Recent advances concerning non-parametric MAP estimators: Order-theoretic perspectives and Γ-convergence

Birzhan Ayanbayev[1] Ilja Klebanov[2] Hefin Lambley[3] Han Cheng Lie[4] **T. J. Sullivan**[3,5]

Mathematical Foundations for Data-driven Engineering
ICMS, Bayes Centre, Edinburgh, UK
9–12 May 2023

[1]Suleyman Demirel University, KZ
[2]Freie Universität Berlin, DE
[3]**University of Warwick, UK**
[4]Universität Potsdam, DE
[5]**Alan Turing Institute, UK**

Motivation for modes

- We often wish to summarise a complicated probability measure μ on a space X by a point estimator $x^\star \in X$ that is "a point of maximum probability under μ", i.e. a **mode** of μ.
  - In the Bayesian perspective on inverse problems (e.g. Stuart, 2010), μ is the posterior measure and $x^\star$ is a **maximum a posteriori estimator**.
  - The analogous problem in molecular dynamics is finding **minimum action paths** of a path measure μ.
  - Many optimisation-based methods (claim to) calculate such points.

Motivation for modes

- We often wish to summarise a complicated probability measure μ on a space X by a point estimator $x^\star \in X$ that is "a point of maximum probability under μ", i.e. a **mode** of μ.
    - In the Bayesian perspective on inverse problems (e.g. Stuart, 2010), μ is the posterior measure and $x^\star$ is a **maximum a posteriori estimator**.
    - The analogous problem in molecular dynamics is finding **minimum action paths** of a path measure μ.
    - Many optimisation-based methods (claim to) calculate such points.
- But...
    - How do we even define modes when X is, say, a separable metric space, without relying on Lebesgue densities etc?
    - There are several candidates in the literature, with incompletely classified relationships and plenty of pathological counterexamples.
    - Are modes stable under perturbations of the problem setup (prior, likelihood, data)?
    - If modes are supposed to be "maximal" or "greatest" in some sense, can order-theoretic concepts be of any help?

Outline

1. Modes: Heuristics, strong modes, weak modes, etc.
2. Some notions from order theory
3. The positive-radius preorder
4. The small-radius limiting preorder
5. Stability of modes via Γ-convergence of OM functionals
6. Closing remarks

Modes: Heuristics, strong modes, weak modes, etc.

Modes: Heuristic definitions

- A **mode** of $\mu \in \mathcal{P}(X)$ should be a point $x^\star \in X$ with "maximum probability under μ".
- If $X = \mathbb{R}^d$ and μ has a (bounded, continuous) density ρ with respect to d-dimensional Lebesgue measure, then

$$
\begin{aligned}
x^\star \text{ is a mode} &\iff x^\star \text{ maximises } \rho \\
&\iff x^\star \text{ minimises } -\log \rho.
\end{aligned}
$$

Modes: Heuristic definitions

- A **mode** of $\mu \in \mathcal{P}(X)$ should be a point $x^\star \in X$ with "maximum probability under μ".
- If $X = \mathbb{R}^d$ and μ has a (bounded, continuous) density ρ with respect to d-dimensional Lebesgue measure, then

$$\begin{aligned} x^\star \text{ is a mode } &\iff x^\star \text{ maximises } \rho \\ &\iff x^\star \text{ minimises } -\log \rho. \end{aligned}$$

- This is all nonsense if μ has no such density, especially so if X has no Lebesgue-like uniform reference measure.
- A reasonably general approach is to examine the μ-mass of the closed metric ball $B_r(x) := \{x' \in X \mid d(x, x') \leqslant r\}$ the small-radius limit $r \to 0$.

Modes: Onsager–Machlup functionals

- Dürr and Bach (1978) proposed that a mode should be understood as a global minimiser of the **Onsager–Machlup (OM) functional** I_μ of μ:

$$\lim_{r \to 0} \frac{\mu(B_r(x))}{\mu(B_r(x'))} = \frac{\exp(-I_\mu(x))}{\exp(-I_\mu(x'))} \quad \text{for } x, x' \in X. \tag{1.1}$$

- I_μ is a "formal negative log-density" for μ, but is in general only defined for points $x, x' \in E \subsetneq X$.
- E.g. the OM functional of a Gaussian measure on a Hilbert space is finite only on the Cameron–Martin space, where it is half the square of the CM norm.
- The treatment of modes as minimisers of I_μ requires considerable care, especially since in sometimes it is not even possible to set $I_\mu := +\infty$ on $X \setminus E$: the ratio in (1.1) may oscillate and fail to converge as $r \to 0$. (We return to this relationship towards the end.)

Modes: Strong modes

Definition 1 (After Dashti et al., 2013)

A point $x^\star \in X$ is a **strong mode** for μ if

$$\lim_{r \to 0} \frac{\mu(B_r(x^\star))}{M_r} = 1, \qquad (1.2)$$
$$M_r := \sup_{x \in X} \mu(B_r(x)).$$

(Separability of X ensures that $\mathrm{supp}(\mu) \neq \varnothing$ and $M_r > 0$.)

- Dashti et al. (2013) established the existence of strong MAP estimators for "nice" posteriors with respect to Gaussian priors on Hilbert spaces X, and showed that they are OM minimisers.
- The proof strategy relied on finding $x_r^\star$ such that $\mu(B_r(x_r^\star)) = M_r$ and then extracting convergent subsequences as $r \to 0$. This approach is surpisingly difficult to generalise even to Gaussian measures on Banach spaces! (Klebanov and Wacker, 2023; Lambley, 2023)

Modes: Weak modes

Any strong mode must lie in $\operatorname{supp}(\mu)$, and the ratio in (1.2) is at most 1 for every choice of $x^\star \in X$, so

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

Modes: Weak modes

Any strong mode must lie in $\operatorname{supp}(\mu)$, and the ratio in (1.2) is at most 1 for every choice of $x^\star \in X$, so

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

Definition 2 (After Helin and Burger, 2015)

A point $x^\star \in X$ is a **(global) weak mode** for μ if

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

Suggestive order-theoretic notation: $x^\star$ is weak mode $\iff$ $x' \preccurlyeq_0 x^\star$ for all $x' \in X$.

- Every strong mode is a weak mode.
- The converse is false in general, although some sufficient conditions for "weak $\implies$ strong" are known. (Lie and Sullivan, 2018; Lambley, 2023)

Some notions from order theory

Preorders

Definition 3

A **preorder** on a set X is a binary relation $\preccurlyeq$ satisfying

- **reflexivity**: for all $x \in X$, $x \preccurlyeq x$; and
- **transitivity**: for all $x, y, z \in X$, if $x \preccurlyeq y$ and $y \preccurlyeq z$, then $x \preccurlyeq z$.

Preorders

Definition 3

A **preorder** on a set X is a binary relation $\preccurlyeq$ satisfying

- **reflexivity**: for all $x \in X$, $x \preccurlyeq x$; and
- **transitivity**: for all $x, y, z \in X$, if $x \preccurlyeq y$ and $y \preccurlyeq z$, then $x \preccurlyeq z$.

- If both of $x \preccurlyeq y$ and $y \preccurlyeq x$ hold, then x and y are called **equivalent** and we write $x \asymp y$.
- If $x \asymp y \implies x = y$, then $\preccurlyeq$ is a **partial order** (but we won't see any partial orders today).
- If at least one of $x \preccurlyeq y$ and $y \preccurlyeq x$ hold, then x and y are called **comparable**; otherwise, they are called **incomparable** and we write $x \parallel y$.
- If every two elements are comparable then $\preccurlyeq$ is a **total preorder** and X is a **chain**.
- A collection of mutually incomparable elements is called an **antichain**.

Greatest and maximal elements

Definition 4

Let X be a set equipped with a preorder $\precsim$.

- $g \in X$ is a **greatest element** if, for every $x \in X$, $g \succsim x$.
- $m \in X$ is a **maximal element** if, whenever $x \in X$ is such that $m \precsim x$, it follows that $m \succsim x$ (and hence that $m \asymp x$).

Greatest and maximal elements

Definition 4

Let X be a set equipped with a preorder $\preccurlyeq$.

- $g \in X$ is a **greatest element** if, for every $x \in X$, $g \succcurlyeq x$.
- $m \in X$ is a **maximal element** if, whenever $x \in X$ is such that $m \preccurlyeq x$, it follows that $m \succcurlyeq x$ (and hence that $m \asymp x$).

- Note that any greatest element is also a maximal element, but it must additionally be comparable to (and dominate) every element of X.
- On the other hand, a maximal element is only required to dominate those elements of X with which it is comparable — potentially a very small subset of X.

$$m \text{ is maximal} \iff \nexists x \in X \text{ s.t. } x \succ m.$$

Upward closures etc.

- We write
$$\uparrow Y := \{x \in X \mid x \succcurlyeq y \text{ for some } y \in Y\}$$
for the **upward closure** of $Y \subseteq X$, and further write, for $y \in X$,
$$\uparrow y := \uparrow\{y\} = \{x \in X \mid x \succcurlyeq y\},$$
so that $\uparrow Y = \bigcup_{y \in Y} \uparrow y$.
- Finally, since many of the preorders we consider will be parametrised by radius $r \geqslant 0$, we will write $\preccurlyeq_r$ for the preorder, $\|_r$ for the induced relation of incomparability, $\uparrow_r Y$ for the upward closure of Y with respect to $\preccurlyeq_r$, etc.

The positive-radius preorder

The positive-radius preorder

Definition 5

Given $r > 0$, define the **positive-radius preorder** $\preccurlyeq_r$ on X by

$$x \preccurlyeq_r y \iff \mu(B_r(x)) \leqslant \mu(B_r(y)).$$

Write $\mathfrak{M}_r$ for the set of $\preccurlyeq_r$-greatest elements, which we call **radius-r modes**.

The positive-radius preorder

Definition 5

Given $r > 0$, define the **positive-radius preorder** $\preccurlyeq_r$ on X by

$$x \preccurlyeq_r y \iff \mu(B_r(x)) \leqslant \mu(B_r(y)).$$

Write $\mathfrak{M}_r$ for the set of $\preccurlyeq_r$-greatest elements, which we call **radius-r modes**.

- In fact, since $\preccurlyeq_r$ is total,

$$\begin{aligned} x^\star \text{ is a radius-}r\text{ mode} &\iff x_r^\star \text{ is greatest} \\ &\iff x_r^\star \text{ is maximal} \\ &\iff \mu(B_r(x_r^\star)) = M_r \\ &\iff \left[(x_n)_{n\in\mathbb{N}} \text{ s.t. } \mu(B_r(x_n)) \to M_r \implies x^\star \in \bigcap_{n\in\mathbb{N}} \uparrow_r x_n \right]. \end{aligned}$$

- Thus, the natural question is: When does $\preccurlyeq_r$ have greatest/maximal elements?

The positive-radius preorder

Existence of radius-r modes

Existence of radius-r modes: An easy case

Proposition 6 (Existence of radius-r modes in compact spaces)

Let X be a compact metric space, let $\mu \in \mathcal{P}(X)$, and let $r > 0$. Then $\preccurlyeq_r$ has at least one radius-r mode $x_r^\star \in X$, i.e. $\mathfrak{M}_r \neq \varnothing$.

Proof. This is a special case of more general results later, but a self-contained proof is given by observing that the map $\mu(B_r(\cdot))\colon X \to [0, 1]$ is upper semicontinuous and hence has at least one global maximiser $x_r^\star$ in the compact space X. ■

Existence of radius-r modes: An easy case

Proposition 6 (Existence of radius-r modes in compact spaces)

Let X be a compact metric space, let $\mu \in \mathcal{P}(X)$, and let $r > 0$. Then $\preccurlyeq_r$ has at least one radius-r mode $x_r^\star \in X$, i.e. $\mathfrak{M}_r \neq \varnothing$.

Proof. This is a special case of more general results later, but a self-contained proof is given by observing that the map $\mu(B_r(\,\cdot\,)) \colon X \to [0, 1]$ is upper semicontinuous and hence has at least one global maximiser $x_r^\star$ in the compact space X. ∎

- However, the earlier description of radius-r modes as points of $\bigcap_{n \in \mathbb{N}} \uparrow_r x_n$ when $\mu(B_r(x_n)) \to M_r$ suggests that we should view the existence of radius-r modes as an intersection problem, especially since it's easy to show that $\uparrow_r x_n$ is always closed, bounded, and "nearly compact".
- A general setting for discussing such questions is offered by measures of non-compactness and the generalised intersection theorem.

Measures of non-compactness

- As the name suggests, a measure of non-compactness is a set function that gives a numerical measure of "how non-compact" a (closed) subset of a metric space is.

Definition 7

Given $A \subseteq X$, its **separation** (or **Istrățescu**) **measure of non-compactness** is

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

Measures of non-compactness

- As the name suggests, a measure of non-compactness is a set function that gives a numerical measure of “how non-compact” a (closed) subset of a metric space is.

Definition 7

Given $A \subseteq X$, its **separation** (or **Istrățescu**) **measure of non-compactness** is

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

- $\gamma(A)$ is an increasing function with respect to inclusion of sets, is finite precisely when A is bounded, and is zero precisely when A is pre-compact.
- The function γ is bi-Lipschitz equivalent with several other measures of non-compactness such as the set (or Kuratowski) measure of non-compactness and the ball (or Hausdorff) measure of non-compactness.
- $\uparrow_r x_n$ is always closed, bounded, and has $\gamma(\uparrow_r x_n) \leqslant 2r$.

Theorem 8 (Generalised intersection theorem)

Let $(A_n)_{n\in\mathbb{N}}$ be a decreasingly nested sequence of non-empty, closed subsets of a topological space X and let $A := \bigcap_{n\in\mathbb{N}} A_n$.

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

Radius-r modes and intersections of upward closures

Theorem 9 (Existence of radius-r modes)

Let X be a separable metric space, $\mu \in \mathcal{P}(X)$, and $r > 0$.

- *Suppose that X has the Heine–Borel property, i.e. that every closed and bounded subset of X is compact. Then $\mathfrak{M}_r$ is non-empty and compact.*
- $\dots$

Proof. Take $(x_n)_n$ with $\mu(B_r(x_n)) \nearrow M_r$. The sets $\uparrow_r x_n$ are non-empty and decreasingly nested. They are also closed and bounded, and hence compact. Their intersection is exactly $\mathfrak{M}_r$. Cantor's intersection theorem now yields that $\mathfrak{M}_r$ is non-empty and compact. ■

Theorem 9 (Existence of radius-r modes)

Let X be a separable metric space, $\mu \in \mathcal{P}(X)$, and $r > 0$.

- ...
- *Suppose that X is complete and that μ is a doubling measure, i.e. there exists a constant $C > 0$ such that*

  $$\text{for all } x \in X \text{ and } r > 0, \quad \mu(B_{2r}(x)) \leqslant C\mu(B_r(x)).$$

  Then $\mathfrak{M}_r$ is non-empty and compact.
- ...

Proof. Every complete metric space with a doubling measure is Heine–Borel (Björn and Björn, 2011, Proposition 3.1). ■

Theorem 9 (Existence of radius-r modes)

Let X be a separable metric space, $\mu \in \mathcal{P}(X)$, and $r > 0$.

- ...
- *Suppose that X is complete and there exists a point $o \in X$ and a function $f \colon (0, \infty)^2 \to (0, \infty)$ such that*

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

  Then $\mathfrak{M}_r$ is non-empty and compact.
- ...

Proof. This condition also implies the Heine–Borel property. ∎

Radius-r modes and intersections of upward closures

Theorem 9 (Existence of radius-r modes)

Let X be a separable metric space, $\mu \in \mathcal{P}(X)$, and $r > 0$.

- $\dots$
- *Suppose that X is complete and that there exists $(x_n)_{n \in \mathbb{N}}$ with $\mu(B_r(x_n)) \nearrow M_r$ and $\gamma(\uparrow_r x_n) \to 0$ as $n \to \infty$. Then $\mathfrak{M}_r$ is non-empty and compact.*
- $\dots$

Proof. Appeal to Kuratowksi's intersection theorem. ∎

Theorem 9 (Existence of radius-r modes)

Let X be a separable metric space, $\mu \in \mathcal{P}(X)$, and $r > 0$.

- $\dots$
- *Suppose that X is complete and that there exists $(x_n)_{n \in \mathbb{N}}$ with $\mu(B_r(x_n)) \nearrow M_r$ and $\operatorname{diam}(\uparrow_r x_n) \to 0$ as $n \to \infty$. Then $\mathfrak{M}_r$ is a singleton.*
- $\dots$

Proof. Appeal to Cantor's intersection theorem for complete metric spaces. ∎

Radius-r modes and intersections of upward closures

Theorem 9 (Existence of radius-r modes)

Let X be a separable metric space, $\mu \in \mathcal{P}(X)$, and $r > 0$.

- ...
- *Suppose that X is a Banach space and that there exists $(x_n)_{n \in \mathbb{N}}$ with $\mu(B_r(x_n)) \nearrow M_r$, and that $\uparrow_r x_n$ is weakly compact for all sufficiently large n. Then $\mathfrak{M}_r$ is non-empty and weakly compact.*
- ...

Proof. Very similar argument, but using the weak topology of X and Cantor's theorem for general Hausdorff topological spaces. ∎

Radius-r modes and intersections of upward closures

Theorem 9 (Existence of radius-r modes)

Let X be a separable metric space, $\mu \in \mathcal{P}(X)$, and $r > 0$.

- *...*
- *Suppose that X is a reflexive Banach space and that there exists $(x_n)_{n \in \mathbb{N}}$ with $\mu(B_r(x_n)) \nearrow M_r$ and that $\uparrow_r x_n$ is convex for all sufficiently large n. Then $\mathfrak{M}_r$ is non-empty, weakly compact, and convex.*

Proof. Each closed, bounded, and convex subset of X is weakly compact. Now use the previous part. ∎

So does *every* μ have a radius-r mode?

The positive-radius preorder

Non-existence of radius-r modes

Measures with no radius-r modes

- Equip $X = \mathbb{N}$ with the following variant of the discrete metric:

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

  In the space $(\mathbb{N}, \Delta)$, distinct points are a unit distance apart, with the exception of each odd number and its successor, which are doubly spaced.

- Equip this space with the measure $\mu := \sum_{k \in \mathbb{N}} 2^{-k} \delta_k \in \mathcal{P}(\mathbb{N})$, where δ_k is the unit Dirac measure centred at $k \in \mathbb{N}$.
- It turns out that $(\mathbb{N}, \Delta, \mu)$ has no radius-1 modes.

Measures with no radius-r modes

Theorem 10

For $1 \leqslant r < 2$, $(\mathbb{N}, \Delta, \mu)$ *has no* $x \in \mathbb{N}$ *with* $\mu(B_r(x)) = M_r = 1$.

Measures with no radius-r modes

Theorem 10

For $1 \leqslant r < 2$, $(\mathbb{N}, \Delta, \mu)$ has no $x \in \mathbb{N}$ with $\mu(B_r(x)) = M_r = 1$.

Proof. For $k \in \mathbb{N}$,

$$\mu(B_r(2k-1)) = \mu(X \setminus \{2k\}) = 1 - 2^{-2k}, \tag{3.1}$$

$$\mu(B_r(2k)) = \mu(X \setminus \{2k-1\}) = 1 - 2^{-(2k-1)}. \tag{3.2}$$

Both (3.1) and (3.2) show that $M_r = 1$; (3.1) shows that no odd number is a radius-r mode; (3.2) shows that no even number is a radius-r mode. Thus, μ has no radius-r mode. ■

Measures with no radius-r modes

Theorem 10

For $1 \leqslant r < 2$, $(\mathbb{N}, \Delta, \mu)$ has no $x \in \mathbb{N}$ with $\mu(B_r(x)) = M_r = 1$.

Proof. For $k \in \mathbb{N}$,

$$\mu(B_r(2k-1)) = \mu(X \setminus \{2k\}) = 1 - 2^{-2k}, \tag{3.1}$$

$$\mu(B_r(2k)) = \mu(X \setminus \{2k-1\}) = 1 - 2^{-(2k-1)}. \tag{3.2}$$

Both (3.1) and (3.2) show that $M_r = 1$; (3.1) shows that no odd number is a radius-r mode; (3.2) shows that no even number is a radius-r mode. Thus, μ has no radius-r mode. ∎

We can also form another space (X, d, μ), a countable union of "spread-out" scaled copies of the previous example such that. . .

Measures with no radius-r modes

Theorem 10

For $1 \leqslant r < 2$, $(\mathbb{N}, \Delta, \mu)$ has no $x \in \mathbb{N}$ with $\mu(B_r(x)) = M_r = 1$.

Proof. For $k \in \mathbb{N}$,

$$\mu(B_r(2k-1)) = \mu(X \setminus \{2k\}) = 1 - 2^{-2k}, \tag{3.1}$$

$$\mu(B_r(2k)) = \mu(X \setminus \{2k-1\}) = 1 - 2^{-(2k-1)}. \tag{3.2}$$

Both (3.1) and (3.2) show that $M_r = 1$; (3.1) shows that no odd number is a radius-r mode; (3.2) shows that no even number is a radius-r mode. Thus, μ has no radius-r mode. ∎

We can also form another space (X, d, μ), a countable union of "spread-out" scaled copies of the previous example such that...

Theorem 11

The space (X, d, μ) is complete, separable, non-atomic, and has no radius-r mode for any $0 < r < 1$.

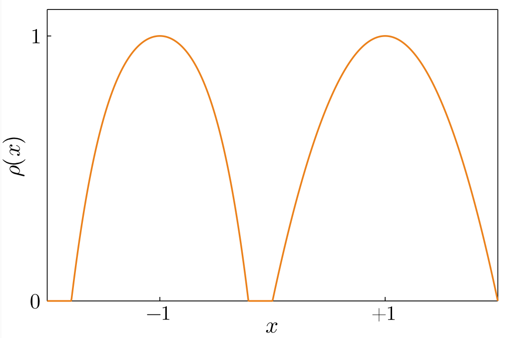
▶ See the construction

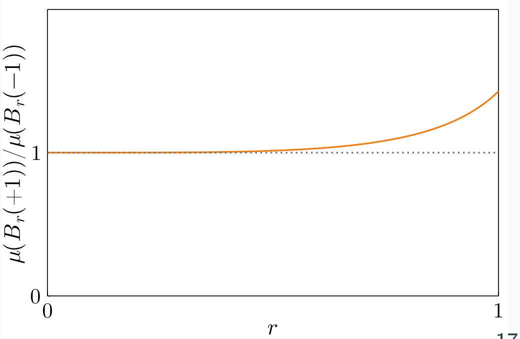
The positive-radius preorder

Limits of radius-r modes

Limits of radius-r modes

- Taking limits of radius-r modes is not straightforward: the limit need not be a strong or weak mode, and not every mode can be represented as the limit of radius-r modes.
- For example, the following measure has a bounded and continuous Lebesgue density and a mode that cannot be represented as the limit of radius-r modes. The problem here is that the points -1 and $+1$ have asymptotically equivalent mass, but the point $+1$ has slightly more mass in each ball $B_r(+1)$; as a result, $+1$ "hides" the other mode -1.





Asymptotic maximising families

Definition 12 (Klebanov and Wacker, 2023)

A net $(x_r)_{r>0} \subseteq X$ is an **asymptotic maximising family** (AMF) if there exists a positive function ε with $\lim_{r \to 0} \varepsilon(r) = 0$ and

$$\frac{\mu(B_r(x_r))}{M_r} \geqslant 1 - \varepsilon(r) \quad \text{for all } r > 0. \tag{3.3}$$

Lemma 13

Let X be a separable metric space and let $\mu \in \mathcal{P}(X)$.

- *If $(x^\star)_{r>0}$ is an AMF with $x^\star \in X$ fixed, then $x^\star$ is a strong mode.*
- *If $x^\star$ is a radius-r mode for all small enough $r > 0$, then $x^\star$ is a strong mode.*
- *If the AMF $(x_r^\star)_{r>0}$ converges to $x^\star$, then $x^\star$ is a generalised strong mode.*

Asymptotic maximising families

Theorem 14 (AMFs and strong modes)

Let X be a complete and separable metric space and let $\mu \in \mathcal{P}(X)$. Let $(x_r)_{r>0}$ be any AMF satisfying (3.3) and let $I := \bigcap_{r>0} \uparrow_r x_r$. Then

- $I \subseteq \operatorname{supp}(\mu)$;
- *every $x^\star \in I$ is a strong (and hence weak and generalised strong) mode for μ;*
- *and if also*

$$0 < r \leqslant s \implies \uparrow_r x_r \subseteq \uparrow_s x_s, \tag{3.4}$$

  then I is non-empty and compact.

Proof. Kuratowski's intersection theorem again... ■

- Hypothesis (3.4) is strong, and fails for the oscillatory examples that we're about to see.

The small-radius limiting preorder

The small-radius limiting preorder

Definition 15 (Small-radius limiting preorder)

Define the **small-radius limiting preorder** $\preccurlyeq_0$ on X by

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

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

The small-radius limiting preorder

Definition 15 (Small-radius limiting preorder)

Define the **small-radius limiting preorder** $\preccurlyeq_0$ on X by

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

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

$$\begin{aligned} x^\star \text{ is a weak mode } &\iff x^\star \text{ is a } \preccurlyeq_0\text{-greatest element} \\ &\iff x^\star \text{ is a globally-comparable } \preccurlyeq_0\text{-maximal element.} \end{aligned}$$

The small-radius limiting preorder

Definition 15 (Small-radius limiting preorder)

Define the **small-radius limiting preorder** $\preccurlyeq_0$ on X by

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

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

$$\begin{aligned} x^\star \text{ is a weak mode} &\iff x^\star \text{ is a } \preccurlyeq_0\text{-greatest element} \\ &\iff x^\star \text{ is a globally-comparable } \preccurlyeq_0\text{-maximal element.} \end{aligned}$$

- $\preccurlyeq_0$ will turn out to be a non-total preorder; "maximal" and "greatest" no longer coincide.
- One could also imagine taking set-theoretic limits of the sets $\preccurlyeq_r \subseteq X \times X$ as $r \to 0$, but this kind of procedure yields relations that are not even transitive!

The small-radius limiting preorder

Incomparability

Incomparability under $\preccurlyeq_0$

Crucially, the preorder $\preccurlyeq_0$ is not necessarily total and incomparability of points $x, x' \in X$ can be characterised as follows:

Lemma 16 (Incomparability in the limiting preorder)

For $x, x' \in X$,

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

Even measures with bounded, continuous Lebesgue densities can admit such incomparable points, and this is particularly worrisome if the incomparable elements are also maximal elements.

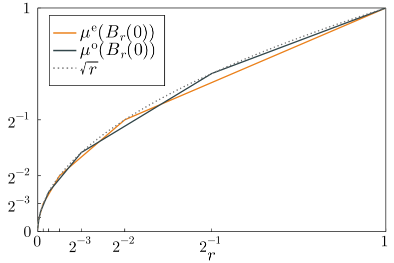
Incomparability: A prototypical example

- A prototypical example of incomparability for absolutely continuous $\mu \in \mathcal{P}(\mathbb{R})$ is given by considering

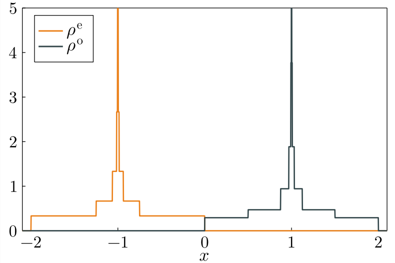
$$\mu = \frac{1}{2}\mu^{\mathrm{e}}(\cdot + 1) + \frac{1}{2}\mu^{\mathrm{e}}(\cdot - 1)$$

where $\mu^{\mathrm{e}}(B_r(0))$ and $\mu^{\mathrm{o}}(B_r(0))$ linearly interpolate $r \mapsto \sqrt{r}$ at the knots a^{-n} with n even and odd respectively, and $a > 1$ is a parameter.

- It is neither difficult nor fun to write down explicit formulæ for the densities ρ^{e} and ρ^{o}.



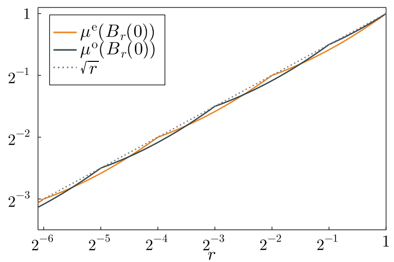
(a) $\mu^{\mathrm{e}}(B_r(0))$ and $\mu^{\mathrm{o}}(B_r(0))$ interpolate $r \mapsto \sqrt{r}$ at the knots a^{-n} with n even/odd.



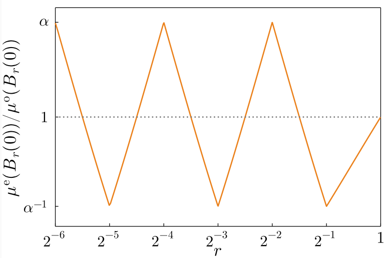
(b) The PDFs $\rho^{\mathrm{e}}(\cdot + 1)$ and $\rho^{\mathrm{o}}(\cdot - 1)$ have singularities like $|\cdot|^{-1/2}$ at -1 and $+1$ respectively.

Incomparability: A prototypical example

- A prototypical example of incomparability for absolutely continuous $\mu \in \mathcal{P}(\mathbb{R})$ is given by considering
$$\mu = \frac{1}{2}\mu^{\mathrm{e}}(\cdot + 1) + \frac{1}{2}\mu^{\mathrm{e}}(\cdot - 1)$$
where $\mu^{\mathrm{e}}(B_r(0))$ and $\mu^{\mathrm{o}}(B_r(0))$ linearly interpolate $r \mapsto \sqrt{r}$ at the knots a^{-n} with n even and odd respectively, and $a > 1$ is a parameter.
- It is neither difficult nor fun to write down explicit formulæ for the densities ρ^{e} and ρ^{o}.



(c) $\mu^{\mathrm{e}}(B_r(0))$ and $\mu^{\mathrm{o}}(B_r(0))$ on a logarithmic scale.



(d) The ratio $\mu(B_r(-1))/\mu(B_r(+1))$ oscillates between α and α^{-1} as $r \to 0$.

Incomparability: A prototypical example

- A prototypical example of incomparability for absolutely continuous $\mu \in \mathcal{P}(\mathbb{R})$ is given by considering
$$\mu = \frac{1}{2}\mu^{\mathrm{e}}(\cdot + 1) + \frac{1}{2}\mu^{\mathrm{e}}(\cdot - 1)$$
where $\mu^{\mathrm{e}}(B_r(0))$ and $\mu^{\mathrm{o}}(B_r(0))$ linearly interpolate $r \mapsto \sqrt{r}$ at the knots a^{-n} with n even and odd respectively, and $a > 1$ is a parameter.
- It is neither difficult nor fun to write down explicit formulæ for the densities ρ^{e} and ρ^{o}.
- For this μ, $\preccurlyeq_0$ has two incomparable maximal elements at ± 1; they are not greatest, nor are they weak, strong, or generalised strong modes.
- One could try to extend $\preccurlyeq_0$ into a total order using the extension theorems of Szpilrajn, Arrow, and Hansson (Szpilrajn, 1930; Hansson, 1968), but such an extension is not uniquely determined: which of ± 1 is greatest becomes a matter of personal choice!

A family of incomparable singularities

- We can define a large family of such measures $\mu_{k,m}$ — parametrised by $k \in \mathbb{N}$, $a > 0$, $m > 0$ — each having a Lebesgue density $\rho_{k,m}$ supported on $[-r(m), r(m)]$ and a singularity "like" $|x|^{-1/2}$ at 0, and with

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

- If $k, k' \geqslant 2$ are coprime and $m, m' > 0$ are arbitrary, then

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

- The idea is now to consider a sum of such measures, with centres in a countable dense subset of $[0, 1]$, masses summing to 1, and cutoff radii $r(m)$ decaying sufficiently rapidly.

Absolutely continuous measures with large antichains

Theorem 17 (Lambley and Sullivan, 2022, Theorem 5.11)

Let $\mu \in \mathcal{P}(\mathbb{R})$ have the Lebesgue density

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

where $\rho_{k,m}$ is the density constructed above with parameter $a = 2$; $k(\ell, i)$ is the $(2^{\ell-1} + i - 1)^{th}$ prime; $m(\ell) := 2^{-2\ell+1}$; and

$$q_{\ell,i} := (2i-1)2^{-\ell} \in D_\ell := \{(2i-1)2^{-\ell} \mid 1 \leqslant i \leqslant 2^{\ell-1}\}.$$

Then the set $D := \biguplus_{\ell \in \mathbb{N}} D_\ell$ of dyadic rationals consists of $\preccurlyeq_0$-maximal elements, is a $\preccurlyeq_0$-antichain, and is dense in $[0, 1]$; there are no $\preccurlyeq_0$-greatest elements.

Absolutely continuous measures with large antichains

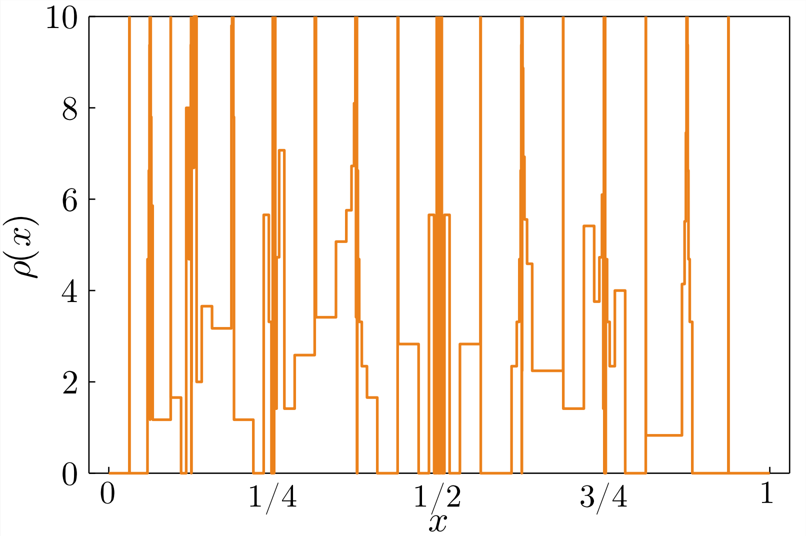


Figure 4.1: Approximation of a density $\rho := \frac{1}{Z} \sum_{k\in\mathbb{N}} \rho_{p_k,R_k}(\cdot - q_k)$ with a countable dense antichain in $[0, 1]$. The plot includes the first 31 densities, i.e. all centres q_k in $\bigcup_{k=1}^{5} D_k$.

Absolutely continuous measures with large antichains

- We can perform a similar construction in which the countable dense antichain D is $\mathbb{Q} \cap [0, 1]$, enumerated using the Farey sequence.
- Similarly, we can construct a measure μ on a separable Hilbert space X that has a density with respect to a non-degenerate Gaussian $\mu_0 \in \mathcal{P}(X)$ and also has a topologically dense antichain.
- In these two cases, we cannot quite prove that all elements of D are $\preccurlyeq_0$-maximal, although we have strong evidence that they are.
- We do know that, for any $q \in D$ and any point $x \in [0, 1]$ where the Lebesgue differentiation theorem holds for μ, that $q \succ_0 x$.
- This leaves a Lebesgue-null set of "awkward" irrationals (Liouville numbers) y that, in principle, *might* satisfy $y \succ_0 q$ for all $q \in D$ — this possibility still needs to be ruled out and is an open problem.

Stability of modes via Γ-convergence of OM functionals

Well-posedness of (Bayesian) inverse problems

- In an **inverse problem** we recover a parameter $u \in X$ from observed data $y \in Y$. Such problems are usually ill posed: the recovered $u^y \in X$ depends sensitively on y, and this sensitivity is worse the "nicer" the forward map $u \mapsto y$ is (and this is why inverse problems need to be regularised).
- In a **Bayesian inverse problem (BIP)**, the recovery of u from y is expressed in the form of a posterior probability distribution $\mu^y \in \mathcal{P}(X)$.
- BIPs *are* well posed (Stuart, 2010; ...; Sprungk, 2020). The posterior μ^y is a stable function of the problem setup — the prior distribution $\mu_0 \in \mathcal{P}(X)$, the observed data $y \in Y$, and the likelihood model $\ell \colon X \to \mathcal{P}(Y)$ — with respect to e.g. the Hellinger, Kullback–Leibler, or Wasserstein distances on $\mathcal{P}(X)$, e.g.

$$\mathrm{KL}(\mu^y \| \mu^{y+\delta y}) \lesssim \|\delta y\| \quad \text{(for fixed } \ell \text{ and } \mu_0\text{)}.$$

- Are the MAP estimators, the "most likely points under μ^y", also stable?

Similarity of modes $\not\Rightarrow$ similarity of measures

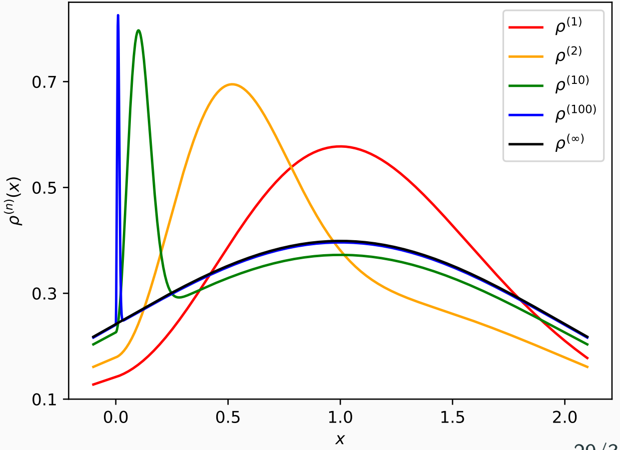
- Unfortunately, closeness of probability measures and closeness of their modes are "orthogonal" questions, even using a strong distance on $\mathcal{P}(X)$ like Kullback–Leibler.
- This is the case even for probability measures on $\mathbb{R}$ with continuous Lebesgue densities, for which a mode is easily defined as a maximiser of the density.
- Obviously, two measures can have very similar (or even the same) modes and yet be very different as measures, e.g. $\mathcal{N}(0,1)$ and $\mathcal{N}(0,10^6)$ or $\mu(E) = \int_E \max(0, 1 - |x|) \, \mathrm{d}x$!
- Perhaps if a sequence of probability measures converges "strongly enough", then their modes will also converge?

Similarity of measures $\not\Rightarrow$ similarity of modes

Consider, for $n \in \mathbb{N}$, $\mu^{(n)} \in \mathcal{P}(\mathbb{R})$ with Lebesgue density

$$\rho^{(n)}(x) := \frac{\exp\left(-\frac{1}{2}(x-1)^2\right) + \mathbb{1}[x \geqslant 0] 4n^2x^2 \exp(-n^2x^2)}{\sqrt{2\pi} + \sqrt{\pi}/n}$$

- The $\rho^{(n)}$ are smooth and uniformly bounded, with unique maximisers at $x_n^\star \approx \frac{1}{n}$.
- Pointwise, $\rho^{(n)} \to \rho^{(\infty)}$, the density of $\mu^{(\infty)} = \mathcal{N}(1,1)$.
- $\mathrm{KL}(\mu^{(\infty)} \| \mu^{(n)}) \approx \frac{1}{n}$.
- But the maximiser of $\rho^{(\infty)}$ is at $x_\infty^\star = 1 \neq 0 = \lim_{n\to\infty} x_n^\star$.



The M property

We formalise a property used implicitly in e.g. Dashti et al. (2013):

Definition 18

We will say that **property** $M(\mu, E)$ holds for $\mu \in \mathcal{P}(X)$ and $E \subseteq X$ if, for some $e \in E$,

$$x \in X \setminus E \implies \lim_{r \to 0} \frac{\mu(B_r(x))}{\mu(B_r(e))} = 0.$$

- Property $M(\mu, E)$ always holds if E is large enough.
- Property $M(\mu, E)$ does not say that $\mu(X \setminus E) = 0$!
- Property $M(\mu, E)$ does say that points outside E cannot qualify as modes of μ.
- Standard example: a Gaussian measure $\mu = \mathcal{N}(0, C)$ on an infinite-dimensional Hilbert space X with infinite-dimensional **Cameron–Martin space** $H(\mu) :=$ **ran** $C^{1/2}$ satisfies property $M(\mu, H(\mu))$ and yet has $\mu(H(\mu)) = 0$.

OM functionals, property M, and modes

- Property $M(\mu, E)$ says that $+\infty$ is a sensible value for the OM functional outside E — as opposed to the mass ratio oscillating so badly that **the limit does not exist** at all.
- This being the case, the next result should be no surprise:

OM functionals, property M, and modes

- Property $M(\mu, E)$ says that $+\infty$ is a sensible value for the OM functional outside E — as opposed to the mass ratio oscillating so badly that **the limit does not exist** at all.
- This being the case, the next result should be no surprise:

Lemma 19 (Ayanbayev et al., 2022a, Prop. 4.1)

Let μ have OM functional $I_\mu \colon E \to \mathbb{R}$ and satisfy property $M(\mu, E)$. Set $I_\mu(x) := +\infty$ for $x \notin E$. Then the weak modes of μ are precisely the minimisers of $I_\mu \colon X \to \mathbb{R} \cup \{+\infty\}$.

- This result gives a rigorous meaning to the claim of Dürr and Bach (1978) that OM minimisers should be seen as "most likely points" in the sense of weak modes.
- The above can be generalised, replacing property $M(\mu, E)$ with the condition that E be **essentially total** with respect to μ (Lambley and Sullivan, 2022).
- Unfortunately, it is **not** generally true that strong modes are OM-minimisers, even when such minimisers exist and property M holds!

Theorem 20 (Γ-convergence and equicoercivity imply convergence of modes; Ayanbayev et al., 2022a, Theorem 4.2)

For $n \in \mathbb{N} \cup \{\infty\}$, let $\mu^{(n)} \in \mathcal{P}(X)$ have OM functional $I_{\mu^{(n)}} \colon E^{(n)} \to \mathbb{R}$ and satisfy property $M(\mu^{(n)}, E^{(n)})$; extend each $I_{\mu^{(n)}}$ to take the value $+\infty$ on $X \setminus E^{(n)}$. Suppose that $(I_{\mu^{(n)}})_{n\in\mathbb{N}}$ is equicoercive and Γ-converges to $I_{\mu^{(\infty)}}$. Then, if $u^{(n)}$ is a weak mode of $\mu^{(n)}$, $n \in \mathbb{N}$, every convergent subsequence of $(u^{(n)})_{n\in\mathbb{N}}$ has as its limit a weak mode of $\mu^{(\infty)}$.

Theorem 20 (Γ-convergence and equicoercivity imply convergence of modes; Ayanbayev et al., 2022a, Theorem 4.2)

For $n \in \mathbb{N} \cup \{\infty\}$, let $\mu^{(n)} \in \mathcal{P}(X)$ have OM functional $I_{\mu^{(n)}} \colon E^{(n)} \to \mathbb{R}$ and satisfy property $M(\mu^{(n)}, E^{(n)})$; extend each $I_{\mu^{(n)}}$ to take the value $+\infty$ on $X \setminus E^{(n)}$. Suppose that $(I_{\mu^{(n)}})_{n \in \mathbb{N}}$ is equicoercive and Γ-converges to $I_{\mu^{(\infty)}}$. Then, if $u^{(n)}$ is a weak mode of $\mu^{(n)}$, $n \in \mathbb{N}$, every convergent subsequence of $(u^{(n)})_{n \in \mathbb{N}}$ has as its limit a weak mode of $\mu^{(\infty)}$.

Proof.

The weak modes are exactly the minimisers of the extended OM functionals, and the rest follows from the fundamental theorem of Γ-convergence (e.g. Braides, 2002; Dal Maso, 1993). ■

Applications of the OM Γ-convergence approach

- In practice, one breaks Γ-convergence and equicoercivity of the posterior OM functionals down into (1) Γ-convergence and equicoercivity of the prior OM functionals plus (2) continuous convergence of log-likelihoods.

Applications of the OM Γ-convergence approach

- In practice, one breaks Γ-convergence and equicoercivity of the posterior OM functionals down into (1) Γ-convergence and equicoercivity of the prior OM functionals plus (2) continuous convergence of log-likelihoods.
- Example 1 (Gaussian priors on Hilbert X): Strong convergence of means and covariance operators yields (1) (Ayanbayev et al., 2022a).
- Example 2 (Besov priors on Besov spaces): Simple convergence of the smoothness and integrability parameters yields (1) (Ayanbayev et al., 2022b).

Applications of the OM Γ-convergence approach

- In practice, one breaks Γ-convergence and equicoercivity of the posterior OM functionals down into (1) Γ-convergence and equicoercivity of the prior OM functionals plus (2) continuous convergence of log-likelihoods.
- Example 1 (Gaussian priors on Hilbert X): Strong convergence of means and covariance operators yields (1) (Ayanbayev et al., 2022a).
- Example 2 (Besov priors on Besov spaces): Simple convergence of the smoothness and integrability parameters yields (1) (Ayanbayev et al., 2022b).
- Such assumptions simultaneously give a convergent sequence of approximate BIPs and a convergent sequence of MAP estimators.

Closing remarks

Closing remarks

- At fixed radius $r > 0$, there is an obvious choice of total preorder, and the order-theoretic point of view opens up attractive proof techniques for the existence of maximal/greatest elements (radius-r modes).
- However, such radius-r modes can fail to exist, even in "nice" spaces.
- In the limit as $r \to 0$, there are several limiting preorders that one could consider.
- Our focus is on $\preccurlyeq_0$, whose greatest elements are weak modes, but is a non-total preorder. Even absolutely continuous measures can admit topologically dense antichains, indicating that a measure must satisfy stringent regularity conditions to be certain of having greatest elements, i.e. weak modes.
- Would applications communities be happy with modes as $\preccurlyeq_0$-maximal points?
- A Γ-convergence theory for OM functionals yields a stability theory for weak MAP estimators — even one compatible with stability for BIPs — albeit without convergence rates.

Thank You!

References

B. Ayanbayev, I. Klebanov, H. C. Lie, and T. J. Sullivan. Γ-convergence of Onsager–Machlup functionals: I. With applications to maximum a posteriori estimation in Bayesian inverse problems. *Inverse Probl.*, 38(2):025005, 32pp., 2022a. doi:10.1088/1361-6420/ac3f81.

B. Ayanbayev, I. Klebanov, H. C. Lie, and T. J. Sullivan. Γ-convergence of Onsager–Machlup functionals: II. Infinite product measures on Banach spaces. *Inverse Probl.*, 38(2):025006, 35pp., 2022b. doi:10.1088/1361-6420/ac3f82.

A. Björn and J. Björn. *Nonlinear Potential Theory on Metric Spaces*, volume 17 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich, 2011. doi:10.4171/099.

A. Braides. *Γ-Convergence for Beginners*, volume 22 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2002. doi:10.1093/acprof:oso/9780198507840.001.0001.

G. Dal Maso. *An Introduction to Γ-Convergence*, volume 8 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser Boston, Inc., Boston, MA, 1993. doi:10.1007/978-1-4612-0327-8.

M. Dashti, K. J. H. Law, A. M. Stuart, and J. Voss. MAP estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Probl.*, 29(9): 095017, 27pp., 2013. doi:10.1088/0266-5611/29/9/095017.

D. Dürr and A. Bach. The Onsager–Machlup function as Lagrangian for the most probable path of a diffusion process. *Comm. Math. Phys.*, 60(2):153–170, 1978. doi:10.1007/BF01609446.

B. Hansson. Choice structures and preference relations. *Synthese*, 18(4):443–458, 1968. doi:10.1007/BF00484979.

T. Helin and M. Burger. Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems. *Inverse Probl.*, 31(8):085009, 22pp., 2015. doi:10.1088/0266-5611/31/8/085009.

I. Klebanov and P. Wacker. Maximum *a posteriori* estimators in ℓ^p are well-defined for diagonal Gaussian priors. *Inverse Probl.*, 39(6):065009, 27pp., 2023. doi:10.1088/1361-6420/acce60.

H. Lambley. Strong maximum a posteriori estimation in Banach spaces with Gaussian priors, 2023. arXiv:2304.13622.

H. Lambley and T. J. Sullivan. An order-theoretic perspective on modes and maximum a posteriori estimation in Bayesian inverse problems, 2022.

H. C. Lie and T. J. Sullivan. Equivalence of weak and strong modes of measures on topological vector spaces. *Inverse Probl.*, 34(11):115013, 22pp., 2018. doi:10.1088/1361-6420/aadef2.

B. Sprungk. On the local Lipschitz stability of Bayesian inverse problems. *Inverse Probl.*, 36(5):055015, 31, 2020. doi:10.1088/1361-6420/ab6f43.

A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numer.*, 19:451–559, 2010. doi:10.1017/S0962492910000061.

E. Szpilrajn. Sur l'extension de l'ordre partiel. *Fund. Math.*, 16:386–389, 1930. doi:10.4064/fm-16-1-386-389.

A non-atomic measure with no radius-r mode for any small r

Building on the example $(\mathbb{N}, \Delta, \mu)$, consider the space

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

equipped with the metric d and probability measure μ given by

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

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

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

A non-atomic measure with no radius-r mode for any small r

Theorem 21

(X, d, μ) as defined above has no radius-r mode for any $0 < r < 1$.

Proof. Fix r and let $n \in \mathbb{N}$ be given by $2^{-n} \leqslant r < 2^{-n+1}$. We consider possible centres $x = (\xi, k, m)$ for various m.

- For $m = n$, similar arguments to the atomic example show that $M_r \geqslant \frac{(2\sigma)^{-n}}{Z}$ but that no such ball realises this supremal mass.
- For $m > n$, $B_r(x) \subseteq \mathbb{R} \times \mathbb{N} \times \{m\}$, and so $\mu(B_r(x)) \leqslant \frac{(2\sigma)^{-m}}{Z} < \frac{(2\sigma)^{-n}}{Z}$.
- For $m < n$, $B_r(x) \subseteq \mathbb{R} \times \{k\} \times \{m\}$, and so we again find that $\mu(B_r(x)) < \frac{(2\sigma)^{-n}}{Z}$.

Hence, $M_r = \frac{(2\sigma)^{-n}}{Z}$ but $\mu(B_r(x)) < M_r$ for all $x \in X$, i.e. μ has no radius-r mode. ∎

◀ Back to main