Convex generalized Fréchet means in a metric tree

Gabriel Romon[1] **and Victor-Emmanuel Brunel**[1]

[1] *CREST, ENSAE, IP Paris*
5 avenue Henry Le Chatelier, 91764 Palaiseau
e-mail: gabriel.romon@ensae.fr; victor-emmanuel.brunel@ensae.fr

Abstract: We are interested in measures of central tendency for a population on a network, which is modeled by a metric tree. The location parameters that we study are generalized Fréchet means obtained by replacing the usual objective $\alpha \mapsto \mathbb{E}[d(\alpha, X)^2]$ with $\alpha \mapsto \mathbb{E}[\ell(d(\alpha, X))]$ where ℓ is a generic convex nondecreasing loss.

We develop a notion of directional derivative in the tree, which helps up locate and characterize the minimizers.

Estimation is performed using a sample analog. We extend to a metric tree the notion of stickiness defined by Hotz et al. (2013). For generalized Fréchet means we obtain a sticky law of large numbers. For the Fréchet median we develop non-asymptotic concentration bounds and sticky central limit theorems.

1. Introduction

1.1. Context

Statisticians commonly model data as an i.i.d. sample from an unknown probability measure μ. There is much interest in the central tendency of μ, i.e., in defining a parameter that is representative of the whole population, and then in estimating this location parameter. When the ambient space is $\mathbb{R}^d$, a prominent measure of central tendency is the mean $\int_{\mathbb{R}^d} x \, d\mu(x)$ and an estimator is the sample mean $\frac{1}{n}\sum_{i=1}^n X_i$, where $X_1, \dots \overset{\text{i.i.d.}}{\sim} \mu$. Fréchet [21] extended the notion of mean to the general setting of metric spaces by leveraging an optimization problem. Given a metric space (E, d) and a Borel probability measure μ on E with finite second moment, a Fréchet mean (or barycenter) of μ is a minimizer of the objective function

$$\begin{aligned} E &\to \mathbb{R} \\ \alpha &\mapsto \int_E d(\alpha, x)^2 d\mu(x). \end{aligned} \tag{1}$$

Only a first moment assumption is actually necessary, however the objective needs to be slightly modified. In many settings the Fréchet mean $\alpha_\star$ exists and is unique [48, 1, 2, 37, 54].

A natural estimator of $\alpha_\star$ is the sample Fréchet mean $\widehat{\alpha}_n$ obtained by minimizing the sample objective $\alpha \mapsto \frac{1}{n}\sum_{i=1}^n d(\alpha, X_i)^2$. Laws of large numbers for $(\widehat{\alpha}_n)_{n\geq 1}$ hold in a variety of spaces [56, 9, 48] and central limit theorems have been developed for Riemannian manifolds [10, 7, 8, 17]. Non-Euclideanity of the space allows for new asymptotic phenomena such as stickiness [29, 31] and smeariness [28, 18]. The non-asymptotic properties of the estimator have attracted much attention recently [42, 3, 34, 55, 15, 19].

Except for laws of large numbers and Riemannian central limit theorems [32, 14], these statistical results and their proofs are specific to the Fréchet mean, i.e., they are tied to the objective function (1). Still, other measures of central tendency are of interest. In the simplest setting of the real line $\mathbb{R}$, a major shortcoming of the sample mean is its lack of robustness to outliers, hence the need for alternatives such as the median. The population Fréchet median can be defined by replacing the squared distance in (1) with $d(\alpha, x)$. In order to cover a variety of location parameters, we study more general objectives of the form $\alpha \mapsto \int_E \ell\big(d(\alpha, x)\big) d\mu(x)$ where $\ell : [0, \infty) \to [0, \infty)$ is a convex nondecreasing function. We refer to minimizers of such an objective as Fréchet-ℓ means.

In exchange for generality in the objective, we constrain the ambient space to be a metric tree T, i.e., an undirected connected acyclic graph with weighted edges, where weights are understood as edge lengths and the distance between two points is the length of the shortest path between them. Metric trees arise in real-life applications, as they are an ideal model for road and communication networks. Tree-shaped networks appear naturally when modeling rivers or sparsely populated areas. A distribution system organized around a unique hub may be described as a star-like network, thus as a tree. In all these settings, the demand for service can occur at random locations across the network and these locations are distributed according to μ. Minimizing $\alpha \mapsto \int_T d(\alpha, x) d\mu(x)$ is then akin to locating a new facility on the network with least average travel time to the demand. This median problem was initially studied in the special case where μ is discrete and supported on the vertices of the network. It gained traction among the operations research community in the 1960s, with an emphasis on the development of efficient algorithms (see, e.g., the surveys [25, 50, 51, 26]). More general objective functions were considered in [46, 12] and the case of non-discrete μ was studied in [36, 11].

A metric tree is a particular instance of a Hadamard space [13], hence results for Fréchet means in general Hadamard spaces (e.g., [48, 6, 34, 14, 15, 19]) apply also to metric trees. There is little statistical literature on Fréchet-ℓ means in the specific setting of metric trees. Basrak [5] focuses on the Fréchet mean in a binary metric tree, and he establishes a central limit theorem for the inductive mean (a different estimator from the sample Fréchet mean). Risser et al. [22, 23] seek to compute Fréchet means on metric graphs, while Hotz et al. [29] develop laws of large numbers and central limit theorems for the Fréchet mean when the ambient space is an open book. A special case of an open book is the m-spider, which can be viewed as a peculiar kind of metric tree.

1.2. Contributions and outline

The goal of this chapter is to investigate the statistical properties of Fréchet-ℓ means in the metric tree T. We describe below how the chapter is organized, and we give a brief summary of our contributions.

- In Section 2 we introduce the precise terminology and setting for our study. By leveraging the geodesic convexity of the objective function ϕ (Proposition 2.5) and the geometry of the tree (Definition 2.6), we develop a notion of directional derivative for ϕ (Definition 2.8). Owing to the geodesic convexity of ϕ, we are able to locate (Proposition 2.12) and characterize (Proposition 2.13) Fréchet-ℓ means according to the signs of directional derivatives.
- In Section 3 we turn to estimation using a sample analog. We observe that the topic of consistency is settled (Lemma 3.3) and we extend the notion of stickiness introduced by Hotz et al. [29] to the metric tree. An arbitrary point $c \in T$ is either sticky, partly sticky or nonsticky according to the signs of directional derivatives at c (Definition 3.4). We show that empirical stickiness is a non-asymptotic phenomenon that happens with exponential probability (Theorem 3.7). As an immediate consequence, we obtain a sticky law of large numbers (Corollary 3.8). Finally, we provide an equivalent definition of stickiness that is stated in terms of robustness to small pertubations of the population distribution (Proposition 3.10).
- In Section 4 we focus on Fréchet medians. We develop more precise statements on the location (Proposition 4.1) and uniqueness (Proposition 4.3) of medians. In the partly sticky case, we

establish central limit theorems (Theorems 4.12 and 4.20) and non-asymptotic concentration bounds (Theorems 4.15 and 4.22).

2. Fréchet ℓ-means in a metric tree

2.1. Terminology and setting

Let us make precise what we mean by a metric tree and introduce further terminology.

Definition 2.1.

1. Let T denote an undirected, connected, acyclic graph with weighted edges (in the usual graph-theoretic sense). The weight of an edge is understood as the length of this edge, i.e., as the distance between the corresponding adjacent vertices. We assume additionally that T has edges of positive finite length and finitely many vertices. We implicitly consider a planar and isometric embedding of T in $\mathbb{R}^2$; T is then equipped with the shortest path metric d: the distance between two points (not necessarily vertices) is the length of the shortest path between them. (T, d) is a metric space, which is referred to as a *metric tree*.
2. A vertex $v \in T$ is a *leaf* if it has exactly one adjacent vertex.
3. Let $m \geq 2$. T is an *m-spider* if the underlying graph-theoric tree has exactly m leaves (and thus 1 central vertex).

Next, we introduce some relevant concepts from metric geometry. Given $x, y \in T$, a constant speed geodesic from x to y is a map γ from some interval $[a, b] \subset \mathbb{R}$ to E such that $\gamma(a) = x$, $\gamma(b) = y$ and $d(\gamma(t_1), \gamma(t_2)) = v|t_1 - t_2|$ for some $v \in [0, \infty)$ and every $t_1, t_2 \in [a, b]$. The real number v is called the speed of the geodesic γ. For the sake of legibility, we will often write γ_t in lieu of $\gamma(t)$. The space (T, d) is uniquely geodesic, meaning that there exists a geodesic from x to y and all the geodesics γ from x to y have the same image. This image is denoted by $[x, y]$ and it is referred to as the geodesic segment joining x and y. We also define open and half-open geodesic intervals (x, y), $[x, y)$, $(x, y]$: for instance $(x, y) = \gamma((a, b))$ for some $a \leq b$ and a geodesic γ from x to y defined on $[a, b]$.

A well-known geometric property of metric trees is that they are CAT(0); see, e.g., [13, Example 1.15(5) p.167]. By our assumptions, (T, d) is also compact and complete, hence it is a compact Hadamard space. In Hadamard spaces it is possible to develop a theory of convex analysis, convex optimization and probability that generalizes to nonlinear settings the classical results known in Hilbert spaces [6].

Definition 2.2. Let (T, d) be a metric tree as above, μ be a probability measure on $(T, \mathcal{B}(T))$ and $\ell : [0, \infty) \to [0, \infty)$ be a convex and nondecreasing function, which we call the *loss function*. We define the *objective function* ϕ

$$\begin{aligned} \phi \colon T &\to \mathbb{R} \\ \alpha &\mapsto \int_T \ell\big(d(\alpha, x)\big) d\mu(x). \end{aligned} \tag{2}$$

Minimizers of ϕ are called *Fréchet ℓ-means of* μ, and we denote by $M(\mu)$ the set of all minimizers.

Example 2.3. Examples of loss functions ℓ include:

1. $\ell : z \mapsto z^p$ where $p \in [1, \infty)$. In this setting, the minimizers of ϕ are called *Fréchet p-means* of μ. In the case $p = 1$ they are referred to as *Fréchet medians*, and when $p = 2$ as *barycenters* or just *Fréchet means*. The corresponding set of minimizers will be denoted specifically by $M_p(\mu)$.

2. $\ell : z \mapsto z^2 \mathbb{1}_{|z|\le c} + (2c|z| - c^2)\mathbb{1}_{|z|>c}$ where $c \ge 0$. It is known as the Huber loss [30].
3. $\ell : z \mapsto 2c^2\big((1+\frac{z^2}{c^2})^{1/2} - 1\big)$ where $c > 0$. It is known as the pseudo-Huber loss, which is a smooth approximation of the standard Huber loss.
4. Consider an arbitrary $f : [0,\infty) \to [0,\infty)$ that is nondecreasing and some $a \ge 0$. The loss is defined via integration:
$$\ell : z \mapsto a + \int_0^z f(t)dt. \tag{3}$$

The following lemma exhibits basic regularity properties of the loss and states that an integral representation such as (3) always exists.

Lemma 2.4. *Let $\ell : [0,\infty) \to [0,\infty)$ be a convex and nondecreasing function.*

1. *The left-derivative $\ell'_- : (0,\infty) \to [0,\infty)$ and right-derivative $\ell'_+ : [0,\infty) \to [0,\infty)$ of ℓ exist and are nondecreasing.*
2. *ℓ is locally Lipschitz.*
3. *For every $z \in [0,\infty)$, $\ell(z) = \ell(0) + \int_0^z \ell'_+(t)dt$.*

In the next proposition, we show that the objective ϕ is well-defined and we provide other foundational properties of ϕ and $M(\mu)$.

Proposition 2.5. 1. *ϕ is well-defined, continuous and convex.*
2. *$M(\mu)$ is a nonempty, closed and convex subset of T.*
3. *ℓ is strictly convex if and only if ℓ'_+ is increasing. In that case, ϕ is strictly convex and $M(\mu)$ is a singleton.*

2.2. Convex calculus in a metric tree

The following section is dedicated to locating and characterizing the minimizers of ϕ.

Given a real-valued function f defined on a vector space E, the variations of f with respect to a reference point $\alpha \in E$ and in a direction $v \in E$ are naturally assessed by restricting f to the half-line $\{\alpha + tv : t \ge 0\} \subset E$ and defining the difference quotient

$$\begin{aligned} q\colon (0,\infty) &\to \mathbb{R} \\ t &\mapsto \frac{f(\alpha+tv) - f(\alpha)}{t}. \end{aligned} \tag{4}$$

If additionally f is convex, then q is nondecreasing and bounded below; its right-sided limit is the directional derivative of f at α in the direction v [27, p.238].

In general the metric space T has no linear structure, and a point $v \in T$ does not carry by itself a notion of direction. However the restriction to the half-line in (4) can be replaced with the restriction to the geodesic segment $[\alpha, v]$, thus we consider the *metric difference quotient*

$$\begin{aligned} Q\colon (0,1] &\to \mathbb{R} \\ t &\mapsto \frac{\phi(\gamma_t) - \phi(\alpha)}{d(\gamma_t, \alpha)}, \end{aligned} \tag{5}$$

where $\gamma : [0,1] \to [\alpha, v]$ denotes the geodesic from α to v. Since $d(\gamma_t, \alpha) = td(v,\alpha)$ and $t \mapsto \phi(\gamma_t)$ is convex, Q is nondecreasing and has a right-sided limit at 0 (which we will see is finite). Before we provide the value of this limit, we need the following definition.

Definition 2.6. Given α and v two distinct elements of T, we let $w_1, \dots, w_m$ denote the leaves of T and we define the subset

$$T_{\alpha \to v} = (\alpha, v] \cup \bigcup_{i \in \{1,\dots,m\}:\ \alpha \notin [v, w_i]} [v, w_i].$$

Alternatively, the metric space $T \setminus \{\alpha\}$ has two path-components and $T_{\alpha\to v}$ is the path-component that contains v. It is also the largest convex subset of T that contains v but not α.

Figure 1 illustrates this definition in two situations: either α is in the interior of an edge, or α is a vertex of T. We stress that α does not belong to $T_{\alpha\to v}$.

(a) α is in the interior of an edge (b) α is a vertex

Fig 1: Illustration for Definition 2.6 in two cases. $T_{\alpha\to v}$ is drawn in red and $T \setminus T_{\alpha\to v}$ is drawn in green.

As shown in the next proposition, the expression for the limit of the metric difference quotient (5) involves the left- and right- derivative of the univariate convex function ℓ, which already played a role in Lemma 2.4.

Proposition 2.7. *Let α and v be two distinct points in T. The following convergence holds:*

$$\frac{\phi(\alpha') - \phi(\alpha)}{d(\alpha', \alpha)} \xrightarrow[\substack{\alpha' \to \alpha \\ \alpha' \in (\alpha, v]}]{} \int_{T \setminus T_{\alpha\to v}} \ell'_+(d(\alpha, x))d\mu(x) - \int_{T_{\alpha\to v}} \ell'_-(d(\alpha, x))d\mu(x). \tag{6}$$

Consequently, the metric difference quotient $Q(t)$ converges to this finite limit as $t \to 0^+$.

Definition 2.8. We refer to the limiting value in (6) as *the directional derivative of ϕ at α towards v* and we denote it by $\phi'_v(\alpha)$.

Remark 2.9. If w is in $T_{\alpha\to v}$ and $w \neq v$, we note that $T_{\alpha\to w} = T_{\alpha\to v}$, thus $\phi'_w(\alpha) = \phi'_v(\alpha)$. The equality between derivatives is expected: if α' is in $(\alpha, w]$ and sufficiently close to α, then α' is in $(\alpha, v]$.

Example 2.10.

1. For Fréchet p-means with $p > 1$,

$$\phi'_v(\alpha) = p\Big(\int_{T \setminus T_{\alpha\to v}} d(\alpha, x)^{p-1} d\mu(x) - \int_{T_{\alpha\to v}} d(\alpha, x)^{p-1} d\mu(x) \Big).$$

2. For Fréchet medians,

$$\begin{aligned} \phi'_v(\alpha) &= \mu(T \setminus T_{\alpha\to v}) - \mu(T_{\alpha\to v}) \\ &= 1 - 2\mu(T_{\alpha\to v}) \\ &= 2\mu(T \setminus T_{\alpha\to v}) - 1. \end{aligned}$$

Remarkably, the directional derivative does not involve the metric d; it is expressed solely in terms of μ.

Remark 2.11. Assessing the directional derivative of ϕ along an edge is not a new idea: [46, 12] perform the computation for the sample Fréchet p-mean, [48] does so for the population Fréchet mean on a m-spider, and [35] for the sample Fréchet mean.

Next, we leverage the geometry of T and the geodesic convexity of ϕ to show a connection between the sign of the directional derivative and the location of the minimizers of ϕ.

Proposition 2.12. *Let α_0 and v two distinct points in T.*

1. *If $\phi'_v(\alpha_0) < 0$, then $M(\mu) \subset T_{\alpha_0 \to v}$.*
2. *If $\phi'_v(\alpha_0) > 0$, then $M(\mu) \subset T \setminus T_{\alpha_0 \to v}$.*
3. *If $\phi'_v(\alpha_0) = 0$, then $\alpha_0 \in M(\mu)$.*

As a consequence, we obtain the following first-order optimality conditions.

Proposition 2.13. *Let $\alpha \in T$.*

1. *The following are equivalent:*
   (a) *$\alpha \in M(\mu)$.*
   (b) *For every $v \in T \setminus \{\alpha\}$, $\phi'_v(\alpha) \geq 0$.*
   (c) *For every neighboring vertex v of α, $\phi'_v(\alpha) \geq 0$.*
2. *Assume that α lies in the interior of an edge $[v, w]$, that $\mu(\{\alpha\})\ell'_+(0) = 0$ and that ℓ is differentiable over $(0, \infty)$. Then*

$$\begin{aligned}\alpha \in M(\mu) &\iff \phi'_v(\alpha) = \phi'_w(\alpha) = 0 \\ &\iff \int_{T_{\alpha \to v}} \ell'(d(\alpha, x))d\mu(x) = \int_{T_{\alpha \to w}} \ell'(d(\alpha, x))d\mu(x).\end{aligned}$$

Remark 2.14. Any v with $\phi'_v(\alpha) < 0$ is called a *descent direction at* α. By Proposition 2.13 we obtain the following alternative, which is well-known in convex optimization over vector spaces: either there exists a descent direction at α, or $\alpha \in M(\mu)$.

Example 2.15. Assume that α lies in the interior of an edge $[v, w]$.

1. For Fréchet p-means with $p > 1$, item 2. of Proposition 2.13 yields

$$\alpha \in M_p(\mu) \iff \int_{T_{\alpha \to v}} d(\alpha, x)^{p-1} d\mu(x) = \int_{T_{\alpha \to w}} d(\alpha, x)^{p-1} d\mu(x).$$

2. For Fréchet medians,

$$\alpha \in M_1(\mu) \iff \mu(T_{\alpha \to v} \cup \{\alpha\}) \geq \frac{1}{2} \text{ and } \mu(T_{\alpha \to w} \cup \{\alpha\}) \geq \frac{1}{2}.$$

This last optimality condition is reminiscent of the classical characterization of a median on $\mathbb{R}$ as any $m \in \mathbb{R}$ that verifies both $\mu\big((-\infty, m]\big) \geq 1/2$ and $\mu\big([m, \infty)\big) \geq 1/2$.

By Proposition 2.5, $M(\mu)$ is a nonempty convex subset of T. Under a mild additional assumption on ℓ we obtain the following more precise statement on the geometry of $M(\mu)$.

Proposition 2.16. *If ℓ is increasing, then $M(\mu)$ is a geodesic segment.*

Example 2.17. When $p > 1$ the loss defining the Fréchet p-mean is strictly convex, hence by Proposition 2.5 $M(\mu)$ is a singleton and thus a geodesic segment. For Fréchet medians ($p = 1$), the loss is not strictly convex, however it is increasing and $M(\mu)$ is a geodesic segment.

3. Estimation of Fréchet ℓ-means and statistical results

3.1. Estimation setting

Definition 3.1. Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. T-valued random elements defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, each with distribution μ. For each $n \geq 1$, we define the *empirical measure* $\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ and the *empirical objective function* $\widehat{\phi}_n : \alpha \mapsto \frac{1}{n}\sum_{i=1}^n \ell(d(\alpha, X_i))$. Minimizers of $\widehat{\phi}_n$ (i.e., elements of $M(\widehat{\mu}_n)$) are called *empirical Fréchet ℓ-means.*

To avoid notational overburden, the directional derivative of $\widehat{\phi}_n$ will be written as $\widehat{\phi}'_v(\alpha)$; the integer n is clear from the context and is therefore omitted.

We face measurability difficulties when stating some of the statistical results below. We resolve these issues in two ways: first, without loss of generality, we assume throughout that the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is complete. Second, we evaluate the probability of (possibly nonmeasurable) subsets of Ω by employing outer and inner measures $\mathbb{P}^*, \mathbb{P}_*$ [52, Chapter 1.2].

In the section on central limit theorems, we consider arbitrary sequences of minimizers $(\widehat{\alpha}_n)_{n\geq 1}$ such that $\widehat{\alpha}_n \in M(\widehat{\mu}_n)$ for each $n \geq 1$. It will be convenient to require in addition that each $\widehat{\alpha}_n$ is measurable between the σ-algebras $\mathcal{F}$ and $\mathcal{B}(T)$. Since T is compact and ϕ is continuous, such a measurable selection exists [4, Theorem 18.19].

3.2. A law of large numbers for sets

In terms of sets, estimation is successful if the stochastic set $M(\widehat{\mu}_n)$ gets closer in some sense to the true set $M(\mu)$ as $n \to \infty$. In the works [56, 49, 9, 32, 43, 20] two modes of convergence are considered for the sequence $(M(\widehat{\mu}_n))_{n\geq 1}$.

Definition 3.2.
1. $(M(\widehat{\mu}_n))_{n\geq 1}$ is *strongly consistent in outer limit* [43] (alternatively, *in Kuratowski upper limit* [20] or *in the sense of Ziezold* [56, 32]) if

$$\mathbb{P}\Big(\bigcap_{n\geq 1} \overline{\bigcup_{p\geq n} M(\widehat{\mu}_p)} \subset M(\mu)\Big) = 1.$$

2. $(M(\widehat{\mu}_n))_{n\geq 1}$ is *strongly consistent in one-sided Hausdorff distance* [43, 20] (alternatively, *in the sense of Bhattacharya–Patrangenaru* [9]) if

$$\mathbb{P}\Big(\sup_{\alpha\in M(\widehat{\mu}_n)} \inf_{\beta\in M(\mu)} d(\alpha, \beta) \xrightarrow[n\to\infty]{} 0\Big) = 1.$$

In [29, 43, 20] each of these statements is regarded as a set-valued strong law of large numbers. A caveat about Definition 3.2 is that these notions of closeness between $M(\widehat{\mu}_n)$ and $M(\mu)$ are only one-sided: there might exist some $\alpha_\star \in M(\mu)$ such that the distance of $\alpha_\star$ to $M(\widehat{\mu}_n)$ remains bounded away from 0 with positive probability. However, in the case of a unique Fréchet ℓ-mean (i.e., $M(\mu) = \{\alpha_\star\}$), Definition 3.2 yields strong consistency in the usual sense: for any sequence of measurable selections $(\widehat{\alpha}_n)_{n\geq 1}$, $\mathbb{P}(d(\widehat{\alpha}_n, \alpha_\star) \to 0) = 1$.

Since the metric space that we consider is compact, the two modes of consistency introduced above are equivalent. In [32, 43], strong consistency is obtained for a wide variety of metric spaces and functions ℓ. As an application of these results to our setting, we obtain the following strong law of large numbers.

Lemma 3.3 (Strong law of large numbers)**.** *$(M(\widehat{\mu}_n))_{n\geq 1}$ is strongly consistent in either of the senses of Definition 3.2.*

3.3. Stickiness

In order to simplify the exposition, throughout this subsection we add the requirement that the loss ℓ be increasing, instead of just nondecreasing as in Definition 2.2.

We leverage the geometry of the metric tree T to describe in more detail how $(M(\widehat{\mu}_n))_{n\geq 1}$ converges to $M(\mu)$. To this end, we adapt the concept of stickiness that was introduced in [29] and further explored in [31, 8].

Definition 3.4. Let $c \in T$ with neighboring vertices $v_1, \dots, v_m$. Depending on which of the following disjoint and exhaustive conditions is satisfied, we say that c is:

- *sticky* if $c \in M(\mu)$ and for every $i \in \{1, \dots, m\}$, $\phi'_{v_i}(c) > 0$,
- *partly sticky* if $c \in M(\mu)$ and there exists some $i \in \{1, \dots, m\}$ such that $\phi'_{v_i}(c) = 0$,
- *nonsticky* if c is not in $M(\mu)$.

Remark 3.5. By Proposition 2.13 c is nonsticky if and only if there exists some (unique) $i \in \{1, \dots, m\}$ such that $\phi'_{v_i}(c) < 0$.

Remark 3.6. Originally Hotz et al. [29] defined stickiness in the setting of the Fréchet mean ($\ell : z \mapsto z^2$) on an open book. For us, the compact metric tree that most resembles an open book is the m-spider with center c and leaves $v_1, \dots, v_m$ (if each branch of the m-spider was unbounded, the space would be an open book with spine $\{c\}$). In [29, Definition 2.10], stickiness is defined according to the sign of the quantity

$$m_i = \int_{(c,v_i]} d(c,x)d\mu(x) - \sum_{j\neq i} \int_{(c,v_j]} d(c,x)d\mu(x),$$

which in our notation is exactly $-\frac{1}{2}\phi'_{v_i}(c)$.

When the directional derivative $\phi'_{v_i}(c)$ is nonzero, Proposition 2.12 helps to locate the minimizers of ϕ. As n grows, it is expected that the empirical counterpart $\widehat{\phi}'_{v_i}(c)$ becomes nonzero and has the same sign as $\phi'_{v_i}(c)$ with high probability. It is then possible to obtain identical localization constraints on the empirical Fréchet ℓ-means. The following theorem makes this intuition precise.

Theorem 3.7 (Nonasymptotic empirical stickiness)**.** *Let $c \in T$ and $n \geq 1$ be fixed.*

1. *If c is sticky, then $M(\mu) = \{c\}$ and*

$$\mathbb{P}_*(M(\widehat{\mu}_n) = \{c\}) \geq 1 - \sum_{i=1}^{m} \exp\left(-\frac{\phi'_{v_i}(c)^2}{2\ell'_+(D)^2} n\right).$$

2. *If c is partly sticky, letting $I = \{i \in \{1,\dots,m\} : \phi'_{v_i}(c) = 0\}$, the set I has cardinality 1 or 2, the inclusions $\{c\} \subset M(\mu) \subset \{c\} \cup \bigcup_{i\in I} T_{c\to v_i}$ hold and*

$$\mathbb{P}_*\big(M(\widehat{\mu}_n) \subset \{c\} \cup \bigcup_{i\in I} T_{c\to v_i}\big) \geq 1 - \sum_{i\notin I} \exp\left(-\frac{\phi'_{v_i}(c)^2}{2\ell'_+(D)^2} n\right).$$

3. *If c is nonsticky with $\phi'_{v_i}(c) < 0$, then $M(\mu) \subset T_{c\to v_i}$ and*

$$\mathbb{P}_*(M(\widehat{\mu}_n) \subset T_{c\to v_i}) \geq 1 - \exp\left(-\frac{\phi'_{v_i}(c)^2}{2\ell'_+(D)^2} n\right).$$

The exponential bounds of Theorem 3.7 combined with a Borel–Cantelli argument lead to the following asymptotic result. When c is sticky, with probability 1 the empirical sets of minimizers $M(\widehat{\mu}_n)$ are eventually equal to $\{c\}$. This justifies the usage of the adjective “sticky”.

Corollary 3.8 (Sticky law of large numbers)**.**

1. *If c is sticky,*
$$\mathbb{P}\big(\exists N \geq 1, \forall n \geq N, M(\widehat{\mu}_n) = \{c\}\big) = 1.$$
2. *If c is partly sticky, letting I as in Theorem 3.7,*
$$\mathbb{P}\big(\exists N \geq 1, \forall n \geq N, M(\widehat{\mu}_n) \subset \{c\} \cup \bigcup_{i\in I} T_{c\to v_i}\big) = 1.$$
3. *If c is nonsticky with $\phi'_{v_i}(c) < 0$,*
$$\mathbb{P}\big(\exists N \geq 1, \forall n \geq N, M(\widehat{\mu}_n) \subset T_{c\to v_i}\big) = 1.$$

Remark 3.9. If it is known a priori that $M(\mu)$ is a singleton (e.g., if ℓ is strictly convex), then by Lemma 3.3, the partly sticky statement of Corollary 3.8 can be refined as

$$\mathbb{P}\big(\exists N \geq 1, \forall n \geq N, M(\widehat{\mu}_n) \subset \bigcup_{i\in I} [c, v_i]\big) = 1.$$

The earlier definition of stickiness involves the signs of the directional derivatives at c. It is therefore stated in terms of the landscape of the objective function ϕ around c. The following proposition provides an equivalent formulation of stickiness: c is sticky if and only if the equality $M(\nu) = \{c\}$ holds for every measure ν that is sufficiently close to μ. The notion of stickiness thus has an interpretation in terms of robustness.

We quantify the closeness between two probability measures ν_1, ν_2 using the total variation metric defined as $\mathrm{TV}(\nu_1,\nu_2) = \sup_{B\in\mathcal{B}(T)} |\nu_1(B) - \nu_2(B)|$ and the 1-Wasserstein metric $W_1(\nu_1,\nu_2) = \sup\{\int_T f(x)d\nu_1(x) - \int_T f(x)d\nu_2(x) : f \text{ is 1-Lipschitz}\}$. Total variation is stronger than 1-Wasserstein, in the sense that $W_1(\nu_1,\nu_2) \leq D\,\mathrm{TV}(\nu_1,\nu_2)$ [53, Theorem 6.15]. Note that closeness in W_1 need not imply closeness in total variation.

Proposition 3.10.
1. *c is sticky if and only if there exists $\varepsilon > 0$ such that for every probability measure ν verifying $\mathrm{TV}(\nu,\mu) \leq \varepsilon$ we have $M(\nu) = \{c\}$.*
2. *Under the additional assumption that ℓ is differentiable with Lipschitz derivative, 1. holds in W_1 instead of* TV*.*

Remark 3.11. Connections between stickiness and robustness under perturbations were already explored in the context of stratified spaces by Huckemann et al. [31, Section 7], Bhattacharya et al. [8, Proposition 2.8] and most recently by Lammers et al. [33].

4. The special case of Fréchet medians

Now, we restrict our focus to Fréchet medians, i.e., the case where the loss is $\ell : z \mapsto z$. Among the operations research community, the median case has generated the most interest, as it is the most intuitive in applications: the practitioner looks for a new facility on the network that minimizes the average travel time to the demand.

4.1. Further descriptive results

In Proposition 2.16 we observed that the set of medians $M_1(\mu)$ was a geodesic segment. Besides, for a discrete measure with uniform weights $\nu = \frac{1}{m}\sum_{i=1}^m \delta_{x_i}$ on the real line $\mathbb{R}$, it is well-known that $M_1(\nu)$ contains at least one of the x_i. We provide a generalization of this fact on a tree: an extremity of $M_1(\mu)$ is a vertex of T, or it is in the support of μ (i.e., the smallest closed subset of T that has μ-probability 1 [39, Theorem 2.1]).

Proposition 4.1. *Let α_1 and α_2 denote the endpoints of the geodesic segment $M_1(\mu)$, i.e., $M_1(\mu) = [\alpha_1, \alpha_2]$. The following inclusion holds: $\{\alpha_1, \alpha_2\} \subset \mathcal{V} \cup \operatorname{supp}(\mu)$, where $\mathcal{V}$ is the set of vertices of T.*

Remark 4.2. Hakimi [24] states a weaker statement: when μ is discrete and supported on $\mathcal{V}$, he proves that $M_1(\mu) \cap \mathcal{V} \neq \emptyset$.

A measure ν on $\mathbb{R}$ has at least two medians if and only if there exists $m_1 < m_2$ such that $\nu\big((-\infty, m_1]\big) = \nu\big([m_2, -\infty)\big) = \frac{1}{2}$ [40, Corollary 2.6]. The next proposition is an extension of this fact to metric trees.

Proposition 4.3. *μ has more than one Fréchet median if and only if there exist G_1, G_2 two disjoint closed convex subsets of T such that $\mu(G_1) = \mu(G_2) = \frac{1}{2}$.*

In the next subsections, a convex subset G is known that contains $M_1(\mu)$. It is fruitful to consider the metric projection on G (for the definition and basic properties of the metric projection on a closed convex subset of a Hadamard space see, e.g., [6, Theorem 2.1.12]) and transform μ into a measure supported on G, hence the following definition.

Definition 4.4. Let G be a closed convex subset of T, and let $\pi : T \to T$ denote the metric projection on G. We denote by $\pi\#\mu$ the pushforward measure of μ by π, and we write $\phi_{\pi\#\mu}$ for the objective function corresponding to $\pi\#\mu$.

Remark 4.5. Although the image of π is G, we define π as a map with codomain T so that $\pi\#\mu$ remains naturally a Borel measure on T.

The following technical proposition gathers statements on $\pi\#\mu$ that will prove useful in the next subsections.

Proposition 4.6.
1. *The set $\pi(T \setminus G)$ is finite. We write $\pi(T \setminus G) = \{v_1, \dots, v_m\}$ and we define the sets $T_i = \pi^{-1}(\{v_i\})$.*
2. *$\pi\#\mu$ is a Borel measure on T. It rewrites explicitly as $\pi\#\mu = \mu_{|\mathring{G}} + \sum_{i=1}^m \mu(T_i)\delta_{v_i}$.*
3. *$M_1(\pi\#\mu)$ is a subset of G.*
4. *ϕ and $\phi_{\pi\#\mu}$ differ by a constant over G. More precisely,*

$$\forall \alpha \in G,\ \phi(\alpha) = \phi_{\pi\#\mu}(\alpha) + \sum_{i=1}^m \int_{T_i} d(v_i, x)\, d\mu(x).$$

5. *The following inclusion holds:* $M_1(\mu) \cap G \subset M_1(\pi\#\mu)$.
6. *Assume that* $M_1(\mu) \subset G$. *Then* $M_1(\mu) = M_1(\pi\#\mu)$.

4.2. Further statistical results

We are now ready to return to the statistical side. In what follows we assume that $M_1(\mu) = \{\alpha_\star\}$, i.e., there is a unique Fréchet median $\alpha_\star$. A sufficient condition for uniqueness was given in Proposition 4.3. We are thus in the classical setting of parameter estimation. However the empirical set $M_1(\widehat{\mu}_n)$ may not be a singleton; we consider therefore an arbitrary sequence of measurable selections $(\widehat{\alpha}_n)_{n\geq 1}$, as explained in Section 3.1.

When $\alpha_\star$ is sticky, Corollary 3.8 asserts that $(\widehat{\alpha}_n)_{n\geq 1}$ converges almost surely to $\alpha_\star$ at an arbitrarily fast rate. From a statistical standpoint the sticky case is thus fully elucidated.

Let $v_1, \dots, v_m$ denote the neighboring vertices of $\alpha_\star$. In Theorem 3.7 it was seen for the partly sticky case that there are at most two i's such that $\phi'_{v_i}(\alpha_\star) = 0$. The properties of $\widehat{\alpha}_n$ must therefore be studied in two distinct cases, which we denominate as follows.

Definition 4.7. Consider the case where $\alpha_\star$ is partly sticky. We say that $\alpha_\star$ is *one-sidedly partly sticky* if there is a unique i such that $\phi'_{v_i}(\alpha_\star) = 0$. Otherwise, we say that $\alpha_\star$ is *two-sidedly partly sticky.*

4.2.1. Improved concentration bounds for Fréchet medians

The exponential bounds in Theorem 3.7 were obtained by estimating probabilities of the type $\mathbb{P}(\widehat{\phi}'_v(\alpha) \leq 0)$, where the population derivative verifies $\phi'_v(\alpha) > 0$. To do so, we employed Hoeffding's inequality for the sum of bounded independent random variables, which yields

$$\mathbb{P}(\widehat{\phi}'_v(\alpha) \leq 0) \leq \exp\left(-\frac{\phi'_v(\alpha)^2}{2\ell'_+(D)^2} n\right). \tag{7}$$

In the case of medians, $\phi'_v(\alpha) = 1 - 2\mu(T_{\alpha\to v})$ and $\mathbb{P}(\widehat{\phi}'_v(\alpha) \leq 0) = \mathbb{P}(\widehat{\mu}_n(T_{\alpha\to v}) \geq \frac{1}{2})$. Since $n\widehat{\mu}_n(T_{\alpha\to v})$ is a sum of n i.i.d. Bernoulli random variables, each with parameter $\mu(T_{\alpha\to v}) = \frac{1-\phi'_v(\alpha)}{2}$, a finer Chernoff bound [16, Theorem 1 and Example 3] provides

$$\begin{aligned}\mathbb{P}(\widehat{\phi}'_v(\alpha) \leq 0) &\leq \left(2\sqrt{\mu(T_{\alpha\to v})(1-\mu(T_{\alpha\to v}))}\right)^n \\ &= \exp\Big(-\frac{1}{2}\big|\ln\big(1-\phi'_v(\alpha)^2\big)\big|\, n\Big),\end{aligned} \tag{8}$$

where the upper bound is 0 if $\phi'_v(\alpha) = 1$. Identical inequalities hold when the signs are reversed (i.e., when $\phi'_v(\alpha) < 0$ and we bound $\mathbb{P}(\widehat{\phi}'_v(\alpha) \geq 0)$ from above). In the median case, $\ell'_+(D) = 1$. Comparing the displays (7) and (8), we note that for every $a \in (-1, 1)$, $|\ln(1-a^2)| \geq a^2$, and as $a \to \pm 1$, a^2 goes to 1 whereas $|\ln(1-a^2)|$ goes to ∞. The bound (8) improves therefore significantly on (7), as it is fully sensitive to the magnitude of $\phi'_v(\alpha)$. In what follows we will state exponential bounds that resemble (8).

4.2.2. The two-sided partly sticky case

Assume without loss of generality that $\phi'_{v_1}(\alpha_\star) = \phi'_{v_2}(\alpha_\star) = 0$ and for all $i > 2$ (if any), $\phi'_{v_i}(\alpha_\star) > 0$. By the law of large numbers in Lemma 3.3, it is known that $\mathbb{P}(d(\widehat{\alpha}_n, \alpha_\star) \to 0) = 1$. By the sticky law of large numbers in Corollary 3.8, we know additionally that $\widehat{\alpha}_n$ lies eventually in the subset $\{c\} \cup \bigcup_{i \in I} T_{c \to v_i}$ with probability 1. As a consequence, $\widehat{\alpha}_n$ is eventually in the geodesic segment $[v_1, v_2]$ with probability 1. The closed convex subset on which we will project the measure μ and the data is therefore $G = [v_1, v_2]$. By assumption it contains the true median $\alpha_\star$.

G is naturally isometric to the compact interval $[-d(\alpha_\star, v_2), d(\alpha_\star, v_1)] \subset \mathbb{R}$, where $\alpha_\star$ is sent on 0. By pushing forward again, this time with target space $\mathbb{R}$, we replace the problem with the analysis of sample medians on the real line. This motivates the next definition and the lemma that follows.

Definition 4.8. Let $\gamma : [-d(\alpha_\star, v_2), d(\alpha_\star, v_1)] \to [v_2, v_1]$ denote the unit-speed geodesic from v_2 to v_1, and let I denote the inverse of γ. For each $n \geq 1$ we define $Y_n = I(\pi(X_n))$, $\widehat{m}_n = I(\pi(\widehat{\alpha}_n))$ and the event $\Omega_n = \big\{\widehat{\alpha}_n \in [v_1, v_2]\big\}$. We denote by ν the pushforward measure $(I \circ \pi)\#\mu$, and by Y a random variable with distribution ν.

Remark 4.9. For convenience, we also use the notation $M_1(\cdot)$ to denote the set of medians of a measure on $\mathbb{R}$ (which is not a metric tree by our definition).

Lemma 4.10.
1. *ν is a Borel measure on $\mathbb{R}$ supported on the compact interval $[-d(\alpha_\star, v_2), d(\alpha_\star, v_1)]$, the Y_n are i.i.d. with distribution ν and $M_1(\nu) = \{0\}$.*
2. *On the event Ω_n, $\widehat{m}_n \in M_1\left(\frac{1}{n}\sum_{k=1}^n \delta_{Y_k}\right)$ and $d(\widehat{\alpha}_n, \alpha_\star) = |\widehat{m}_n - 0| = |\widehat{m}_n|$.*
3. *$\mathbb{P}(\Omega_n) \geq 1 - \exp\left(-\frac{1}{2}\left|\ln\left(1 - 4\mu\big((\alpha_\star, v_1)\big)^2\right)\right| n\right) - \exp\left(-\frac{1}{2}\left|\ln\left(1 - 4\mu\big((\alpha_\star, v_2)\big)^2\right)\right| n\right)$.*

Before we can state a central limit theorem, we define the following function.

Definition 4.11. The *two-sided branch mass function* Δ is

$$\begin{aligned} \Delta \colon [-d(\alpha_\star, v_2), d(\alpha_\star, v_1)] &\to [0, \infty) \\ t &\mapsto \mu\big((\alpha_\star, \gamma_t)\big). \end{aligned}$$

Δ plays an important role: the next result shows that its rate of decay as $t \to 0$ drives the rate of convergence of $\widehat{\alpha}_n$ and the asymptotic distribution of a properly rescaled version of $\widehat{m}_n$.

Theorem 4.12 (Two-sided sticky central limit theorem)**.** *Assume that Δ has the following asymptotic expansion as $t \to 0$:*

$$\Delta(t) = Kt^a + o(t^a),$$

for some constants $a > 0$ and $K > 0$. Let Z denote a random variable with standard normal distribution.

1. *$n^{1/(2a)}\widehat{m}_n$ converges in distribution to the random variable $\operatorname{sgn}(Z)\left(\frac{|Z|}{2K}\right)^{1/a}$.*
2. *$n^{1/(2a)} d(\widehat{\alpha}_n, \alpha_\star)$ converges in distribution to the random variable $\left(\frac{|Z|}{2K}\right)^{1/a}$.*

Corollary 4.13. *Assume that Δ is differentiable at 0 with positive derivative. Then $\sqrt{n}\widehat{m}_n$ is asymptotically normal with asymptotic variance $\frac{1}{4[\Delta'(0)]^2}$.*

Remark 4.14. As is the case on $\mathbb{R}$ [44, Theorem 2.3.3.A], under a differentiability condition the empirical Fréchet median exhibits normal fluctuations around the population median and converges at the speed $\sqrt{n}$. The convergence gets arbitrarily fast or slow according to the rate of decay of Δ at 0.

Δ also plays a key role in the concentration bound stated next.

Theorem 4.15. *Let $n \geq 1$ be fixed. For t such that $0 < t \leq \min(d(\alpha_\star, v_1), d(\alpha_\star, v_2))$, the following concentration bound holds:*

$$\mathbb{P}(d(\widehat{\alpha}_n, \alpha_\star) \geq t) \leq \exp\big(-\tfrac{1}{2}\big|\ln\big(1-4\Delta^2(t)\big)\big|\ n\big) + \exp\big(-\tfrac{1}{2}\big|\ln\big(1-4\Delta^2(-t)\big)\big|\ n\big). \tag{9}$$

More generally, for every $t > 0$:

$$\begin{aligned}\mathbb{P}(d(\widehat{\alpha}_n, \alpha_\star) \geq t) \leq \quad & \mathbb{1}_{t \leq d(\alpha_\star, v_1)} e^{-\frac{n}{2}|\ln(1-4\Delta^2(t))|} + \mathbb{1}_{t \leq d(\alpha_\star, v_2)} e^{-\frac{n}{2}|\ln(1-4\Delta^2(-t))|} \\ & + (\mathbb{1}_{t > d(\alpha_\star, v_1)} + \mathbb{1}_{t > d(\alpha_\star, v_2)})\mathbb{P}(\Omega_n^c).\end{aligned}$$

Remark 4.16. The condition $\phi'_{v_1}(\alpha_\star) = \phi'_{v_2}(\alpha_\star) = 0$ rewrites as $\mu(T_{\alpha_\star \to v_1}) = \mu(T_{\alpha_\star \to v_2}) = \frac{1}{2}$. Consequently, for a fixed $t \in [-d(\alpha_\star, v_2), d(\alpha_\star, v_1)]$, we have $\Delta(t) \leq \frac{1}{2}$, and as $\Delta(t)$ gets closer to $\frac{1}{2}$ the term $\exp\big(-\frac{1}{2}\big|\ln\big(1-4\Delta^2(t)\big)\big|\ n\big)$ in (9) gets closer to 0.

4.2.3. The one-sided partly sticky case

We turn to the partly sticky case with $\phi'_{v_1}(\alpha_\star) = 0$ and for all $i \geq 2$ (if any), $\phi'_{v_i}(\alpha_\star) > 0$. We proceed similarly as in the two-sided partly sticky case. Here, the closed convex subset on which we project is $G = [\alpha_\star, v_1]$.

Definition 4.17. Let $\gamma_1 : [0, d(\alpha_\star, v_1)] \to [\alpha_\star, v_1]$ denote the unit-speed geodesic from $\alpha_\star$ to v_1, and let I denote the inverse of γ_1. For each $n \geq 1$ we define $Y_n = I(\pi(X_n))$, $\widehat{m}_n = I(\pi(\widehat{\alpha}_n))$ and the event $\Omega_n = \big\{\widehat{\alpha}_n \in [\alpha_\star, v_1]\big\}$. We denote by ν the pushforward measure $(I \circ \pi)\#\mu$, and by Y a random variable with distribution ν.

Lemma 4.18.
1. *ν is a Borel measure on $\mathbb{R}$ supported on the compact interval $[0, d(\alpha_\star, v_1)]$, the Y_n are i.i.d. with distribution ν and $M_1(\nu) = \{0\}$.*
2. *On the event Ω_n, $\widehat{m}_n \in M_1\Big(\frac{1}{n}\sum_{k=1}^n \delta_{Y_k}\Big)$ and $d(\widehat{\alpha}_n, \alpha_\star) = |\widehat{m}_n - 0| = \widehat{m}_n$.*
3. *$\mathbb{P}(\Omega_n) \geq \exp\big(-\frac{1}{2}\big|\ln\big(1-4\mu\big((\alpha_\star, v_1)\big)^2\big)\big|\ n\big) - \sum_{i=2}^m \exp\Big(-\frac{1}{2}\big|\ln\big(1-\phi'_{v_i}(\alpha_\star)^2\big)\big|\ n\Big)$.*

Definition 4.19. For each $i \in \{1, \dots, m\}$ we define $\gamma_i : [0, d(\alpha_\star, v_i)] \to [\alpha_\star, v_i]$ the unit-speed geodesic from $\alpha_\star$ to v_i and the i-th *branch mass function*

$$\begin{aligned}\delta_i \colon [0, d(\alpha_\star, v_i)] &\to [0, \infty) \\ t &\mapsto \mu\big((\alpha_\star, \gamma_{i,t})\big).\end{aligned}$$

Theorem 4.20 (One-sided sticky central limit theorem)**.** *Assume that δ_1 has the following expansion as $t \to 0^+$:*

$$\delta_1(t) = Kt^a + o(t^a),$$

for some constants $a > 0$ and $K > 0$. Let Z denote a random variable with standard normal distribution.

1. *$n^{1/(2a)}\widehat{m}_n$ converges in distribution to the random variable $\frac{1}{2K}\max(0, Z)^{1/a}$.*
2. *$n^{1/(2a)} d(\widehat{\alpha}_n, \alpha_\star)$ converges in distribution to the random variable $\frac{1}{2K}\max(0, Z)^{1/a}$.*

Remark 4.21. The rate of convergence $n^{1/(2a)}$ is the same as in the two-sided partly sticky case. In contrast however, the fluctuations are one-sided along the edge $[\alpha_\star, v_1]$.

Theorem 4.22. *Let $n \geq 1$ be fixed. For t such that $0 < t \leq \min_{1 \leq i \leq m} d(\alpha_\star, v_i)$, the following concentration bound holds:*

$$\mathbb{P}(d(\widehat{\alpha}_n, \alpha_\star) \geq t) \leq \exp\big(-\tfrac{1}{2}\big|\ln\big(1 - 4\delta_1^2(t)\big)\big|\ n\big) + \sum_{i=2}^{m} \exp\big(-\tfrac{1}{2}\big|\ln\big(1 - [2\delta_i(t) + \phi'_{v_i}(\alpha_\star)]^2\big)\big|\ n\big). \quad (10)$$

More generally, for every $t > 0$:

$$\begin{aligned}\mathbb{P}(d(\widehat{\alpha}_n, \alpha_\star) \geq t) \leq \quad & \mathbb{1}_{t \leq d(\alpha_\star, v_1)} e^{-\frac{n}{2}|\ln(1-4\delta_1^2(t))|} + \sum_{i=2}^{m} \mathbb{1}_{t \leq d(\alpha_\star, v_i)} e^{-\frac{n}{2}|\ln(1-[2\delta_i(t)+\phi'_{v_i}(\alpha_\star)]^2)|} \\ & + \sum_{i=1}^{m} \mathbb{1}_{t > d(\alpha_\star, v_i)} \mathbb{P}(\Omega_n^c).\end{aligned}$$

Remark 4.23. Since $\phi'_{v_1}(\alpha_\star) = 0$, half of the total mass from μ is on $T_{\alpha_\star \to v_1}$, while the other half is shared among the other $m-1$ branches departing from $\alpha_\star$. If the branch in direction v_i with $i \geq 2$ has very low mass, i.e., if $\mu(T_{\alpha_\star \to v_i})$ is small, then $\phi'_{v_i}(\alpha_\star)$ is close to 1 and the contribution of the term $\exp\big(-\frac{1}{2}\big|\ln\big(1 - [2\delta_i(t) + \phi'_{v_i}(\alpha_\star)]^2\big)\big|\ n\big)$ in (10) is negligible.

References

[1] Afsari, B. (2011). Riemannian L^p center of mass: existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* **139** 655–673. MR2736346

[2] Agueh, M. and Carlier, G. (2011). Barycenters in the Wasserstein space. *SIAM J. Math. Anal.* **43** 904–924. MR2801182

[3] Ahidar-Coutrix, A., Le Gouic, T. and Paris, Q. (2020). Convergence rates for empirical barycenters in metric spaces: curvature, convexity and extendable geodesics. *Probab. Theory Related Fields* **177** 323–368. MR4095017

[4] Aliprantis, C. D. and Border, K. C. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Third ed. Springer, Berlin. MR2378491

[5] Basrak, B. (2010). Limit theorems for the inductive mean on metric trees. *J. Appl. Probab.* **47** 1136–1149. MR2752884

[6] Bačák, M. (2014). *Convex Analysis and Optimization in Hadamard Spaces. De Gruyter Series in Nonlinear Analysis and Applications* **22**. De Gruyter, Berlin. MR3241330

[7] Bhattacharya, A. and Bhattacharya, R. (2012). *Nonparametric inference on manifolds. Institute of Mathematical Statistics (IMS) Monographs* **2**. Cambridge University Press, Cambridge With applications to shape spaces. MR2934285

[8] Bhattacharya, R. and Lin, L. (2017). Omnibus CLTs for Fréchet means and nonparametric inference on non-Euclidean spaces. *Proc. Amer. Math. Soc.* **145** 413–428. MR3565392

[9] Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. MR1962498

[10] Bhattacharya, R. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann. Statist.* **33** 1225–1259. MR2195634

[11] Blanquero, R. and Carrizosa, E. (2013). Solving the median problem with continuous demand on a network. *Comput. Optim. Appl.* **56** 723–734. MR3128755

[12] Brandeau, M. L. and Chiu, S. S. (1988). Parametric facility location on a tree network with an L_p-norm cost function. *Transportation Sci.* **22** 59–69. MR929024

[13] Bridson, M. R. and Haefliger, A. (1999). *Metric Spaces of Non-Positive Curvature. Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **319**. Springer-Verlag, Berlin. MR1744486

[14] Brunel, V.-E. (2023). Geodesically convex M-estimation in metric spaces. In *Proceedings of Thirty Sixth Conference on Learning Theory* (G. Neu and L. Rosasco, eds.). *Proceedings of Machine Learning Research* **195** 2188–2210. PMLR.

[15] Brunel, V.-E. and Serres, J. (2023). Concentration of empirical barycenters in metric spaces. *arXiv preprint arXiv:2303.01144*.

[16] Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statistics* **23** 493–507. MR57518

[17] Eltzner, B., Galaz-García, F., Huckemann, S. and Tuschmann, W. (2021). Stability of the cut locus and a central limit theorem for Fréchet means of Riemannian manifolds. *Proc. Amer. Math. Soc.* **149** 3947–3963. MR4291592

[18] Eltzner, B. and Huckemann, S. F. (2019). A smeary central limit theorem for manifolds with application to high-dimensional spheres. *Ann. Statist.* **47** 3360–3381. MR4025745

[19] Escande, P. (2023). On the Concentration of the Minimizers of Empirical Risks. *arXiv preprint arXiv:2304.00809*.

[20] Evans, S. N. and Jaffe, A. Q. (2023). Limit Theorems for Fréchet Mean Sets. *Bernoulli*. Forth-

coming, https://www.e-publications.org/ims/submission/BEJ/user/submissionFile/49299?confirm=a1a771f4 (version: 2023-09-08).

[21] Fréchet, M. (1947). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. Henri Poincaré* **10** 215–310.

[22] Gadat, S., Gavra, I. and Risser, L. (2018). How to calculate the barycenter of a weighted graph. *Math. Oper. Res.* **43** 1085–1118. MR3884563

[23] Gavra, I. and Risser, L. (2018). Online Barycenter Estimation of Large Weighted Graphs. *arXiv:1803.11137*.

[24] Hakimi, S. L. (1964). Optimum locations of switching centers and the absolute centers and medians of a graph. *Oper. Res.* **12** 450–459.

[25] Handler, G. Y. and Mirchandani, P. B. (1979). *Location on networks theory and algorithms. MIT Press Series in Signal Processing, Optimization, and Control.* MIT Press, Cambridge, Mass.-London. MR525692

[26] Hansen, P., Labbé, M., Peeters, D. and Thisse, J.-F. (1987). Single facility location on networks. In *Surveys in combinatorial optimization (Rio de Janeiro, 1985). North-Holland Math. Stud.* **132** 113–145. North-Holland, Amsterdam. MR878777

[27] Hiriart-Urruty, J.-B. and Lemaréchal, C. (1993). *Convex Analysis and Minimization Algorithms. I. Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **305**. Springer-Verlag, Berlin Fundamentals. MR1261420

[28] Hotz, T. and Huckemann, S. (2015). Intrinsic means on the circle: uniqueness, locus and asymptotics. *Ann. Inst. Statist. Math.* **67** 177–193. MR3297863

[29] Hotz, T., Huckemann, S., Le, H., Marron, J. S., Mattingly, J. C., Miller, E., Nolen, J., Owen, M., Patrangenaru, V. and Skwerer, S. (2013). Sticky central limit theorems on open books. *Ann. Appl. Probab.* **23** 2238–2258. MR3127934

[30] Huber, P. J. and Ronchetti, E. M. (2009). *Robust Statistics*, second ed. *Wiley Series in Probability and Statistics*. John Wiley & Sons, Inc., Hoboken, NJ. MR2488795

[31] Huckemann, S., Mattingly, J. C., Miller, E. and Nolen, J. (2015). Sticky central limit theorems at isolated hyperbolic planar singularities. *Electron. J. Probab.* **20** no. 78, 34. MR3371437

[32] Huckemann, S. F. (2011). Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth. *Ann. Statist.* **39** 1098–1124. MR2816349

[33] Lammers, L., Van, D. T., Nye, T. M. W. and Huckemann, S. F. ([2023] ©2023). Types of Stickiness in BHV Phylogenetic Tree Spaces and Their Degree. In *Geometric science of information. Part I. Lecture Notes in Comput. Sci.* **14071** 357–365. Springer, Cham. MR4638275

[34] Le Gouic, T., Paris, Q., Rigollet, P. and Stromme, A. J. (2023). Fast convergence of empirical barycenters in Alexandrov spaces and the Wasserstein space. *J. Eur. Math. Soc. (JEMS)* **25** 2229–2250. MR4592868

[35] McCormack, A. and Hoff, P. (2022). The Stein effect for Fréchet means. *Ann. Statist.* **50** 3647–3676. MR4524511

[36] Nkansah, P. T. and David, H. T. (1986). Network median problems with continuously distributed demand. *Transportation Sci.* **20** 213–219. MR849974

[37] Ohta, S.-i. (2012). Barycenters in Alexandrov spaces of curvature bounded below. *Adv. Geom.* **12** 571–587. MR3005101

[38] Papadopoulos, A. (2014). *Metric Spaces, Convexity and Nonpositive Curvature*, second ed. *IRMA Lectures in Mathematics and Theoretical Physics* **6**. European Mathematical Society (EMS), Zürich. MR3156529

[39] Parthasarathy, K. R. (1967). *Probability measures on metric spaces. Probability and Mathe-*

matical Statistics **No. 3**. Academic Press, Inc., New York-London. MR226684

[40] ROMON, G. (2022). Statistical properties of approximate geometric quantiles in infinite-dimensional Banach spaces. *arXiv:2211.00035*.

[41] RUDIN, W. (1987). *Real and Complex Analysis*, Third ed. McGraw-Hill Book Co., New York. MR924157

[42] SCHÖTZ, C. (2019). Convergence rates for the generalized Fréchet mean via the quadruple inequality. *Electron. J. Stat.* **13** 4280–4345. MR4023955

[43] SCHÖTZ, C. (2022). Strong laws of large numbers for generalizations of Fréchet mean sets. *Statistics* **56** 34–52. MR4392146

[44] SERFLING, R. J. (1980). *Approximation theorems of mathematical statistics. Wiley Series in Probability and Mathematical Statistics*. John Wiley & Sons, Inc., New York. MR595165

[45] SHAO, J. (2003). *Mathematical Statistics*, second ed. *Springer Texts in Statistics*. Springer-Verlag, New York. MR2002723

[46] SHIER, D. R. and DEARING, P. M. (1983). Optimal locations for a class of nonlinear, single-facility location problems on a network. *Oper. Res.* **31** 292–303. MR698570

[47] SHIRYAEV, A. N. (2016). *Probability. 1*, third ed. *Graduate Texts in Mathematics* **95**. Springer, New York. MR3467826

[48] STURM, K.-T. (2003). Probability measures on metric spaces of nonpositive curvature. In *Heat kernels and analysis on manifolds, graphs, and metric spaces (Paris, 2002). Contemp. Math.* **338** 357–390. Amer. Math. Soc., Providence, RI. MR2039961

[49] SVERDRUP-THYGESON, H. (1981). Strong law of large numbers for measures of central tendency and dispersion of random variables in compact metric spaces. *Ann. Statist.* **9** 141–145. MR600540

[50] TANSEL, B. C., FRANCIS, R. L. and LOWE, T. J. (1983). Location on networks: a survey. I. The p-center and p-median problems. *Management Sci.* **29** 482–497. MR704593

[51] TANSEL, B. C., FRANCIS, R. L. and LOWE, T. J. (1983). Location on networks: a survey. I. The p-center and p-median problems. *Management Sci.* **29** 482–497. MR704593

[52] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics*. Springer-Verlag, New York. MR1385671

[53] VILLANI, C. (2009). *Optimal transport. Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **338**. Springer-Verlag, Berlin Old and new. MR2459454

[54] YOKOTA, T. (2016). Convex functions and barycenter on CAT(1)-spaces of small radii. *J. Math. Soc. Japan* **68** 1297–1323. MR3523548

[55] YUN, H. and PARK, B. U. (2023). Exponential concentration for geometric-median-of-means in non-positive curvature spaces. *Bernoulli* **29** 2927–2960. MR4632126

[56] ZIEZOLD, H. (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the Eighth European Meeting of Statisticians (Tech. Univ. Prague, Prague, 1974), Vol. A* 591–602. Reidel, Dordrecht-Boston, Mass. MR501230

Appendix A: Proofs

A.1. Proofs for Section 2

Lemma A.1. *The function ℓ defined in (3) is convex.*

Proof of Lemma A.1. The following inequality between slopes holds: for any $0 \leq z_1 < z_2 < z_3$,

$$\frac{\ell(z_2) - \ell(z_1)}{z_2 - z_1} = \frac{1}{z_2 - z_1} \int_{z_1}^{z_2} f(t)dt \leq f(z_2) \leq \frac{1}{z_3 - z_2} \int_{z_2}^{z_3} f(t)dt = \frac{\ell(z_3) - \ell(z_2)}{z_3 - z_2}, \tag{11}$$

thus ℓ is convex [38, Proposition 6.2.1]. □

Proof of Lemma 2.4. 1. Since ℓ is convex, it has finite left- and right-derivative at each $z > 0$, with ℓ'_- and ℓ'_+ being nondecreasing [38, Proposition 6.2.7]. Since ℓ is nondecreasing, ℓ'_- and ℓ'_+ are nonnegative, and the function $z \mapsto \big(\ell(z) - \ell(0)\big)/z$ is bounded below by 0, hence $\ell'_+(0)$ is finite.

2. For a fixed $Z > 0$ and $0 \leq z_1 < z_2 \leq Z$, we have the estimate

$$0 \leq \frac{\ell(z_2) - \ell(z_1)}{z_2 - z_1} \leq \ell'_-(z_2) \leq \ell'_-(Z),$$

thus ℓ is locally Lipschitz.

3. By the last point, ℓ is absolutely continuous on compact intervals. Let $Z > 0$ be fixed. By the fundamental theorem of calculus [41, Theorem 7.18], ℓ is differentiable a.e. on $[0, Z]$ and for every $z \in [0, Z]$, $\ell(z) - \ell(0) = \int_0^z f_Z(t)dt$, where f_Z denotes a derivative of ℓ. Since f_Z and ℓ'_+ coincide a.e. on $[0, Z]$, we obtain $\ell(z) - \ell(0) = \int_0^z \ell'_+(t)dt$. □

Proof of Proposition 2.5. 1. Since T is bounded and ℓ is nondecreasing, $\ell(d(\alpha, x)) \leq \ell(D)$, hence ϕ is well-defined. That ϕ is continuous follows from continuity of ℓ (seen in Lemma 2.4) and the dominated convergence theorem. Since T is Hadamard, by [38, Example 8.4.7 (i)] the map $\alpha \mapsto d(\alpha, x)$ is convex for each $x \in T$, and by the convexity and monotonicity of ℓ we obtain convexity of $\alpha \mapsto \ell(d(\alpha, x))$. That ϕ is convex follows by integration.

2. T is compact and ϕ is continuous, hence $M(\mu)$ is nonempty. By [6, Example 2.1.3], $M(\mu)$ is closed and convex.

3. If ℓ'_+ is increasing, the inequality (11) between slopes is strict, hence ℓ is strictly convex. If ℓ'_+ is not increasing, there exists an open interval $I \subset [0, \infty)$ where ℓ'_+ is equal to some constant C. By [38, Proposition 6.2.7], ℓ'_- is also equal to C, hence ℓ is differentiable over I with derivative C, thus ℓ is affine over I and ℓ is not strictly convex.

We suppose now that ℓ is strictly convex. Assume for the sake of contradiction that ϕ is not strictly convex: there exists a geodesic $\gamma : [0, 1] \to T$ and $t \in (0, 1)$ such that $\phi(\gamma_t) = (1-t)\phi(\gamma_0) + t\phi(\gamma_1)$, i.e.,

$$0 = \int_T \big((1-t)\ell(d(\gamma_0, x)) + t\ell(d(\gamma_1, x)) - \ell(d(\gamma_t, x))\big) d\mu(x).$$

The function $x \mapsto (1-t)\ell(d(\gamma_0, x)) + t\ell(d(\gamma_1, x)) - \ell(d(\gamma_t, x))$ is thus nonnegative and has integral 0. Consequently, there exists $x_\star \in T$ such that

$$\ell(d(\gamma_t, x_\star)) = (1-t)\ell(d(\gamma_0, x_\star)) + t\ell(d(\gamma_1, x_\star)). \tag{12}$$

Since ℓ is strictly convex and nondecreasing, the function $\alpha \mapsto \ell(d(\alpha, x_\star))$ is strictly convex as well; this contradicts (12). By [38, Proposition 8.4.5] $M(\mu)$ is a singleton. □

Proof of Proposition 2.7. Decomposing the distance $d(\alpha', x)$ with respect to the location of x in the tree, we obtain the equality

$$\begin{aligned}\phi(\alpha') - \phi(\alpha) = &\int_{T\setminus T_{\alpha\to v}} \Big(\ell\big(d(\alpha,x)+d(\alpha,\alpha')\big) - \ell\big(d(\alpha,x)\big)\Big)d\mu(x) \\ &+ \int_{T_{\alpha'\to v}} \Big(\ell\big(d(\alpha,x)-d(\alpha,\alpha')\big) - \ell\big(d(\alpha,x)\big)\Big)d\mu(x) \\ &+ \int_{(\alpha,\alpha']} \Big(\ell(d(\alpha',x)) - \ell(d(\alpha,x))\Big)d\mu(x).\end{aligned} \tag{13}$$

To obtain the limit (6), we consider a sequence $(\alpha'_n)_{n\geq 1}$ of points in $(\alpha, v]$ that converges to α, and we apply the dominated convergence theorem to each integral in Equation (13). The domination follows from the following estimate: for $\alpha' \neq \alpha$ and $x \in T$, by the convexity of ℓ:

$$\frac{\ell(d(\alpha',x)) - \ell(d(\alpha,x))}{d(\alpha',\alpha)} \leq \ell'_+(d(\alpha',x))\frac{d(\alpha',x)-d(\alpha,x)}{d(\alpha',\alpha)} \leq \ell'_+(D),$$

hence by symmetry

$$\frac{|\ell(d(\alpha',x)) - \ell(d(\alpha,x))|}{d(\alpha',\alpha)} \leq \ell'_+(D).$$

□

Proof of Proposition 2.12. 1. Since $\phi'_v(\alpha_0) < 0$, there exists a one-sided neighborhood N of α_0 such that $N \subset (\alpha_0, v]$ and $\alpha' \in N \implies \phi(\alpha') < \phi(\alpha_0)$. For the sake of contradiction assume the existence of $\alpha_\star \in M(\mu) \cap (T \setminus T_{\alpha_0\to v})$. Fix some $\alpha' \in N$ and let $\gamma : [0,1] \to [\alpha', \alpha_\star]$ be the geodesic from α' to $\alpha_\star$. For some $t \in (0,1)$, $\gamma(t) = \alpha_0$, thus

$$\phi(\alpha_0) = \phi(\gamma_t) \leq (1-t)\phi(\alpha') + t\phi(\alpha_\star) < (1-t)\phi(\alpha_0) + t\phi(\alpha_\star),$$

hence $\phi(\alpha_0) < \phi(\alpha_\star)$, a contradiction.

2. There exists a one-sided neighborhood N of α_0 such that $N \subset (\alpha_0, v]$ and $\alpha' \in N \implies \phi(\alpha') > \phi(\alpha_0)$. For the sake of contradiction assume the existence of $\alpha_\star \in M(\mu) \cap T_{\alpha_0\to v}$. Let $\gamma : [0,1] \to [\alpha_0, \alpha_\star]$ be the geodesic from α_0 to $\alpha_\star$. For small enough positive t, $\gamma(t)$ is in N, thus

$$\phi(\alpha_0) < \phi(\gamma_t) \leq (1-t)\phi(\alpha_0) + t\phi(\alpha_\star),$$

hence $\phi(\alpha_0) < \phi(\alpha_\star)$, a contradiction.

3. We let $\alpha \in T \setminus \{\alpha_0\}$ be arbitrary and we show that $\phi(\alpha_0) \leq \phi(\alpha)$.

Consider first the case where $\alpha \in T_{\alpha_0\to v}$. Letting $\gamma : [0,1] \to [\alpha_0, \alpha]$ be the geodesic from α_0 to α, $\phi(\gamma_t) \leq (1-t)\phi(\alpha_0) + t\phi(\alpha)$, thus for each $t > 0$

$$\frac{\phi(\gamma_t) - \phi(\alpha_0)}{t} \leq \phi(\alpha) - \phi(\alpha_0). \tag{14}$$

Since for small enough t we have $\gamma_t \in (\alpha_0, v]$, passing to the limit yields $\phi'_v(\alpha_0) \leq \phi(\alpha) - \phi(\alpha_0)$, hence $\phi(\alpha_0) \leq \phi(\alpha)$.

If $T \setminus (T_{\alpha_0\to v} \cup \{\alpha_0\})$ is empty, the proof is over. Otherwise we pick w in this set and we consider the case where $\alpha \in T \setminus (T_{\alpha_0\to v} \cup \{\alpha_0\})$. With the geodesic γ from α_0 to α, Equation (14) still holds and taking the limit yields $\phi'_w(\alpha_0) \leq \phi(\alpha) - \phi(\alpha_0)$. For α', α'' such that $\alpha' \in (\alpha_0, v]$ and $\alpha'' \in (\alpha_0, w]$, letting ψ denote the geodesic from α'' to α' and $t = d(\alpha_0, \alpha'')/d(\alpha', \alpha'')$, the convexity inequality $\phi(\psi_t) \leq (1-t)\phi(\alpha'') + t\phi(\alpha')$ rewrites as

$$\frac{\phi(\alpha_0) - \phi(\alpha')}{d(\alpha_0, \alpha')} \leq \frac{\phi(\alpha'') - \phi(\alpha_0)}{d(\alpha'', \alpha_0)}. \tag{15}$$

Taking limits, we obtain $0 = -\phi'_v(\alpha_0) \leq \phi'_w(\alpha_0)$, hence $0 \leq \phi(\alpha) - \phi(\alpha_0)$. □

Proof of Proposition 2.13. 1. If $\alpha \in M(\mu)$, condition (b) follows from nonnegativity of the numerator in (6). Suppose that every neighboring node v satisfies $\phi'_v(\alpha) \geq 0$. If all the $\phi'_v(\alpha)$ are positive, then by combining the inclusions of Proposition 2.12 we have $M(\mu) = \{\alpha\}$. Otherwise $\phi'_v(\alpha) = 0$ for some v and Proposition 2.12 yields $\alpha \in M(\mu)$.

2. For convenience, let $S(\alpha) = \int_T \ell'_+(d(\alpha, x))d\mu(x)$. By the assumption, α has two neighboring nodes: v and w. By 1.(c) and the differentiability of ℓ, $\alpha \in M(\mu)$ if and only if

$$\int_{T_{\alpha \to v}} \ell'(d(\alpha, x))d\mu(x) \leq \frac{S(\alpha)}{2} \quad \text{and} \quad \int_{T_{\alpha \to w}} \ell'(d(\alpha, x))d\mu(x) \leq \frac{S(\alpha)}{2}.$$

Note additionally that

$$S(\alpha) = \int_{T_{\alpha \to v}} \ell'(d(\alpha, x))d\mu(x) + \int_{T_{\alpha \to w}} \ell'(d(\alpha, x))d\mu(x) + \int_{\{\alpha\}} \ell'_+(d(\alpha, x))d\mu(x).$$

Since $\mu(\{\alpha\})\ell'_+(0) = 0$, the rightmost integral is 0, and the claim follows. □

Proof of Proposition 2.16. Assume for the sake of contradiction that $M(\mu)$ is not a geodesic segment. Then it contains a 3-spider G with center c (a vertex of T) and outer vertices v_1, v_2, v_3 (which may not be vertices of T). For $i \in \{1, 2, 3\}$, since $c \in M(\mu)$ we must have $\phi'_{v_i}(c) \geq 0$. Furthermore, since $v_i \in M(\mu)$ the second localization constraint of Proposition 2.12 implies $\phi'_{v_i}(c) = 0$. Since ℓ'_+ is nonnegative and $\ell'_+(z) \geq \ell'_-(z)$ holds for each $z > 0$ we obtain the bound:

$$\begin{aligned} 0 = \phi'_{v_i}(c) &= \int_{\{c\}} \ell'_+(d(c, x))d\mu(x) + \int_{T \setminus (T_{c \to v_i} \cup \{c\})} \ell'_+(d(c, x))d\mu(x) - \int_{T_{c \to v_i}} \ell'_-(d(c, x))d\mu(x) \\ &\geq \int_{T \setminus \{c\}} \ell'_-(d(c, x))d\mu(x) - 2\int_{T_{c \to v_i}} \ell'_-(d(c, x))d\mu(x), \end{aligned}$$

hence $\int_{T_{c \to v_i}} \ell'_-(d(c, x))d\mu(x) \geq \frac{1}{2}\int_{T \setminus \{c\}} \ell'_-(d(c, x))d\mu(x)$. Summing these inequalities, we find

$$\int_{T \setminus \{c\}} \ell'_-(d(c, x))d\mu(x) \geq \sum_{i=1}^{3} \int_{T_{c \to v_i}} \ell'_-(d(c, x))d\mu(x) \geq \frac{3}{2}\int_{T \setminus \{c\}} \ell'_-(d(c, x))d\mu(x), \tag{16}$$

which yields $\int_{T \setminus \{c\}} \ell'_-(d(c, x))d\mu(x) = 0$ and $\mathbb{1}_{T \setminus \{c\}}(x)\ell'_-(d(c, x)) = 0$ for μ-almost every x. Since ℓ is increasing, $\ell_-(z) > 0$ holds for each $z > 0$, thus $\mu = \delta_c$ and $\phi'_{v_1}(c) = 1$. This is a contradiction, hence $M(\mu)$ is a geodesic segment. □

A.2. Proofs for Section 3

Proof of Lemma 3.3. [32, Theorem A.3] or [43, Theorem 3.2]. $\square$

Proof of Theorem 3.7. 1. Let $i \in \{1, \dots, m\}$ be fixed. For each $k \in \{1, \dots, n\}$ we define the random variables

$$Y_k = \mathbb{1}_{T \setminus T_{c \to v_i}}(X_k)\ell'_+(d(c, X_k)) - \mathbb{1}_{T_{c \to v_i}}(X_k)\ell'_-(d(c, X_k)),$$

so that $\widehat{\phi}'_{v_i}(c) = \frac{1}{n}\sum_{k=1}^n Y_k$, the Y_k are i.i.d. and $|Y_k| \leq \ell'_+(d(c, X_k)) \leq \ell'_+(D)$. Since c is sticky, $\phi'_{v_i}(c) > 0$ and by Hoeffding's inequality,

$$\begin{aligned} \mathbb{P}(\widehat{\phi}'_{v_i}(c) \leq 0) &= \mathbb{P}\big(-\widehat{\phi}'_{v_i}(c) - (-\phi'_{v_i}(c)) \geq \phi'_{v_i}(c)\big) \\ &\leq \exp\left(-\frac{\phi'_{v_i}(c)^2}{2\ell'_+(D)^2} n\right). \end{aligned}$$

We note that $\ell'_+(D) > 0$ since ℓ is increasing. Proposition 2.12 yields the equality $M(\mu) = \{c\}$ as well as the inclusion $\left\{\bigcap_{i=1}^m \widehat{\phi}'_{v_i}(c) > 0\right\} \subset \{M(\widehat{\mu}_n) = \{c\}\}$. By subadditivity of $\mathbb{P}^*$,

$$\mathbb{P}_*(M(\widehat{\mu}_n) = \{c\}) \geq 1 - \sum_{i=1}^m \exp\left(-\frac{\phi'_{v_i}(c)^2}{2\ell'_+(D)^2} n\right).$$

2. For the partly sticky case, the implication $i \notin I \implies \phi'_{v_i}(c) > 0$ and the equality

$$\bigcap_{i \notin I}(T \setminus T_{c \to v_i}) = \{c\} \cup \bigcup_{i \in I} T_{c \to v_i}$$

justify the inclusion $M(\mu) \subset \{c\} \cup \bigcup_{i \in I} T_{c \to v_i}$. The bound on the cardinality of I follows from the argument that led to (16). The rest of the proof is similar to the sticky case.

3. The proof in the nonsticky case is also similar to the sticky case. $\square$

Proof of Corollary 3.8. We only deal with the sticky case, as the others are similar. Fix some $\varepsilon > 0$. Since the series $\sum_{n \geq 1}\sum_{i=1}^m \exp\left(-\frac{\phi'_{v_i}(c)^2}{2\ell'_+(D)^2} n\right)$ has a finite sum, there exists $p \geq 1$ such that $\sum_{n=p}^\infty \sum_{i=1}^m \exp\left(-\frac{\phi'_{v_i}(c)^2}{2\ell'_+(D)^2} n\right) \leq \varepsilon$. Note the inclusion $\limsup_n\{M(\widehat{\mu}_n) \neq \{c\}\} \subset \bigcup_{n \geq p}\{M(\widehat{\mu}_n) \neq \{c\}\}$. Since $\mathbb{P}^*$ is monotonic and subadditive,

$$\mathbb{P}^*\big(\limsup_n\{M(\widehat{\mu}_n) \neq \{c\}\}\big) \leq \sum_{n=p}^\infty \mathbb{P}^*\big(\{M(\widehat{\mu}_n) \neq \{c\}\}\big).$$

By Theorem 3.7, $\mathbb{P}^*\big(\{M(\widehat{\mu}_n) \neq \{c\}\}\big) \leq \sum_{i=1}^m \exp\left(-\frac{\phi'_{v_i}(c)^2}{2\ell'_+(D)^2} n\right)$ and finally

$$\mathbb{P}^*\big(\limsup_n\{M(\widehat{\mu}_n) \neq \{c\}\}\big) \leq \varepsilon.$$

This inequality is valid for every ε, therefore $\mathbb{P}^*\big(\limsup_n\{M(\widehat{\mu}_n) \neq \{c\}\}\big) = 0$ and $\mathbb{P}_*\big(\liminf_n\{M(\widehat{\mu}_n) = \{c\}\}\big) = 1$. By the completeness assumption on the probability space, the set $\liminf_n\{M(\widehat{\mu}_n) = \{c\}\}$ is in $\mathcal{F}$ and has probability 1.

$\square$

Proof of Proposition 3.10. In this proof we will say for convenience that c *is μ-sticky* if c is sticky when the measure under study is μ, and we say similarly that c *is μ-partly sticky*.

1. Suppose c is μ-sticky and define $\varepsilon = \min_j \phi'_{v_j}(c)/\big(4\ell'_+(D)\big)$. Let ν be such that $\mathrm{TV}(\nu,\mu) \le \varepsilon$, let $\widetilde{\phi}$ denote the objective function associated to ν and fix some $i \in \{1,\dots,m\}$. Since the function

$$\varphi : x \mapsto \mathbb{1}_{T\setminus T_{c\to v_i}}(x)\ell'_+(d(c,x)) - \mathbb{1}_{T_{c\to v_i}}(x)\ell'_-(d(c,x))$$

is bounded by $\ell'_+(D)$, we have the estimate

$$|\widetilde{\phi}'_{v_i}(c) - \phi'_{v_i}(c)| = \left|\int_T \varphi(x)d\nu(x) - \int_T \varphi(x)d\mu(x)\right| \le 2\ell'_+(D)\,\mathrm{TV}(\nu,\mu) \le \frac{\phi'_{v_i}(c)}{2},$$

which implies $\widetilde{\phi}'_{v_i}(c) > 0$, thus c is ν-sticky and $M(\nu) = \{c\}$.

Conversely for some $\varepsilon > 0$, we suppose that $\forall \nu, \mathrm{TV}(\nu,\mu) \le \varepsilon \implies M(\nu) = \{c\}$ and we assume for the sake of contradiction that c is not μ-sticky. Since $c \in M(\mu)$, c is μ-partly sticky and there exists i with $\phi'_{v_i}(c) = 0$, i.e., $\int_T \varphi(x)d\mu(x) = 0$. Since ℓ is increasing, $\ell'_-(d(c,v_i)) > 0$ hence $\varphi(v_i) < 0$. Next, define the mixture measure $\nu = (1-\varepsilon)\mu + \varepsilon\delta_{v_i}$ so that

$$\widetilde{\phi}'_{v_i}(c) = (1-\varepsilon)\phi'_{v_i}(c) + \varepsilon\varphi(v_i) = \varepsilon\varphi(v_i) < 0 \tag{17}$$

and $\mathrm{TV}(\nu,\mu) \le \varepsilon$. By our initial assumption, the closeness of the measures implies $M(\nu) = \{c\}$, which contradicts the inequality (17).

3. If ℓ is differentiable and ℓ' is M-Lipschitz, then φ is $2M$-Lipschitz. The previous arguments readily adapt with the 1-Wasserstein distance by leveraging the equality $\mathrm{TV}(\nu_1,\nu_2) = \frac{1}{2}\sup\{\int_T f(x)d\nu_1(x) - \int_T f(x)d\nu_2(x) \mid f : T \to [-1,1] \text{ is measurable}\}$ [47, Lemma 1 p.432].

□

A.3. Proofs for Section 4

Proof of Proposition 4.1. Assume for the sake of contradiction that α_1 is neither a vertex nor a point in the support. Then α_1 lies in the interior of an edge $[v,w]$, where

$$[v,\alpha_1) \cap M_1(\mu) = \emptyset. \tag{18}$$

Since $T\setminus\mathrm{supp}(\mu)$ is open, so is the intersection $\big(T\setminus\mathrm{supp}(\mu)\big)\cap(v,w)$, which contains α_1. Consequently, we can pick some $\alpha \in \big(T\setminus\mathrm{supp}(\mu)\big)\cap(v,\alpha_1)$ that verifies $\mu\big([\alpha,\alpha_1]\big) = 0$. This last equality implies $\mu(T_{\alpha\to v}) = \mu(T_{\alpha_1\to v})$ and $\mu(T_{\alpha\to w}) = \mu(T_{\alpha_1\to w})$, thus $\phi'_v(\alpha) = \phi'_v(\alpha_1)$ and $\phi'_w(\alpha) = \phi'_w(\alpha_1)$. Therefore α verifies the same optimality conditions as α_1 and we must have $\alpha \in M_1(\mu)$. However, by construction $\alpha \in (v,\alpha_1)$, which contradicts (18). We proceed identically with α_2. □

Proof of Proposition 4.3. Assume $\alpha_1 \ne \alpha_2$ are elements of $M_1(\mu)$. It is easily seen that $T\setminus T_{\alpha_1\to\alpha_2}$ and $T\setminus T_{\alpha_2\to\alpha_1}$ are disjoint, closed and convex subsets of T. A necessary optimality condition for α_1 is $\phi'_{\alpha_2}(\alpha_1) \ge 0$, which rewrites as $\mu(T\setminus T_{\alpha_1\to\alpha_2}) \ge \frac{1}{2}$. Symmetrically, $\mu(T\setminus T_{\alpha_2\to\alpha_1}) \ge \frac{1}{2}$ hence $\mu(T\setminus T_{\alpha_1\to\alpha_2}) = \mu(T\setminus T_{\alpha_2\to\alpha_1}) = \frac{1}{2}$.

Conversely, assume the existence of such G_1, G_2. Since T is compact the distance between subsets $d(G_1,G_2)$ is positive and attained for some $\alpha_1 \in G_1$, $\alpha_2 \in G_2$, i.e., $d(G_1,G_2) = d(\alpha_1,\alpha_2) > 0$. Since $G_1 \subset T\setminus T_{\alpha_1\to\alpha_2}$ and $G_2 \subset T\setminus T_{\alpha_2\to\alpha_1}$, we obtain $\mu(T\setminus T_{\alpha_1\to\alpha_2}) = \mu(T\setminus T_{\alpha_2\to\alpha_1}) = \frac{1}{2}$, thus $\phi'_{\alpha_2}(\alpha_1) = \phi'_{\alpha_1}(\alpha_2) = 0$, hence $\{\alpha_1,\alpha_2\} \subset M_1(\mu)$. □

Proof of Proposition 4.6. 1. G equipped with the induced metric is a metric tree. By the finiteness assumption on T, the tree G also has finitely many nodes. For $x \in T \setminus G$, $\pi(x)$ is clearly among the nodes of G, hence $\pi(T \setminus G)$ is finite.

2. By [6, Theorem 2.1.12] π is 1-Lipschitz, hence continuous and $\pi\#\mu$ is a Borel measure on T. Given a Borel subset B of T, note that $\pi^{-1}(B)$ rewrites as the disjoint union $\pi^{-1}(B\cap\mathring{G})\cup\bigcup_{i=1}^{m}\pi^{-1}(B\cap\{v_i\})$, with $\pi^{-1}(B\cap\{v_i\}) = T_i$ if $v_i \in B$ and $\pi^{-1}(B\cap\{v_i\}) = \emptyset$ otherwise.

3. Let $\alpha \in T \setminus G$ and assume w.l.o.g. that $\alpha \in T_1$. For any $y \in G$ we have the decomposition $d(\alpha, y) = d(\alpha, v_1) + d(v_1, y)$, thus

$$\phi_{\pi\#\mu}(\alpha) = \int_T \big(d(\alpha, v_1) + d(v_1, \pi(x))\big)d\mu(x) = d(\alpha, v_1) + \phi_{\pi\#\mu}(v_1) > \phi_{\pi\#\mu}(v_1) = \phi_{\pi\#\mu}(\pi(\alpha)).$$

As a consequence, any minimizer of $\phi_{\pi\#\mu}$ lies in G.

4. Fix $\alpha \in G$. We leverage the explicit form of $\pi\#\mu$ and we decompose the distance $d(\alpha, x)$ for $x \in T_i$:

$$\begin{aligned}\phi_{\pi\#\mu}(\alpha) &= \int_{\mathring{G}} d(\alpha, x)d\mu(x) + \sum_{i=1}^{m}\mu(T_i)d(\alpha, v_i)\\ &= \int_{\mathring{G}} d(\alpha, x)d\mu(x) + \sum_{i=1}^{m}\int_{T_i}\big(d(\alpha, x) - d(x, v_i)\big)d\mu(x)\\ &= \phi(\alpha) - \sum_{i=1}^{m}\int_{T_i} d(v_i, x)d\mu(x).\end{aligned}$$

5. Let $\alpha \in M_1(\mu)\cap G$. By 4., α is in $\arg\min_{\alpha\in G}\phi(\alpha)$ and this set is equal to $M_1(\pi\#\mu)$ by 3.

6. By points 3. and 4., $M_1(\mu) = \arg\min_{\alpha\in G}\phi(\alpha) = \arg\min_{\alpha\in G}\phi_{\pi\#\mu}(\alpha) = M_1(\pi\#\mu)$. □

Proof of Lemma 4.10. 1. By Proposition 4.6, $M_1(\pi\#\mu) = M_1(\mu) = \{\alpha_\star\}$, thus $M_1(\nu) = \{0\}$.

2. On the event Ω_n, $\widehat{\alpha}_n \in M_1(\widehat{\mu}_n)\cap[v_1, v_2]$ hence $\widehat{\alpha}_n \in M_1(\pi\#\widehat{\mu}_n)$, which rewrites as $\widehat{\alpha}_n \in M_1\Big(\frac{1}{n}\sum_{k=1}^{n}\delta_{\pi(X_k)}\Big)$. On Ω_n we have therefore $\widehat{m}_n \in M_1\Big(\frac{1}{n}\sum_{k=1}^{n}\delta_{Y_k}\Big)$ and $d(\widehat{\alpha}_n, \alpha_\star) = |\widehat{m}_n - 0| = |\widehat{m}_n|$.

3. Since $\phi'_{v_1}(\alpha_\star) = \phi'_{v_2}(\alpha_\star) = 0$ we have $\mu(T_{\alpha_\star\to v_1}) = \mu(T_{\alpha_\star\to v_2}) = \frac{1}{2}$, hence

$$\mu\big(T \setminus (T_{\alpha_\star\to v_1}\cup T_{\alpha_\star\to v_2})\big) = 0.$$

As a consequence, with probability 1, $\widehat{\phi}'_{v_i}(\alpha_\star) = 1$ holds for every $i > 2$. Moreover,

$$0 > \phi'_{\alpha_\star}(v_1) = 1 - 2\big(\mu([\alpha_\star, v_1)) + \mu(T_{\alpha_\star\to v_2})\big) = -2\mu((\alpha_\star, v_1)).$$

By (8), $\mathbb{P}(\widehat{\phi}'_{\alpha_\star}(v_1) \geq 0) \leq \exp\Big(-\frac{1}{2}\big|\ln\big(1 - \phi'_{\alpha_\star}(v_1)^2\big)\big|n\Big)$. Finally, note that $\mathbb{P}(\Omega_n) \geq \mathbb{P}_*(M_1(\widehat{\mu}_n) \subset [v_1, v_2])$ and perform a union bound to obtain the claim. □

Proof of Theorem 4.12. 1. For each $n \geq 1$ let $Y_{(1)} \leq \ldots \leq Y_{(n)}$ denote the order statistics of the sample $Y_1, \ldots, Y_n$. It is well-known that the set of real medians $M_1\Big(\frac{1}{n}\sum_{k=1}^{n}\delta_{Y_k}\Big)$ is the singleton $\{Y_{\left(\lfloor\frac{n}{2}\rfloor+1\right)}\}$ when n is odd and the interval $[Y_{\left(\lfloor\frac{n}{2}\rfloor\right)}, Y_{\left(\lfloor\frac{n}{2}\rfloor+1\right)}]$ when n is even.

We follow a similar path as the proof of [45, Theorem 5.10] for real quantiles. Fix $t > 0$ and let us determine the limit of $\mathbb{P}(n^{1/(2a)}\widehat{m}_n < t)$. We start with the upper bound

$$\mathbb{P}(\{n^{1/(2a)}\widehat{m}_n < t\} \ \cap\ \Omega_n) \leq \mathbb{P}\Big(\big\{Y_{\left(\lfloor \frac{n}{2} \rfloor\right)} < \frac{t}{n^{1/(2a)}}\big\} \ \cap\ \Omega_n\Big) \leq \mathbb{P}\Big(\sum_{k=1}^{n} \mathbb{1}_{Y_k < \frac{t}{n^{1/(2a)}}} \geq \Big\lfloor \frac{n}{2} \Big\rfloor\Big).$$

Letting $B_n = \sum_{k=1}^n \mathbb{1}_{Y_k < \frac{t}{n^{1/(2a)}}}$, $C_n = \frac{B_n - \mathbb{E}[B_n]}{\mathbb{V}[B_n]}$ and $p_n = \mathbb{P}(Y < \frac{t}{n^{1/(2a)}})$ we obtain

$$\mathbb{P}(\{n^{1/(2a)}\widehat{m}_n < t\} \ \cap\ \Omega_n) \leq \mathbb{P}\Big(C_n \geq \frac{\lfloor \frac{n}{2} \rfloor - np_n}{\sqrt{np_n(1-p_n)}}\Big) = F_{-C_n}\Big(\frac{np_n - \lfloor \frac{n}{2} \rfloor}{\sqrt{np_n(1-p_n)}}\Big). \tag{19}$$

Note that $\lim_n p_n = P(Y \leq 0) = \frac{1}{2}$ and as n goes to infinity,

$$p_n - \frac{1}{2} = \mathbb{P}\big(Y \in (0, t/n^{1/(2a)})\big) = \Delta(t/n^{1/(2a)}) = Kt^a n^{-1/2} + o(n^{-1/2}),$$

therefore

$$\frac{np_n - \lfloor \frac{n}{2} \rfloor}{\sqrt{np_n(1-p_n)}} \xrightarrow[n\to\infty]{} 2Kt^a.$$

By the Lyapunov central limit theorem [45, Example 1.33], C_n converges in distribution to a standard normal, hence so does $-C_n$. By Pólya's theorem [45, Proposition 1.16], $\sup_{x\in\mathbb{R}} |F_{-C_n}(x) - \Phi(x)| \xrightarrow[n\to\infty]{} 0$ (where Φ denotes the cdf of the standard normal distribution) and the RHS of (19) converges to $\Phi(2Kt^a)$. Moreover $\mathbb{P}(\Omega_n) \to 0$, hence

$$\limsup_n \mathbb{P}(n^{1/(2a)}\widehat{m}_n < t) = \limsup_n \mathbb{P}(\{n^{1/(2a)}\widehat{m}_n < t\} \ \cap\ \Omega_n) \leq \Phi(2Kt^a). \tag{20}$$

Now, we turn to the lower bound

$$\mathbb{P}(\{n^{1/(2a)}\widehat{m}_n < t\} \ \cap\ \Omega_n) \geq 1 - \mathbb{P}\Big(\big\{Y_{\left(\lfloor \frac{n}{2} \rfloor + 1\right)} \geq \frac{t}{n^{1/(2a)}}\big\} \ \cap\ \Omega_n\Big) \geq 1 - \mathbb{P}\Big(\sum_{k=1}^{n} \mathbb{1}_{Y_k \geq \frac{t}{n^{1/(2a)}}} \geq \frac{n}{2}\Big)$$

and by the exact same techniques we find that the RHS converges to $\Phi(2Kt^a)$, thus

$$\liminf_n \mathbb{P}(n^{1/(2a)}\widehat{m}_n < t) = \liminf_n \mathbb{P}(\{n^{1/(2a)}\widehat{m}_n < t\} \ \cap\ \Omega_n) \geq \Phi(2Kt^a).$$

Combining with (20) we obtain

$$\forall t > 0,\ \ \mathbb{P}(n^{1/(2a)}\widehat{m}_n < t) \xrightarrow[n\to\infty]{} \Phi(2Kt^a). \tag{21}$$

Next, fix $u > 0$ and an integer $k \geq 1$. Observe that

$$\limsup_n \mathbb{P}(n^{1/(2a)}\widehat{m}_n \leq u) \leq \limsup_n \mathbb{P}(n^{1/(2a)}\widehat{m}_n < u + \tfrac{1}{k}) \overset{(21)}{=} \Phi(2K(u + \tfrac{1}{k})^a).$$

Letting $k \to \infty$ yields $\limsup_n \mathbb{P}(n^{1/(2a)}\widehat{m}_n \leq u) \leq \Phi(2Ku^a)$. Furthermore

$$\liminf_n \mathbb{P}(n^{1/(2a)}\widehat{m}_n \leq u) \geq \liminf_n \mathbb{P}(n^{1/(2a)}\widehat{m}_n < u) \overset{(21)}{=} \Phi(2Ku^a),$$

thus

$$\forall u > 0,\ \mathbb{P}(n^{1/(2a)}\widehat{m}_n \le u) \xrightarrow[n\to\infty]{} \Phi(2Ku^a). \tag{22}$$

Finally, fix $t \le 0$ and note that

$$\mathbb{P}(\{n^{1/(2a)}\widehat{m}_n \le t\} \ \cap\ \Omega_n) = \mathbb{P}\Big(\sum_{k=1}^{n} \mathbb{1}_{Y_k \le \frac{t}{n^{1/(2a)}}} \ge \Big\lfloor \frac{n}{2} \Big\rfloor\Big).$$

Letting $B_n = \sum_{k=1}^{n} \mathbb{1}_{Y_k \le \frac{t}{n^{1/(2a)}}}$, $C_n = \frac{B_n - \mathbb{E}[B_n]}{\mathbb{V}[B_n]}$ and $p_n = \mathbb{P}(Y \le \frac{t}{n^{1/(2a)}})$ we have

$$\mathbb{P}(\{n^{1/(2a)}\widehat{m}_n \le t\} \ \cap\ \Omega_n) \le \mathbb{P}\Big(C_n \ge \frac{\lfloor \frac{n}{2} \rfloor - np_n}{\sqrt{np_n(1-p_n)}}\Big) = F_{-C_n}\Big(\frac{np_n - \lfloor \frac{n}{2} \rfloor}{\sqrt{np_n(1-p_n)}}\Big).$$

Note that $\lim_n p_n = P(Y \le 0) = \frac{1}{2}$ and as n goes to infinity,

$$\frac{1}{2} - p_n = \mathbb{P}\big(Y \in (t/n^{1/(2a)}, 0)\big) = \Delta(t/n^{1/(2a)}) = K|t|^a n^{-1/2} + o(n^{-1/2}),$$

from which we derive the convergence

$$\frac{np_n - \lfloor \frac{n}{2} \rfloor}{\sqrt{np_n(1-p_n)}} \xrightarrow[n\to\infty]{} -2K|t|^a.$$

The rest of the proof is similar to what was done for $t \ge 0$ and we find

$$\forall t \le 0,\ \mathbb{P}(n^{1/(2a)}\widehat{m}_n \le t) \xrightarrow[n\to\infty]{} \Phi(-2K|t|^a). \tag{23}$$

Combining (22) and (23), $n^{1/(2a)}\widehat{m}_n$ converges in distribution to a random variable with cdf $t \mapsto \Phi(2K \operatorname{sgn}(t)|t|^a)$, hence to the random variable $\operatorname{sgn}(Z)\left(\frac{|Z|}{2K}\right)^{1/a}$.

2. On the event Ω_n, we have the equality $d(\widehat{\alpha}_n, \alpha_\star) = |\widehat{m}_n|$. The convergence in distribution of $n^{1/(2a)}\widehat{m}_n$ and the estimate $\mathbb{P}(\Omega_n) \to 0$ are enough to obtain the claim. □

Proof of Corollary 4.13. Since Δ is differentiable it has the Taylor expansion $\Delta(t) = \Delta'(0)t + o(t)$, thus Theorem 4.12 applies with $a = 1$ and $K = \Delta'(0)$. □

Proof of Theorem 4.15. It was seen in the proof of Lemma 4.10 that $\mu(T \setminus (T_{\alpha_\star\to v_1} \cup T_{\alpha_\star\to v_2})) = 0$, thus $\mathbb{P}(\widehat{\alpha}_n \in \{\alpha\} \cup T_{\alpha_\star\to v_1} \cup T_{\alpha_\star\to v_2}) = 1$. Next, note that

$$\begin{aligned}\mathbb{P}(d(\widehat{\alpha}_n, \alpha_\star) \ge t,\ \widehat{\alpha}_n \in \{\alpha\} \cup T_{\alpha_\star\to v_1}) &\le \mathbb{1}_{t \le d(\alpha_\star, v_1)}\mathbb{P}(\widehat{\alpha}_n \notin T_{\gamma_t\to\alpha_\star}) + \mathbb{1}_{t > d(\alpha_\star, v_1)}\mathbb{P}(\Omega_n^c)\\ &\le \mathbb{1}_{t \le d(\alpha_\star, v_1)}\mathbb{P}\big(\widehat{\phi}'_{\alpha_\star}(\gamma_t) \ge 0\big) + \mathbb{1}_{t > d(\alpha_\star, v_1)}\mathbb{P}(\Omega_n^c)\end{aligned}$$

and since $\phi'_{\alpha_\star}(\gamma_t) = 1 - 2\mu(T_{\gamma_t\to\alpha_\star}) = 1 - 2\mu\big(T_{\alpha_\star\to v_2} \cup [\alpha_\star, \gamma_t)\big) = 1 - 2(\frac{1}{2} + \Delta(t)) = -2\Delta(t)$ we obtain the bound

$$\mathbb{P}(d(\widehat{\alpha}_n, \alpha_\star) \ge t,\ \widehat{\alpha}_n \in \{\alpha\} \cup T_{\alpha_\star\to v_1}) \le \mathbb{1}_{t \le d(\alpha_\star, v_1)}e^{-\frac{n}{2}|\ln(1-4\Delta^2(t))|} + \mathbb{1}_{t > d(\alpha_\star, v_1)}\mathbb{P}(\Omega_n^c).$$

Proceeding similarly with v_2 finishes the proof. □

Proof of Theorem 4.20. Similar to the proof of Theorem 4.12. □

Proof of Theorem 4.22. Similar to the proof of Theorem 4.22. □