((0, 1), M)$ is a length-minimizing constant-speed geodesic if and only if it satisfies for all $u \leq v \leq w$ in $[0, 1]$ that

$$E_p^{u,v}(c(u), c(v)) + E_p^{v,w}(c(v), c(w) - E_p^{u,w}(c(u), c(w)) = 0.$$

Equivalently, by the definitions of E and E_p,

$$\mathbb{E}\big[E^{u,v}(c(u), c(v)) + E^{v,w}(c(v), c(w) - E^{u,w}(c(u), c(w))\big] = 0,$$

where $\mathbb{E}$ is the expectation with respect to the Lebesgue measure on $(0, 1)$. Equivalently, the following property holds almost surely: for all rational numbers $u \leq v \leq w$ in $[0, 1]$,

$$E^{u,v}(c(u), c(v)) + E^{v,w}(c(v), c(w) - E^{u,w}(c(u), c(w)) = 0.$$

By A.3 this implies for almost every $\omega \in (0, 1)$ that the sample path

$$[0, 1] \cap \mathbb{Q} \ni u \mapsto c(u, \omega)$$

is parameterized by constant speed. In particular, any such sample path can be extended continuously to all real numbers in $[0, 1]$. Thus, we have established that c has a version whose sample paths are almost surely constant-speed minimizing geodesics. Moreover, this property is equivalent to the previous ones. □

On finite probability spaces, the statement about geodesics in 4.4 extends to $p = 1$ if the constant-speed condition is omitted, as shown in 3.5. However, this is not the case on infinite probability spaces, as the following example shows.

Example 4.5 (Discontinuity of sample paths). *Constant-speed minimizing geodesics in $L^1((0, 1), M)$ may have discontinuous sample paths.*

Proof. Let $M = \mathbb{R}$. The curve

$$c \colon [0, 1] \times (0, 1) \to M, \qquad c(t, \omega) \coloneqq \mathbb{1}_{[t,1]}(\omega)$$

is a constant-speed minimizing geodesic in $L^1((0, 1), M)$, but none of its sample paths are continuous. □

Definition 4.6 (probability distributions). Let $\mathcal{P}^p(M)$ denote the space of all probability distributions P on M which satisfy for one (equivalently, all) $o \in M$ that $\|d(o, \cdot)\|_{L^p(P)} < \infty$. We endow $\mathcal{P}^p(M)$ with the *Wasserstein metric*,

$$\bar{d}_p(P, Q) = \inf_R \|d(\cdot, \cdot)\|_{L^p(R)}, \quad P, Q \in \mathcal{P}^p(M),$$

where the infimum is over all probability distributions R on $M \times M$ with marginals P, Q. Moreover, we write $\mathcal{P}_n(M)$ for the subset of all *atomic probability distributions* of the form $\frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, where δ_{x_i} is the Dirac measure centered at $x_i \in M$.

As an aside, the set $\mathcal{P}_n(M)$ of atomic distributions can equivalently be characterized as the set of $\{0, 1/n, \ldots, 1\}$-valued probability measures. This equivalence uses the separability of M and is shown in A.6. The following lemma identifies samples with probability distributions, namely, with their *empirical laws*.

Lemma 4.7 (Samples as probability distributions). *For any $n \in \mathbb{N}$, the sample space $(M^n/S_n, \bar{d}_p)$ is isometric to the space $(\mathcal{P}_n(M), \bar{d}_p)$ of atomic probability distributions.*

Proof. Samples $\bar{x} = \pi(x) \in M^n/S_n$ are naturally identified with atomic probability distributions $P = \frac{1}{n} \sum_{i=1}^n \delta_{x_i} \in \mathcal{P}_n(M)$. If $\bar{y} = \pi(y) \in M^n/S_n$ is another sample with corresponding probability distribution $Q = \frac{1}{n} \sum_{i=1}^n \delta_{y_i} \in \mathcal{P}_n(M)$, then

$$\bar{d}_p(\bar{x}, \bar{y}) = \min_{\pi(x)=\bar{x}, \pi(y)=\bar{y}} d_p(x, y) = \min_{\pi(x)=\bar{x}, \pi(y)=\bar{y}} \|d(x, y)\|_{L^p(\{1,\ldots,n\})}$$
$$= \min_R \|d(\cdot, \cdot)\|_{L^p(R)},$$

where the last minimum is over all atomic probability distributions $R \in \mathcal{P}_n(M \times M)$ with marginal laws P and Q. By Birkhoff's theorem, one may equivalently take the minimum over the larger set of all (not necessarily atomic) probability distributions R on $M \times M$ with marginal laws P and Q [44, Proposition 1.3.1]. This shows that the right-hand side equals $\bar{d}_p(P, Q)$. Therefore, the identification of samples with probability distributions is an isometry. $\square$

Lemma 4.8 (Infinite samples). *The sample spaces $(M^n/S_n, \bar{d}_p)$ are isometrically embedded in the complete path-metric space $(\mathcal{P}^p(M), \bar{d}_p)$. For M locally compact, they converge to $(\mathcal{P}^p(M), \bar{d}_p)$ in the following sense: for any compact $K \subset \mathcal{P}^p(M)$,*

$$\lim_{n \to \infty} \sup_{P \in K} \inf_{Q \in M^n/S_n} \bar{d}_p(P, Q) = 0.$$

Here M^n/S_n is identified with the subset $\mathcal{P}_n(M)$ of $\mathcal{P}^p(M)$ using 4.7.

Proof. The sample space $(M^n/S_n, \bar{d}_p)$ is isometrically embedded in $(\mathcal{P}^p(M), \bar{d}_p)$ as a consequence of 4.7. It is well-known that the Wasserstein metric $\bar{d}_p$ on $\mathcal{P}_p(M)$ is a complete path metric [51, Theorem 6.18 and Corollary 7.22]. It remains to prove the convergence. Let $\epsilon > 0$. As K is compact, there are $m \in \mathbb{N}$ and $P_1, \ldots, P_m \in K$ such that the open $\bar{d}_p$-balls $B_{\epsilon/2}(P_i)$ cover K. For each $i \in \{1, \ldots, m\}$, the empirical distributions of P_i converge to P_i in the Wasserstein distance $\bar{d}_p$ [44, Proposition 2.2.6]. Therefore, there are distributions $Q_1, \ldots, Q_m \in \mathcal{P}_n(M)$ for some $n \in \mathbb{N}$ such that $\bar{d}_p(P_i, Q_i) \leq \epsilon/2$ for all $i \in \{1, \ldots, m\}$. It follows that every $P \in K$ is ϵ-close to some distribution in $\mathcal{P}_n(M)$. $\square$

Recall from 3.2 that the sample space $(M^n/S_n, \bar{d}_p)$ is the path-metric quotient of the configuration space (M^n, d_p) with respect to the action of permutation group of $\{1, \ldots, n\}$. A similar statement applies to infinite sample and configuration spaces, as shown in the following lemma. In analogy to 3.2, let $\pi \colon L^p((0, 1), M) \to \mathcal{P}^p(M)$ be the map from random variables to their law or, in more analytic terms, the push-forward of the Lebesgue measure along the given measurable function. Moreover, let $\mathrm{Aut}((0, 1))$ be the automorphism group of the probability space $(0, 1)$, i.e., the group of bi-measurable measure-preserving functions from $(0, 1)$ to itself.

Lemma 4.9 (Quotient structure). *The Wasserstein metric $\bar{d}_p$ on $\mathcal{P}^p(M)$ is a quotient metric:*

$$\bar{d}_p(P, Q) = \inf_{\pi(x)=P, \pi(y)=Q} d_p(x, y) = \inf_{\sigma \in \mathrm{Aut}((0,1))} d_p(x, y \circ \sigma),$$

where $P, Q \in \mathcal{P}^p(M)$ and $x, y \in L^p((0,1), M)$ with $\pi(x) = P$, $\pi(y) = Q$.

Proof. The first equality holds because any coupling R in the definition 4.6 of the Wasserstein metric is the joint law of some random variables $x, y \in L^p((0,1), M)$. The second equality holds because the action of $\mathrm{Aut}((0,1))$ is nearly transitive on the fibers of π in the following sense [12, Lemma 6.4]: for all $x, y \in L^p((0,1), M)$ with $\pi(x) = \pi(y)$ and all $\epsilon > 0$, there exists $\sigma \in \mathrm{Aut}((0,1))$ such that $d_p(x, y \circ \sigma) \leq \epsilon$. $\qquad\square$

The following lemma generalizes 3.6 from finite to infinite configurations and samples, respectively.

Theorem 4.10 (Geodesics between infinite samples). *Let M be a connected complete locally compact path-metric space. Then any minimizing geodesic in the infinite sample space $(\mathcal{P}^p(M), \bar{d}_p)$ is the projection of a minimizing geodesic in the configuration space $(L^p(\Omega, M), d_p)$, which we call its horizontal lift.*

Proof. This is proven in [51, Corollary 7.22] along the same lines as 3.6, i.e., using Lagrangian energy-action pairs. The horizontal lift is called displacement interpolation there. $\qquad\square$

Skeleta and orbit-type strata of finite sample spaces M^n/S_n were defined in 2.1 and 2.7, respectively. Via the isometry 4.7 to atomic probability distributions and the isometric embedding 4.8 into p-integrable probability distributions, one obtains straight-forward extensions to skeleta and orbit-type strata of infinite sample spaces, as defined next.

Definition 4.11 (Infinite skeleta and orbit-type strata). For any $q \in \mathbb{N}$, the q-*skeleton* in the infinite-sample space $\mathcal{P}^p(M)$ is the subset $\mathcal{P}(M)_q$ of all probability distributions whose support is a set of at most q points. Similarly, for any partition $(\mathbf{w}) \coloneqq (w_1 \geq \cdots \geq w_q)$ of 1 consisting of non-negative real numbers w_i summing up to 1, the $(\mathbf{w})$-*stratum* in the infinite-sample space $\mathcal{P}^p(M)$ is the subset of all $P = \sum_{i=1}^q w_i \delta_{x_i} \in \mathcal{P}(M)_q$ with distinct points x_i. The measure P is called *regular* if the points x_i are distinct and the weights w_i are strictly positive.

5. Means and polymeans

In this section, we generalize Fréchet means [25] and k-means [40] to *polymeans* using the path-metric structure of sample space. Background and further references on Fréchet means can be found in the textbook [47]. Throughout this section, we consider the configuration space (M^n, d_p) and sample space $(M^n/S_n, \bar{d}_p)$ of a connected complete path-metric space (M, d) for some $n \in \mathbb{N}$ and $p \in [1, \infty)$. The following definition introduces polymeans as metric projections onto certain subsets of sample space M^n/S_n, namely q-skeleta $(M^n/S_n)_q$ (see 2.2) or $(\mathbf{k})$-strata $(M^n/S_n)_{(\mathbf{k})}$ (see 2.8).

Definition 5.1 (Polymeans). For any $q \in \mathbb{N}$, a q-*mean* of a sample is a $\bar{d}_p$-nearest point in the q-skeleton of sample space. Similarly, for any partition $(\mathbf{k})$ of n, a $(\mathbf{k})$-*mean* of a sample is a $\bar{d}_p$-nearest point in the closure of the $(\mathbf{k})$-stratum.

Recall that the q-skeleton is closed, and the closure of the $(\mathbf{k})$-stratum is the union of all $(\mathbf{k}')$-strata with $(\mathbf{k}') \leq (\mathbf{k})$. This ensures the existence of q-means and $(\mathbf{k})$-means, as shown next. One should be aware that a q-mean might consist of less than q distinct points, and similarly a $(\mathbf{k})$-mean might have orbit type $(\mathbf{k}')$ with $(\mathbf{k}') \leq (\mathbf{k})$.

Lemma 5.2 (Existence of polymeans). *If M is a complete locally compact path-metric space, then every sample $\bar{x} \in M^n/S_n$ has a q-mean and a $(\mathbf{k})$-mean, for each $q \in \mathbb{N}_{>0}$ and orbit type $(\mathbf{k}) \coloneqq (k_1 \geq \cdots \geq k_q)$.*

Proof. For sufficiently large $r > 0$, the closed ball $B_r(\bar{x})$ has non-empty intersection with the q-skeleton. By the Hopf–Rinow theorem A.2, this intersection is compact and therefore contains a point of minimal $\bar{d}_p$-distance to $\bar{x}$. The argument for the $(\mathbf{k})$-stratum is similar. $\square$

Generic configurations have unique polymeans, as shown next. Here generic is understood in a measure-theoretic sense, i.e., up to null sets with respect to a given Riemannian volume form.

Lemma 5.3 (Uniqueness of polymeans). *Let M be a complete finite-dimensional Riemannian manifold, and assume that $p = 2$. Then the configurations $x \in M^n$ such that $\pi(x)$ has more than one q-mean or more than one $(\mathbf{k})$-mean are a null set with respect to the Riemannian volume form.*

Proof. We consider M^n as a complete Riemannian manifold with Riemannian distance d_2. Let K be the q-skeleton or the $(\mathbf{k})$-stratum in M^n, and let C be the set of all points in M^n whose distance to K is realized by more than one geodesic (sometimes called the medial axis). At any point in C, the squared distance function to K is non-differentiable [41, Remark 3.6]. These points of non-differentiability constitute a C^2-rectifiable set [41, Proposition 3.7]. Thus, its subset C has vanishing measure. $\square$

We next show that the definition of polymeans extends the definition of Fréchet p-means.

Example 5.4 (Fréchet means). *Fréchet means correspond exactly to 1-means or, equivalently, (n)-means, where (n) denotes the trivial partition.*

Proof. Recall that the 1-skeleton in sample space M^n/S_n consists of all $\bar{y} = \pi(y, \ldots, y)$ with $y \in M$ and coincides with the orbit-type stratum $(M^n/S_n)_{(n)}$, where (n) denotes the partition of n of length 1. Thus, 1-means coincide with (n)-means and minimize, for a given $\bar{x} = \pi(x)$ in M^n/S_n, the functional

$$\bar{d}_p(\bar{x}, \bar{y}) = \left(\frac{1}{n} \sum_{i=1}^{n} d(x_i, y)^p \right)^{1/p}$$

over all $\bar{y} = \pi(y, \ldots, y)$ in the 1-skeleton of M^n/S_n. Minimizers of the right-hand side, seen as a function of $y \in M$, are exactly Fréchet means. Thus, a point $y \in M$ is a Fréchet mean of a configuration $x \in M^n$ if and only if the sample $\pi((y, \ldots, y)) \in M^n/S_n$ is a 1-mean, or equivalently an (n)-mean, of $\pi(x) \in M^n/S_n$. $\square$

k-mean clustering remains a very popular method in cluster analysis, more than 60 years after [40, 32]. Like the Fréchet p-mean, it can be generalized with the

power p of the distance [54]. We show below that this corresponds to our geometric definition of polymeans.

Example 5.5 (k-means). *q-means correspond exactly to k-means clustering for $k = q \in \mathbb{N}$.*

Proof. Let $\bar{x}, \bar{y} \in M^n/S_n$ with $\bar{y}$ belonging to the q-skeleton. Then there are lifts $x, y \in M^n$ such that $\pi(x) = \bar{x}$, $\pi(y) = \bar{y}$, and $d_p(x, y) = \bar{d}_p(\bar{x}, \bar{y})$. The set $\{1, \ldots, n\}$ can be partitioned into non-empty subsets $A_1, \ldots, A_q$ such that $y_i = y_j$ for any $i, j \in S_k$ and $k \in \{1, \ldots, q\}$. Then

$$n\bar{d}_p(\bar{x}, \bar{y})^p = \sum_{i=1}^{q} \sum_{j \in A_i} d(x_j, y_i)^p.$$

The left-hand side is minimized by q-means $\bar{y}$, and the right-hand side is minimized by partitions $A_1, \ldots, A_k$ and k-means $(y_1, \ldots, y_k)$ with $k = q$. Therefore, the q-mean and k-mean problems are equivalent. As an aside, the q-mean vector $\bar{y}$ does not encode the optimal correspondence between points x_i and y_i, and the k-mean vector $(y_1, \ldots, y_k)$ does not encode the multiplicities $\#A_i$. However, this information can be retrieved easily by matching each point x_j to the nearest point y_i. $\qquad\square$

Definition 5.6 (Clusters). A *clustering* of a sample $\bar{x} \in M^n/S_n$ is a representation $\bar{x} = \bar{x}_1 \sqcup \cdots \sqcup \bar{x}_q := \pi((x_1, \ldots, x_q))$, where $\bar{x}_i = \pi(x_i) \in M^{k_i}/S_{k_i}$ for some partition $k_1 + \cdots + k_q = n$ with $k_i \in \mathbb{N}_{>0}$ and $q \in \mathbb{N}_{>0}$. In this situation, $\bar{x}_i$ are called *clusters* or *sub-samples* of sizes k_i.

Lemma 5.7 (Polymeans as clusters). *If $\bar{y}$ is a q-mean of $\bar{x}$, then there are clusterings $\bar{x} = \bar{x}_1 \sqcup \cdots \sqcup \bar{x}_q$ and $\bar{y} = \bar{y}_1 \sqcup \cdots \sqcup \bar{y}_q$ such that each $\bar{y}_i$ is a 1-mean of $\bar{x}_i$. Moreover, if $\bar{y}$ is a $(\mathbf{k})$-mean of $\bar{x}$ with $(\mathbf{k}) := (k_1 \geq \cdots \geq k_q)$, then the partition can be chosen such that each cluster $\bar{x}_i$ has size k_i.*

Proof. Let $A_1, \ldots, A_q$ be a partition of $\{1, \ldots, n\}$ as in the proof of 5.6. Then the clusterings $\bar{x}_i = \pi((x_j)_{j \in A_i})$ and $\bar{y}_i = \pi((y_j)_{j \in A_i})$ have the desired property. $\qquad\square$

Lemma 5.7 exhibits polymeans as *weighted means*, where the weights correspond to the cluster sizes, normalized by the total number of samples. The same interpretation is obtained by identifying polymeans with atomic measures via 4.7. In some situations it may be advantageous to consider *unweighted polymeans*, which encode only the locations but not the weights of the clusters. The following definition describes q such clusters located at mutually distinct points $y_1, \ldots, y_q \in M$. Recall that the ensemble of such mutually distinct point configurations modulo permutations is the regular stratum $(M^q/S_q)_{\text{reg}}$.

Definition 5.8 (Unweighted q-means). For any $q \in \mathbb{N}$, an *unweighted q-mean* of a sample $\bar{x} = \pi(x) \in M^n/S_n$ is a regular q-sample $\bar{z} \in (M^q/S_q)_{\text{reg}}$ which minimizes the functional

$$(M^q/S_q)_{\text{reg}} \ni \bar{z} = \pi(z) \mapsto \sum_{i=1}^{n} \min_{j \in \{1, \ldots, q\}} d(x_i, z_j)^p.$$

Unweighted q-means may fail to exist for a given $q \in \mathbb{N}_{>0}$ because the regular stratum $(M^q/S_q)_{\text{reg}}$ is not closed. It is, however, open and dense. Thus, for any given $q \in \mathbb{N}_{>0}$, there always exists an unweighted q'-mean with $q' \leq q$. The

definitions of weighted and unweighted polymeans are consistent with each other in the following sense.

Lemma 5.9 (Relation between weighted and unweighted q-means)**.** *Let $\bar{x} \in M^n/S_n$, and let $z_1, \ldots, z_q$ be distinct points in M. Then $\pi(z_1, \ldots, z_q) \in (M^q/S_q)_{reg}$ is an unweighted q-mean of $\bar{x}$ if and only if*

$$\pi((\underbrace{z_1, \ldots, z_1}_{k_1 \ times}, \ldots, \underbrace{z_q, \ldots, z_q}_{k_q \ times})) \in M^n/S_n$$

is a q-mean of $\bar{x}$ for some integer weights k_i summing up to n.

Proof. This easily follows from the definitions. $\qquad\square$

Skeleta and orbit-type strata in infinite sample space $\mathcal{P}^p(M)$ were defined in 4.10. This yields the following straight-forward extensions to polymeans of infinite samples.

Definition 5.10 (Population polymeans)**.** A *population q-mean* of an infinite sample $P \in \mathcal{P}^p(M)$ is a $\bar{d}_p$-nearest point in the q-skeleton of $\mathcal{P}^p(M)$. Similarly, for any partition $(\mathbf{k}) := (k_1 \geq \cdots \geq k_q)$ consisting of non-negative real numbers k_i summing up to 1, a *population $(\mathbf{k})$-mean* of $P \in \mathcal{P}^p(M)$ is a $\bar{d}_p$-closest point in the $(\mathbf{k})$-stratum of $\mathcal{P}^p(M)$. Moreover, an *unweighted population q-mean* of $P \in \mathcal{P}^p(M)$ is a $\bar{d}_p$-closest point in the regular stratum of $\mathcal{P}_q(M)$.

6. Random samples

Throughout this section, we consider the configuration space (M^n, d_p) and sample space $(M^n/S_n, \bar{d}_p)$ of a separable complete path-metric space (M, d) for some $n \in \mathbb{N}$ and $p \in [1, \infty)$. We use the letter $\mathcal{P}$ to designate probability distributions. Thus, $\mathcal{P}(M^n/S_n)$ is the set of probability distributions on sample space, and $\mathcal{P}(M^n)$ is the set of all probability distributions on configuration space. Moreover, we write $\mathcal{P}(M^n)_{S_n}$ for the subset of *symmetric* probability distributions, where symmetry means S_n-invariance.

Lemma 6.1 (Distributions of samples)**.** *Probability distributions on sample space M^n/S_n correspond exactly to symmetric probability distributions on configuration space M^n.*

Proof. We claim that the projection from configuration onto sample space induces a bijection

$$\mathcal{P}(M^n)_{S_n} \ni P \mapsto \pi_* P \in \mathcal{P}(M^n/S_n).$$

To prove the claim, we will construct an inverse of this map by randomization over the S_n-orbit using the probability kernel

$$K \colon M^n \ni x \mapsto \frac{1}{n!} \sum_{\sigma \in S_n} \delta_{x_\sigma} \in \mathcal{P}(M^n)_{S_n}.$$

This kernel is S_n-invariant and consequently descends to a probability kernel

$$(1) \qquad \bar{K} \colon M^n/S_n \ni \bar{x} = \pi(x) \mapsto \frac{1}{n!} \sum_{\sigma \in S_n} \delta_{x_\sigma} \in \mathcal{P}(M^n)_{S_n},$$

which maps samples $\bar{x}$ to uniform distributions on their fibers $\pi^{-1}(x)$ in configuration space. The two kernels are related by $K = \bar{K} \circ \pi$. For any probability

distribution $\bar{P}$ on M^n/S_n, we write $\int \bar{K}(\bar{x})\bar{P}(\mathrm{d}\bar{x})$ for the composition of the kernel $\bar{K}$ with the probability distribution $\bar{P}$. Formally, this is a measure-valued Pettis integral. Then the map

$$(2) \qquad \mathcal{P}(M^n/S_n) \ni \bar{P} \mapsto \int \bar{K}(\bar{x})\bar{P}(\mathrm{d}\bar{x}) \in \mathcal{P}(M^n)_{S_n}$$

is an inverse to the map π_* because

$$\pi_* \int \bar{K}(\bar{x})\bar{P}(\mathrm{d}\bar{x}) = \int \pi_*\big(\bar{K}(\bar{x})\big)\bar{P}(\mathrm{d}\bar{x}) = \int \delta_{\bar{x}}\bar{P}(\mathrm{d}\bar{x}) = \bar{P},$$

$$\int \bar{K}(\bar{x})(\pi_* P)(\mathrm{d}\bar{x}) = \int \bar{K}\big(\pi(x)\big)P(\mathrm{d}x) = \int K(x)P(\mathrm{d}x)$$

$$= \frac{1}{n!}\sum_{\sigma \in S_n} \int \delta_{x_\sigma}P(\mathrm{d}x) = \frac{1}{n!}\sum_{\sigma \in S_n}(r_\sigma)_* P = P,$$

where $r_\sigma \colon M^n \ni x \mapsto x_\sigma \in M^n$ is the action of the permutation σ on the configuration space, and where the last equality follows from the symmetry of P. $\square$

Hewitt and Savage [27, Section 12] characterized the set of extremal points within the convex set of symmetric probability distributions on M^n, for short, *extremal distributions*. Moreover, they proved that every symmetric probability distribution is a mixture of extremal distributions and called such mixtures *presentable*. As a corollary to Lemma 6.1, one obtains an elementary proof of these facts. The more widely studied case of infinite configurations is discussed in 6.3 and 6.4.

Corollary 6.2 (Finite Hewitt–Savage theorem)**.** *The extremal points in the convex set* $\mathcal{P}(M^n)_{S_n}$ *of symmetric distributions are exactly of the form* $\frac{1}{n!}\sum_{\sigma \in S_n}\delta_{x_\sigma}$, $x \in M^n$. *Moreover, all symmetric probability distributions on* M^n *are presentable.*

Proof. The map (6.1.2) is a linear bijection and therefore maps extremal points in its domain to extremal points in its range. The extremal points in the domain are easily identified as the Dirac measures. The image of a Dirac measure $\delta_{\bar{x}}$ with $\bar{x} = \pi(x) \in M^n/S_n$ is the distribution $\frac{1}{n!}\sum_{\sigma \in S_n}\delta_{x_\sigma}$. The range of the map (6.1.2) consists of mixtures of such distributions, i.e., presentable distributions. Moreover, as (6.1.2) is surjective, all symmetric distributions are presentable. $\square$

The following lemma characterizes distributions of infinite samples, thereby generalizing the corresponding result 6.1 for finite samples. The full permutation group $S_{\mathbb{N}}$ of the natural numbers is too large for our purpose. Instead, we consider the *infinite permutation group* $S_{(\mathbb{N})} := \bigcup_{n \in \mathbb{N}} S_n$, which acts upon the *infinite configuration space* $M^{\mathbb{N}} := \prod_{n \in \mathbb{N}} M$. A probability distribution on $M^{\mathbb{N}}$ is called *symmetric* if it is $S_{(\mathbb{N})}$-invariant, and the set of symmetric distributions is denoted by $\mathcal{P}(M^{\mathbb{N}})_{S_{(\mathbb{N})}}$. The correct space of *infinite samples*, which leads to a generalization of 6.1, is not the quotient space $M^{\mathbb{N}}/S_{(\mathbb{N})}$, but the space $\mathcal{P}(M)$. This is demonstrated in Example 6.5 and is in line with the limiting result 4.8.

Lemma 6.3 (Distributions of infinite samples)**.** *Probability distributions on the infinite sample space* $\mathcal{P}(M)$ *correspond exactly to symmetric probability distributions on the configuration space* $M^{\mathbb{N}}$.

Proof. For some fixed point $o \in M$, define a projection from infinite configuration space to infinite sample space as follows:

$$\pi \colon M^{\mathbb{N}} \to \mathcal{P}(M), \qquad \pi(x) \coloneqq \begin{cases} \displaystyle\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}, & \text{if the weak limit exists,} \\ \delta_o, & \text{otherwise,} \end{cases}$$

The push-forward along this projection restricts to the following map from symmetric distributions to probability distributions on infinite sample space $\mathcal{P}(M)$:

$$\pi_* \colon \mathcal{P}(M^{\mathbb{N}})_{S_{(\mathbb{N})}} \to \mathcal{P}(\mathcal{P}(M)).$$

We claim that the map π_* is an inverse of the map

$$\mathcal{P}(\mathcal{P}(M)) \ni Q \mapsto \int_{\mathcal{P}(M)} P^{\mathbb{N}} \, Q(\mathrm{d}P) \in \mathcal{P}(M^{\mathbb{N}})_{S_{(\mathbb{N})}},$$

where $P^{\mathbb{N}} \coloneqq \bigotimes_{n \in \mathbb{N}} P$ denotes the product distribution on $M^{\mathbb{N}}$, and where the integral is a measure-valued Pettis integral. Note that the distributions on the right-hand side are laws of conditionally i.i.d. sequences of M-valued random variables. To prove the claim, we appeal to the infinite-sample version 6.4 of the Hewitt–Savage theorem, which states that symmetric distributions coincide exactly with presentable distributions, i.e., with Pettis integrals as above. For any $P \in \mathcal{P}(M)$, the weak law of large numbers implies $\pi(x) = P$ for $P^{\mathbb{N}}$-almost every $x \in M^{\mathbb{N}}$. This implies $\pi_*(P^{\mathbb{N}}) = \delta_P$. Consequently, every $Q \in \mathcal{P}(\mathcal{P}(M))$ satisfies

$$\pi_* \left(\int_{\mathcal{P}(M)} P^{\mathbb{N}} \, Q(\mathrm{d}P) \right) = \int_{\mathcal{P}(M)} \pi_*(P^{\mathbb{N}}) \, Q(\mathrm{d}P) = \int_{\mathcal{P}(M)} \delta_P \, Q(\mathrm{d}P) = Q.$$

This proves the claim and establishes the desired one-to-one correspondence.   □

The above proof uses the well-known Hewitt–Savage theorem [27], which is a generalization of Corollary 6.2 to infinite sample spaces. As before, *presentable distributions* are defined as mixtures of *extremal distributions*, i.e., of extremal points in the convex set of symmetric distributions.

Theorem 6.4 (Infinite Hewitt–Savage theorem [27])**.** *The extremal points in the convex set $\mathcal{P}(M^{\mathbb{N}})_{S_{(\mathbb{N})}}$ of symmetric distributions are exactly the product distributions $P^{\mathbb{N}} \coloneqq \bigotimes_{n \in \mathbb{N}} P$ with $P \in \mathcal{P}(M)$. Moreover, all symmetric distributions on $M^{\mathbb{N}}$ are presentable.*

This result is asymptotically consistent with its finite-sample counterpart 6.2. Indeed, by the Diaconis–Freedman theorem [18] symmetric distributions on M^n are close to mixtures of product distributions for large n. More precisely, the total variation distance from k-dimensional marginal distributions of elements of $\mathcal{P}(M^n)_{S_n}$ to mixtures of product distributions is at most $k(k-1)/n$.

The following example shows that the correspondence 6.3 between probability distributions on sample space and symmetric probability distributions on configuration space fails if the sample space is defined as $M^{\mathbb{N}}/S_{(\mathbb{N})}$ instead of $\mathcal{P}(M)$.

Example 6.5 (Infinite sample space)**.** *There is a probability distribution on $M^{\mathbb{N}}/S_{(\mathbb{N})}$ which does not correspond to any symmetric probability distribution on $M^{\mathbb{N}}$.*

Proof. In analogy to 6.3 we say that a probability distribution Q on $M^{\mathbb{N}}/S_{(\mathbb{N})}$ corresponds to a symmetric probability distribution P on $M^{\mathbb{N}}$ if $Q = \pi_* P$, where $\pi \colon M^{\mathbb{N}} \to M^{\mathbb{N}}/S_{(\mathbb{N})}$ is the canonical projection. For any such P, the weak limit $\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}$ exists for P-almost every $x \in M^{\mathbb{N}}$, as shown in the proof of 6.3. Moreover, this limit is invariant under the action of $S_{(\mathbb{N})}$ on $M^{\mathbb{N}}$ because every permutation in $S_{(\mathbb{N})}$ affects only finitely many indices. Thus, if Q corresponds to some P, then the limit $\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{\bar{x}_i}$ is well-defined and exists for Q-almost every $\bar{x} \in M^{\mathbb{N}}/S_{(\mathbb{N})}$. However, it is easy to construct a distribution Q on $M^{\mathbb{N}}/S_{(\mathbb{N})}$ which does not have this property. Indeed, assuming that M contains at least two points, one may construct a sequence of points $x_i \in M$ such that $\frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}$ does not converge weakly as $n \to \infty$. Then $Q := \delta_{\bar{x}}$ with $\bar{x} := \pi(x)$ is the desired counter-example. $\qquad\square$

We next investigate random samples and random configurations. For this purpose, we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which all random variables are defined. A *random configuration* is a random variable in M^n or $M^{\mathbb{N}}$, depending on the finite versus infinite case. Similarly, a *random sample* is a random variable in M^n/S_n or $\mathcal{P}(M)$, respectively. A random configuration is called *exchangeable* if its law is symmetric, i.e., invariant under permutations in S_n or $S_{(\mathbb{N})}$, respectively. The following characterization is analogous to 6.1–6.4.

Corollary 6.6 (Random configurations and samples). *Random samples correspond exactly (possibly after passing to an extended probability space) to exchangeable configurations, which in turn correspond exactly to conditionally i.i.d. M-valued random variables. This statement applies to finite and infinite configurations and samples, respectively.*

Proof. This can be shown in analogy to 6.1–6.4, working with random variables instead of their laws. The extension of the probability space is necessary, unless the given probability space is already sufficiently rich, for implementing the random ordering in the proof of 6.1 and the i.i.d. sampling in the proof of 6.3. $\qquad\square$

7. Asymptotic properties of polymeans

Polymeans, similar to Fréchet means [29], satisfy a law of large numbers and a central limit theorem under suitable conditions, as shown next. We refer to 5.1, 5.8, and 5.10 for their definition. Throughout this section, (M, d) is a separable complete connected path-metric space, $p \in [1, \infty)$, and $q \in \mathbb{N}_{>0}$. The space M, as well topological products and quotients thereof, are endowed with the corresponding Borel sigma algebras. For some probability distribution $P \in \mathcal{P}^p(M)$, we consider a sequence of independent P-distributed random variables $(x_i)_{i \in \mathbb{N}}$ defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The corresponding n-samples are denoted by $\bar{x}_n := \pi(x_1, \ldots, x_n) \in M^n/S_n$. We write $\mu_n \subset (M^n/S_n)_q$ for the set of q-means of $\bar{x}_n$, $\bar{y}_n \in (M^n/S_n)_q$ for a measurable selection of q-means of $\bar{x}_n$, and $\bar{z}_n \in (M^q/S_q)_{\mathrm{reg}}$ for a measurable selection of unweighted q-means of $\bar{x}_n$. It will be convenient to identify the samples $\bar{x}_n, \bar{y}_n, \bar{z}_n$ with their empirical laws P_n, Q_n, R_n, respectively, using the isometry 4.7 between M^n/S_n and $\mathcal{P}_n(M)$. The population counterparts of the above empirical objects are denoted by $\mu_0, \bar{y}_0, \bar{z}_0, Q_0$, and R_0, respectively. Note that all of these objects belong to one and the same path-metric space $\mathcal{P}^p(M)$ thanks to the isometric embedding 4.8 of finite into infinite sample spaces.

Definition 7.1 (Strong consistency [55]). The empirical q-means μ_n are called *strongly consistent* estimators for the set μ_0 of population q-means if

$$\mathbb{P}\left[\bigcap_{n=1}^{\infty}\overline{\bigcup_{k=n}^{\infty}\mu_k}\subseteq\mu_0\right]=1.$$

Note that strong consistency is equivalent to the following statement: with probability 1, any accumulation point of the sets μ_n belongs to μ_0.

Lemma 7.2 (Strong consistency). *The empirical q-means μ_n are strongly consistent estimators for the population q-means μ_0.*

This statement is a consequence of the Gamma-convergence of the functionals which are minimized by μ_n and μ_0, respectively, as shown in the following proof. A similar argument is used in [55] and [29, Theorem A.3]. These proofs are longer because implications of Gamma-convergence are re-proven there.

Proof. The empirical q-means μ_n are the minimizers of the functional

$$F_n\colon \mathcal{P}^p(M)_q\to\mathbb{R}_+,\qquad F_n(Q)=\begin{cases}\bar{d}_p(P_n,Q), & Q\in(M^n/S_n)_q,\\ \infty, & Q\notin(M^n/S_n)_q.\end{cases}$$

Similarly, the population q-means μ are the minimizers of the functional

$$F\colon \mathcal{P}^p(M)_q\to\mathbb{R}_+,\qquad\qquad F(Q)=\bar{d}_p(P,Q).$$

The empirical laws P_n converge to the population law P in the Wasserstein metric $\bar{d}_p$ by [44, Proposition 2.2.6]. We claim that this implies Gamma-convergence $F_n\to F$. To prove the claim, note that for any converging sequence $Q_n\to Q$ in $\mathcal{P}^p(M)_q$,

$$F(Q)=\bar{d}_p(P,Q)=\lim_{n\to\infty}\bar{d}_p(P_n,Q_n)\leq\liminf_{n\to\infty}F_n(Q_n).$$

Moreover, any $Q\in\mathcal{P}^p(M)_q$ can be approximated in the $\bar{d}_p$-distance by a sequence $Q_n\in(M^n/S_n)_q$. Indeed, Q is of the form $Q=\sum_{i=1}^q w_i\delta_{x_i}$ for some $x_i\in M$ and $w_i\in[0,1]$, and the approximations Q_n may be defined by rounding the weights to the nearest multiples of $1/n$. For any such approximating sequence $Q_n\to Q$ one has

$$F(Q)=\bar{d}_p(P,Q)=\lim_{n\to\infty}\bar{d}_p(P_n,Q_n)=\lim_{n\to\infty}F_n(Q_n).$$

This proves that F_n Gamma-converges to F. Thus, the accumulation points of F_n-minimizers are F-minimizers, which is exactly strong consistency. $\qquad\square$

If the empirical q-means are strongly consistent and the population q-mean is unique, then any measurable selection Q_n of empirical q-means converges in probability to the population q-mean Q_0. In this situation one may inquire about the rate of convergence $Q_n\to Q_0$. As an auxiliary first step, the following lemma shows that Q_n possesses the same best-approximation property as Q_0, up to some error terms. Controlling these error terms leads to the convergence rate established subsequently in 7.4.

Lemma 7.3 (Error bound). *Assume that $P\in\mathcal{P}^{2p}(M)$, let $Q_0\in\mathcal{P}(M)_q$ be a q-mean of P, assume that Q_0 is distinct from P, and for each $n\in\mathbb{N}$, let $Q_n\in\mathcal{P}_n(M)_q$ be a q-mean of the empirical law P_n. Then*

$$\bar{d}_p(P,Q_n)-\bar{d}_p(P,Q_0)\leq\bar{d}_p(P_n,P)+O_{\mathbb{P}}(n^{-1/2}).$$

Proof. Let $K\colon M \to \mathcal{P}(M)$ be an optimal transport map from P to Q_0, i.e.,

$$Q_0 = \int_M K(x)P(\mathrm{d}x), \quad \bar{d}_p(P, Q_0) = \left(\int_M \int_M d(x,y)^p K(x, \mathrm{d}y)P(\mathrm{d}x)\right)^{1/p}.$$

Such a transport map can be obtained from an optimal coupling between P and Q_0 via disintegration. Then K is also a transport map between P_n and $\tilde{Q}_n$, where

$$\tilde{Q}_n \coloneqq \int_M K(x)P_n(\mathrm{d}x) \in \mathcal{P}_n(M)_q.$$

By the triangle inequality and the best-approximation property of the polymeans,

$$\bar{d}_p(P, Q_n) \le \bar{d}_p(P_n, Q_n) + \bar{d}_p(P_n, P) \le \bar{d}_p(P_n, \tilde{Q}_n) + \bar{d}_p(P_n, P)$$

$$\le \left(\int_M \int_M d(x,y)^p K(x, \mathrm{d}y)P_n(\mathrm{d}x)\right)^{1/p} + \bar{d}_p(P_n, P).$$

Rewriting the right-hand side using the defining properties of K leads to the estimate

$$\bar{d}_p(P, Q_n) \le \left(\bar{d}_p(P, Q_0)^p + \int_M \int_M d(x,y)^p K(x, \mathrm{d}y)(P_n - P)(\mathrm{d}x)\right)^{1/p} + \bar{d}_p(P_n, P).$$

By the central limit theorem, the random variables

$$n^{1/2} \int_M \int_M d(x,y)^p K(x, \mathrm{d}y)(P_n - P)(\mathrm{d}x)$$

converge in distribution to a normal random variable. As $d_p(P, Q_0) > 0$, this establishes the lemma. The central limit theorem may be applied thanks to the square-integrability condition

$$\int_M \left(\int_M d(x,y)^p K(x, \mathrm{d}y)\right)^2 P(\mathrm{d}x) \le \int_M \int_M d(x,y)^{2p} K(x, \mathrm{d}y)P(\mathrm{d}x)$$

$$\le \sum_{y \in \mathrm{supp}(Q_0)} \int_M d(x,y)^{2p} P(\mathrm{d}x) < \infty. \qquad \square$$

The bound in 7.3 involves the Wasserstein distance $\bar{d}_p(P_n, P)$ between a distribution P and the empirical distribution P_n of an n-sample, which is itself a random variable. On $M = \mathbb{R}^d$ it has been shown for distributions P with sufficiently many moments that $\|\bar{d}_p(P_n, P)\|_{L^p(\Omega)}$ is of the order $n^{-1/\max\{d, 2p\}}$, with an additional logarithmic factor if $d = 2p$ [24, Theorem 1]. This paper also gives references for improved rates under more stringent conditions on P. The case of non-flat M is largely open.

Using Lemma 7.3, the following theorem bounds the rate at which the empirical q-means Q_n converge to the population q-mean Q_0. Besides the distance $\bar{d}_p(P_n, P)$, it also involves a real number α, which quantifies the coercivity of the Wasserstein distance $\bar{d}_p(P, \cdot)$ near a minimizer Q_0 in the q-skeleton and depends on the subspace geometry of the q-skeleton within Wasserstein space.

Theorem 7.4 (Convergence rate). *Let $P \in \mathcal{P}^{2p}(M)$, let $Q_n \in \mathcal{P}_n(M)_q$ be a sequence of q-means of P_n converging in probability to a population q-mean $Q_0 \in \mathcal{P}(M)_q$, and assume for some $\alpha > 0$ and $c > 0$ that*

$$\bar{d}_p(P, Q) - \bar{d}_p(P, Q_0) \ge c\bar{d}_p(Q, Q_0)^\alpha$$

for all $Q \in (M^n/S_n)_q$ *near* Q_0. *Then*

$$\bar{d}_p(Q_n, Q_0) = O_{\mathbb{P}}(\bar{d}_p(P_n, P)^{1/\alpha}) + O_{\mathbb{P}}(n^{-1/(2\alpha p)}).$$

Proof. The error bound 7.3 together with the assumption on the distance function imply that

$$c\bar{d}_p(Q_n, Q_0)^\alpha \leq \bar{d}_p(P, Q_n) - \bar{d}_p(P, Q_0) \leq \bar{d}_p(P_n, P) + O_{\mathbb{P}}(n^{-1/(2p)}).$$

Taking the α-th root establishes the theorem. $\square$

It remains open if weighted q-means are asymptotically normal after a suitable rescaling. However, we will answer this question affirmatively for unweighted q-means, defined in 5.8. Note that these are strongly consistent thanks to the strong consistency 7.2 of weighted q-means.

Definition 7.5 (Asymptotic normality). Assume that M, and consequently also the regular stratum $(M^q/S_q)_{\text{reg}} = \mathcal{P}_q(M)_{\text{reg}}$, is a manifold. Fix a regular sample $R_0 \in \mathcal{P}_q(M)_{\text{reg}}$ and a symmetric bilinear form Σ on the tangent space at R_0. Then a sequence $R_1, R_2, \ldots$ of random elements in $\mathcal{P}_q(M)$ is called *asymptotically normal* with mean R_0 and covariance Σ if for some (equivalently, every) coordinate chart (U, u) around R_0, the sequence $\sqrt{n} \mathbb{1}_U(R_n) u(R_n)$ converges in law to a normal distribution $\mathcal{N}(0, u_*(\Sigma))$.

The chart independence in this definition is a consequence of the delta method [50, Theorem 3.1]. We then get the following asymptotic result.

Theorem 7.6 (Asymptotic normality). *Let M be a manifold with Riemannian path metric d and assume that conditions (1)–(2) in the proof below hold true. Then any sequence $R_1, R_2, \ldots$ of unweighted q-means of $P_1, P_2, \ldots$, which converges in probability to a unique unweighted population q-mean R_0, is asymptotically normal.*

Proof. As before, we use 4.7 to identify the unweighted q-means $R_0, R_1, R_2, \ldots \in \mathcal{P}_q(M)_{\text{reg}}$ with the corresponding q-samples $\bar{z}_0, \bar{z}_1, \bar{z}_2, \ldots \in (M^q/S_q)_{\text{reg}}$. In the notation of [21] and [29], and in line with Definition 5.8 of unweighted q-means, we define the Fréchet functional

$$\bar{\rho}: M \times (M^q/S_q)_{\text{reg}} \ni (x, \bar{z}) = (x, \pi(z)) \mapsto \min_{i \in \{1, \ldots, q\}} d(x, z_i)^p.$$

Then the unweighted q-means $\bar{z}_n$ minimize the functional

$$P_n\bar{\rho}: (M^q/S_q)_{\text{reg}} \ni \bar{z} \mapsto \int_M \bar{\rho}(x, \bar{z}) P_n(\mathrm{d}x),$$

and the unweighted population q-mean $\bar{z}_0$ minimizes the functional

$$P\bar{\rho}: (M^q/S_q)_{\text{reg}} \ni \bar{z} \mapsto \int_M \bar{\rho}(x, \bar{z}) P(\mathrm{d}x).$$

To verify the conditions of [21] we make the following assumptions:

(1) The following sets have zero probability under P:
$$\{\bar{z}_{0,1}, \ldots, \bar{z}_{0,q}\}, \quad \text{Cut}(\bar{z}_{0,1}) \cup \cdots \cup \text{Cut}(\bar{z}_{0,q}),$$
$$\{x \in M : \exists i \neq j \in \{1, \ldots, q\} : d(x, \bar{z}_{0,i}) = d(x, \bar{z}_{0,j}) = \rho(x, \bar{z}_0)\}.$$

(2) The function $P\bar{\rho}$ defined above has a non-degenerate Hessian at $\bar{z}_0$.

Note that the first assumption guarantees for P-almost every $x \in M$ the existence of the Riemannian gradient of the function $\bar\rho(x, \cdot)$ at $\bar z_0$. Indeed, the only points x where the gradient may fail to exist are the points $\bar z_{0,i}$, their cut loci $\mathrm{Cut}(\bar z_{0,i})$, and the locations which are closest to more than one $\bar z_{0,i}$. A further condition of [21] to be verified for all $x \in M$ is that the function $\bar\rho(x, \cdot)$ is uniformly continuous on bounded domains with respect to the metric $\bar d_p$ on M^q/S_q. This follows from the estimate

$$|\bar\rho(x, \pi(z')) - \bar\rho(x, \pi(z))| \leq \max_{i\in\{1,\ldots,q\}} \min_{j\in\{1,\ldots,q\}} d(z_i, z_j')^p \leq q\bar d_p(\pi(z), \pi(z'))^p.$$

Thus, we have verified the conditions of [21, Theorem 11], and it follows that the sequence $\bar z_n$ or equivalently R_n is asymptotically normal. $\qquad\square$

The asymptotic normality of unweighted q-means generalizes from independent to exchangeable observations $x_1, x_2, \ldots$ under certain conditions. Equivalently, as shown in 6.6, the observations can be seen as random elements in an infinite sample space.

Corollary 7.7 (Asymptotic normality, exchangeable observations). *Theorem 7.6 extends to exchangeable sequences of (not necessarily independent) observations $x_1, x_2, \ldots$, provided that condition (1) in the proof below is satisfied.*

Proof. By the infinite Hewitt–Savage theorem 6.4 and its Corollary 6.6, the exchangeable sequence $x_1, x_2, \ldots$ is i.i.d. conditionally on some sigma algebra $\mathcal{G}$. It follows from 7.6 that conditionally on $\mathcal{G}$, the sequence $R_1, R_2, \ldots$ is asymptotically normal with mean R_0 and covariance Σ, for some $\mathcal{G}$-measurable symmetric bilinear form Σ on the tangent space of $(M^q/S_q)_{\mathrm{reg}}$ at R_0. The covariance Σ can be computed explicitly as follows. Let $\bar\rho$ be defined as in the proof of 7.6, and recall that the gradient of the function $\bar\rho(x, \cdot)$ evaluated at R_0 exists for P-almost every $x \in M$. Therefore, for any $i \in \mathbb{N}_{>0}$, one may define the random variable X_i as the gradient of the random function $\bar\rho(x_i, \cdot)$ evaluated at R_0. Accordingly, X_i is a random variable with values in the tangent space of $(M^q/S_q)_{\mathrm{reg}}$ at R_0. Let $\bar H$ denote the Hessian of the function $P\bar\rho$ at R_0. Thanks to the non-degeneracy assumption in 7.6, $\bar H$ is an automorphism on the tangent space of $(M^q/S_q)_{\mathrm{reg}}$ at R_0, and we denote its inverse by $\bar H^{-1}$. Then the covariance Σ is given by [21, Theorem 11]

$$\Sigma = \frac{1}{4}\,\mathrm{Cov}[\bar H^{-1}(X_1) \otimes \bar H^{-1}(X_1)|\mathcal{G}].$$

To ensure that Σ is deterministic, we make the following assumption:

(1) $\qquad \mathbb{E}[X_1] = 0, \qquad \mathrm{Cov}(X_1, X_2) = 0, \qquad \mathrm{Cov}(X_1 \otimes X_1, X_2 \otimes X_2) = 0.$

Define $B = \mathbb{E}[X_1 \otimes X_1]$ and $C = \mathrm{Cov}[X_1 \otimes X_1]$. Then the relations

$$0 = \mathbb{E}[X_1 \otimes X_2] = \mathbb{E}[\mathbb{E}[X_1 \otimes X_2|\mathcal{G}]] = \mathbb{E}[\mathbb{E}[X_1|\mathcal{G}]^{\otimes 2}],$$
$$C = \mathbb{E}[(X_1 \otimes X_1 - B) \otimes (X_2 \otimes X_2 - B)] = \mathbb{E}[\mathbb{E}[X_1 \otimes X_1 - B|\mathcal{G}]^{\otimes 2}],$$

show that (1) is equivalent to

$$\mathbb{E}[X_1|\mathcal{G}] = 0, \qquad \mathbb{E}[X_1 \otimes X_1|\mathcal{G}] = B.$$

Therefore, $\Sigma = (\bar H^{-1} \otimes \bar H^{-1})(B)$ is deterministic, as claimed. As R_0 and Σ are deterministic, the sequence $R_1, R_2, \ldots$ is not only conditionally but also unconditionally asymptotically normal. See [14, Theorem 9.2.1] for further details in the Euclidean case. $\qquad\square$

APPENDIX A.

Definition A.1 (Path metrics [26, 10]). In any metric space (M, d), for any real numbers $s \leq t$, the *length* of a continuous curve $c \colon [s, t] \to M$ is defined as

$$\ell(c) = \sup_{\substack{n \in \mathbb{N} \\ s = u_0 \leq \cdots \leq u_n = t}} \sum_{i=0}^{n-1} d\big(c(u_i), c(u_{i+1})\big) \in [0, \infty].$$

The curve is said to have *constant speed* $v \in \mathbb{R}_{\geq 0}$ if $\ell(c|_{[u_1, v_1]}) = v|u_1 - u_2|$ for all $s \leq u_1 < u_2 \leq t$. The metric space is called a *path-metric space* if the distance between any pair of points equals the infimum of the lengths of continuous curves joining the points. A *minimizing geodesic* in a metric space (M, d) is a continuous curve whose length equals the distance between its end points. A *geodesic* is a curve whose restriction to any sufficiently small subinterval is a minimizing geodesic.

Theorem A.2 (Hopf–Rinov theorem [26, 1.9]). *If (M, d) is a connected complete locally compact path-metric space then:*

    (1) Closed balls are compact, or, equivalently, each bounded closed subset of M is compact.

    (2) Any two points can be joined by a minimizing geodesic.

Theorem A.3 (Characterization of path metrics [26, Theorem 1.8]). *The following properties of a metric space (M, d) are equivalent:*

    (1) For any points $x, y \in M$ and $r > 1/2$ there exists a point $z \in M$ such that

$$\max\{d(x, z), d(z, y)\} \leq r d(x, y).$$

    (2) For all $x, y \in M$ and $r_1, r_2 > 0$ with $r_1 + r_2 \leq d(x, y)$ we have

$$d(B(x, r_1), B(y, r_2)) := \inf\{d(x', y') : d(x', x) \leq r_1, d(y', y) \leq r_2\}$$
$$\leq d(x, y) - r_1 - r_2.$$

Every path-metric space has these properties. Conversely, a complete metric space with property (1) *or* (2) *is a path-metric space.*

Definition A.4 (Lagrangian actions). Following [51, Definition 7.11], a *Lagrangian energy–action pair* (E, A) on a topological space M is a family of energy functionals $E^{s,t} \colon M \times M \to \mathbb{R}$ and action functionals $A^{s,t} \colon C([s, t], M) \to \mathbb{R}$, indexed by real numbers $s \leq t$, which satisfies the following three properties:

    (1) for all $r \leq s \leq t$, $A^{r,s} + A^{s,t} = A^{r,t}$,

    (2) for all $s \leq t$ and $x, y \in M$,

$$E^{s,t}(x, y) = \inf_{\substack{c \in C([s,t], M) \\ c(s) = x, c(t) = y}} A^{s,t}(c).$$

    (3) for all $s \leq t$ and $c \in C([s, t], M)$,

$$A^{s,t}(c) = \sup_{\substack{n \in \mathbb{N} \\ s = u_0 \leq \cdots \leq u_n = t}} \sum_{i=0}^{n-1} E^{u_i, u_{i+1}}\big(c(u_i), c(u_{i+1})\big).$$

Curves which assume the minimum in (2) are called *minimizing curves* for (E, A).

Examples of Lagrangian energy–action pairs on path-metric spaces (M, d) are (d, ℓ) as well as the functionals described in the following lemma, which are related to the Riemannian or Finsler energy.

Lemma A.5 (Lagrangian actions)**.** *For any path-metric space (M, d) and $p \in (1, \infty)$, the following defines a Lagrangian energy–action pair (E, A):*

$$E^{s,t}(x, y) = \frac{d(x, y)^p}{|s - t|^{p-1}}, \qquad A^{s,t}(c) = \sup_{\substack{n \in \mathbb{N} \\ s = u_0 \leq \cdots \leq u_n = t}} \sum_{i=0}^{n-1} \frac{d(c(u_i), c(u_{i+1}))^p}{|u_i - u_{i+1}|^{p-1}}.$$

Minimizing curves for (E, A) are exactly constant-speed minimizing geodesics.

Proof. Properties (1) and (3) of Lagrangian actions hold by definition. Property (2) can be verified as follows: as (M, d) is a path-metric space, the definition of the energy implies for any real numbers $s \leq t$ and points $x, y \in M$ that

$$E^{s,t}(x, y) = \inf_{\substack{c \in C([s,t], M) \\ c(s) = x, c(t) = y}} \sup_{\substack{n \in \mathbb{N} \\ s = u_0 \leq \cdots \leq u_n = t}} \frac{\left(\sum_{i=0}^{n-1} d(c(u_i), c(u_{i+1})) \right)^p}{|s - t|^{p-1}}.$$

Estimating the right-hand side using Hölder's inequality yields

$$E^{s,t}(x, y) \leq \inf_{\substack{c \in C([s,t], M) \\ c(s) = x, c(t) = y}} \sup_{\substack{n \in \mathbb{N} \\ s = u_0 \leq \cdots \leq u_n = t}} \sum_{i=0}^{n-1} \frac{d(c(u_i), c(u_{i+1}))^p}{|s - t|^{p-1}} = \inf_{\substack{c \in C([s,t], M) \\ c(s) = x, c(t) = y}} A^{s,t}(c).$$

For constant-speed curves, Hölder's inequality is an equality. Moreover, any continuous curve can be reparameterized to constant speed. Therefore, the preceding inequality is actually an equality. This shows (2).

The statement about minimizing curves hinges on the following Hölder inequality: for all $u \leq v \leq w$ in the domain of a continuous curve $c \colon [s, t] \to M$,

$$d(c(u), c(v)) + d(c(v), c(w))$$
$$= \frac{d(c(u), c(v))}{|u - v|^{(p-1)/p}} |u - v|^{(p-1)/p} + \frac{d(c(v), c(w))}{|v - w|^{(p-1)/p}} |v - w|^{(p-1)/p}$$
$$\leq \left(E^{u,v}(c(u), c(v)) + E^{v,w}(c(v), c(w)) \right)^{1/p} |u - w|,$$

with equality if and only if the vector $\left(d(c(u), c(v)), d(c(v), c(w)) \right)$ in $\mathbb{R}^2$ is parallel to the vector $(v - u, w - v)$.

Let $c \colon [s, t] \to M$ be a continuous curve. Then c is a minimizing geodesic with constant speed if and only if it satisfies for all $u \leq v \leq w$ in $[s, t]$ that

$$d(c(u), c(v)) + d(c(v), c(w)) = d(c(u), c(w)),$$
$$d(c(u), c(v)) = d(c(s), c(t)) |u - v|.$$

Equivalently, by the above Hölder inequality, it holds for all $u \leq v \leq w$ in $[s, t]$ that

$$E^{u,v}(c(u), c(v)) + E^{v,w}(c(v), c(w)) = E^{u,w}(c(u), c(w)),$$

which means that c minimizes the energy–action pair (E, A). $\qquad\square$

Lemma A.6 (Atomic distributions)**.** *Let M be a metric space or, more generally, a first-countable space. Then the set $\mathcal{P}_n(M)$ coincides with the set of $\{0, 1/n, \ldots, 1\}$-valued probability distributions on M.*

Proof. Clearly, every distribution in $\mathcal{P}_n(M)$ takes values in $\{0, 1/n, \ldots, 1\}$. Conversely, assume that P is a $\{0, 1/n, \ldots, 1\}$-valued probability distribution. Let $x \in M$, and let $(U_i)_{i \in \mathbb{N}}$ be a decreasing basis of open neighborhoods of x. If $\min_{i \in \mathbb{N}} P(U_i)$ vanishes, then it vanishes for sufficiently large i, and consequently

x does not belong to the support of P. Otherwise, $P(\{x\}) = \min_{i \in \mathbb{N}} P(U_i) \geq \frac{1}{n}$, which can be the case for only finitely many $x \in M$. Therefore, the support of P is a finite set. It follows that P is a weighted sum of Dirac measures at distinct points in M. Necessarily, the weights are multiples of $1/n$. $\qquad\square$

Acknowledgements

References

[1] B. Afsari. Riemannian l^p center of mass: Existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(02):655–673, 2 2011.

[2] D. Alekseevsky, A. Kriegl, M. Losik, and P. W. Michor. The Riemannian geometry of orbit spaces—the metric, geodesics, and integrable systems. *Publ. Math. Debrecen*, 62(3-4):247–276, 2003. Dedicated to Professor Lajos Tamássy on the occasion of his 80th birthday.

[3] S. Alexander, V. Kapovitch, and A. Petrunin. *An invitation to Alexandrov geometry: CAT(0) spaces*. Springer Briefs in Mathematics. Springer, Cham, 2019.

[4] M. Arnaudon and L. Miclo. Means in complete manifolds: uniqueness and approximation. *ESAIM: Probability and Statistics*, 18:185–206, 2014.

[5] R. Bhattacharya and V. Patrangenaru. Nonparametric estimation of location and dispersion on Riemannian manifolds. *Journal of Statistical Planning and Inference*, 108(1-2):23–35, 11 2002.

[6] R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds. *The Annals of Statistics*, 31(1):1–29, 2 2003.

[7] R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds—II. *The Annals of Statistics*, 33(3):1225–1259, 6 2005.

[8] L. J. Billera, S. P. Holmes, and K. Vogtmann. Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.*, 27(4):733–767, 2001.

[9] J. R. Blum, H. Chernoff, M. Rosenblatt, and H. Teicher. Central Limit Theorems for Interchangeable Processes. *Canadian Journal of Mathematics*, 10:222–229, 1958/ed.

[10] D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[11] P. Buser and H. Karcher. *Gromov's almost flat manifolds*. Number 81 in Astérisque. Société mathématique de France, 1981.

[12] P. Cardaliaguet. Notes on mean field games (from P.-L. Lions' lectures at collège de france). Available on the website of Collège de France http://www.college-de-france.fr.

[13] H. Chernoff and H. Teicher. A Central Limit Theorem for Sums of Interchangeable Random Variables. *Annals of Mathematical Statistics*, 29(1):118–130, 3 1958.

[14] Y. S. Chow and H. Teicher. *Probability Theory: Independence, Interchangeability, Martingales*. Springer Texts in Statistics. Springer, New York, third edition, 1997.

[15] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimension*. Cambridge University Press, 2 edition, 2014.

[16] B. de Finetti. La Prévision: Ses Lois Logiques, Ses Sources Subjectives. *Annales de l'Institut Henri Poincaré*, 17:1–68, 1937.

[17] C. Dellacherie. Ensembles analytiques: théorèmes de séparation et applications. In *Séminaire de Probabilités IX Université de Strasbourg*, volume 465 of *Lecture Notes in Mathematics*, pages 336–372. Springer, 1975.

[18] P. Diaconis and D. Freedman. Finite Exchangeable Sequences. *The Annals of Probability*, 8(4):745–764, 1980. Publisher: Institute of Mathematical Statistics.

[19] I. Dryden and K. Mardia. *Statistical Shape Analysis*. John Wiley & Sons, 1998.

[20] B. Eltzner. Geometrical Smeariness - A new Phenomenon of Fréchet Means. *arXiv:1908.04233*, 8 2019.

[21] B. Eltzner and S. F. Huckemann. A smeary central limit theorem for manifolds with application to high-dimensional spheres. *Annals of Statistics*, 47(6):3360–3381, 12 2019.

[22] A. Figalli, T. O. Gallouët, and L. Rifford. On the Convexity of Injectivity Domains on Nonfocal Manifolds. *SIAM Journal on Mathematical Analysis*, 47(2):969–1000, 1 2015.

[23] B. D. Finetti. *Funzione caratteristica di un fenomeno aleatorio*. Società anonima tipografica, 1930.

[24] N. Fournier and A. Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3-4):707–738, 2015.

[25] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancie. *Ann. Inst. H. Poincaré*, 10:215–310, 1948.

[26] M. Gromov. *Metric structures for Riemannian and non-Riemannian spaces, Based on the 1981 French original, With appendices by M. Katz, P. Pansu and S. Semmes. Translated from the French by Sean Michael Bates*, volume 152 of *Progress in Mathematics*. Birkhäuser, Boston, 1999.

[27] E. Hewitt and L. J. Savage. Symmetric measures on Cartesian products. *Transactions of the American Mathematical Society*, 80(2):470–501, 1955.

[28] T. Hotz, S. Huckemann, H. Le, J. S. Marron, J. C. Mattingly, E. Miller, J. Nolen, M. Owen, V. Patrangenaru, and S. Skwerer. Sticky central limit theorems on open books. *The Annals of Applied Probability*, 23(6):2238–2258, 12 2013.

[29] S. F. Huckemann. Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *Ann. Statist*, page 2011, 2011.

[30] S. F. Huckemann. On the meaning of mean shape: manifold stability, locus and the two sample test. *Annals of the Institute of Statistical Mathematics*, 64(6):1227–1259, 12 2012.

[31] S. Hundrieser, B. Eltzner, and S. F. Huckemann. Finite Sample Smeariness of Fréchet Means and Application to Climate. *arXiv:2005.02321*, 5 2020.

[32] A. K. Jain. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8):651–666, 6 2010.

[33] P. E. Jupp and K. V. Mardia. A unified view of the theory of directional statistics, 1975-1988. *International Statistical Review*, 57(3):261–294, 1989.

[34] H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541, 1977.

[35] D. G. Kendall. Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces. *Bull. London Math. Soc.*, 16(2):81–121, 3 1984.

[36] D. G. Kendall, D. Barden, T. K. Carne, and H. Le, editors. *Shape & Shape Theory*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA, 10 1999.

[37] W. S. Kendall and H. Le. Limit theorems for empirical Fréchet means of independent and non-identically distributed manifold-valued random variables. *Brazilian Journal of Probability and Statistics*, 25(3):323–352, 11 2011.

[38] M. Klass and H. Teicher. The Central Limit Theorem for Exchangeable Random Variables Without Moments. *Annals of Probability*, 15(1):138–153, 1 1987.

[39] K. Kuwae and T. Shioya. Variational convergence over metric spaces. *Transactions of the American Mathematical Society*, 360(1):35–75, 2008.

[40] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. University of California Press, 1967. ISSN: 0097-0433.

[41] Mantegazza and Mennucci. Hamilton—Jacobi Equations and Distance Functions on Riemannian Manifolds. *Applied Mathematics & Optimization*, 47(1):1–25, 12 2002.

[42] J. S. Marron and A. M. Alonso. Overview of object oriented data analysis. *Biometrical Journal*, 56(5):732–753, 9 2014.

[43] P. W. Michor. *Topics in differential geometry*, volume 93 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2008.

[44] V. M. Panaretos and Y. Zemel. *An Invitation to Statistics in Wasserstein Space*. Springer Nature, 2020.

[45] X. Pennec. Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. *J. Math. Imaging Vis.*, 25(1):127–154, 2006.

[46] X. Pennec. Curvature effects on the empirical mean in Riemannian and affine Manifolds: A non-asymptotic high concentration expansion in the small-sample regime. *arXiv:1906.07418*, 6 2019.

[47] X. Pennec, S. Sommer, and P. T. Fletcher. *Riemannian Geometric Statistics in Medical Image Analysis*. Elsevier, 2020.

[48] K.-T. Sturm. Probability measures on metric spaces of nonpositive curvature. In P. Auscher, T. Coulhon, and A. Grigor'yan, editors, *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces, Paris, France*, volume 338, pages 357–390. American Mathematical Society, Providence, Rhode Island,

2003.

[49] P. K. Turaga and A. Srivastava, editors. *Riemannian Computing in Computer Vision*. Springer International Publishing, Cham, 2016.

[50] A. W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

[51] C. Villani. *Optimal transport: old and new*. Springer, 2008.

[52] Wikipedia contributors. Orbifolds — Wikipedia, the free encyclopedia, 2020. [Online; accessed 13-July-2020].

[53] S. Willard. *General topology*. Courier Corporation, 2004.

[54] J. Xu and K. Lange. Power k-Means Clustering. In K. Chaudhuri and R. Salakhutdinov, editors, *International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6921–6931, Long Beach, California, USA, 6 2019. PMLR.

[55] H. Ziezold. On Expected Figures and a Strong Law of Large Numbers for Random Elements in Quasi-Metric Spaces. In J. Kožešnik, editor, *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the 1974 European Meeting of Statisticians: Held at Prague, from August 18 to 23, 1974*, pages 591–602. Springer Netherlands, Dordrecht, 1977.

Nanyang Technological University, Singapore
Email address: philipp.harms@ntu.edu.sg

Faculty of Mathematics, University of Vienna, Austria
Email address: peter.michor@univie.ac.at

Universtité Côte d'Azur and Inria, Sophia Antipolis, France
Email address: xavier.pennec@inria.fr

Department of Computer Science, University of Copenhagen, Denmark
Email address: sommer@di.ku.dk