An order-theoretic perspective on modes and maximum a posteriori estimation in Bayesian inverse problems

Hefin Lambley

Mathematics Institute, University of Warwick, Coventry, CV4 7AL (hefin.lambley@warwick.ac.uk)

21st September, 2022

Based on the dissertation submitted to the University of Warwick in April 2022 in partial fulfilment of the requirements for the degree of Master of Mathematics, with some minor corrections and amendments.

Contents

Acknowledgements

This work would not be possible without the support of my supervisor, Tim Sullivan, who introduced me to the fascinating world of modes and provided many valuable comments. I also thank Ilja Klebanov and Philipp Wacker for helpful discussions over the course of the project.

1 Introduction

The predictions of a Bayesian model are described through the posterior probability distribution. It is often desirable to summarise the posterior using a point estimate which represents the most plausible value for the parameters of the model.

A classical point estimate in Bayesian inference is the maximum a posteriori (MAP) estimator, or mode, which is conventionally defined to be a point maximising the posterior density. This notion is adequate when the posterior admits a continuous, bounded Lebesgue density and a maximiser of the density exists, but it is not suited to situations where the measure admits no density. Bayesian inverse problems often use infinite-dimensional parameter spaces, leading to *nonparametric* models. In infinite dimensions, the posterior measure cannot be defined with respect to the Lebesgue measure, because there is no analogue of Lebesgue measure in infinite-dimensional spaces.

More generally, modes can be defined in terms of small-ball probabilities under the posterior measure, as proposed by Stuart (2010, § 5.3). This idea developed into the strong mode of Dashti et al. (2013). Given a probability measure μ on a metric space X, a strong mode is a point $x \in X$ such that

$$\lim_{r \searrow 0} \frac{\mu(B_r(x))}{\sup_{y \in X} \mu(B_r(y))} = 1.$$

Extending this, Helin and Burger (2015) proposed the concept of a weak mode in a normed space. This paper focuses on the global weak mode, defined by Ayanbayev et al. (2022a) for metric spaces, which modifies the original definition of Helin and Burger and addresses some pathologies. A point $x \in \operatorname{supp}(\mu)$ is a global weak mode if, for all $x' \in X$,

$$\limsup_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} \leq 1. \tag{1.1}$$

We can view (1.1) as a binary relation comparing the probabilities of small balls around the points x and x'. Inspired by this, we develop an ordering $\succeq_0$ on the metric space X where $\succeq_0$-greatest elements are global weak modes. Using the ordering, pathologies such as non-existence of modes can be understood as a lack of $\succeq_0$-comparability between elements. This paper constructs examples to show that the set of $\succeq_0$-incomparable elements can be dense in the space X, including examples where X is a separable Hilbert space or a sufficiently regular separable Banach space.

The plan for this paper is as follows.

- Section 2 gives motivation from the study of Bayesian inverse problems.
- Section 3 recalls the definitions of modes used in this paper, and provides a brief overview of order-theoretic terminology.

- Section 4 studies the *radius-r ordering*, which arises by defining $x \succeq_r x' \iff \mu(B_r(x)) \geq \mu(B_r(x'))$, and proves results on the existence of *radius-r modes*.
- Section 5 derives the ordering $\succeq_0$ discussed in the introduction and proves the correspondence between global weak modes and $\succeq_0$-greatest elements. It also proves results on equivalence of strong and weak modes using the ordering.
- Section 6 reviews some alternative binary relations induced by measures and compares their features to the ordering $\succeq_0$ from Section 5.
- Section 7 studies comparability under $\succeq_0$ in more detail, with various examples of when incomparability can arise.
- Section 8 studies the connections between Onsager–Machlup (OM) functionals and the ordering $\succeq_0$.
- Section 9 concludes the paper and states some open questions.

2 Motivation from Bayesian inverse problems

The need for strong and weak modes arises from the study of Bayesian inverse problems, where small-ball modes generalise the usual definition of a MAP estimator. To motivate the study of modes in terms of small-ball probabilities, we consider a simple Bayesian inverse problem following the setup of Cotter et al. (2009).

Assume that observations $y \in \mathbb{R}^n$ arise from an unknown $x \in X$, through an *observation operator* $\mathcal{G} \colon X \to \mathbb{R}^n$. The observation operator is often based on a forward model[1] and, in practice, the observations y are corrupted by observational noise. For simplicity, we consider a model $y = \mathcal{G}(x) + \sigma$ with Gaussian noise $\sigma \sim N(0, \Sigma)$, where Σ is a positive-definite covariance matrix. As $y \mid x \sim N(\mathcal{G}(x), \Sigma)$, the likelihood function $L(y \mid x)$ is given by

$$L(y \mid x) \propto \exp\left(-\frac{1}{2}|y - \mathcal{G}(x)|_\Sigma^2\right),$$

where $|x|_\Sigma^2 = x^T\Sigma^{-1}x$. Bayesian inverse problems typically take X to be a Banach or Hilbert space of functions, such as $C([0,1], \mathbb{R})$ or $L^2(\Omega)$ for some domain Ω. The Bayesian approach requires a choice of prior distribution μ_0; it is common to choose a Gaussian prior on X. Cotter et al. (2009) prove that the posterior measure μ^y for $x \mid y$ is absolutely continuous with respect to μ_0, and has Radon–Nikodym derivative

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(x) \propto \exp\left(-\frac{1}{2}\,|y - \mathcal{G}(x)|_\Sigma^2\right). \tag{2.1}$$

[1]In many Bayesian inverse problems, the forward model is a partial differential equation (PDE). Stuart (2010, § 3) gives examples of Bayesian inverse problems where the observations y arise as pointwise data from a PDE model, including models of fluid dynamics and the heat equation.

The frequentist approach typically uses the maximum likelihood estimator: a point $x \in X$ maximising the likelihood $L(y \mid x)$. This is not adequate for inverse problems—many problems of interest are ill-posed, requiring regularisation through the prior to ensure solutions exist and are stable under perturbation.

A variety of approaches exist to summarise the posterior using a point estimate (Kaipio and Somersalo, 2005, § 3.1.1). While this paper focuses on the MAP estimator, the conditional mean $\mathbb{E}[x \mid y]$ is a good alternative (Burger and Lucka, 2014). However, it can be extremely challenging in practice to compute the conditional mean: after discretising the problem, the expectation is generally an integral in a high-dimensional space, which must be estimated using Markov chain Monte Carlo (MCMC) sampling (Lucka, 2012). MCMC methods often require significant tuning and computational resources to yield reasonable approximations to the desired integral, giving reason to prefer the MAP estimator in practice.

Dashti et al. (2013) and Wacker (2020) focus on MAP estimation in the situation that μ_0 is a Gaussian measure on a Banach or Hilbert space.

In this setting, the posterior cannot be expressed in terms of a density with respect to Lebesgue measure, necessitating the use of small-ball modes. While the posterior μ^y has a density with respect to the prior, it is no longer sufficient to maximise this density: no prior measure can assign the same mass to every ball in infinite dimensions, so a maximiser of the posterior density (2.1) may not genuinely be a point of maximum probability.

3 Preliminaries

In the paper, we work in the setting of metric spaces equipped with Borel probability measures. This section states the notation and assumptions used throughout the paper.

In a metric space (X, d), the closed ball of radius r is $B_r(x) := \{y \in X : d(x, y) \leq r\}$. Similarly, the open ball of radius r is denoted $B_r^\circ(x)$. The indicator function of a subset $A \subseteq X$ is denoted $\mathbf{1}_A$. The Lebesgue measure on $\mathbb{R}^n$ is denoted λ. After equipping X with a probability measure μ, we define the topological support of μ by

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

We denote the collection of all Borel probability measures on X by $\mathcal{P}(X)$.

Assumptions. Unless specified otherwise, all results assume that:

(i) (X, d) is a Polish metric space, that is a complete and separable metric space; and

(ii) μ is a probability measure on the Borel σ-algebra $\mathcal{B}(X)$.

As X is separable, the support $\operatorname{supp}(\mu)$ cannot be empty (Aliprantis and Border, 2006, Theorem 12.14). This rules out pathological measures where, for all $x \in X$ and small r, we have $\mu(B_r(x)) = 0$; the small-ball approach to modes would be hopeless in such circumstances.

It is often convenient to introduce the following assumption on the measure μ, which ensures that mass is not concentrated on the boundary of a ball.

Definition 3.2. A **spherically non-atomic measure** μ satisfies

$$\mu(B_r(x)) = \mu(B_r^\circ(x)) \text{ for all } x \in X \text{ and } r > 0. \qquad \blacklozenge$$

Example 3.3. (a) Lebesgue measure is clearly spherically non-atomic: the boundary of any metric ball has measure zero.

(b) Any non-degenerate Gaussian measure on a separable Banach space is spherically non-atomic, because $\mu(\partial B_r^\circ(x)) = 0$ for any open ball (Agapiou et al., 2018, Lemma 6.1). $\blacklozenge$

We prove in Proposition 4.2 that for such measures, $(x, r) \mapsto \mu(B_r(x))$ is continuous. In a Banach space, where the closure of $B_r^\circ(x)$ is $B_r(x)$, spherical non-atomicity is equivalent to each ball being a μ-continuity set: a set whose boundary has measure zero.[2]

We denote the supremal ball mass over all balls of radius r by

$$M_r := \sup_{x \in X} \mu(B_r(x)),$$

which is strictly positive for all $r > 0$ as the support of μ is non-empty.

3.1 Mode theory

The small-ball definition of a mode as defined by Dashti et al. (2013) has been generalised in several ways. This paper will largely focus on global weak modes in the sense of Ayanbayev et al. (2022a). We recall the definitions of strong and weak modes used in the literature[3] and the relations between these concepts.

Definition 3.4. A **strong mode** of μ (Ayanbayev et al., 2022a, Definition 3.6) is any point $x \in X$ such that

$$\lim_{r \searrow 0} \frac{\mu(B_r(x))}{M_r} = 1. \qquad \blacklozenge$$

[2]See Bogachev (2007, p. 186) for some properties of continuity sets.

[3]Much of the literature uses open balls for the definition of strong and weak modes. This paper differs by using closed balls for the definitions instead: the advantage of this choice will become clear in Section 4 when we prove that $x \mapsto \mu(B_r(x))$ is upper semicontinuous using closed balls.

Strong modes always lie in the support of μ, so

$$x \in \operatorname{supp}(\mu) \text{ is a strong mode } \iff \limsup_{r \searrow 0} \frac{M_r}{\mu(B_r(x))} \leq 1 \iff \liminf_{r \searrow 0} \frac{\mu(B_r(x))}{M_r} \geq 1.$$

The limit of the ratio $\mu(B_r(x))/M_r$ may fail to exist, as in Example 7.8.

Definition 3.5. A **global weak mode** of μ (Ayanbayev et al., 2022a, Definition 3.7) is any point $x \in \operatorname{supp}(\mu)$ satisfying, for all points $x' \in X$,

$$\limsup_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} \leq 1. \qquad \blacklozenge$$

The ratio is well-defined for all $r > 0$ as $x \in \operatorname{supp}(\mu)$. Section 5.2 examines when strong and global weak modes are equivalent. The next result verifies the implication that strong modes are weak, which follows immediately from the definitions.

Proposition 3.6. *Any strong mode is a global weak mode.*

Proof. Let $x \in X$ be a strong mode. Then $x \in \operatorname{supp}(\mu)$, and for any $x' \in X$,

$$\limsup_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = \limsup_{r \searrow 0} \frac{\mu(B_r(x'))}{M_r} \times \frac{M_r}{\mu(B_r(x))} \leq \limsup_{r \searrow 0} \frac{M_r}{\mu(B_r(x))} = 1.$$

The last inequality follows because $\mu(B_r(x'))/M_r \leq 1$ always holds. $\blacksquare$

Remark 3.7. Helin and Burger (2015) and Lie and Sullivan (2018) consider the the concepts of E-strong and E-weak modes in a topological vector space X. These differ from strong and global weak modes by only allowing shifts of ball centres by vectors in a subspace $E \subseteq X$. Ayanbayev et al. (2022a, Remark 3.8) note that the original definition by Helin and Burger for E-weak modes did not use a $\limsup$, which leads to counterexamples to the claim E-strong $\implies$ E-weak of Helin and Burger (2015, Lemma 3). Using the limit superior resolves this issue and ensures all strong modes are weak modes.

Clason et al. (2019) extended the definition of a strong mode to consider sequences of ball centres. The next example shows the benefit of this approach, particularly for probability measures with discontinuities in their densities.

Example 3.8. Let μ be the probability measure with Lebesgue density $\rho(x) = 2x\mathbf{1}_{[0,1]}(x)$, as shown in Fig. 1. This measure has no strong or global weak mode:

- If $x \in (0,1)$, then x is not a global weak mode. For $r < |1-x|/2$, the point $x' := 1 - |1-x|/2$ satisfies $\mu(B_r(x')) = 4\left(1 - \frac{|1-x|}{2}\right) r$, while $\mu(B_r(x)) = 4xr = 4\,(1 - |1-x|)\, r$. Hence,

$$\limsup_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = \frac{1 - |1-x|/2}{1 - |1-x|} > 1.$$

- Neither 0 nor 1 are global weak modes. We have $\mu(B_r(0)) = r^2$ and $\mu(B_r(1)) = r(2-r)$, so 0 cannot be a global weak mode (as $\mu(B_r(1))/\mu(B_r(0)) \to \infty$) and 1 cannot be a global weak mode (take $x' = 1 - \varepsilon$ for ε sufficiently small). ◆

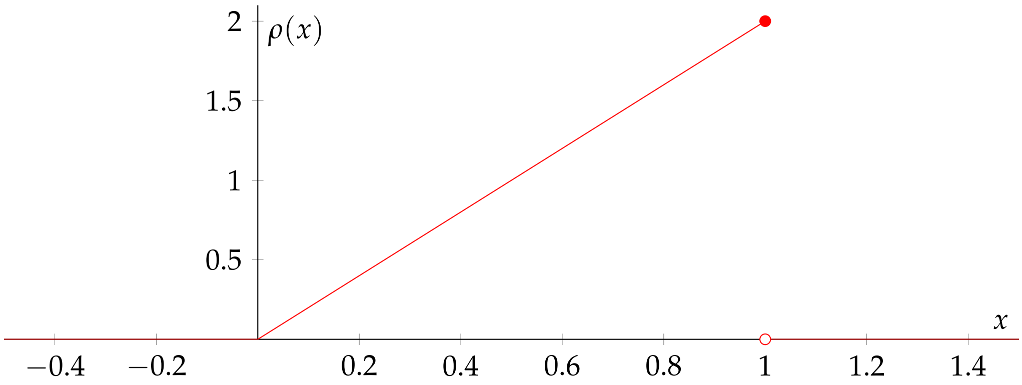

Figure 1: Density $\rho(x) = 2x\mathbf{1}_{[0,1]}(x)$ in Example 3.8.

While 1 is not a global weak mode, it is a generalised mode in the sense of Clason et al.

Definition 3.9. A **generalised (strong) mode** of μ (Clason et al., 2019, Definition 2.3) is any point $x \in X$ such that, for all sequences $(r_n) \searrow 0$, there exists an approximating sequence $(x_n)_{n\in\mathbb{N}} \to x$ satisfying

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

Strong modes are always generalised modes, using the constant sequence $x_n = x$ as an approximating sequence. Example 3.8 shows that generalised modes are not necessarily strong or weak modes.

3.2 Order theory

To study the ordering $\succeq_0$ discussed in Section 1, we need some basic notions from order theory. An excellent introduction is given by Davey and Priestley (2002).

Definition 3.10. A **preorder** (Aliprantis and Border, 2006, §1.3) on a set X is a binary relation $\succeq$ satisfying the following:

(a) (transitivity) if $x \succeq y$ and $y \succeq z$, then $x \succeq z$; and

(b) (reflexivity) for any $x \in X$, $x \succeq x$.

A **total** preorder also satisfies:

(c) (totality) for any $x, y \in X$, at least one of the relations $x \succeq y$ or $y \succeq x$ holds. ◆

Preorders differ from partial orders because they lack antisymmetry: it is possible for both $x \succeq y$ and $y \succeq x$ to hold for distinct elements $x \neq y \in X$.

The **(principal) upward closure** of a point $x \in X$ in a preorder $\succeq$ (Davey and Priestley, 2002, Definition 1.27) is defined by $\uparrow x := \{y \in X : y \succeq x\}$.

We can treat binary relations as subsets of the Cartesian product $X \times X$, equipped with a suitable product metric, such as the supremum metric.

Definition 3.11. A preorder $\succeq$ is called **continuous** (Aliprantis and Border, 2006, p. 44) if the set $\succeq$ is closed in $X \times X$. Similarly, we say $\succeq$ is **upper semicontinuous** if, for all $y \in X$, the upward closure $\uparrow y = \{x \in X : x \succeq y\}$ is closed. ◆

In non-total preorders, there is an important distinction between maximal and greatest elements. The following definitions are based on Davey and Priestley (2002, Definition 1.23).

Definition 3.12. Let $\succeq$ be a preorder on X. A point $x \in X$ is

(i) **greatest** if $x \succeq x'$ for all $x' \in X$; and

(ii) **maximal** if, for any $x' \in X$, $x' \succeq x \implies x \succeq x'$. ◆

In any preorder, greatest elements must also be maximal. In a total preorder, every maximal element is a greatest element—so the set of maximal elements coincides with the set of greatest elements.

Preorders impose fewer restrictions than partial and total orders: distinct points in a preorder may be equivalent, or may fail to be comparable at all. To describe these situations, we use the standard terminology of order theory, following Davey and Priestley (2002).

Definition 3.13. Under the analytic ordering $\succeq_0$, the points $x, x' \in X$ are

(i) **incomparable**, denoted $x \parallel_0 x'$, if $x \not\succeq_0 x'$ and $x' \not\succeq_0 x$; and

(ii) **equivalent**, denoted $x \sim_0 x'$, if $x \succeq_0 x'$ and $x' \succeq_0 x$. ◆

A total preorder has no incomparable elements by definition.

4 Fixed-radius preorder

This section studies the ordering $\succeq_r$ arising through the comparison of small-ball probabilities around points for a fixed ball radius $r > 0$. The main goal is to establish conditions which guarantee the existence of a *radius-r mode*: a point $x \in X$ such that

$$\mu(B_r(x)) = M_r. \tag{4.1}$$

4.1 Small-ball maps

Finding a radius-r mode amounts to maximising the map $x \mapsto \mu(B_r(x))$. More generally, we can consider $(x, r) \mapsto \mu(B_r(x))$ as a function of two variables; under mild assumptions, this map is jointly continuous. Knowing that this map is continuous simplifies proving results on the existence of radius-r modes in the next section.

Definition 4.1. Given the map $(x, r) \mapsto \mu(B_r(x))$, and a fixed $x \in X$, the restriction $r \mapsto \mu(B_r(x))$ is called the **radial cumulative distribution function** (RCDF) at x. ♦

The RCDF is increasing by monotonicity of probability. We now prove some continuity properties for the map $(x, r) \mapsto \mu(B_r(x))$.

Proposition 4.2. *(i) For any $r > 0$, the map $x \mapsto \mu(B_r(x))$ is upper semicontinuous.*

(ii) The RCDF $r \mapsto \mu(B_r(x))$ is differentiable for almost all $r > 0$.

(iii) Suppose that μ is spherically non-atomic. For any $r > 0$, the map $x \mapsto \mu(B_r(x))$ is continuous. Similarly, for any $x \in X$, the RCDF $r \mapsto \mu(B_r(x))$ is continuous.

(iv) Suppose that μ is spherically non-atomic. The map $(x, r) \mapsto \mu(B_r(x))$ is jointly continuous.

Proof. (i) Fix $x \in X$, and let $(x_n)_{n \in \mathbb{N}}$ be any sequence such that $(x_n) \to x$. By the reverse Fatou lemma, as the functions $\mathbf{1}_{B_r(x_n)}$ are bounded above by the integrable function $\mathbf{1}_X$, we can show that $x \mapsto \mu(B_r(x))$ is sequentially upper semicontinuous:

$$\begin{aligned}
\limsup_{n \to \infty} \mu(B_r(x_n)) &= \limsup_{n \to \infty} \int_X \mathbf{1}_{B_r(x_n)} \, \mathrm{d}\mu \\
&\leq \int_X \limsup_{n \to \infty} \mathbf{1}_{B_r(x_n)} \, \mathrm{d}\mu \\
&\leq \int_X \mathbf{1}_{B_r(x)} \, \mathrm{d}\mu \qquad \text{(Lemma A.1)} \\
&= \mu(B_r(x)).
\end{aligned}$$

(ii) By monotonicity of probability, the RCDF is an increasing function, so it is differentiable almost everywhere (Bruckner, 1978, p. 66).

(iii) Arguing as in (i), we find that $x \mapsto \mu(B_r^\circ(x))$ is lower semicontinuous[4]. By assumption, the measures of open and closed balls coincide, so $x \mapsto \mu(B_r(x))$ is both upper and lower semicontinuous, and hence continuous.

To verify that the RCDF is continuous at any $x \in X$, note that $r \mapsto \mu(B_r(x))$ right-continuous with left limits. Hence, if the RCDF is discontinuous at r^*, then

$$\lim_{r \searrow r^*} \mu(B_r(x)) = \mu(B_{r^*}(x)) \text{ and } \lim_{r \nearrow r^*} \mu(B_r(x)) < \mu(B_{r^*}(x))$$

[4] A similar argument is used by Lie and Sullivan (2018, Lemma 3.3).

because the left limit exists and the RCDF is increasing. Continuity of measure means that if $(r_n) \nearrow r^*$, then

$$\mu(B^\circ_{r^*}(x)) = \mu\left(\bigcup_{n\in\mathbb{N}} B_{r_n}(x)\right) = \lim_{n\to\infty} \mu(B_{r_n}(x)) < \mu(B_{r^*}(x)).$$

Spherical non-atomicity gives $\mu(B_{r^*}(x)) = \mu(B^\circ_{r^*}(x))$, yielding a contradiction.

(iv) By (iii), the map $(x,r) \mapsto \mu(B_r(x))$ is separately continuous; the RCDF $r \mapsto \mu(B_r(x))$ is also increasing. Separate continuity and monotonicity in one variable imply that the map is jointly continuous[5]. ■

Remark 4.3. Non-degenerate Gaussian measures are spherically non-atomic, as discussed in Example 3.3, but the behaviour of Gaussian measures is exceptionally subtle. Gaussian measures are not continuous in all directions[6] by Proposition 3.2.8 of Bogachev (2010). Consequently, if μ is a Gaussian measure on an infinite-dimensional separable Banach space X, there must exist a Borel set A such that

$$x \mapsto \mu(A + x)$$

is not continuous in x.

4.2 Definition and properties

Definition 4.4. The **fixed-radius ordering** $\succeq_r$ on X is the binary relation defined by

$$x \succeq_r x' \iff \mu(B_r(x)) \geq \mu(B_r(x')).$$

A greatest element of $\succeq_r$ is called a **radius-r mode**. ♦

We will denote the upward closure of $x \in X$ under the fixed-radius ordering by $\uparrow_r x$.

Proposition 4.5. *The fixed-radius ordering $\succeq_r$ satisfies the following properties:*

(i) $\succeq_r$ is a total preorder;

(ii) $\succeq_r$ is upper semicontinuous (in the sense of Definition 3.11); and

(iii) $x \in X$ is greatest (or, equivalently, maximal) in $\succeq_r$ if and only if $\mu(B_r(x)) = M_r$.

Proof. (i) Reflexivity and transitivity follow because $\geq$ is reflexive and transitive on $\mathbb{R}$. Totality follows by the trichotomy of $>$ on $\mathbb{R}$: if $\mu(B_r(x)) \ngeq \mu(B_r(x'))$, then $\mu(B_r(x')) \geq \mu(B_r(x))$.

[5] Young (1910) proved this for $f\colon \mathbb{R}^2 \to \mathbb{R}$. The argument generalises to functions of the form $f\colon X \times (0,\infty) \to [0,1]$ where X is a topological space (see Grushka 2019, Theorem 3.1).

[6] in the sense of Bogachev 2010, Definition 3.1.2

(ii) This follows using Proposition 4.2: the set $\uparrow_r y$ is the preimage of $[\mu(B_r(y)), \infty)$ under the map $x \mapsto \mu(B_r(x))$, so it is closed because the map is upper semicontinuous.

(iii) The point x is $\succeq_r$-greatest if and only if $\mu(B_r(x)) \geq \mu(B_r(x'))$ for all $x' \in X$. Taking the supremum over all $x' \in X$, the result follows. ■

The upward closures $\uparrow_r x$ are, under mild assumptions, closed and bounded.

Proposition 4.6. *For any $x \in X$, we have that:*

(i) the upward closure $\uparrow_r x$ is a closed subset of X; and

(ii) if $\mu(B_r(x)) > 0$, then $\uparrow_r x$ is bounded.

Proof. (i) This follows because the fixed-radius preorder $\succeq_r$ is upper semicontinuous.

(ii) We prove the result by contradiction. Fix a point $p \in \uparrow_r x$. As $\uparrow_r x$ is unbounded, there must exist a sequence $(x_n)_{n\in\mathbb{N}}$ of elements of $\uparrow_r x$ such that $d(x_n, p) \to \infty$. By passing to subsequences, we can insist that $d(x_{n+1}, p) > d(x_n, p) + 2r$ for all $n \in \mathbb{N}$. Under this assumption, for any $m < n$, we have $B_r(x_n) \cap B_r(x_m) = \emptyset$, because if $y \in B_r(x_n) \cap B_r(x_m)$, then by the triangle inequality

$$d(x_n, p) \leq d(x_n, y) + d(y, x_m) + d(x_m, p) \leq 2r + d(x_m, p).$$

By assumption, $d(x_n, p) > 2(n-m)r + d(x_m, p)$, implying that $2(n-m)r < 2r$: a contradiction. Therefore, the sequence $(B_r(x_n))_{n\in\mathbb{N}}$ is a sequence of pairwise disjoint balls, and as $\mu(B_r(x_n)) \geq \mu(B_r(x))$ for all $n \in \mathbb{N}$, we have

$$\mu(X) \geq \sum_{n\in\mathbb{N}} \mu(B_r(x)) = \infty,$$

which contradicts the fact that X is a probability space. ■

Some care must be taken when studying radius-r modes: it is not even necessary that radius-r modes lie in $\operatorname{supp}(\mu)$.

Example 4.7. Take the continuous probability density $\rho(x) \propto \max(0, 1 - 4(x-1)^2) + \max(0, 1 - 4(x+1)^2)$ on $\mathbb{R}$. A radius-1 mode is located at 0, which is not in the support of the probability measure induced by ρ. ♦

4.3 Existence of greatest elements

With the definition of the fixed-radius ordering established, we consider the question of whether a radius-r mode exists. Equivalently, we ask whether the supremum M_r in (4.1) is attained. The extreme-value theorem states that a continuous function on a compact set attains its maximum; this result remains true for upper semicontinuous functions, giving the following result.

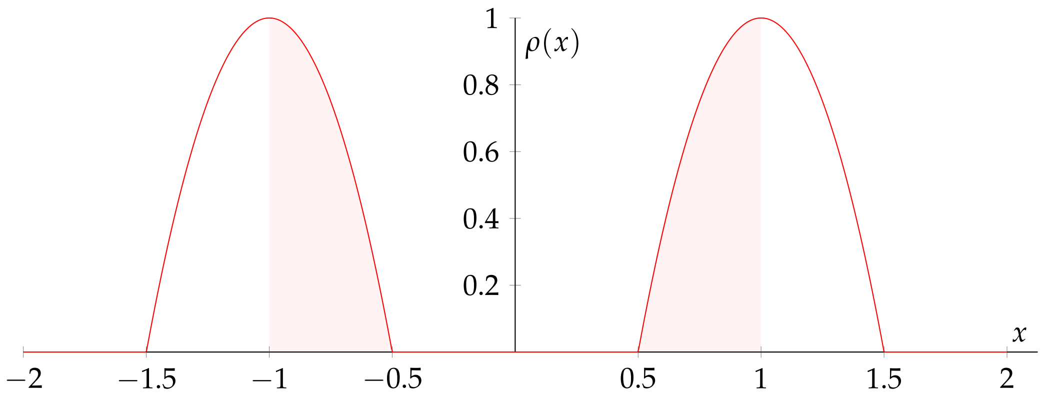

Figure 2: Density $\rho(x)$ from Example 4.7. The radius-1 ball centred at 0 attains the supremal mass M_1.

Proposition 4.8. *Let X be compact. Then a radius-r mode exists for any $r > 0$.*

Proof. Proposition 4.2(i) proves that $x \mapsto \mu(B_r(x))$ is upper semicontinuous. Hence, the function attains its maximum on the compact set X. ∎

To extend this result, we can argue as in the proof that upper semicontinuous functions attain their maximum used by Aliprantis and Border 2006, Theorem 2.43. The proof considers the upward closures $\uparrow_r x$ for each $x \in X$, and uses the the finite intersection property[7] to prove that the intersection of all upward closures $\uparrow_r x$ is non-empty. Any point in this intersection must be a radius-r mode. We modify this argument, observing that

$$\bigcap_{y \in X} \uparrow_r y = \bigcap_{y \succeq_r x} \uparrow_r y \quad \text{for any } x \in X. \tag{4.2}$$

Proposition 4.9. *Suppose that, for some $x \in X$, the upward closure $\uparrow_r x$ is compact. Then there exists a radius-r mode in X.*

Proof. Take the collection $\mathcal{C} = \{\uparrow_r y : y \succeq_r x\}$. This is a collection of closed subsets of $\uparrow_r x$ satisfying the finite intersection property. Hence, as $\uparrow_r x$ is compact, the intersection over all $C \in \mathcal{C}$ is non-empty, and (4.2) shows that any point lying in this intersection is a $\succeq_r$-greatest element. ∎

Recall that X is proper, or equivalently has the Heine–Borel property, if closed and bounded sets are compact. This property can be used to prove the existence of radius-r modes, because, when $\mu(B_r(x)) > 0$, the upward closure $\uparrow_r x$ is closed and bounded, and hence compact.

Theorem 4.10. *Suppose that X has the Heine–Borel property. Then X has a radius-r mode.*

[7] This property is an equivalent characterisation of compactness; see Munkres (2000, § 26).

Proof. Pick $y \in \operatorname{supp}(\mu) \neq \emptyset$. Then $\uparrow_r y$ is closed and bounded by Proposition 4.6. Using the Heine–Borel property, $\uparrow_r y$ is compact; the result follows by Proposition 4.9. ∎

This result shows that radius-r modes always exist in finite-dimensional Banach spaces such as $\mathbb{R}^n$. However, infinite-dimensional Banach spaces never possess the Heine–Borel property—the unit ball is closed and bounded but not compact in infinite dimensions. Williamson and Janos (1987) prove that if a metric space X is σ-compact and locally compact, then there is a metric d compatible with the topology of X such that (X, d) has the Heine–Borel property, yielding a large class of spaces for which we can apply Theorem 4.10.

While the Heine–Borel property fails for infinite-dimensional Banach spaces, it can hold in other topological vector spaces without a norm structure, such as Montel spaces[8]. Treves (1967, § 50–51) shows that a number of function spaces, including the Schwartz space $\mathcal{S}(\mathbb{R}^n)$, are Montel spaces; Theorem 4.10 therefore applies for such spaces.

The next result uses a restrictive assumption on the upward closures—namely that they shrink in diameter—to prove the existence of a unique radius-r mode. For this result to hold, the sequence $\mu(B_r(x_n))$ must be increasing so that the upward closures are nested.

Proposition 4.11. *Suppose that $(x_n)_{n\in\mathbb{N}}$ satisfies $\mu(B_r(x_n)) \nearrow M_r$ and $\operatorname{diam}(\uparrow_r x_n) \searrow 0$. Then a unique radius-$r$ mode exists.*

Proof. As $\big(\mu(B_r(x_n))\big)_{n\in\mathbb{N}}$ is increasing, the sets $\uparrow_r x_n$ are nested; they are also closed by Proposition 4.6. As $\operatorname{diam}(\uparrow_r x_n) \searrow 0$, Cantor's intersection theorem (Aliprantis and Border, 2006, Theorem 3.7) proves that

$$\bigcap_{n\in\mathbb{N}} \uparrow_r x_n \text{ is a singleton.}$$

As $\mu(B_r(x_n)) \nearrow M_r$, the point in the intersection is a radius-r mode. ∎

To find a radius-r mode, it is sufficient to take an approximating sequence $(x_n)_{n\in\mathbb{N}}$ with $\mu(B_r(x_n)) \to M_r$ and find a convergent subsequence $(x_{n_k})_{k\in\mathbb{N}}$. The limit of this subsequence must be a radius-r mode, so it is natural to ask whether $(x_n)_{n\in\mathbb{N}}$ could actually have no convergent subsequences. In a complete metric space, Lemma A.2 implies that, if $(x_n)_{n\in\mathbb{N}}$ is such a sequence, then there is a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ with the property that, for some $\delta > 0$, the balls $\big(B_\delta(x_{n_k})\big)_{k\in\mathbb{N}}$ are all disjoint.

If $\big(\mu(B_\delta(x_{n_k}))\big)_{k\in\mathbb{N}}$ is increasing, the space X would have infinite measure because the δ-balls are disjoint. Although $\big(\mu(B_r(x_{n_k}))\big)_{k\in\mathbb{N}}$ is an increasing sequence, this does not imply the same for the sequence of δ-balls. To relate the measure of r-balls and δ-balls, we consider doubling measures. Our definition follows Heinonen (2001, § 1.4) and Björn and Björn (2011, § 3).

[8] These are Hausdorff, locally convex, barrelled topological vector spaces; see Treves (1967, § 34).

Definition 4.12. A **(globally) doubling measure** μ is a measure of full support with a constant $C > 0$ such that $\mu(B_r(x)) \geq C\mu(B_{2r}(x))$ for all $x \in X$ and $r > 0$. ◆

Proposition 4.13. *Let μ be any doubling measure. Then a radius-r mode exists.*

Proof. Let $(x_n)_{n\in\mathbb{N}}$ be an approximating sequence to M_r. Suppose for contradiction it has no convergent subsequence. Then let $\delta > 0$ and $(x_{n_k})_{k\in\mathbb{N}}$ be given from Lemma A.2. We can certainly find $m \in \mathbb{N}$ such that $\delta \geq 2^{-m}r$. The balls $(B_{2^{-m}r}(x_n))_{n\in\mathbb{N}}$ are disjoint because the balls $(B_\delta(x_n))_{n\in\mathbb{N}}$ are disjoint. The doubling property implies

$$\sum_{n=1}^{\infty} \mu(B_{2^{-m}r}(x_n)) \geq \sum_{n=1}^{\infty} C^m \mu(B_r(x_n)) = \infty,$$

but this contradicts the assumption that $\mu(X) = 1$. ■

Complete metric spaces admitting a doubling measure must have the Heine–Borel property (Björn and Björn, 2011, Proposition 3.1), so an argument based on doubling measures is no extension of our existing results. It was not necessary for the measure μ to be doubling to apply the isolation argument: it may be possible to argue very carefully to extend this to a larger class of measures. Clason et al. (2019) discuss the η-annular decay property of Björn et al. (2017), which is closely related to the concept of a doubling measure. Any measure with the global η-annular decay property is doubling.

4.4 Approximation by compact sets

As X is a Polish space, and μ is a Borel probability measure on X, the measure μ is always **regular** (Cohn, 2013, Proposition 9.1.12). Inner regularity gives that, for any measurable set $A \subseteq X$ and any $\varepsilon > 0$, there exists a compact set $K \subseteq A$ such that $\mu(A \setminus K) < \varepsilon$. The following result allows us to construct a nested sequence of compact sets $K_1 \subseteq K_2 \subseteq \cdots$ such that $\mu(X \setminus K_n) \to 0$.

Proposition 4.14. *Let $(\varepsilon_n)_{n\in\mathbb{N}} \searrow 0$. There exists an increasing sequence of compact sets $(K_n)_{n\in\mathbb{N}}$ with $\mu(K_n) \geq 1 - \varepsilon_n$.*

Proof. Using inner regularity, we can construct a sequence $(\tilde{K}_n)_{n\in\mathbb{N}}$ with $\mu(\tilde{K}_n) \geq 1 - \varepsilon_n$. The result follows by defining the sets

$$K_n = \bigcup_{i=1}^{n} \tilde{K}_i$$

which are compact, nested, and satisfy $\mu(K_n) \geq \mu(\tilde{K}_n) \geq 1 - \varepsilon_n$. ■

At this point, we might hope to use Proposition 4.8 which guarantees that each K_n has a radius-r mode, then study sequences $(x_n)_{n\in\mathbb{N}}$ where x_n has maximal ball mass out of all points in K_n. We can prove that $\mu(B_r(x_n)) \to M_r$ using spherical non-atomicity; this requires that the measure have full support, so that the sequence of compact sets cannot 'miss' any open set.

Proposition 4.15. *Let μ be spherically non-atomic and suppose that* $\operatorname{supp}(\mu) = X$. *Let $(K_n)_{n\in\mathbb{N}}$ be a sequence of compact sets as constructed in Proposition 4.14. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence such that x_n maximises $x \mapsto \mu(B_r(x))$ over K_n. Then*

$$\sup_{n\in\mathbb{N}} \mu(B_r(x_n)) = M_r.$$

Proof. Suppose for contradiction that, for some $\varepsilon > 0$, we have $\sup_{n\in\mathbb{N}} \mu(B_r(x_n)) = M_r - \varepsilon$. As M_r is the supremal ball mass over all $x \in X$, we can find $y \in X$ with $\mu(B_r(y)) > M_r - \varepsilon/2$. As $x \mapsto \mu(B_r(x))$ is continuous, we can find $\delta > 0$ such that for $x \in B_\delta(y)$, we have $\mu(B_r(x)) > M_r - \varepsilon$. As $\operatorname{supp}(\mu) = X$, and $\mu(\cup_n K_n) = 1$, $\cup_n K_n$ must be dense in X by Lemma A.3, so $(\cup_n K_n) \cap B_\delta(y) \neq \varnothing$, giving a contradiction. ■

The main difficulty with this approach is ensuring that the sequence $(x_n)_{n\in\mathbb{N}}$ from the previous result has a convergent subsequence. As discussed previously, it is not obvious that this must happen at all, so the approach of approximating by compact sets encounters the same difficulties as before.

4.5 Weak topology

When X is an infinite-dimensional Banach space, it does not possess the Heine–Borel property, but the vector space structure means that we can consider properties of an approximating sequence $(x_n)_{n\in\mathbb{N}}$ under the weak topology. In reflexive Banach spaces, every sequence has a weakly convergent subsequence by the Banach–Alaoglu theorem. However, the weak limit does not, *a priori*, have to attain the supremal ball mass.

To show that the weak limit of an approximating sequence does attain the supremum, we can try to show that $x \mapsto \mu(B_r(x))$ is weakly upper semicontinuous, or that some monotone function of this map is semicontinuous.

To show that a functional is weakly semicontinuous, one solution is to assume **quasi-convexity** (Prékopa, 1995, § 4) of the functional—the property that the preimage of $(-\infty, c)$ is convex for any $c \in \mathbb{R}$. If f is a strongly lower semicontinuous and quasi-convex functional, then $f^{-1}((-\infty, c))$ is both strongly closed and convex, and hence weakly closed (Rudin, 1973, Theorem 3.12). This ensures that f is weakly lower semicontinuous.

Lemma 4.16. *Let X be a reflexive Banach space and let $r > 0$. Suppose that $x \mapsto -\log\left(\mu(B_r(x))\right)$ is quasi-convex. Then it is is weakly lower semicontinuous, so a radius-r mode exists.*

Proof. The map $x \mapsto -\log\left(\mu(B_r(x))\right)$ is strongly lower semicontinuous because log is continuous. By assumption, the map is also quasi-convex, and therefore it is weakly lower semicontinuous. Take a minimising sequence $(x_n)_{n\in\mathbb{N}}$, which must be bounded because upward closures are bounded. The Banach–Alaoglu theorem shows that this sequence must have a weakly convergent subsequence $(x_{n_k})_{k\in\mathbb{N}} \rightharpoonup x$. Weak lower semicontinuity gives that x is a radius-r mode, because

$$\liminf_{k\to\infty} -\log\left(\mu(B_r(x_{n_k}))\right) \geq -\log\left(\mu(B_r(x))\right). \qquad \blacksquare$$

We now verify that log-concave measures satisfy the conditions of Lemma 4.16, and hence have radius-r modes.

Definition 4.17. A **log-concave measure** is a measure on a Banach space X satisfying

$$\mu\left(\lambda A + (1-\lambda)B\right) \geq \mu(A)^{\lambda}\mu(B)^{1-\lambda}$$

for $\lambda \in [0,1]$ and measurable sets $A, B \subseteq X$. ♦

Proposition 4.18. *Let X be a reflexive Banach space and let μ be a log-concave measure. Then a radius-r mode exists.*

Proof. The Minkowski sum $\lambda B_r(x) + (1-\lambda)B_r(y)$ satisfies

$$\lambda B_r(x) + (1-\lambda)B_r(y) \subseteq B_r(\lambda x + (1-\lambda)y).$$

This follows because, for $\lambda a + (1-\lambda)b \in \lambda B_r(x) + (1-\lambda)B_r(y)$, we have $\|\lambda(a-x) + (1-\lambda)(b-y)\| \leq r$. Hence, the measure μ satisfies

$$\mu\left(B_r(\lambda x + (1-\lambda)y)\right) \geq \mu\left(\lambda B_r(x) + (1-\lambda)B_r(y)\right) \geq \mu(B_r(x))^{\lambda}\mu(B_r(y))^{1-\lambda}.$$

Taking logarithms and negating, we obtain that $x \mapsto -\log\left(\mu(B_r(x))\right)$ is convex, and hence quasi-convex. The result follows by Lemma 4.16. $\blacksquare$

Examples of log-concave measures include measures on $\mathbb{R}^n$ with log-concave densities (Bogachev, 1998, Theorem 1.8.4) and centred Gaussian measures on separable Hilbert spaces (Preiss et al., 2021, p. 7).

Remark 4.19. Agapiou et al. (2018, § 2.2) study MAP estimation for inverse problems with Besov priors, which are log-concave. They prove that log-concavity is sufficient for *local* modes to coincide with *global* modes—although their usage of the term 'global weak mode' is different to ours, as Agapiou et al. (2018) use the term to mean an E-weak mode for some $E \subseteq X$.

5 Global weak modes and the analytic ordering

This section establishes the analytic ordering $\succeq_0$ described in Section 1 and its connection to global weak modes. It also discusses equivalence of strong and weak modes through the language of the analytic ordering—in particular, that the existence of a strong mode is sufficient to ensure all global weak modes are strong modes.

5.1 Definition and properties

Definition 5.1. The **analytic ordering** $\succeq_0$ on a metric space X equipped with a measure $\mu \in \mathcal{P}(X)$ is the binary relation defined as follows.

(i) For $x, x' \in \operatorname{supp}(\mu)$,

$$x \succeq_0 x' \iff \limsup_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} \leq 1.$$

(ii) For $x \in \operatorname{supp}(\mu)$ and $x' \notin \operatorname{supp}(\mu)$, $x \succeq_0 x'$ and $x' \not\succeq_0 x$.

(iii) For $x, x' \notin \operatorname{supp}(\mu)$, both $x \succeq_0 x'$ and $x' \succeq_0 x$. ♦

Proposition 5.2. *Let X be a metric space and suppose that $\mu \in \mathcal{P}(X)$. Then $\succeq_0$ is a preorder.*

Proof. Reflexivity for $x \notin \operatorname{supp}(\mu)$ follows by definition. If $x \in \operatorname{supp}(\mu)$, then it is immediate that $x \succeq_0 x$, because

$$\limsup_{r \searrow 0} \frac{\mu(B_r(x))}{\mu(B_r(x))} = 1.$$

For transitivity, suppose that $x \succeq_0 y$ and $y \succeq_0 z$. Suppose first that $z \notin \operatorname{supp}(\mu)$. If $x \notin \operatorname{supp}(\mu)$, transitivity follows by property (iii). If $x \in \operatorname{supp}(\mu)$, then $x \succeq_0 z$ by property (ii) of the definition. Now suppose that $z \in \operatorname{supp}(\mu)$. Then $x, y \in \operatorname{supp}(\mu)$ because no point outside the support can be greater than a point inside the support. Hence,

$$\limsup_{r \searrow 0} \frac{\mu(B_r(z))}{\mu(B_r(x))} \leq \limsup_{r \searrow 0} \frac{\mu(B_r(z))}{\mu(B_r(y))} \times \limsup_{r \searrow 0} \frac{\mu(B_r(y))}{\mu(B_r(x))} \leq 1. \qquad \blacksquare$$

Incomparable and equivalent pairs of elements may be characterised in terms of small-ball limits as follows.

Proposition 5.3. *Under the analytic ordering $\succeq_0$, the points $x, x' \in \operatorname{supp}(\mu)$ are*

(i) incomparable if and only if

$$\liminf_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} < 1 < \limsup_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))}.$$

(ii) equivalent if and only if

$$\lim_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = 1.$$

Proof. (i) Suppose x and x' are incomparable. As $x \not\succeq_0 x'$ and $x' \not\succeq_0 x$, we must have

$$\limsup_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} > 1 \text{ and } \limsup_{r \searrow 0} \frac{\mu(B_r(x))}{\mu(B_r(x'))} > 1. \tag{5.1}$$

Using the second inequality in (5.1) and Lemma A.4, we obtain the desired result. Conversely,

$$\liminf_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} < 1 \implies \limsup_{r \searrow 0} \frac{\mu(B_r(x))}{\mu(B_r(x'))} > 1,$$

and hence we have that $x \not\succeq_0 x'$ and $x' \not\succeq_0 x$.

(ii) Suppose that $x \sim_0 x'$. As in (i),

$$x' \succeq_0 x \iff \limsup_{r \searrow 0} \frac{\mu(B_r(x))}{\mu(B_r(x'))} \leq 1 \iff \liminf_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} \geq 1. \tag{5.2}$$

We also have $x \succeq_0 x'$; combining this with (5.2), we obtain

$$\lim_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = 1.$$

Conversely, if the limit is one, it follows immediately that $x \succeq_0 x'$ and $x' \succeq_0 x$. ■

The next result proves that $\succeq_0$-greatest elements correspond exactly to global weak modes, which is the most important property of the analytic ordering.

Theorem 5.4. *A point $x \in X$ is $\succeq_0$-greatest if and only if it is a global weak mode.*

Proof. Suppose that $x \in X$ is a greatest element. Lemma A.5 gives that $x \in \operatorname{supp}(\mu)$. Given any $x' \in X$, either:

(i) $x' \notin \operatorname{supp}(\mu)$, so $\mu(B_r(x')) = 0$ for sufficiently small r, and hence

$$\limsup_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = 0; \text{ or}$$

(ii) $x \in \operatorname{supp}(\mu)$, and because $x \succeq_0 x'$,

$$\limsup_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} \leq 1. \tag{5.3}$$

Hence, x is a global weak mode. Conversely, if $x \in \operatorname{supp}(\mu)$ is a global weak mode, then x is $\succeq_0$-greatest because, for $x' \notin \operatorname{supp}(\mu)$, we have $x \succeq_0 x'$, and for $x' \in \operatorname{supp}(\mu)$, the inequality (5.3) gives $x \succeq_0 x'$ by definition. ■

Maximal elements in $\succeq_0$ also have an interpretation in terms of small-ball probabilities.

Proposition 5.5. *Suppose that $x \in X$ is $\succeq_0$-maximal. Then, for any $x' \in X$,*

$$\liminf_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} \leq 1.$$

Proof. As x is maximal, it must lie in $\operatorname{supp}(\mu)$, or else any $x' \in \operatorname{supp}(\mu)$ would satisfy $x' \succeq_0 x$ and $x \not\succeq_0 x'$. If $x' \notin \operatorname{supp}(\mu)$, then

$$\liminf_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = 0.$$

If $x' \in \operatorname{supp}(\mu)$, then either $x \sim_0 x'$, giving the result immediately, or $x' \not\succeq_0 x$. In this case, the result follows by Lemma A.4 because

$$x' \not\succeq_0 x \iff \limsup_{r \searrow 0} \frac{\mu(B_r(x))}{\mu(B_r(x'))} > 1 \iff \liminf_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} < 1. \qquad \blacksquare$$

In any preorder, greatest elements are maximal, so all global weak modes—being $\succeq_0$-greatest elements—are $\succeq_0$-maximal.

Lemma 5.6. *(i) Let $x \in \operatorname{supp}(\mu)$ and suppose that $x \sim_0 x'$. Then $x' \in \operatorname{supp}(\mu)$.*

(ii) Let x be a strong mode. Then $x \sim_0 x'$ if and only if x' is a strong mode.

(iii) Let x be a global weak mode. Then $x \sim_0 x'$ if and only if x' is a global weak mode.

Proof. (i) If $x' \notin \operatorname{supp}(\mu)$, then $x' \not\succeq_0 x$, contradicting the assumption that $x \sim_0 x'$.

(ii) Suppose that $x \sim_0 x'$. By (i), $x' \in \operatorname{supp}(\mu)$, and using Proposition 5.3,

$$\lim_{r \searrow 0} \frac{\mu(B_r(x'))}{M_r} = \lim_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} \lim_{r \searrow 0} \frac{\mu(B_r(x))}{M_r} = 1.$$

Conversely, if x' is a strong mode, then

$$\lim_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = \lim_{r \searrow 0} \frac{\mu(B_r(x'))}{M_r} \lim_{r \searrow 0} \frac{M_r}{\mu(B_r(x))} = 1,$$

so $x \sim_0 x'$ by Proposition 5.3.

(iii) If $x \sim_0 x'$, then $x' \succeq_0 x$, so x' is also $\succeq_0$-greatest and hence a global weak mode. Conversely, if x' is $\succeq_0$-greatest, then $x' \succeq_0 x$. As x is $\succeq_0$-greatest, we also have $x \succeq_0 x'$, so equivalence follows. $\blacksquare$

Upward closures in the analytic ordering are not always closed and bounded, unlike in the fixed-radius preorder.

Example 5.7. (i) The set of global weak modes of the measure μ with Lebesgue density $\rho(x) = \mathbf{1}_{[0,1]}(x)$ is $(0,1)$, so $\uparrow_0 x = (0,1)$ for any $x \in (0,1)$, showing that the upward closure need not be closed in the analytic ordering.

(ii) Define the measure μ by the Lebesgue density

$$\rho(x) = \sum_{n \in \mathbb{N}} \mathbf{1}_{[n-2^{-2n}, n+2^{-2n}]}.$$

Every $n \in \mathbb{N}$ is a strong mode, so $\mathbb{N} \subseteq \uparrow_0 1$; hence $\uparrow_0 1$ is not bounded. ◆

5.2 Equivalence of mode types

This section derives results on the equivalence of strong, weak and generalised modes. Lie and Sullivan (2018) considered equivalence of E-strong and E-weak modes (as discussed in Remark 3.7), using a uniformity condition to guarantee this. The following result uses a similar argument to prove that elements which are $\succeq_r$-greatest for all small r are strong modes.

Proposition 5.8. *Suppose that there exists $x \in X$ and $r^* > 0$ such that x is $\succeq_r$-greatest for all $r \in (0, r^*)$. Then x is a strong mode.*

Proof. By Proposition 4.5, for all $r \in (0, r^*)$, we have

$$\frac{\mu(B_r(x))}{M_r} = 1.$$

Taking limits as $r \searrow 0$, we obtain that x is a strong mode. ■

If there is a convergent net of radius-r modes, the resulting limit is a generalised mode.

Proposition 5.9. *Suppose that $(x_r)_{r \in (0,r^*)}$ is a net such that each x_r is a radius-r mode, and $x_r \to x \in X$. Then x is a generalised mode.*

Proof. For each $r \in (0, r^*)$,

$$\frac{\mu(B_r(x_r))}{M_r} = 1,$$

so taking the limit as $r \searrow 0$, we have that x is a generalised mode. ■

Example 3.8 provides a counterexample to the claim that the limit of radius-r modes is necessarily a strong or global weak mode. Using the analytic ordering, we can prove that if a strong mode exists, the concepts of strong and global weak modes are equivalent.

Theorem 5.10. *Suppose that $x \in X$ is a strong mode. Then every global weak mode is a strong mode.*

Proof. Let $y \in X$ be a global weak mode. We must have that $y \succeq_0 x$; Proposition 3.6 shows that x is also a global weak mode, so $x \succeq_0 y$. As y is equivalent to a strong mode, it is strong by Lemma 5.6. ∎

This proves that when a strong mode exists, strong and global weak modes are equivalent. Hence, for a measure μ, either:

- μ has at least one strong mode, so all global weak modes are strong modes; or
- μ has no strong modes, but still has global weak modes (as in Example B.5 of Ayanbayev et al. 2022a); or
- μ has no strong or global weak modes (as in Example 3.8).

Clason et al. (2019, Theorem 3.1) prove necessary and sufficient conditions for generalised modes to be strong modes.

Theorem 5.11 (Clason et al. 2019). *Let $x \in X$ be a generalised mode. Then x is a strong mode if and only if for all sequences $(r_n)_{n\in\mathbb{N}} \to 0$, there exists an approximating sequence $(x_n)_{n\in\mathbb{N}} \to x$ with*

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

6 Orderings induced by measures

This section investigates other orders on X based on the measure μ. Many of the relations we derive appear natural, but do not have desirable properties such as transitivity. One particularly useful ordering is the set-theoretic ordering $\sqsupseteq_0$, which is a subset of the analytic ordering $\succeq_0$ that is derived as a limit of the fixed-radius preorders $\succeq_r$.

6.1 Set-theoretic ordering

The fixed-radius preorders $\succeq_r$ are simple to work with: they are total preorders and existence of a radius-r mode can be guaranteed in many cases. The set-theoretic ordering is one reasonable way to take the limit of the orders $(\succeq_r)_{r>0}$, observing that if $x \succeq_r x'$ for all small r, then x should be greater than x' in any reasonable limiting order.

Given a collection $(A_r)_{r>0}$ of subsets of $X \times X$, the set-theoretic limit inferior is

$$\liminf_{r\searrow 0} A_r = \bigcup_{R>0} \bigcap_{r<R} A_r.$$

This set limit agrees exactly with the notion described in the previous paragraph.

Definition 6.1. The **set-theoretic limiting ordering** is the relation

$$\sqsupseteq_0 = \liminf_{r\searrow 0} \succeq_r. \qquad \blacklozenge$$

Lemma A.6 shows that the intersection of preorders and the union of nested preorders forms a preorder, so $\sqsupseteq_0$ is a preorder. The relation $x \sqsupseteq_0 x'$ holds precisely when there exists some $r^* > 0$ (depending on x and x') such that $x \succeq_r x'$ for all $r \in (0, r^*)$. As r^* can depend on x and x', this relation can be rather subtle: for example, it is not necessary for a greatest element of $\sqsupseteq_0$ to be a radius-r mode for any $r > 0$.

Example 6.2 (Following an example of Klebanov (2022)). Define the probability measure μ on $\mathbb{R}$ by the Lebesgue density[9]

$$\rho(x) = \frac{1}{Z}\left(\frac{1}{2\sqrt{2}}|x|^{-\frac{1}{2}}\mathbf{1}_{[-\frac{1}{2},\frac{1}{2}]} + \sum_{n=1}^{\infty} 2^{\frac{n+3}{2}}\mathbf{1}_{[n-2^{-(n+1)},n+2^{-(n+1)}]}\right),$$

where $Z < \infty$ is a normalisation constant so that ρ is a probability density, as shown in Figure 3. It is straightforward to verify that:

$$\mu(B_r(0)) = \frac{\sqrt{2r}}{Z} \quad \text{for } r \leq \frac{1}{2},$$
$$\mu(B_r(n)) = \frac{2^{\frac{n+3}{2}}}{Z} r \quad \text{for } r \leq 2^{-(n+1)}.$$

In particular, $\mu(B_r(x)) \in \Theta(r)$ as $r \to 0$ for $x \neq 0$, so $0 \sqsupseteq_0 x$ for any $x \in \mathbb{R}$. Even though 0 is $\sqsupseteq_0$-greatest, it is not a radius-r mode for any $r \leq \frac{1}{4}$. To see this, pick the unique value of $n \in \mathbb{N}$ such that $2^{-(n+1)} < r \leq 2^{-n}$. Then $\mu(B_r(n-1)) = \frac{2^{\frac{n+2}{2}}}{Z} r > \frac{\sqrt{2r}}{Z} = \mu(B_r(0))$, so 0 cannot be a radius-r mode. ♦

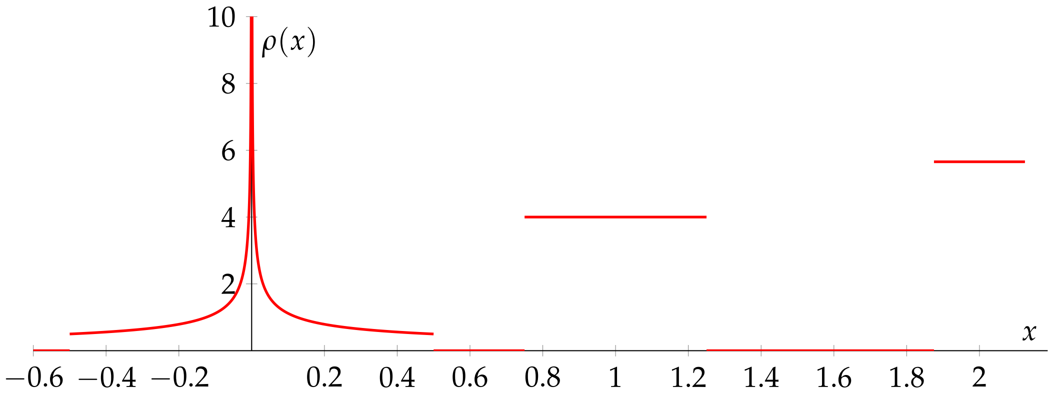

Figure 3: Unnormalised density $\rho(x)$ from Example 6.2.

Proposition 6.3. *The ordering $\sqsupseteq_0$ satisfies the following properties:*

(i) if $x \sqsupseteq_0 x'$, then $x \succeq_0 x'$, so the ordering $\sqsupseteq_0$ is a subset of $\succeq_0$; and

[9] In the original version of this manuscript, Example 6.2 contained an error. I. Klebanov kindly pointed this out and showed me a correct example; the measure given here is very similar to that construction.

(ii) for $x, x' \in \operatorname{supp}(\mu)$,

$$\limsup_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} < 1 \implies x \sqsupseteq_0 x'.$$

Proof. (i) As $x \sqsupseteq_0 x'$, there exists $r^* > 0$ such that $x \succeq_r x'$ for all $r \in (0, r^*)$, If $x \notin \operatorname{supp}(\mu)$, then $\mu(B_r(x)) = 0$ for sufficiently small r, so $x' \notin \operatorname{supp}(\mu)$ and hence $x \sim_0 x'$. If $x \in \operatorname{supp}(\mu)$, we have that $x \succeq_0 x'$ because

$$\limsup_{r \searrow 0} \underbrace{\frac{\mu(B_r(x'))}{\mu(B_r(x))}}_{\leq 1 \text{ as } x \succeq_r x' \text{ for small } r < r^*} \leq 1.$$

(ii) As the limit superior is less than 1, there exists $r^* > 0$ such that, for all $r \in (0, r^*)$,

$$\frac{\mu(B_r(x'))}{\mu(B_r(x))} < 1,$$

and therefore $x \succeq_r x'$ for all $r \in (0, r^*)$. ■

The difference between the analytic ordering and the set-theoretic ordering arises when

$$\limsup_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = 1. \tag{6.1}$$

The ratio of $\mu(B_r(x'))$ and $\mu(B_r(x))$ may be strictly greater than one for all $r > 0$, but taking the limit superior we obtain (6.1). Proposition 5.8 proved that if $x \in X$ is greatest in $\succeq_r$ for all $r \in (0, r^*)$, then x is a strong mode. But this does not imply $\sqsupseteq_0$-greatest elements are strong modes; as shown by Example 6.2. Hence, greatest elements of $\sqsupseteq_0$ do not correspond to any of the small-ball mode definitions previously discussed.

6.2 Analytic $\liminf$ ordering

Modifying the analytic ordering $\succeq_0$ to use the $\liminf$ gives a new relation $\trianglerighteq_0$ defined by

$$x \trianglerighteq_0 x' \iff \liminf_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} \leq 1.$$

Proposition 5.5 proved that the $\liminf$ criterion is satisfied for $\succeq_0$-maximal elements, giving a seemingly weaker relation. However, $\trianglerighteq_0$ fails to be transitive: it is not always true that

$$\liminf_{r \searrow 0} \frac{\mu(B_r(z))}{\mu(B_r(x))} \leq \liminf_{r \searrow 0} \frac{\mu(B_r(z))}{\mu(B_r(y))} \times \liminf_{r \searrow 0} \frac{\mu(B_r(y))}{\mu(B_r(x))}.$$

There also may not be a $\succeq_0$-maximal element, as shown by the next example.

Example 6.4. Define the probability density $\rho\colon \mathbb{R} \to [0, \infty)$ by

$$\rho(x) = \sum_{n \in \mathbb{N}} 2^n \mathbf{1}_{[n-2^{-4n}, n+2^{-4n}]}.$$

Let μ denote the corresponding probability measure on $\mathbb{R}$. We see that $n+1 \succeq_0 n$ and $n \not\succeq_0 n+1$ for all $n \in \mathbb{N}$, so no $n \in \mathbb{N}$ is $\succeq_0$-maximal. For any $x \in \mathbb{R}$, there exists $n \in \mathbb{N}$ with $n \succeq_0 x$ and $x \not\succeq_0 n$, so there is no $\succeq_0$-maximal element—in other words, no $\trianglerighteq_0$-greatest element. ♦

6.3 Kuratowski limits

Section 6.1 proved that the set-theoretic limit of the fixed-radius preorders $\succeq_r$ was weaker than the analytic ordering $\succeq_0$, in the sense that $\sqsupseteq_0$ was a subset of $\succeq_0$. A stronger form of set-theoretic limit is the lower closed limit, or Kuratowski limit, of a net of sets. The following definitions are based on those of Beer (1993, § 5.2).

Definition 6.5. Let $(\succeq_r)_{r>0}$ be any net of preorders on X, viewed as subsets of $X \times X$.

(i) The **lower closed limit** $\mathrm{Li}_{r \searrow 0} \succeq_r$ is the set of all points $(x, y) \in X \times X$ such that, for each open neighbourhood U of (x, y), there exists $R > 0$ such that, for all $r < R$, $U \cap \succeq_r \neq \varnothing$.

(ii) The **upper closed limit** $\mathrm{Ls}_{r \searrow 0} \succeq_r$ is the set of all points $(x, y) \in X \times X$ such that there exists a sequence $(x_{r_n}, y_{r_n}) \in \succeq_{r_n}$ with $(x_{r_n}, y_{r_n}) \to (x, y)$. ♦

The pair (x, y) lies in the lower closed limit $\mathrm{Li}_{r \searrow 0} \succeq_r$ if and only if there exists $r^* > 0$ and $(x_r, y_r) \to (x, y)$ such that $(x_r, y_r) \in \succeq_r$ for $r \leq r^*$ (Beer, 1993, Lemma 5.2.7).

Definition 6.6. The **metric limiting relation** $\succeq_0^{\mathrm{m}}$ is the binary relation defined by

$$\succeq_0^{\mathrm{m}} = \operatorname*{Li}_{r \searrow 0} \succeq_r. \qquad ♦$$

Reflexivity of $\succeq_0^{\mathrm{m}}$ always holds, because $(x, x) \in \succeq_r$ for all $r > 0$. However, straightforward examples as given below can cause transitivity to break down.

Example 6.7. Let μ be the probability measure on $\mathbb{R}$ with Lebesgue density $\rho = \mathbf{1}_{[0,1]}$. The net $x_r = 1 - r$ satisfies $x_r \to 1$ and $\mu(B_r(x_r)) = 2r$, so $1 \succeq_0^{\mathrm{m}} p$ for any $p \in (0, 1)$, using the net (x_r, p) in the definition. Similarly, the net $x'_r = 1 + r$ satisfies $x'_r \to 1$ and $\mu(B_r(x'_r)) = 0$, so $p' \succeq_0^{\mathrm{m}} 1$ for any $p' \in \mathbb{R} \setminus [0, 1]$ using the net (p', x'_r) in the definition. However, $p' \not\succeq_0^{\mathrm{m}} p$, so transitivity does not hold. ♦

In the previous example, changing Li to Ls in the definition of $\succeq_0^{\mathrm{m}}$ would still lead to a violation of transitivity. Hence, while the metric relation is attractive in some respects, it does not have the very desirable property of transitivity.

6.4 Extension theorems

The previous subsections developed orders which often failed to be transitive or total. It is possible to take the **transitive closure** of any binary relation R to obtain a transitive binary relation R^t with $R \subseteq R^t$ (Davey and Priestley, 2002, p. 31). While tempting, taking the transitive closure of $\trianglerighteq_0$ or $\succeq_0^{\mathrm{m}}$ gives new orders that are far less meaningful from the perspective of mode theory, as shown by the following example.

Example 6.8. Let $\succeq$ be the transitive closure of $\succeq_0^{\mathrm{m}}$ in Example 6.7. For $x, x' \in \mathbb{R}$, there are three cases:

(1) $x, x' \in [0, 1]$. Then $x \succeq_0^{\mathrm{m}} x'$ and $x' \succeq_0^{\mathrm{m}} x$, because $x \succeq_r x'$ and $x' \succeq_r x$ for all small r. Hence, x and x' are equivalent in the transitive closure.

(2) $x, x' \in \mathbb{R} \setminus [0, 1]$. Then $x \succeq_0^{\mathrm{m}} x'$ and $x' \succeq_0^{\mathrm{m}} x$ by a similar argument to (1) as $\mu(B_r(x)) = \mu(B_r(x'))$ for sufficiently small r.

(3) $x \in [0, 1]$ and $x' \in \mathbb{R} \setminus [0, 1]$. Then, by the argument in Example 6.7, $x \succeq_0^{\mathrm{m}} 1$ and $1 \succeq_0^{\mathrm{m}} x'$, so $x \succeq x'$ in the transitive closure. Similarly, $x' \succeq_0^{\mathrm{m}} 1$ and $1 \succeq_0^{\mathrm{m}} x$, so $x' \succeq x$.

Hence, the transitive closure is the entire space $X \times X$. ♦

Hansson (1968, Lemma 3) proved that any preorder can be extended to a total preorder[10]. This is not an adequate solution to incomparability—extending $\succeq_0$ to a total preorder breaks the main property that $\succeq_0$-greatest elements are global weak modes.

7 Incomparability in the analytic ordering

We turn our attention to the problem of incomparability, which arises in the analytic ordering when, for some $x, x' \in \operatorname{supp}(\mu)$,

$$\liminf_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} < 1 < \limsup_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))}. \tag{7.1}$$

In other words, incomparability arises when the ratio of the RCDFs around x and x' oscillates above and below one as $r \to 0$. In the order $\succeq_0$, it is interesting to study sets of mutually comparable and incomparable elements; such sets are called *chains* and *antichains* respectively (Davey and Priestley, 2002, Definition 1.3).

Definition 7.1. In a preorder $\succeq$ on X, a set $P \subseteq X$ is

(i) a **chain** if the elements of P are mutually comparable, i.e. $x, y \in P \implies x \not\parallel_0 y$; and

[10] A similar result for partial orders is known as Szpilrajn's theorem (Davey and Priestley, 2002, p. 244).

(ii) an **antichain** if the elements of P are mutually incomparable, i.e. $x \neq y \in P \implies x \parallel_0 y$. ♦

This section investigates when elements in the ordering $\succeq_0$ are comparable and how large an antichain can be under $\succeq_0$. A summary of the results we prove is as follows:

- **Measures on finite and countable spaces.**
  - in finite or discrete spaces, all elements are comparable (Proposition 7.4); and
  - in countable spaces, an antichain can be countable (Example 7.5).
- **Absolutely continuous measures on $\mathbb{R}^n$.**
  - antichains can be dense (Corollary 7.10); and
  - for strictly positive densities, antichains have measure zero (Proposition 7.13).
- **Separable Hilbert spaces and some separable Banach spaces.**
  - antichains can be dense (Corollary 7.12).

7.1 Comparability and growth rates

When the RCDFs $r \mapsto \mu(B_r(x))$ and $r \mapsto \mu(B_r(x'))$ grow at sufficiently different rates, we can guarantee comparability between x and x'.

Proposition 7.2. *Let $x, x' \in X$ and suppose that, as $r \searrow 0$,*

$$\mu(B_r(x)) \in \Theta(\Phi(r)) \text{ and } \mu(B_r(x')) \in \Theta(\Psi(r)), \tag{7.2}$$

for increasing functions $\Phi\colon (0,1) \to (0,\infty)$ and $\Psi\colon (0,1) \to (0,\infty)$. Suppose further that

$$\lim_{r \searrow 0} \frac{\Psi(r)}{\Phi(r)} = 0.$$

Then $x \succeq_0 x'$ and $x' \not\succeq_0 x$, and hence the points are comparable.

Proof. By (7.2), there exists $r^* > 0$ such that, for all $r \in (0, r^*)$,

$$\frac{\mu(B_r(x'))}{\mu(B_r(x))} \leq \frac{C\Psi(r)}{C'\Phi(r)}, \tag{7.3}$$

for some absolute constants $C, C' > 0$, using that $\mu(B_r(x')) \in O(\Psi(r))$ and $\mu(B_r(x)) \in \Omega(\Phi(r))$. Taking the limit superior of (7.3), we have

$$\limsup_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} \leq 0,$$

which proves that $x \succeq_0 x'$. ∎

The main requirement for the proof is the Ω bound on $\Psi(r)$ and the O bound on $\Phi(r)$: the assumption of a Θ bound on both functions is stronger than necessary.

Corollary 7.3. *If $\mu(B_r(x)) \in \Theta(r^\alpha)$ and $\mu(B_r(x')) \in \Theta(r^\beta)$, then x and x' may only be incomparable if $\alpha = \beta$.*

Proof. Without loss of generality, take $\alpha < \beta$. Then, by Proposition 7.2, as

$$\lim_{r \searrow 0} \frac{r^\beta}{r^\alpha} = \lim_{r \searrow 0} r^{\beta - \alpha} = 0,$$

we must have $x \succeq_0 x'$, so x and x' are comparable. ∎

The previous result is not even close to classifying all possible growth rates of RCDFs: a classic example of a function which is not $\Theta(r^\alpha)$ for any $\alpha \in \mathbb{R}$ is $f(r) = r\log(r)$.

Proposition 7.4. *(a) Let X be finite. Any two elements x and $x' \in X$ are comparable.*

(b) Let X be countable and $\operatorname{supp}(\mu) = X$. Then there are infinitely many points $x \in X$ which are comparable with all $x' \in X$.

Proof. If X is finite or countable, the measure μ must be atomic. If $x \in X$ is isolated, it must be comparable with all elements $x' \in X$: there must exist an open ball $B_r^\circ(x) = \{x\}$ for some $r > 0$, and so, provided $x \in \operatorname{supp}(\mu)$,

$$\lim_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = \frac{1}{\mu(\{x\})} \lim_{r \searrow 0} \mu(B_r(x')).$$

This limit exists because $r \mapsto \mu(B_r(x'))$ is increasing and takes values in teh bounded interval $[0, 1]$, so x and x' must be comparable (Proposition 5.3).

(a) Finite metric spaces are discrete, so every point is isolated. For any $x \in X$, either $x \in \operatorname{supp}(\mu)$, in which case $x \succeq_0 x'$ for any $x' \notin \operatorname{supp}(\mu)$ and x is comparable with $x' \in \operatorname{supp}(\mu)$; or $x \notin \operatorname{supp}(\mu)$, in which case x is equivalent to $x' \notin \operatorname{supp}(\mu)$ and $x' \succeq_0 x$ for $x' \in \operatorname{supp}(\mu)$.

(b) The space X is complete and can be expressed as a disjoint union of closed sets

$$X = \bigsqcup_{n \in \mathbb{N}} \{x_n\},$$

where $(x_n)_{n \in \mathbb{N}}$ is any enumeration of X. By the Baire category theorem, one of these singletons has non-empty interior, so at least one of the sets $\{x_n\}$ is open, meaning that x_n is isolated. Repeating the argument inductively by removing the isolated points, there must be a countable number of isolated points. The result follows. ∎

The following example shows how incomparability can arise in a countable space.

Example 7.5. Let X be the closure of the set $\{-1+2^{-n} : n \in \mathbb{N}\} \cup \{1-2^{-n} : n \in \mathbb{N}\}$ with the natural metric inherited from $\mathbb{R}$. Define a measure μ on X by

$$\mu = \sum_{k=1}^{\infty} 3 \times 2^{-2k-1} \delta_{-1+2^{-2k+1}} + \sum_{k=1}^{\infty} 3 \times 2^{-2k-2} \delta_{1-2^{-2k}}.$$

It can be verified that the RCDFs at the points -1 and $+1$ satisfy

$$\liminf_{r \searrow 0} \frac{\mu(B_r(+1))}{\mu(B_r(-1))} = \frac{1}{2} < 2 = \limsup_{r \searrow 0} \frac{\mu(B_r(+1))}{\mu(B_r(-1))}. \qquad \blacklozenge$$

7.2 Incomparable RCDFs

To induce incomparability between elements x and x' in the analytic ordering, we require RCDFs around x and x' whose ratio oscillates above and below one infinitely often as $r \searrow 0$. This section describes a family of RCDFs $(Z_k)_{k \in \mathbb{N}}$ such that, for all $i \neq j$,

$$\liminf_{r \searrow 0} \frac{Z_j(r)}{Z_i(r)} < 1 < \limsup_{r \searrow 0} \frac{Z_j(r)}{Z_i(r)}. \tag{7.4}$$

If these RCDFs can be realised through a probability measure μ, then incomparability can arise in the analytic ordering. This approach is very flexible: any function which agrees with the RCDFs Z_k at the points $\{2^{-n} : n \in \mathbb{N}\}$, for example, will give the required behaviour of (7.4). We also give an example to show that the RCDFs $(Z_k)_{k \in \mathbb{N}}$ can be realised as absolutely continuous probability measures on $\mathbb{R}$.

Lemma 7.6. *Let $(p_k)_{k \in \mathbb{N}}$ be the increasing sequence of primes, and let $Z_k \colon (0, \frac{1}{2}] \to (0,1)$ be any function such that, for all $n \in \mathbb{N}$,*

$$Z_k(2^{-n}) = \begin{cases} 2^{-n/2} & p_k \nmid n \\ \sqrt{2} \times 2^{-n/2} & p_k \mid n. \end{cases}$$

Then for any $i \neq j$,

$$\liminf_{r \searrow 0} \frac{Z_j(r)}{Z_i(r)} < 1 < \limsup_{r \searrow 0} \frac{Z_j(r)}{Z_i(r)}.$$

Proof. Let $(n_k)_{k \in \mathbb{N}}$ be a sequence[11] such that $p_i \mid n_k$ for all $k \in \mathbb{N}$ but $p_j \nmid n_k$ for all $k \in \mathbb{N}$. Similarly, let $(m_k)_{k \in \mathbb{N}}$ be a sequence such that $p_i \nmid m_k$ for all $k \in \mathbb{N}$ but $p_j \mid m_k$ for all $k \in \mathbb{N}$. Then

$$\frac{Z_j(m_k)}{Z_i(m_k)} = \frac{\sqrt{2} \times 2^{-m_k/2}}{2^{-m_k/2}} = \sqrt{2} \text{ and } \frac{Z_j(n_k)}{Z_i(n_k)} = \frac{2^{-n_k/2}}{\sqrt{2} \times 2^{-n_k/2}} = \frac{1}{\sqrt{2}}.$$

[11] We can always find such sequences: take $n_k = (kp_j + 1)p_i$. Then $p_j \mid n_k$ if and only if $p_j \mid p_i$; but p_i and p_j are coprime.

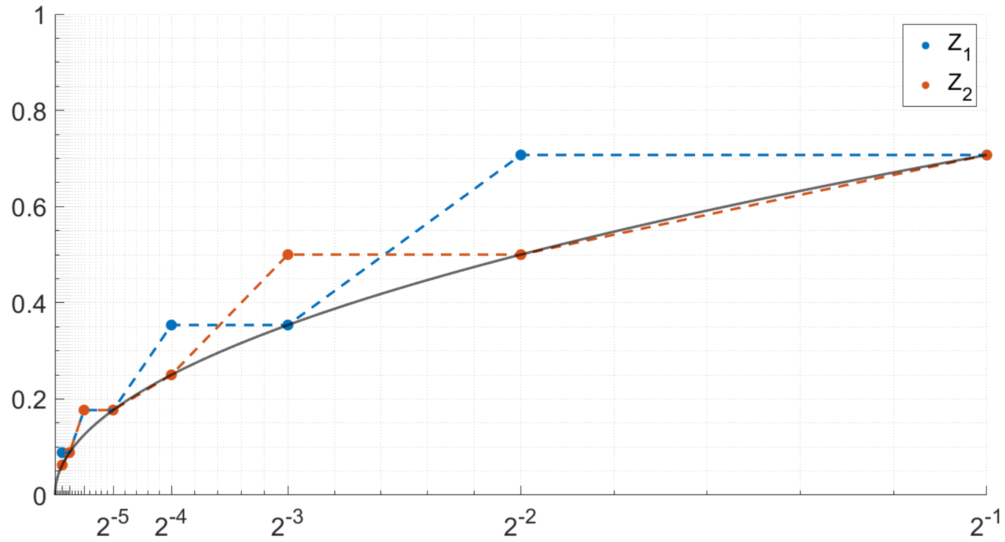

Figure 4: The RCDFs $Z_1(r)$ and $Z_2(r)$ plotted at the knots 2^{-n}. In this case, $p_1 = 2$ and $p_2 = 3$, so $Z_1(r) = \sqrt{2}r^{1/2}$ when $r = 2^{-2n}$. The dashed lines linearly interpolate between the knots. The grey line is the function $r \mapsto \sqrt{r}$. Note that a line interpolating between two points on $r \mapsto \sqrt{r}$ lies below the function because $r \mapsto \sqrt{r}$ is concave.

Therefore,

$$\liminf_{r \searrow 0} \frac{Z_j(r)}{Z_i(r)} \le \frac{1}{\sqrt{2}} < 1 \text{ and } \limsup_{r \searrow 0} \frac{Z_j(r)}{Z_i(r)} \ge \sqrt{2} > 1. \qquad \blacksquare$$

To realise the RCDFs $(Z_k)_{k \in \mathbb{N}}$ as probability measures on $\mathbb{R}$, we follow a method of I. Klebanov and linearly interpolate between the points $Z_k(2^{-n})$ for $n \in \mathbb{N}$, then compute a density compatible with the RCDF. An example of this method is given by Ayanbayev et al. (2022a, Example B.2).

Example 7.7. Linearly interpolating between the knots 2^{-n} of the RCDF Z_k, we obtain

$$Z_k(r) = \begin{cases} 0 & r = 0 \\ 2^{\frac{n+1}{2}}(\sqrt{2}-1)r + 2^{-\frac{n+1}{2}}(2-\sqrt{2}) & r \in [2^{-(n+1)}, 2^{-n}], p_k \nmid n+1 \text{ and } p_k \nmid n \\ 2^{-n/2} & r \in [2^{-(n+1)}, 2^{-n}], p_k \mid n+1 \\ 2^{\frac{n+1}{2}} r & r \in [2^{-(n+1)}, 2^{-n}], p_k \mid n. \end{cases}.$$

A compatible Lebesgue density is given by $\rho_{k,r} \colon \mathbb{R} \to [0, \infty)$, for $r \in (0, 1/2]$, by

$$\rho_{k,r}(x) = \begin{cases} 2^{\frac{n-1}{2}}(\sqrt{2}-1) & |x| \le r \text{ and } |x| \in (2^{-(n+1)}, 2^{-n}], p_k \nmid n+1 \text{ and } p_k \nmid n \\ 0 & |x| \le r \text{ and } |x| \in (2^{-(n+1)}, 2^{-n}], p_k \mid n+1 \\ 2^{\frac{n-1}{2}} & |x| \le r \text{ and } |x| \in (2^{-(n+1)}, 2^{-n}], p_k \mid n \\ 0 & x = 0 \text{ or } |x| > r. \end{cases}$$

The parameter r will be called the **truncation parameter**. Fig. 4 and Fig. 5 show the behaviour of Z_k and $\rho_{k,r}$. It is straightforward to verify that, when $2^{-n} \leq r$,

$$\int_{B_{2^{-n}}(0)} \rho_{k,r}(t)\,\mathrm{d}t = Z_k(2^{-n}) \quad \text{and} \quad \int_{X\setminus B_r(0)} \rho_{k,r}(t)\,\mathrm{d}t = 0. \qquad \blacklozenge$$

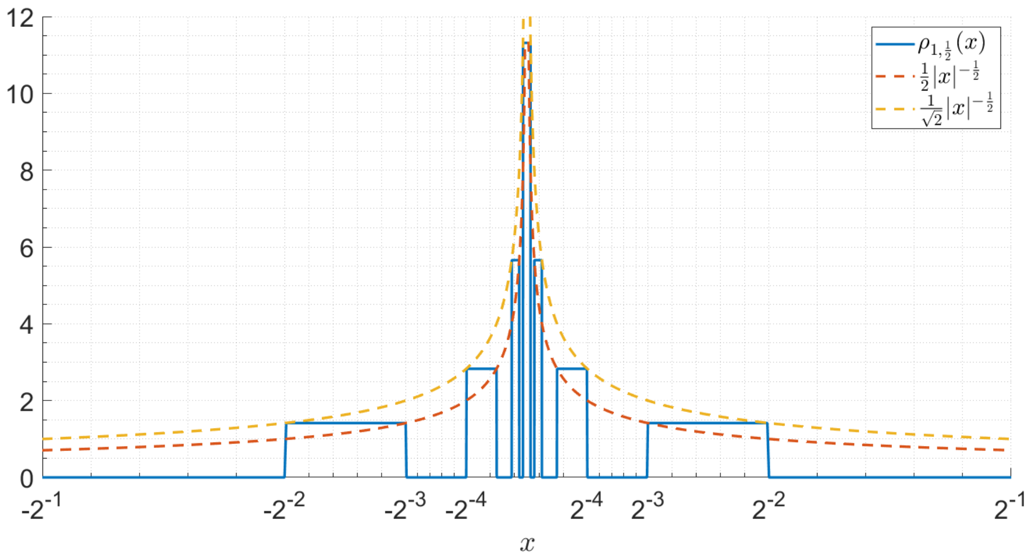

Figure 5: The density $\rho_{1,\frac{1}{2}}$, which is a sum of step functions. See also the RCDF in Fig. 4.

The next example shows how the densities $\rho_{k,r}$ can induce $\succeq_0$-incomparable elements.

Example 7.8. Take the probability density $\rho(x) = \frac{1}{2\sqrt{2}}\left(\rho_{1,2^{-1}}\left(x+\frac{1}{2}\right) + \rho_{2,2^{-1}}\left(x-\frac{1}{2}\right)\right)$, shown in Fig. 6. Let μ be the corresponding probability measure. For $r < \frac{1}{2}$, we have

$$\frac{\mu\left(B_r(\frac{1}{2})\right)}{\mu\left(B_r(-\frac{1}{2})\right)} = \frac{Z_2(r)}{Z_1(r)}.$$

Lemma 7.6 gives that $\frac{1}{2} \parallel_0 -\frac{1}{2}$, because

$$\liminf_{r\searrow 0} \frac{\mu\left(B_r(\frac{1}{2})\right)}{\mu\left(B_r(-\frac{1}{2})\right)} < 1 < \limsup_{r\searrow 0} \frac{\mu\left(B_r(\frac{1}{2})\right)}{\mu\left(B_r(-\frac{1}{2})\right)}. \tag{7.5}$$

Neither $\frac{1}{2}$ nor $-\frac{1}{2}$ are $\succeq_0$-greatest elements, because they are not comparable with every other element. No other point is a global weak mode: for $x \notin \operatorname{supp}\mu$, the analytic ordering is defined such that $x \not\succeq_0 \frac{1}{2} \in \operatorname{supp}\mu$, and for $x \in \operatorname{supp}\mu$,

$$\limsup_{r\searrow 0} \frac{\mu\left(B_r(\frac{1}{2})\right)}{\mu(B_r(x))} > 1.$$

Hence, μ has no global weak mode. This also gives an example of when the limit in the definition of a strong mode can fail to exist. ⧫

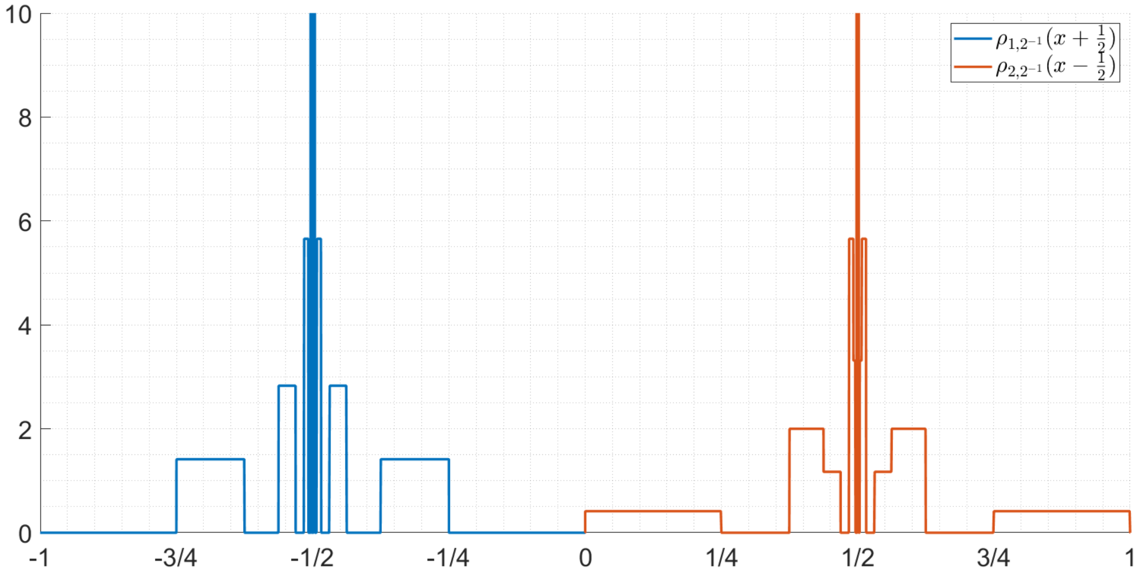

Figure 6: Density $\rho(x) = \frac{1}{2\sqrt{2}} \left(\rho_{1,2^{-1}}(x + \frac{1}{2}) + \rho_{2,2^{-1}}(x - \frac{1}{2}) \right)$ of Example 7.8.

7.3 Dense antichains

Example 7.8 constructed a probability measure with an antichain containing two elements. This section develops examples of measures with antichains which are dense in the space X. The ingredients required to construct such examples are relatively simple: it must be possible to create a family $(\mu_k)_{k\in\mathbb{N}}$ of finite measures which induce incomparability, and the space must admit a countable dense subset with an enumeration satisfying a lower bound on the distance between terms.

Theorem 7.9. *Suppose that $(X, \|\cdot\|)$ is a separable Banach space with a countable dense subset E enumerated as $(q_k)_{k\in\mathbb{N}}$. Suppose that, for any $q \in E$, the enumeration satisfies the separation bound*[12]

$$\phi(n; q) = \inf_{\substack{k\leq n \\ q_k\neq q}} \|q - q_k\| \in \Omega(n^{-1/2}) \text{ as } n \to \infty. \tag{7.6}$$

Suppose also that X admits a family $(\mu_k)_{k\in\mathbb{N}}$ of finite measures satisfying:

(i) (truncation) $\mu_k(X \setminus B_{2^{-k}}(0)) = 0$;

(ii) (incomparability of RCDFs*) for all $n \geq k$, $\mu_k(B_{2^{-n}}(0)) = Z_k(2^{-n})$; and*

[12]We call enumerations satisfying this bound **Farey enumerations**, inspired by the Farey sequence $\mathscr{F}_n$ enumerating $\mathbb{Q} \cap [0, 1]$ which satisfies this distance bound. Lemma A.8 proves Farey enumerations hold in more general settings.

(iii) (dominated ball mass away from mode) if $x \neq 0$, then

$$\lim_{r \searrow 0} \frac{\mu_k(B_r(x))}{\sqrt{r}} = 0.$$

Then X admits a probability measure μ with a countable dense antichain.

Proof. Define the measure μ by

$$\mu(A) = \frac{1}{Z} \sum_{k \in \mathbb{N}} \mu_k(A - q_k),$$

where $Z = \sum_{k \in \mathbb{N}} \mu_k(X)$ is a normalisation constant. Properties (i) and (ii) imply $\mu_k(X) = Z_k(2^{-k}) \leq \sqrt{2} \times 2^{k/2}$, so μ is a well-defined probability measure. We will prove that, for any $i \neq j$, we have $q_i \not\geq_0 q_j$, because

$$\limsup_{r \searrow 0} \frac{\mu(B_r(q_j))}{\mu(B_r(q_i))} > 1.$$

To do this, we use Lemma A.12 to prove that

$$\limsup_{r \searrow 0} \frac{\mu(B_r(q_j))}{\mu(B_r(q_i))} = \limsup_{r \searrow 0} \frac{\mu_j(B_r(0))}{\mu_i(B_r(0))} \geq \sqrt{2},$$

where the right-hand inequality follows by the construction of μ_i and μ_j to have RCDFs at 0 matching Z_i and Z_j respectively.

The essential idea is that $\mu(B_r(q_k)) \sim \frac{1}{Z}\mu_k(B_r(0))$ as $r \to 0$, so the limits agree. To apply Lemma A.12, we need that, for any $i \in \mathbb{N}$,

$$\lim_{r \searrow 0} \frac{\mu(B_r(q_i))}{\frac{1}{Z}\mu_i(B_r(0))} = 1.$$

This follows from Lemma A.14, which uses the distance bound ϕ on the $(q_k)_{k \in \mathbb{N}}$, the truncation properties of the measures $(\mu_k)_{k \in \mathbb{N}}$, and the domination property. Hence, we have

$$\limsup_{r \searrow 0} \frac{\mu(B_r(q_j))}{\mu(B_r(q_i))} = \limsup_{r \searrow 0} \frac{\mu_j(B_r(0))}{\mu_i(B_r(0))} \geq \sqrt{2},$$

because, for $r = 2^{-n} \leq 2^{-i}$, property (ii) gives $\mu_i(B_r(q_i)) = Z_i(2^{-n})$. Lemma 7.6 proved that the lim sup of the ratio of two distinct Z_k was at least $\sqrt{2}$, so the result follows. ∎

Corollary 7.10. *There is an absolutely continuous probability measure on $\mathbb{R}$ possessing a countable dense antichain.*

Proof. Lemma A.8 proves the existence of an enumeration $(q_k)_{k \in \mathbb{N}}$ of $\mathbb{Q}$ satisfying the required distance bound. With $\rho_{k,r}$ as in Example 7.7, define

$$\mu_k(A) = \int_A \rho_{k,2^{-k}}(t)\,\mathrm{d}t.$$

Example 7.7 showed the truncation property and that the RCDF at 0 agrees with Z_k. Property (iii) follows if $x \neq 0$ because, for $r < |x|/2$,

$$\int_{B_r(x)} \rho_{k,r}(t)\,\mathrm{d}t \leq 2r \sup_{t\in[x-r,x+r]} \rho_{k,r}(t).$$

The supremum is bounded above by an absolute constant $C > 0$ because any $t \in [x - r, x + r]$ must be at least a distance $|x|/2$ away from the singularity at 0. Hence

$$\lim_{r\searrow 0} \frac{\mu_k(B_r(x))}{\sqrt{r}} = 0. \qquad \blacksquare$$

There is a Farey enumeration $(q_k)_{k\in\mathbb{N}}$ of $\mathbb{Q}\cap[0,1]$ by Lemma A.7. Using the same measures μ_k as in Corollary 7.10, we can construct measures on $[0,1]$ where $\mathbb{Q}\cap[0,1]$ is an antichain. Fig. 7 gives an approximation to the density associated with the measure μ derived in Theorem 7.9 and Corollary 7.10 in the interval $[0,1]$.

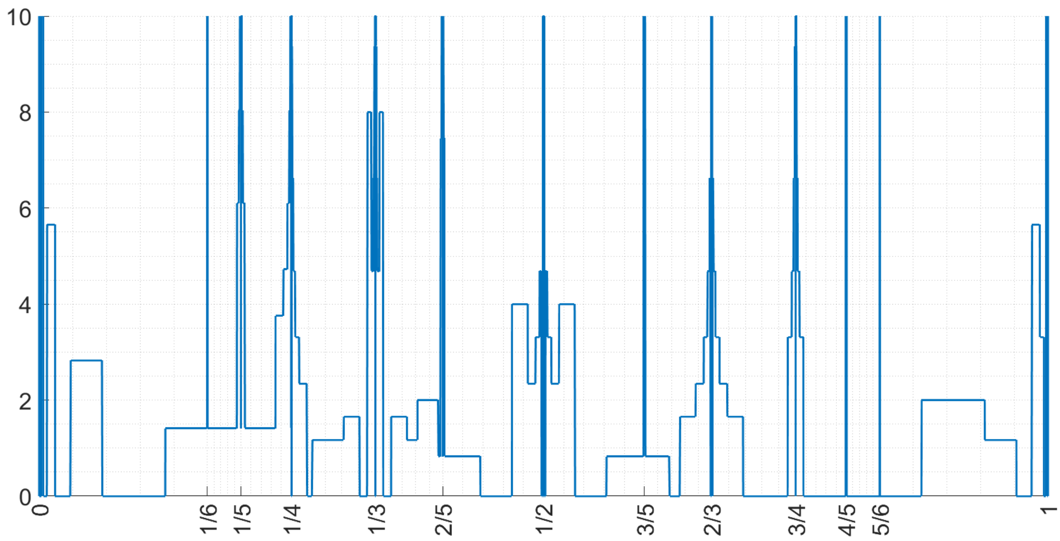

Figure 7: Approximation of the density of μ constructed in Theorem 7.9, where $X = [0,1]$. The measures μ_k are defined by the density $\rho_{k,2^{-k}}$. The enumeration $(q_k)_{k\in\mathbb{N}}$ is a Farey enumeration of $[0,1]\cap\mathbb{Q}$. For plotting purposes, measures are only placed at rationals up to denominator 6. The true density has a singularity at every rational in $[0,1]$.

Corollary 7.11. *Let X be an infinite-dimensional separable real Hilbert space. There is a probability measure possessing a countable dense antichain.*

Proof. Let μ_0 be the Gaussian measure on X defined in Lemma A.15, which is centred and non-degenerate[13]. Let $H(\mu_0)$ be the Cameron–Martin space of μ_0. The Cameron–Martin space must be dense by Theorem 3.6.1 of Bogachev (1998). As X is separable, $H(\mu_0)$ must also be separable. We also know that $H(\mu_0)$ is a Hilbert space, so we can pick an orthonormal basis $(\psi_k)_{k\in\mathbb{N}}$ of $H(\mu_0)$. Lemma A.8 derives a countable dense subset of $H(\mu_0)$ defined in terms of the basis $(\psi_k)_{k\in\mathbb{N}}$ satisfying the required distance bound.

We now construct a family of measures $(\mu_k)_{k\in\mathbb{N}}$ defined by densities with respect to μ_0 which satisfy the required properties in Theorem 7.9. To do this, we construct a density which is constant on annuli around 0. More precisely, define

$$\rho_k = \sum_{j=k}^{\infty} \underbrace{\frac{Z_k(2^{-j}) - Z_k(2^{-j-1})}{\mu_0(B_{2^{-j}}(0)) - \mu_0(B_{2^{-j-1}}(0))}}_{c_j} \mathbf{1}_{B_{2^{-j}}(0)\setminus B_{2^{-j-1}}(0)}.$$

The measure μ constructed by Theorem 7.9 shifts the measures μ_k by vectors q_k in the Cameron–Martin space, so $\mu \ll \mu_0$. We now verify that μ_k has the desired properties.

(i) By the monotone convergence theorem, as ρ_k is a sum of non-negative functions,

$$\int_{X\setminus B_{2^{-k}}(0)} \rho_k(t)\,\mathrm{d}t = \sum_{j=k}^{\infty} \int_{X\setminus B_{2^{-k}}(0)} c_j \mathbf{1}_{B_{2^{-j}}(0)\setminus B_{2^{-j-1}}(0)}(t)\,\mathrm{d}t = 0.$$

(ii) We have, for $n \geq k$, that

$$\begin{aligned}\int_{B_{2^{-n}}(0)} \rho_k(t)\,\mathrm{d}t &= \sum_{j=k}^{\infty} c_j \int_{B_{2^{-n}}(0)} \mathbf{1}_{B_{2^{-j}}(0)\setminus B_{2^{-j-1}}(0)}(t)\,\mathrm{d}t \\ &= \sum_{j=n}^{\infty} \Big(Z_k(2^{-j}) - Z_k(2^{-j-1}) \Big) = Z_k(2^{-n}).\end{aligned}$$

The sum in the last line starts at n, because the annuli $B_{2^{-j}}(0) \setminus B_{2^{-j-1}}(0)$ do not lie in $B_{2^{-n}}(0)$ for $k \leq j < n$.

(iii) Fix $x \neq 0$. Then for $r < \|x\|/2$,

$$\int_{B_r(x)} \rho_k(t)\,\mathrm{d}t \leq \mu_0(B_r(x)) \sup_{t:\|x-t\|\leq r} \rho_k(t). \tag{7.7}$$

Using that μ_0 is Gaussian and $\mu_0(B_r(0)) \in o(\sqrt{r})$ from Lemma A.15,

$$\limsup_{r\searrow 0} \frac{\mu_k(B_r(x))}{\sqrt{r}} \leq \underbrace{\limsup_{r\searrow 0} \frac{\mu_k(B_r(x))}{\mu_0(B_r(x))}}_{\text{finite by (7.7)}} \times \underbrace{\limsup_{r\searrow 0} \frac{\mu_0(B_r(x))}{\mu_0(B_r(0))}}_{\leq 1 \text{ by Anderson's inequality}} \times \underbrace{\limsup_{r\searrow 0} \frac{\mu_0(B_r(0))}{\sqrt{r}}}_{= 0 \text{ by Lemma A.15}}.$$

[13] A Gaussian measure is called *non-degenerate* if it has full support (Bogachev, 1998, Definition 3.6.2).

Combining the limits, we obtain the desired result that

$$\limsup_{r \searrow 0} \frac{\mu_k(B_r(x))}{\sqrt{r}} = 0. \qquad \blacksquare$$

It is more difficult to extend the proof to separable Banach spaces:

- A Farey enumeration in Hilbert space was found using an orthonormal basis. Separable Banach spaces may not even possess Schauder bases[14], so we will only consider Banach spaces with a **p-Riesz basis** (Christensen et al., 2021, Definition 1.2), which is a sequence $(\psi_k)_{k\in\mathbb{N}}$ satisfying $\overline{\operatorname{span}}(\{\psi_k : k \in \mathbb{N}\}) = X$ such that, for some $A, B > 0$ and for any $N \in \mathbb{N}$ and $(c_k)_{k=1}^N$,

$$A\left(\sum_{k=1}^{N} |c_k|^p\right)^{\frac{1}{p}} \leq \left\|\sum_{k=1}^{N} c_k\psi_k\right\| \leq B\left(\sum_{k=1}^{N} |c_k|^p\right)^{\frac{1}{p}}.$$

  There are many examples of Banach spaces with p-Riesz bases: for example, the canonical basis for the sequence space ℓ^p is a p-Riesz basis.

- The measures (μ_k) in Corollary 7.11 were formed by reweighting a Gaussian measure μ_0 on Hilbert space with $\mu_0(B_r(0)) \in o(\sqrt{r})$. The small-ball decay of μ_0 is necessary to ensure that the RCDFs of the reweighted measure satisfy $\mu_k(B_r(x)) \in o(\sqrt{r})$ for $x \neq 0$. Informally, we need μ_k to have a single 'peak' at 0.

  For Gaussian measures on Banach spaces, the small-ball asymptotics $\mu_0(B_r(0))$ are not as easy to estimate. To work around this, we assume that a Gaussian measure μ_0 with $\mu_0(B_r(0)) \in o(\sqrt{r})$ exists. This condition is satisfied, for example, by the Wiener measure on $C([0,1],\mathbb{R})$ (Lifshits, 1995, p. 261).

Corollary 7.12. *Let X be an infinite-dimensional separable Banach space with a p-Riesz basis $(\psi_k)_{k\in\mathbb{N}}$. Suppose that X admits a Gaussian measure μ_0 with $\mu_0(B_r(0)) \in o(\sqrt{r})$. There is a probability measure $\mu \ll \mu_0$ on X possessing a countable dense antichain.*

Proof. The proof is identical to Corollary 7.11. ■

7.4 Radon–Nikodym densities

In inverse problems, measures with Radon–Nikodym densities of the form

$$\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(u) \propto \exp\big(-\Phi(u)\big)$$

often arise, where Φ is continuous, as discussed in Section 2. This section studies techniques to relate the analytic ordering on ν with the analytic ordering on μ.

[14] As proved by Enflo (1973).

7.4.1 Lebesgue densities

For $X \subseteq \mathbb{R}^n$, it is natural to consider absolutely continuous probability measures (with respect to Lebesgue measure). In this case, we can use the Lebesgue differentiation theorem (Stein and Shakarchi, 2009, p. 104), which says that if $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ is locally integrable, then, for λ-almost all $x \in \mathbb{R}^n$,

$$\lim_{r \searrow 0} \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} f(y)\, \mathrm{d}y = f(x). \tag{7.8}$$

If f is also continuous, then (7.8) holds everywhere, meaning that measures with continuous densities have particularly nice properties. We state the results for measures on $\mathbb{R}$, but the idea readily generalises to $\mathbb{R}^n$.

Proposition 7.13. *Suppose that $\rho \colon \mathbb{R} \to (0, \infty)$ is the Lebesgue density of μ. Then*

$$\lim_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = \frac{\rho(x')}{\rho(x)}$$

holds for λ-almost all $x, x' \in X$. Furthermore, if ρ is continuous, it holds for all $x, x' \in X$.

In particular, for any probability measure μ defined by a strictly positive Lebesgue density, λ-almost all points are $\succeq_0$-comparable.

Proof. By the Lebesgue differentiation theorem, for λ-almost any $x \in \mathbb{R}$ (respectively all $x \in \mathbb{R}$ when ρ is continuous),

$$\lim_{r \searrow 0} \frac{1}{2r} \int_{x-r}^{x+r} \rho(t)\, \mathrm{d}t = \rho(x) \in (0, \infty),$$

and hence

$$\lim_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = \lim_{r \searrow 0} \frac{\mu(B_r(x'))}{2r} \times \lim_{r \searrow 0} \frac{2r}{\mu(B_r(x))} = \frac{\rho(x')}{\rho(x)}$$

holds almost everywhere (respectively everywhere), proving the desired result. ■

Corollary 7.14. *Suppose that $\rho \colon (0, \infty) \to \mathbb{R}$ is the Lebesgue density of μ and suppose that ρ is continuous. Then the analytic ordering $\succeq_0$ has no incomparable elements.*

The proof of Proposition 7.13 sheds light on why we had to exclude points where the density is zero, in order to prevent limits of the form $0/0$. Even insisting that $\operatorname{supp}(\mu) = X$ is not sufficient to guarantee that ρ is strictly positive λ-almost everywhere in X, as shown by the following example.

Example 7.15. Rudin (1987, §2, Exercise 8) gives an example of a set $E \subseteq [0, 1]$ with Lebesgue measure $0 < \lambda(E) < 1$ such that

$$\mu(A) := \frac{1}{\lambda(E)} \lambda(A \cap E)$$

assigns positive measure to any open set, but has density $\rho(x) = \mathbf{1}_E(x)$, which is zero on the set $[0,1] \setminus E$ of positive measure. ⧫

Without restricting the density to be strictly positive and finite everywhere, we can prove a weaker bound on the size of an antichain.

Proposition 7.16. *Let μ be a probability measure defined by the density $\rho \colon \mathbb{R} \to [0, +\infty]$. There must exist a $\succeq_0$-chain $V \subseteq U$ with $\lambda(V) > 0$, so $\mathbb{R}$ cannot be an antichain.*

Proof. As ρ is a probability density, it must be strictly positive on some open subset $U \subseteq X$. The density $\rho|_U$ satisfies the conditions of Proposition 7.13, so

$$\lim_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = \frac{\rho(x')}{\rho(x)} \text{ for } \lambda\text{-almost all } x, x' \in U.$$

Let $V \subseteq U$ be the set in Proposition 7.13 for which the Lebesgue differentiation theorem holds. This must be a $\succeq_0$-chain by construction, giving the desired result. ■

7.4.2 Gaussian priors

Tišer (1988) proved a Gaussian differentiation theorem similar to the Lebesgue differentiation theorem used in the previous section. The theorem holds only for Gaussian measures γ on separable Hilbert space where the eigenvalues of the covariance operator decay sufficiently quickly. Preiss et al. (2021, Corollary 3) prove that the differentiation theorem fails if the eigenvalues of the covariance operator decay as k^{-s} for $s \in (1, 6/5)$. Furthermore, they prove the existence of a function $f \in L^p(\gamma)$ for all $p \geq 1$ such that, for all $x \in X$,

$$\lim_{r \searrow 0} \frac{1}{\gamma(B_r(x))} \int_{B_r(x)} f \, \mathrm{d}\gamma = \infty.$$

In practice, the restrictive assumptions of the Gaussian differentiation theorem mean that results in the style of Proposition 7.13 do not transfer well to the infinite-dimensional case.

8 Onsager–Machlup functionals

For many measures, it is possible to define an Onsager–Machlup (OM) functional I on a subset $E \subseteq X$ which gives an expression for the limit

$$\lim_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = \frac{e^{-I(x')}}{e^{-I(x)}} \text{ for } x, x' \in E.$$

OM functionals roughly play the role of the negative logarithm of the density, although the lack of Lebesgue measure in infinite-dimensional space means this cannot be interpreted literally. OM functionals of Gaussian measures are treated in detail by Bogachev

(1998, § 4.7). In the context of Bayesian inverse problems, OM functionals were studied recently by Ayanbayev et al. (2022a,b).

8.1 Definition and properties

Definition 8.1. Given $E \subseteq \operatorname{supp}(\mu)$, the map $I \colon E \to \mathbb{R}$ is called an **Onsager–Machlup functional** (Ayanbayev et al., 2022a, Definition 3.1) if, for all $x, x' \in E$,

$$\lim_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = \frac{e^{-I(x')}}{e^{-I(x)}}.$$

Property $\boldsymbol{M(\mu, E)}$ is satisfied for the measure μ and set E if, for some $x^* \in E$,

$$x \in X \setminus E \implies \lim_{r \searrow 0} \frac{\mu(B_r(x))}{\mu(B_r(x^*))} = 0. \qquad \blacklozenge$$

OM functionals are not unique: addition of a constant gives another valid OM functional. For measures on $\mathbb{R}^n$ with a Lebesgue density, the interpretation of the OM functional as the negative logarithm of the density can be made precise.

Proposition 8.2. *Let $X \subseteq \mathbb{R}^n$, and suppose that μ is a probability measure on X defined by a Lebesgue density $\rho \colon X \to [0, \infty]$. Then there exists a set $E \subseteq \{x \in X : \rho(x) > 0\}$ with $\lambda(\{x \in X : \rho(x) > 0\} \setminus E) = 0$ such that $I \colon E \to \mathbb{R}$ defined by $I(x) = -\log \rho(x)$ is an OM functional.*

Proof. By the Lebesgue differentiation theorem, there is a set $A \subseteq X$ with $\lambda(A) = 0$ such that, for all $x \in X \setminus A$, the density ρ is finite and

$$\lim_{r \searrow 0} \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} \rho(t) \, \mathrm{d}t = \rho(x).$$

Defining $E = \{x \in X : \rho(x) > 0\} \cap (X \setminus A)$, we find that

$$\lim_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = \frac{\rho(x')}{\rho(x)} \quad \text{for all } x, x' \in E.$$

Hence, $I(x) = -\log \rho(x)$ is an OM functional. The measure of $\{x \in X : \rho(x) > 0\} \setminus E$ follows from the fact that $\lambda(A) = 0$. ■

8.2 OM functionals and the analytic ordering

It is possible to construct measures where no OM functional $I \colon E \to \mathbb{R}$ has property $M(\mu, E)$. This can occur even when $\succeq_0$ is a total order.

Example 8.3. Using the notation of Example 7.7, let $\rho(x) \propto \rho_{1, \frac{1}{2}}(x + 1) + \frac{1}{2\sqrt{2}} \rho_{2, \frac{1}{2}}(x - 1)$

and let μ be the corresponding probability measure on $\mathbb{R}$. We obtain

$$\frac{\mu(B_r(1))}{\mu(B_r(-1))} = \frac{1}{2\sqrt{2}} \frac{Z_2(r)}{Z_1(r)}.$$

The construction of Z_k ensures that

$$\liminf_{r \searrow 0} \frac{Z_2(r)}{Z_1(r)} = \frac{1}{\sqrt{2}} \quad \text{and} \quad \limsup_{r \searrow 0} \frac{Z_2(r)}{Z_1(r)} = \sqrt{2}.$$

Hence, -1 and 1 are comparable because

$$\liminf_{r \searrow 0} \frac{\mu(B_r(1))}{\mu(B_r(-1))} = \frac{1}{8} \quad \text{and} \quad \limsup_{r \searrow 0} \frac{\mu(B_r(1))}{\mu(B_r(-1))} = \frac{1}{2}.$$

If $E \subseteq \operatorname{supp}(\mu)$ contains both -1 and 1, there cannot be an OM functional defined on E because the limit of the ratio $Z_2(r)/Z_1(r)$ does not exist. If $1 \in E$ and $-1 \notin E$, or vice versa, property $M(\mu, E)$ fails because there is no $x^* \in E$ such that

$$\lim_{r \searrow 0} \frac{\mu(B_r(1))}{\mu(B_r(x^*))} = 0.$$

A similar problem arises if $1 \notin E$ and $-1 \notin E$. Therefore, no OM functional with property $M(\mu, E)$ exists for this measure. ♦

If an OM functional $I \colon E \to \mathbb{R}$ exists and property $M(\mu, E)$ holds, then incomparability only arises between points which, intuitively, never had any hope of being global weak modes.

Lemma 8.4. *Suppose that $I \colon E \to \mathbb{R}$ is an OM functional for μ. Then:*

(i) any two points in E are comparable; and

(ii) for any $x \in E$ and $x' \in X \setminus E$, we always have $x \succeq_0 x'$.

Proof. (i) This follows because, for $x, x' \in E$,

$$\lim_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = \exp(I(x) - I(x')).$$

By Proposition 5.3, as the lim sup and lim inf must agree, x and x' are comparable.

(ii) Using Ayanbayev et al. (2022a, Lemma B.1(a)), we have $x \succeq_0 x'$, because

$$x \in E \text{ and } x' \in X \setminus E \implies \lim_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = 0. \qquad \blacksquare$$

An OM functional with property $M(\mu, E)$ may exist even when some points are incomparable under $\succeq_0$.

Example 8.5. Following the method of Example 7.7, define the RCDFs F_1 and F_2 at the knots 2^{-n} by

$$F_1(2^{-n}) = \begin{cases} 2^{-2n} & 2 \nmid n \\ \frac{2^{-2n}}{4} & 2 \mid n \end{cases} \quad \text{and} \quad F_2(2^{-n}) = \begin{cases} 2^{-2n} & 3 \nmid n \\ \frac{2^{-2n}}{4} & 3 \mid n. \end{cases}$$

Extend F_1 and F_2 to the domain $(0, \frac{1}{2}]$ by linearly interpolating. Compatible densities on $\mathbb{R}$ are given by

$$\rho_1(x) = \begin{cases} \frac{3}{8}2^{-n} & |x| \in (2^{-n}, 2^{-n+1}] \\ & \text{and } 2 \mid n \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \rho_2(x) = \begin{cases} \frac{15}{8}2^{-n} & |x| \in (2^{-n}, 2^{-n+1}] \text{ and } 3 \mid n \\ \frac{3}{8}2^{-n} & |x| \in (2^{-n}, 2^{-n+1}] \text{ and } 3 \mid n+1 \\ 0 & \text{otherwise.} \end{cases}$$

The probability measure μ induced by the density $\rho \propto \rho_1(x+1) + \rho_2(x-1)$ has incomparable elements: it is straightforward to verify that $1 \parallel_0 -1$ because

$$\liminf_{r \searrow 0} \frac{F_1(r)}{F_2(r)} < 1 < \limsup_{r \searrow 0} \frac{F_1(r)}{F_2(r)}.$$

Hence, incomparable points need not lie at a singularity as in the example of Ayanbayev et al. (2022a), and incomparability can occur even for probability measures with bounded densities. By considering the small-ball probabilities of the step functions away from -1 and 1, we can see that the measure μ possesses global weak modes, and property $M(\mu, E)$ holds for sets E excluding an open interval around -1 and 1. ♦

8.3 Weak M-property

We can weaken the M-property of Ayanbayev et al. (2022a) while preserving the main theorem that minimisers of an OM functional $I \colon E \to \mathbb{R}$ with property $M(\mu, E)$ are global weak modes of μ.

Definition 8.6. Given $\varnothing \neq E \subseteq \operatorname{supp}(\mu)$, the weak M-property $\mathrm{WM}(\mu, E)$ is satisfied if:

(i) for all $x \in E$ and $x' \in X \setminus E$, we have $x \succeq_0 x'$, i.e.

$$\limsup_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} \leq 1; \text{ and}$$

(ii) for all $x' \in X \setminus E$, there exists $x \in E$ such that $x' \not\succeq_0 x$, i.e.

$$\liminf_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} < 1.$$

♦

Proposition 8.7. *Let $\varnothing \neq E \subseteq \operatorname{supp}(\mu)$. Then,* *(i) if property $M(\mu, E)$ holds, then property* $\mathrm{WM}(\mu, E)$ *also holds; and*

(ii) if property $\mathrm{WM}(\mu, E)$ *holds and $x \in X \setminus E$, then x is not $\succeq_0$-greatest, and hence not a global weak mode.*

Proof. (i) Lemma B.1(a) of Ayanbayev et al. (2022a) gives

$$x \in E \text{ and } x' \in X \setminus E \implies \lim_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = 0,$$

so properties (i) and (ii) of Definition 8.6 follow immediately.

(ii) By property (ii), there exists $x' \in E$ such that $x \not\succeq_0 x'$, so x is not $\succeq_0$-greatest. ■

As in Ayanbayev et al. (2022a), when property $\mathrm{WM}(\mu, E)$ is satisfied, we define the extended OM functional $I \colon X \to \overline{\mathbb{R}}$, taking values in the extended real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$, taking values in the extended real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ by setting $I(x) = +\infty$ for $x \in X \setminus E$.

Theorem 8.8. *Suppose that $I \colon E \to \mathbb{R}$ is an OM functional for μ satisfying property* $\mathrm{WM}(\mu, E)$. *Then*

$$x \in X \text{ is a global weak mode of } \mu \iff x \text{ minimises the extended OM functional } I.$$

Proof. Suppose that $x \in X$ is a global weak mode. Then x is $\succeq_0$-greatest, so $x \in E$. For any $x' \in E$, we have $I(x) \leq I(x')$ because

$$1 \geq \lim_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = \exp\left(I(x) - I(x')\right).$$

For $x' \in X \setminus E$, we have that $I(x) < \infty = I(x')$. Hence, x minimises the extended OM functional I. Conversely, any minimiser of I must lie in E, as I is finite only on E. Property $\mathrm{WM}(\mu, E)$ gives that $x \succeq_0 x'$ for $x' \in X \setminus E$. For $x' \in E$, we have $x \succeq_0 x'$ because

$$\limsup_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = \exp\left(I(x) - I(x')\right) \leq 1. \qquad ■$$

The measure in Example 8.3 had a global weak mode, but it did not admit an OM functional with property $M(\mu, E)$. Property $\mathrm{WM}(\mu, E)$ resolves this issue: as long as there is a global weak mode, an OM functional with property $\mathrm{WM}(\mu, E)$ exists for which the minimisers are precisely the global weak modes. This OM functional may only be trivially defined on the set of global weak modes, however.

Proposition 8.9. *Suppose that μ has a global weak mode. Then there exists a set $\varnothing \neq E \subseteq \operatorname{supp}(\mu)$ such that an OM functional $I \colon E \to \mathbb{R}$ exists, and property* $\mathrm{WM}(\mu, E)$ *holds.*

Proof. Let E be the set of global weak modes of μ. By hypothesis, $E \neq \emptyset$, and for $x, x' \in E$, we have

$$\lim_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} = 1.$$

Define the OM functional $I \colon E \to \mathbb{R}$ by $I(x) = c$ for some $c \in \mathbb{R}$. Property $\mathrm{WM}(\mu, E)$ holds: any $x \in E$ is greatest, so $x \succeq_0 x'$ for all $x' \in X \setminus E$, and we chose E to be the set of all global weak modes, so there can be no point in $X \setminus E$ greater than all elements of E. ■

9 Conclusion and final remarks

The order-theoretic perspective is an intuitive and convenient way to study global weak modes, and some situations where a global weak mode fails to exist can be understood more easily using the concept of incomparability. The examples of Section 7 show that antichains in the analytic order can be large: even topologically dense in the space X. The order-theoretic framework also makes properties such as transitivity and comparability explicit; future generalisations or modifications of the definition of a mode may benefit from having these properties.

There are several open questions to be resolved in future work. Section 4 proved various conditions which guarantee that a radius-r mode exists. However, the interesting case of Bayesian posteriors arising from a Gaussian prior does not fall under any of these conditions, so it remains open whether radius-r modes always exist in this setting. Frustratingly, I know of no counterexample of a probability measure without a radius-r mode. Without a counterexample, we cannot rule out the possibility that radius-r modes always exist. A good setting to search for such an example is an infinite-dimensional Banach or Hilbert space, where our current results do not guarantee a radius-r mode.

Section 7 proved that measures admitting a countable dense antichain exist in any Hilbert space, and in Banach spaces with p-Riesz basis and a suitable Gaussian measure. It would be natural to extend the result to any Banach space and remove the conditions imposed in Corollary 7.12.

This work has emphasised that both strong and global weak modes can be pathological. It is unclear whether there is a better definition which avoids some of these issues. Section 6 proved that some obvious modifications lead to worse orderings from the perspective of mode theory, showing that finding a good definition is a somewhat subtle problem. In future work, it would be worthwhile to investigate reasonable ways to define modes through an ordering on the space X which is transitive and total, while ensuring that the ordering reflects the measure structure so that greatest elements can meaningfully be considered modes.

References

Agapiou, Sergios et al. (2018). 'Sparsity-promoting and edge-preserving maximum *a posteriori* estimators in non-parametric Bayesian inverse problems'. In: *Inverse Problems* 34.4, 045002. DOI: 10.1088/1361-6420/aaacac.

Aliprantis, Charalambos D. and Border, Kim C. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. 3rd ed. Springer. DOI: 10.1007/978-3-662-03961-8.

Ayanbayev, Birzhan et al. (2022a). 'Γ-convergence of Onsager–Machlup functionals: I. With applications to maximum a posteriori estimation in Bayesian inverse problems'. In: *Inverse Problems* 38.2, 025005. DOI: 10.1088/1361-6420/ac3f81.

– (2022b). 'Γ-convergence of Onsager–Machlup functionals: II. Infinite product measures on Banach spaces'. In: *Inverse Problems* 38.2, 025006. DOI: 10.1088/1361-6420/ac3f82.

Beer, Gerald (1993). *Topologies on Closed and Closed Convex Sets*. Vol. 268. Mathematics and Its Applications. Springer. DOI: 10.1007/978-94-015-8149-3.

Björn, Anders and Björn, Jana (2011). *Nonlinear Potential Theory on Metric Spaces*. Tracts in Mathematics 17. European Mathematical Society. DOI: 10.4171/099.

Björn, Anders et al. (2017). 'The annular decay property and capacity estimates for thin annuli'. In: *Collectanea Mathematica* 68.2, pp. 229–241. DOI: 10.1007/s13348-016-0178-y.

Bogachev, Vladimir (1998). *Gaussian Measures*. Vol. 62. Mathematical Surveys and Monographs. American Mathematical Society. DOI: 10.1090/surv/062.

– (2007). *Measure Theory*. Vol. 1. 2 vols. Springer. DOI: 10.1007/978-3-540-34514-5.

– (2010). *Differentiable Measures and the Malliavin Calculus*. Vol. 164. Mathematical Surveys and Monographs. American Mathematical Society. DOI: 10.1090/surv/164.

Bruckner, Andrew M. (1978). *Differentiation of real functions*. Lecture notes in mathematics 659. Berlin Heidelberg: Springer. ISBN: 978-0-387-08910-2.

Burger, Martin and Lucka, Felix (2014). 'Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators'. In: *Inverse Problems* 30.11, 114004. DOI: 10.1088/0266-5611/30/11/114004.

Christensen, Ole et al. (2021). 'On Approximate Operator Representations of Sequences in Banach Spaces'. In: *Complex Analysis and Operator Theory* 15.3, p. 54. DOI: 10.1007/s11785-021-01106-6.

Clason, Christian et al. (2019). 'Generalized modes in Bayesian inverse problems'. In: *SIAM/ASA Journal on Uncertainty Quantification* 7.2, pp. 652–684. DOI: 10.1137/18M1191804.

Cohn, Donald L. (2013). *Measure Theory*. 2nd ed. Birkhäuser Advanced Texts. Birkhäuser. DOI: 10.1007/978-1-4614-6956-8.

Cotter, S. L. et al. (2009). 'Bayesian inverse problems for functions and applications to fluid mechanics'. In: *Inverse Problems* 25.11, 115008. DOI: 10.1088/0266-5611/25/11/115008.

Dashti, Masoumeh et al. (2013). 'MAP Estimators and Their Consistency in Bayesian Nonparametric Inverse Problems'. In: *Inverse Problems* 29.9, 095017. DOI: 10.1088/0266-5611/29/9/095017.

Davey, B. A. and Priestley, H. A. (2002). *Introduction to Lattices and Order*. 2nd ed. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511809088.

Enflo, Per (1973). 'A counterexample to the approximation problem in Banach spaces'. In: *Acta Mathematica* 130, pp. 309–317. DOI: 10.1007/BF02392270.

Grushka, Ya I. (2019). 'On Monotonous Separately Continuous Functions'. In: *Applied General Topology* 20.1, p. 75. DOI: 10.4995/agt.2019.9817.

Hansson, Bengt (1968). 'Choice Structures and Preference Relations'. In: *Synthese* 18.4, pp. 443–458. ISSN: 15730964. URL: http://www.jstor.org/stable/20114617.

Hardy, G. H. and Wright, E. M. (2008). *An Introduction to the Theory of Numbers*. 6th ed. Oxford University Press. ISBN: 978-0-19-921985-8.

Heinonen, Juha (2001). *Lectures on Analysis on Metric Spaces*. Universitext. New York, NY: Springer New York. DOI: 10.1007/978-1-4613-0131-8.

Helin, Tapio and Burger, Martin (2015). 'Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems'. In: *Inverse Problems* 31.8, 085009. DOI: 10.1088/0266-5611/31/8/085009.

Kaipio, Jari and Somersalo, Erkki (2005). *Statistical and Computational Inverse Problems*. Vol. 160. Applied Mathematical Sciences. New York: Springer–Verlag. DOI: 10.1007/b138659.

Klebanov, Ilja (2022). *Personal communication*.

Lie, Han Cheng and Sullivan, T. J. (2018). 'Equivalence of weak and strong modes of measures on topological vector spaces'. In: *Inverse Problems* 34.11, 115013. DOI: 10.1088/1361-6420/aadef2.

Lifshits, M. A. (1995). *Gaussian Random Functions*. Mathematics and Its Applications 322. Dordrecht: Springer. DOI: 10.1007/978-94-015-8474-6.

Lucka, Felix (2012). 'Fast Markov chain Monte Carlo sampling for sparse Bayesian inference in high-dimensional inverse problems using L1-type priors'. In: *Inverse Problems* 28.12, 125012. DOI: 10.1088/0266-5611/28/12/125012.

Munkres, James R. (2000). *Topology*. 2nd ed. Upper Saddle River, NJ: Prentice Hall. ISBN: 0-13-181629-2.

Preiss, David et al. (2021). 'A set of positive Gaussian measure with uniformly zero density everywhere'. In: *Journal of the European Mathematical Society* 23.7, pp. 2439–2466. DOI: 10.4171/JEMS/1058.

Prékopa, András (1995). *Stochastic Programming*. Dordrecht: Springer Netherlands. DOI: 10.1007/978-94-017-3087-7.

Rudin, Walter (1973). *Functional Analysis*. 1st ed. Higher Mathematics. McGraw-Hill. ISBN: 978-0-07-054236-5.

– (1987). *Real and Complex Analysis*. 3rd ed. New York: McGraw-Hill. ISBN: 978-0-07-054234-1.

Smoryński, Craig (1991). *Logical Number Theory I*. Universitext. Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-642-75462-3.

Stein, Elias M. and Shakarchi, Rami (2009). *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*. Princeton University Press. DOI: 10.1515/9781400835560.

Stuart, A. M. (2010). 'Inverse problems: A Bayesian perspective'. In: *Acta Numerica* 19, pp. 451–559. DOI: 10.1017/S0962492910000061.

Tišer, Jaroslav (1988). 'Differentiation Theorem for Gaussian Measures on Hilbert Space'. In: *Transactions of the American Mathematical Society* 308.2, pp. 655–666. DOI: 10.2307/2001096.

Treves, François (1967). *Topological Vector Spaces, Distributions and Kernels*. 2nd ed. Vol. 25. Pure and Applied Mathematics. Burlington: Elsevier Science & Technology. ISBN: 978-0-08-087337-4.

Wacker, Philipp (2020). 'MAP estimators for nonparametric Bayesian inverse problems in Banach spaces'. In: *arXiv*. URL: http://arxiv.org/abs/2007.12760.

Williamson, Robert and Janos, Ludvik (1987). 'Constructing metrics with the Heine–Borel property'. In: *Proceedings of the American Mathematical Society* 100.3, pp. 567–573. DOI: 10.1090/S0002-9939-1987-0891165-X.

Young, W. (1910). ‘A note on monotone functions’. In: *The Quarterly Journal of Pure and Applied Mathematics (Oxford Ser.)* 41, pp. 79–87.

Appendix

A Supporting results

Supporting results for Section 4

Lemma A.1. *Let $(x_n)_{n\in\mathbb{N}} \to x$ be a convergent sequence in X. Fix $r > 0$. Then*

$$\limsup_{n\to\infty} \mathbf{1}_{B_r(x_n)} \leq \mathbf{1}_{B_r(x)}.$$

Proof. Pick any point p such that $\limsup_{n\to\infty} \mathbf{1}_{B_r(x_n)}(p) = 1$. Then there is an increasing sequence $(n_k)_{k\in\mathbb{N}}$ such that $p \in B_r(x_{n_k})$ for all k. By the triangle inequality

$$d(x,p) \leq d(x,x_{n_k}) + d(x_{n_k},p) \leq r + d(x,x_{n_k}).$$

As $(x_{n_k}) \to x$, taking the infimum over $k \in \mathbb{N}$ yields

$$d(x,p) \leq \inf_{k\in\mathbb{N}} \Big(r + d(x,x_{n_k}) \Big) = r.$$

Hence, $p \in B_r(x)$; this proves that $\limsup_{n\to\infty} \mathbf{1}_{B_r(x_n)}(p) = 1 \implies \mathbf{1}_{B_r(x)}(p) = 1$. ∎

Lemma A.2 (Isolation). *Suppose that $(x_n)_{n\in\mathbb{N}}$ is a sequence with no convergent subsequence. Then there exists $\delta > 0$ and a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ such that the balls $(B_\delta(x_{n_k}))_{k\in\mathbb{N}}$ are all disjoint.*

Proof. The subspace $Y = \{x_n : n \in \mathbb{N}\} \subseteq X$ is closed because $(x_n)_{n\in\mathbb{N}}$ has no convergent subsequences, so Y contains all of its limit points. The subspace Y cannot be compact, or else $(x_n)_{n\in\mathbb{N}}$ would have a convergent subsequence. As Y is closed but not compact, it must fail to be totally bounded.

We construct a sequence $(x_{n_k})_{k\in\mathbb{N}}$ inductively such that the balls $B_{\delta/2}(x_{n_k})$ are disjoint. Suppose that we have selected the first k terms, so that $x_{n_1}, \ldots, x_{n_k}$ all have disjoint $(\delta/2)$-balls. The balls $(B^\circ_\delta(x_{n_i}))_{i=1}^k$ cannot cover Y because Y is not totally bounded. Hence, we can pick

$$n_{k+1} = \min \left\{ n > n_k : x_n \notin \bigcup_{i=1}^{k} B^\circ_\delta(x_{n_i}) \right\} < \infty.$$

By construction, for $1 \leq i \leq k$, we have $d(x_{n_i}, x_{n_{k+1}}) \geq \delta$. The triangle inequality implies that, for any $p \in B_r(x_{n_i})$,

$$d(x_{n_{k+1}}, p) \geq d(x_{n_i}, x_{n_{k+1}}) - d(p, x_{n_i}) > \delta/2,$$

so $B_{\delta/2}(x_{n_{k+1}})$ is disjoint from the previous $\delta/2$-balls. The result follows by picking x_{n_1} arbitrarily from the sequence. ∎

Lemma A.3. *Suppose that μ is a finite measure and* $\text{supp}(\mu) = X$. *If $A \subseteq X$ satisfies $\mu(X \setminus A) = 0$, then A is dense in X.*

Proof. Let $U \subseteq X$ be a non-empty open set, and suppose that $A \cap U = \emptyset$. We obtain a contradiction: μ has full support, so $\mu(U) > 0$, but $\mu(X) + \mu(U) = \mu(A) + \mu(U) = \mu(A \sqcup U) \leq \mu(X)$. ∎

Supporting results for Section 5

Lemma A.4. *Adopting the convention that $\frac{1}{+\infty} = 0$ and $\frac{1}{0} = +\infty$, for $x, x' \in$* $\text{supp}(\mu)$,

$$\left(\limsup_{r \searrow 0} \frac{\mu(B_r(x'))}{\mu(B_r(x))} \right)^{-1} = \liminf_{r \searrow 0} \frac{\mu(B_r(x))}{\mu(B_r(x'))}.$$

Proof. Suppose that the limit superior is zero. Then, for all $\varepsilon > 0$, there exists $r^* > 0$ such that

$$\sup_{r < r^*} \frac{\mu(B_r(x'))}{\mu(B_r(x))} < \varepsilon \iff \inf_{r < r^*} \frac{\mu(B_r(x))}{\mu(B_r(x'))} > \frac{1}{\varepsilon},$$

so the limit inferior diverges to infinity. Now suppose that the limit superior is non-zero and finite. In particular, this means that

$$S_R := \sup_{r < R} \frac{\mu(B_r(x'))}{\mu(B_r(x))}$$

must be strictly positive. Hence, for all $r \in (0, R)$,

$$\frac{\mu(B_r(x))}{\mu(B_r(x'))} \geq \frac{1}{S_R},$$

and this lower bound is tight. Taking limits as $R \searrow 0$ gives the desired result. If the limit superior is infinite, there must exist a sequence $(r_n) \searrow 0$ such that

$$\frac{\mu(B_{r_n}(x'))}{\mu(B_{r_n}(x))} > n \iff \frac{\mu(B_{r_n}(x))}{\mu(B_{r_n}(x'))} < 1/n,$$

so the limit inferior of the right hand side is zero. ∎

Lemma A.5. *A point $x \in X$ is $\succeq_0$-greatest in X if and only if $x \in$* $\text{supp}(\mu)$ *and x is a greatest element of* $\text{supp}(\mu)$, *that is $x \succeq_0 x'$ for all $x' \in$* $\text{supp}(\mu)$.

Proof. Suppose that x is greatest in X. Then x must lie in $\text{supp}(\mu)$, because a point outside the support cannot be greater than points in the support, which would contradict the assumption that x is $\succeq_0$-greatest. As x is greatest in X, it is greatest in $\text{supp}(\mu)$. Conversely, suppose that $x \in \text{supp}(\mu)$ is greatest in $\text{supp}(\mu)$. Then, for $x' \notin \text{supp}(\mu)$, $x \succeq_0 x'$ by definition, so x is $\succeq_0$-greatest in X. ∎

Supporting results for Section 6

Lemma A.6. *(a) Let $(\succeq_r)_{r\in(0,r^*)}$ be a collection of preorders on X. Then*

$$\bigcap_{r\in(0,r^*)} \succeq_r \text{ is a preorder.}$$

(b) Let $(\succeq_r)_{r\in(0,r^*)}$ be a collection of preorders on X with $\succeq_r \subseteq \succeq_s$ for $r \leq s$. Then

$$\bigcup_{r\in(0,r^*)} \succeq_r \text{ is a preorder.}$$

Proof. (a) Reflexivity follows because $(x,x) \in \succeq_r$ for all $r \in (0,r^*)$. For transitivity, suppose that (x,y) and (y,z) lie in the intersection. This implies that (x,y) and (y,z) lie in $\succeq_r$ for all $r \in (0,r^*)$. As $\succeq_r$ is a preorder, $(x,z) \in \succeq_r$ for all $r \in (0,r^*)$. Hence, (x,z) lies in the intersection.

(b) Reflexivity follows as in (a). Suppose (x,y) and (y,z) lie in the union. Using the nesting property, we can find $r > 0$ such that both $(x,y) \in \succeq_r$ and $(y,z) \in \succeq_r$. As $\succeq_r$ is a preorder, (x,z) lies in $\succeq_r$ by transitivity. ∎

Supporting results for Section 7

The **Farey sequence** $\mathscr{F}_n$ is defined to be the increasing sequence of fractions in $[0,1]$ with denominator at most n. For example, $\mathscr{F}_4 = \{0, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, 1\}$. Hardy and Wright (2008) prove in Theorem 331 that

$$|\mathscr{F}_n| \sim \frac{3n^2}{\pi^2} \text{ as } n \to \infty. \tag{A.1}$$

In the following results, we enumerate the set $\mathbb{Q} \cap [0,1]$ using the sequence $(q_k)_{k\in\mathbb{N}}$, which we construct by enumerating the rationals in increasing order of denominator (ignoring rationals already counted), so that $\{q_1, \dots, q_{|\mathscr{F}_n|}\} = \mathscr{F}_n$.

Lemma A.7. *Enumerate the rationals in $[0,1] \cap \mathbb{Q}$ using the sequence $(q_k)_{k\in\mathbb{N}}$ formed by enumerating the rationals as described above. Fix $\alpha/\beta =: q \in [0,1] \cap \mathbb{Q}$. The fill distance*

$$\phi(n;q) := \inf_{\substack{k\leq n \\ q_k \neq q}} |q - q_k|$$

satisfies the following properties:

(i) $\phi(n;q)$ is decreasing; and

(ii) $\phi(n;q) \in \Omega(n^{-1/2})$ as $n \to \infty$.

Proof. (i) This follows from the definition of ϕ as an infimum.

(ii) By (A.1), there exists $n_0 \in \mathbb{N}$ and $c_1 \leq c_2$ such that, for all $n \geq n_0$, the bound $c_1 n^2 \leq |\mathscr{F}_n| \leq c_2 n^2$ holds. For $n \geq c_1 n_0^2$, this gives

$$n'(n) := \left| \mathscr{F}_{\left\lceil \left(\frac{n}{c_1}\right)^{1/2} \right\rceil} \right| \geq n.$$

The value of $\phi(n')$ can be computed directly because $\{q_1, \ldots, q_{n'}\}$ is precisely the Farey sequence $\mathscr{F}_{\lceil (n/c_1)^{1/2} \rceil}$ (enumerated in a different order). As ϕ is decreasing, we obtain that

$$\phi(n; q) \geq \phi(n'; q) \geq \frac{1}{\beta \lceil (n/c_1)^{1/2} \rceil} \geq \frac{c_1^{1/2}}{2\beta n^{1/2}} =: C_1 n^{-1/2},$$

because the distance of the rational $\frac{\alpha}{\beta}$ from a rational with denominator at most n' is bounded below by $\frac{1}{\beta n'}$. Hence, $\phi(n) \in \Omega(n^{-1/2})$. ∎

More generally, we say that a countable metric space E has a **Farey enumeration** if there is an enumeration $(q_k)_{k \in \mathbb{N}}$ of E such that, for any $q \in E$,

$$\phi(n; q) = \inf_{\substack{k \leq n \\ q_k \neq q}} d(q, q_k) \in \Omega(n^{-1/2}).$$

A Farey enumeration may not necessarily exist: an example is given by the space $E = \overline{\{2^{-n} : n \in \mathbb{N}\}}$. To prove that Farey enumerations exist in $\mathbb{Q}$, we use the **box enumeration** (Smoryński, 1991, p. 21), which is the bijection $P \colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ defined by

$$P(x, y) = 1 + \max((x-1)^2, (y-1)^2) + (x-1) + \begin{cases} 0 & x \leq y \\ x - y & y < x. \end{cases}$$

Lemma A.8. *There exists a Farey enumeration in:*

(i) $\mathbb{Q} \cap [0, \infty)$ *(using the Euclidean metric);*

(ii) $\mathbb{Q}$ *(using the Euclidean metric);*

(iii) $\mathbb{Q}^N$ *for any* $N \geq 1$ *(using any metric);*

(iv) any infinite-dimensional separable Hilbert space X *(using the metric induced by the inner product); and*

(v) any infinite-dimensional separable Banach space X *with a* p*-Riesz basis* $(\psi_k)_{k \in \mathbb{N}}$ *(using the metric induced by the norm).*

Proof. (i) Define the sequence $(\tilde{q}_k)_{k \in \mathbb{N}}$, using $(a_k, b_k) := P^{-1}(k)$, by

$$\tilde{q}_1 = 0, \ \tilde{q}_k = \frac{a_k}{b_k}.$$

This enumerates all the rationals in $\mathbb{Q} \cap [0, \infty)$, but may repeat certain elements. Let $(q_k)_{k\in\mathbb{N}}$ be the subsequence of $(\tilde{q}_k)_{k\in\mathbb{N}}$ formed by deleting previously enumerated terms. The box enumeration ensures that no rational with denominator greater than n is enumerated before all rationals in $\mathbb{Q} \cap [0, 1]$ with denominator at most n are enumerated. This means that the rationals $q_1, \ldots, q_{|\mathscr{F}_n|}$ have denominator at most n. For a fixed $q \in \mathbb{Q} \cap [0, \infty)$ with $q = \alpha/\beta$, we must have

$$\phi(|\mathscr{F}_n|; q) \geq \frac{1}{\beta n}.$$

The argument follows using the bound (A.1) exactly as in Lemma A.7.

(ii) Let $(s_k)_{k\in\mathbb{N}}$ denote the enumeration from (i). Enumerate $\mathbb{Q}$ by the sequence

$$q_1 = 0,\ q_{2k} = s_{k+1},\ q_{2k+1} = -s_{k+1}.$$

This sequence never repeats any element, and $q_1, \ldots, q_{|\mathscr{F}_n|}$ contains no rationals with denominator greater than n. The argument as in part (i) completes the proof.

(iii) We proceed by induction on N to derive an enumeration on $\mathbb{Q}^N$ where the components of the first $|\mathscr{F}_n|$ terms have denominator at most n. The base case is proved in (ii), giving an enumeration $(q_k^{(1)})_{k\in\mathbb{N}}$ of $\mathbb{Q}$. For $N > 1$, use the box enumeration P: let $(a_k, b_k) = P^{-1}(k)$ and define

$$q_k^{(N)} = \left(q_{a_k}^{(N-1)}, q_{b_k}^{(1)}\right).$$

The box enumeration ensures that, for each of the terms $q_1^{(N)}, \ldots, q_{|\mathscr{F}_n|}^{(N)}$, both a_k and b_k are at most $|\mathscr{F}_n|$. The induction hypothesis gives that the first $|\mathscr{F}_n|$ terms of $(q_k^{(N-1)})_{k\in\mathbb{N}}$ and $(q_k^{(1)})_{k\in\mathbb{N}}$ must all have denominator at most n in each component. Given a fixed $z = (\alpha_1/\beta_1, \ldots, \alpha_N/\beta_N) \in \mathbb{Q}^N$ and any $q_k^{(N)} \neq z$, we see that $q_k^{(N)}$ and z must differ in at least one component, which we denote i. Defining $\beta = \max_{1\leq j\leq N} \beta_j$, we have

$$\|q_k^{(N)} - z\|_\infty \geq |q_{k,i}^{(N)} - z_i|.$$

If $k \leq |\mathscr{F}_n|$, then $q_{k,i}^{(N)}$ is a rational distinct from z_i with denominator at most n. Hence,

$$\|q_k^{(N)} - z\|_\infty \geq \frac{1}{\beta n}.$$

Following the argument in Lemma A.7, we can verify

$$\phi_\infty^{(N)}(n; z) := \inf_{\substack{k\leq n \\ q_k^{(N)}\neq z}} \|q_k^{(N)} - z\|_\infty \in \Omega(n^{-1/2}).$$

The bound for an arbitrary metric follows using the equivalence of metrics on $\mathbb{R}^N$.

(iv) Let $(\psi_k)_{k\in\mathbb{N}}$ be an orthonormal basis of X. A countable dense subset of X is

$$E = \bigcup_{j\in\mathbb{N}} \underbrace{\left\{ \sum_{i=1}^{j} \alpha_i \psi_i : \alpha_i \in \mathbb{Q} \right\}}_{:=E_j}. \tag{A.2}$$

We enumerate E by combining enumerations for each E_j; the sets E_j are not disjoint so we must take care to ensure the enumeration we derive does not repeat terms. Given the canonical bijection $h_j\colon \mathbb{Q}^j \to E_j$ defined by $h_j((\alpha_1,\ldots,\alpha_j)) = \sum_{i=1}^{j} \alpha_i\psi_i$ and the enumeration $(q_k^{(j)})_{k\in\mathbb{N}}$ from (iii), define $(s_k^{(j)})_{k\in\mathbb{N}} = h_j(q_k^{(j)})$. As in (iii), use the box enumeration to define $(a_k, b_k) = P^{-1}(k)$ and define

$$q_k = s_{a_k}^{(b_k)}.$$

If $k \le |\mathscr{F}_n|$, then the coefficients of q_k are rationals with denominator at most n. Fix $z = (\alpha_1/\beta_1, \ldots, \alpha_N/\beta_N, 0, \ldots) \in E$ and let $\beta = \max_{1\le i\le N} \beta_i$. Then any $q_k \ne z$ must differ from z in at least one component, which we denote i. We see

$$\|q_k - z\| \ge |q_{k,i} - z_i| \ge \frac{1}{\beta n}. \tag{A.3}$$

The result follows as in (iii).

(v) We use the dense subset (A.2), where $(\psi_k)_{k\in\mathbb{N}}$ is now the given p-Riesz basis. Enumerate E using the sequence $(q_k)_{k\in\mathbb{N}}$ constructed just as in (iv). The bound (A.3) still holds for a fixed $z \in E$, using the assumption that $(\psi_k)_{k\in\mathbb{N}}$ is a p-Riesz basis, because

$$\left\| \sum_{j=1}^{i} (q_{k,j} - z_j)\psi_j \right\| \ge A \left(\sum_{j=1}^{i} |q_{k,j} - z_j| \right)^{\frac{1}{p}} \ge A|q_{k,i} - z_i| \ge \frac{A}{\beta n}. \qquad \blacksquare$$

Given a Farey enumeration $(q_k)_{k\in\mathbb{N}}$ of a countable metric space E, we can enhance the bound as follows.

Lemma A.9. *Let E be a countable metric space with Farey enumeration $(q_k)_{k\in\mathbb{N}}$. Fix $q \in E$. There exists $n_1 \in \mathbb{N}$ such that, for all $n \ge n_1$ with $n \ne m$,*

$$d(q, q_n) - 2^{-n} > 0.$$

Proof. Using that $\phi(n; q) \in \Omega(n^{-1/2})$, there exists n_0 such that, for all $n \ge n_0$,

$$c_1 n^{-1/2} \le \phi(n; q). \tag{A.4}$$

We therefore obtain the bound $d(q, q_n) - 2^{-n} \ge c_1 n^{-1/2} - 2^{-n}$. As $2^{-n} \in o(n^{-1/2})$, we can find $N \in \mathbb{N}$ such that $n \geq N \implies 2^{-n} \leq c_1 n^{-1/2}$. Take $n_1 = \max(n_0, N)$; when $\max(n_0, N) = m$, set $n_1 = m + 1$, so that $n_1 \neq m$. ■

Lemma A.10. *Let E be a countable metric space with a Farey enumeration $(q_k)_{k\in\mathbb{N}}$. Fix $q \in E$. Let n_1 be chosen as in Lemma A.9 and define, for $n \geq n_1$,*

$$\psi(n; q) := \inf_{\substack{n_1 \leq k \leq n \\ q_k \neq q}} \left(d(q, q_k) - 2^{-k} \right).$$

Then:

(i) $\psi(n_1; q) > 0$;

(ii) ψ is decreasing; and

(iii) there exists a constant $c > 0$ such that, for all $n \geq n_1$, $\psi(n; q) \geq cn^{-1/2}$.

Proof. (i) By construction, $n_1 \neq m$ and $d(q, q_{n_1}) - 2^{-n_1} > 0$ by Lemma A.9.

(ii) This is immediate from the definition of ψ.

(iii) For a fixed $n_1 \leq k \leq n$ with $q_k \neq q$,

$$\begin{aligned} d(q, q_k) - 2^{-k} &\geq \phi(k; q) - 2^{-k} \\ &\geq c_1 k^{-1/2} - 2^{-k} \qquad \text{(using that } n_1 \geq n_0\text{, so (A.4) holds)} \\ &\geq c_1 k^{-1/2} - \frac{n_1^{1/2}}{2^{n_1}} k^{-1/2} =: ck^{-1/2}. \end{aligned}$$

The last line follows because $2^{-k} \leq \frac{n_1^{1/2}}{2^{n_1}} k^{-1/2}$ always holds for $k \geq n_1$. ■

Perturbations of RCDFs

Lemma A.11. *Let $f\colon (0,1) \to [0, \infty)$ and $g\colon (0,1) \to [0, \infty)$. Then*

$$\lim_{r \searrow 0} f(r) = A \in (0, \infty) \implies \limsup_{r \searrow 0} f(r)g(r) = A \limsup_{r \searrow 0} g(r). \tag{A.5}$$

Proof. Fix $0 < \varepsilon < A$. Using (A.5), there exists $r^* > 0$ such that, for all $r \in (0, r^*)$, we have $|f(r) - A| < \varepsilon$. Therefore,

$$\lim_{R \searrow 0} \left(\sup_{r \leq \min(R, r^*)} f(r)g(r) \right) \leq \lim_{R \searrow 0} \left((A + \varepsilon) \sup_{r \leq \min(R, r^*)} g(r) \right) = (A + \varepsilon) \limsup_{r \searrow 0} g(r),$$

and similarly

$$\lim_{R \searrow 0} \left(\sup_{r \leq \min(R, r^*)} f(r)g(r) \right) \geq \lim_{R \searrow 0} \left((A - \varepsilon) \sup_{r \leq \min(R, r^*)} g(r) \right) = (A - \varepsilon) \limsup_{r \searrow 0} g(r).$$

Taking $\varepsilon \searrow 0$ proves the desired result. ∎

Lemma A.12 (Perturbation lemma). *Let $F, G, H, K \colon (0,1) \to (0,\infty)$, and suppose that*

$$\limsup_{r \searrow 0} \frac{F(r)}{G(r)} = \alpha > 1.$$

Define $\varepsilon_1(r) = H(r) - F(r)$ and $\varepsilon_2(r) = K(r) - G(r)$. Then

$$\lim_{r \searrow 0} \frac{\varepsilon_1(r)}{F(r)} = 0 \ \textit{and} \ \lim_{r \searrow 0} \frac{\varepsilon_2(r)}{G(r)} = 0 \implies \limsup_{r \searrow 0} \frac{H(r)}{K(r)} = \alpha > 1.$$

Proof. Using the definitions of H and K, we have

$$\lim_{r \searrow 0} \frac{H(r)}{F(r)} = 1 + \lim_{r \searrow 0} \frac{\varepsilon_1(r)}{F(r)} = 1 \quad \text{and} \quad \lim_{r \searrow 0} \frac{K(r)}{G(r)} = 1 + \lim_{r \searrow 0} \frac{\varepsilon_2(r)}{K(r)} = 1.$$

By Lemma A.11, we obtain

$$\limsup_{r \searrow 0} \frac{H(r)}{K(r)} = \left(\lim_{r \searrow 0} \frac{K(r)}{G(r)} \times \frac{F(r)}{H(r)} \right) \limsup_{r \searrow 0} \frac{F(r)}{G(r)} = \alpha.$$ ∎

Lemma A.13. *Suppose that $Z \colon (0, 1/2] \to (0,\infty)$ is increasing and satisfies $Z(2^{-n}) \geq 2^{n/2}$. Then $Z(r) \geq \sqrt{r/2}$.*

Proof. As Z is increasing, $Z(r) \geq Z(\lfloor \log_2(r) \rfloor) \geq 2^{\lfloor \log_2(r) \rfloor} \geq 2^{(\log_2(r)-1)/2}$. This gives the required bound. ∎

Lemma A.14. *Let X be a separable Banach space with a Farey enumeration $(q_k)_{k \in \mathbb{N}}$. Let $(\mu_k)_{k \in \mathbb{N}}$ be measures on X satisfying the properties of Theorem 7.9, and define*

$$\mu(A) = \frac{1}{Z} \sum_{k \in \mathbb{N}} \mu_k(A - q_k),$$

where $Z = \sum_{k \in \mathbb{N}} \mu_k(X)$. Then, for any $i \in \mathbb{N}$,

$$\lim_{r \searrow 0} \frac{\mu(B_r(q_i))}{\frac{1}{Z}\mu_i(B_r(0))} = 1.$$

Proof. Using the definition of μ, we have

$$\frac{\mu(B_r(q_i))}{\frac{1}{Z}\mu_i(B_r(0))} = 1 + \frac{\sum_{k \neq i} \mu_k(B_r(q_i - q_k))}{\mu_i(B_r(0))}.$$

We now show that the limit of the second term is zero. Lemma A.10 proves that, given the Farey enumeration $(q_k)_{k\in\mathbb{N}}$, there exists $n_1 \in \mathbb{N}$ such that, for all $n \geq n_1$,

$$\psi(n; q_i) := \inf_{\substack{n_1 \leq k \leq n \\ k \neq i}} \|q_i - q_k\| - 2^{-k} \geq cn^{-1/2} \text{ for some } c > 0.$$

As $\psi(n; q_i) \geq cn^{-1/2}$ for $n \geq n_1$, we see that, if $n_2(r) := \lfloor (c/r)^2 \rfloor$, then

$$n_1 \leq n \leq n_2(r) \implies \psi(n; q_i) \geq \psi(n_2(r); q_i) \geq r.$$

This means that, for $n_1 \leq n \leq n_2(r)$ and $n \neq i$, we have the distance bound $\|q_i - q_n\| \geq 2^{-n} + r$, so $\mu_n(B_r(q_n - q_i)) = 0$. Using this, we have

$$\sum_{k \neq i} \mu_k(B_r(q_k - q_i)) \leq \sum_{\substack{k=1 \\ k \neq i}}^{n_1 - 1} \mu_k(B_r(q_k - q_i)) + \sum_{\substack{k=n_2(r)+1 \\ k \neq i}}^{\infty} \mu_k(B_r(q_k - q_i)).$$

The RCDF $\mu_i(B_r(0))$ satisfies the bound $\mu_i(B_r(0)) \geq \sqrt{r/2}$ by Lemma A.13. Hence, using property (iii) of the assumptions of Theorem 7.9, we have

$$\sum_{\substack{k=1 \\ k \neq i}}^{n_1 - 1} \lim_{r \searrow 0} \frac{\mu_k(B_r(q_k - q_i))}{\mu_i(B_r(0))} \leq \sqrt{2} \sum_{\substack{k=1 \\ k \neq i}}^{n_1 - 1} \lim_{r \searrow 0} \frac{\mu_k(B_r(q_k - q_i))}{\sqrt{r}} = 0.$$

To bound the second sum, we use the truncation property from the assumptions of Theorem 7.9, which implies that $\mu_k(X) \leq \sqrt{2} \times 2^{-k/2}$. Hence,

$$\begin{aligned}
\lim_{r \searrow 0} \frac{1}{\mu_i(B_r(0))} \sum_{\substack{k=n_2(r)+1 \\ k \neq i}}^{\infty} \mu_k(B_r(q_k - q_i)) &\leq \lim_{r \searrow 0} \frac{1}{\mu_i(B_r(0))} \sum_{\substack{k=n_2(r)+1 \\ k \neq i}} \sqrt{2} \times 2^{-k/2} \\
&\leq \sqrt{2} \lim_{r \searrow 0} \frac{1}{\sqrt{r/2}} \frac{2^{-(n_2(r)+1)/2}}{1 - 2^{-1/2}},
\end{aligned}$$

where the final line bounds the sum of the geometric series and uses that $\mu_i(B_r(0)) \geq \sqrt{r/2}$. Using the definition of $n_2(r)$, we have

$$\lim_{r \searrow 0} \frac{1}{\mu_i(B_r(0))} \sum_{\substack{k=n_2(r)+1 \\ k \neq i}}^{\infty} \mu_k(B_r(q_k - q_i)) \leq \frac{2}{1 - 2^{-1/2}} \lim_{r \searrow 0} \frac{2^{-(c/r)^2/2}}{\sqrt{r}} = 0.$$

The desired limit has been proven. ∎

Gaussian measures in Hilbert spaces

Lemma A.15. *Let X be an infinite-dimensional separable real Hilbert space. There exists a centred non-degenerate Borel Gaussian measure μ_0 on X such that*

$$\mu_0(B_r(0)) \sim C \exp(-ar^{-2}) \text{ as } r \to 0$$

for some constants $a > 0$ and $C > 0$. In particular, $\mu_0(B_r(0)) \in o(\sqrt{r})$ as $r \to 0$.

Proof. Let $(\psi_k)_{k\in\mathbb{N}}$ be an orthonormal basis of X. Let $h\colon \ell^2(\mathbb{R}) \to X$ denote the canonical isometric isomorphism $h(x) = \sum_{k=1}^{\infty} x_k \psi_k$. We will construct a measure on $\ell^2(\mathbb{R})$ with the desired properties, then push it forward to X using h. Let

$$\mu = \bigotimes_{k=1}^{\infty} N(0, \sigma_k^2),$$

where $\sigma_k = k^{-2}/2$. The Gaussian measure μ is centred and non-degenerate. The construction of μ as a product measure allows us to explicitly compute the small-ball probabilities $\mu(\{x \in \ell^2(\mathbb{R}) : \|x\|_2 \le r\})$. To do this, observe that if $\xi_k \sim N(0, \sigma_k^2)$ are independent and $\xi = (\xi_1, \xi_2, \dots)$ is a random vector in $\mathbb{R}^\infty$, then

$$\|\xi\|_2 \le r \iff \underbrace{\sum_{k=1}^{\infty} |\xi_k|^2}_{:=\zeta^2} \le r^2.$$

Following Lifshits (1995, § 18, Example 2), we find

$$\mathbb{P}[\zeta^2 \le r^2] \sim C \exp\left(-(\pi/4)^2 r^{-2}\right) \text{ as } r \to 0,$$

for some $C > 0$. Using the isometric isomorphism h, we can define μ_0 by the push-forward $\mu_0(A) = \mu(h^{-1}(A))$. We see $\mu_0(X) = \mu(\ell^2(\mathbb{R})) = 1$, so μ_0 is a probability measure. The measure μ_0 is Gaussian[15] because, for any $f \in X^*$, we have $\mu_0 \circ f^{-1} = \mu \circ (f \circ h)^{-1}$; $f \circ h$ is a bounded linear functional on ℓ^2 so $\mu_0 \circ f^{-1} = \mu \circ (f \circ h)^{-1}$ is Gaussian. The measure μ_0 is centred because, for any $f \in X^*$,

$$\int_X f \, \mathrm{d}\mu_0 = \int_{\ell^2} h^{-1} \circ f \, \mathrm{d}\mu = (h^{-1} \circ f)(0)$$

as $h^{-1} \circ f$ is a bounded linear functional and μ is a centred Gaussian. The measure μ_0 is non-degenerate because μ is non-degenerate. We see that $h^{-1}(B_r(0))$ is the closed ball around 0 of radius r in the ℓ^2 norm, so the small-ball probability estimate remains. ∎

[15] A measure μ on X is Gaussian if, for any functional $f \in X^*$, the push-forward measure $\mu \circ f^{-1}$ is a Gaussian measure on $\mathbb{R}$ (Lifshits, 1995, § 8).