Electronic Journal of Probability

Electron. J. Probab. **20** (2015), no. 78, 1–34.
ISSN: 1083-6489 DOI: 10.1214/EJP.v20-3887

Sticky central limit theorems at isolated hyperbolic planar singularities

Stephan Huckemann* Jonathan C. Mattingly† Ezra Miller‡
James Nolen§

Abstract

We derive the limiting distribution of the barycenter b_n of an i.i.d. sample of n random points on a planar cone with angular spread larger than 2π. There are three mutually exclusive possibilities: (i) (*fully sticky* case) after a finite random time the barycenter is almost surely at the origin; (ii) (*partly sticky* case) the limiting distribution of $\sqrt{n}b_n$ comprises a point mass at the origin, an open sector of a Gaussian, and the projection of a Gaussian to the sector's bounding rays; or (iii) (*nonsticky* case) the barycenter stays away from the origin and the renormalized fluctuations have a fully supported limit distribution—usually Gaussian but not always. We conclude with an alternative, topological definition of stickiness that generalizes readily to measures on general metric spaces.

Keywords: Fréchet mean; central limit theorem; law of large numbers; stratified space; nonpositive curvature.
AMS MSC 2010: 60B99; 60F05.
Submitted to EJP on October 26, 2014, final version accepted on July 18, 2015.
Supersedes arXiv:1410.6879v2.

Introduction

It has recently been observed that large samples from well-behaved probability distributions on metric spaces that are not smooth Riemannian manifolds are sometimes constrained to lie in subsets of low dimension, and that central limit theorems in such cases consequently behave non-classically, with components of limiting distributions supported on thin subsets of the sample space [14, 2, 4]. Our results here continue this line of investigation with the first complete description of "sticky" behavior at a singularity of codimension 2.

*Felix Bernstein Institute for Mathematical Statistics in the Biosciences, University of Göttingen, Goldschmidtstrasse 7, 37077 Göttingen, Germany. E-mail: huckeman@math.uni-goettingen.de
†Mathematics Department and Department of Statistical Science, Duke University, Durham, NC 27708, USA. E-mail: jonm@math.duke.edu
‡Mathematics Department, Duke University, Durham, NC 27708, USA. www.math.duke.edu/~ezra
§Mathematics Department, Duke University, Durham, NC 27708, USA. E-mail: nolen@math.duke.edu

More precisely, we prove laws of large numbers (Theorem 1.12; see Section 5 for proofs and more details) as well as central limit theorems (Section 1.4; proofs in Section 6) for Fréchet means of probability distributions (Definitions 1.6 and 1.7) on metric spaces possessing the simplest geometric singularities in codimension 2. The spaces are surfaces homeomorphic to the Euclidean plane $\mathbb{R}^2$ and metrically flat locally everywhere except at a single *cone point*, where the angle sum—the length of a circle of radius 1—exceeds 2π (see Section 1.1 for precise definitions). Thus the surface is planar, the singularity is isolated, and its geometry is hyperbolic, in the sense of negatively curved; hence the title of this paper.

The asymptotic behavior splits into three cases, called *fully sticky*, *partly sticky*, and *nonsticky* (Definition 1.8 and Proposition 1.10), according to whether the mean lies stably at the singularity (Theorem 1.13), unstably at the singularity (Theorem 1.14), or away from the singularity (Theorem 1.15), respectively. Specific examples illustrating the sticky phenomena, including subtle non-local effects of the singular negative curvature when the mean lies in the smooth stratum (Example 2.5), occupy Section 2. In contrast to the usual strong law asserting almost-sure convergence of empirical means to a population mean, sticky strong laws deal also with the limiting behavior of supports of the laws of empirical means. In the sticky case this support degenerates in some specified sense already in finite random time (Theorem 1.12). Our sticky central limit theorems assert that the limiting distributions are mixtures of parts of Gaussians and collapsed (i.e., projected) parts of Gaussians. Even in the nonsticky case, the limiting laws can fail to be Gaussian (Example 2.5), which may come as a surprise: although the space is locally Euclidean near the mean, the conclusion of Theorem 2.3 of Bhattacharya and Patrangenaru (2005) can nonetheless not be valid.

Concluding our analysis is a topological characterization of stickiness for measures on isolated planar hyperbolic singularities (Theorem 7.6), as opposed to the algebraic one in terms of moments (Definitions 1.7 and 1.8) used for the rest of the exposition. Thinking topologically leads to a very general notion of stickiness (Definition 7.10), which we include with an eye toward sampling from more general geometrically or topologically singular spaces. We have in mind stratified spaces (see [9] or [18]), suitably metrized, noting that (for example) every real semialgebraic variety admits a canonical Whitney stratification with finitely many semialgebraic strata [8, Section 2.7].

A motivating example of such stratified sample spaces comes from evolutionary biology, where the objects are phylogenetic trees. The space of such objects is CAT(0) (or equivalently, globally nonpositively curved) [7] and therefore has many desirable features where geometric probability is concerned [20]. [2] treat the space $\mathcal{T}_4$ of phylogenetic trees with four leaves.[1] The singularity of $\mathcal{T}_4$ at its center cone point is a (non-disjoint) union of a certain number of copies of an isolated planar singularity with angle sum $5\pi/2 > 2\pi$. Therefore some features of our results are present in the central limit theorem at the cone point of $\mathcal{T}_4$ [2, Theorem 5.2], which identifies the support of the limit measure in each right-angled orthant as a cone over an interval. However, the limit measure exhibits additionally non-classical behavior at the boundary of its support, where mass concentrates on the edges and even more on the origin. The simpler nature of an isolated planar singularity, which lacks the global combinatorial complexity of tree space, allows us to discover these boundary components and characterize them by identifying the limit measure as the convex projection of a Gaussian distribution (Theorem 1.14).

While the strong law of large numbers on quasi-metric spaces by [22] and on manifolds by [5] requires the existence of a population mean, meaning square-integrability of

[1] As this draft was completed, the preprint [3] was posted. The results there are proved for arbitrary numbers of leaves but restrict to singularities in codimension 0 and 1.

the underlying law, for fully sticky strong laws the existence of a population mean is not necessary: no square-integrability is required. Curiously, a (fully) sticky central limit theorem can consequently hold in the absence of any population mean at all (Example 2.4). That said, the greater challenge consists in the partly sticky case; as is the case for the multivariate Central Limit Theorem, as well as for that on manifolds by [10, 11, 6] or on certain stratified spaces by [16], square-integrability is still required.

In addition to the theoretical interest in the asymptotic behavior of means on stratified spaces, another driving motivation comes from the need to accordingly devise inferential statistical methods for applications based on the asymptotic behavior of Fréchet sample means and similar mean quantities, e.g. [12, 1, 17, 19]. This type of development is exemplified, in the form of confidence intervals on the spider, by [15].

Many parts of this paper are rather technical—though elementary—and require the buildup of notation in Sections 3 and 4, as we fold the isolated planar singularity onto $\mathbb{R}^2$. The behavior of first moments under folding and rotation is essential to understand the limiting location of barycenters on the singular space $\mathcal{K}$ (which we call the *kale*), and their limiting laws on $\mathcal{K}$ as well as on $\mathbb{R}^2$, which are described by certain sectors where a first folded moment is non-negative. A list of notion is given in Section 8.

1 Basic definitions and principal results

1.1 Isolated hyperbolic planar singularities

The *kale* is the space

$$\mathcal{K} = \big((0,\infty) \times (\mathbb{R}/\alpha\mathbb{Z})\big) \cup \{\mathbf{0}\} \tag{1.1}$$

where $\alpha > 2\pi$ is the *angle sum* at the isolated point $\mathbf{0}$, called the *origin*, the sole point at which the metric is not locally Euclidean. Points are specified by polar coordinates $p = (r,\theta) \in \mathcal{K}$ for a *radius* $r > 0$ and *angle* $\theta \in \mathbb{R}/\alpha\mathbb{Z}$, and the origin is often expressed as $\mathbf{0} = (0,0)$ or $\mathbf{0} = (0,\theta)$ for any $\theta \in \mathbb{R}/\alpha\mathbb{Z}$; that is, the origin is viewed as lying at zero radius along every ray emanating from it. The circle $\mathbb{R}/\alpha\mathbb{Z}$, a group under addition, has the natural uniform metric defined by

$$|\theta' - \theta|_\alpha = \min_{n\in\mathbb{Z}} |n\alpha + \theta' - \theta|.$$

Note that $|\theta' - \theta|_\alpha \le \alpha/2$. Denote by $d(p_1,p_2)$ the metric on $\mathcal{K}$ defined by

$$d\big((r_1,\theta_1),(r_2,\theta_2)\big)^2 = \begin{cases} (r_1+r_2)^2 & \text{if } |\theta_1-\theta_2|_\alpha \ge \pi, \\ r_1^2 + r_2^2 - 2r_1r_2\cos\big(|\theta_1-\theta_2|_\alpha\big) & \text{if } |\theta_1-\theta_2|_\alpha \le \pi. \end{cases}$$

When one of the points is the origin—so one of the radii vanishes—both cases apply, and in that situation the distance equals the other radius. Geometrically, $\mathcal{K}$ is the metric cone over a circle of length α placed at distance 1 from the cone point $\mathbf{0}$.

If we allowed $\alpha = 2\pi$, then this construction would yield $\mathcal{K} = \mathbb{R}^2$ with the Euclidean metric. If we allowed $\alpha < 2\pi$, then this construction would be a right circular ("ice cream") cone with angle sum α at the apex. The cases where the angle sum α is bigger than, equal to, or smaller than 2π correspond to the curvature at the origin being negative, flat, or positive, respectively. The name "kale" derives from the negative curvature of that particular leafy vegetable.

Definition 1.1. *From now on, write $|\theta' - \theta|$ for $|\theta' - \theta|_\alpha$, the role of α being understood. When $\big|\theta' - \theta\big| \le \pi$, we identify $\theta' - \theta$ with a number in the closed interval $[-\pi,\pi]$. Specifically, there is a unique integer n such that $\theta' - \theta + n\alpha \in [-\pi,\pi]$, and in this case we set*

$$\theta' - \theta = \theta' - \theta + n\alpha \in [-\pi,\pi].$$

Definition 1.1 implies that when $|\theta' - \theta| \leq \pi$, the intervals of length $|\theta' - \theta|$ with endpoints θ and θ', closed or open at either end, are all well defined in $\mathbb{R}/\alpha\mathbb{Z}$. In fact, even the interval $[\theta - \pi, \theta + \pi] \subset \mathbb{R}/\alpha\mathbb{Z}$ is well defined for all $\theta \in \mathbb{R}/\alpha\mathbb{Z}$, because $\alpha > 2\pi$. If $|\theta - \theta'| \leq \pi$, then the intervals $[\theta, \theta'] = [\theta', \theta]$ coincide as subsets of $\mathbb{R}/\alpha\mathbb{Z}$; it matters not whether $\theta - \theta' < 0$ or $\theta - \theta' > 0$.

Definition 1.2. *If $I \subset \mathbb{R}/\alpha\mathbb{Z}$ is any interval of angles, define the* sector

$$C_I = \{(r, \theta) \in \mathcal{K} \mid r \geq 0 \text{ and } \theta \in I\}$$

that is the cone over I from the origin. (If I is closed, then C_I is a closed subset of $\mathcal{K}$.)

Definition 1.3. *For a fixed angle $\theta \in \mathbb{R}/\alpha\mathbb{Z}$, the* folding map *$F_\theta \colon \mathcal{K} \to \mathbb{R}^2$ is determined by*

$$F_\theta(r', \theta') = \begin{cases} 0 & \text{if } r' = 0 \\ (r', \theta' - \theta) & \text{if } r' > 0 \text{ and } |\theta' - \theta| \leq \pi \\ (r', \pi) & \text{if } r' > 0 \text{ and } |\theta' - \theta| \geq \pi. \end{cases}$$

Here we are using polar coordinates for both $\mathcal{K}$ and $\mathbb{R}^2$; later we will sometimes use cartesian coordinates for the image of F_θ. Observe that when $|\theta' - \theta| = \pi$ the second and third cases agree. A simple geometric description of the folding map is given in terms of light and shadow as follows, cf. also Figure 1.

Definition 1.4. *The open set*

$$\mathcal{I}_\theta = \big\{(r, \theta') \in \mathcal{K} \mid r > 0 \text{ and } |\theta' - \theta| > \pi\big\} \subset \mathcal{K}$$

is the part of $\mathcal{K}$ invisible *from the angle θ. The complement $\mathcal{K} \setminus \mathcal{I}_\theta$ is the part* visible from *θ. The complement $\mathcal{K} \setminus \overline{\mathcal{I}_\theta}$ of the closure of the invisible part is* fully visible*, and the set $\overline{\mathcal{I}_\theta} \setminus \mathcal{I}_\theta$ of boundary points outside of $\mathcal{I}_\theta$ is* partly visible*. The* shadow *of any set $A \subseteq \mathbb{R}/\alpha\mathbb{Z}$ is*

$$\mathcal{I}_A = \bigcup_{\theta \in A} \mathcal{I}_\theta.$$

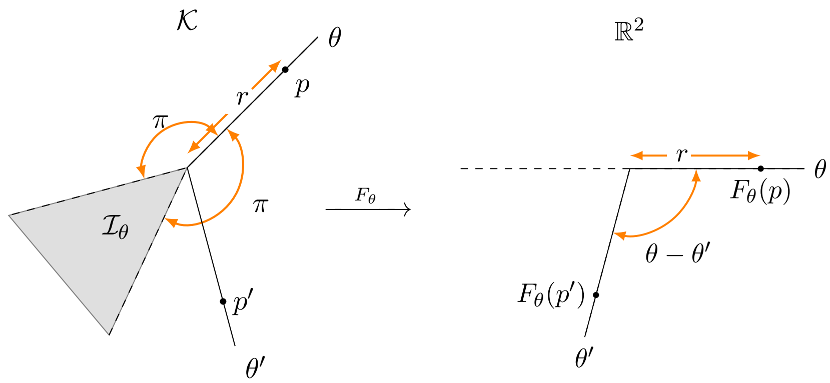

Figure 1: Fix points $p \neq 0 \neq p'$ with angles θ and θ' on the kale $\mathcal{K}$. Left: The shadow $\mathcal{I}_\theta$ of p is the interior of the sector of points whose shortest paths to p pass through the origin. In other words, as seen from p, the origin casts the shadow $\mathcal{I}_\theta$. All these points are *invisible* from p. (For future reference, with notation as in (1.18) and Lemma 4.3, including the upper dashed line gives $\mathcal{I}_\theta^+$ and including the lower dashed line gives $\mathcal{I}_\theta^-$.) Right: Under the folding map F_θ centered at angle θ, the shadow collapses to the negative horizontal axis.

The terminology referring to (in)visibility and shadow is motivated as follows. Imagine placing a light source at a point $p = (r, \theta)$. If rays of light (geodesics) in $\mathcal{K}$ are obstructed by the origin, then $\mathcal{I}_\theta$ is the set of points in the shadow cast by the origin. Alternatively, imagine light emanating from sources within $\mathcal{I}_\theta$: an observer at (r, θ) is not able to resolve the image, since all light rays arriving at the observer have merged at the origin.

Remark 1.5. The folding map F_θ is the unique continuous map $\mathcal{K} \to \mathbb{R}^2$ that preserves all distances from points on the ray at angle θ to other points in $\mathcal{K}$; c.f. Lemma 3.2. In particular, it preserves radius from the origin. The folding map F_θ collapses the part of $\mathcal{K}$ invisible from θ to the negative horizontal axis of $\mathbb{R}^2$ and takes the fully visible part of $\mathcal{K}$ bijectively to the complement of the negative horizontal axis.

The folding map F_θ is the "logarithm map" from $\mathcal{K}$ to the tangent space at any point with positive radius along the ray at angle θ. In smooth manifolds, log maps are right inverse to exponential maps, the latter being globally defined on the tangent space at a point p, while the former is only defined in a neighborhood of p. Here, singularity of the metric at $\mathbf{0} \in \mathcal{K}$ prevents exp from being well defined, whereas uniqueness of geodesics in $\mathcal{K}$ (that is, the absence of a cut locus) makes the log map globally defined on $\mathcal{K}$.

1.2 Barycenters and folded first moments

Let μ be a Borel probability measure on $\mathcal{K}$. Our main results concern statistics of random points drawn independently from the measure μ on $\mathcal{K}$. We assume throughout that μ satisfies the integrability condition

$$\bar{r} := \int_{\mathcal{K}} d(\mathbf{0}, p)\, d\mu(p) < \infty. \tag{1.2}$$

Because $\mathcal{K}$ is not a linear space, the mean of a probability distribution on $\mathcal{K}$ cannot be defined using addition, as it can be in $\mathbb{R}^2$. Instead, we use the notion of barycenter of a distribution μ. If the second moment condition (square-integrability)

$$\int_{\mathcal{K}} d(\mathbf{0}, q)^2\, d\mu(q) < \infty \tag{1.3}$$

holds, then the function $\Gamma\colon \mathcal{K} \to \mathbb{R}$ defined by

$$\Gamma(p) = \frac{1}{2} \int_{\mathcal{K}} d(p, q)^2\, d\mu(q) \tag{1.4}$$

is finite for all $p \in \mathcal{K}$, and it has a unique minimizer (proved later, at Corollary 4.13). This leads to the following definition.

Definition 1.6. *Under the second moment condition (1.3), the unique minimizer of Γ is the* barycenter *of μ, denoted by $\bar{b}$.*

It is possible to extend this definition in a consistent way to the setting where only the integrability condition (1.2) holds for μ rather than the stronger square-integrability condition (1.3); see Definition 1.11. For now, we only say enough to state this generalization of Definition 1.6, postponing the full discussion to Section 4.

Under the folding map $F_\theta : \mathcal{K} \to \mathbb{R}^2$, the measure μ pushes forward to a probability measure $\tilde{\mu}_\theta = \mu \circ F_\theta^{-1}$ on $\mathbb{R}^2$. The family of measures $\{\tilde{\mu}_\theta\}_{\theta \in \mathbb{R}/\alpha\mathbb{Z}}$ on $\mathbb{R}^2$ allows us to deduce properties of the measure μ on $\mathcal{K}$. For points $z \in \mathbb{R}^2$, we typically use cartesian coordinates $z = (z_1, z_2)$; the context should prevent any confusion with the radial representation (r, θ) of points in $\mathcal{K}$. Back in $\mathbb{R}^2$, denote by $e_1 = (1, 0)$ and $e_2 = (0, 1)$ the standard basis vectors, and by "$\cdot$" the standard inner product. The mean of $\tilde{\mu}_\theta$ in $\mathbb{R}^2$ can be defined in the usual way, as follows.

Definition 1.7. *For $\theta \in \mathbb{R}/\alpha\mathbb{Z}$, the* first moment *of μ folded about θ (or equivalently, the* mean *of μ folded about θ) is*

$$m_\theta = \int_{\mathbb{R}^2} z \, d\tilde{\mu}_\theta(z) = \int_{\mathcal{K}} F_\theta(p) \, d\mu(p) = e_1 m_{\theta,1} + e_2 m_{\theta,2}$$

where

$$m_{\theta,i} = e_i \cdot m_\theta = \int_{\mathcal{K}} e_i \cdot F_\theta(p) \, d\mu(p) \quad \textit{for } i = 1, 2.$$

The integrability condition (1.2) implies that the first moment m_θ is finite and that $\theta \mapsto m_\theta$ is continuous.

Definition 1.8. *Fix a probability distribution μ on $\mathcal{K}$ and let $K \subset \mathbb{R}/\alpha\mathbb{Z}$ be the subset on which $m_{\theta,1} \geq 0$. The distribution μ is*

(i) fully sticky *if K is empty;*
(ii) partly sticky *if K is non-empty and $m_{\theta,1} = 0$ on its entirety; and*
(iii) nonsticky *if K has non-empty interior and $m_{\theta,1} > 0$ on int(K).*

The measure μ is sticky *if it is either fully sticky or partly sticky. When μ is partly sticky, a direction θ is* sticky *if $m_{\theta,1} < 0$ and* fluctuating *if $m_{\theta,1} \geq 0$.*

Notice that since $\theta \mapsto m_{\theta,1}$ is continuous, the set K from Definition 1.8 is always a closed set. To rule out certain pathologies, we always assume the following nondegeneracy condition.

Assumption 1.9. *The measure μ is* nondegenerate *in the sense that*

$$\mu(R_{\theta,\theta'}) < 1 \textit{ for all angles } \theta, \theta' \textit{ such that } |\theta - \theta'| \geq \pi, \tag{1.5}$$

where for angles $\theta, \theta' \in \mathbb{R}/\alpha\mathbb{Z}$,

$$R_{\theta,\theta'} = \{(r, \theta) \mid r \geq 0\} \cup \{(r, \theta') \mid r \geq 0\},$$

the union of the two rays at angles θ and θ'.

If nondegeneracy does not hold, then $\mu(R_{\theta,\theta'}) = 1$ for some pair of angles $\theta, \theta' \in \mathbb{R}/\alpha\mathbb{Z}$ such that $|\theta - \theta'| \geq \pi$: all of the mass is concentrated on two rays separated by an angle of at least π. Since $|\theta - \theta'| \geq \pi$ means that $(1, \theta') \in \overline{\mathcal{I}_\theta}$ (or equivalently that $(1, \theta) \in \overline{\mathcal{I}_{\theta'}}$), it is not hard to show that this scenario is metrically equivalent the case of $\mathcal{K} = \mathbb{R}$.

The terms fully sticky, partly sticky, and nonsticky in Definition 1.8 are mutually exclusive. The following result shows that under minimal assumptions, every distribution is covered by one of these three cases; this is essentially Proposition 4.11.

Proposition 1.10. *If μ is a probability measure on $\mathcal{K}$ that is integrable (1.2) and nondegenerate (1.5), then μ is either fully sticky, partly sticky, or nonsticky. Furthermore, if μ is partly sticky, then the interval $[A, B]$ on which $m_{\theta,1} \geq 0$ has length $|A - B| < \pi$; if μ is nonsticky, then $|A - B| \leq \pi$ and the function $\theta \mapsto m_{\theta,1}$ is strictly concave on its interior (A, B).*

We are now in a position to generalize the concept of barycenter in $\mathcal{K}$ to the setting where μ only satisfies the integrability condition (1.2) but not the square-integrability condition (1.3).

Definition 1.11. *If the probability distribution μ satisfies* (1.2) *and is sticky (either fully or partly sticky), then set the* mean *of μ equal to the origin $\mathbf{0}$. If μ is nonsticky, then set the* mean *of μ equal to the point $(m_{\theta',1}, \theta') \in \mathcal{K}$, where θ' maximizes the function $\theta \mapsto m_{\theta,1}$.*

In light of Proposition 1.10, the mean of μ is well defined for all distributions that satisfy the integrability and nondegeneracy assumptions; the second moment condition used in the definition of the barycenter is not necessary to define a mean. In Corollary 4.13 we show that when the barycenter is defined, the mean of μ coincides with its barycenter.

1.3 Empirical measures and the law of large numbers

For a given set of points $\{p_n\}_{n=1}^N \subset \mathcal{K}$, define the empirical measure

$$\mu_N = \frac{1}{N}\sum_{n=1}^N \delta_{p_n},$$

the averaged sum of unit measures supported on the points p_n. This is a Borel probability measure on $\mathcal{K}$, and all results of the previous section apply to μ_N. Let $b_N = b(p_1, \ldots, p_N)$ be the barycenter of μ_N:

$$b_N = b(p_1, \ldots, p_N) = \underset{p\in\mathcal{K}}{\arg\min}\Big(\frac{1}{2N}\sum_{n=1}^N d(p, p_n)^2\Big), \tag{1.6}$$

uniquely defined (by Corollary 4.13). For $\theta \in \mathbb{R}/\alpha\mathbb{Z}$, write $\eta_{\theta,N} \in \mathbb{R}^2$ for the *folded average*

$$\eta_{\theta,N} = \frac{1}{N}\sum_{n=1}^N F_\theta(p_n). \tag{1.7}$$

The folded first moments of μ_N, which we denote by $m_\theta^N \in \mathbb{R}^2$, are defined by

$$m_\theta^N = e_1\, m_{\theta,1}^N + e_2\, m_{\theta,2}^N$$

where

$$m_{\theta,i}^N = e_i \cdot m_\theta^N = \int_{\mathcal{K}} e_i \cdot F_\theta(p)\; d\mu_N(p) \quad \text{for } i = 1, 2.$$

Comparing these formulas to (1.7), the folded average is evidently equivalent to the first moment of the empirical measure:

$$\eta_{\theta,N} = m_\theta^N \quad \text{ for all } \theta \in \mathbb{R}/\alpha\mathbb{Z}. \tag{1.8}$$

An important issue in our analysis is whether the folded average $\eta_{\theta,N}$ is close to the folded barycenter $F_\theta b_N$, that is, whether "averaging commutes with folding". These two points in $\mathbb{R}^2$ may not coincide; the relation between $\eta_{\theta,N}$ and $F_\theta b_N$ is addressed later in Lemma 4.15.

Henceforth, let $\{p_n\}_{n=1}^N$ be a collection of independent random points on $\mathcal{K}$, each distributed according to μ. More precisely, let $\{p_n(\omega) \mid n = 1, \ldots, N\}$ be a collection of independent, identically distributed $\mathcal{K}$-valued random variables, each distributed according to μ over a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Their barycenter $b_N(\omega) = b(p_1(\omega), \ldots, p_N(\omega)) \in \mathcal{K}$ is also a random variable taking values in $\mathcal{K}$. For each $\theta \in \mathbb{R}/\alpha\mathbb{Z}$, let $m_\theta^N = m_\theta^N(\omega)$ be the random first moments associated with the empirical measures $\mu_N = \mu_N(\omega) = \frac{1}{N}\sum_{n=1}^N \delta_{p_n(\omega)}$. As before, denote by m_θ the deterministic folded means of μ in Definition 1.7. For any angle θ,

$$\mathbb{E}[m_\theta^N] = \frac{1}{N}\sum_{n=1}^N \int_{\mathcal{K}} F_\theta(p_n)\, d\mu(p_n) = \frac{1}{N}\sum_{n=1}^N \int_{\mathbb{R}^2} z\, d\tilde{\mu}_\theta(z) = m_\theta.$$

By the usual strong law of large numbers for $\mathbb{R}^2$-valued random variables,

$$m_\theta^N \to m_\theta \ \mathbb{P}\text{-almost surely as } N \to \infty, \text{for all } \theta \in \mathbb{R}/\alpha\mathbb{Z}. \tag{1.9}$$

Translating back into a law of large numbers in $\mathcal{K}$ for the random barycenters b_N, the behavior in the first two cases is strikingly different than the typical law of large numbers in a Euclidean space. The following result is proved in Section 5.

Theorem 1.12 (Law of Large Numbers on $\mathcal{K}$). *Assume that μ satisfies the integrability condition (1.2). Exactly one of the following holds, depending on how sticky μ is.*

1. *(Fully sticky) The mean of μ is $\mathbf{0}$ and there exists a random integer N^* such that the barycenter b_N from (1.6) satisfies $b_N(\omega) = \mathbf{0}$ for all $N \geq N^*(\omega)$, $\mathbb{P}$-almost surely.*
2. *(Partly sticky) The mean of μ is $\mathbf{0}$ and $b_N(\omega) \to \mathbf{0}$ almost surely as $N \to \infty$. Furthermore, if $[A, B]$ is the interval of fluctuating directions from Definition 1.8(ii) and Proposition 1.10, and I is an open interval of angles containing $[A, B]$ then there exists a random integer N^* such that $b_N(\omega) \in C_I$ (Definition 1.2) for all $N \geq N^*(\omega)$, almost surely.*
3. *(Nonsticky) The mean $\bar{b}$ of μ is not $\mathbf{0}$ and $b_N(\omega) \to \bar{b}$ almost surely as $N \to \infty$.*

The theorem implies that for all of the sticky directions θ, the empirical mean b_N stops fluctuating after some random but finite N^* along the ray $\{(r, \theta) \mid r \geq 0\}$; this is the phenomenon that we refer to as “stickiness". In fluctuating directions, the empirical mean b_N continues to vary as $N \to \infty$, although the magnitude of the movement goes to zero asymptotically.

1.4 Central Limit Theorems

The central limit theorems in this section describe the asymptotic behavior of the properly normalized fluctuations of b_N about the mean of μ. Due to the non-standard nature of the sticky law of large numbers, it is not surprising that the central limit theorem also takes a different form in sticky cases. Even in the nonsticky case, the central limit theorem is non-standard. Each of the three possibilities in Proposition 1.10 is covered in a separate theorem; these three theorems are proved in Section 6.

1.4.1 Fully sticky case

The simplest case is the fully sticky case, where there are asymptotically no fluctuations in any direction. On $\mathcal{K}$ define the scaling $\beta(r, \theta) = (\beta r, \theta)$ for arbitrary $\beta \geq 0$ such that $F_{\theta'}\big(\beta(r, \theta)\big) = \beta F_{\theta'}(r, \theta)$ for all $\theta, \theta' \in \mathbb{R}/\alpha\mathbb{Z}$ and $r, \beta \geq 0$. Let ν_N denote the distribution of the rescaled empirical means:

$$\nu_N(U) = \mathbb{P}\big(\sqrt{N} b_N(\omega) \in U\big), \text{ where } b_N \text{ is the empirical barycenter (1.6)} \tag{1.10}$$

for all Borel sets $U \subset \mathcal{K}$.

Theorem 1.13. *If a probability measure μ on $\mathcal{K}$ is fully sticky, then the rescaled empirical mean measures $\{\nu_N\}_{N=1}^\infty$ from (1.10) converge in the total variation norm (and hence weakly) to the point measure $\delta_\mathbf{0}$ as $N \to \infty$. In particular, for any bounded function $\phi : \mathcal{K} \to \mathbb{R}$,*

$$\lim_{N \to \infty} \int_\mathcal{K} \phi(p) d\nu_N(p) = \phi(\mathbf{0}). \tag{1.11}$$

In this fully sticky case, the term “Central Limit Theorem” is a bit of a misnomer, since there are no asymptotic fluctuations. In fact, Theorem 1.13 would still be true if we replace $\sqrt{N}$ in (1.10) with any increasing function of N.

The next two cases require a bit more notation and setup.

1.4.2 Partly sticky case

Assume the second moment condition (1.3). Since the mean $\bar{b}$ of μ lies at the origin $\mathbf{0}$ in the partly sticky case, again consider the rescaled empirical measure ν_N defined by (1.10). The limit of ν_N is another measure on $\mathcal{K}$, constructed as follows.

Let θ^* and $\rho \in [0, \pi/2)$ be such that $[\theta^* - \rho, \theta^* + \rho] = [A, B]$ where $[A, B]$ is the interval of fluctuating directions (Definition 1.8.ii and Proposition 1.10). Let g denote the law of the multivariate normal random variable on $\mathbb{R}^2$ having mean zero and covariance matrix

$$\Sigma = \int_{\mathbb{R}^2} yy^T d\tilde{\mu}_{\theta^*}(y). \tag{1.12}$$

This matrix is well defined due to the square-integrability condition (1.3).

Denote by $D_\rho \subset \mathbb{R}^2$ the closed sector

$$D_\rho = \big\{(r\cos\vartheta, r\sin\vartheta) \in \mathbb{R}^2 \mid r \geq 0 \text{ and } -\rho \leq \vartheta \leq \rho\big\} \tag{1.13}$$

and by $\hat{P}_\rho : \mathbb{R}^2 \to D_\rho$ the *convex projection* onto D_ρ:

$$\hat{P}_\rho(q) = \underset{z \in D_\rho}{\arg\min}\, d_2(q, z), \tag{1.14}$$

where $d_2(z, w) : \mathbb{R}^2 \times \mathbb{R}^2 \to [0, \infty)$ denotes the Euclidean metric in $\mathbb{R}^2$. Since $|A - B| < \pi$, the folding map F_{θ^*} takes the sector $C_{[A,B]}$ (Definition 1.2) bijectively to D_ρ. It is possible that $\rho = 0$ or equivalently $A = B$, in which case $C_{[A,B]}$ and D_ρ are rays.

Finally, define the measure h_{θ^*} on $\mathcal{K}$ by

$$h_{\theta^*} = g \circ \hat{P}_\rho^{-1} \circ F_{\theta^*}, \tag{1.15}$$

where $g \circ \hat{P}_\rho^{-1}$ is the pushforward of the normal measure g, whose covariance matrix is defined in (1.12), under the projection $\hat{P}_\rho$ to D_ρ. Figure 2 illustrates the construction in an example.

Theorem 1.14. *If a measure μ on $\mathcal{K}$ is partly sticky and square-integrable* (1.3)*, then the rescaled empirical mean measures $\{\nu_N\}_{N=1}^\infty$ from (1.10) converge weakly to the measure h_{θ^*} from (1.15) as $N \to \infty$, where θ^* is the midpoint of the interval K in Definition 1.8. That is, for any continuous, bounded function $\phi : \mathcal{K} \to \mathbb{R}$,*

$$\lim_{N\to\infty} \int_{\mathcal{K}} \phi(p) d\nu_N(p) = \int_{\mathcal{K}} \phi(p) dh_{\theta^*}(p). \tag{1.16}$$

The measure h_{θ^*} is supported on the closed sector $C_{[A,B]}$. The limit distribution h_{θ^*} can be decomposed into a singular part and an absolutely continuous part:

$$h_{\theta^*} = h_{sing} + h_{abs}.$$

The absolutely continuous part is the restriction of a Gaussian to the set $\mathrm{int}(C_{[A,B]}) = C^+_{(A,B)}$, which is the interior of the closed sector $C_{[A,B]}$:

$$h_{abs}(V) = g \circ F_{\theta^*}\big(V \cap C^+_{(A,B)}\big).$$

When $A = B$, the sector $C_{[A,B]}$ has no interior and $h_{abs} = 0$. The singular part h_{sing} is supported on the boundary $\partial C_{[A,B]}$, and it includes an atom $w\delta_0(p)$ at the origin with weight

$$w = g\left(\big\{(r\cos\vartheta, r\sin\vartheta) \in \mathbb{R}^2 \mid r > 0 \text{ and } \vartheta \in [\rho + \pi/2, 3\pi/2 - \rho]\big\}\right).$$

However, not all of the mass in the singular part lies at the origin; h_{sing} also distributes mass continuously on the edges of the sector $C_{[A,B]}$. In particular,

$$h_{sing}\big(\partial C_{[A,B]} \backslash \{0\}\big) = g\big(\big\{(r\cos\vartheta, r\sin\vartheta) \in \mathbb{R}^2 \,|\, r > 0, \vartheta \in [\rho, \rho + \pi/2) \cup (3\pi/2 - \rho, 2\pi - \rho]\big\}\big).$$

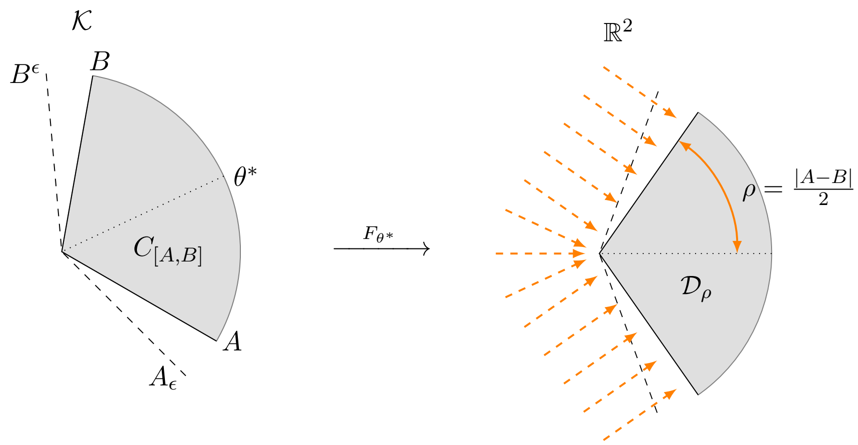

Figure 2: Partly sticky case. Left: $C_{[A,B]}$ is that sector in $\mathcal{K}$ centered at θ^* that is spanned by the angles θ for which $m_{\theta,1} = 0$. For N larger than a finite but random number, $b_N \in C_{[A_\epsilon,B^\epsilon]}$ almost surely. Right: D_ρ is the bijective image of $C_{[A,B]}$ under the folding map centered at θ^*. With a Gaussian g centered at $0 \in \mathbb{R}^2$, up to the bijection, the limiting measure is g on $\text{int}(D_\rho)$ and the pushforward of g on $\mathbb{R}^2 \setminus D_\rho$ to ∂D_ρ under the convex projection $\hat{P}_\rho$. The dashed arrows show the directions of this convex projection.

1.4.3 Nonsticky case

When μ is nonsticky, the mean of μ is $\bar{b} = (r^*, \theta^*) \in \mathcal{K}$, where $r^* = m_{\theta^*,1} > 0$ and θ^* is the unique angle for which

$$m_{\theta^*,1} = \max_\theta m_{\theta,1}.$$

In particular this means that $\bar{b} \neq \mathbf{0}$, so the limit measure obtained by renormalizing fluctuations of b_N lives on the tangent space of $\bar{b}$, which is isomorphic to $\mathbb{R}^2$, not $\mathcal{K}$ as in sticky cases.

With θ^* fixed, the family of random variables $\{m_{\theta^*}^N\}_{N=1}^\infty$ satisfies a standard central limit theorem in $\mathbb{R}^2$. Specifically, let g be the law of a multivariate normal random variable on $\mathbb{R}^2$ with zero mean and covariance matrix

$$\Sigma = \int_{\mathbb{R}^2} (y - F_{\theta^*}\bar{b})(y - F_{\theta^*}\bar{b})^T d\tilde{\mu}_{\theta^*}(y).$$

This matrix is well defined under the square-integrability condition (1.3). The standard central limit theorem implies that as $N \to \infty$ the law of the random variable

$$\sqrt{N}\left(m_{\theta^*}^N - F_{\theta^*}\bar{b}\right)$$

in $\mathbb{R}^2$ converges weakly to g.

Although is it reasonable to expect that $F_{\theta^*}b_N$ would satisfy the same central limit theorem, this might in fact not be the case, depending on whether the closed shadow $\overline{\mathcal{I}_{\theta^*}}$ carries mass. Define $\kappa \geq 0$ to be the random variable

$$\kappa(\omega) = \begin{cases} \dfrac{w^+(\theta^*)}{r^*} & \text{if } e_2 \cdot F_{\theta^*}b_N(\omega) < 0, \\[2ex] \dfrac{w^-(\theta^*)}{r^*} & \text{if } e_2 \cdot F_{\theta^*}b_N(\omega) > 0, \\[2ex] 0 & \text{else,} \end{cases} \tag{1.17}$$

where (cf. Figure 1 for $\mathcal{I}_\theta^\pm$)

$$w^\pm(\theta) = \int_{\mathcal{I}_\theta^\pm} d(\mathbf{0}, p)\, d\mu(p),$$

and

$$\begin{aligned} \mathcal{I}_\theta^+ &= \mathcal{K} \setminus \{(r,\theta') \mid r > 0 \text{ and } -\pi \le \theta' - \theta < \pi\}, \\ \mathcal{I}_\theta^- &= \mathcal{K} \setminus \{(r,\theta') \mid r > 0 \text{ and } -\pi < \theta' - \theta \le \pi\}. \end{aligned} \tag{1.18}$$

On the Borel sets W in $\mathbb{R}^2$ define the family

$$\tilde{\nu}_N(W) = \mathbb{P}\big(\sqrt{N}(e_1 \cdot F_{\theta^*} b_N - r^*, (1+\kappa) e_2 \cdot F_{\theta^*} b_N) \in W\big)$$

of measures indexed by N. If $\mu(\overline{\mathcal{I}_{\theta^*}}) = 0$, then $\kappa = 0$ and

$$\tilde{\nu}_N(W) = \mathbb{P}\big(\sqrt{N}(F_{\theta^*} b_N - F_{\theta^*} \bar{b}) \in W\big),$$

since $F_{\theta^*} \bar{b} = (r^*, 0)$.

Theorem 1.15. *If μ is nonsticky and square-integrable* (1.3)*, then the measures $\{\tilde{\nu}_N\}_{N=1}^\infty$ converge weakly to g as $N \to \infty$. That is, for any continuous, bounded function $\phi : \mathbb{R}^2 \to \mathbb{R}$,*

$$\lim_{N\to\infty} \int_{\mathbb{R}} \phi(z) d\tilde{\nu}_N(z) = \int_{\mathbb{R}} \phi(z) dg(z). \tag{1.19}$$

When $w^+(\theta^*) = w^-(\theta^*)$, Theorem 1.15 implies that $\sqrt{N}(F_{\theta^*} b_N - F_{\theta^*} \bar{b})$ is Gaussian in the limit as $N \to \infty$. When $w^+(\theta^*) = w^-(\theta^*) > 0$, the fluctuation of $F_{\theta^*} b_N - F_{\theta^*} \bar{b}$ in the e_2 direction is smaller than the fluctuation of $m_{\theta^*,2}^N$; this is due to the presence of mass in the closed shadow $\overline{\mathcal{I}_{\theta^*}}$. On the other hand, if $w^+(\theta^*) \neq w^-(\theta^*)$, then $\sqrt{N}(F_{\theta^*} b_N - F_{\theta^*} \bar{b})$ is not Gaussian in the limit; see Example 2.5.

2 Examples

Here are a few examples illustrating some phenomena described by the limit theorems.

Example 2.1 (Partly sticky)**.** Fix $\alpha > 2\pi$ and $\theta^* \in \mathbb{R}/\alpha\mathbb{Z}$. Let $K \ge 3$ be an odd integer. Let μ be the sum of K atoms having mass $1/K$ at the points

$$q_k = (1, \theta^* + \frac{2\pi}{K} k), \quad k = -\frac{K-1}{2}, \ldots, 0, \ldots, \frac{K-1}{2} \in \mathbb{Z}.$$

That is,

$$\mu = \frac{1}{K} \sum_{k=1}^K \delta_{q_k}.$$

In this case $m_\theta \le 0$ for all $\theta \in \mathbb{R}/\alpha\mathbb{Z}$, while $m_\theta = 0$ if and only if $|\theta - \theta^*| \le \pi/K$. The limit distribution h_{θ^*} is supported on the sector

$$C_{[-\frac{\pi}{K}, \frac{\pi}{K}]} = \Big\{(r,\theta) \,\Big|\, r \ge 0 \text{ and } -\frac{\pi}{K} \le \theta - \theta^* \le \frac{\pi}{K}\Big\},$$

including a singular part at the origin with weight $1 - \frac{2}{K}\lfloor\frac{K+2}{4}\rfloor - \frac{1}{K}$ and a singular part on $\partial C_{[-\frac{\pi}{K}, \frac{\pi}{K}]} \setminus \{0\}$, with weight $\frac{2}{K}\lfloor\frac{K+2}{4}\rfloor$ cf. Figure 3. The limit distribution does not vary with α, given that $\alpha > 2\pi$.

In the limit, Example 2.1 gives the following.

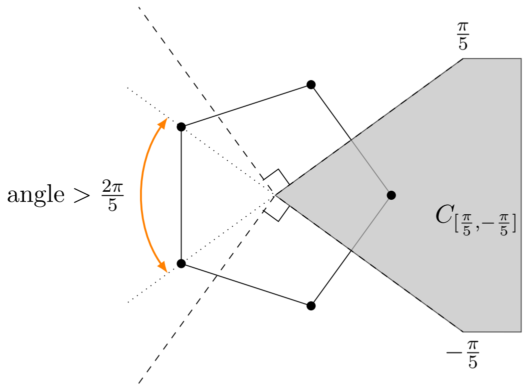

Figure 3: Example 2.1 in the case $K = 5$, with $\theta^* = 0$.

Example 2.2 (Partly sticky with singular limit measure)**.** Fix $\alpha > 2\pi$ and $\theta^* \in \mathbb{R}/\alpha\mathbb{Z}$. Suppose μ is uniform on the set

$$S_1 = \{(r, \theta) \in \mathcal{K} \mid r = 1 \text{ and } -\pi < \theta - \theta^* < \pi\}.$$

Then $m_\theta \leq 0$ for all $\theta \in \mathbb{R}/\alpha\mathbb{Z}$, while $m_\theta = 0$ only for $\theta = \theta^*$. The limit distribution h_{θ^*} puts an atom of mass $1/2$ at the origin, and half a Gaussian on the ray $\{(r, \theta^*) \mid r > 0\}$. In particular, h_{θ^*} has no absolutely continuous part. As in Example 2.1, the limit distribution does not vary with α, given that $\alpha > 2\pi$.

Example 2.3 (Embedding the spider)**.** Suppose $\alpha > K\pi$. Then there are angles $\theta_k \in \mathbb{R}/\alpha\mathbb{Z}$ for $k = 1, \dots, K$ such that $|\theta_k - \theta_j| > \pi$ for all $j \neq k$. Working with measures supported on the union of the rays at angles $\theta_1, \dots, \theta_K$ is equivalent to working with probability distributions on the spider with K legs—that is, an open book of dimension 1 with K leaves, cf. [14]—by mapping the ray $\{(r, \theta_k) \in \mathcal{K} \mid r > 0\}$ to a leg of the spider.

Example 2.4 (Full stickiness without square-integrability)**.** Let $d\sigma = r dr \otimes d\theta$ denote the canonical measure on $\mathcal{K}$. Here, dr denotes the usual Lebesgue measure on $[0, \infty)$ and $d\theta$ the canonical quotient measure on $\mathbb{R}/\alpha\mathbb{Z}$. With arbitrary but fixed $1 < \beta < 2$ let μ be the measure on $\mathcal{K}$ with density

$$g(r, \theta) = \frac{2\beta}{\alpha(\beta + 2)} \times \begin{cases} 1 & \text{if } 0 \leq r \leq 1 \\ \frac{1}{r^{\beta+2}} & \text{if } 1 \leq r < \infty. \end{cases}$$

The integrability condition (1.2) is satisfied with $\bar{r} = \frac{2\beta(\beta+2)}{3(\beta+1)(\beta-1)}$. Moreover, $m_{\theta,1} = (2\pi - \alpha)\bar{r} < 0$ for all $\theta \in \mathbb{R}/\alpha\mathbb{Z}$. By virtue of Theorem 1.13, there is a random integer N^* such that $b_N = \mathbf{0}$ for all $N \geq N^*$ almost surely. On the other hand, square-integrability does not hold, as $\int_{\mathcal{K}} r_p^2 \, d\mu(p) = \infty$, and hence $\bar{b}$ is not defined.

Example 2.5 (Non-Gaussian behavior in the nonsticky case)**.** Fix $t > 3$ and let μ be the distribution on $\mathcal{K}$ which puts mass $1/5$ at each of the points

$$p_1 = (t, 0), \quad p_2 = (1, \pi/2), \quad p_3 = (1, \pi), \quad p_4 = (2, -\pi), \quad p_5 = (1, -\pi/2).$$

The points p_3 and p_4 lie on the boundary of $\mathcal{I}_{\theta=0}$, so under the folding map F_θ with $\theta = 0$, the points p_3 and p_4 collapse onto the axis $(-\infty, 0) \times \{0\}$; points p_2 and p_5 map to the vertical axis $\{0\} \times \mathbb{R}$. We compute:

$$m_{\theta=0,1} = \frac{1}{5}(t + 0 - 1 - 2 + 0) = \frac{t - 3}{5} > 0.$$

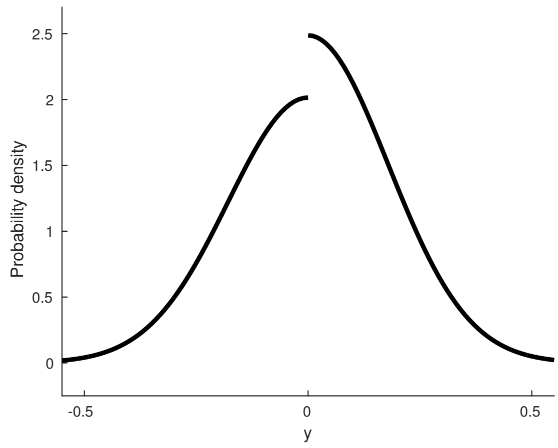

Figure 4: Depicted is the marginal in Example 2.5 for $t = 5$, i.e. the vertical component $y = \sqrt{N} e_2 \cdot F_0 b_N$ of the folded empirical means multiplied by $\sqrt{N}$. For these, $y \to \frac{1}{3}\mathbb{I}_{Z \geq 0} Z + \frac{2}{3}\mathbb{I}_{Z \leq 0} Z$ asymptotically in distribution as $N \to \infty$ where $Z \sim \mathcal{N}\Big(0, 2/5\Big)$.

The push-forward $\tilde{\mu}_\theta = \mu \circ F_\theta^{-1}$ has symmetry about the x-axis when $\theta = 0$, which implies that $m_{\theta=0,2} = 0$. By the results of Section 4 below, this implies $\theta \mapsto m_{\theta,1}$ is maximized at $\theta = \theta^* = 0$. However,

$$w^+(\theta^*) = \frac{1}{5} d(\mathbf{0}, p_3)^2 = \frac{1}{5} \quad \text{and} \quad w^-(\theta^*) = \frac{1}{5} d(\mathbf{0}, p_4)^2 = \frac{4}{5}$$

in this case. As a consequence of Theorem 1.15 and the subsequent remarks, the limit distribution of $\sqrt{N}(F_{\theta^*} b_N - F_{\theta^*} \bar{b})$ on $\mathbb{R}^2$ is non-degenerate and not Gaussian, cf. Figure 4.

3 Folding isolated hyperbolic planar singularities

This section elaborates on the geometric structure of the kale $\mathcal{K}$ defined in (1.1).

Lemma 3.1 (Openness of visibility)**.** *If p is fully visible from the angle θ_0 then it is fully visible from all θ sufficiently close to θ_0. The same is true for invisibility.*

Proof. The sets $\mathcal{I}_\theta$ and $\mathcal{K} \setminus \overline{\mathcal{I}_\theta}$ are open. □

Recall that $d_2(z, w) : \mathbb{R}^2 \times \mathbb{R}^2 \to [0, \infty)$ denotes the Euclidean metric in $\mathbb{R}^2$. The following lemma follows easily from the definitions of F_θ and the metric d on $\mathcal{K}$.

Lemma 3.2. *For any two points $p_1, p_2 \in \mathcal{K}$ and any angle $\theta \in \mathbb{R}/\alpha\mathbb{Z}$,*

$$d_2\big(F_\theta(p_1), F_\theta(p_2)\big) \leq d(p_1, p_2),$$

with strict inequality if $p_1 \in \mathcal{I}_\theta$, $p_2 \in \mathcal{K} \setminus \overline{\mathcal{I}_\theta}$, and p_2 has an angle different from θ. Moreover, for any $p \in \mathcal{K}$ and $\theta \in \mathbb{R}/\alpha\mathbb{Z}$,

$$d_2\big(F_\theta(r, \theta), F_\theta(p)\big) = d\big((r, \theta), p\big).$$

Lemma 3.3. *If $(1, \theta') \in \overline{\mathcal{I}_\theta}$ and $p \in \mathcal{K}$, then*

$$e_1 \cdot F_\theta(p) \leq -e_1 \cdot F_{\theta'}(p). \tag{3.1}$$

If $(1, \theta') \in \mathcal{I}_\theta$ then equality holds precisely when $p \in R_{\theta,\theta'}$.

Remark 3.4. The conditions $(1,\theta') \in \overline{\mathcal{I}_\theta}$ and $(1,\theta') \in \mathcal{I}_\theta$ could equivalently be expressed as $|\theta - \theta'| \geq \pi$ and $|\theta - \theta'| > \pi$, respectively; in particular, they are symmetric in θ and θ'.

Proof of Lemma 3.3. Both assertions are obvious for $p = \mathbf{0}$. Hence assume $p \neq \mathbf{0}$, i.e. that $p = (r, \hat{\theta}) \in \mathcal{K}$ with $r > 0$. Then

$$e_1 \cdot F_\theta(p) = r \cos\big(\min\{|\hat{\theta} - \theta|, \pi\}\big),$$

and similarly with θ' in place of θ. The statement of the lemma is symmetric in θ and θ' by Remark 3.4, so without loss of generality assume $|\hat{\theta} - \theta| \geq |\hat{\theta} - \theta'|$.

Then $e_1 \cdot F_\theta(p)$ and $e_1 \cdot F_{\theta'}(p)$ are both negative—and thus (3.1) with strict inequality is automatic—unless $|\hat{\theta} - \theta'| \leq \pi/2$. Henceforth assume $|\hat{\theta} - \theta'| \leq \pi/2$. Then $|\hat{\theta} - \theta| \geq \pi/2$ because $|\theta - \theta'| \geq \pi$.

If $|\hat{\theta} - \theta| \geq \pi$, then the left side of (3.1) is $-r$ while the right side is $-r\cos|\hat{\theta} - \theta'|$; the cosine is nonnegative because $|\hat{\theta} - \theta'| \leq \pi/2$, and it achieves the value 1 only when $\hat{\theta} = \theta'$, which is when $p \in R_{\theta,\theta'}$, as desired.

The only remaining case is where $|\hat{\theta} - \theta'| \leq \pi/2 \leq |\hat{\theta} - \theta| < \pi$. Since $(1,\theta') \in \overline{\mathcal{I}}_\theta$ but $|\hat{\theta} - \theta| < \pi$, the ray $\hat{\theta}$ must lie between θ' and θ, in the sense that $|\theta - \theta'| = |\theta - \hat{\theta}| + |\hat{\theta} - \theta'|$ and passing through this angle from θ' to θ hits the ray at $\hat{\theta}$ along the way. This picture is easily drawn in the Euclidean plane $\mathbb{R}^2$, with θ' along the horizontal axis, $\hat{\theta}$ in the first (northeast) quadrant, and $\theta - \pi$ between θ' and $\hat{\theta}$, possibly equal to θ' but never $\hat{\theta}$. (The reflection of this picture across the horizontal axis is possible, as well, but as cosine is an even function it changes none of the algebra.) Using $\theta - \pi$ instead of θ is handy because $-e_1 \cdot F_\theta(p)$ is the cosine of the angle β between $\theta - \pi$ and $\hat{\theta}$. The desired result follows because $\beta \leq |\hat{\theta} - \theta'| \leq \pi/2$ and cosine is strictly decreasing on the interval $[0, \pi/2]$ while $\beta = |\hat{\theta} - \theta'|$ only when $\theta - \theta' = \pi$, which is the case $(1,\theta') \in \overline{\mathcal{I}}_\theta \setminus \mathcal{I}_\theta$. $\square$

4 Barycenters and first moments of probability measures on the kale

This section describes properties of the functions $\theta \mapsto m_\theta$ and $\theta \mapsto m_\theta^N$; the behavior of these functions aids in understanding how the barycenters b_N behave in the limit $N \to \infty$. Recall that the barycenter is the minimizer of $\Gamma(p)$, defined in (1.4). To motivate what comes next and better explain the connection between barycenters and the first component $m_{\theta,1}$ of folded means m_θ, we recall the analogous calculation for $\mathbb{R}^n$. Define $\gamma\colon \mathbb{R}^n \to [0,\infty)$ by

$$\gamma(x) = \frac{1}{2}\int_{\mathbb{R}^n} \|x - y\|^2 d\nu(y)$$

for a given probability measure ν on $\mathbb{R}^n$. The barycenter of ν in this Euclidean setting is the point $x \in \mathbb{R}^n$ that minimizes $\gamma(x)$. Observe that

$$\|x - y\|^2 = \|x\|^2 - 2x \cdot y + \|y\|^2 = \|x\|^2 - 2\|x\|(\hat{x} \cdot y) + \|y\|^2,$$

where $\hat{x} = x/\|x\|$ is the unit vector in the direction of x. Hence if ν is square-integrable, and

$$\gamma(x) = \frac{1}{2}\|x\|^2 - \|x\| \int_{\mathbb{R}^n} (\hat{x} \cdot y)\, d\nu(y) + \gamma(0), \tag{4.1}$$

then the minimizer of γ lies in the direction $\hat{x}$ that maximizes

$$m \cdot \hat{x} = \int_{\mathbb{R}^n} (\hat{x} \cdot y)\, d\nu(y) \tag{4.2}$$

and at a distance from the origin equal to the maximum value of (4.2). Here $m \in \mathbb{R}^n$ is the mean of ν. Hence if $\hat{x}^*$ is the maximizing direction, then the barycenter can be written in polar coordinates $(r, \hat{x})$ as $(m \cdot \hat{x}^*, \hat{x}^*)$. From this it follows that the solution is the usual mean in Euclidean space. Even when the term $\gamma(0)$ in (4.1) is infinite, it is reasonable to take this as the definition of mean. To make the maximization of (4.2) well defined, one only needs to assume ν is integrable rather than square-integrable.

A similar calculation can be done in the kale setting. Since the folding map rotates the direction θ back to the direction e_1 in the Euclidean plane, $m_{\theta,1}$ is exactly analogous to (4.2). The following lemma proves the expression analogous to (4.1) in the setting of $\mathcal{K}$.

Lemma 4.1. *Suppose a measure μ is square-integrable (1.3). Then for all points $(r, \theta) \in \mathcal{K}$,*

$$\Gamma(r, \theta) = \frac{r^2}{2} - r\, m_{\theta,1} + \Gamma(\mathbf{0}).$$

Remark 4.2. As a consequence of $\|F_\theta(r, \theta)\| \le r$, the pushforward $\tilde{\mu}_\theta = \mu \circ F_\theta^{-1}$ is also square-integrable when μ is.

Proof of Lemma 4.1. For $p = (r, \theta)$, using Lemma 3.2,

$$\begin{aligned}
\Gamma(p) &= \frac{1}{2} \int_{\mathcal{K}} d_2(F_\theta p, F_\theta q)^2 \, d\mu(q) \\
&= \frac{1}{2} \int_{\mathcal{K}} \left(|e_1 \cdot F_\theta p - e_1 \cdot F_\theta q|^2 + |e_2 \cdot F_\theta p - e_2 \cdot F_\theta q|^2 \right) d\mu(q) \\
&= \frac{1}{2} \int_{\mathcal{K}} \left(|r - e_1 \cdot F_\theta q|^2 + |e_2 \cdot F_\theta q|^2 \right) d\mu(q) \\
&= \frac{r^2}{2} - r \int_{\mathcal{K}} e_1 \cdot F_\theta q \, d\mu(q) + \int_{\mathcal{K}} |\left(e_1 \cdot F_\theta q|^2 + |e_2 \cdot F_\theta q|^2 \right) d\mu(q).
\end{aligned} \qquad \square$$

Motivated by a need to understand properties of the function $m_{\theta,1}$, we now explore its differentiability. Define one-sided derivatives of $g\colon \mathbb{R}/\alpha\mathbb{Z} \to \mathbb{R}$ at $\theta \in \mathbb{R}/\alpha\mathbb{Z}$ by

$$D_\theta^+ g(\theta) = \lim_{\substack{\theta' \to \theta \\ \theta' \in (\theta, \theta + \pi)}} \frac{g(\theta') - g(\theta)}{\theta' - \theta} \quad \text{and} \quad D_\theta^- g(\theta) = \lim_{\substack{\theta' \to \theta \\ \theta' \in (\theta - \pi, \theta)}} \frac{g(\theta') - g(\theta)}{\theta' - \theta}.$$

Recall Definition 1.1 of the (not necessarily positive) real number $\theta' - \theta$. When the one-sided derivatives agree, write $\frac{d}{d\theta} g(\theta)$ or $g'(\theta)$ as usual.

Lemma 4.3. *The function $m_{\theta,1}\colon \mathbb{R}/\alpha\mathbb{Z} \to \mathbb{R}$ is continuously differentiable, and*

$$\frac{d}{d\theta} m_{\theta,1} = m_{\theta,2}.$$

Moreover, for every $\theta \in \mathbb{R}/\alpha\mathbb{Z}$, the one-sided derivatives $D_\theta^\pm \frac{dm_{\theta,1}}{d\theta} = D_\theta^\pm m_{\theta,2}$ exist and satisfy

$$D_\theta^\pm \frac{dm_{\theta,1}}{d\theta} = D_\theta^\pm m_{\theta,2} = -m_{\theta,1} - \int_{\mathcal{I}_\theta^\mp} d(\mathbf{0}, p)\, d\mu(p) = -m_{\theta,1} - w^\mp(\theta) \tag{4.3}$$

where $w^\pm(\theta)$ and $\mathcal{I}_\theta^\pm$ are as in (1.18), *cf. Figure 1. In particular since $\mathcal{I}_\theta \subset \mathcal{I}_\theta^\pm$,*

$$D_\theta^\pm \frac{dm_{\theta,1}}{d\theta} \le -m_{\theta,1} - \int_{\mathcal{I}_\theta} d(\mathbf{0}, p)\, d\mu(p) \tag{4.4}$$

holds for all $\theta \in \mathbb{R}/\alpha\mathbb{Z}$.

Proof. For $\theta' \in \mathbb{R}/\alpha\mathbb{Z}$, define functions $f_{\theta'} : \mathbb{R}/\alpha\mathbb{Z} \to [-1, 1]$ and $g_{\theta'} : \mathbb{R}/\alpha\mathbb{Z} \to [-1, 1]$ by

$$f_{\theta'}(\theta) = \cos\left(\min\{|\theta - \theta'|, \pi\} \right) \quad \text{and} \quad g_{\theta'}(\theta) = \begin{cases} \sin(\theta' - \theta) & \text{if } |\theta - \theta'| \le \pi \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$m_{\theta,1} = \int_{\mathcal{K}} r_p\, f_{\theta_p}(\theta)\, d\mu(p) \quad \text{and} \quad m_{\theta,2} = \int_{\mathcal{K}} r_p\, g_{\theta_p}(\theta)\, d\mu(p), \tag{4.5}$$

where $p = (r_p, \theta_p)$. Each function f_{θ_p} is continuously differentiable. The integrability condition (1.2) and the dominated convergence theorem imply that

$$\frac{dm_{\theta,1}}{d\theta} = \int_{\mathcal{K}} r_p\, f'_{\theta_p}(\theta) d\mu(p) = \int_{\mathcal{K}\setminus\mathcal{I}_\theta} r_p\, \sin(\theta_p - \theta) d\mu(p) = m_{\theta,2}.$$

Each g_{θ_p} has one-sided derivatives:

$$D^+_\theta g_{\theta_p}(\theta) = \begin{cases} -\cos(\theta_p - \theta) & \text{if } \theta_p - \pi \le \theta < \theta_p + \pi, \text{ i.e. } -\pi < \theta_p - \theta \le \pi \\ 0 & \text{otherwise,} \end{cases}$$

$$D^-_\theta g_{\theta_p}(\theta) = \begin{cases} -\cos(\theta_p - \theta) & \text{if } \theta_p - \pi < \theta \le \theta_p + \pi \text{ i.e. } -\pi \le \theta_p - \theta < \pi \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, by the dominated convergence theorem, $m_{\theta,2}$ also has one-sided derivatives at every $\theta \in \mathbb{R}/\alpha\mathbb{Z}$:

$$\begin{aligned} D^+_\theta m_{\theta,2} = \int_{\mathcal{K}} r_p\, D^+_\theta g_{\theta_p}(\theta)\, d\mu(p) &= -\int_{\mathcal{K}\setminus\mathcal{I}^-_\theta} r_p \cos(\theta_p - \theta)\, d\mu(p) \\ &= -m_{\theta,1} - \int_{\mathcal{I}^-_\theta} r_p\, d\mu(p). \end{aligned}$$

Similarly,

$$\begin{aligned} D^-_\theta m_{\theta,2} = \int_{\mathcal{K}} r_p\, D^-_\theta g_{\theta_p}(\theta)\, d\mu(p) &= -\int_{\mathcal{K}\setminus\mathcal{I}^+_\theta} r_p \cos(\theta_p - \theta)\, d\mu(p) \\ &= -m_{\theta,1} - \int_{\mathcal{I}^+_\theta} r_p\, d\mu(p). \end{aligned}$$

In particular, (4.4) holds for all $\theta \in \mathbb{R}/\alpha\mathbb{Z}$. □

Corollary 4.4. *Let $A \ne B$ and $|A - B| \le \pi$. If $m_{\theta,1} = 0$ for all $\theta \in [A, B]$ then $\mu(\mathcal{I}_\theta) = 0$ for all $\theta \in [A, B]$.*

Proof. For $\theta \in (A, B)$ this an immediate consequence of (4.4), since

$$0 = D^\pm_\theta \frac{dm_{\theta,1}}{d\theta} \le -\int_{\mathcal{I}_\theta} d(\mathbf{0}, p)\, d\mu(p) \le 0.$$

When $0 < B - A \le \pi$, the D^+_θ and D^-_θ versions of this calculation remain valid for the endpoints $\theta = A$ and $\theta = B$, respectively, and swapped when $0 < A - B \le \pi$. □

Example 4.5. The assertion of Corollary 4.4 is wrong when $A = B$, i.e. when $\theta^* \in \mathbb{R}/\alpha\mathbb{Z}$ with $m_{\theta^*,1} = 0$ is isolated. To see this, consider μ having point masses of weight $1/3$ at $(1, \theta^*)$ as well as at $(1/2, \theta^* + \pi + \epsilon)$ and $(1/2, \theta^* - \pi - \epsilon)$ with $0 < \epsilon < \alpha/2 - \pi$. Then $\mu(\mathcal{I}_{\theta^*}) = 2/3$ while $m_{\theta,1} < 0$ for all $\theta \ne \theta^*$ and $m_{\theta^*,1} = 0$.

Example 4.6. The shadow of an angle θ with $m_{\theta,1} > 0$ may carry mass. Changing the first point in Example 4.5 into $(2, \theta^*)$ yields $m_{\theta^*,1} = 1/3 > 0$ and $\mu(\mathcal{I}_{\theta^*}) = 2/3$.

Recalling the definition $w^\pm(\theta)$ from (1.18), observe that $0 \le w^\pm(\theta) \le \bar{r}$ holds for all θ because the integrand is nonnegative and $\mathcal{I}^\pm_\theta \subset \mathcal{K}$. Also, as a consequence of (4.5),

$$\|m_\theta\| = \sqrt{m^2_{\theta,1} + m^2_{\theta,2}} \le \sqrt{2}\bar{r}. \tag{4.6}$$

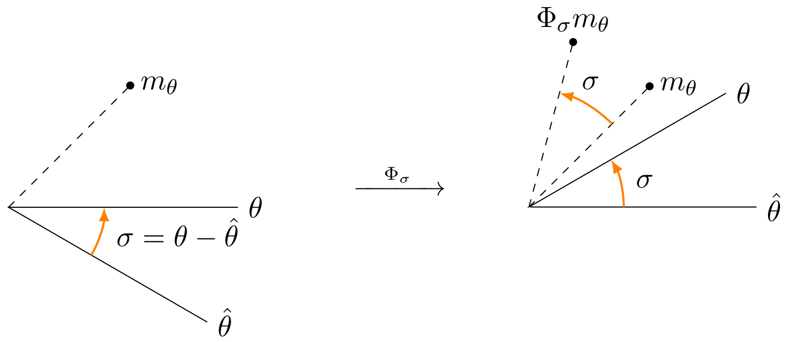

Figure 5: The rotation. If all the shadows from θ to $\hat{\theta}$ carry no mass then $\Phi_{\hat{\theta}-\theta} m_\theta = m_{\hat{\theta}}$ by Lemma 4.8.

Since μ is a probability measure, due to σ-additivity, only countably many of the rays $\{(r,\theta) \mid 0 \le r < \infty\}$ for $\theta \in \mathbb{R}/\alpha\mathbb{Z}$ carry positive mass of μ. Consequently, w^+ and w^- are continuous almost everywhere with respect to the understood measure on $\mathbb{R}/\alpha\mathbb{Z}$ induced by Lebesgue measure on $[0,\alpha)$, and so is $\theta \mapsto D^\pm m_{\theta,2}$. Furthermore, $w^+(\theta) = w^-(\theta)$ for almost every angle θ.

Corollary 4.7. *Let $\hat{\theta} \in \mathbb{R}/\alpha\mathbb{Z}$ and $\theta \in [\hat{\theta}-\pi, \hat{\theta}+\pi]$. Then*

$$m_{\theta,1} = m_{\hat{\theta},1}\cos(\theta-\hat{\theta}) + m_{\hat{\theta},2}\sin(\theta-\hat{\theta}) - \int_{\hat{\theta}}^{\theta} w(\psi)\sin(\theta-\psi)\,d\psi \tag{4.7}$$

and

$$m_{\theta,2} = -m_{\hat{\theta},1}\sin(\theta-\hat{\theta}) + m_{\hat{\theta},2}\cos(\theta-\hat{\theta}) - \int_{\hat{\theta}}^{\theta} w(\psi)\cos(\theta-\psi)\,d\psi. \tag{4.8}$$

where

$$w(\psi) = \int_{\mathcal{I}_\psi} d(\mathbf{0},p)\,d\mu(p). \tag{4.9}$$

Proof. Since w^+ and w^- are equal for almost every angle, we have

$$\int_{\hat{\theta}}^{\theta} w^\pm(\psi)\sin(\theta-\psi)\,d\psi = \int_{\hat{\theta}}^{\theta} w(\psi)\sin(\theta-\psi)\,d\psi$$

and similarly for the integral in (4.8). Equation (4.7) then follows from (4.3) using integration by parts along $\psi \in (\hat{\theta},\theta)$:

$$m_{\psi,1}\sin(\theta-\psi) = -D_\psi^\pm m_{\psi,2}\sin(\theta-\psi) - w^\mp(\psi)\sin(\theta-\psi).$$

Equation (4.8) follows from

$$D_\psi^\pm m_{\psi,2}\cos(\theta-\psi) = -m_{\psi,1}\cos(\theta-\psi) - w^\mp(\psi)\cos(\theta-\psi).$$

□

For an angle $\sigma \in \mathbb{R}$ define the rotation $\Phi_\sigma : \mathbb{R}^2 \to \mathbb{R}^2$ by $\Phi_\sigma(r,\psi) = (r,\psi+\sigma)$ in polar coordinates, cf. Figure 5. As usual, denote by $\|\cdot\|$ the standard Euclidean norm on $\mathbb{R}^2$. Recall Definition 1.4, specifically $\mathcal{I}_A$ for an interval A.

Lemma 4.8. *Let $\hat\theta \in \mathbb{R}/\alpha\mathbb{Z}$. For all $\theta \in \mathbb{R}/\alpha\mathbb{Z}$ with $|\theta - \hat\theta| \le \pi$,*

$$\|\Phi_{\hat\theta-\theta} m_{\hat\theta} - m_\theta\| \le |\theta - \hat\theta| \int_{\mathcal{I}_{(\hat\theta,\theta)}} d(\mathbf{0}, p)\, d\mu(p).$$

In particular, $\|\Phi_{\hat\theta-\theta} m_{\hat\theta} - m_\theta\| \le |\theta - \hat\theta|\bar r$. Also, if $\mu(\mathcal{I}_{(\hat\theta,\theta)}) = 0$, then $\Phi_{\hat\theta-\theta} m_{\hat\theta} = m_\theta$.

Proof. Suppose $\theta \in [\hat\theta - \pi, \hat\theta + \pi]$. Then

$$\Phi_{\hat\theta-\theta} m_{\hat\theta} = \begin{pmatrix} \cos(\theta-\hat\theta) & \sin(\theta-\hat\theta) \\ -\sin(\theta-\hat\theta) & \cos(\theta-\hat\theta) \end{pmatrix} \begin{pmatrix} m_{\hat\theta,1} \\ m_{\hat\theta,2} \end{pmatrix}.$$

So, by Corollary 4.7,

$$\begin{aligned} \|\Phi_{\hat\theta-\theta} m_{\hat\theta} - m_\theta\|^2 &= \left| \int_{\hat\theta}^{\theta} w(\psi) \sin(\theta-\psi)\, d\psi \right|^2 + \left| \int_{\hat\theta}^{\theta} w(\psi) \cos(\theta-\psi)\, d\psi \right|^2 \\ &\le |\theta-\hat\theta| \int_{\hat\theta}^{\theta} w(\psi)^2 \sin^2(\theta-\psi)\, d\psi + |\theta-\hat\theta| \int_{\hat\theta}^{\theta} w(\psi)^2 \cos^2(\theta-\psi)\, d\psi \\ &= |\theta-\hat\theta| \int_{\hat\theta}^{\theta} w(\psi)^2\, d\psi \le |\theta-\hat\theta|^2 \sup_{\psi\in(\hat\theta,\theta)} w(\psi)^2. \end{aligned}$$

The assertion follows now from

$$\sup_{\psi\in(\theta,\hat\theta)} w(\psi) = \sup_{\psi\in(\theta,\hat\theta)} \int_{I_\psi} d(\mathbf{0},p)\, d\mu(p) \le \int_{\mathcal{I}_{(\theta,\hat\theta)}} d(\mathbf{0},p)\, d\mu(p).$$ □

Lemma 4.9. *If $\theta, \theta' \in \mathbb{R}/\alpha\mathbb{Z}$ with $(1,\theta') \in \overline{\mathcal{I}_\theta}$, then $m_{\theta,1} \le -m_{\theta',1}$. If $(1,\theta') \in \mathcal{I}_\theta$, then $m_{\theta,1} = -m_{\theta',1}$ if and only if $\mu(R_{\theta,\theta'}) = 1$.*

Proof. This follows from Lemma 3.3 and the definition of m_θ and $m_{\theta'}$:

$$\begin{aligned} m_{\theta,1} &= \int_{R_{\theta,\theta'}} e_1 \cdot F_\theta(p)\, d\mu(p) + \int_{\mathcal{K}\setminus R_{\theta,\theta'}} e_1 \cdot F_\theta(p)\, d\mu(p) \\ &= -\int_{R_{\theta,\theta'}} e_1 \cdot F_{\theta'}(p)\, d\mu(p) + \int_{\mathcal{K}\setminus R_{\theta,\theta'}} e_1 \cdot F_\theta(p)\, d\mu(p) \text{ by equality case in Lemma 3.3} \\ &\le -\int_{R_{\theta,\theta'}} e_1 \cdot F_{\theta'}(p)\, d\mu(p) - \int_{\mathcal{K}\setminus R_{\theta,\theta'}} e_1 \cdot F_{\theta'}(p)\, d\mu(p) \text{ by Lemma 3.3} \\ &= -m_{\theta',1}. \end{aligned}$$

If $(1,\theta') \in \mathcal{I}_\theta$, then equality holds only if $\mu(\mathcal{K} \setminus R_{\theta,\theta'}) = 0$. □

Corollary 4.10. *The nondegeneracy condition (Assumption 1.9) implies that $m_{\theta,1} < -m_{\theta',1}$ whenever $(1,\theta') \in \mathcal{I}_\theta$, or in other words, whenever $|\theta - \theta'| > \pi$.*

Proposition 4.11. *Assuming integrability (1.2) and nondegeneracy (Definition 1.9), the subset of $\mathbb{R}/\alpha\mathbb{Z}$ on which $m_{\theta,1} \ge 0$ is a closed interval that is exactly one of the following:*

(i) empty,

(ii) of length $< \pi$, with $m_{\theta,1} = 0$ on its entirety, or

(iii) of length $\le \pi$, with $m_{\theta,1}$ strictly concave (and hence strictly positive) on its interior.

The length of the interval depends on μ as well as on α.

Proof. In any case, Corollary 4.10 implies that $\min m_{\theta,1} < 0$. Henceforth assume case (i) does not hold, so the set K of points where $m_{\theta,1} \geq 0$ is nonempty. Because $m_{\theta,1}$ is continuous, the subset $K \subset \mathbb{R}/\alpha\mathbb{Z}$ is closed and $m_{\theta,1} = 0$ on its boundary.

First suppose that $\max m_{\theta,1} > 0$, the goal being to reach conclusion (iii). Then K contains distinct points A and B where $m_{\theta,1} = 0$. Corollary 4.10 implies that $|B - A| \leq \pi$. Lemma 4.3 implies that $m_{\theta,1}$ is strictly concave whenever $m_{\theta,1} > 0$. Hence we can and do assume that

$$m_{B,2} = \frac{d}{d\theta}\Big|_{\theta=B} m_{\theta,1} < 0 < \frac{d}{d\theta}\Big|_{\theta=A} m_{\theta,1} = m_{A,2},$$

using Lemma 4.3 and the fact that $m_{\theta,1} > 0$ whenever $\theta \in (A, B)$. Now (4.7) implies that

$$m_{\theta,1} = \begin{cases} m_{A,2} \sin(\theta - A) - \displaystyle\int_A^\theta w(\psi) \sin(\theta - \psi)\, d\psi & \text{if } \theta \in (A - \pi, A) \\ m_{B,2} \sin(\theta - B) - \displaystyle\int_B^\theta w(\psi) \sin(\theta - \psi)\, d\psi & \text{if } \theta \in (B, B + \pi), \end{cases} \tag{4.10}$$

and both of these are strictly negative. Since also $m_{\theta,1} < 0$ for all $\theta \notin (\theta' - \pi, \theta' + \pi)$ for all $\theta' \in [A, B]$ by Corollary 4.10, conclusion (iii) follows when $\max m_{\theta,1} > 0$.

Finally, assume $\max m_{\theta,1} = 0$. Fix a left boundary point A a right boundary point B of K; note that $A = B$ is possible. Corollary 4.10 again teaches that $B - A \leq \pi$. By hypothesis,

$$m_{B,2} = \frac{d}{d\theta}\Big|_{\theta=B} m_{\theta,1} = 0 = \frac{d}{d\theta}\Big|_{\theta=A} m_{\theta,1} = m_{A,2}.$$

Hence (4.7) takes the forms

$$m_{\theta,1} = \int_\theta^A w(\psi) \sin(\theta - \psi)\, d\psi \quad \text{and} \quad m_{\theta,1} = -\int_B^\theta w(\psi) \sin(\theta - \psi)\, d\psi.$$

These formulas, plus the choices of A and B as left and right endpoints, imply that $m_{\theta,1} < 0$ for all $\theta \in [A - \pi, A) \cup (B, B + \pi]$. In words, every left endpoint of K is preceded by, and every right endpoint of K is followed by, an interval of length at least π on which $m_{\theta,1} < 0$. Since $|A - B| \leq \pi$, the interval $[A, B]$ contains no endpoints of K other than A and B themselves. Therefore $m_{\theta,1} = 0$ for all $\theta \in [A, B]$. Corollary 4.10 prevents $m_{\theta,1} \geq 0$ for θ outside of $[A - \pi, B + \pi]$. Except for showing the strict inequality $|B - A| < \pi$, this completes the proof that $\max m_{\theta,1} = 0$ forces conclusion (ii).

Suppose, then, that $|B - A| = \pi$. Corollary 4.4 implies that $\mu(C_{[A,B]}) = 1$. If θ^* is the midpoint of the interval $[A, B]$, then the measure $\tilde{\mu}_{\theta^*}$ is supported on the half-space $H^+ = \{(z_1, z_2) \in \mathbb{R}^2 \mid z_1 \geq 0\}$. But $m_{1,\theta} = 0$ for all $\theta \in [A, B]$, whence $\tilde{\mu}_{\theta^*}$ is actually supported on a single line ∂H^+. This contradicts the non-degeneracy hypothesis. Therefore $|B - A| < \pi$, as desired. □

Corollary 4.12. *If $\max_\theta m_{\theta,1} > 0$, then there is a unique angle θ^* at which the maximum is attained: $m_{\theta^*,1} = \max_\theta m_{\theta,1}$. Furthermore, $m_{\theta^*,2} = 0$ for that angle.*

Proof. The claim concerning $m_{\theta^*,1}$ is immediate from the concavity in Proposition 4.11. The fact that $m_{\theta^*,2} = 0$ follows from the first claim of Lemma 4.3: $m_{\theta,2} = \frac{d}{d\theta} m_{\theta,1}$. □

Corollary 4.13. *Assume square-integrability (1.3). If $\max_\theta m_{\theta,1} \leq 0$ then $\Gamma(p)$ attains its minimum at the unique point $\bar{b} = \mathbf{0}$. If $\max_\theta m_{\theta,1} > 0$, then $\Gamma(p)$ attains its minimum at the unique point $\bar{b} = (m_{\theta^*,1}, \theta^*)$, where*

$$\theta^* = \arg\max_\theta m_{\theta,1}.$$

In either case, the mean of μ in Definition 1.11 coincides with the barycenter of μ.

Proof. Use the explicit expression for $\Gamma(p)$ from Lemma 4.1 and Corollary 4.12; minimize over r and θ. $\square$

Corollary 4.14. *Assume square-integrability (1.3). If there is $(r',\theta')\in\mathcal{K}$ with $r'\geq 0$ and $m_{\theta'}=(r',0)$ then $\bar{b}=(r',\theta')$.*

Proof. When $r'=m_{\theta',1}>0$, Proposition 4.11(iii) holds, and θ' lies interior to the closed interval $[A,B]$ there. Due to Corollary4.13, $m_{\theta,1}$ attains a unique maximum at $\theta^*\in(A,B)$. Moreover, $m_{\theta'}=(r',0)$ and $m_{\theta',2}=0$ imply, with Lemma 4.3, that

$$\frac{d}{d\theta}m_{1,\theta}\Big|_{\theta=\theta'}=m_{\theta',2}=0. \tag{4.11}$$

By strict concavity in Proposition 4.11(iii), $\theta^*=\theta'$, so $\bar{b}=(r',\theta')$ by Corollary 4.13.

The case $r'=m_{\theta',1}=0$ can only occur in cases (ii) and (iii) of Proposition 4.11, with θ' being an endpoint of the closed interval in case (iii) and anywhere in the closed interval in case (ii). Since (4.11) holds nonetheless, strict concavity in case (iii) cannot be. Consequently, $m_{\theta,1}\leq 0$ for all θ. Therefore, by Corollary 4.13, $\bar{b}=(0,\theta')=(r',\theta')$. $\square$

We conclude this section with important estimates relating folded averages $\eta_{\theta,N}$ from (1.7) to folded barycenters $F_\theta b_N$ of empirical distributions on $\mathcal{K}$.

Lemma 4.15. *Suppose that $b_N=(\hat{r},\hat{\theta})$ with $\hat{r}>0$. If $\theta\in\mathbb{R}/\alpha\mathbb{Z}$ and $|\theta-\hat{\theta}|\leq\pi$, then*

$$\|\eta_{\theta,N}-F_\theta b_N\|\leq\frac{|\theta-\hat{\theta}|}{N}\sum_{p_n\in\mathcal{I}_{(\theta,\hat{\theta})}}d(\mathbf{0},p_n).$$

In particular, $\eta_{\hat{\theta},N}=F_{\hat{\theta}}b_N$. Also, if $p_n\notin\mathcal{I}_{(\theta,\hat{\theta})}$ for all $n=1,\dots,N$, then $\eta_{\theta,N}=F_\theta b_N$.

Proof. This is a consequence of Lemma 4.8 applied to the measure μ^N and the associated first moments $m^N_\theta=\eta_{N,\theta}$. By Corollary 4.13, $b_N=(m^N_{\hat{\theta},1},\hat{\theta})$ and $m^N_{\hat{\theta},2}=0$ because $\hat{r}>0$. Hence $F_{\hat{\theta}}b_N=(m^N_{\hat{\theta},1},0)=m^N_{\hat{\theta}}$. As $|\theta-\hat{\theta}|\leq\pi$, in polar coordinates $F_\theta b_N=(\hat{r},\hat{\theta}-\theta)=\Phi_{\hat{\theta}-\theta}F_{\hat{\theta}}b_N$, so

$$\|F_\theta b_N-\eta_{\theta,N}\|=\|\Phi_{\hat{\theta}-\theta}F_{\hat{\theta}}b_N-\eta_{\theta,N}\|=\|\Phi_{\hat{\theta}-\theta}m^N_{\hat{\theta}}-m^N_\theta\|.$$

Therefore, by Lemma 4.8,

$$\|F_\theta b_N-\eta_{\theta,N}\|\leq|\theta-\hat{\theta}|\int_{\mathcal{I}_{(\theta,\hat{\theta})}}d(\mathbf{0},p)\mu_N(p)=\frac{|\theta-\hat{\theta}|}{N}\sum_{p_n\in\mathcal{I}_{(\theta,\hat{\theta})}}d(\mathbf{0},p_n). \qquad\square$$

The following is a special version of Corollary 4.14.

Corollary 4.16. *If $\eta_{\theta',N}=(r',0)\in\mathbb{R}^2$ with $r'\geq 0$, then $b_N=(r',\theta')$.*

5 Proof of the sticky law of large numbers

The standard law of large numbers for folded averages in $\mathbb{R}^2$ states that $m^N_\theta\to m_\theta$ as $N\to\infty$. It holds uniformly in θ, as follows.

Lemma 5.1. *For any $\epsilon>0$, there is a random integer $N^*_\epsilon(\omega)$ such that $\max_{\theta\in\mathbb{Z}/\alpha\mathbb{Z}}\|m^N_\theta-m_\theta\|\leq\epsilon$ for all $N\geq N^*_\epsilon(\omega)$, $\mathbb{P}$ almost surely.*

Proof. Fix $\epsilon > 0$ and an integer $n > \max\left(24\alpha\bar{r}/\epsilon, \alpha/\pi\right)$, and let $\theta_k = \alpha k/n + \alpha\mathbb{Z}$ for $k = 0, \dots, n-1$. Then $|\theta_k - \theta_{k+1}| = \alpha/n < \pi$. For any $\theta \in [\theta_k, \theta_{k+1})$,

$$\|\Phi_{\theta_k - \theta} z - z\| \leq \frac{\alpha\|z\|}{n} \text{ for any } z \in \mathbb{R}^2$$

as well as by Lemma 4.8,

$$\|\Phi_{\theta_k - \theta} m_{\theta_k} - m_\theta\| \leq \frac{\alpha\bar{r}}{n}. \tag{5.1}$$

Hence, making also use of (4.6),

$$\|m_{\theta_k} - m_\theta\| \;\leq\; \|m_{\theta_k} - \Phi_{\theta_k - \theta} m_{\theta_k}\| + \|\Phi_{\theta_k - \theta} m_{\theta_k} - m_\theta\| \;\leq\; 4\,\frac{\alpha\bar{r}}{n}. \tag{5.2}$$

By the law of large numbers (1.9), there is nullset $\mathcal{M}_1$ and an integer $N_1(\omega)$ such that $\|m^N_{\theta_k} - m_{\theta_k}\| \leq \epsilon/3$ for all $N \geq N_1(\omega)$, all $k \in \{0, \dots, n-1\}$, and all $\omega \in \Omega \setminus \mathcal{M}_1$. Similarly, by the law of large numbers there is also a nullset $\mathcal{M}_2$ and an integer $N_2(\omega)$ such that

$$0 \leq \overline{r^N} := \int_{\mathcal{K}} d(\mathbf{0}, p)\, d\mu_N(p) \leq 2\bar{r}$$

for all $N \geq N_2(\omega)$ and all $\omega \in \Omega \setminus \mathcal{M}_2$. Applying (5.2) to the empirical moments gives thus

$$\|m^N_{\theta_k} - m^N_\theta\| \leq \frac{4\alpha\overline{r^N}}{n} \leq 8\frac{\alpha\bar{r}}{n}$$

for all $N \geq N_2(\omega)$. Finally,

$$\begin{aligned}\|m^N_\theta - m_\theta\| &\leq \|m^N_\theta - m^N_{\theta_k}\| + \|m^N_{\theta_k} - m_{\theta_k}\| + \|m_{\theta_k} - m_\theta\| \\ &\leq 8\frac{\alpha\bar{r}}{n} + \frac{\epsilon}{3} + 4\frac{\alpha\bar{r}}{n} < \epsilon\end{aligned}$$

for all $N \geq N^*_\epsilon(\omega) = \max(N_1(\omega), N_2(\omega))$ and $\Omega \ni \omega \notin \mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2$. $\square$

Given a set of angles $T \subset \mathbb{R}/\alpha\mathbb{Z}$, define the set

$$C^+_T = \{(r, \theta) \in \mathcal{K} \mid r > 0 \text{ and } \theta \in T\} = C_T \setminus \{\mathbf{0}\}, \tag{5.3}$$

which excludes the origin from the sector C_T (Definition 1.2).

Theorem 5.2. *Let $T \subset \mathbb{R}/\alpha\mathbb{Z}$ be a closed subset such that $m_{\theta,1} < 0$ for all $\theta \in T$. Then there is a random integer $N^*(\omega)$ such that*

$$b_N(\omega) \notin C^+_T \textit{ for all } N \geq N^*(\omega)$$

holds $\mathbb{P}$-almost surely. In particular, if μ is fully sticky then there is a random integer $N^*(\omega)$ such that $b_N = \mathbf{0}$ for all $N \geq N^*(\omega)$, $\mathbb{P}$-almost surely. Similarly, if μ is partly sticky and $T \subset \mathbb{R}/\alpha\mathbb{Z}$ is any open interval containing the maximal interval where $m_{\theta,1} = 0$, as described in Propositions 1.10 and 4.11, then $b_N \in C_T$ for all $N \geq N^*(\omega)$, $\mathbb{P}$-almost surely.

Proof. Since T is closed and $m_{\theta,1}$ is continuous, there is $\epsilon > 0$ such that $\sup_{\theta\in T} m_{\theta,1} < -\epsilon < 0$. By Lemma 5.1 there is a random integer $N^*(\omega)$ such that $m^N_{\theta,1} < -\epsilon/2$ for all $\theta \in T$, almost surely for all $N \geq N^*(\omega)$. Now, b_N is the unique minimizer of

$$p \mapsto \Gamma_N(p) = \frac{1}{2}\int_{\mathcal{K}} d(p, q)^2\, d\mu_N(q).$$

Since the empirical measures $\{\mu_N\}_{N=1}^{\infty}$ are square-integrable (even if μ is not),

$$\Gamma_N(r,\theta) = \frac{r^2}{2} - r\, m_{\theta,1}^N + \Gamma_N(0) \tag{5.4}$$

by Lemma 4.1. Therefore, if $\theta \in T$, and $r > 0$, and $N \geq N^*(\omega)$, then almost surely

$$\Gamma_N(r,\theta) > \frac{r^2}{2} + \frac{\epsilon}{2} r + \Gamma_N(0) > \Gamma_N(0).$$

Hence the minimizer b_N lies outside of C_T^+ almost surely. □

By a very similar argument, Corollary 4.13 and Lemma 5.1 together imply the following, which we state without proof. It also is a consequence of the strong law of [22].

Theorem 5.3. *Suppose that* $\max_\theta m_{\theta,1} = m_{\theta^*,1} > 0$. *Let* $T \subset \mathbb{R}/\alpha\mathbb{Z}$ *be any open interval of length* $\leq \pi$ *containing* θ^*. *There is a random integer* $N^*(\omega)$ *such that*

$$b_N(\omega) \in C_T^+ \text{ for all } N \geq N^*(\omega)$$

holds $\mathbb{P}$*-almost surely. In particular, if* μ *is nonsticky then for any* $\epsilon \in (0, \pi/2)$*, the empirical barycenter* b_N *lies in* $C^+_{(\theta^*-\epsilon,\theta^*+\epsilon)}$ *for all* $N > N^*(\omega)$, $\mathbb{P}$*-almost surely.*

We now give the proof the law of large numbers on $\mathcal{K}$ (Theorem 1.12) by collecting various results we have already proved.

Proof of Theorem 1.12. The fully sticky case is immediate from Theorem 5.2. Consider the partly sticky case. By Corollary 4.13 applied to the empirical measure μ_N, the empirical barycenter is $b_N = \mathbf{0}$ or $b_N = (m^N_{\theta^*,1}, \theta^*)$ where θ^* maximizes $\theta \mapsto m^N_{\theta,1}$. Combining this fact with Lemma 5.1 leads to the conclusion that

$$\limsup_{N\to\infty} d(b_N, \mathbf{0}) = \limsup_{N\to\infty} m^N_{\theta^*,1} \leq \max_\theta m_{1,\theta}$$

holds $\mathbb{P}$-almost surely. In the partly sticky case, $m_{1,\theta} \leq 0$ for all θ. Thus $b_N \to \mathbf{0}$ holds $\mathbb{P}$-almost surely. The other statements in the partly sticky case follow from Theorem 5.2.

Finally, consider the nonsticky case. Convergence $b_N \to \bar{b}$ again follows from the representation $b_N = (m^N_{\theta^*,1}, \theta^*)$ where θ^* maximizes $\theta \mapsto m^N_{\theta,1}$. By Lemma 5.1 $\mathbb{P}$-almost surely any maximizer θ^N of $\theta \mapsto m^N_{\theta,1}$ converges, as $N \to \infty$, to the maximizer of $\theta \mapsto m_{\theta,1}$, which is unique in the nonsticky case. By definition of $\bar{b}$, this implies that $b_N \to \bar{b}$, $\mathbb{P}$-almost surely. □

6 Proofs of the central limit theorems

This section contains proofs of the three central limit theorems: Theorem 1.13, Theorem 1.14, and Theorem 1.15. First comes the fully sticky case, which follows almost immediately from Theorem 1.12.

Proof of Theorem 1.13. Let N^* be the random integer from Theorem 1.12, which has the property that, $\mathbb{P}$-almost surely, $b_N = 0$ for all $N \geq N^*(\omega)$. If $\phi : \mathcal{K} \to \mathbb{R}$ is any bounded function then

$$\begin{aligned}\Big| \int \phi(p)\, d\nu_N(p) - \phi(\mathbf{0}) \Big| &= \big| \mathbb{E}\phi(b_N) - \phi(\mathbf{0}) \big| \\ &= \big| \mathbb{E}(\phi(b_N) - \phi(\mathbf{0})) \mathbf{1}_{N<N^*} \big| \leq 2 \big(\sup_{p\in\mathcal{K}} |\phi(p)| \big) \mathbb{P}(N < N^*).\end{aligned}$$

Since N^* is almost surely finite, $\mathbb{P}(N < N^*) \to 0$ as $N \to \infty$ which concludes the proof. Since the bound on the right hand side depends only on the supremum norm of ϕ, the bound also implies convergence in the total variation norm. □

Next comes the proof of the central limit theorem in the partly sticky case.

Proof of Theorem 1.14. Let $K = [A, B]$ be the interval on which $m_{\theta,1} = 0$, so $m_{\theta,1} < 0$ for all $\theta \notin [A, B]$ by hypothesis. Recall that θ^* is the midpoint of this interval. Let $\epsilon \in (0, \pi/4)$. By Theorem 5.2 there is an integer $N^*(\omega, \epsilon)$ such that, almost surely, $b_N(\omega) \in C_{[A_\epsilon, B^\epsilon]}$ if $N \geq N^*(\omega, \epsilon)$, where $A_\epsilon = A - \epsilon$ and $B^\epsilon = B + \epsilon$. Since ν_N is the distribution of the random variable $\sqrt{N} b_N$ on $\mathcal{K}$,

$$\lim_{N\to\infty} \nu_N(C_{[A_\epsilon, B^\epsilon]}) = 1.$$

Therefore

$$\begin{aligned}
&\lim_{N\to\infty} \left(\int_{\mathcal{K}} \phi_1(p)\, d\nu_N(p) - \int_{\mathcal{K}} \phi_2(p)\, d\nu_N(p) \right) \\
&\qquad = \lim_{N\to\infty} \left(\int_{C_{[A_\epsilon, B^\epsilon]}} \phi_1(p)\, d\nu_N(p) - \int_{C_{[A_\epsilon, B^\epsilon]}} \phi_2(p)\, d\nu_N(p) \right)
\end{aligned}$$

holds for any bounded continuous function $\phi_1, \phi_2 : \mathcal{K} \to \mathbb{R}$. For this reason it suffices to prove (1.16) for continuous bounded functions differing only on $C_{[A_\epsilon, B^\epsilon]}$. Such functions are of the form $\phi = \varphi \circ F_{\theta^*}$ where $\varphi : \mathbb{R}^2 \to \mathbb{R}$ is continuous and bounded.

Using the convex projection $\hat{P}_\rho$ from (1.14) for $\rho = \frac{1}{2}|A - B|$, let ζ_N denote the measure on $\mathbb{R}^2$ defined by $\mathbb{P}(\sqrt{N} \hat{P}_\rho(\eta_{\theta^*,N}) \in W) = \zeta_N(W)$ for Borel sets $W \subset \mathbb{R}^2$. Then $\mathbb{E}[\eta_{\theta^*,N}] = 0$, because $m_\theta = \mathbb{E}[\eta_{\theta,N}]$ for all $\theta \in \mathbb{R}/\alpha\mathbb{Z}$ and $m_{\theta^*} = 0$ by hypothesis. Recalling Remark 4.2, which guarantees square-integrability, the standard CLT for $\eta_{\theta^*,N}$ in $\mathbb{R}^2$ implies that the law of $\sqrt{N}\eta_{\theta^*,N}$ converges to g, the law of the multivariate normal with covariance (1.12). Thus

$$\lim_{N\to\infty} \int_{\mathbb{R}^2} \varphi(z) d\zeta_N(z) = \int_{\mathbb{R}^2} \varphi(z) d(g \circ \hat{P}_\rho^{-1}(z)) \tag{6.1}$$

holds for any continuous bounded function $\varphi : \mathbb{R}^2 \to \mathbb{R}$. We claim that for any $\delta > 0$ there is an integer N_δ such that

$$\mathbb{P}(\sqrt{N} \| F_{\theta^*} b_N - \hat{P}_\rho \eta_{\theta^*,N} \| > \delta) \leq \delta \tag{6.2}$$

holds for all $N \geq N_\delta$. This estimate and (6.1) imply that

$$\lim_{N\to\infty} \int_{\mathbb{R}^2} \varphi(z) d\tilde{\nu}_N(z) = \int_{\mathbb{R}^2} \varphi(z) d(g \circ \hat{P}_\rho^{-1}(z)) \tag{6.3}$$

where $\tilde{\nu}_N = \nu_N \circ F_{\theta^*}^{-1}$ is the law of $\sqrt{N} F_{\theta^*} b_N$ on $\mathbb{R}^2$.

Recall that $F_{\theta^*} : C_{[A_\epsilon, B^\epsilon]} \to D_{\rho+\epsilon}$ is bijective, where the sector $D_{\rho+\epsilon} \subset \mathbb{R}^2$ is defined by replacing ρ with $\rho + \epsilon$ in (1.13), and $\nu_N(C_{[A_\epsilon, B^\epsilon]}) \to 1$ as $N \to \infty$. Combining this with (6.3) leads to the conclusion that (1.16) holds for the continuous bounded function $\phi = \varphi \circ F_{\theta^*}$:

$$\begin{aligned}
\lim_{N\to\infty} \int_{\mathcal{K}} \varphi(F_{\theta^*}(p)) d\nu_N(p) &= \lim_{N\to\infty} \int_{\mathbb{R}^2} \varphi(z) d\tilde{\nu}_N(z) \\
&= \int_{\mathbb{R}^2} \varphi(z) d(g \circ \hat{P}_\rho^{-1}(z)) \\
&= \int_{\mathcal{K}} \varphi(F_{\theta^*}(p)) d(g \circ \hat{P}_\rho^{-1} \circ F_{\theta^*}(p)).
\end{aligned}$$

It remains to prove (6.2) by estimating $\| F_{\theta^*} b_N - \hat{P}_\rho \eta_{\theta^*,N} \|$.

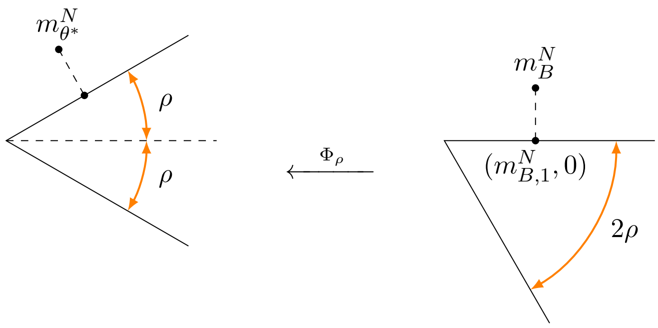

Figure 6: Detail for the proof of (6.2): Convex projection commutes with rotation.

First, suppose $b_N = (r, \hat\theta) \in C^+_{[A,B]}$. If $A = B$ then $\hat\theta = \theta^*$ and thus $\eta_{\theta^*,N} = F_{\theta^*} b_N = \hat P_\rho F_{\theta^*} b_N$ by Lemma 4.15. Now assume $A \neq B$. Then $\mu(\mathcal{I}_\theta) = 0$ for all $\theta \in [\theta^*, \hat\theta]$ by Corollary 4.4, as by hypothesis $m_\theta = 0$ for all $\theta \in [\theta^*, \hat\theta]$ and $|\hat\theta - \theta^*| \le |B - A| < \pi$. This implies that also $\nu_N(\mathcal{I}_\theta) = 0$ for all $\theta \in [\theta^*, \hat\theta]$. Since $r > 0$, Lemma 4.15 implies that $\eta_{\theta^*,N} = F_{\theta^*} b_N = \hat P_\rho F_{\theta^*} b_N$, as desired.

For the remainder of the proof, let $\epsilon \in (0, \pi/4)$ and assume $b_N \in C_{[A_\epsilon, B^\epsilon]}$ but $b_N \notin C^+_{[B,A]}$. Suppose $b_N = (r, \hat\theta)$ with $\hat\theta \in [B, B^\epsilon]$ and $r \ge 0$; the case $\hat\theta \in [A_\epsilon, A]$ is treated in the same way. By Corollary 4.13 and Lemma 4.3, $m^N_{\hat\theta,1} = r$ and $m^N_{\hat\theta,2} = 0$. Denote by

$$w_N(s) = \frac{1}{N} \sum_{p_n \in \mathcal{I}_s} d(\mathbf{0}, p_n),$$

the sample analog of $w(s)$ from (4.9). Utilizing the second equation in Corollary 4.7,

$$m^N_{B,2} = m^N_{\hat\theta,1} \sin(\hat\theta - B) + \int_B^{\hat\theta} w_N(\psi) \cos(\psi - B)\, d\psi,$$

which implies that $m^N_{B,2} \ge 0$. Moreover, by the first equation of Corollary 4.7,

$$m^N_{B,1} = m^N_{\hat\theta,1} \cos(\hat\theta - B) - \int_B^{\hat\theta} w_N(\psi) \sin(\psi - B)\, d\psi.$$

Therefore $m^N_{\hat\theta,1} \ge m^N_{B,1} \ge 0$. Similarly also

$$\begin{aligned} m^N_{\hat\theta,1} &= m^N_{B,1} \cos(\hat\theta - B) + m^N_{B,2} \sin(\hat\theta - B) - \int_B^{\hat\theta} w(\psi) \sin(\hat\theta - \psi)\, d\psi \\ &\le m^N_{B,1} + \epsilon m^N_{B,2}. \end{aligned}$$

This shows that $r \in [m^N_{B,1}, m^N_{B,1} + \epsilon m^N_{B,2}]$. For later use, note that for $\epsilon > 0$ sufficiently small,

$$m^N_{B,2} \le (m^N_{\hat\theta,1} + \bar r)\epsilon \le (m^N_{B,1} + \bar r)\epsilon + m^N_{B,2}\epsilon^2 \le 3\bar r \epsilon. \tag{6.4}$$

Observe that $\Phi_\rho m^N_B = m^N_{\theta^*}$. If $A = B$ this is obvious because $\rho = 0$ and $\theta^* = B$. If $A \neq B$, then this follows from Lemma 4.8 because $\nu^N(\mathcal{I}_\theta) = 0$ for all $\theta \in [\theta^*, B]$, due to $\mu(I_\theta) = 0$. Therefore $\hat P_\rho m^N_{\theta^*} = \hat P_\rho \Phi_\rho m^N_B = (m^N_{B,1}, \rho)$ in polar coordinates, because convex projection commutes with rotation, cf. Figure 6. In conjunction with $F_{\theta^*} b_N = (r, \hat\theta - \theta^*)$, therefore

$$\begin{aligned}
\|F_{\theta^*} b_N - \hat{P}_\rho m^N_{\theta^*}\|^2 &= \big(r\cos(\hat\theta-\theta^*) - m^N_{B,1}\cos\rho\big)^2 + \big(r\sin(\hat\theta-\theta^*) - m^N_{B,1}\sin\rho\big)^2 \\
&= r^2 + (m^N_{B,1})^2 - 2rm^N_{B,1}\cos(\hat\theta - B) \\
&= (r - m^N_{B,1})^2 + 2rm^N_{B,1}\big(1-\cos(\hat\theta-B)\big) \\
&\le (\epsilon m^N_{B,2})^2 + (m^N_{B,1} + \epsilon m^N_{B,2}) m^N_{B,1}\epsilon^2 \\
&= \epsilon^2(m^N_{B,1} + m^N_{B,2})^2 - \epsilon^2 m^N_{B,1} m^N_{B,2}(2-\epsilon).
\end{aligned}$$

By applying the same argument when $\hat\theta \in [A_\epsilon, A]$, upon noting that $m^N_{A,1}, m^N_{A,2} \le 0$, we conclude that for ϵ sufficiently small and $b_N \in C_{[A_\epsilon,A)\cup(B,B^\epsilon]}$,

$$\|F_{\theta^*} b_N - \hat{P}_\rho m^N_{\theta^*}\| \le \epsilon\left(m^N_{B,1} + m^N_{B,2} - m^N_{A,1} - m^N_{A,2}\right).$$

Let $X_N = m^N_{B,1} + m^N_{B,2} - m^N_{A,1} - m^N_{A,2}$; each term in this sum is the average of N independent random variables in $\mathbb{R}^2$, and each term has zero mean since $\mathbb{E}(m^N_A) = m_A = 0$ and $\mathbb{E}(m^N_B) = m_B = 0$, by hypothesis. The Chebychev inequality implies

$$\begin{aligned}
\mathbb{P}\Big(b_N \in C_{[A_\epsilon,B^\epsilon]}, \sqrt{N}\big\|F_{\theta^*} b_N - \hat{P}_\rho \eta_{\theta^*,N}\big\| > \delta\Big) &\le \mathbb{P}\Big(\sqrt{N}\epsilon|X_N| > \delta\Big) \\
&\le \frac{\epsilon^2\mathbb{E}\big(X_N^2\big)N}{\delta^2} \\
&\le \frac{\delta}{2} \text{ for } \epsilon = \sqrt{C\delta^3/2} \qquad (6.5)
\end{aligned}$$

by square-integrability with a constant C that depends only on μ.

By Theorem 5.2 there is an integer $N^*(\omega,\epsilon)$ such that $b_N \in C_{[B_\epsilon,A^\epsilon]}$ if $N \ge N^*(\omega,\epsilon)$ for almost surely all ω. In particular, given $\delta > 0$ there is an integer $N_{\epsilon,\delta}$ such that

$$\mathbb{P}(b_N \in C_{[B_\epsilon,A^\epsilon]}) \ge 1 - \delta/2 \text{ for all } N \ge N_{\epsilon,\delta}.$$

Setting $N_\delta = N_{\epsilon,\delta}$ for $\epsilon = \sqrt{C\delta^3/2}$ with (6.5), the above yields the desired claim (6.2). □

We conclude with the proof of the central limit theorem for the nonsticky case.

Proof of Theorem 1.15. In the nonsticky case, the barycenter of μ, denoted $\bar b$, is equal to $(r^*,\theta^*) \in \mathcal{K}$ where $r^* = m_{\theta^*,1} > 0$ and θ^* is the unique angle that maximizes $\theta \mapsto m_{\theta,1}$. By Theorem 5.3, $b_N \in C^+_{[\theta^*-\epsilon,\theta^*+\epsilon]}$ for all N sufficiently large, given any fixed $\epsilon \in (0,\pi/2)$.

The standard CLT for $m^N_{\theta^*}$ in $\mathbb{R}^2$ implies that the law of $\sqrt{N}(m^N_{\theta^*} - F_{\theta^*}\bar b)$ converges weakly to g. In cartesian coordinates, $F_{\theta^*}\bar b = (e_1 \cdot F_{\theta^*}\bar b, e_2 \cdot F_{\theta^*}\bar b) = (r^*, 0)$. Therefore, to show that the law of the random vector $\sqrt{N}\big(e_1\cdot F_{\theta^*}b_N - r^*, (1+\kappa)e_2\cdot F_{\theta^*}b_N\big)$ also converges weakly to g as $N\to\infty$ (the random variable $\kappa \ge 0$ was defined at (1.17)), it suffices to show that for any $\delta > 0$,

$$\lim_{N\to\infty}\mathbb{P}\Big(\sqrt{N}\big\|\big(e_1\cdot F_{\theta^*}b_N, (1+\kappa)e_2\cdot F_{\theta^*}b_N\big) - m^N_{\theta^*}\big\| \ge \delta\Big) = 0. \qquad (6.6)$$

Recall from (5.4) that the empirical mean b_N is the unique minimizer of

$$(r,\theta) \mapsto \Gamma(r,\theta) = \frac{r^2}{2} - rm^N_{\theta,1} + \Gamma_N(0).$$

That is, if $b_N = (\hat r, \hat\theta)$, then $\hat\theta$ is the unique maximizer of the function

$$\theta \mapsto f(\theta) = \frac{(m^N_{\theta,1})^2}{2}.$$

The first objective is to show that $|\hat{\theta} - \theta^*| = O(1/\sqrt{N})$, meaning that for any $\epsilon > 0$ there are constants N_ϵ, $C_\epsilon > 0$ such that $\mathbb{P}\left(\sqrt{N}|\hat{\theta} - \theta^*| > C_\epsilon\right) \le \epsilon$ for all $N \ge N_\epsilon$. Using Corollary 4.7, write $m^N_{\theta,1}$ in terms of θ^*:

$$m^N_{\theta,1} = m^N_{\theta^*,1}\cos(\theta - \theta^*) + m^N_{\theta^*,2}\sin(\theta - \theta^*) - \int_{\theta^*}^{\theta} w_N(\psi)\sin(\theta - \psi)\,d\psi, \tag{6.7}$$

where

$$w_N(\psi) = \frac{1}{N}\sum_{p_n \in \mathcal{I}_\psi} d(\mathbf{0}, p_n).$$

Because $m^N_{\theta^*}$ satisfies the central limit theorem and because $m_{\theta^*,2} = 0$, this implies that

$$\begin{aligned} m^N_{\theta,1} &= m_{\theta^*,1}\cos(\theta - \theta^*) + O(1/\sqrt{N})\cos(\theta - \theta^*) \\ &\quad + O(1/\sqrt{N})\sin(\theta - \theta^*) - \int_{\theta^*}^{\theta} w_N(\psi)\sin(\theta - \psi)\,d\psi. \end{aligned} \tag{6.8}$$

For $|\theta - \theta^*| < \pi/2$ the function

$$\theta \mapsto \int_{\theta^*}^{\theta} -w_N(\psi)\sin(\theta - \psi)\,d\psi \le 0$$

has a maximum at $\theta = \theta^*$. In view of this and (6.8), we conclude that the angle $\hat{\theta}$ at which the maximum in $\theta \mapsto m^N_{\theta,1}$ is attained must satisfy $|\hat{\theta} - \theta^*| \le O(1/\sqrt{N})$.

Now we compare $F_{\theta^*}b_N$ to $m^N_{\theta^*}$ to derive (6.6). Recall that $\Phi_\sigma : \mathbb{R}^2 \to \mathbb{R}^2$ denotes rotation by angle σ (defined just before Lemma 4.8). When $|\hat{\theta} - \theta^*| < \pi$ (which happens almost surely as $N \to \infty$) we have $F_{\theta^*}b_N = \Phi_{\hat{\theta}-\theta^*}F_{\hat{\theta}}b_N$. Therefore, by Lemma 4.15, we have

$$F_{\theta^*}b_N = \Phi_{\hat{\theta}-\theta^*}F_{\hat{\theta}}b_N = \Phi_{\hat{\theta}-\theta^*}m^N_{\hat{\theta}} \tag{6.9}$$

for N large enough. By Corollary 4.7 we also have

$$m^N_{\hat{\theta}} = \Phi^{-1}_{\hat{\theta}-\theta^*}m^N_{\theta^*} - V \tag{6.10}$$

where $V = (V_{1,N}, V_{2,N})$ is the vector with components

$$V_{1,N} = \int_{\theta^*}^{\hat{\theta}} w_N(\psi)\sin(\hat{\theta} - \psi)\,d\psi, \qquad V_{2,N} = \int_{\theta^*}^{\hat{\theta}} w_N(\psi)\cos(\hat{\theta} - \psi)\,d\psi.$$

Hence

$$e_1 \cdot F_{\theta^*}b_N - e_1 \cdot m^N_{\theta^*} = -e_1 \cdot \Phi_{\hat{\theta}-\theta^*}V = -\cos(\hat{\theta} - \theta^*)V_{1,N} + \sin(\hat{\theta} - \theta^*)V_{2,N} \tag{6.11}$$

and

$$e_2 \cdot F_{\theta^*}b_N - e_2 \cdot m^N_{\theta^*} = -e_2 \cdot \Phi_{\hat{\theta}-\theta^*}V = \sin(\hat{\theta} - \theta^*)V_{1,N} - \cos(\hat{\theta} - \theta^*)V_{2,N} \tag{6.12}$$

for N sufficiently large. Using the fact that $|\hat{\theta} - \theta^*| \le O(1/\sqrt{N})$, we find that $|V_{1,N}| = O(1/N)$ and $|V_{2,N}| = O(1/\sqrt{N})$: indeed,

$$0 \le \inf_\psi w_N(\psi) \le \sup_\psi w_N(\psi) \le \frac{1}{N}\sum_{n=1}^{N} d(\mathbf{0}, p_n)$$

and the latter converges to $\bar{r} < \infty$ (recall (1.2)) almost surely as $N \to \infty$. Hence, with probability one,

$$V_{1,N} \le 2\bar{r}\int_{\theta^*}^{\hat{\theta}} \sin(\hat{\theta} - \psi)\,d\psi \le \bar{r}(\hat{\theta} - \theta^*)^2, \quad \text{and} \qquad |V_{2,N}| \le 2\bar{r}|\hat{\theta} - \theta^*|$$

hold for all N large enough. Applying this at (6.11) and using $|\hat{\theta}-\theta^*| \le O(1/\sqrt{N})$, we obtain

$$e_1 \cdot F_{\theta^*} b_N - e_1 \cdot m^N_{\theta^*} = O(1/N),$$

by which we mean that for any $\epsilon > 0$, there is C_ϵ such that

$$\limsup_{N\to\infty} \mathbb{P}\left(|e_1 \cdot F_{\theta^*} b_N - e_1 \cdot m^N_{\theta^*}| \ge C_\epsilon/N\right) \le \epsilon. \tag{6.13}$$

In particular, $e_1 \cdot F_{\theta^*} b_N - e_1 \cdot m^N_{\theta^*}$ is $o(1/\sqrt{N})$ in the sense of (6.6).

To complete the proof of (6.6), we must show that $(1+\kappa)e_2 \cdot F_{\theta^*} b_N - e_2 \cdot m^N_{\theta^*}$ is $o(1/\sqrt{N})$, as well. We will use (6.12) and a more subtle estimate of $V_{2,N}$ and of $\hat{\theta}-\theta^*$. From (6.10) and the fact that $e_2 \cdot m^N_{\hat{\theta}} = m^N_{\hat{\theta},2} = 0$ (by Lemma 4.15), we have

$$\begin{aligned} 0 = m^N_{\hat{\theta},2} &= -m^N_{\theta^*,1}\sin(\hat{\theta}-\theta^*) + m^N_{\theta^*,2}\cos(\hat{\theta}-\theta^*) - V_{2,N} \\ &= -m^N_{\theta^*,1}(\hat{\theta}-\theta^*) + m^N_{\theta^*,2} + O(N^{-1}) - V_{2,N}. \end{aligned} \tag{6.14}$$

(We used $|\hat{\theta}-\theta^*| \le O(1/\sqrt{N})$ again in the last equality.) As the next lemma shows, the integral term $V_{2,N}$ is approximated by $r^*\kappa(\hat{\theta}-\theta^*)$, where the random variable κ was defined at (1.17): $r^*\kappa = w^-(\theta^*)$ if $\hat{\theta} > \theta^*$, and $r^*\kappa = w^+(\theta^*)$ if $\hat{\theta} < \theta^*$, and $r^*\kappa = 0$ if $\hat{\theta} = \theta^*$.

Lemma 6.1. *Let $\hat{\theta}$ be the angular coordinate of b_N. Let U^+_N be the event that $\hat{\theta} > \theta^*$, and let U^-_N be the event that $\hat{\theta} < \theta^*$. If Z_N is the random variable*

$$\begin{aligned} Z_N = \mathbb{I}_{U^+_N} \cdot &\left| \int_{\theta^*}^{\hat{\theta}} w_N(\psi)\cos(\hat{\theta}-\psi)\,d\psi - w^-(\theta^*)(\hat{\theta}-\theta^*) \right| \\ + \mathbb{I}_{U^-_N} \cdot &\left| \int_{\theta^*}^{\hat{\theta}} w_N(\psi)\cos(\hat{\theta}-\psi)\,d\psi - w^+(\theta^*)(\hat{\theta}-\theta^*) \right|, \end{aligned} \tag{6.15}$$

then Z_N is $o(1/\sqrt{N})$ in probability as $N\to\infty$: for any $\delta > 0$,

$$\lim_{N\to\infty} \mathbb{P}(Z_N > \delta/\sqrt{N}) = 0.$$

By combining Lemma 6.1 and (6.14), we derive

$$0 = -m^N_{\theta^*,1}(\hat{\theta}-\theta^*) + m^N_{\theta^*,2} - r^*\kappa(\hat{\theta}-\theta^*) + O(1/N) + o(1/\sqrt{N}), \tag{6.16}$$

and thus

$$\hat{\theta}-\theta^* = \frac{m^N_{\theta^*,2}}{m^N_{\theta^*,1} + r^*\kappa} + o(1/\sqrt{N}). \tag{6.17}$$

Recalling that $|V_{1,N}| = O((\hat{\theta}-\theta^*)^2) = O(1/N)$, we now combine (6.12) with Lemma 6.1 and (6.17) to obtain

$$\begin{aligned} e_2 \cdot F_{\theta^*} b_N - m^N_{\theta^*,2} &= -V_{2,N} + O(1/N) \\ &= -r^*\kappa(\hat{\theta}-\theta^*) + o(1/\sqrt{N}) \\ &= -m^N_{\theta^*,2}\frac{r^*\kappa}{m^N_{\theta^*,1} + r^*\kappa} + o(1/\sqrt{N}). \end{aligned} \tag{6.18}$$

In the case $w^\pm(\theta^*) = 0$, we have $\kappa = 0$, so (6.6) follows from (6.18) and (6.13). However, when $w^\pm(\theta^*) \ne 0$, (6.18) implies that

$$e_2 \cdot F_{\theta^*} b_N - m^N_{\theta^*,2} = -m^N_{\theta^*,2}\frac{r^*\kappa}{r^* + r^*\kappa} + o(1/\sqrt{N}),$$

because $m^N_{\theta^*,1} \to r^*$ and $m^N_{\theta^*,2} \to m_{\theta^*,2} = 0$ as $N \to \infty$. Therefore,

$$(1+\kappa)e_2 \cdot F_{\theta^*} b_N - m^N_{\theta^*,2} = o(1/\sqrt{N}).$$

This and (6.13) imply (6.6), as desired. Except for the proof of Lemma 6.1, the proof of Theorem 1.15 is complete. □

Proof of Lemma 6.1. We will restrict our attention to the case that $\hat{\theta} > \theta^*$ (in the event U_N^+, which is equivalent to $e_2 \cdot F_{\theta^*} b_N > 0$); the other case is analyzed in the same way. We decompose the integral as

$$\begin{aligned}
\int_{\theta^*}^{\hat{\theta}} w_N(\psi) \cos(\hat{\theta} - \psi)\, d\psi \quad &= \quad w^-(\theta^*) \int_{\theta^*}^{\hat{\theta}} \cos(\hat{\theta} - \psi)\, d\psi \\
&\quad + \left(w_N^-(\theta^*) - w^-(\theta^*)\right) \int_{\theta^*}^{\hat{\theta}} \cos(\hat{\theta} - \psi)\, d\psi \\
&\quad + \int_{\theta^*}^{\hat{\theta}} \left(w_N(\psi) - w_N^-(\theta^*)\right) \cos(\hat{\theta} - \psi)\, d\psi \\
&= \quad T_1 + T_2 + T_3, \qquad (6.19)
\end{aligned}$$

where

$$w_N^-(\psi) = \frac{1}{N} \sum_{p_n \in \mathcal{I}_\psi^-} d(\mathbf{0}, p_n).$$

Now we estimate each of the terms T_1, T_2, and T_3, in (6.19) using the fact that $\hat{\theta} - \theta^* = O(1/\sqrt{N})$, which was proved independently of this lemma. First,

$$\begin{aligned}
T_1 \quad &= \quad w^-(\theta^*) \int_{\theta^*}^{\hat{\theta}} \cos(\hat{\theta} - \psi)\, d\psi \\
&= \quad w^-(\theta^*) \left((\hat{\theta} - \theta^*) + O((\hat{\theta} - \theta^*)^3) \right) = w^-(\theta^*)(\hat{\theta} - \theta^*) + O(1/N).
\end{aligned}$$

For T_2, we apply the CLT to

$$\left(w_N^-(\theta^*) - w^-(\theta^*)\right) = \frac{1}{N} \sum_{n=1}^{N} \left(d(\mathbf{0}, p_n) \mathbb{I}_{\mathcal{I}_{\theta^*}^-}(p_n) - \mathbb{E}[d(\mathbf{0}, p_n) \mathbb{I}_{\mathcal{I}_{\theta^*}^-}(p_n)] \right)$$

which is a sum of independent, identically-distributed random variables with zero mean and finite variance (due to square integrability condition (1.3)). Hence $\mathrm{Var}\left(w_N^-(\theta^*) - w^-(\theta^*)\right) = O(1/N)$ and $\left(w_N^-(\theta^*) - w^-(\theta^*)\right) = O(1/\sqrt{N})$, which implies that

$$T_2 = \left(w_N^-(\theta^*) - w^-(\theta^*)\right) \int_{\theta^*}^{\hat{\theta}} \cos(\hat{\theta} - \psi)\, d\psi = O(1/\sqrt{N}) O(\hat{\theta} - \theta^*) = O(1/N).$$

Finally, we show that the term $T_3 = \int_{\theta^*}^{\hat{\theta}} \left(w_N(\psi) - w_N^-(\theta^*)\right) \cos(\hat{\theta} - \psi)\, d\psi$ is $o(1/\sqrt{N})$. If $\mathcal{I}_\psi \Delta \mathcal{I}_{\theta^*}^- = (\mathcal{I}_\psi \cup \mathcal{I}_{\theta^*}^-) \setminus (\mathcal{I}_\psi \cap \mathcal{I}_{\theta^*}^-)$ denotes the symmetric difference of the shadows, then

$$\left| w_N(\psi) - w_N^-(\theta^*) \right| \le \frac{1}{N} \sum_{n=1}^{N} \left(d(\mathbf{0}, p_n) \mathbb{I}_{\mathcal{I}_\psi \Delta \mathcal{I}_{\theta^*}^-}(p_n) \right).$$

Recall that we are assuming $\theta^* \le \hat{\theta}$. We may also assume that $|\hat{\theta} - \theta^*| < \min(\alpha - \pi, \pi/2)$ (which happens with probability approaching 1 as $N \to \infty$), then $\mathcal{I}_\psi \Delta \mathcal{I}_{\theta^*}^- \subset \mathcal{I}_{\hat{\theta}} \Delta \mathcal{I}_{\theta^*}^-$ for

$\psi \in (\theta^*, \hat{\theta})$, and therefore

$$\begin{aligned} |T_3| &\leq \int_{\theta^*}^{\hat{\theta}} \left| w_N(\psi) - w_N^-(\theta^*) \right| \cos(\hat{\theta} - \psi)\, d\psi \\ &\leq \frac{1}{N} \sum_{n=1}^N d(\mathbf{0}, p_n) \mathbb{I}_{\mathcal{I}_{\hat{\theta}} \Delta \mathcal{I}_{\theta^*}^-}(p_n) \int_{\theta^*}^{\hat{\theta}} \cos(\hat{\theta} - \psi)\, d\psi \\ &\leq \frac{1}{N} \sum_{n=1}^N d(\mathbf{0}, p_n) \mathbb{I}_{\mathcal{I}_{\hat{\theta}} \Delta \mathcal{I}_{\theta^*}^-}(p_n) |\hat{\theta} - \theta^*|. \end{aligned} \tag{6.20}$$

Fix $\epsilon > 0$ small and let $C_\epsilon > 0$ be such that $\mathbb{P}(\hat{\theta} \in [\theta^* - \frac{C_\epsilon}{\sqrt{N}}, \theta^* + \frac{C_\epsilon}{\sqrt{N}}]) \geq 1 - \epsilon$ for all N large enough. Then, with probability exceeding $1 - 2\epsilon$, we have

$$\mathbb{I}_{U_N^+} \cdot |T_3| \leq \frac{C_\epsilon}{\sqrt{N}} \frac{1}{N} \sum_{n=1}^N d(\mathbf{0}, p_n) \mathbb{I}_{\mathcal{I}_{\hat{\theta}} \Delta \mathcal{I}_{\theta^*}^-}(p_n) \leq \frac{C_\epsilon}{\sqrt{N}} \frac{1}{N} \sum_{n=1}^N d(\mathbf{0}, p_n) \mathbb{I}_{\mathcal{I}_{\bar{\theta}_N^\epsilon} \Delta \mathcal{I}_{\theta^*}^-}(p_n)$$

for N large enough, where the angle $\bar{\theta}_N^\epsilon = \theta^* + \frac{C_\epsilon}{\sqrt{N}}$ is now non-random. The random variables

$$\xi_n = d(\mathbf{0}, p_n) \mathbb{I}_{\mathcal{I}_{\bar{\theta}_N^\epsilon} \Delta \mathcal{I}_{\theta^*}^-}(p_n), \quad n = 1, 2, \ldots, N$$

are independent and identically distributed with mean and variance

$$\mathbb{E}[\xi_n] = \int_{\mathcal{I}_{\bar{\theta}_N^\epsilon} \Delta \mathcal{I}_{\theta^*}^-} d(\mathbf{0}, p)\, d\mu(p), \qquad \mathrm{Var}(\xi_n) = \mathbb{E}[\xi_n^2] - \mathbb{E}[\xi_n]^2 \leq \int_{\mathcal{I}_{\bar{\theta}_N^\epsilon} \Delta \mathcal{I}_{\theta^*}^-} d(\mathbf{0}, p)^2\, d\mu(p).$$

Due to the square integrability condition (1.3), both $\mathbb{E}[\xi_n]$ and $\mathrm{Var}(\xi_n)$ are finite. Moreover, since

$$\mathcal{I}_{\bar{\theta}_N^\epsilon} \Delta \mathcal{I}_{\theta^*}^- = \left\{ (r, \theta) \in \mathcal{K} \mid r > 0,\ \ \theta \in (\theta^* - \pi, \bar{\theta}_N^\epsilon - \pi) \cup [\bar{\theta}_N^\epsilon + \pi, \theta^* + \pi) \right\}$$

we have

$$\bigcap_{N \geq 1} \mathcal{I}_{\bar{\theta}_N^\epsilon} \Delta \mathcal{I}_{\theta^*}^- = \emptyset.$$

Hence, $\mu(\mathcal{I}_{\bar{\theta}_N^\epsilon} \Delta \mathcal{I}_{\theta^*}^-) \to 0$ as $N \to \infty$, and

$$\lim_{N \to \infty} \int_{\mathcal{I}_{\bar{\theta}_N^\epsilon} \Delta \mathcal{I}_{\theta^*}^-} d(\mathbf{0}, p)\, d\mu(p) = 0, \qquad \lim_{N \to \infty} \int_{\mathcal{I}_{\bar{\theta}_N^\epsilon} \Delta \mathcal{I}_{\theta^*}^-} d(\mathbf{0}, p)^2\, d\mu(p) = 0. \tag{6.21}$$

Thus, both $\mathbb{E}[\xi_n]$ and $\mathrm{Var}(\xi_n)$ vanish as $N \to \infty$. Consequently, for any $\delta > 0$,

$$\limsup_{N \to \infty} \mathbb{P}\left(|T_3| \geq \frac{\delta}{\sqrt{N}}, \quad \hat{\theta} > \theta^* \right) \leq 2\epsilon + \limsup_{N \to \infty} \mathbb{P}\left(C_\epsilon \frac{1}{N} \sum_{n=1}^N \xi_n > \delta \right) = 2\epsilon. \tag{6.22}$$

As $\epsilon > 0$ is arbitrary, we conclude that $T_3 = o(1/\sqrt{N})$. The result now follows by combining these estimates of T_1, T_2, and T_3.

Note: the reason we prove that Z_N is $o(1/\sqrt{N})$ rather than a stronger statement like $Z_N \leq O(1/N)$, is that we have no control over the rate at which $\mu(\mathcal{I}_{\bar{\theta}_N^\epsilon} \Delta \mathcal{I}_{\theta^*}^-) \to 0$ as $N \to \infty$ or on the rate of convergence in (6.21), unless we make more assumptions about μ.

□

7 Topological definition of sticky mean

7.1 Topological version for kale

Let $\mathcal{M}_1$ be the set of all finite Borel measures μ on $\mathcal{K}$ satisfying the integrability condition (1.2). This section considers how the mean (or barycenter) of a measure $\mu \in \mathcal{M}_1$ varies under perturbations of the measure. For this reason, we temporarily modify the notation for $m_{\theta,1}$ to $m_{\theta,1}(\mu)$, to reflect the measure μ being used. It is then easy to see that for $\mu, \nu \in \mathcal{M}_1$,

$$m_{\theta,1}(\mu + \epsilon\nu) = m_{\theta,1}(\mu) + \epsilon m_{\theta,1}(\nu). \tag{7.1}$$

Two measures $\mu, \nu \in \mathcal{M}_1$ are considered equivalent if they differ only in their total mass, meaning that there is a constant $c > 0$ with $\mu = c\nu$. Denote the space of equivalence classes by $\widetilde{\mathcal{M}}_1$. Endow $\mathcal{M}_1$ with the topology generated by the Wasserstein metric defined by

$$\rho(\mu, \nu) = \sup_{f \in \mathrm{Lip}_1} \left(\int f d\mu - \int f d\nu \right),$$

where Lip_1 is the set of real-valued, Lipschitz-continuous functions on $\mathcal{K}$ with Lipschitz constant 1. This topology extends to $\widetilde{\mathcal{M}}_1$ by declaring the distance between μ and ν to be the Wasserstein distance $\rho(\mu, \nu)$ when μ and ν are normalized so that $\mu(\mathcal{K}) = \nu(\mathcal{K}) = 1$.

Now comes the first in a sequence of results leading us to a definition of sticky and nonsticky that is more topological than Definition 1.8.

Lemma 7.1. *Let $\mu \in \widetilde{\mathcal{M}}_1$ be fully sticky. There exists an open neighborhood U of μ so that $\nu \in U$ implies (i) ν is fully sticky and (ii) μ and ν have the same mean.*

Proof. Since the function $e_1 \cdot F_\theta : \mathcal{K} \to \mathbb{R}$ is in Lip_1, Lemma 3.2 yields

$$\sup_\theta |m_{\theta,1}(\mu) - m_{\theta,1}(\nu)| \le \rho(\mu, \nu) \tag{7.2}$$

for any two measures $\mu, \nu \in \widetilde{\mathcal{M}}_1$. If μ is fully sticky, then there exists $\epsilon > 0$ so that $m_{\theta,1}(\mu) \le -\epsilon < 0$ for all θ. Therefore, if $\rho(\mu, \nu) \le \epsilon/2$ then $m_{\theta,1}(\nu) \le -\epsilon/2 < 0$ holds for all θ. Hence, by Definition 1.8, ν is also fully sticky. Since all fully sticky measures on the kale $\mathcal{K}$ have mean $\mathbf{0}$, we conclude that μ and ν have the same means. $\square$

Lemma 7.2. *The set of fully sticky measures is an open subset of $\widetilde{\mathcal{M}}_1$, as is the set of nonsticky measures.*

Proof. The statement for fully sticky measures is contained in Lemma 7.1. On the other hand, by Definition 1.8 the nonsticky measures are characterized by $m_{\theta,1}$ being strictly positive for an open range of θ. Let μ be a nonsticky measure with $m_{\theta,1}(\mu) > 2\epsilon$ for $\theta \in (A, B)$, for some $\epsilon > 0$. If $\nu \in B_\epsilon(\mu) \subset \widetilde{\mathcal{M}}_1$ then (7.2) implies that for all $\theta \in (A, B)$,

$$m_{\theta,1}(\nu) \ge \inf_{\theta \in (A,B)} m_{\theta,1}(\mu) - \rho(\mu, \nu) > 2\epsilon - \epsilon.$$

Therefore all $\nu \in B_\epsilon(\mu)$ are also nonsticky. $\square$

Definition 7.3. *Fix a measure $\mu \in \mathcal{M}_1$. A measure $\nu \in \mathcal{M}_1$, thought of as a direction, is*

1. sticky for μ *if μ and $\mu + \epsilon\nu$ have the same mean for all sufficiently small $\epsilon > 0$;*
2. fluctuating for μ *if μ and $\mu + \epsilon\nu$ have different means for all sufficiently small $\epsilon > 0$.*

Since normalization does not change whether a measure is sticky, partly sticky, or nonsticky, one could replace $\mu + \epsilon\nu$ by $(1-\epsilon)\mu + \epsilon\nu$ in the above definition. The latter has the advantage of producing a probability measure if both μ and ν were initially so.

It is convenient to have a specific class of perturbations at our disposal. Note that for the unit measure δ_p supported at the point $p = (r, \theta')$,

$$m_{\theta,1}(\delta_p) = \begin{cases} r\cos(\theta - \theta') & \text{if } |\theta - \theta'| < \pi \\ -r & \text{if } |\theta - \theta'| \geq \pi. \end{cases} \tag{7.3}$$

Lemma 7.4. *Any nonsticky or partly sticky $\mu \in \mathcal{M}_1$ has a fluctuating direction in $\mathcal{M}_1$.*

Proof. Let (r^*, θ^*) be the mean of μ. When μ is partly sticky, $r^* = 0$ and θ^* is any value; when μ is nonsticky, $r^* = m_{\theta^*,1} > 0$ and $\theta \mapsto m_{\theta,1}(\mu)$ attains its maximum at the unique point θ^*. Fix any radius $r > 0$ with $r \neq r^*$, and set $\mu_\epsilon = (1-\epsilon)\mu + \epsilon\delta_{(r,\theta^*)}$. By (7.1) and (7.3), $\theta \mapsto m_{\theta,1}(\mu_\epsilon)$ now has its unique maximum at θ^*, but $m_{\theta^*,1}(\mu_\epsilon) \neq m_{\theta^*,1}(\mu)$ because $r \neq r^*$. Hence μ and μ_ϵ have different means, so the direction $\delta_{(r,\theta^*)}$ is fluctuating for μ. $\square$

Lemma 7.5. *If $\mu \in \mathcal{M}_1$ is partly sticky then μ has a sticky direction (other than $\nu = \mu$).*

Proof. Since μ is partly sticky, $m_{\theta,1}(\mu) \leq 0$ for all θ. Let ν be any fully sticky measure and define $\mu_\epsilon = (1-\epsilon)\mu + \epsilon\nu$. Since ν is fully sticky, $m_{\theta,1}(\nu) < 0$ for all θ, and hence $m_{\theta,1}(\mu_\epsilon) < 0$ for all θ as long as $\epsilon > 0$. Therefore μ_ϵ is fully sticky, and the means of μ_ϵ and μ coincide at $\mathbf{0}$ for all $\epsilon \in [0,1]$. Thus $\nu \neq \mu$ is a sticky direction for μ. $\square$

The above lemmas combine with the fact that all measures in $\widetilde{\mathcal{M}}_1$ are either fully sticky, partly sticky, or nonsticky (Proposition 4.11 and Definition 1.8) to prove the following theorem, which could be seen as an alternative definition of the terms "fully sticky", "partly sticky", and "nonsticky" for finite measures on $\mathcal{K}$.

Theorem 7.6. *Let $\mathcal{S} \subset \widetilde{\mathcal{M}}_1$ be the open subset of fully sticky measures. A measure $\mu \in \widetilde{\mathcal{M}}_1$ is*

1. *fully sticky (i.e. $\mu \in \mathcal{S}$) if and only if there is an open neighborhood of μ so that all measures in that neighborhood have the same mean as μ. Equivalently, a measure μ is fully sticky if and only if all directions $\nu \in \widetilde{\mathcal{M}}_1$ are sticky for μ.*
2. *partly sticky if and only if $\mu \in \partial\mathcal{S}$, the topological boundary of $\mathcal{S}$. Equivalently, a measure μ is partly sticky if and only if every open neighborhood of μ contains open sets U and V such that $\nu \in V \implies \nu$ has the same mean as μ and $\nu \in U \implies \mu$ and ν have different means.*
3. *nonsticky if and only if $\mu \in \widetilde{\mathcal{M}}_1 \setminus \overline{\mathcal{S}}$, the compliment of the closure of $\mathcal{S}$. Equivalently, a measure μ is nonsticky if and only if no open neighborhood of μ contains an open set U consisting of measures with the same mean as μ.*

Remark 7.7. As N gets large, the empirical measure

$$\mu_N = \frac{1}{N}\sum_{n=1}^{N} \delta_{p_n}$$

converges to μ in the topology generated by ρ if the p_n are chosen independently and according to μ. (For instance combine [21, Theorem 6.9] and the standard weak convergence of empirical measures.) If μ is sticky then eventually μ_N lies in a neighborhood of μ in which all measures have the same mean. On the other hand, if μ is nonsticky then nearby measures have different means than μ and hence the mean of μ_N fluctuates. When μ is partly sticky, sometimes μ_N lies in a set of measures sharing their mean with μ, and sometimes it lies in a set of measures having different means than μ.

Remark 7.8. Endowing $\mathcal{M}_1$ instead with the topology generated by the open neighborhoods

$$U_{\mu,\epsilon} = \{\nu \in \mathcal{M}_1 \mid \max_{\theta} |m_{\theta,1}(\mu) - m_{\theta,1}(\nu)| < \epsilon\}$$

maintains the truth of the above results. However, using the standard weak topology on measures, which is finer, would cause the topological characterization of stickiness to fail.

7.2 Topological definition for arbitrary metric spaces

Suppose $\mathcal{K}$ is a metric space, and let $\mathcal{M}$ be a set of probability measures on $\mathcal{K}$.

Example 7.9. When $\mathcal{M} = \widetilde{\mathcal{M}}_1$ is the set of Borel probability measures on $\mathcal{K}$ satisfying the integrability condition (1.2), different topologies on $\mathcal{M}$ are induced by the Wasserstein metric and by the sets $U_{\mu,\epsilon}$ in Remark 7.8. The standard weak topology is yet another possibility.

Definition 7.10. *Let $\mathcal{M}$ be a set of measures on a metric space $\mathcal{K}$ with the metric topology. Assume $\mathcal{M}$ has a given topology. A* mean *is a continuous assignment $\mathcal{M} \to \{$closed subsets of $\mathcal{K}\}$. A measure μ* sticks *to a closed subset $C \subseteq \mathcal{K}$ if every neighborhood of μ in $\mathcal{M}$ contains a nonempty open subset consisting of measures whose mean sets are contained in C.*

Remark 7.11. Regarding the topology on the set of closed subsets of $\mathcal{K}$, implicit in Definition 7.10, we have in mind the topology induced by the *Hausdorff distance*:

$$d(A, B) = \max\left\{\sup_{a \in A} d(a, B)\ ,\ \sup_{b \in B} d(b, A)\right\}.$$

That is, $d(A, B)$ is the farthest a point of A is from B or the farthest a point of B is from A, whichever is greater. Other topologies on the set of closed subsets of $\mathcal{K}$ are possible, such as the “pointed Hausdorff topology", which is compact and locally compact.

Continuity implies that the mean of μ is contained in C if μ sticks to C.

Example 7.12. This paper has investigated measures on the kale $\mathcal{K}$, which can stick to the subset $C = \{\mathbf{0}\}$ consisting of the origin. The notion of “mean” here is Definition 1.11, which assigns to each measure a single point; this assignment is continuous by Lemma 4.3.

In spaces of interest, integrability conditions, such as those in Section 1 here, would imply existence of means. However, means in general metric spaces—even nice ones such as compact Riemannian manifolds—need not be single points. In other words, the general analogue of the minimization problem in Section 1.2 could have multiple solutions. For instance the mean set of the uniform measure on a sphere is equal to that entire sphere, whereas each sample mean is unique almost surely (cf. Remark 2.6 in [5]). In Section 5 of [13] there is an example of a measure on the circle where the mean set is a proper circular arc. In fact, this can be viewed as the limiting case of measures with unique means, the central limit theorems for which feature arbitrarily slow convergence rates. Uniqueness of means for the kale stem from its negative curvature; see [20, Proposition 4.3], for example.

Remark 7.13. In the language of earlier sections, Definition 7.10 only sets forth the notion of “sticky”, which includes both the sticky and partly sticky cases. In the generality of Definition 7.10, it would be said that a measure μ *fully sticks to* C if some open neighborhood of μ consists entirely of measures whose means are contained within C. It would not be required that the means (closed subsets of C) of the measures in such a neighborhood should equal the mean of μ or even intersect it at all. In the case where $\mathcal{K}$

is an open book [14], for example, means are unique and measures can stick to the spine, but nothing prevents the mean of a sticky measure from moving along the spine.

The set of partly sticky measures would be defined as those that are sticky but not fully sticky. Definition 7.10 implies that the set of partly sticky measures is the topological boundary of the set of sticky measures.

It remains open to characterize which metric spaces—among, say, the topologically stratified spaces (see [9] or [18]), to be concrete—admit measures that stick to subsets of measure 0. Given such a sticky situation, first goals would be to prove laws of large numbers and central limit theorems, contrasting the fully, partly, and nonsticky cases. The limiting measures in such results would be singular analogues of Gaussian distributions; it is not clear what properties of Gaussian distributions are the right ones to lift so as to characterize the building blocks of limiting measures in general.

8 List of Notation

$d(p,q)$	Metric on $\mathcal{K}$. See Section 1.1.
F_θ	The folding map, from $\mathcal{K}$ to $\mathbb{R}^d$, at angle θ. Definition 1.3.
$\mathcal{I}_\theta$	The shadow of angle θ; an open subset of $\mathcal{K}$. Definition 1.4.
μ	A probability measure on $\mathcal{K}$.
μ_N	The empirical measure for points $p_1,\dots,p_N \in \mathcal{K}$. See Section 1.3.
b_N	The barycenter of a (random) set of points $p_1,\dots,p_N \in \mathcal{K}$. See (1.6).
$\bar{b}$	Population barycenter. See Definition 1.6.
$\tilde{\mu}_\theta$	The pushforward $\mu \circ F_\theta^{-1}$ of μ under F_θ; a measure on $\mathbb{R}^2$.
m_θ	First moment of measure μ folded about angle θ. Definition 1.7.
m_θ^N	First moment of the empirical measure μ_N folded about angle θ. Definition 1.7.
$m_{\theta,1}$, $m_{\theta,2}$	Components of $m_\theta = (m_{\theta,1}, m_{\theta,2}) \in \mathbb{R}^2$.
$\eta_{\theta,N}$	Folded average, equivalent to m_θ^N. See (1.8).
ν_N	Distribution of rescaled empirical means, a probability measure on $\mathcal{K}$. See (1.10).
$\kappa(\omega)$	A random variable related to the CLT in the non-sticky case. See (1.17).
$w^\pm(\theta)$	Constants which depend on the measure μ. See (1.17).
$\mathcal{I}_\theta^\pm$	Shadow at angle θ including part of the boundary. See (1.18).
Φ_σ	Rotation in $\mathbb{R}^2$ by angle σ. See Lemma 4.8.
$\bar{r}$	Constant bounding first moments of the measure μ. See (1.2).
$\hat{P}_\rho$	Convex projection onto a sector in $\mathbb{R}^2$. See (1.14).
g	Gaussian measure on $\mathbb{R}^2$ with mean zero, covariance Σ. See Sec. 1.4.2 and 1.4.3.

References

[1] Aydın, B., Pataki, G., Wang, H., Bullitt, E., Marron, J.: A principal component analysis for trees. *The Annals of Applied Statistics* **3**, (2009), 1597-1615. MR-2752149

[2] Barden, D., Le, H., Owen, M.: Central limit theorems for Fréchet means in the space of phylogenetic trees. *Electron. J. Probab* **18**, (2013), 1-25. MR-3035753

[3] Barden, D., Le, H., Owen, M.: Limiting behaviour of Fréchet means in the space of phylogenetic trees. arXiv:1409.7602v1 (2014)

[4] Basrak, B.: Limit theorems for the inductive mean on metric trees. *Journal of Applied Probability* **47**, (2010), 1136-1149. MR-2752884

[5] Bhattacharya, R. N., and Patrangenaru, V.: Large sample theory of intrinsic and extrinsic sample means on manifolds I. *The Annals of Statistics* **31**, (2003), 1-29. MR-1962498

[6] Bhattacharya, R. N. and Patrangenaru, V.: Large sample theory of intrinsic and extrinsic sample means on manifolds II. *The Annals of Statistics* **33**, (2005), 1225-1259. MR-2195634

[7] Billera, L., Holmes, S., Vogtmann, K.: Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics* **27**, (2001), 733–767. MR-1867931

[8] Gibson, C. G., Wirthmüller, K., du Plessis, A. A., Looijenga, E.: *Topological Stability of Smooth Mappings*. Lecture Notes in Mathematics, **552**, Springer-Verlag, 1976. MR-0436203

[9] Goresky, M. and MacPherson, R.: *Stratified Morse theory*. Ergebnisse der Mathematik und ihrer Grenzgebiete (3), **14**, Springer-Verlag, 1988. MR-0932724

[10] Hendriks, H. and Landsman Z.: Asymptotic behaviour of sample mean location for manifolds. *Statistics & Probability Letters* **26**, (1996), 169-178. MR-1381468

[11] Hendriks, H. and Landsman Z.: Mean location and sample mean location on manifolds: asymptotics, tests, confidence regions. *Journal of Multivariate Analysis* **67**, (1998), 227-243. MR-1659156

[12] Holmes, S.: Statistics for phylogenetic trees. *Theoretical population biology* **63**, (2003), 17-32.

[13] Hotz, T. and Huckemann, S.: Intrinsic means on the circle: uniqueness, locus and asymptotics. *Annals of the Institute of Statistical Mathematics*, (2014), 1-17.

[14] Hotz, T., Huckemann, S., Le, H., Marron, J. S., Mattingly, J., Miller, E., Nolen, J., Owen, M., Patrangenaru, V., Skwerer, S.: Sticky central limit theorems on open books. *Annals of Applied Probability*, (2013), 2238-2258. MR-3127934

[15] Hotz, T. and Le H. Confidence regions in spiders. Preprint (2014), to appear in *Oberwolfach Reports*.

[16] Huckemann, S.: Inference on 3D Procrustes means: Tree boles growth, rank-deficient diffusion tensors and perturbation models. *Scandinavian Journal of Statistics* **38**, (2011), 424-446. MR-2833839

[17] Nye, T.: Principal components analysis in the space of phylogenetic trees. *The Annals of Statistics* **39**, (2011), 2716-2739. MR-2906884

[18] Pflaum, M.J.: *Analytic and Geometric Study of Stratified Spaces*. Lecture Notes in Mathematics, **1768**, Springer-Verlag, 2001. MR-1869601

[19] Skwerer, S., Bullitt, E., Huckemann, S., Miller, E., Oguz, I., Owen, M., Patrangenaru, V., Provan, S., Marron, J.: Tree-oriented analysis of brain artery structure. *Journal of Mathematical Imaging and Vision*. **50** (2014), 126-143. MR-3233139

[20] Sturm, K.-T.: Probability measures on metric spaces of nonpositive curvature. In *Heat kernels and analysis on manifolds, graphs, and metric spaces: lecture notes from a quarter program on heat kernels, random walks, and analysis on manifolds and graphs*, Contemporary Mathematics, **338**, (2003) 357-390. MR-2039961

[21] Villani, C.: Optimal Transport. Old and new. Springer-Verlag, 2009. MR-2459454

[22] Ziezold, H.: Expected figures and a strong law of large numbers for random elements in quasi-metric spaces. *Transaction of the 7th Prague Conference on Information Theory, Statistical Decision Function and Random Processes* **A**, (1977), 591-602. MR-0501230

Acknowledgments. The authors acknowledge support through the Niedersachsen Vorab of the Volkswagen Foundation (SH) as well as grants from the US National Science Foundation: DMS-0854879 (JCM); DMS-1001437 (EM); DMS-1007572, DMS-1351653 (JN); and the Statistical and Applied Mathematical Sciences Institute, SAMSI, DMS-1127914 (SH).