SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. 11, No. 4, pp. 1195–1224

An Order-Theoretic Perspective on Modes and Maximum A Posteriori Estimation in Bayesian Inverse Problems*

Hefin Lambley† and T. J. Sullivan‡

Abstract. It is often desirable to summarize a probability measure on a space X in terms of a mode, or MAP estimator, i.e., a point of maximum probability. Such points can be rigorously defined using masses of metric balls in the small-radius limit. However, the theory is not entirely straightforward: the literature contains multiple notions of mode and various examples of pathological measures that have no mode in any sense. Since the masses of balls induce natural orderings on the points of X, this article aims to shed light on some of the problems in nonparametric MAP estimation by taking an order-theoretic perspective, which appears to be a new one in the inverse problems community. This point of view opens up attractive proof strategies based upon the Cantor and Kuratowski intersection theorems; it also reveals that many of the pathologies arise from the distinction between greatest and maximal elements of an order, and from the existence of incomparable elements of X, which we show can be dense in X, even for an absolutely continuous measure on $X = \mathbb{R}$.

Key words. Bayesian inverse problems, local behavior of measures, maximum a posteriori estimation, modes of probability measures, orders on metric spaces

MSC codes. 06F99, 28A75, 28C15, 60B05, 62F10, 62R20

DOI. 10.1137/22M154243X

1. Introduction.

In diverse applications such as statistical inference and the analysis of transition paths of random dynamical systems it is desirable to summarize a complicated probability measure μ on a space X by a single distinguished point $x^\star \in X$ that is, in some sense, a "point of maximum probability" under μ—i.e., a *mode* or, in the Bayesian context, a *maximum a posteriori* (*MAP*) *estimator*. Many optimization-based approaches to inverse problems (e.g., Tikhonov regularization of the misfit) aim to calculate or approximate modes, at least heuristically understood. Over the last decade, it has become common to define modes in terms of masses of metric balls in the limit as the ball radius tends to zero, since this makes sense even when X is a very general—possibly infinite-dimensional—space, as is often the case for modern inference problems [26].

*Received by the editors December 21, 2022; accepted for publication (in revised form) May 14, 2023; published electronically October 20, 2023.
https://doi.org/10.1137/22M154243X
Funding: This work has been partially supported by the DFG through project 415980428. The first author is supported by the Warwick Mathematics Institute Centre for Doctoral Training and received funding from the University of Warwick and the UK Engineering and Physical Sciences Research Council (grant EP/W524645/1).
†Mathematics Institute, University of Warwick, Coventry, CV4 7AL, United Kingdom (hefin.lambley@warwick.ac.uk).
‡Mathematics Institute and School of Engineering, University of Warwick, Coventry, CV4 7AL, United Kingdom, and Alan Turing Institute, 96 Euston Road, London, NW1 2DB, United Kingdom (t.j.sullivan@warwick.ac.uk).

However, this "small balls" theory of modes is not entirely straightforward. There are various definitions—e.g., the strong mode of [9], the generalized strong mode of [7], the weak mode of [16]—with various subtle distinctions among them. Even the existence theory for modes is not entirely straightforward: there are already examples in the literature, and this article will supply further examples, of relatively simple probability measures that have no mode. It can even be the case that the average of two disjointly supported unimodal probability measures may have no mode.

The purpose of this article is to formulate the notion of a mode in an order-theoretic manner and thereby to clarify some of these pathologies in the theory of modes. We claim that this is a natural step to take in view of the heuristic understanding of modes as "most probable points."

With an order-theoretic point of view, many of the difficulties can be seen to arise from the distinction between greatest and maximal elements of a preordered set $(X,\preccurlyeq)$ when the preorder $\preccurlyeq$ is not total, i.e., when there exist *incomparable* $x,x' \in X$ for which neither $x \preccurlyeq x'$ nor $x \succcurlyeq x'$ holds. Simply put, a greatest element must dominate every other element of X, whereas a maximal element need only dominate those with which it is comparable; for a total preorder, maximal and greatest elements coincide. Motivated by the needs of inverse problems theory, current notions of modes correspond to greatest elements. However, many preorders lack maximal elements, and even those that have maximal elements may lack greatest elements; this is exactly the situation of the examples discussed in Example 5.7 and Theorem 5.11. Thus, one might argue that current notions of mode are order-theoretically "too strong," and perhaps maximal elements should be considered as modes, but possibly these are "too weak" for the needs of applications communities. We hope that the present article will stimulate discussion on this point.

Outline of the paper. The rest of this paper is structured as follows.

Section 2 sets out basic notation for the rest of the paper, including a brief recap of necessary concepts from functional analysis, measure theory, and order theory.

Section 3 gives an overview of related work in this area, in particular the "small balls" approach to defining MAP estimators for nonparametric statistical inverse problems.

Section 4 introduces and analyzes the total preorder $\preccurlyeq_r$ on X induced by the μ-measures of metric balls of fixed radius $r > 0$. Because the preorder $\preccurlyeq_r$ is total, its maximal elements are also greatest and can be seen as approximate "radius-r modes" for μ. We are able to provide several criteria for the existence of such radius-r modes $x_r^\star$ (Theorem 4.6) as well as examples of measures that admit none (Examples 4.7 and 4.8). As a prelude to the next section, we also consider the convergence of $x_r^\star$ as $r \to 0$ (Theorems 4.11 and 4.12).

In section 5 we attempt to take the limit as $r \to 0$ of the preorders $\preccurlyeq_r$ to define a preorder $\preccurlyeq_0$ whose greatest elements will be weak modes of μ. However, because the preorder $\preccurlyeq_0$ is not total, the distinction between greatest and maximal elements becomes important. Incomparable maximal elements are particularly troubling because their maximality means that one would like to think of them as candidate modes, yet their incomparability means that one cannot actually say which is "most probable" and hence a bona fide mode, as in Example 5.7. We show that antichains (collections of mutually incomparable elements) can be topologically dense in X even when μ is absolutely continuous with respect to Lebesgue measure on $X \subseteq \mathbb{R}$ (Theorem 5.11). We also show that measures with a continuous Lebesgue density may have incomparable elements, but never incomparable *maximal* elements (Example 5.9 and Proposition 5.14).

Section 6 gives some closing remarks, while technical supporting results can be found in section SM1, and section SM2 discusses some alternatives to the limiting preorder $\preccurlyeq_0$ of section 5 and illustrates their shortcomings.

2. Problem setting and notation.

2.1. Spaces of interest. Throughout, unless noted otherwise, X will be a metric space with metric d; we write $\mathscr{B}(X)$ for its Borel σ-algebra, i.e., the one generated by the closed balls $B_r(x) := \{x' \in X \mid d(x,x') \leqslant r\}$, $x \in X$, $r \geqslant 0$; we also write $\mathring{B}_r(x) := \{x' \in X \mid d(x,x') < r\}$ for the corresponding open ball. We will often assume that X is complete and separable, and occasionally we will specialize to the case of X being a separable Banach or Hilbert space.

2.2. Measures of noncompactness and intersection theorems. Our approach in section 4 will make much use of measures of noncompactness and intersection theorems; see [24, sections 7.5–7.8] for a thorough treatment of these concepts and their properties.

Briefly, given $A \subseteq X$, its *separation* (or *Istrăţescu*) *measure of noncompactness* is

$$\gamma(A) := \inf \left\{ r \geqslant 0 \,\middle|\, \text{there is no } (x_n)_{n\in\mathbb{N}} \subseteq A \text{ with } \inf_{\substack{m,n\in\mathbb{N}\\ m\neq n}} d(x_m,x_n) \geqslant r \right\}. \tag{2.1}$$

This is an increasing function with respect to inclusion of sets, is finite precisely when A is bounded, and is zero precisely when A is precompact. The function γ is bi-Lipschitz equivalent with several other measures of noncompactness such as the set (or Kuratowski) measure of noncompactness and the ball (or Hausdorff) measure of noncompactness.

Theorem 2.1 (generalized intersection theorem). *Let $(A_n)_{n\in\mathbb{N}}$ be a decreasingly nested sequence of nonempty, closed subsets of a topological space X and let $A := \bigcap_{n\in\mathbb{N}} A_n$.*

(a) *(Cantor) If each A_n is compact, then A is nonempty and compact.*
(b) *(Cantor) If X is a complete metric space and* $\operatorname{diam}(A_n) \to 0$ *as $n \to \infty$, then A is a singleton.*
(c) *(Kuratowski) If X is a complete metric space and $\gamma(A_n) \to 0$ as $n \to \infty$, then A is nonempty and compact.*

2.3. Measure-theoretic concepts. Given a metric space X, $\mathscr{P}(X)$ denotes the set of all probability measures on $\mathscr{B}(X)$. Absolute continuity of μ with respect to ν is denoted $\mu \ll \nu$. The *topological support* of $\mu \in \mathscr{P}(X)$ is

$$\operatorname{supp}(\mu) := \{x \in X \mid \text{for all } r > 0,\ \mu(B_r(x)) > 0\}, \tag{2.2}$$

which is always a closed subset of X, and is nonempty when X is separable (or, equivalently, second countable or Lindelöf) [2, Theorem 12.14].

The n-dimensional Lebesgue measure on $\mathbb{R}^n$ will be denoted λ^n.

The quantity $\mu(B_r(x))$ will play a major role in this work, especially when thought of as a function of $r > 0$ for various choices of $x \in X$; we shall call the map $r \mapsto \mu(B_r(x))$ the *radial cumulative distribution function* (RCDF) and some of its key properties are given in Lemma SM1.1 and Corollary SM1.4.

2.4. Order-theoretic concepts. We summarize here some basic terms from order theory; for a comprehensive introduction to order theory, see, e.g., [10].

In the course of this work, the set X will be equipped with various *preorders* $\preccurlyeq$, i.e., relations satisfying both

(a) *reflexivity*: for all $x \in X$, $x \preccurlyeq x$; and
(b) *transitivity*: for all $x, y, z \in X$, if both $x \preccurlyeq y$ and $y \preccurlyeq z$, then $x \preccurlyeq z$.

For any such preorder, we will write $x \asymp x'$ if both $x \preccurlyeq x'$ and $x \succcurlyeq x'$ hold true, in which case x and x' are called *equivalent*[1] in the preorder; we write $x \prec x'$ if $x \preccurlyeq x'$ but $x \not\succcurlyeq x'$.

If at least one of $x \preccurlyeq x'$ and $x \succcurlyeq x'$ holds true, then we call x and x' *comparable*; if neither holds, then we call them *incomparable* and write $x \parallel x'$. A preorder $\preccurlyeq$ is *total* or *linear* if there are no incomparable elements. A subset of X on which $\preccurlyeq$ is total is called a *chain,* and a subset for which every two distinct elements are incomparable is called an *antichain.*

We highlight and contrast two notions of a "biggest" element for a preorder.

Definition 2.2. *Let X be a set equipped with a preorder $\preccurlyeq$.*
(a) *$g \in X$ is a* greatest element *if, for every $x \in X$, $g \succcurlyeq x$.*
(b) *$m \in X$ is a* maximal element *if, whenever $x \in X$ is such that $m \preccurlyeq x$, it follows that $m \succcurlyeq x$ (and hence $m \asymp x$). Equivalently, m is maximal if there is no $x \in X$ with $x \succ m$.*
(c) *$u \in X$ is an* upper bound *for $A \subseteq X$ if, for all $x \in A$, $u \succcurlyeq x$.*

Note in particular that a greatest element is also a maximal element, but it must additionally be comparable to (and dominate) every element of X. On the other hand, a maximal element is only required to dominate those elements of X with which it is comparable, and those elements could constitute a rather small subset of X.

The most famous statement about the existence of maximal elements is *Zorn's lemma*: under the axiom of choice, if $(X, \preccurlyeq)$ is a preordered space in which every chain $Y \subseteq X$ has an upper bound, then X has at least one maximal element. However, Zorn's lemma says nothing about the existence of greatest elements.

We write $\uparrow Y := \{x \in X \mid x \succcurlyeq y \text{ for some } y \in Y\}$ for the *upward closure* of $Y \subseteq X$ and further write, for $y \in X$, $\uparrow y := \uparrow\{y\} = \{x \in X \mid x \succcurlyeq y\}$, so that $\uparrow Y = \bigcup_{y \in Y} \uparrow y$.

Finally, since many of the preorders we consider will be parametrized by radius $r \geqslant 0$, we will write $\preccurlyeq_r$ for the preorder, $\parallel_r$ for the induced relation of incomparability, $\uparrow_r Y$ for the upward closure of Y with respect to $\preccurlyeq_r$, etc.

3. Overview of related work. Modes, loosely understood as points of maximum probability, arise in many areas of pure and applied mathematics. Two application domains where modes are particularly prominent are the analysis of the transition paths of random dynamical systems and the Bayesian approach to inverse problems.

The random dynamical systems setting is exemplified by mathematical models of chemical reactions using diffusion processes. One is typically interested in the (rare) transitions of the process from one energy well or metastable state to another, and in particular one wishes to understand the transition paths that a diffusion process is most likely to take. This amounts

[1]A preorder $\preccurlyeq$ is called a *partial order* if it is *antisymmetric,* i.e., if $x \asymp x' \implies x = x'$, but almost none of the preorders that we consider will actually be partial orders.

to a study of the modes of the law μ of the diffusion process on the associated path space X; e.g., for a molecule consisting of n atoms in three-dimensional space, $X = C([0,T];\mathbb{R}^{3n})$. The modes of μ are understood as *minimum-action paths,* and the behavior of μ near the mode is quantified using Freidlin–Wentzell theory or large deviations theory [11, 13, 14].

In the Bayesian approach to inverse problems [17, 26], the reconstruction of an X-valued parameter of interest from observed Y-valued data is expressed in the form of a probability measure $\mu \in \mathscr{P}(X)$, the *posterior distribution.* In many modern inverse problems, particularly those coupled to partial differential equations, the space X is an infinite-dimensional function space or a high-dimensional discretization of such a space, e.g., via a system of finite elements.

The posterior measure μ arises from three ingredients: a *prior measure* $\mu_0 \in \mathscr{P}(X)$, which encodes (subjective) beliefs about the parameter that are held in advance of knowing the observation mechanism or the specific data that are observed; a *likelihood model,* i.e., a family of probability measures $L(\,\cdot\,|x) \in \mathscr{P}(Y)$, one for each $x \in X$, which models how observed data would be expected to arise if the parameter value x were the truth; and a specific observed instance of the data, a point $y \in Y$. Strictly speaking, the posterior measure μ is defined as the disintegration (conditional distribution) of the joint measure $\nu(\mathrm{d}x,\mathrm{d}y) := L(\mathrm{d}y|x)\mu_0(\mathrm{d}x) \in \mathscr{P}(X \times Y)$ along the y-fiber [6]. For simplicity, however, we often concentrate on the case that μ has a density with respect to μ_0 given by Bayes' formula,

$$\mu(\mathrm{d}x) = \frac{\exp(-\Phi(x;y))\,\mu_0(\mathrm{d}x)}{\int_X \exp(-\Phi(x';y))\,\mu_0(\mathrm{d}x')}, \tag{3.1}$$

where $\Phi \colon X \times Y \to \mathbb{R}$ is called the *potential.* In simple settings with $\dim Y < \infty$, the Lebesgue probability density of $L(\,\cdot\,|x)$ is proportional to $\exp(-\Phi(x;\,\cdot\,))$ and Φ can be interpreted as a nonnegative misfit functional. The case of infinite-dimensional data, $\dim Y = \infty$, is considerably more subtle and does not generally admit a density for μ with respect to μ_0 as in (3.1); see, e.g., [26, Remark 3.8] and [21, Remark 9].

Since the full posterior distribution μ can be a rather intractable object, it is often desirable to have access to a convenient point summary: the two principal such point estimators are the *conditional mean estimator* (i.e., the mean of μ) and a *maximum a posteriori estimator* (i.e., a mode, or point of maximum probability, for μ), and here we focus on this second approach. Heuristically, at least when $X = \mathbb{R}^d$, a MAP estimator is just an essential maximizer of the Lebesgue density of μ, i.e. a minimizer of the sum of $\Phi(\,\cdot\,;y)$ and the negative logarithm of the Lebesgue density of μ_0. However, this definition is not effective if we have no access to Lebesgue densities; in particular, it makes no sense when $\dim X = \infty$ [27].

To handle the general infinite-dimensional case, various definitions of modes/MAP estimators have been advanced over recent years, and we summarize them here.[2] One approach [12] is to understand a mode of the path measure μ of a diffusion process as a minimizer of the *Onsager–Machlup (OM) functional* I_μ of μ, which is defined by the relation

$$\lim_{r \to 0} \frac{\mu(B_r(x))}{\mu(B_r(x'))} = \frac{\exp(-I_\mu(x))}{\exp(-I_\mu(x'))} \quad \text{for } x, x' \in X. \tag{3.2}$$

[2]The definitions of [7, 9, 16] were all stated in the case of a separable Banach space X, but they generalize easily to the metric setting, as given here. Also, their definitions used open rather than closed balls.

In some sense, I_μ is a formal negative log-density for μ, but it is in general only a partially defined extended-real-valued function. For example, the OM functional of a Gaussian measure on a Hilbert space is finite only on the Cameron–Martin space. The rigorous interpretation of modes as minimizers of I_μ requires considerable care, especially since in some cases it is not even possible to assign $+\infty$ as an exceptional value for I_μ: the ratio in (3.2) may oscillate and fail to converge as $r \to 0$.

A *strong mode* of μ was defined by [9] to be any $x^\star \in X$ such that

$$\lim_{r\to 0} \frac{\mu(B_r(x^\star))}{M_r} = 1, \tag{3.3}$$

$$M_r := \sup_{x\in X} \mu(B_r(x)). \tag{3.4}$$

(By Corollary SM1.2, separability of X ensures that $\operatorname{supp}(\mu) \neq \varnothing$ and $M_r > 0$.) Any strong mode must lie in $\operatorname{supp}(\mu)$, and the ratio in (3.3) is at most 1 for every choice of $x^\star \in X$, so

$$x^\star \text{ is a strong mode} \iff \liminf_{r\to 0} \frac{\mu(B_r(x^\star))}{M_r} \geqslant 1 \iff \limsup_{r\to 0} \frac{M_r}{\mu(B_r(x^\star))} \leqslant 1. \tag{3.5}$$

However, [7] observed that even elementary absolutely continuous measures on $\mathbb{R}$ such as $\mu(E) := \int_{E\cap[-1,1]} |x| \, \mathrm{d}x$ do not have strong modes, even though the Lebesgue density of μ is clearly maximized at ± 1. Therefore, they call $x^\star \in X$ a *generalized strong mode* if, for every positive null sequence $(r_n)_{n\in\mathbb{N}}$, there exists a sequence $(x_n)_{n\in\mathbb{N}}$ converging to $x^\star$ such that

$$\lim_{n\to\infty} \frac{\mu(B_{r_n}(x_n))}{M_{r_n}} = 1. \tag{3.6}$$

Motivated by (3.5), $x^\star \in \operatorname{supp}(\mu) \subseteq X$ is called a *weak mode* [16] if[3]

$$\limsup_{r\to 0} \frac{\mu(B_r(x'))}{\mu(B_r(x^\star))} \leqslant 1 \text{ for all } x' \in X. \tag{3.7}$$

As a point of terminology, [16] were primarily interested in the restricted case that $x' \in x^\star + E$, where $x^\star \in E$ and E is a topologically dense linear subspace of a Banach space X, and [23] later called this case an *E-weak mode*. Conversely, [3] call $x^\star$ satisfying (3.7) a *global weak mode*. Since we are only going to consider global weak modes, we can simply call them *weak modes* without any ambiguity.

Under the assumption that the OM functional I_μ of μ is real-valued on $\varnothing \neq E \subseteq X$ and

$$\text{for some } x \in E \text{ and all } x' \in X \setminus E, \quad \lim_{r\to 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = 0, \tag{3.8}$$

which [3, Definition 3.1] call *property $M(\mu, E)$*, I_μ can be regarded as having the value $+\infty$ on $X \setminus E$ and weak modes are precisely minimizers of this extended version of I_μ. This enabled [3, 4] to establish a stability and convergence theory for weak modes in terms of the

[3]In fact, [16] used "lim" in place of "lim sup" in (3.7), implicitly assuming the existence of the limit. However, as [3] observe, this yields an unsatisfying definition because it excludes the case in which the ratio oscillates, while remaining bounded away from unity, from being a weak mode. The desirable implication "strong mode $\implies$ weak mode" fails for the original "lim" version of the definition, but holds for the "lim sup" version.

Γ-convergence and equicoercivity of the associated OM functionals.[4] Furthermore, as we show below in Lemma 5.3, weak modes are exactly the greatest elements of a natural preorder $\preccurlyeq_0$ on X, namely the one induced by the limiting ratios of masses of balls in the small-radius limit (Definition 5.1).

There are also local versions of the strong and weak modes [1], in which $x^\star$ is only compared to points in a sufficiently small ball $B_\delta(x^\star)$, $\delta > 0$, analogous to local maximizers of the Lebesgue probability density function/local minimizers of the OM functional.

For μ of the form (3.1) with μ_0 Gaussian and X Hilbert, [9] proved that μ has a strong mode by studying maximizers of the radius-r ball mass $x \mapsto \mu(B_r(x))$ for fixed $r > 0$—which we call radius-r modes in section 4—and arguing that a sequence of such maximizers must converge to a strong mode. The arguments of [9] assume the existence of radius-r modes without proof; in subsection 4.2, we prove results on the existence of radius-r modes in various settings but also provide examples that have no such radius-r modes. Despite the contributions of [9], [19], and [18], among others—and our own offerings—a surprising amount is still unknown about the existence of radius-r modes, let alone weak and strong modes, even for "nicely" reweighted Gaussian measures on Banach spaces.

4. The positive-radius preorder.

4.1. Definition and basic properties.

A probability measure on a metric space X induces a family of preorders on X, one for each positive radius, in a very straightforward way.

Definition 4.1 (positive-radius preorder). *Let X be a metric space and let $\mu \in \mathscr{P}(X)$. For each $r > 0$, define a relation $\preccurlyeq_r$ on X by*

$$x \preccurlyeq_r x' \iff \mu(B_r(x)) \leqslant \mu(B_r(x')). \tag{4.1}$$

It is almost trivial to verify that $\preccurlyeq_r$ satisfies the axioms for a preorder. We will write $x \asymp_r x'$ if both $x \preccurlyeq_r x'$ and $x \succcurlyeq_r x'$ hold, and $x \parallel_r x'$ if neither $x \preccurlyeq_r x'$ nor $x \succcurlyeq_r x'$ holds. In fact, though, incomparability never arises for this preorder: totality of the usual order $\leqslant$ on $\mathbb{R}$ implies totality of $\preccurlyeq_r$ on X. Totality implies that the maximal and greatest elements of X with respect to $\preccurlyeq_r$ coincide (Lemma 4.3), which simplifies the discussion considerably.

Upward closures with respect to $\preccurlyeq_r$ are notably well behaved. In particular, Lemma 4.2(b) says that the relation $\preccurlyeq_r$ is *upper semicontinuous* [2, p. 44].

Lemma 4.2 (closedness, boundedness, and noncompactness of upward closures). *Let X be a metric space, let $\mu \in \mathscr{P}(X)$, and fix $r > 0$.*

(a) *For each $t \geqslant 0$, $\{x' \in X \mid \mu(B_r(x')) \geqslant t\}$ is closed.*
(b) *For each $x \in X$, $\uparrow_r x := \{x' \in X \mid x' \succcurlyeq_r x\}$ is closed.*
(c) *For each $t > 0$, $\{x' \in X \mid \mu(B_r(x')) \geqslant t\}$ is bounded, with separation measure of noncompactness $\gamma(\{x' \in X \mid \mu(B_r(x')) \geqslant t\}) \leqslant 2r$.*
(d) *For each $x \in X$ with $\mu(B_r(x)) > 0$, $\uparrow_r x$ is bounded with $\gamma(\uparrow_r x) \leqslant 2r$.*

[4] Frustratingly, while there are some situations in which strong modes can be characterized as minimizers of OM functionals [1, 9], there are also situations in which this correspondence breaks down, even when property $M(\mu, E)$ holds [3, Example B.5].

Proof. Claim (a) is immediate from the upper semicontinuity of the map $x \mapsto \mu(B_r(x))$ (Lemma SM1.1(a)), and (b) is a special case of claim (a).

Now fix $t > 0$ and suppose for a contradiction that $(x_n)_{n\in\mathbb{N}}$ is an unbounded sequence in $\{x' \in X \mid \mu(B_r(x')) \geqslant t\}$. By passing to a subsequence if necessary, we may assume that $d(x_n, x_{n'}) > 2r$ for all distinct $n, n' \in \mathbb{N}$. We thus obtain the contradiction that

$$1 = \mu(X) \geqslant \mu\left(\biguplus_{n\in\mathbb{N}} B_r(x_n)\right) = \sum_{n\in\mathbb{N}} \mu(B_r(x_n)) \geqslant \sum_{n\in\mathbb{N}} t = \infty.$$

This shows that $\{x' \in X \mid \mu(B_r(x')) \geqslant t\}$ must be bounded and also that it admits no infinite subset with separation $2r$, thus establishing (c), of which (d) is a special case. ■

4.2. Existence and absence of greatest elements. Our first aim is to establish existence of greatest elements for $\preccurlyeq_r$, which we also call *radius-r modes*. Such points can be seen as approximate modes[5] with respect to the positive radius/spatial resolution r; only in the next section will we attempt to take the limit as $r \searrow 0$.

Lemma 4.3 now gives several equivalent conditions for a point to be a radius-r mode. The intersection criterion (d) will prove especially helpful in what follows, in the sense that we establish existence of radius-r modes by showing that intersections of this type are nonempty.

Lemma 4.3 (characterization of radius-r modes). *Let X be any metric space, let $\mu \in \mathscr{P}(X)$, and let $r > 0$. As in (3.4), let $M_r := \sup_{x\in X} \mu(B_r(x))$. Then the following are equivalent and if one (and hence any) holds, then $x_r^\star \in X$ is called a* radius-r mode:

(a) *$x_r^\star$ is a $\preccurlyeq_r$-maximal element;*
(b) *$x_r^\star$ is a $\preccurlyeq_r$-greatest element;*
(c) *$x_r^\star \in \bigcap_{x\in X} \uparrow_r x$;*
(d) *$x_r^\star \in \bigcap_{n\in\mathbb{N}} \uparrow_r x_n$ for some sequence $(x_n)_{n\in\mathbb{N}} \subseteq X$ with $\mu(B_r(x_n)) \nearrow M_r$ as $n \to \infty$;*
(e) *$\mu(B_r(x_r^\star)) = M_r$.*

Proof. ((a) $\iff$ (b)) This equivalence holds because $\preccurlyeq_r$ is a total preorder.

((b) $\iff$ (c)) This equivalence is simply a restatement of the definition of being greatest.

((c) $\implies$ (d)) This implication is obvious, since $\bigcap_{n\in\mathbb{N}} \uparrow_r x_n \supseteq \bigcap_{x\in X} \uparrow_r x$.

((d) $\implies$ (e)) Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in X with $\mu(B_r(x_n)) \nearrow M_r$ and let $x_r^\star \in \bigcap_{n\in\mathbb{N}} \uparrow_r x_n$. Then, for each n, $\mu(B_r(x_r^\star)) \geqslant \mu(B_r(x_n))$, and taking the limit as $n \to \infty$ shows that $\mu(B_r(x_r^\star)) \geqslant M_r$. The definition of M_r implies that $\mu(B_r(x_r^\star)) \leqslant M_r$, and so $\mu(B_r(x_r^\star)) = M_r$.

((e) $\implies$ (b)) Suppose that $x_r^\star$ has $\mu(B_r(x_r^\star)) = M_r$. Then, for any $x \in X$, $\mu(B_r(x)) \leqslant M_r$, i.e., $x \preccurlyeq_r x_r^\star$. Thus, $x_r^\star$ is $\preccurlyeq_r$-greatest. ■

A very simple existence result for radius-r modes is the following.

Proposition 4.4 (existence of radius-r modes in compact spaces). *Let X be a compact metric space, let $\mu \in \mathscr{P}(X)$, and let $r > 0$. Then $\preccurlyeq_r$ has at least one radius-r mode $x_r^\star \in X$.*

Proof. This is a special case of Theorem 4.6(a), and it also follows from [2, Theorem 2.44], but a self-contained proof is given by observing that the map $\mu(B_r(\,\cdot\,))\colon X \to [0,1]$ is upper

[5]The intuition that radius-r modes are approximate modes must be treated skeptically. For example, consider $\mu \in \mathscr{P}(X)$ with bimodal continuous Lebesgue density $\rho(x) \propto \max\{0, 1-4(x-1)^2\} + \max\{0, 1-4(x+1)^2\}$, for which a radius-1 mode is located at 0, which is neither a maximizer of ρ nor even in $\operatorname{supp}(\mu)$.

semicontinuous (Lemma SM1.1(a)) and hence has at least one global maximizer $x_r^\star$ in the compact space X. ∎

We now adopt a very different approach to establishing the existence of radius-r modes, one based on applying intersection theorems to upward closures with respect to $\preccurlyeq_r$. We begin with a very general lemma; when Lemma 4.5 is used in practice, $\mathcal{T}$ will often be the metric topology, but another useful case is the weak topology of a Banach space.

Lemma 4.5. *Let X be a separable metric space, $\mu \in \mathscr{P}(X)$, and $r > 0$. Suppose that $\mathcal{T}$ is a topology on X such that, for some sequence $(x_n)_{n\in\mathbb{N}} \subset X$ with $\mu(B_r(x_n)) \nearrow M_r > 0$, $\uparrow_r x_n$ is $\mathcal{T}$-closed and $\mathcal{T}$-compact for all sufficiently large n. Then the set $\mathfrak{M}_r$ of radius-r modes for μ is nonempty, $\mathcal{T}$-compact, and $\mathfrak{M}_r = \bigcap_{n\in\mathbb{N}} \uparrow_r x_n$.*

Proof. Separability of X implies that $M_r > 0$. Let $(x_n)_{n\in\mathbb{N}}$ be such that $\mu(B_r(x_n)) \nearrow M_r$ as $n \to \infty$. The sets $\uparrow_r x_n$ are nonempty; since the sequence $(\mu(B_r(x_n)))_{n\in\mathbb{N}}$ is increasing, $\uparrow_r x_{n+1} \subseteq \uparrow_r x_n$ for each n, i.e., they are decreasingly nested; by hypothesis, for sufficiently large n, they are also $\mathcal{T}$-closed and $\mathcal{T}$-compact. Therefore, by Cantor's intersection theorem (Theorem 2.1(a)), their intersection is nonempty and $\mathcal{T}$-compact. This intersection is precisely the set $\mathfrak{M}_r$ of radius-r modes, as already shown by Lemma 4.3. ∎

Theorem 4.6 (existence of radius-r modes). *Let X be a separable metric space, $\mu \in \mathscr{P}(X)$, and $r > 0$. Let $\mathfrak{M}_r$ denote the set of radius-r modes for μ.*

(a) *Suppose that X has the Heine–Borel property, i.e., that every closed and bounded subset of X is compact. Then $\mathfrak{M}_r$ is nonempty and compact.*
(b) *Suppose that X is complete and that μ is a doubling measure, i.e., there exists a constant $C > 0$ such that $\mu(B_{2r}(x)) \leqslant C\mu(B_r(x))$ for all $x \in X$ and $r > 0$. Then $\mathfrak{M}_r$ is nonempty and compact.*
(c) *Suppose that X is complete and there exists a point $o \in X$ and a function $f\colon (0,\infty)^2 \to (0,\infty)$ such that*

$$\text{for all } x \in B_R(o), \quad \mu(B_\delta(x)) \geqslant f(\delta, R) > 0. \tag{4.2}$$

Then $\mathfrak{M}_r$ is nonempty and compact.
(d) *Suppose that X is complete and that there exists $(x_n)_{n\in\mathbb{N}}$ with $\mu(B_r(x_n)) \nearrow M_r$ and $\gamma(\uparrow_r x_n) \to 0$ as $n \to \infty$. Then $\mathfrak{M}_r$ is nonempty and compact.*
(e) *Suppose that X is complete and that there exists $(x_n)_{n\in\mathbb{N}}$ with $\mu(B_r(x_n)) \nearrow M_r$ and $\operatorname{diam}(\uparrow_r x_n) \to 0$ as $n \to \infty$. Then $\mathfrak{M}_r$ is a singleton.*
(f) *Suppose that X is a Banach space and that there exists $(x_n)_{n\in\mathbb{N}}$ with $\mu(B_r(x_n)) \nearrow M_r$ and that $\uparrow_r x_n$ is weakly compact for all sufficiently large n. Then $\mathfrak{M}_r$ is nonempty and weakly compact.*
(g) *Suppose that X is a reflexive Banach space and that there exists $(x_n)_{n\in\mathbb{N}}$ such that $\mu(B_r(x_n)) \nearrow M_r$ and $\uparrow_r x_n$ is convex for all sufficiently large n. Then $\mathfrak{M}_r$ is nonempty, weakly compact, and convex.*

Proof.
(a) Lemma 4.2 ensures that every upward closure $\uparrow_r x$, $x \in X$, is closed and bounded in the metric topology on X. The claim now follows from Lemma 4.5.

(b) By [5, Proposition 3.1], any complete metric space with a doubling measure has the Heine–Borel property. The claim now follows from (a).

(c) Let $R>0$ and $\delta>0$ be arbitrary. The lower bound (4.2) implies that there cannot be an infinite set of pairwise-disjoint balls $B_\delta(x_n)$, $n\in\mathbb{N}$, with centers $x_n\in B_R(o)$ since, if there were, then we would obtain the contradiction

$$1\geqslant\mu(B_{R+\delta}(o))\geqslant\mu\left(\biguplus_{n\in\mathbb{N}}B_\delta(x_n)\right)=\sum_{n\in\mathbb{N}}\mu(B_\delta(x_n))\geqslant\sum_{n\in\mathbb{N}}f(\delta,R)=\infty.$$

Since $\delta>0$ was arbitrary, $\gamma(B_R(o))=0$, i.e., $B_R(o)$ is compact. Now, given any closed and bounded set $A\subseteq X$, choose $R>0$ large enough that $A\subseteq B_R(o)$ to see that A must be compact. Therefore, X has the Heine–Borel property. The claim now follows from (a).

(d) The claim follows from Kuratowski's intersection theorem (Theorem 2.1(c)).

(e) As already observed, by Lemma 4.2, each upward closure $\uparrow_r x_n$ is both closed and bounded in the metric topology and they are decreasingly nested. Since X is complete, Cantor's intersection theorem (Theorem 2.1(b)) yields that $\mathfrak{M}_r=\bigcap_{n\in\mathbb{N}}\uparrow_r x_n=\{x_r^\star\}$ for some $x_r^\star\in X$.

(f) This is simply Lemma 4.5 in the special case that $\mathcal{T}$ is the weak topology of the separable Banach space X.

(g) Each closed, bounded, and convex subset of the separable, reflexive Banach space X is necessarily weakly compact, and so the claim follows from (f). ■

Theorem 4.6 is by no means universally applicable, and indeed there are measures that have no radius-r modes, as the next two examples show.

Example 4.7 (an atomic measure with no radius-r mode for $1\leqslant r<2$). Let $X=\mathbb{N}$ be equipped with the following variant of the discrete metric:

$$\Delta(k,\ell):=\begin{cases}0 & \text{if } k=\ell,\\ 2 & \text{if } \min\{k,\ell\} \text{ is odd and } \max\{k,\ell\}=\min\{k,\ell\}+1,\\ 1 & \text{otherwise.}\end{cases} \tag{4.3}$$

In the space (X,Δ), distinct points are a unit distance apart, with the exception of each odd number and its successor, which are doubly spaced. Equip this space with the measure $\mu:=\sum_{k\in\mathbb{N}}2^{-k}\delta_k\in\mathscr{P}(X)$, where δ_k is the unit Dirac measure centered at k. For arbitrary $k\in\mathbb{N}$,

$$\mu(B_1(2k-1))=\mu(X\setminus\{2k\})=1-2^{-2k}, \tag{4.4}$$

$$\mu(B_1(2k))=\mu(X\setminus\{2k-1\})=1-2^{-(2k-1)}. \tag{4.5}$$

Both (4.4) and (4.5) show that $M_1=1$; (4.4) shows that no odd number is a radius-1 mode; (4.5) shows that no even number is a radius-1 mode. Thus, μ has no radius-1 mode at all.

Similar arguments also show that μ has no radius-r mode for $1\leqslant r<2$; for $r\geqslant 2$, every point of X is a radius-r mode; for $0<r<1$, the point $1\in X$ is the unique radius-r mode.

It is interesting to relate this example to Theorem 4.6. In this setting, for each $k \in X$, $\uparrow_1 k \supseteq \{k, k+2, k+4, \dots\}$. This set is noncompact with $\gamma(\uparrow_1 k) \geqslant 1$, since it contains an infinite 1-separated sequence. Thus, neither Theorem 4.6(a) nor (d) can apply. Also, although the space (X, Δ) is complete,[6] Theorem 4.6(e) does not apply because $\operatorname{diam}(\uparrow_1 k) \geqslant 1$.

Example 4.8 (a nonatomic measure with no radius-r mode for any $0 < r < 1/8$). Building on the ideas of Example 4.7, consider the space

$$X := \left\{ (\xi, k, m) \in \mathbb{R} \times \mathbb{N}^2 \,\middle|\, |\xi| \leqslant 2^{-k-m-1} \right\}, \tag{4.6}$$

equipped with the metric d and probability measure μ given by

$$d((\xi, k, m), (\eta, \ell, n)) := \begin{cases} 2 & \text{if } m \neq n, \\ 2^{-m} \Delta(k, \ell) & \text{if } m = n \text{ and } k \neq \ell, \\ |\xi - \eta| & \text{if } m = n \text{ and } k = \ell, \end{cases} \tag{4.7}$$

$$\mu\left(\biguplus_{k,m \in \mathbb{N}} (E_{k,m} \times \{k\} \times \{m\}) \right) := \frac{1}{Z} \sum_{k,m \in \mathbb{N}} \sigma^{-m} \lambda^1 (E_{k,m} \cap [-2^{-k-m-1}, 2^{-k-m-1}]), \tag{4.8}$$

for $E_{k,m} \in \mathscr{B}(\mathbb{R})$, where λ^1 is one-dimensional Lebesgue measure, $1/2 < \sigma < 1$ is a scaling parameter, and the normalization constant is $Z := \sum_{m \in \mathbb{N}} (2\sigma)^{-m} \in (1, \infty)$.

Now let $0 < r < 1/8$ be arbitrary and let $n \in \mathbb{N}$ be uniquely determined by $2^{-n} \leqslant r < 2^{-n+1}$. We now determine M_r and whether or not it can be attained by the masses of balls $B_r(x)$, where $x = (\xi, k, m)$ has $m = n$, $m < n$, or $m > n$, respectively. Note that, since $r < 2$, the first case of (4.7) implies that $B_r(x) \subseteq \mathbb{R} \times \mathbb{N} \times \{m\}$.

(i) First suppose that $m = n$. For odd $k \in \mathbb{N}$,

$$\mu(B_r(x)) = \mu(\mathbb{R} \times (\mathbb{N} \setminus \{k+1\}) \times \{m\}) = \frac{(2\sigma)^{-m}}{Z} (1 - 2^{-(k+1)}).$$

Taking the limit as $k \to \infty$ shows that $M_r \geqslant \frac{(2\sigma)^{-n}}{Z}$ but that no such ball realizes this supremal mass. The case of even k is similar, just as in Example 4.7.

(ii) For $m > n$, since $B_r(x) \subseteq \mathbb{R} \times \mathbb{N} \times \{m\}$, it follows that x is not a radius-r mode because

$$\mu(B_r(x)) \leqslant \frac{(2\sigma)^{-m}}{Z} < \frac{(2\sigma)^{-n}}{Z} \leqslant M_r.$$

(iii) If $m < n$, then $r < 2^{-n} < 2^{-m}$, and so the second case of (4.7) ensures that $B_r(x) \subseteq \mathbb{R} \times \{k\} \times \{m\}$. The mass of such a ball is maximized by the case $\xi = 0$, $k = 1$, $m = n - 1$, in which case the ball (which is a single line segment) has mass

[6] Just as in the case of the discrete metric, in this space, the properties of being a Cauchy sequence, being a convergent sequence, and being eventually constant all coincide.

$$\begin{aligned}
\mu(B_r(x)) &= \frac{\sigma^{-m}}{Z}\lambda^1([-r,r]\cap[-2^{-k-m-1},2^{-k-m-1}]) \\
&\leqslant \sigma\frac{\sigma^{-n}}{Z}\lambda^1([-2^{-n},2^{-n}]\cap[-2^{-n-1},2^{-n-1}]) \\
&= \sigma\frac{(2\sigma)^{-n}}{Z} < \frac{(2\sigma)^{-n}}{Z},
\end{aligned}$$

where the last inequality follows from the fact that $\sigma < 1$.
Hence, $M_r = \frac{(2\sigma)^{-n}}{Z}$ but $\mu(B_r(x)) < M_r$ for all $x \in X$, i.e., μ has no radius-r mode.

Thus, while Theorem 4.6 on the existence of radius-r modes covers a variety of well-behaved spaces and measures, Examples 4.7 and 4.8 show that existence cannot be guaranteed for general spaces and measures.

Before moving on, we mention one interesting intermediate case, motivated by applications to inverse problems, namely countable product measures on weighted ℓ^p spaces (and isometric linear images of such spaces). This is a broad class that includes Gaussian, Besov [8, 22], and Cauchy measures [28]. It turns out that reweightings of such measures always have radius-r modes, i.e., Bayesian posteriors with such measures as priors always have radius-r MAP estimators. We defer the precise statements to Theorem SM1.8 and Corollary SM1.9 in subsection SM1.2 because they do not have a particularly order-theoretic flavor.

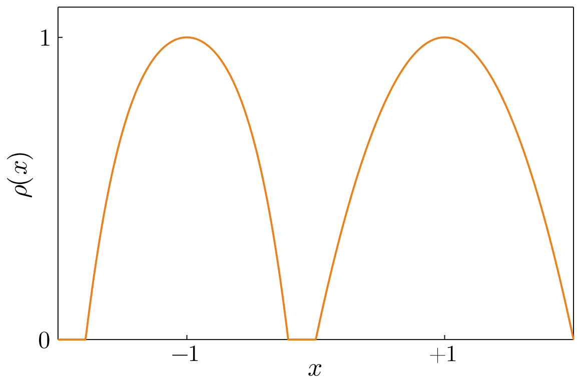
Also, while we do prove the existence of radius-r modes, we do not consider taking limits as $r \to 0$ to obtain true MAP estimators for such posteriors. The main difficulty here lies in proving that such a family $(x_r^\star)_{r>0}$ is bounded, so that a weakly convergent subsequence can be extracted; this is not true in general, so one must argue using properties of the prior and likelihood (e.g., when the prior is Gaussian or Besov).

Indeed, the whole question of taking limits of radius-r modes is a sensitive one and is the topic of the next section.

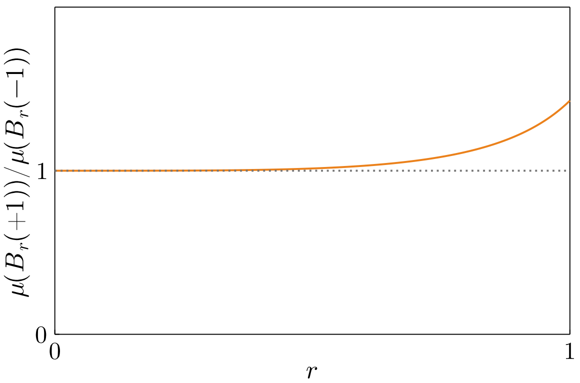
4.3. Convergence of greatest and near-greatest elements. If radius-r modes $x_r^\star$ do exist for each $r > 0$, it is then natural to ask whether sequences of radius-r modes can approximate true modes, e.g., strong or weak modes. This approach is used by [9, Theorem 3.5] to obtain strong modes for Bayesian posteriors arising from Gaussian priors.

However, we have seen that existence of radius-r modes can be difficult to prove, and in some cases no radius-r modes exist (Examples 4.7 and 4.8). Taking limits of radius-r modes is also problematic for more general measures: the limit need not be a strong or weak mode, and not every mode can be represented as the limit of radius-r modes. Thus, one cannot hope to use the approach of taking limits of radius-r modes to find true modes if there is no correspondence between modes and limits of radius-r modes. Nevertheless, we show that some of the difficulties can be overcome using *asymptotic maximizing families* (AMFs) as proposed by [18].

To illustrate the problem and motivate the introduction of AMFs, we first give an example of a measure with a bounded and continuous Lebesgue density possessing a mode that cannot be represented as the limit of radius-r modes. The problem here is that balls around the points ± 1 have asymptotically equivalent mass, but each ball around $+1$ has slightly more mass than the corresponding ball around -1; as a result, $+1$ "hides" the other mode -1.



(a) Unnormalised density (4.9) of the measure in Example 4.9. The point $+1$ is the unique radius-r mode for all sufficiently small r.



(b) The ratio $\mu(B_r(+1))/\mu(B_r(-1))$ converges to 1 as $r \to 0$, but it is strictly greater than 1 for any $r > 0$.

Figure 4.1. *Not every strong mode is the limit of a sequence of radius-r modes: in Example 4.9, -1 is a strong mode but $+1$ is the unique radius-r mode for all small r.*

Example 4.9. Define $\mu \in \mathscr{P}(\mathbb{R})$ by the Lebesgue density as shown in Figure 4.1, given by

$$\rho(x) \propto \max\{1-(x-1)^2, 0\} + \max\{1-(x+1)^2-(x+1)^4, 0\}. \tag{4.9}$$

When r is sufficiently small, $\mu(B_r(+1)) = 2r - \frac{2}{3}r^3$ and $\mu(B_r(-1)) = 2r - \frac{2}{3}r^3 - \frac{2}{5}r^5$, so there is a unique radius-r mode at $+1$. However, $+1$ and -1 are both strong modes: $+1$ is a radius-r mode for all sufficiently small r, so it is a strong mode (Theorem 4.11) and -1 is a strong mode because

$$\lim_{r\to 0} \frac{\mu(B_r(-1))}{M_r} = \lim_{r\to 0} \frac{\mu(B_r(+1))}{M_r} \lim_{r\to 0} \frac{\mu(B_r(-1))}{\mu(B_r(+1))} = 1.$$

Instead of representing modes as limits of radius-r modes—which might not be possible—one may consider families of points that are nearly greatest, which always exist, and try to take limits of such families.

Definition 4.10. *Let X be a metric space and let $\mu \in \mathscr{P}(X)$. A net $(x_r)_{r>0} \subseteq X$ is an AMF if there exists a positive function ε with $\lim_{r\to 0} \varepsilon(r) = 0$ and*

$$\frac{\mu(B_r(x_r))}{M_r} \geqslant 1 - \varepsilon(r) \quad \textit{for all } r > 0. \tag{4.10}$$

Note that every measure admits an AMF satisfying (4.10), even if the function ε is specified in advance, which is sometimes advantageous.

The following results shed light on the subtleties involved in taking limits of radius-r modes, or, more generally, AMFs.

Theorem 4.11. *Let X be a separable metric space and let $\mu \in \mathscr{P}(X)$.*
(a) *If $(x^\star)_{r>0}$ is an AMF with $x^\star \in X$ fixed, then $x^\star$ is a strong mode.*
(b) *If $x^\star$ is a radius-r mode for all small enough $r > 0$, then $x^\star$ is a strong mode.*
(c) *If the AMF $(x_r^\star)_{r>0}$ converges to $x^\star$, then $x^\star$ is a generalized strong mode.*

Proof.
(a) As $\mu(B_r(x^\star)) \geqslant (1-\varepsilon(r))M_r$, it is immediate that $x^\star$ is a strong mode, because

$$1 \geqslant \lim_{r\to 0} \frac{\mu(B_r(x^\star))}{M_r} \geqslant \lim_{r\to 0}(1-\varepsilon(r)) = 1.$$

(b) This is immediate from (a) as $(x^\star)_{r>0}$ forms an AMF.
(c) This is precisely [7, Lemma 2.4]. ∎

The claim in (c)—which requires that the net $(x_r^\star)_{r>0}$ converges to $x^\star$ along every subsequence—cannot be made stronger without additional hypotheses: the limit $x^\star$ need not be a strong or weak mode (as can be seen in Example 5.4(a), for which $x^\star = 1$ is the limit of an AMF which is neither a strong mode nor a weak mode). Furthermore, one cannot weaken the hypotheses of (c) further: the points ± 1 in Example 5.7 are limit points of an AMF but they are not even generalized modes.

The general question of classifying measures μ for which limits of radius-r modes are strong modes is still open, although [9] show that reweightings of Gaussian measures on Hilbert spaces enjoy this property, and [18] show the same for some Gaussian measures on sequence spaces.

Under an additional nesting assumption, intersection arguments can be applied to AMFs to yield the existence of several kinds of modes.

Theorem 4.12 (AMFs and strong modes). *Let X be a complete and separable metric space and let $\mu \in \mathscr{P}(X)$. Let $(x_r)_{r>0}$ be any AMF, i.e., any net satisfying (4.10), and let $I := \bigcap_{r>0} \uparrow_r x_r$. Then*
(a) *$I \subseteq \operatorname{supp}(\mu)$;*
(b) *every $x^\star \in I$ is a strong (and hence weak and generalized strong) mode for μ;*
(c) *and if also*

$$0 < r \leqslant s \implies \uparrow_r x_r \subseteq \uparrow_s x_s, \tag{4.11}$$

then I is nonempty and compact.

Proof. Let $x^\star \in I$. For all sufficiently small $r > 0$, it follows that $\mu(B_r(x^\star)) \geqslant \mu(B_r(x_r)) \geqslant M_r(1-\varepsilon(r)) > 0$, and so $x^\star \in \operatorname{supp}(\mu)$, which establishes (a). Furthermore, since $x^\star \in \uparrow_r x_r$ for each r,

$$1 \geqslant \frac{\mu(B_r(x^\star))}{M_r} \geqslant \frac{\mu(B_r(x_r))}{M_r} \geqslant 1 - \varepsilon(r).$$

Taking the limit as $r \to 0$ throughout shows that $x^\star$ is a strong mode (and hence also a weak and generalized strong mode) for μ, establishing (b). (Alternatively, one may observe that $(x^\star)_{r>0}$ is a constant AMF and appeal to Theorem 4.11(a).)

For (c), let $(r_n)_{n\in\mathbb{N}}$ be some null sequence of radii. The nesting hypothesis (4.11) implies that $I = \bigcap_{n\in\mathbb{N}} \uparrow_{r_n} x_{r_n}$. For each n, $\uparrow_{r_n} x_{r_n}$ is nonempty and, by Lemma 4.2, is closed and bounded with $\gamma(\uparrow_{r_n} x_{r_n}) \leqslant 2r_n$. This, together with the nesting hypothesis (4.11) and Kuratowski's intersection theorem (Theorem 2.1(c)), ensures that I is nonempty and compact. ∎

Remark 4.13.

(a) The nesting hypothesis (4.11), in conjunction with Lemma 4.2, ensures that the AMF $(x_r)_{r>0}$—and indeed any family of greatest elements $(x_r^\star)_{r>0}$—must be bounded. This means that Theorem 4.12 does not apply to measures such as Example 5.4(b), for which the radius-r modes "escape to infinity" as $r \to 0$. Hypothesis (4.11) also fails for measures displaying oscillatory behavior of the kind discussed in Example 5.7.

(b) Theorem 4.12 is not sharp, in the sense that there can exist modes $x^\star \notin \bigcap_{r>0} \uparrow_r x_r$. See Example 4.9 for an example of this situation with

$$x_r \equiv +1, \quad \uparrow_r x_r = \{+1\}, \quad x^\star = -1 \notin \bigcap_{r>0} \uparrow_r x_r.$$

5. Preorders in the small-radius limit.

One would like to think of $x^\star \in X$ as a mode of $\mu \in \mathscr{P}(X)$ if $x^\star$ is a greatest or maximal element of X with respect to the preorder $\preccurlyeq_r$ "in the limit as $r \to 0$" in some sense. However, is such a limiting preorder well defined? Must this preorder have greatest or maximal elements?

In fact, there are several candidates for a small-radius limiting preorder and it appears that each of them has at least one undesirable feature. This work will focus on the analytic small-radius limiting preorder $\preccurlyeq_0$, to be defined shortly (Definition 5.1). This preorder has the advantage that its greatest elements are weak modes; however, it has the disadvantage that it is not total, i.e., the existence of greatest elements is not guaranteed, and indeed the collection of incomparable elements may be rather large. We claim that this is a small price to pay: we show in section SM2 that the alternative definitions are even more ill behaved.

5.1. Definition and basic properties.

Definition 5.1 (small-radius limiting preorder). *Let X be a metric space and let $\mu \in \mathscr{P}(X)$. Define a preorder $\preccurlyeq_0$ on X by*

$$x \preccurlyeq_0 x' \iff \limsup_{r\to 0} \frac{\mu(B_r(x))}{\mu(B_r(x'))} \leqslant 1 \iff \liminf_{r\to 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} \geqslant 1 \tag{5.1}$$

if both $x, x' \in \operatorname{supp}(\mu)$. Additionally, as exceptional cases, $x \preccurlyeq_0 x'$ is defined to be false for $x \in \operatorname{supp}(\mu)$ and $x' \notin \operatorname{supp}(\mu)$, and $x \preccurlyeq_0 x'$ is defined to be true for $x \notin \operatorname{supp}(\mu)$ and $x' \in X$.

It is relatively straightforward to verify that $\preccurlyeq_0$, as defined above, is a preorder on X; the only subtleties are correct handling of points outside the support and the use of the upper bound (but not equality)

$$\limsup_{r\to 0} \frac{\mu(B_r(x))}{\mu(B_r(y))} \frac{\mu(B_r(y))}{\mu(B_r(z))} \leqslant \limsup_{r\to 0} \frac{\mu(B_r(x))}{\mu(B_r(y))} \limsup_{r\to 0} \frac{\mu(B_r(y))}{\mu(B_r(z))} \tag{5.2}$$

when verifying transitivity. As usual, we will write $x \asymp_0 x'$ if both $x \preccurlyeq_0 x'$ and $x \succcurlyeq_0 x'$ hold, and $x \parallel_0 x'$ if neither $x \preccurlyeq_0 x'$ nor $x \succcurlyeq_0 x'$ holds.

The appeal of the preorder $\preccurlyeq_0$ is that its greatest elements are exactly the weak modes of μ, as defined in (3.7), as the next two results show.

Lemma 5.2 (properties of $\preccurlyeq_0$-maximal elements). *Let X be separable and let $\mu \in \mathscr{P}(X)$.*

(a) *If $x^\star$ is $\preccurlyeq_0$-maximal, then $x^\star \in \operatorname{supp}(\mu)$.*

(b) *The point* $x^\star \in \operatorname{supp}(\mu)$ *is* $\preccurlyeq_0$*-maximal if and only if any* $x \in X$ *satisfies either*

$$\liminf_{r\to 0} \frac{\mu(B_r(x))}{\mu(B_r(x^\star))} < 1 \text{ or } \lim_{r\to 0} \frac{\mu(B_r(x))}{\mu(B_r(x^\star))} = 1. \tag{5.3}$$

Proof.

(a) Suppose that $x^\star$ is maximal but, for a contradiction, suppose also that $x^\star \notin \operatorname{supp}(\mu)$. As X is separable, take $x \in \operatorname{supp}(\mu) \neq \varnothing$. By the exceptional cases of Definition 5.1, $x^\star \prec_0 x$, contradicting the assumption that $x^\star$ is maximal.

(b) Let $x, x' \in \operatorname{supp}(\mu)$. It is straightforward to verify from the definitions that

$$x \prec_0 x' \iff \liminf_{r\to 0} \frac{\mu(B_r(x))}{\mu(B_r(x'))} < 1, \tag{5.4}$$

$$x \asymp_0 x' \iff \lim_{r\to 0} \frac{\mu(B_r(x))}{\mu(B_r(x'))} = 1. \tag{5.5}$$

Suppose first that $x^\star$ is maximal, and let $x \in X$ be arbitrary. If $x \notin \operatorname{supp}(\mu)$, then $\mu(B_r(x)) = 0$ for all sufficiently small r, so (5.3) holds. If $x \in \operatorname{supp}(\mu)$, then maximality of $x^\star$ implies that either $x \prec_0 x^\star$ or $x \asymp_0 x^\star$, from which (5.3) follows.

Conversely, suppose that $x \in X$ satisfies $x^\star \preccurlyeq_0 x$. The exceptional cases in Definition 5.1 imply that $x \in \operatorname{supp}(\mu)$. Hence, by (5.5), $x^\star \asymp_0 x$, proving that $x^\star$ is $\preccurlyeq_0$-maximal. ∎

Lemma 5.3 (characterization of weak modes). *Let* X *be separable and let* $\mu \in \mathscr{P}(X)$*. Then the following are equivalent:*

(a) $x^\star \in X$ *is a weak mode for* μ*;*

(b) $x^\star \in X$ *is a* $\preccurlyeq_0$*-greatest element;*

(c) $x^\star \in X$ *is a* $\preccurlyeq_0$*-maximal element that is comparable with every other* $x' \in X$*.*

Proof. ((a) $\Longrightarrow$ (b)) Suppose that $x^\star$ is a weak mode for μ. Then, by definition (see (3.7)), it follows that $x \preccurlyeq_0 x^\star$ for each $x \in \operatorname{supp}(\mu)$. As $x^\star \in \operatorname{supp}(\mu)$, any point $x \notin \operatorname{supp}(\mu)$ satisfies $x \preccurlyeq_0 x^\star$ by the special cases in the definition of $\preccurlyeq_0$.

((b) $\Longrightarrow$ (a)) Suppose that $x^\star$ is a $\preccurlyeq_0$-greatest element. Then $x^\star \in \operatorname{supp}(\mu)$ by Lemma 5.2. Hence, for $x' \in \operatorname{supp}(\mu)$, (3.7) holds because $x' \preccurlyeq_0 x^\star$. For $x' \notin \operatorname{supp}(\mu)$, we obtain

$$\frac{\mu(B_r(x'))}{\mu(B_r(x^\star))} = 0 \text{ for sufficiently small } r,$$

proving that $x^\star$ is a weak mode.

((b) $\iff$ (c)) This is obvious, since the defining property of being greatest is exactly the property of being maximal and globally comparable. ∎

The preorder $\preccurlyeq_0$ does have some shortcomings. One is that, in contrast to $\preccurlyeq_r$ with $r > 0$ (Lemma 4.2), upward closures under $\preccurlyeq_0$ need be neither closed nor bounded.

Example 5.4.

(a) For an example of a nonclosed upward closure under $\preccurlyeq_0$, similar in spirit to the example of [7] mentioned in section 3, let $\mu \in \mathscr{P}(\mathbb{R})$ have the Lebesgue density $\rho \colon \mathbb{R} \to \mathbb{R}$,

$\rho(x) := 2x\mathbb{1}[0 \leqslant x \leqslant 1]$, with $\operatorname{supp}(\mu) = [0,1]$, and consider $y \in \mathbb{R}$. If $y < 0$ or $y > 1$, then $y \notin \operatorname{supp}(\mu)$ and $\uparrow_0 y = \mathbb{R}$. For $0 \leqslant y \leqslant 1/2$, $\uparrow_0 y = [y,1]$, which is closed. However, for $1/2 < y < 1$,

$$\lim_{r\to 0} \frac{\mu(B_r(1))}{\mu(B_r(y))} = \lim_{r\to 0} \frac{2r - r^2}{4yr} = \frac{1}{2y} < 1$$

and so $\uparrow_0 y = [y, 1)$, which is not closed. Finally, $\uparrow_0 1 = [1/2, 1]$, which is closed.

(b) For an example of an unbounded upward closure under $\preccurlyeq_0$, let $\mu \in \mathscr{P}(\mathbb{R})$ have the unbounded Lebesgue density $\rho\colon \mathbb{R} \to \mathbb{R}$,

$$\rho(x) := \sum_{n\in\mathbb{N}} n\mathbb{1}[n - \tfrac{2^{-n-1}}{n} \leqslant x \leqslant n + \tfrac{2^{-n-1}}{n}].$$

That is, ρ consists of a sum of disjoint indicator functions centred on the natural numbers $n \in \mathbb{N}$, each having mass 2^{-n} and height n. Then, for any $x, y \in \mathbb{N}$ with $x > y$,

$$\lim_{r\to 0} \frac{\mu(B_r(x))}{\mu(B_r(y))} = \lim_{r\to 0} \frac{2xr}{2yr} = \frac{x}{y} > 1$$

and so $\uparrow_0 y \supseteq \mathbb{N} \cap [y, \infty)$.

Example 5.4 furnishes two examples of measures with no $\preccurlyeq_0$-maximal element, let alone a $\preccurlyeq_0$-greatest element (weak mode), or strong mode. However, these examples are relatively tame: there is no mode simply because any candidate mode $x^\star$ is dominated by some other point x'. The *real* shortcoming and subtlety of $\preccurlyeq_0$ is that it is not total—that is, the order admits incomparable elements—and we make this the topic of the next subsection.

5.2. Criteria for incomparability and comparability. For $r > 0$, totality of $\preccurlyeq_r$ followed immediately from Definition 4.1. This is certainly not so obvious for $\preccurlyeq_0$. Indeed, what is immediate from Definition 5.1 is that $\preccurlyeq_0$-incomparable elements can be characterized as follows.

Lemma 5.5 (incomparability in the limiting preorder). *For $x, x' \in X$,*

$$x \parallel_0 x' \iff x, x' \in \operatorname{supp}(\mu) \text{ and } \liminf_{r\to 0} \frac{\mu(B_r(x))}{\mu(B_r(x'))} < 1 < \limsup_{r\to 0} \frac{\mu(B_r(x))}{\mu(B_r(x'))}. \tag{5.6}$$

In the other direction, we can give a (very strong) sufficient condition for two points to be comparable under $\preccurlyeq_0$.

Lemma 5.6. *Let X be any metric space and let $\mu \in \mathscr{P}(X)$. Suppose that on some interval $(0, r^\star)$, the function $r \mapsto \mu(B_r(x))/\mu(B_r(x'))$ is uniformly continuous for x, $x' \in \operatorname{supp}(\mu)$. Then x and x' are $\preccurlyeq_0$-comparable.*

Proof. The ratio function $\mu(B_r(x))/\mu(B_r(x'))$ can be uniquely extended to a uniformly continuous function on $[0, r^\star]$ [2, Lemma 3.11]. By continuity, the limit of the ratio function as $r \to 0$ must exist; the result follows by Lemma 5.5. ∎

The previous two lemmas hint at a way to construct concrete examples of measures with incomparable points under $\preccurlyeq_0$: one must choose the masses around two points such that the ratio of the masses of balls around such points oscillates as $r \to 0$.

We now construct such a measure on $\mathbb{R}$ with a Lebesgue density and two $\preccurlyeq_0$-incomparable maximal points, neither of which is $\preccurlyeq_0$-greatest. (Subsection 5.3 will supply even more extreme and general examples, but it is pedagogically useful to consider a simpler construction first.) The idea is to construct a density so that the measure $\mu \in \mathscr{P}(\mathbb{R})$ induced by it has a specific behavior around the points $x = \pm 1$. In this case, the density is chosen so that $r \mapsto \mu(B_r(x))$ piecewise linearly interpolates the function $r \mapsto \sqrt{r}$ through either the interpolation knots $r = a^{-n}$ with $n \in \mathbb{N}$ even or the interpolation knots $r = a^{-n}$ with $n \in \mathbb{N}$ odd, where $a > 1$ is chosen arbitrarily. It turns out that these mild perturbations of the integrable singularity $\rho(x) \propto |x|^{-1/2}$ produce "incomparable modes."

Example 5.7 (an absolutely continuous measure on $\mathbb{R}$ with incomparable maximal points and neither weak nor generalized modes, after an example of I. Klebanov). Let X be any Borel-measurable subset of $\mathbb{R}$ containing $[-2, 2]$. Fix $a > 1$ and, as illustrated in Figure 5.1, define $\mu^{\mathrm{e}}, \mu^{\mathrm{o}} \in \mathscr{P}(X)$ via their Lebesgue densities $\rho^{\mathrm{e}}, \rho^{\mathrm{o}} \colon X \to [0, \infty]$,

$$\rho^{\mathrm{e}}(x) := \begin{cases} 0 & \text{if } |x| > 1, \\ \dfrac{a^{n/2}(1 - a^{-1})}{2(1 - a^{-2})} & \text{if } a^{-2-n} \leqslant |x| \leqslant a^{-n} \text{ for even } n \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}, \\ \infty & \text{if } x = 0, \end{cases}$$

and

$$\rho^{\mathrm{o}}(x) := \begin{cases} 0 & \text{if } |x| > 1, \\ \dfrac{1 - a^{-1/2}}{2(1 - a^{-1})} & \text{if } a^{-1} \leqslant |x| \leqslant 1, \\ \dfrac{a^{n/2}(1 - a^{-1})}{2(1 - a^{-2})} & \text{if } a^{-2-n} \leqslant |x| \leqslant a^{-n} \text{ for odd } n \in \mathbb{N}, \\ \infty & \text{if } x = 0, \end{cases}$$

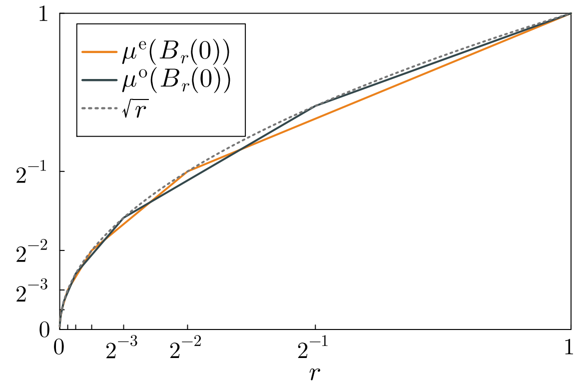
so that the RCDFs are

$$\mu^{\mathrm{e}}(B_r(0)) = \begin{cases} 1 & \text{if } r \geqslant 1, \\ a^{-1-n/2} + (r - a^{-2-n}) \dfrac{a^{n/2}(1 - a^{-1})}{1 - a^{-2}} & \text{if } a^{-2-n} \leqslant r \leqslant a^{-n} \text{ for even } n \in \mathbb{N}_0, \\ 0 & \text{if } r = 0, \end{cases}$$

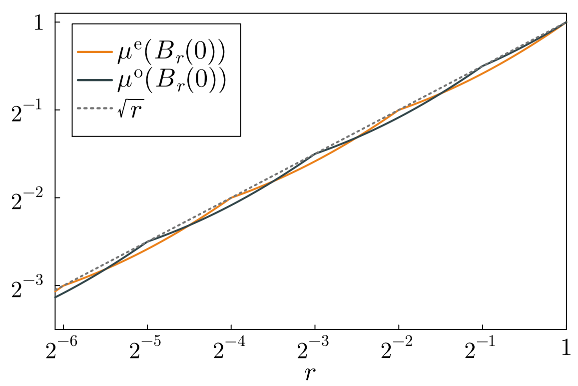
and

$$\mu^{\mathrm{o}}(B_r(0)) = \begin{cases} 1 & \text{if } r \geqslant 1, \\ a^{-1/2} + (r - a^{-1}) \dfrac{1 - a^{-1/2}}{1 - a^{-1}} & \text{if } a^{-1} \leqslant r \leqslant 1, \\ a^{-1-n/2} + (r - a^{-2-n}) \dfrac{a^{n/2}(1 - a^{-1})}{1 - a^{-2}} & \text{if } a^{-2-n} \leqslant r \leqslant a^{-n} \text{ for odd } n \in \mathbb{N}, \\ 0 & \text{if } r = 0. \end{cases}$$

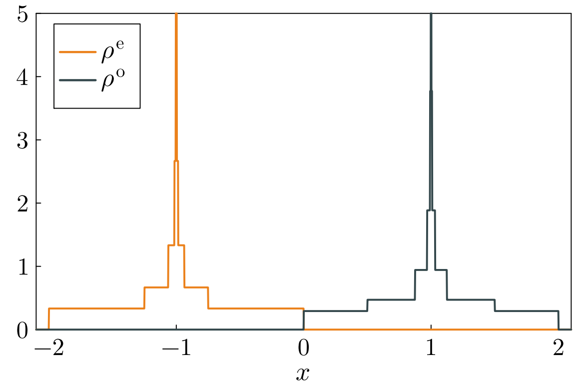
We now consider the probability measure $\mu := \frac{1}{2}\mu^{\mathrm{e}}(\,\cdot + 1) + \frac{1}{2}\mu^{\mathrm{o}}(\,\cdot - 1) \in \mathscr{P}(X)$ with Lebesgue density $\rho := \frac{1}{2}\rho^{\mathrm{e}}(\,\cdot + 1) + \frac{1}{2}\rho^{\mathrm{o}}(\,\cdot - 1)$.



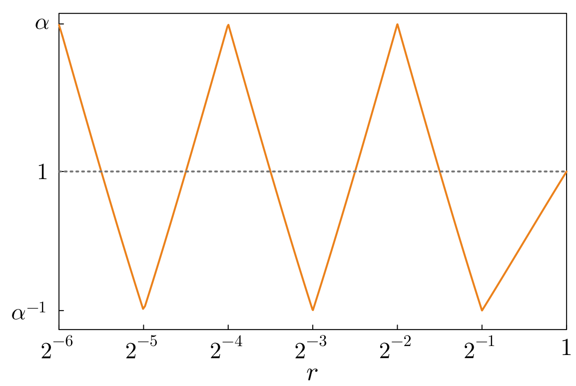
(a) The RCDFs $\mu^{\mathrm{e}}(B_r(0))$ and $\mu^{\mathrm{o}}(B_r(0))$ (shown on a linear scale) interpolate between the knots a^{-n} to give mild perturbations of the function $r \mapsto \sqrt{r}$.



(b) The RCDFs $\mu^{\mathrm{e}}(B_r(0))$ and $\mu^{\mathrm{o}}(B_r(0))$ (shown with a logarithmic scale for r) agree with the function $r \mapsto \sqrt{r}$ at the knots a^{-n} for even n and odd n respectively.



(c) The probability density functions $\rho^{\mathrm{e}}(\cdot + 1)$ and $\rho^{\mathrm{o}}(\cdot - 1)$ have singularities which behave like $|\cdot|^{-1/2}$ at -1 and $+1$ respectively.



(d) The ratio ${}^{\mu(B_r(-1))}/_{\mu(B_r(+1))}$ (shown with a logarithmic scale for r) oscillates between α and α^{-1} as $r \to 0$, so the lim inf of the ratio is below 1 and the lim sup is above 1.

Figure 5.1. *Illustration of the measures defined in Example 5.7 for the parameter choice $a = 2$.*

We first observe that $\pm 1 \succcurlyeq_0 x$ for any $x \neq \pm 1$. For sufficiently small $r > 0$, both ρ^{e} and ρ^{o} are bounded above by a constant on $B_r(x) = [x - r, x + r]$, so that $\mu(B_r(x)) \leqslant cr$ for some $c \geqslant 0$. On the other hand, by construction, both $\mu(B_r(-1))$ and $\mu(B_r(+1))$ are asymptotically equivalent to ${}^{\sqrt{r}}/_{2}$ as $r \to 0$, from which it follows that $\pm 1 \succcurlyeq_0 x$.

However, -1 and $+1$ are incomparable. Observe that, for $r = a^{-n}$ with $n \in \mathbb{N}$ even,

$$\frac{\mu(B_r(-1))}{\mu(B_r(+1))} = \alpha := \frac{a+1}{2a^{1/2}} > 1,$$

whereas for $r = a^{-n}$ with $n \in \mathbb{N}$ odd, this ratio of ball masses takes the value α^{-1}, and, for all $r > 0$, it lies in the interval $[\alpha^{-1}, \alpha]$, all of which can be verified easily from the interpolation formulae for $\mu^{\mathrm{e}}(B_r(0))$ and $\mu^{\mathrm{o}}(B_r(0))$. Lemma 5.5 now implies that $-1 \parallel_0 +1$, since

$$\alpha^{-1} = \liminf_{r\to 0} \frac{\mu(B_r(-1))}{\mu(B_r(+1))} < 1 < \limsup_{r\to 0} \frac{\mu(B_r(-1))}{\mu(B_r(+1))} = \alpha.$$

Thus, the preorder $\preccurlyeq_0$ induced by μ has two incomparable maximal elements, namely ± 1, and has no greatest elements, and hence μ has no weak modes (Lemma 5.3).

We now check that $+1$ and -1 are not generalized modes. Let $r_n := a^{-2n}$, and suppose that $x_n \to 1$ as $n \to \infty$. Choose N large enough that, for all $n \geqslant N$, $|x_n - 1| < 1/2$ and $r_n < 1/2$. As the density $\rho^{\mathrm{o}}(\,\cdot\, - 1)$ is a symmetric singularity around $+1$, it follows that $\mu(B_{r_n}(x_n)) \leqslant \mu(B_{r_n}(+1))$. As $M_{r_n} = \mu(B_{r_n}(-1))$, we obtain that

$$\liminf_{n\to\infty} \frac{\mu(B_{r_n}(x_n))}{M_{r_n}} \leqslant \liminf_{n\to\infty} \frac{\mu(B_{r_n}(+1))}{M_{r_n}} = \alpha^{-1} < 1.$$

This proves that $+1$ is not a generalized mode; a similar argument with $(r_n)_{n\in\mathbb{N}} = a^{-2n+1}$ proves that -1 is not a generalized mode.

Finally, suppose that $x \neq \pm 1$, and let $(r_n)_{n\in\mathbb{N}}$ be any null sequence. Let $\varepsilon := \min\{|x - 1|, |x+1|\}$. Suppose that $x_n \to x$ as $n \to \infty$. There must exist $N \in \mathbb{N}$ such that, for all $n \geqslant N$, $|x_n - 1| > \varepsilon/2$ and $|x_n + 1| > \varepsilon/2$. The Lebesgue density of μ is bounded on $\mathbb{R} \setminus (B_{\varepsilon/2}(+1) \cup B_{\varepsilon/2}(-1))$ by some constant $C > 0$, so $\mu(B_{r_n}(x_n)) \leqslant C r_n$ for $n \geqslant N$. As $M_{r_n} \in \Theta(r_n^{1/2})$ as $n \to \infty$, it follows that

$$\liminf_{n\to\infty} \frac{\mu(B_{r_n}(x_n))}{M_{r_n}} = 0,$$

so x is not a generalized mode.

Example 5.7 illustrates a difficulty with weak modes, and one whose cause can be traced to incomparability: if the space X is partitioned into disjoint positive-mass sets A and B, existence of modes for μ restricted to (or conditioned upon) A and B individually cannot ensure existence of a mode for μ, since the modes of $\mu|_A$ and $\mu|_B$ may be $\preccurlyeq_0$-incomparable.

Thus, while ± 1 are intuitively modes and have Lebesgue density $+\infty$, the measure μ has no modes in any of the senses defined in section 3. We emphasize that one cannot simply declare all points with Lebesgue density $+\infty$ to be modes, since this would place all singularities of the density on the same footing, which is clearly undesirable if one singularity is genuinely "smaller" than the other in the sense that the RCDFs around these points are, say, $\sqrt{r}$ and $2\sqrt{r}$, and so the smaller one ought not to be considered a mode.

As suggested in the introduction, this example could be interpreted as evidence that maximal—rather than greatest—elements of a preorder are good candidates for modes. Indeed, from the order-theoretic perspective, maximal elements appear to be just as reasonable as greatest elements, and we hope that this encourages further study of whether maximal elements are sufficient for applications.

The extension theorems of Szpilrajn [29] and Arrow (as proved by Hansson [15]) assert that any nontotal preorder $\preccurlyeq$ can be extended to a total preorder $\preccurlyeq'$. Thus, given the nontotality of $\preccurlyeq_0$, one might hope to resolve all these issues by defining a mode of μ to be a $\preccurlyeq_0'$-greatest element. Unfortunately, such a total extended preorder is not uniquely determined and so such a definition of a mode would not be well defined: for the measure μ of Example 5.7, there are total extensions $\preccurlyeq_0'$ of $\preccurlyeq_0$ yielding each of the three situations

$$-1 \preccurlyeq_0' +1 \not\preccurlyeq_0' -1, \quad +1 \preccurlyeq_0' -1 \not\preccurlyeq_0' +1, \quad \text{and } -1 \asymp_0' +1.$$

That is, which (if any) of ± 1 counts as a mode would seem to be a matter of personal choice.

Finally, we note that similar ideas could be used to construct incomparable points that are not $\preccurlyeq_0$-maximal, but such examples have less importance for the theory of modes.

5.3. Absolutely continuous measures with dense antichains. Example 5.7 can be easily extended to construct a measure $\mu \in \mathscr{P}(\mathbb{R})$ with any finite number of mutually incomparable $\preccurlyeq_0$-maximal elements, none of which are greatest elements. Indeed, it is natural to wonder how bad the situation of incomparability can be, and in particular how large an antichain can be. This section's main result, Theorem 5.11, shows that μ may have a topologically dense antichain consisting of maximal elements (and mutually incomparable would-be modes are "nearly everywhere"), even when μ has a Lebesgue density; from the perspective of geometric measure theory, the notable point here is that there is no need to resort to singular measures.

We begin with the following straightforward proposition.

Proposition 5.8. *Let X be a finite or discrete metric space and let $\mu \in \mathscr{P}(X)$. Then $\preccurlyeq_0$ has no incomparable elements.*

Proof. Let $x, x' \in \operatorname{supp}(\mu)$. As X is discrete, the measure μ must be atomic, so

$$\lim_{r \to 0} \frac{\mu(B_r(x))}{\mu(B_r(x'))} = \frac{\mu(\{x\})}{\mu(\{x'\})}.$$

As the limit exists, the ratio does not oscillate on either side of unity as $r \to 0$, so comparability of x and x' follows from Lemma 5.5. ■

Proposition 5.8 shows that any measure on a finite metric space induces a total order $\preccurlyeq_0$. We now show that incomparability can arise even in very simple settings, such as in a countable metric space or on the real line with a continuous, bounded Lebesgue density.

Example 5.9.

(a) Let X be the closure of the set $\{-1 + 2^{-n} \mid n \in \mathbb{N}\} \cup \{1 - 2^{-n} \mid n \in \mathbb{N}\}$ with the Euclidean metric inherited from $\mathbb{R}$. Define the measure $\mu \in \mathscr{P}(X)$ by

$$\mu := \frac{1}{Z} \sum_{k=1}^{\infty} 2^{-4k+1} \delta_{1-2^{-4k+1}} + 2^{-4k-1} \delta_{-1+2^{-4k}},$$

where $Z > 0$ is a normalization constant. Then $+1$ and -1 are incomparable because

$$\frac{\mu(B_{2^{-4k+1}}(+1))}{\mu(B_{2^{-4k+1}}(-1))} = \frac{(15Z)^{-1} \times 2^{-4k+1}}{4 \times (15Z)^{-1} \times 2^{-4k+1}} = \frac{1}{4}, \tag{5.7}$$

$$\frac{\mu(B_{2^{-4k-1}}(+1))}{\mu(B_{2^{-4k-1}}(-1))} = \frac{\mu(B_{2^{-4(k+1)+3}}(+1))}{\mu(B_{2^{-4k-1}}(-1))} = \frac{4 \times (15Z)^{-1} \times 2^{-4(k+1)+3}}{(15Z)^{-1} \times 2^{-4k-1}} = 4. \tag{5.8}$$

(b) Let $X = \mathbb{R}$ and define the densities

$$\Delta_{w,h}(x) = \begin{cases} h\left(1 - \frac{|x|}{w}\right) & |x| \leqslant w, \\ 0 & \text{otherwise}, \end{cases}$$

which have total mass wh and are supported on the interval $[-w, w]$. By analogy with part (a), let $\mu \in \mathscr{P}(\mathbb{R})$ have the continuous, bounded Lebesgue density

$$\rho(x) = \frac{1}{Z} \sum_{k \in \mathbb{N}} \Delta_{2^{-4k+1},1}(x - (1 - 2^{-4k+2})) + \Delta_{2^{-4k-1},1}(x - (-1 + 2^{-4k})),$$

where $Z > 0$ is a normalization constant. As (5.7) and (5.8) remain true for the measure μ in this example, the points ± 1 are incomparable.

While Example 5.9(b) shows that even a measure with a continuous, bounded Lebesgue density may have an antichain, we show in Proposition 5.14 that this antichain is never at the "top" of the order as in Example 5.7.

Examples such as Examples 5.7 and 5.9 can be extended to show that an antichain may be countably infinite. To do so, we first introduce a family of "coprime" oscillatory RCDFs to generalize the RCDFs μ^{e} and μ^{o} of Example 5.7.

Proposition 5.10 (a family of oscillatory RCDFs). *Fix $a > 1$ and a natural number $k \geqslant 2$. Construct the Lebesgue densities $\rho_k \colon \mathbb{R} \to [0, \infty]$ as in* Figure 5.2(a)*, defined by*

$$\rho_k(x) := \begin{cases} 0 & \text{if } |x| > a^{-1}, \\ \frac{1}{2} a^{\frac{n+1}{2}} & \text{if } a^{-n-1} < |x| \leqslant a^{-n} \text{ for } n \in \mathbb{N} \text{ with } k \mid n, \\ 0 & \text{if } a^{-n-1} < |x| \leqslant a^{-n} \text{ for } n \in \mathbb{N} \text{ with } k \mid n+1, \\ \frac{1}{2} a^{\frac{n}{2}} \left(\frac{1 - a^{-1/2}}{1 - a^{-1}} \right) & \text{if } a^{-n-1} < |x| \leqslant a^{-n} \text{ for } n \in \mathbb{N} \text{ with } k \nmid n \text{ and } k \nmid n+1, \\ \infty & \text{if } x = 0, \end{cases}$$

and, given $m > 0$, define the corresponding truncated densities $\rho_{k,m}(x) := \rho_k(x) \mathbb{1}\,[|x| \leqslant r(m)]$, with the truncation radius $r(m)$ chosen such that

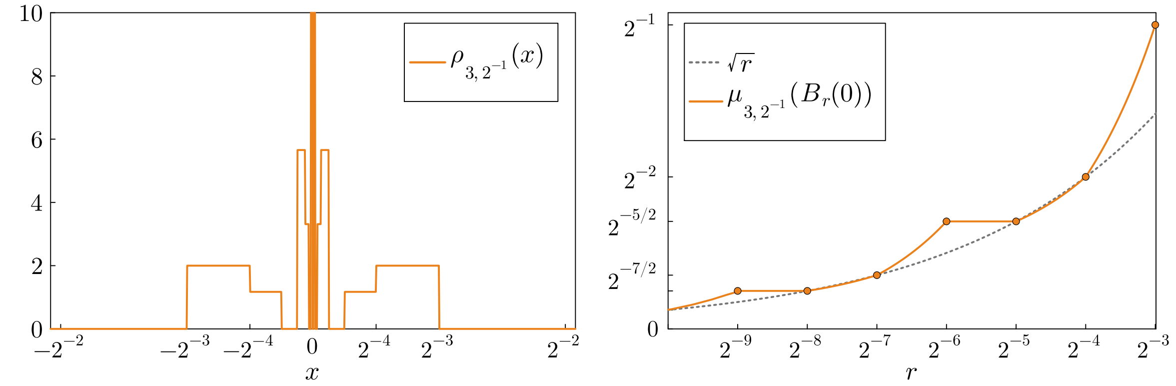
$$r(m) := \inf \left\{ s \,\middle|\, \int_{B_s(0)} \rho_k(t) \, \mathrm{d}t = m \right\}.$$

Write $\mu_{k,m}$ for the measure on the real line with $\mu_{k,m}(\mathbb{R}) = m$ and Lebesgue density $\rho_{k,m}$.

(a) *The RCDF $s \mapsto \mu_{k,m}(B_s(0))$ linearly interpolates between the knots*

$$\{(a^{-n}, a^{-n/2}) \mid n \in \mathbb{N},\ k \nmid n\} \cup \{(a^{-n}, a^{1/2 - n/2}) \mid n \in \mathbb{N},\ k \mid n\} \cup \{(0,0)\},$$

*until truncated at radius $r(m)$ (*Figure 5.2(b)*) and has formula*



(a) As in Example 5.7, the density $\rho_{k,m}$ is based on perturbations of the singularity $|\cdot|^{-1/2}$ and has mass m.

(b) The RCDF $\mu_{k,m}(B_{\cdot}(0))$ linearly interpolates between the knots $r = a^{-n}$, $n \in \mathbb{N}$ (marked as circles) to obtain the desired perturbations of the "growth rate" $\sqrt{r}$.

Figure 5.2. *Example of a density $\rho_{k,m}$ and RCDF $\mu_{k,m}$ from Proposition 5.10 with $a = 2$.*

$$\mu_{k,m}(B_s(0)) = \begin{cases} \mu_{k,m}(B_{r(m)}(0)) & \text{if } s > r(m), \\ a^{\frac{n+1}{2}} s & \text{if } a^{-n-1} < s \leqslant a^{-n} \text{ for } n \in \mathbb{N} \text{ with } k \mid n, \\ a^{-\frac{n}{2}} & \text{if } a^{-n-1} < s \leqslant a^{-n} \text{ for } n \in \mathbb{N} \text{ with } k \mid n+1, \\ \frac{1-a^{-1/2}}{1-a^{-1}} \left(a^{\frac{n}{2}} s + a^{-\frac{n}{2}}\right) & \text{if } a^{-n-1} < s \leqslant a^{-n} \text{ for } n \in \mathbb{N} \text{ with } k \nmid n \text{ and } k \nmid n+1, \\ 0 & \text{if } s = 0. \end{cases}$$

(b) *In particular, if $a^{-n} \leqslant r(m)$,*

$$\mu_{k,m}(B_{a^{-n}}(0)) = \begin{cases} a^{-n/2} & \text{if } k \nmid n, \\ a^{1/2-n/2} & \text{if } k \mid n. \end{cases}$$

(c) *Given distinct coprime integers $k, k' \geqslant 2$ and arbitrary $m, m' > 0$,*

$$\liminf_{s \to 0} \frac{\mu_{k,m}(B_s(0))}{\mu_{k',m'}(B_s(0))} < 1 < \limsup_{s \to 0} \frac{\mu_{k,m}(B_s(0))}{\mu_{k',m'}(B_s(0))}.$$

(d) *Provided $s \leqslant r(m)$, we have $\sqrt{s/a} \leqslant \mu_{k,m}(B_s(0)) \leqslant \sqrt{as}$.*
(e) *The truncation radius satisfies $r(m) \leqslant am^2$.*
(f) *The density ρ_k satisfies $\rho_k(t) \leqslant t^{-1/2}$ for all $t \in \mathbb{R}$.*

Proof.
(a) The formula for the RCDF follows by integrating the density $\rho_{k,m}$.
(b) The value at the knots a^{-n} follows from (a).
(c) We exploit the fact that k and k' are coprime, so the sequence $(n_i)_{i \in \mathbb{N}} = (ik' - 1)k \nearrow \infty$ is divisible by k but not k', and the sequence $(m_i)_{i \in \mathbb{N}} = (ik - 1)k' \nearrow \infty$ is divisible by

k' but not k. For sufficiently large i, $a^{-n_i} \leqslant \min\{r(m), r(m')\}$, and hence by (b) we obtain

$$\frac{\mu_{k,m}(B_{a^{-n_i}}(0))}{\mu_{k',m'}(B_{a^{-n_i}}(0))} = \frac{a^{1/2 - n_i/2}}{a^{-n_i/2}} = a^{\frac{1}{2}}.$$

Similarly, for i sufficiently large such that $a^{-m_i} \leqslant \min\{r(m), r(m')\}$,

$$\frac{\mu_{k,m}(B_{a^{-m_i}}(0))}{\mu_{k',m'}(B_{a^{-m_i}}(0))} = \frac{a^{-m_i/2}}{a^{1/2 - m_i/2}} = a^{-\frac{1}{2}}.$$

As these hold for all i sufficiently large, and a^{-n_i} and a^{-m_i} converge to zero, the desired inequality follows.

(d) The lower bound follows because, for $s \leqslant r(m)$,

$$\mu_{k,m}(B_s(0)) \geqslant \mu_{k,m}(B_{a^{-\lfloor -\log_a(s) \rfloor}}(0)) \geqslant a^{-\lfloor -\log_a(s) \rfloor/2} \geqslant \sqrt{s/a},$$

where the penultimate inequality uses (b); the upper bound is easily verified from the construction of $\mu_{k,m}$ as a linear interpolation of the knots.

(e) As $\int_{B_s(0)} \rho_k(t)\,\mathrm{d}t \geqslant \sqrt{s/a}$, it follows that $\int_{B_{am^2}(0)} \rho_k(t)\,\mathrm{d}t \geqslant m$, and hence $r \leqslant am^2$.

(f) This is easily verified from the expression for ρ_k. ■

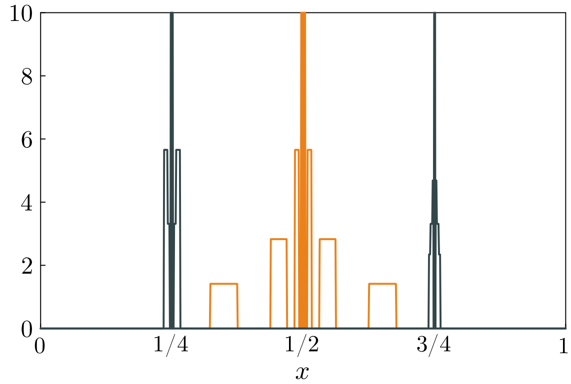
We now use Proposition 5.10 to show that a maximal antichain of a measure can be topologically dense even in the apparently well-behaved case of an absolutely continuous probability measure on the real line. Our example shows that the set of $\preccurlyeq_0$-maximal elements might be very different to the set of $\preccurlyeq_0$-greatest elements: the measure we construct has a dense set of maximal elements, yet it does not possess any greatest element because none of those maximal elements is globally comparable.

In spirit, the idea is much the same as Example 5.7: center mutually incomparable compactly supported singularities at a dense collection of points $\{q_k\}_{k\in\mathbb{N}}$. This is much more subtle, however, as one must take care to ensure that the points q_k are distant enough from one another that the singularities neither interfere with each other nor accumulate too much mass at a point outside of the dense set. Here, this is achieved by taking the q_k to be multiples of powers of two, a case that is easily analyzed but quite sparse. Indeed, we write D for the set of dyadic rationals, which we write as the disjoint union over the *levels* $D_\ell := \{(2i-1)2^{-\ell} \,|\, 1 \leqslant i \leqslant 2^{\ell-1}\}$. By a slight abuse of terminology, we also describe the sum of the densities centered at points in D_ℓ as the ℓth level of the measure.

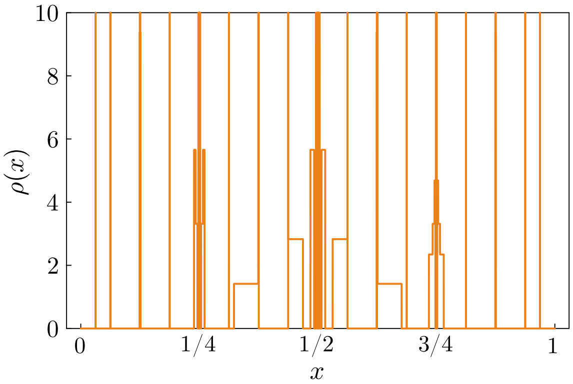
Theorem 5.11 (an absolutely continuous measure on $\mathbb{R}$ with a countable dense antichain). *Let $\mu \in \mathscr{P}(\mathbb{R})$ have the Lebesgue density $\rho \colon \mathbb{R} \to \mathbb{R}$ as shown in Figure 5.3, defined by*

$$\rho(x) := \sum_{\ell=1}^{\infty} \sum_{i=1}^{2^{\ell-1}} \rho_{k(\ell,i),m(\ell)}(x - q_{\ell,i}),$$

where $\rho_{k,m}$ is the density constructed in Proposition 5.10 with parameter $a = 2$; $k(\ell,i)$ is the $(2^{\ell-1}+i-1)$th prime; $m(\ell) := 2^{-2\ell+1}$; and $q_{\ell,i} := (2i-1)2^{-\ell} \in D_\ell$. Then,



(a) The density ρ is constructed as a sum of the prototype densities $\rho_{k,m}$. The orange density is $\rho_{2,2^{-1}}(\cdot - 1/2)$, and the grey densities are $\rho_{3,2^{-3}}(\cdot - 1/4)$ and $\rho_{5,2^{-3}}(\cdot - 3/4)$.



(b) Approximation of the density, truncated at the fifth level (i.e. with the densities centred at all dyadic rationals of the form $c2^{-n}$ with $n \leqslant 5$).

Figure 5.3. *The density ρ from Theorem 5.11 for which the dyadic rationals are an antichain.*

(a) *the* ℓth *level of the measure μ, consisting of all densities centered at points in D_ℓ, has mass $2^{-\ell}$, and hence μ is a probability measure;*
(b) *the set of dyadic rationals $D = \left\{(2i-1)2^{-\ell} \,\middle|\, \ell \in \mathbb{N},\ 1 \leqslant i \leqslant 2^{\ell-1}\right\}$ is a $\preccurlyeq_0$-antichain;*
(c) *every element of D is $\preccurlyeq_0$-maximal.*

Proof.
(a) By construction, each density in level ℓ has mass $m(\ell) = 2^{-2\ell+1}$, and there are $2^{\ell-1}$ densities, giving a total mass of $2^{-\ell}$. It follows that μ is a probability measure as $\int_{\mathbb{R}} \rho(x)\,\mathrm{d}x = \sum_{\ell\in\mathbb{N}} 2^{-\ell} = 1$.
(b) Take distinct elements $q_{\ell,i}, q_{\ell',i'} \in D$. It suffices to check that $q_{\ell,i} \not\preccurlyeq_0 q_{\ell',i'}$, as one can swap $q_{\ell,i}$ and $q_{\ell',i'}$ to obtain that $q_{\ell,i} \parallel_0 q_{\ell',i'}$. Asymptotically, $\mu(B_r(q_{\ell,i})) \sim \mu_{k(\ell,i),m(\ell)}(B_r(0))$ as $r \to 0$, and likewise $\mu(B_r(q_{\ell',i'})) \sim \mu_{k(\ell',i'),m(\ell')}(B_r(0))$ (Lemma SM1.12(a)). The identity $\limsup_{r\to 0} f(r)g(r) = \limsup_{r\to 0} f(r) \lim_{r\to 0} g(r)$ now yields

$$\begin{aligned}
&\limsup_{r\to 0} \frac{\mu(B_r(q_{\ell,i}))}{\mu(B_r(q_{\ell',i'}))} \\
&\quad = \limsup_{r\to 0} \frac{\mu_{k(\ell,i),m(\ell)}(B_r(0))}{\mu_{k(\ell',i'),m(\ell')}(B_r(0))} \lim_{r\to 0} \frac{\mu(B_r(q_{\ell,i}))}{\mu_{k(\ell,i),m(\ell)}(B_r(0))} \frac{\mu_{k(\ell',i'),m(\ell')}(B_r(0))}{\mu(B_r(q_{\ell',i'}))} \\
&\quad = \limsup_{r\to 0} \frac{\mu_{k(\ell,i),m(\ell)}(B_r(0))}{\mu_{k(\ell',i'),m(\ell')}(B_r(0))} > 1,
\end{aligned}$$

where the final line follows by the construction of the oscillatory RCDFs in Proposition 5.10(c) as $k(\ell,i)$ and $k(\ell',i')$ are distinct primes. This proves that $q_{\ell,i} \not\preccurlyeq_0 q_{\ell',i'}$ as claimed, from which incomparability follows.
(c) To show that $q \in D$ is maximal, it suffices to check that $q \not\preccurlyeq_0 x$ for any $x \in [0,1] \setminus D$; part (b) proves that $q \not\preccurlyeq_0 x$ when $x \in D$. To prove this, we must characterize the behavior of the RCDF $\mu(B_r(x))$; this depends on the properties of the binary representation of x

and in particular on a quantity we call the *dyadic irrationality exponent* $\beta_2(x) \in [1, \infty)$ (Definition SM1.10). If $\beta_2(x) < 4$, then $\mu(B_r(x)) \in o(r^{1/2})$ (Lemma SM1.12(b)); as $\mu(B_r(q)) \in \Theta(r^{1/2})$ by the construction of the density centered at q, it follows that $q \not\preccurlyeq_0 x$ because

$$\limsup_{r \to 0} \frac{\mu(B_r(x))}{\mu(B_r(q))} = 0.$$

If $\beta_2(x) > 4$, then x is approximated particularly well by a sequence of dyadic rationals, so there exists a sequence of scales as $r \to 0$ such that the RCDF $\mu(B_r(x))$ behaves much like its approximating dyadic rational. In fact, this approximation is so good that $q \parallel_0 x$ (Lemma SM1.12(d)) for the same reason that two dyadic rationals are incomparable. In the critical case $\beta_2(x) = 4$, there exist examples with $q \parallel_0 x$ and examples where $x \prec_0 q$, but in either case we can still verify that $q \not\preccurlyeq_0 x$ (Lemma SM1.12(c)) as required. This proves that no $x \in [0, 1] \setminus D$ can dominate any $q \in D$, completing the proof. ∎

Remark 5.12.

(a) The proof shows that the dyadic rationals do not form a maximal antichain in the sense of setwise inclusion: points with $\beta_2(x) > 4$ are also incomparable with the dyadic rationals; thus, the cardinality of a maximal antichain is at least $\aleph_0$. On the other hand, the Lebesgue differentiation theorem implies that any antichain has Lebesgue measure zero (see also Proposition 5.14(a)), so one cannot expect to find a larger antichain in a measure-theoretic sense.

(b) Our construction is not limited to this specific dense set and enumeration, or even to absolutely continuous measures on the real line; for example, one can reweight a Gaussian measure on a separable Hilbert space H to have a similar RCDF to our prototypical measures $\mu_{k,m}$ at the point 0, then place such measures at points in a dense subset of H. Another possibility is to argue as in Theorem 5.11 using $\mathbb{Q} \cap [0, 1]$ as the dense set; the behavior then depends on the usual number-theoretic irrationality exponent[7] instead of the dyadic irrationality exponent $\beta_2(x)$, but one still obtains a dense antichain containing all rationals in $[0, 1]$. Some of the technical steps are described in more detail in [20, section 7.3 and Appendix A].

5.4. Essential totality. The need for a $\preccurlyeq_0$-greatest element to be globally comparable is a nontrivial one, and it can fail rather dramatically, e.g., when the maximal elements form a dense antichain as in Theorem 5.11. Such examples could be criticized as somewhat artificial, but we feel that they highlight the importance of checking for incomparability and developing technical conditions on the measure which prevent it.

One could rule out incomparability if $\preccurlyeq_0$ were total, but this is not true in general, and checking this condition is often difficult in practice. We propose a somewhat weaker condition, where one can tolerate incomparability away from the "top" of the preorder, as long as any candidate for a maximal element is also globally comparable.

[7] For further details about the irrationality exponent, traditionally denoted $\mu(x)$, see, e.g., [25].

Our condition of *essential totality* can be interpreted as an order-theoretic generalization of the M-property of [3]; recall (3.8). A motivating example is that of a Gaussian measure μ on an infinite-dimensional space X: the Cameron–Martin space $H(\mu)$ is an essentially total subspace where a maximal element must lie, and any element of the Cameron–Martin space is globally comparable using the OM functional and property $M(\mu, H(\mu))$.

Definition 5.13. *Let X be a metric space and let $\mu \in \mathscr{P}(X)$. A nonempty subset $E \subseteq X$ is μ-essentially total if*

(a) *any two elements of E are comparable (i.e., E is a $\preccurlyeq_0$-chain);*
(b) *for any $x \in E$ and $x' \in X \setminus E$, $x' \preccurlyeq_0 x$; and*
(c) *for any $x' \in X \setminus E$, there exists $x \in E$ such that $x' \prec_0 x$.*

Condition (b) says that if $x^\star \in E$ is an upper bound on E, then it is $\preccurlyeq_0$-greatest; (c) says that no element in $X \setminus E$ can be greatest. We emphasize, though, that there is no need for E to be a large set in any measure-theoretic or topological sense.

Proposition 5.14 (examples of essentially total subsets).

(a) *Suppose that $X \subseteq \mathbb{R}^n$ is open and that $\mu \in \mathscr{P}(X)$ has continuous density $\rho\colon X \to [0, \infty)$ with respect to λ^n. Then $E := \{x \in X \mid \rho(x) > 0\}$ is μ-essentially total, and $I_\mu(x) := -\log \rho(x)$ is an OM functional with domain E.*
(b) *Suppose that $\mu \in \mathscr{P}(X)$ has an OM functional $I_\mu\colon E \to \mathbb{R}$ and property $M(\mu, E)$ holds. Then E is μ-essentially total.*
(c) *Suppose more generally that $\mu_0 \in \mathscr{P}(X)$ has an OM functional $I_{\mu_0}\colon E \to \mathbb{R}$ and property $M(\mu_0, E)$ holds, and that $\mu \in \mathscr{P}(X)$ has Radon–Nikodym derivative*

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(x) \propto \exp(-\Phi(x))$$

for some locally uniformly continuous potential $\Phi\colon X \to \mathbb{R}$. Then E is μ-essentially total, and $I_\mu(x) := I_{\mu_0}(x) + \Phi(x)$ is an OM functional for μ.

Proof.

(a) The Lebesgue differentiation theorem implies that for any $x \in X$,

$$\lim_{r \to 0} \frac{\mu(B_r(x))}{\lambda^n(B_r(x))} = \rho(x).$$

For any x and $x' \in E$, one can pick r sufficiently small such that $B_r(x)$ and $B_r(x')$ lie in the open set X. This implies that $\lambda^n(B_r(x)) = \lambda^n(B_r(x'))$, and so

$$\lim_{r \to 0} \frac{\mu(B_r(x))}{\mu(B_r(x'))} = \lim_{r \to 0} \frac{\mu(B_r(x))}{\lambda^n(B_r(x))} \lim_{r \to 0} \frac{\lambda^n(B_r(x'))}{\mu(B_r(x'))} = \frac{\rho(x)}{\rho(x')}. \tag{5.9}$$

Hence, E is a chain and I_μ is an OM functional on E. When $x' \in X \setminus E$, one can still apply the Lebesgue differentiation theorem to obtain

$$\lim_{r \to 0} \frac{\mu(B_r(x'))}{\lambda^n(B_r(x'))} = 0,$$

so an argument similar to that in (5.9) proves that $x' \prec_0 x$ for any $x \in E$.

(b) The existence of an OM functional I_μ proves that E is a chain. Using the M-property and [3, Lemma B.1], for $x' \in X \setminus E$ and $x \in E$, we must have $x' \prec_0 x$, because

$$\lim_{r \to 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = 0.$$

(c) By [3, Lemma B.8], I_μ is an OM functional for μ and property $M(\mu, E)$ holds. The result follows by (b). ■

Proposition 5.15. *Let X be a metric space and let $\mu \in \mathscr{P}(X)$. Suppose that $\varnothing \neq E \subseteq X$ is μ-essentially total.*

(a) *Any $\preccurlyeq_0$-maximal element must lie in E and is $\preccurlyeq_0$-greatest.*
(b) *If μ admits an OM functional $I_\mu \colon E \to \mathbb{R}$, then*

$$x^\star \text{ is } \preccurlyeq_0\text{-greatest} \iff x^\star \in E \text{ and } x^\star \text{ minimizes } I_\mu.$$

Proof.

(a) A maximal element $x^\star$ must lie in E, or else one could find $x \in E$ such that $x^\star \prec_0 x$ by essential totality, contradicting the maximality of $x^\star$. Conditions (a) and (b) of essential totality together imply that $x^\star$ is globally comparable, so it must be greatest (Lemma 5.3).
(b) Using the OM functional for E, one finds that

$$\begin{aligned} x^\star \in E \text{ is an upper bound for } E &\iff \lim_{r \to 0} \frac{\mu(B_r(x))}{\mu(B_r(x^\star))} \leqslant 1 \text{ for all } x \in E \\ &\iff \frac{e^{-I_\mu(x)}}{e^{-I_\mu(x^\star)}} \leqslant 1 \text{ for all } x \in E \\ &\iff x^\star \text{ minimizes } I_\mu. \end{aligned}$$

If $x^\star$ is $\preccurlyeq_0$-greatest, then $x^\star \in E$ by (a), and the previous implications prove that $x^\star$ minimizes I_μ. Conversely, the definition of essential totality ensures that an upper bound for E is $\preccurlyeq_0$-greatest, proving the reverse implication. ■

The variational characterization of weak modes as minimizers of the OM functional generalizes the result of [3, Proposition 4.1] to essentially total subsets. Specializing to the case of a continuous Lebesgue density on an open set (Proposition 5.14(a)) recovers the intuitive result that $x^\star$ is a weak mode if and only if it is a global maximizer of ρ. The situation is more subtle if X is not open: the measure in Example 5.4(a) restricted to $X = [0, 1]$ has a continuous Lebesgue density maximized at $x^\star = 1$, but $x^\star$ is not a weak mode.

Proposition 5.14(c) on reweightings of well-behaved measures has the significant corollary that maximal elements are always greatest when the measure is a Bayesian posterior as in (3.1) arising from a Gaussian prior. This is highly reassuring from the perspective of applications: pathological examples in the style of Theorem 5.11 with nongreatest maximal elements do not occur in Bayesian posteriors for well-behaved inverse problems.

6. Closing remarks. This article has proposed that modes of probability measures should be understood as greatest or maximal elements of preorders that are defined using the masses of metric balls.

At fixed radius $r > 0$, there is an obvious choice of total preorder, and the order-theoretic point of view opens up attractive proof techniques for the existence of maximal/greatest elements (radius-r modes) (Theorem 4.6). However, we have also seen that such radius-r modes can fail to exist (Examples 4.7 and 4.8), which provides further justification for the use of asymptotic maximizing families as proposed by [18], and we are able to contribute to the convergence analysis of such families as $r \to 0$ (Theorems 4.11 and 4.12).

In the limit as $r \to 0$, there are several limiting preorders that one could consider. The one on which we have focused, whose greatest elements are weak modes, is a nontotal preorder. Indeed, we have shown that even absolutely continuous measures can admit topologically dense antichains (Theorem 5.11), indicating that a measure must satisfy stringent regularity conditions to be certain of having greatest elements, i.e., weak modes.

As noted in the introduction, we hope that this article will stimulate further discussion in the community about the "correct" definition of a mode. We argue that there is a tension between the order-theoretic desire for modes to be merely *maximal* elements of some preorder and an application-driven desire for modes to be *greatest* elements. To some extent, this tension can be avoided if one works only with particularly nice measures that display no oscillatory properties or that satisfy criteria such as essential totality, thus keeping all pathologies away from the "top" of the preorder.

Further useful new definitions of modes may be introduced and one would hope that they correspond to preorders. However, as explored in section SM2, it may well be that such definitions only induce nontransitive *relations*. In such cases, the loss of transitivity is not necessarily fatal, so long as it is kept away from the "top" of the relation, so that maximal/greatest elements may be defined.

On a high level, it would be interesting to know whether or not there can exist a function assigning to every (sufficiently well-behaved) measure $\mu \in \mathscr{P}(X)$ a total preorder $\preccurlyeq^{\mu}$ whose maximal or greatest elements are useful modes for μ. This would appear to be a major open question that will involve much further investigation.

Acknowledgments. The authors would like to thank David Bate, Adam Epstein, Ilja Klebanov, Florian Theil, and Philipp Wacker for helpful discussions.

REFERENCES

[1] S. Agapiou, M. Burger, M. Dashti, and T. Helin, *Sparsity-promoting and edge-preserving maximum a posteriori estimators in non-parametric Bayesian inverse problems*, Inverse Problems, 34 (2018), 045002, https://doi.org/10.1088/1361-6420/aaacac.

[2] C. D. Aliprantis and K. C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 3rd ed., Springer, Berlin, 2006, https://doi.org/10.1007/3-540-29587-9.

[3] B. Ayanbayev, I. Klebanov, H. C. Lie, and T. J. Sullivan, Γ-*convergence of Onsager-Machlup functionals:* I. *With applications to maximum a posteriori estimation in Bayesian inverse problems*, Inverse Problems, 38 (2022), 025005, https://doi.org/10.1088/1361-6420/ac3f81.

[4] B. Ayanbayev, I. Klebanov, H. C. Lie, and T. J. Sullivan, Γ-*convergence of Onsager-Machlup functionals:* II. *Infinite product measures on Banach spaces*, Inverse Problems, 38 (2022), 025006, https://doi.org/10.1088/1361-6420/ac3f82.

[5] A. Björn and J. Björn, *Nonlinear Potential Theory on Metric Spaces*, EMS Tracts Math. 17, European Mathematical Society, Zürich, 2011, https://doi.org/10.4171/099.

[6] J. T. Chang and D. Pollard, *Conditioning as disintegration*, Stat. Neerl., 51 (1997), pp. 287–317, https://doi.org/10.1111/1467-9574.00056.

[7] C. Clason, T. Helin, R. Kretschmann, and P. Piiroinen, *Generalized modes in Bayesian inverse problems*, SIAM/ASA J. Uncertain. Quantif., 7 (2019), pp. 652–684, https://doi.org/10.1137/18M1191804.

[8] M. Dashti, S. Harris, and A. M. Stuart, *Besov priors for Bayesian inverse problems*, Inverse Probl. Imaging, 6 (2012), pp. 183–200, https://doi.org/10.3934/ipi.2012.6.183.

[9] M. Dashti, K. J. H. Law, A. M. Stuart, and J. Voss, *MAP estimators and their consistency in Bayesian nonparametric inverse problems*, Inverse Problems, 29 (2013), 095017, https://doi.org/10.1088/0266-5611/29/9/095017.

[10] B. A. Davey and H. A. Priestley, *Introduction to Lattices and Order*, 2nd ed., Cambridge University Press, New York, 2002, https://doi.org/10.1017/CBO9780511809088.

[11] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed., Appl. Math. 38 (New York), Springer, New York, 1998, https://doi.org/10.1007/978-3-642-03311-7.

[12] D. Dürr and A. Bach, *The Onsager-Machlup function as Lagrangian for the most probable path of a diffusion process*, Comm. Math. Phys., 60 (1978), pp. 153–170, https://doi.org/10.1007/BF01609446.

[13] W. E. W. Ren and E. Vanden-Eijnden, *Minimum action method for the study of rare events*, Comm. Pure Appl. Math., 57 (2004), pp. 637–656, https://doi.org/10.1002/cpa.20005.

[14] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, 2nd ed., Grundlehren Math. Wiss. 260, Springer, New York, 1998, https://doi.org/10.1007/978-1-4612-0611-8.

[15] B. Hansson, *Choice structures and preference relations*, Synthese, 18 (1968), pp. 443–458, https://doi.org/10.1007/BF00484979.

[16] T. Helin and M. Burger, *Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems*, Inverse Problems, 31 (2015), 085009, https://doi.org/10.1088/0266-5611/31/8/085009.

[17] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Appl. Math. Sci. 160, Springer, New York, 2005, https://doi.org/10.1007/b138659.

[18] I. Klebanov and P. Wacker, *Maximum a posteriori estimators in ℓ^p are well-defined for diagonal Gaussian priors*, Inverse Problems, 39 (2023), 065009, https://doi.org/10.1088/1361-6420/acce60.

[19] R. Kretschmann, *Nonparametric Bayesian Inverse Problems with Laplacian Noise*, Ph.D. thesis, Universität Duisburg-Essen, 2019, https://doi.org/10.17185/duepublico/70452.

[20] H. Lambley, *An Order-Theoretic Perspective on Modes and Maximum a Posteriori Estimation in Bayesian Inverse Problems*, MA4K9 Research Project (master's dissertation), University of Warwick, 2022.

[21] S. Lasanen, *Non-Gaussian statistical inverse problems. Part* I*: Posterior distributions*, Inverse Probl. Imaging, 6 (2012), pp. 215–266, https://doi.org/10.3934/ipi.2012.6.215.

[22] M. Lassas, E. Saksman, and S. Siltanen, *Discretization-invariant Bayesian inversion and Besov space priors*, Inverse Probl. Imaging, 3 (2009), pp. 87–122, https://doi.org/10.3934/ipi.2009.3.87.

[23] H. C. Lie and T. J. Sullivan, *Equivalence of weak and strong modes of measures on topological vector spaces*, Inverse Problems, 34 (2018), 115013, https://doi.org/10.1088/1361-6420/aadef2.

[24] E. Malkowsky and V. Rakočević, *Advanced Functional Analysis*, CRC Press, Boca Raton, FL, 2019, https://doi.org/10.1201/9780429442599.

[25] A. N. Parshin and I. R. Shafarevich, eds., *Number Theory* IV, Encyclopaedia Math. Sci. 44, Springer, Berlin, 1998, https://doi.org/10.1007/978-3-662-03644-0.

[26] A. M. Stuart, *Inverse problems: A Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559, https://doi.org/10.1017/S0962492910000061.

[27] V. N. Sudakov, *Linear sets with quasi-invariant measure*, Dokl. Akad. Nauk SSSR, 127 (1959), pp. 524–525.

[28] T. J. Sullivan, *Well-posed Bayesian inverse problems and heavy-tailed stable quasi-Banach space priors*, Inverse Probl. Imaging, 11 (2017), pp. 857–874, https://doi.org/10.3934/ipi.2017040.

[29] E. Szpilrajn, *Sur l'extension de l'ordre partiel*, Fund. Math., 16 (1930), pp. 386–389, https://doi.org/10.4064/fm-16-1-386-389.