# Γ-convergence of Onsager–Machlup functionals: I. With applications to maximum *a posteriori* estimation in Bayesian inverse problems

**Birzhan Ayanbayev[1]** ⓘ**, Ilja Klebanov[2]** ⓘ**, Han Cheng Lie[3]** ⓘ
**and T J Sullivan[1],*** ⓘ

[1] Mathematics Institute and School of Engineering, University of Warwick, Coventry, CV4 7AL, United Kingdom
[2] Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany
[3] Institut für Mathematik, Universität Potsdam, Campus Golm, Haus 9, Karl-Liebknecht-Straße 24–25, Potsdam OT Golm 14476, Germany

E-mail: birzhan.ayanbayev@warwick.ac.uk, klebanov@zedat.fu-berlin.de, hanlie@uni-potsdam.de and t.j.sullivan@warwick.ac.uk

CrossMark

## Abstract

The Bayesian solution to a statistical inverse problem can be summarised by a mode of the posterior distribution, i.e. a maximum *a posteriori* (MAP) estimator. The MAP estimator essentially coincides with the (regularised) variational solution to the inverse problem, seen as minimisation of the Onsager–Machlup (OM) functional of the posterior measure. An open problem in the stability analysis of inverse problems is to establish a relationship between the convergence properties of solutions obtained by the variational approach and by the Bayesian approach. To address this problem, we propose a general convergence theory for modes that is based on the Γ-convergence of OM functionals, and apply this theory to Bayesian inverse problems with Gaussian and edge-preserving Besov priors. Part II of this paper considers more general prior distributions.

---

*Author to whom any correspondence should be addressed.

(Some figures may appear in colour only in the online journal)

## 1. Introduction

In diverse applications such as Bayesian inference and the transition path analysis of diffusion processes, it is important to be able to summarise a probability measure $\mu$ on a possibly infinite-dimensional space $X$ by a single distinguished point of $X$—a point of maximum probability under $\mu$ in some sense, i.e. a *mode* of $\mu$. If $\mu$ is an absolutely continuous measure on a finite-dimensional Euclidean space $X$, then the modes of $\mu$ are the maximisers of its Lebesgue density. In the Bayesian statistical context, if $\mu$ is the posterior measure, then the modes of $\mu$ are precisely the maximum *a posteriori* (MAP) estimators. If $X$ is an infinite-dimensional Banach space $X$, then a Lebesgue density is not available. In this case it has become common to define modes using the posterior probabilities of norm balls in the small-radius limit. Under suitable conditions, such modes admit a variational characterisation as the minimisers of an appropriate Onsager–Machlup (OM) functional. Heuristically, such an OM functional plays the role of the negative logarithm of the 'Lebesgue density' of $\mu$, but the rigorous formulation of this relationship requires some care. In a statistical context, this variational characterisation of modes suggests a connection between the fully Bayesian approach and the regularised variational approach to inverse problems: the negative logarithm of the prior acts as a regulariser for the misfit (i.e. for the negative log-likelihood).
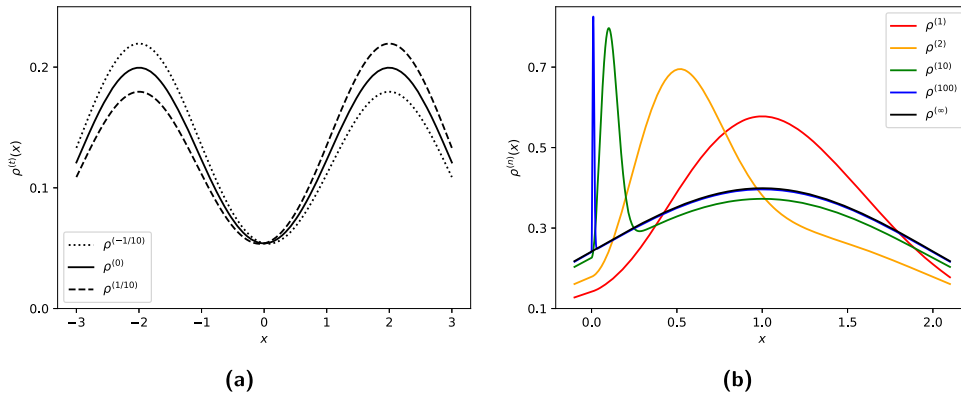
A significant challenge to exploiting this connection is the lack of a suitable convergence theory. This is because the stability properties of MAP estimators are poorly understood. In particular, it is not known under what circumstances mild perturbations of the setup of a Bayesian inverse problem (BIP) lead to mild perturbations of the posterior distribution *and* to mild perturbations of its MAP estimators. Typical examples of perturbations include those arising from finite-dimensional truncation of an infinite-dimensional prior; numerical approximation of an ideal forward operator within the likelihood, e.g. the solution operator of a differential equation; perturbation of observed data; or limiting procedures such as small-noise limits.

In the last decade, beginning with the seminal work of Stuart (2010), many articles have studied the well-posedness and stability of BIPs in function spaces. However, in general, the stability of the posterior and of its MAP estimators are 'orthogonal' questions: two posterior probability measures can be arbitrarily close in a strong sense such as Kullback–Leibler (relative entropy) distance and still have MAP estimators that are at constant distance from one another; conversely, even equality of MAP estimators says nothing about the similarity of the full posteriors. The situation vis-a-vis convergence is even less satisfying, as the following examples show:

**Example 1.1.**

(a) The normal distributions $\mu^{(1)} := \mathcal{N}(0, 1)$ and $\mu^{(\sigma)} := \mathcal{N}(0, \sigma^2)$ on $\mathbb{R}$ have the same unique mode at 0, but can be very far apart in the Kullback–Leibler sense:

$$\mathrm{KL}\left(\mu^{(1)} \| \mu^{(\sigma)}\right) = \frac{\sigma^{-1} - 1 + \log \sigma}{2} \to +\infty \quad \text{as } \sigma \to 0 \quad \text{or} \quad \sigma \to +\infty,$$

$$(1.1)$$

**Figure 1.** (a) The densities $\rho^{(t)}$ from (1.2) for $t \approx 0$ are close in the Kullback–Leibler sense but their modes are far apart, at $x \approx \pm r = \pm 2$ respectively. However, the cluster points of the modes of $\rho^{(t)}$ as $t \to 0$ yield the modes of $\rho^{(0)}$. (b) The densities $\rho^{(n)}$ from (1.3) for $n \in \{1, 2, 10, 100, \infty\}$. For $n \in \mathbb{N}$, the unique mode of $\rho^{(n)}$ is at $x \approx \frac{1}{n}$, whereas the unique mode of $\rho^{(\infty)}$ is at $x = 1$, even though $\rho^{(n)} \to \rho^{(\infty)}$ pointwise.

(b) In the other direction, fix a large $r \gg 1$ and consider for $t \in (-1, 1)$ the following Gaussian mixture distribution $\mu^{(t)}$ on $\mathbb{R}$ with Lebesgue density $\rho^{(t)} : \mathbb{R} \to [0, \infty)$ illustrated in figure 1(a) and given by

$$\rho^{(t)}(x) := \frac{(1 + t) \exp\left(-\frac{1}{2}(x - r)^2\right) + (1 - t) \exp\left(-\frac{1}{2}(x + r)^2\right)}{2\sqrt{2\pi}}. \qquad (1.2)$$

For large values of $r$, $\rho^{(t)}$ has two local maximisers near $r$ and $-r$. When $t > 0$, the local maximiser near $r$ is the unique mode; when $t < 0$, the local maximiser near $-r$ is the unique mode. However, $\mu^{(t)}$ and $\mu^{(-t)}$ are very close in the Kullback–Leibler sense: e.g. for $r = 5$, $\mathrm{KL}\left(\mu^{(t)} \| \mu^{(-t)}\right) \approx 10^{5/4} t^{9/4}$ as $t \to 0$.

One might argue that at least the cluster points as $t \to 0$ of the modes of the measures $\mu^{(t)}$ yield the two modes at $x = \pm r$ of the symmetric Gaussian mixture $\mu^{(0)}$. However, even this situation cannot be expected to hold true in general, as the next example shows.

(c) For $n \in \mathbb{N}$, let $\mu^{(n)}$ be the probability measure on $\mathbb{R}$ with Lebesgue density $\rho^{(n)} : \mathbb{R} \to [0, \infty)$ illustrated in figure 1(b) and given by

$$\rho^{(n)}(x) := \frac{\exp\left(-\frac{1}{2}(x - 1)^2\right) + 1[x \geqslant 0] 4n^2 x^2 \exp(-n^2 x^2)}{\sqrt{2\pi} + \sqrt{\pi}/n}, \qquad (1.3)$$

where $1[P] := 1$ if $P$ is true and $1[P] := 0$ if $P$ is false. The densities converge pointwise but not uniformly to the Gaussian distribution $\mu^{(\infty)} = \mathcal{N}(1, 1)$ with density $\rho^{(\infty)}(x) \propto \exp\left(-\frac{1}{2}(x - 1)^2\right)$, which has a unique mode at $x = 1$. Convergence in the Kullback–Leibler sense also holds, with $\mathrm{KL}\left(\mu^{(\infty)} \| \mu^{(n)}\right) \approx \frac{1}{n}$. However, each $\mu^{(n)}$ has a unique mode at approximately $x \approx \frac{1}{n}$ and the unique cluster point of this sequence of modes is $x = 0 \neq 1$. Thus, even in finite-dimensional settings, pointwise convergence of densities (and hence of OM functionals) does not imply convergence of modes.

Given the variational characterisation of modes as minimisers of an OM functional, it seems natural to assess the convergence (and hence stability) of modes using a notion of convergence

that is appropriate for variational problems, i.e. one for which the convergence of functionals implies the convergence of minimisers. The notion of $\Gamma$-convergence, as introduced by De Giorgi and collaborators from the 1970s onwards (De Giorgi 2006), fulfils exactly this rôle and, in particular, overcomes the shortcomings of pointwise convergence as illustrated in example 1.1(c). Indeed, if the densities in example 1.1(c) were uniformly convergent, then they would be continuously convergent and hence $\Gamma$-convergent as well, and the pathological non-convergence of modes would have been avoided. Therefore, the strategy followed by this article consists in the following:

(a) Formulate the problem of finding modes of probability measures on potentially infinite-dimensional spaces as a variational problem for the associated OM functionals; and

(b) Study the $\Gamma$-convergence properties of such problems, in order to obtain criteria for the convergence and stability of such modes.

The remainder of this article is structured as follows. Section 2 gives an overview of related work in the theory of modes for measures on infinite-dimensional spaces and section 3 sets out some notation and basic results for the rest of the article. Section 4 explores the correspondence between modes and minimisers of OM functionals, and hence demonstrates that $\Gamma$-convergence of OM functionals is the correct notion of convergence to ensure convergence of modes. Section 5 develops this idea in two prototypical settings, namely Gaussian and Besov-1 measures, which are frequently used as Bayesian prior distributions; these results can then be transferred to measures that are absolutely continuous with respect to these paradigmatic examples and can be interpreted as the corresponding posterior measures. More general prior measures, which include Cauchy measures and Besov-$p$ measures for $1 \leqslant p \leqslant 2$ are treated in part II of this paper (Ayanbayev *et al* 2021). In section 6, these ideas are then applied to the convergence and stability of MAP estimators for BIPs, for which Gaussian and Besov-1 measures are prototypical prior distributions. Some conclusions and suggestions for further work are given in section 7. Standard definitions and results relating to $\Gamma$-convergence are collected in appendix A and technical supporting results are given in appendix B.

## 2. Overview of related work

In stochastic analysis and mathematical physics, the interpretation of the minimisers of OM functionals over path spaces as most probable paths appears to be due to Dürr and Bach (1978). The OM functionals of diffusion processes, and hence the determination of MAP paths, have been considered by e.g. Zeitouni (1989) and Dembo and Zeitouni (1991). It is important to note that simply determining the OM functional on some $n$-dimensional approximation space and then taking a limit as $n \to \infty$ can fail to yield the correct OM functional as defined in terms of ratios of small ball probabilities by (3.3). This is because the space on which the OM functional is finite can be 'smoother' than the full space on which small ball probabilities are defined (Dashti *et al* 2013).

Recent years have seen a growing interest in the well-posedness and stability of BIPs in function spaces, a perspective proposed by a seminal article of Stuart (2010) that has stimulated many follow-on works and generalisations (e.g. Dashti *et al* (2012), Hosseini (2017), Latz (2020), Sprungk (2020), Sullivan (2017)). There is also complementary theory of discretisation invariance, sometimes referred to as the 'Finnish school', which in some sense treats the finite-dimensional discrete versions of BIPs as the primary objects of interest but pays careful attention to their limiting properties as the discretisation dimension tends to infinity (e.g. Lasanen (2012a, 2012b), Lassas and Siltanen (2004), Lassas *et al* (2009), Lehtinen *et al* (1989)).

However, the robustness studies in these works have focussed on the robustness of the posterior measure in a distributional sense such as the Hellinger, Kullback–Leibler, or Wasserstein sense and, as example 1.1 shows, these are insufficient to ensure robustness of modes or MAP estimators. Our results can be seen as contributions to this field in the sense that they establish stability/convergence of MAP estimators in a setting that is not limited to finite-dimensional or even linear spaces.

In the BIP context, the definition of a MAP estimator as the centre of a norm ball that has maximum posterior probability in a small-radius limit appears to be due to Dashti *et al* (2013). As Dashti *et al* (2013) note, a similar definition of a *maximal point* was given earlier by Hegland (2007), but that analysis was implicitly limited to the finite-dimensional setting, since it assumed finiteness of the Cameron–Martin norm. The context of Dashti *et al* (2013) was limited to a separable Hilbert space[4] $X$ equipped with a Bayesian posterior measure $\mu$ that was absolutely continuous with respect to a centred non-degenerate Gaussian reference measure $\mu_0$. In this setting, Dashti *et al* (2013) established the existence of MAP estimators and characterised them as the minimisers of the OM functional, which they further identified as the sum of the log-likelihood and the OM functional of $\mu_0$. When read in the context of a general probability measure on a metric space, rather than the original setting of a Bayesian posterior on a Hilbert space, the definition of Dashti *et al* (2013) is essentially the definition of a *strong mode* (definition 3.6).

The work of Dashti *et al* (2013) has been extended in multiple ways. Dunlop and Stuart (2016) proved the connection between MAP estimators and OM functionals in the setting of piecewise continuous inversion, where the prior is defined in terms of a combination of Gaussian random fields. Recently, Kretschmann (2019) has corrected some technical deficiencies of Dashti *et al* (2013).

The definition of a (strong) MAP estimator for $\mu$ given by Dashti *et al* (2013) was relaxed to that of a *weak MAP estimator* by Helin and Burger (2015), in which comparisons between the masses of balls are only performed for balls whose centres differ by an element of a topologically dense subspace $E$ of $X$. Helin and Burger (2015) showed that this weak MAP estimator has a close relationship with the zeroes of the logarithmic derivative $\beta_h^\mu := \frac{d(d_h\mu)}{d\mu}$ of $\mu$, where

$$d_h\mu(A) := \lim_{t \to 0} \frac{\mu(A + th) - \mu(A)}{t} \quad \text{for measurable } A \subseteq X,$$

is the Fomin (directional) derivative of $\mu$ in the direction $h \in X$. The initial analysis of the weak MAP estimator relied upon the existence of a continuous representative for $\beta_h^\mu$, which could not be guaranteed for several important applications, notably the edge-preserving Besov prior with $p = 1$. By focussing on the Radon–Nikodym derivative $r_h^\mu := \frac{d\mu(\cdot - h)}{d\mu}$ instead of $\beta_h^\mu$ instead of $\beta_h^\mu$, the analysis of Agapiou *et al* (2018) remedied this shortcoming, posed the definitions and results in more general terms of modes of probability measures rather than MAP estimators of Bayesian posteriors, and also considered local (rather than global) strong and weak modes. The equivalence of strong and weak modes when $E$ is dense in $X$ and under a *uniformity condition* on $\mu$ was established by Lie and Sullivan (2018), who worked in the more general context of measures on metrisable topological vector spaces.

Finally, we mention that the definition of a strong mode is unsuitable for probability measures with bounded support, and especially to such measures with essential discontinuities in the density. Examples of such measures include uniform measures on bounded subsets of

---

[4] Dashti *et al* (2013) stated their results in separable Banach spaces. However, as pointed out by Wacker (2020), their proof techniques are valid only in the case that $X$ is a Hilbert space and not in the Banach space case.

*X*. The definition is unsuitable because it excludes 'obvious' modes on the boundary of the support. The recent *generalised mode* of Clason *et al* (2019) addresses this deficiency.

## 3. Preliminaries and notation

### 3.1. General notation and assumptions

Throughout this article, $X$ will denote either a topological space, a metric space, a separable Banach or a Hilbert space. When thought of as a measurable space, $X$ will be equipped with its Borel $\sigma$-algebra $\mathcal{B}(X)$, which is generated by the collection of all open sets. When $X$ is a metric space, we write $B_r(x)$ for the open ball in $X$ of radius $r$ centred on $x$. The set of all probability measures on $(X, \mathcal{B}(X))$ will be denoted $\mathcal{P}(X)$; we denote typical probability measures by $\mu$, $\mu_0$, $\mu^{(n)}$ for $n \in \mathbb{N} \cup \{\infty\}$ etc. The topological support of $\mu \in \mathcal{P}(X)$ defined on a metric space $X$ is

$$\mathrm{supp}(\mu) := \{x \in X \,|\, \text{for all } \ r > 0, \ \mu(B_r(x)) > 0\}, \tag{3.1}$$

which is always a closed subset of $X$.

We write $\overline{\mathbb{R}}$ for the extended real line $\mathbb{R} \cup \{\pm\infty\}$, i.e. the two-point compactification of $\mathbb{R}$.

For $0 < p \leqslant \infty$, we write $\ell^p := \ell^p(\mathbb{N})$ for the real sequence space of $p$th-power summable sequences, and bounded sequences in the case $p = \infty$. Furthermore, given $\gamma = (\gamma_k)_{k\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}_{>0}$, we write

$$\ell^p_\gamma := \left\{ h \in \mathbb{R}^{\mathbb{N}} \,\middle|\, (h_k/\gamma_k)_{k\in\mathbb{N}} \in \ell^p \right\}, \qquad \|h\|_{\ell^p_\gamma} := \|(h_k/\gamma_k)_{k\in\mathbb{N}}\|_{\ell^p}, \tag{3.2}$$

for the corresponding weighted $\ell^p$ space. It is well known that $\| \cdot \|_{\ell^p}$ (and hence $\| \cdot \|_{\ell^p_\gamma}$) is a complete quasinorm when $p > 0$, a Banach norm when $p \geqslant 1$, and a Hilbert norm when $p = 2$.

### 3.2. OM functionals

We recall here the definition of an OM functional for a measure. The minimisers of OM functionals will turn out to be the modes of the measure (see section 4). We also stress a property that is already implicitly used without a name in the modes literature, one that essentially ensures that the modes must lie in the domain of the OM functional.

**Definition 3.1.** Let $X$ be a metric space and let $\mu \in \mathcal{P}(X)$. We say that $I = I_\mu = I_{\mu,E} : E \to \mathbb{R}$, with[5] $\varnothing \neq E \subseteq \mathrm{supp}(\mu) \subseteq X$, is an **Onsager–Machlup functional** (OM functional) for $\mu$ if

$$\lim_{r \searrow 0} \frac{\mu(B_r(x_1))}{\mu(B_r(x_2))} = \exp(I(x_2) - I(x_1)) \quad \text{for all } x_1, x_2 \in E. \tag{3.3}$$

We say that **property $M(\mu, E)$** is satisfied if, for some $x^\star \in E$,

$$x \in X \setminus E \implies \lim_{r \searrow 0} \frac{\mu(B_r(x))}{\mu(B_r(x^\star))} = 0, \tag{3.4}$$

and in this situation we extend $I$ to a function $I : X \to \overline{\mathbb{R}}$ with $I(x) := +\infty$ for $x \in X \setminus E$.

**Remark 3.2 (The *M*-property).** Lemma B.1 establishes some basic facts about property $M(\mu, E)$. In particular, lemma B.1(a) shows that property $M(\mu, E)$ does not depend on the choice of reference point $x^\star \in E$, provided that $\mu$ has an OM functional on $E$.

---

[5] We insist on defining the real-valued version of $I$ only on a subset of $\mathrm{supp}(\mu)$ since, for $x_2 \notin \mathrm{supp}(\mu)$, the ratio in the limit in (3.3) is infinite or even undefined for small enough $r > 0$.

Every measure $\mu$ admits an OM functional if $E$ is taken to be small enough, e.g. a singleton subset of supp($\mu$). Therefore, there is a natural desire to have $E$ be 'maximal' in some sense. Property $M(\mu, E)$ means that the set $E$ is the 'maximal' set on which the OM functional assumes finite values.

It is tempting but incorrect to read property $M(\mu, E)$ as saying that $\mu$ somehow concentrates upon $E$. A straightforward counterexample is given by any non-degenerate Gaussian measure $\mu$ with infinite-dimensional Cameron–Martin space $H(\mu)$, such as the law $\mu$ of standard Brownian motion on $X = C([0, 1]; \mathbb{R})$ with $H(\mu) = H^1([0, 1]; \mathbb{R})$. In this situation, property $M(\mu, H(\mu))$ holds (Dashti *et al* 2013, lemma 3.7) and yet $\mu(H(\mu)) = 0$ (Bogachev 1998, theorem 2.4.7). Rather, the purpose of property $M(\mu, E)$ is to ensure that the global weak modes (see definition 3.7) of $\mu$ lie in $E$ and are precisely the minimisers of its extended OM functional (see proposition 4.1). Furthermore, it is essentially the lim sup part of the limit in (3.4) that ensures this; example B.2 shows that if we weaken (3.4) by considering the limit inferior instead of the limit, then—even for very simple choices of $E$—the desired correspondence may break down.

**Remark 3.3 (Topological considerations).** Note that in defining the OM functional here and various notions of mode/MAP estimator later on, we use *open* balls (following e.g. Dashti *et al* (2013) and Agapiou *et al* (2018)) rather than closed balls $\bar{B}_\varepsilon(x)$ (following e.g. Bogachev (1998)). However, proposition B.3 shows that these two notions yield the same definition of OM functionals and global weak modes.

**Remark 3.4 (Uniqueness of OM functionals).** Note that OM functionals are at best unique up to the addition of real constants. Whenever we talk about $\Gamma$-convergence and equicoercivity of sequences of OM functionals, which are at the core of this work, we always mean the existence of representatives that fulfil these properties. Further, whenever we apply results that require both $\Gamma$-convergence and equicoercivity (such as theorem A.3), we need to make sure that the *same* representatives can be chosen for both properties.

**Remark 3.5 (OM functionals and changes of metric).** Unfortunately, the choice of metric on a space $X$ can affect the OM functional of a measure $\mu$ on $X$, even beyond the non-uniqueness alluded to in remark 3.4, and even if the two metrics are Lipschitz equivalent. An explicit example of this is furnished by the finite measure $\mu$ of Lie and Sullivan (2018), (example 5.6); see example B.4 for details.

### 3.3. Modes and MAP estimators

In finite-dimensional spaces, for probability measures that are either purely discrete or possess a continuous Lebesgue density, modes (as points of maximum probability) are easily defined as being global maximisers of the probability mass function or probability density function as appropriate. For probability measures on infinite-dimensional spaces, however, the situation is more delicate as there is no infinite-dimensional analogue of Lebesgue measure to serve as a uniform reference. Therefore, it has become common to define modes by examining the masses of norm balls in the small-radius limit. The following definition of a *strong mode* is a slight generalisation of the definition of a *MAP estimator* for a Bayesian posterior measure on a normed space as given by Dashti *et al* (2013).

**Definition 3.6.** Let $X$ be a metric space. A **strong mode** of $\mu \in \mathcal{P}(X)$ is any $u \in X$ satisfying

$$\lim_{r \searrow 0} \frac{\mu(B_r(u))}{M_r} = 1, \tag{3.5}$$

where $M_r := \sup_{w \in X} \mu(B_r(w)) \in (0, 1]$. Since $\mu(B_r(u)) \leqslant M_r$, the ratio inside the limit in (3.5) is at most one, and so it is equivalent to define a strong mode as being any $u \in X$ for which

$$\lim_{r \searrow 0} \frac{\mu(B_r(u))}{M_r} \geqslant 1 \quad \text{or} \quad \liminf_{r \searrow 0} \frac{\mu(B_r(u))}{M_r} \geqslant 1 \quad \text{or} \quad \limsup_{r \searrow 0} \frac{M_r}{\mu(B_r(u))} \leqslant 1.$$

A related notion of mode for a measure is the *weak mode* or *weak MAP estimator* (Helin and Burger 2015, definition 4), which are points that dominate all other points within an affine subspace, in terms of small ball probabilities. Since we are only interested in *global* weak modes, we simplify the definition slightly and at the same time generalise this concept to metric spaces. The original definition relies on subtraction and thus only applies in the case of linear spaces.

**Definition 3.7.** For a metric space $X$, a **global weak mode** of $\mu \in \mathcal{P}(X)$ is any $u \in \text{supp}(\mu)$ satisfying, for any point $u' \in X$,

$$\limsup_{r \searrow 0} \frac{\mu(B_r(u'))}{\mu(B_r(u))} \leqslant 1. \tag{3.6}$$

**Remark 3.8.** The definition (3.6) differs further from that of Helin and Burger (2015), (definition 4) in that we use a limit superior instead of a limit. We suspect that this is the way it was intended to be defined—if a point dominates every other point in terms of small ball probabilities (in the sense that the ratio in (3.6) becomes $\leqslant 1$ for sufficiently small $r > 0$), it should be called a weak mode. This suspicion is based on Helin and Burger (2015), (lemma 3), where the authors prove that every strong mode is a weak mode, which is clearly a desirable property given the terminology 'strong' and 'weak', but their proof is incorrect given their original definition. The reason is that the ratio in (3.6) can drop below $\leqslant 1$ as $r \searrow 0$ for certain points $u, u' \in X$ without the limit existing (e.g. it might oscillate between $0$ and $\frac{1}{2}$), such that $u$ cannot be a weak mode in the original definition, while it can still be a strong mode. Example B.2 shows that such oscillations can occur and a slight modification provides a concrete counterexample to Helin and Burger (2015), (lemma 3).

**Lemma 3.9.** *Let $\mu \in \mathcal{P}(X)$ and let $u \in \text{supp}(\mu)$ be a strong mode of $\mu$. Then $u$ is a global weak mode of $\mu$.*

**Proof.** Since $u \in \text{supp}(\mu)$ is a strong mode of $\mu$, we obtain, for any point $u' \in X$,

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

Sufficient conditions for the converse implication are given by Lie and Sullivan (2018). The relationship between modes and OM functionals will be examined in section 4.

### 3.4. Generalised inverses

We adopt the following definition of the Moore–Penrose pseudoinverse of an operator (Engl *et al* 1996, definition 2.2):

**Definition 3.10.** For a bounded linear operator $A : X \to Y$ between Hilbert spaces $X$ and $Y$, the **Moore–Penrose pseudoinverse** $A^\dagger$ of $A$ is the unique extension of $(A|_{(\ker A)^\perp})^{-1}$ to a (generally

unbounded) linear operator $A^\dagger : \operatorname{ran} A \oplus (\operatorname{ran} A)^\perp \to X$ subject to the restriction that $\ker A^\dagger = (\operatorname{ran} A)^\perp$.

**Remark 3.11.** For $y \in \operatorname{ran} A \oplus (\operatorname{ran} A)^\perp$,

$$A^\dagger y = \arg \min \left\{ \|x\|_X \mid x \text{ minimises } \|Ax - y\| \right\}.$$

In particular, for $y \in \operatorname{ran} A$, $A^\dagger y$ is the minimum-norm solution of $Ax = y$ (Engl *et al* 1996, theorem 2.5).

**Remark 3.12.** For a self-adjoint and positive semi-definite (SPSD) and compact operator $C = \sum_{n \in \mathbb{N}} \sigma_n^2 e_n \otimes e_n : X \to X$ on a Hilbert space $X$, $(e_n)_{n \in \mathbb{N}}$ being an orthonormal system in $X$ and $\sigma_n \geqslant 0$ for each $n \in \mathbb{N}$, we denote the SPSD operator square root of $C$ by $C^{1/2}$ and furthermore set

$$C^{\dagger/2} := (C^{1/2})^\dagger = \sum_{n \in \mathbb{N} \,:\, \sigma_n \neq 0} \sigma_n^{-1} e_n \otimes e_n.$$

Note that $(C^\dagger)^{1/2}$ can differ from $(C^{1/2})^\dagger$ since it may have a smaller domain.

## 4. Modes, OM functionals, and their convergence

The purpose of this section is to firmly establish the intuitively plausible relationship between the modes of a probability measure $\mu$ and its OM functional $I$, namely that the global weak modes of $\mu$ are exactly the global minimisers of $I$. Once this is done, it is a relatively simple matter to give sufficient conditions for the global weak modes of a sequence of measures to converge to the global weak modes of a limiting measure: $\Gamma$-convergence and equicoercivity of the associated OM functionals.

**Proposition 4.1 (Global weak modes and OM functionals).** *Let $X$ be a metric space and let $I : E \to \mathbb{R}$ be an OM functional for $\mu \in \mathcal{P}(X)$, defined on a nonempty subset $E \subseteq X$ with property $M(\mu, E)$. Then $u \in E$ is a global weak mode of $\mu$ if and only if $u$ is a minimiser of the extended OM functional $I : X \to \overline{\mathbb{R}}$.*

**Proof.** By property $M(\mu, E)$ and lemma B.1(c), any global weak mode of $\mu$ must lie in $E$, and in addition any minimiser of $I : X \to \overline{\mathbb{R}}$ must also lie in $E$, where $E$ is the set on which $I$ takes real values. Let $u \in E$ be arbitrary. Then $\lim_{r \searrow 0} \frac{\mu(B_r(u'))}{\mu(B_r(u))}$ exists for any $u' \in E$ by definition of $I$ being the OM functional, and the same limit exists and equals 0 for $u' \in X \backslash E$, because property $M(\mu, E)$ holds. Thus,

$$
\begin{aligned}
u \in E \text{ is a global weak mode} &\iff \text{for all } u' \in X, \ \limsup_{r \searrow 0} \frac{\mu(B_r(u'))}{\mu(B_r(u))} \leqslant 1 \\
&\iff \text{for all } u' \in X, \ \lim_{r \searrow 0} \frac{\mu(B_r(u'))}{\mu(B_r(u))} \leqslant 1 \\
&\iff \text{for all } u' \in X, \ \exp\left(I(u) - I(u')\right) \leqslant 1 \\
&\iff \text{for all } u' \in X, \ I(u) \leqslant I(u'),
\end{aligned}
$$

as claimed. □

Property $M(\mu, E)$ was essential in the above argument in order for points outside $E$ to be treated in a consistent way. Recall that, in definition 3.1, for a given $\mu \in \mathcal{P}(X)$, we initially defined the OM functional of $\mu$ to be a function $I : E \to \mathbb{R}$. Only under property $M(\mu, E)$ can we sensibly extend $I$ to a $\overline{\mathbb{R}}$-valued function on $X$ by setting $I(x) := +\infty$ for $x \in X \backslash E$. The motivation for this extension is that, by lemma B.1(c), no point of $X \backslash E$ can be a global weak mode for $\mu$, and hence cannot be a strong mode for $\mu$.

Unfortunately, without additional assumptions, an analogous result to proposition 4.1 cannot hold for strong modes, as demonstrated in example B.5. The main idea behind the measure $\mu$ constructed therein is that $u = 1$ 'dominates' any other (fixed) point $u' \in X = \mathbb{R}$ in the limit $r \searrow 0$, i.e. $\lim_{r \searrow 0} \frac{\mu(B_r(u'))}{\mu(B_r(u))} \leqslant 1$, hence $u = 1$ is a global weak mode; but for certain arbitrarily small radii $r_n$, $n \in \mathbb{N}$, there exist points $u_n \in X$ that 'dominate' $u$ by a margin, in fact $\liminf_{r \searrow 0} \frac{\mu(B_r(u))}{M_r} \leqslant \frac{1}{\sqrt{2}}$, hence $u = 1$ cannot be a strong mode. Moreover, for $E = \mathbb{N} \subseteq X$, property $M(\mu, E)$ holds and an OM functional $I_{\mu,E} : E \to \mathbb{R}$ exists and has $u = 1$ as its minimizer. The construction is based on suitably chosen singularities of the Lebesgue density $\rho$ of $\mu$.

The following result, which is an almost immediate consequence of the preceding discussion, provides clear criteria for the convergence of global weak modes along sequences of probability measures. (Definitions and basic properties of $\Gamma$-convergence, equicoercivity, etc are collected in appendix A.)

**Theorem 4.2 ($\Gamma$-convergence and equicoercivity imply convergence of modes).**
*Let $X$ be a metric space and let, for $n \in \mathbb{N} \cup \{\infty\}$, $\mu^{(n)} \in \mathcal{P}(X)$ have OM functionals $I^{(n)} : E^{(n)} \to \mathbb{R}$, such that property $M(\mu^{(n)}, E^{(n)})$ is satisfied. Extend each $I^{(n)}$ to take the value $+\infty$ on $X \backslash E^{(n)}$. Suppose that the sequence $(I^{(n)})_{n \in \mathbb{N}}$ is equicoercive and $\Gamma$-converges to $I^{(\infty)}$. Then, if $u^{(n)}$ is a global weak mode of $\mu^{(n)}$, $n \in \mathbb{N}$, every convergent subsequence of $(u^{(n)})_{n \in \mathbb{N}}$ has as its limit a global weak mode of $\mu^{(\infty)}$.*

**Proof.** By proposition 4.1, the global weak modes of $\mu^{(n)}$ are precisely the minimisers of the extended version of $I^{(n)}$, $n \in \mathbb{N} \cup \{\infty\}$. The rest follows immediately from the fundamental theorem of $\Gamma$-convergence (theorem A.3). $\qquad\qquad\square$

It is instructive to reconsider the earlier example 1.1(c) in light of theorem 4.2. The problem in example 1.1(c)—in which the unique modes of the measures $\mu^{(n)}$ fail to cluster at the unique mode of the limiting measure $\mu^{(\infty)}$—can now be recognised as being due to the fact that although pointwise convergence of Lebesgue densities and OM functionals holds, $\Gamma$-convergence does not. Therefore, theorem 4.2 does not apply to that example and there is no reason for modes to converge in this case.

Theorem 4.2 is, of course, a highly general result. For it to be useful in specific situations, one must prove property $M(\mu^{(n)}, E^{(n)})$ and identify the form of the OM functional $I^{(n)}$ for every $n \in \mathbb{N}$. In addition, one must verify both the $\Gamma$-convergence and equicoercivity properties of the sequence $(I^{(n)})_{n \in \mathbb{N}}$. In the next section, we do this for Gaussian measures and Besov-1 probability measures, which are commonly used as priors in the context of BIPs.

## 5. $\Gamma$-convergence of OM functionals for Gaussian and Besov-1 priors

This section illustrates the preceding general theory of convergence of modes via $\Gamma$-convergence of OM functionals by means of two key examples, namely Gaussian and Besov $B_1^s$ measures, both of which commonly arise as prior distributions in BIPs. Besov $B_p^s$-priors with $1 \leqslant p \leqslant 2$, Cauchy priors, and more general product measures are treated in a unified way in

part II of this paper (Ayanbayev *et al* 2021). The convergence of modes (MAP estimators) for *posterior* distributions will be discussed in section 6.

### *5.1. Gaussian measures*

As a natural first case, we consider the $\Gamma$-convergence of the OM functionals of Gaussian measures—and we call attention to the fact that we consider Gaussian measures with possibly indefinite covariance operators. It is almost folklore that the OM functional of a Gaussian measures is half the square of the associated Cameron–Martin norm; a precise formulation of this result is the following.

**Theorem 5.1 (OM functional of a Gaussian on a separable Banach space).** *Let $\mu$ be a centred Gaussian measure on a separable Banach space $X$. Let $H(\mu)$ be the Cameron–Martin space of $\mu$, with Cameron–Martin norm $\|\cdot\|_{H(\mu)}$. Then, for all $h, k \in H(\mu)$,*

$$\lim_{r \searrow 0} \frac{\mu(B_r(h))}{\mu(B_r(k))} = \exp\left(\frac{1}{2}\|k\|^2_{H(\mu)} - \frac{1}{2}\|h\|^2_{H(\mu)}\right). \tag{5.1}$$

*In particular, the OM functional for $\mu$ on the Cameron–Martin space $H(\mu)$ is half the square of the Cameron–Martin norm.*

**Proof.** This is a special case of Bogachev (1998), (corollary 4.7.8), in which the cylindrical $\sigma$-algebra $\mathcal{E}(X)$ and the Borel $\sigma$-algebra $\mathcal{B}(X)$ coincide by the separability of $X$, the measurable seminorm $q$ is the ambient norm $\|\cdot\|_X$, the $q$-ball $V_r$ is the ball $B_r(0) \in \mathcal{B}(X)$, and the projection $\pi_q$ is the identity due to the definiteness of $q(\cdot) = \|\cdot\|_X$. Note that Bogachev (1998) works with closed balls, this difference being inconsequential in view of proposition B.3. $\qquad\square$

**Corollary 5.2.** *Let $\mu = \mathcal{N}(0, C)$ be a centred Gaussian measure on a separable Hilbert space $X$, where the covariance $C$ is interpreted as an SPSD operator on $X$. Then the (extended) OM functional $I_\mu : X \to \overline{\mathbb{R}}$ of $\mu$ is given by*

$$I_\mu(u) = \begin{cases} \dfrac{1}{2}\|C^{\dagger/2}u\|^2_X & \text{for } u \in H(\mu) = \operatorname{ran} C^{1/2}, \\ +\infty & \text{otherwise.} \end{cases}$$

**Proof.** By Bogachev (1998), (section 2.3, p. 49), the reproducing kernel Hilbert space $X^*_\mu := \overline{X^*}^{L^2(\mu)}$ of $\mu$ can be identified with the weighted Hilbert space of sequences

$$\ell^2_C := \left\{ x = (x_n) \,\middle|\, \sum_{n \in \mathbb{N}} \sigma_n^2 x_n^2 < \infty \right\}, \qquad \|x\|_{\ell^2_C} := \left(\sum_{n \in \mathbb{N}} \sigma_n^2 x_n^2\right)^{1/2}. \tag{5.2}$$

Further, after extending $C$ naturally to $X^*_\mu$, the Cameron–Martin space coincides with the image of $X^*_\mu$ under $C$, i.e. $H(\mu) = C(X^*_\mu)$.

Now let $C = \sum_{n \in \mathbb{N}} \sigma_n^2 e_n \otimes e_n$, $\sigma_n \geqslant 0$, be the eigenvalue decomposition of its covariance operator $C$ with complete orthonormal system $(e_n)_{n \in \mathbb{N}}$ and let $u = \sum_{n \in \mathbb{N}} u_n e_n \in H(\mu)$. Since $H(\mu) = C(X^*_\mu)$, there exists $x = (x_n)_{n \in \mathbb{N}} \in \ell^2_C$, such that $u_n = \sigma_n^2 x_n$ for all $n \in \mathbb{N}$, and, by Bogachev (1998), (lemma 2.4.1) and remark 3.12,

$$\|u\|_{H(\mu)}^2 = \|x\|_{\ell_C^2}^2 = \sum_{n \in \mathbb{N}} \sigma_n^2 x_n^2 = \sum_{n \in \mathbb{N} \,:\, \sigma_n \neq 0} \frac{u_n^2}{\sigma_n^2} = \|C^{\dagger/2} u\|_X^2.$$

The claim now follows from theorem 5.1 and from Dashti *et al* (2013), (lemma 3.7), where property $M(\mu, H(\mu))$ is established[6]. $\qquad\square$

**Remark 5.3.** Note that the notation in Bogachev (1998), (section 2.3, p. 49) is slightly imprecise, since the space $\ell_C^2$ in (5.2) is, in general, only a pre-Hilbert space (and $\|\cdot\|_{\ell_C^2}$ is just a seminorm). To be rigorous, one would need to consider the quotient space of $\ell_C^2$ after factoring out the subspace $\{x \,|\, \|x\|_{\ell_C^2} = 0\}$. This detail has no influence on the proof of corollary 5.2.

**Corollary 5.4.** *Let $\mu_0 = \mathcal{N}(0, C)$ be a centred Gaussian measure on a separable Hilbert space $X$, where the covariance $C$ is interpreted as an SPSD operator on $X$, and $\mu = \mathcal{N}(m, C)$. Then the OM functional $I_\mu : X \to \overline{\mathbb{R}}$ of $\mu$ is given by*

$$I_\mu(u) = \begin{cases} \dfrac{1}{2}\|(u - m)\|_{H(\mu_0)}^2 = \dfrac{1}{2}\|C^{\dagger/2}(u - m)\|_X^2 & \text{for } (u - m) \in H(\mu_0) = \operatorname{ran} C^{1/2}, \\ +\infty & \text{otherwise.} \end{cases}$$

**Proof.** This follows directly from theorem 5.1 and corollary 5.2. $\qquad\square$

We now give the main result of this section, that the strong (norm) convergence of means and covariance operators of Gaussian measures is sufficient to ensure that their associated OM functionals are $\Gamma$-convergent and equicoercive.

**Theorem 5.5 ($\Gamma$-convergence and equicoercivity of OM functionals for Gaussian measures).** *Let $X$ be a separable Hilbert space and $\mu^{(n)} = \mathcal{N}(m^{(n)}, C^{(n)})$ and $\mu = \mathcal{N}(m, C)$ be Gaussian measures on $X$ such that $m^{(n)} \to m$ in $X$ and $C^{(n)} \to C$ with respect to the operator norm. Then $I_\mu = \Gamma\text{-}\lim_{n \to \infty} I_{\mu^{(n)}}$. Furthermore, the sequence $(I_{\mu^{(n)}})_{n \in \mathbb{N}}$ is equicoercive.*

**Remark 5.6.** Since all the Gaussian OM functionals $I_{\mu^{(n)}}$ are quadratic forms, and homogeneity is preserved by $\Gamma$-limits (Braides 2006, proposition 2.13), it is not surprising that $\Gamma\text{-}\lim_{n \to \infty} I_{\mu^{(n)}}$ is quadratic—the point here is to check that the quadratic forms $\Gamma\text{-}\lim_{n \to \infty} I_{\mu^{(n)}}$ and $I_\mu$ agree, and moreover with careful attention to the possibility of indefinite covariances.

**Proof of theorem 5.5.** Let $A := C^{1/2}$ and $A_n := (C^{(n)})^{1/2}$. Further, let $(e_k)_{k \in \mathbb{N}}$ be an orthonormal eigenbasis of $A$, $A = \sum_{k \in \mathbb{N}} \sigma_k \, e_k \otimes e_k$ with $\sigma_k \geqslant 0$, and, for any vector $w \in X$, let $w_k := \langle w, e_k \rangle_X$ denote its $k$th component in that basis.

Let $(u^{(n)})_{n \in \mathbb{N}}$ be a sequence in $X$ that converges to $u \in X$. If $\liminf_{n \to \infty} I_{\mu^{(n)}}(u^{(n)}) = \infty$, then there is nothing to prove. Therefore, define $\mathcal{I} := \liminf_{n \to \infty} I_{\mu^{(n)}}(u^{(n)}) \in \mathbb{R}$. There exists a subsequence of $(u^{(n)})_{n \in \mathbb{N}}$, which for simplicity we also denote by $(u^{(n)})_{n \in \mathbb{N}}$, such that $u^{(n)} - m^{(n)} \in \operatorname{ran} A_n$ for each $n \in \mathbb{N}$ and $I_{\mu^{(n)}}(u^{(n)}) \xrightarrow[n \to \infty]{} \mathcal{I}$ (note that $I_{\mu^{(n)}}(u^{(n)}) = \infty$ unless $u^{(n)} - m^{(n)} \in \operatorname{ran} A_n$).

Now let $\varepsilon > 0$ and $v^{(n)} := A_n^{\dagger}(u^{(n)} - m^{(n)})$, $n \in \mathbb{N}$. Without loss of generality (possibly, by a further thinning of the subsequence) and using corollary 5.4, we may assume

---

[6] The statement of Dashti *et al* (2013), (lemma 3.7) may be understood as a weaker statement than needed for the $M(\mu, H(\mu))$ (lim inf in place of lim in (3.4)). However, their proof clearly shows the stronger statement as we use it here. Further, as pointed out by Wacker (2020), their proof is not valid in the Banach space setting, but works in the Hilbert space setting, which is sufficient for our purposes.

$\frac{1}{2}\|v^{(n)}\|_X^2 = I_{\mu^{(n)}}(u^{(n)}) \leqslant \mathcal{I} + \varepsilon$ for each $n \in \mathbb{N}$. Define $\mathcal{K} := \{k \in \mathbb{N} | \sigma_k > 0\}$ and the sequences $\hat{v} = (v_k)_{k \in \mathbb{N}}$ and $\lfloor \hat{v}^{(n)} \rfloor = (\lfloor v_k^{(n)} \rfloor)_{k \in \mathbb{N}}$, $n \in \mathbb{N}$, by

$$
v_k := \begin{cases} \dfrac{u_k - m_k}{\sigma_k} & \text{if } k \in \mathcal{K}, \\ 0 & \text{otherwise}, \end{cases} \qquad \lfloor v_k^{(n)} \rfloor := \begin{cases} v_k^{(n)} & \text{if } k \in \mathcal{K}, \\ 0 & \text{otherwise}. \end{cases}
$$

To prove the $\Gamma$-lim inf inequality, and with corollary 5.4 in mind, we must show that

(a) $\hat{v} \in \ell^2$ and therefore $v := \sum_{k \in \mathbb{N}} v_k e_k \in X$;
(b) $Av = u - m$ and therefore $u - m \in \operatorname{ran} A$;
(c) $\frac{1}{2}\|v\|_X^2 \leqslant \mathcal{I} + \varepsilon$ and therefore

$$
I_\mu(u) = \frac{1}{2}\|C^{\dagger/2}(u - m)\|_X^2 \leqslant \frac{1}{2}\|v\|_X^2 \leqslant \mathcal{I} = \liminf_{n \to \infty} I_{\mu^{(n)}}(u^{(n)}),
$$

by using the fact that $\varepsilon > 0$ is arbitrary and by using remark 3.11.

Since $\|A_n - A\| \to 0$, $\|u^{(n)} - u\|_X \to 0$ and $\|v^{(n)}\|_X \leqslant M_\varepsilon := \sqrt{2\mathcal{I} + 2\varepsilon}$ for all $n \in \mathbb{N}$, we obtain

$$
\begin{aligned}
\|u - m - Av^{(n)}\|_X &\leqslant \|(u - m) - (u^{(n)} - m^{(n)})\|_X + \|u^{(n)} - m^{(n)} - Av^{(n)}\|_X \\
&= \|u - u^{(n)}\|_X + \|m - m^{(n)}\|_X + \|(A_n - A)v^{(n)}\|_X \\
&\leqslant \|u - u^{(n)}\|_X + \|m - m^{(n)}\|_X + \|A_n - A\|\|v^{(n)}\|_X \\
&\xrightarrow[n \to \infty]{} 0.
\end{aligned}
$$

The above convergence implies componentwise convergence: $|u_k - m_k - \sigma_k v_k^{(n)}| \xrightarrow[n \to \infty]{} 0$ for each $k \in \mathbb{N}$ or, equivalently, $\lfloor v_k^{(n)} \rfloor \xrightarrow[n \to \infty]{} v_k$ for each $k \in \mathbb{N}$ and $u_k - m_k = 0$ for all $k \notin \mathcal{K}$. Since, for each $n \in \mathbb{N}$, $\lfloor \hat{v}^{(n)} \rfloor \in \ell^2$ with $\|\lfloor \hat{v}^{(n)} \rfloor\|_{\ell^2} \leqslant \|v^{(n)}\|_X \leqslant M_\varepsilon$, lemma B.6 implies that $\hat{v} \in \ell^2$ and $v \in X$ with $\|\hat{v}\|_{\ell^2} = \|v\|_X \leqslant M_\varepsilon$, proving (a) and (c). Since $u_k - m_k = 0$ for all $k \notin \mathcal{K}$, we obtain $Av = \sum_{k \in \mathcal{K}} (u_k - m_k)e_k = \sum_{k \in \mathbb{N}} (u_k - m_k)e_k = u - m$, proving (b) and finalising the proof of the $\Gamma$-lim inf inequality.

For the $\Gamma$-lim sup inequality, first note that, if $u - m \notin \operatorname{ran} A$, then $I_\mu(u) = \infty$, and there is nothing to prove since we may choose $u^{(n)} := u$ for all $n \in \mathbb{N}$. If $u - m \in \operatorname{ran} A$, let $v := A^\dagger(u - m)$ and $u^{(n)} := m^{(n)} + A_n v$. Then $\|u^{(n)} - u\|_X \leqslant \|m^{(n)} - m\|_X + \|A_n - A\|\|v\|_X \xrightarrow[n \to \infty]{} 0$. Since $v$ is some solution of $A_n x = u^{(n)} - m^{(n)}$ and $A_n^\dagger(u^{(n)} - m^{(n)})$ is its minimum norm solution (cf remark 3.11), corollary 5.4 implies

$$
I_\mu(u) = \frac{1}{2}\|v\|_X^2 \geqslant \frac{1}{2}\|A_n^\dagger(u^{(n)} - m^{(n)})\|_X^2 = I_{\mu^{(n)}}(u^{(n)}),
$$

for each $n \in \mathbb{N}$, finalising the proof of the $\Gamma$-lim sup inequality.

In order to prove equicoercivity of the sequence $(I_{\mu^{(n)}})_{n \in \mathbb{N}}$, let $t \in \mathbb{R}$ and

$$
K_t := \bigcup_{n \in \mathbb{N}} K_t^{(n)},
$$

$$
K_t^{(n)} := I_{\mu^{(n)}}^{-1}([-\infty, t]) = \left\{ u \in m^{(n)} + \operatorname{ran} A_n \;\middle|\; \frac{1}{2}\|A_n^\dagger(u - m^{(n)})\|_X^2 \leqslant t \right\} = m^{(n)} + A_n \overline{B}_{\sqrt{2t}}(0),
$$

where we used corollary 5.4. We will now show that $K_t$ is (sequentially) precompact. To this end, let $(u^{(\nu)})_{\nu \in \mathbb{N}}$ be a sequence in $K_t$. If $u^{(\nu)} \in K_t^{(n)}$ infinitely often for some $n \in \mathbb{N}$, there is nothing to prove, since $A_n$ is a compact operator for each $n \in \mathbb{N}$. Otherwise, there exist subsequences $(u^{(\nu_j)})_{j \in \mathbb{N}}$ and $(\mu^{(n_j)})_{j \in \mathbb{N}}$ such that $u^{(\nu_j)} \in K_t^{(n_j)}$ for each $j \in \mathbb{N}$. Hence, $u^{(\nu_j)} - m^{(n_j)} \in \operatorname{ran} A_{n_j}$ for each $j \in \mathbb{N}$ and the points $v^{(j)} := A_{n_j}^{\dagger}(u^{(\nu_j)} - m^{(n_j)})$ are uniformly bounded, $\|v^{(j)}\|_X \leqslant \sqrt{2t}$ for $j \in \mathbb{N}$. Since $A$ is a compact operator, the sequence $(w^{(j)})_{j \in \mathbb{N}}$ given by $w^{(j)} := Av^{(j)}$ has a subsequence that converges to some element $w \in X$. For simplicity, we denote this subsequence by $(w^{(j)})_{j \in \mathbb{N}}$. It follows that

$$
\|u^{(\nu_j)} - m - w\|_X \leqslant \|u^{(\nu_j)} - m^{(n_j)} - w^{(j)}\|_X + \|m^{(n_j)} - m\|_X + \|w^{(j)} - w\|_X
$$

$$
\leqslant \underbrace{\|A_{n_j} - A\|}_{\to 0} \underbrace{\|v^{(j)}\|_X}_{\leqslant \sqrt{2t}} + \underbrace{\|m^{(n_j)} - m\|_X}_{\to 0} + \underbrace{\|w^{(j)} - w\|_X}_{\to 0}
$$

$$
\xrightarrow[j \to \infty]{} 0,
$$

and so $(u^{(\nu)})_{\nu \in \mathbb{N}}$ has a convergent subsequence. Hence, $\overline{K}_t$ is compact with $I_{\mu^{(n)}}^{-1}([-\infty, t]) \subseteq \overline{K}_t$ for each $n \in \mathbb{N}$, finalising the proof of equicoercivity. $\qquad\square$

The following corollary is a direct consequence of theorems 4.2 and 5.5:

**Corollary 5.7.** *Let $X$, $\mu$, $(\mu^{(n)})_{n \in \mathbb{N}}$ be as in theorem 5.5. If $u^{(n)}$ is a global weak mode of $\mu^{(n)}$, $n \in \mathbb{N}$, then every convergent subsequence of $(u^{(n)})_{n \in \mathbb{N}}$ has as its limit a global weak mode of $\mu$.*

### 5.2. $B_1^s$-Besov measures

We now establish analogous results to those of the previous section for the class of Besov-1 measures. Besov-1 measures and Gaussian measures on infinite-dimensional spaces are analogous to Laplace distributions and normal distributions on $\mathbb{R}$. Besov-1 measures have been used as sparsity-promoting or edge-preserving priors[7] in inverse problems (Agapiou *et al* 2018, Dashti *et al* 2012, Lassas *et al* 2009).

Throughout this subsection, we use the following notation[8]:

**Assumption 5.8.** Let $s \in \mathbb{R}$, $d \in \mathbb{N}$, $\eta > 0$, $t := s - d(1 + \eta)$ and assume that $\tau := (s/d + 1/2)^{-1} > 0$. The parameter $s$ is thought of as a 'smoothness parameter' and $d$ as a 'spatial dimension'. Define $\gamma_0 := 1$ and $\gamma, \delta \in \mathbb{R}^{\mathbb{N}}$ by

$$
\gamma_k := k^{1 - \frac{1}{\tau}}, \qquad \delta_k := k^{2 + \eta - \frac{1}{\tau}}, \quad k \in \mathbb{N},
$$

and let $\mu_k \in \mathcal{P}(\mathbb{R})$ for $k \in \mathbb{N} \cup \{0\}$ have the Lebesgue density

$$
\frac{\mathrm{d}\mu_k}{\mathrm{d}u}(u) = Z_1 \, \gamma_k^{-1} \exp(-|u/\gamma_k|), \quad Z_1 := \left( \int_{\mathbb{R}} \exp(-|x|) \, \mathrm{d}x \right)^{-1} = \frac{1}{2}.
$$

---

[7] Strictly speaking, regularisation using the Besov-1 norm promotes edge-preservation for the MAP estimator but not for samples from the full posterior distribution.

[8] Typically, Besov measures are introduced on the space $L^2(\mathbb{T}^d)$; the same construction that we use for the components of a random sequence in $\mathbb{R}^{\mathbb{N}}$ is used for the components of a random Fourier or wavelet expansion in $L^2(\mathbb{T}^d)$. In our definition, the dimension $d$ becomes superfluous and one could work with $\tilde{s} := s/d$, but we continue to use the classical notation in order to reduce confusion.

We define the Besov measure $B_1^s$ as follows, using notation that is an adaptation of that of Dashti *et al* (2012) and Agapiou *et al* (2018).

**Definition 5.9 (Sequence space Besov measures and Besov spaces).** Using assumption 5.8, we call $\mu := \bigotimes_{k \in \mathbb{N}} \mu_k$ a (**sequence space**) **Besov measure** on $\mathbb{R}^{\mathbb{N}}$ and write $B_1^s := \mu$. The corresponding **Besov space** is the weighted sequence space $(X_1^s, \| \cdot \|_{X_1^s}) := (\ell_\gamma^1, \| \cdot \|_{\ell_\gamma^1})$.

Since it is the summability parameter $p$ (in this case, $p = 1$) that most strongly affects the qualitative properties of the measure, we often refer simply to a 'Besov-1 measure' (or Besov-$p$ measure in the general case) for any measure in the above class, regardless of the values of $s$, $d$, etc.

**Lemma 5.10.** *Let $\mu = B_1^s$ be the Besov measure defined above and $X = X_1^t = \ell_\delta^1$. Then $\mu(X) = 1$.*

**Proof.** This is a restatement of Lassas *et al* (2009), (lemma 2) for particular case $p = 1$. □

From now on we will consider the Besov measure $\mu = B_1^s$ on the normed spaces $X = X_1^t = \ell_\delta^1$. This is possible since, by Ayanbayev *et al* (2021), (lemma B.1), $\mathcal{B}(\ell_\delta^1) \subseteq \mathcal{B}(\mathbb{R}^{\mathbb{N}})$, where we consider the product topology on $\mathbb{R}^{\mathbb{N}}$.

**Proposition 5.11.** *Let $\mu = B_1^s$ be a $B_1^s$-Besov measure on the space $X = X_1^t = \ell_\delta^1$. Then, for $E = X_1^s = \ell_\gamma^1$, property $M(\mu, E)$ is satisfied and the OM functional $I_\mu : X \to \overline{\mathbb{R}}$ of $\mu$ is given by*

$$I_\mu(u) = \begin{cases} \|u\|_{X_1^s} & \text{for } u \in E, \\ \infty & \text{otherwise.} \end{cases} \tag{5.3}$$

**Proof.** The OM functional formula on $E$ follows from Agapiou *et al* (2018), (theorem 3.9), while property $M(\mu, E)$ follows from Ayanbayev *et al* (2021), (theorem 4.9). The assumptions of this theorem are fulfiled, given definition 5.9 and lemma 5.10. □

**Remark 5.12.** Proposition 5.11 uses and extends (Agapiou *et al* 2018, theorem 3.9). The authors write that 'the space $B_1^s(\mathbb{T}^d)$ here, is the largest space on which the OM functional is defined'. This claim is intuitively true, since $\|h\|_{X_1^s} = +\infty$ if $h \notin X_1^s = E$, and in our notation $X_1^s$ corresponds to $B_1^s(\mathbb{T}^d)$. However, one must not *a priori* exclude the possibility that $I_\mu$ can have a different formula outside of $E$. Property $M(\mu, E)$ in the above proof is one way to guarantee that the claim is true.

We now give a $\Gamma$-convergence and equicoercivity result for sequences of Besov-1 measures with converging smoothness parameters.

**Theorem 5.13 ($\Gamma$-convergence and equicoercivity of OM functionals for Besov-1 measures).** *Let $\mu^{(n)} := B_1^{s^{(n)}}$, $n \in \mathbb{N} \cup \{+\infty\}$, be centred Besov measures such that $s^{(n)} \to s^{(\infty)}$. Then there exists $n_0 \in \mathbb{N}$ such that, for each $n \geqslant n_0$, $\mu^{(n)}(\ell_{\delta^{(\infty)}}^1) = 1$ and we therefore consider these measures on $X = X_1^{t^{(\infty)}} = \ell_{\delta^{(\infty)}}^1$ (after dropping the first $n_0 - 1$ measures). Then, for the OM functionals $I_{\mu^{(n)}} = \| \cdot \|_{X_1^{s^{(n)}}} : X \to \overline{\mathbb{R}}$, $n \in \{n' \in \mathbb{N} | n' \geqslant n_0\} \cup \{+\infty\}$, given by (5.3), the sequence $(I_{\mu^{(n)}})_{n \geqslant n_0}$ is equicoercive and $I_{\mu^{(\infty)}} = \Gamma\text{-lim}_{n \to \infty} I_{\mu^{(n)}}$.*

**Proof.** Since $s^{(n)} \to s^{(\infty)}$ there exists $n_0 \in \mathbb{N}$ such that, for $n \geqslant n_0$, $|s^{(n)} - s^{(\infty)}| \leqslant \frac{d\eta^{(\infty)}}{2}$. Therefore, for $n \geqslant n_0$, we may choose $\eta^{(n)} > 0$ such that $t^{(n)} = s^{(n)} - d(1 + \eta^{(n)}) = s^{(\infty)} - d(1 + \eta^{(\infty)}) = t^{(\infty)}$ and consider $\mu^{(n)}$ as a measure on $X = X_1^{t^{(\infty)}} = \ell_{\delta^{(\infty)}}^1$ by lemma 5.10. Without loss of generality, we assume $n_0 = 1$ from now on in order to simplify notation. Since $s^{(n)} \geqslant \bar{s} := s^{(\infty)} - \frac{d\eta^{(\infty)}}{2}$,

$$\gamma_k^{(n)} = k^{-\frac{s^{(n)}}{d} + \frac{1}{2}} \leqslant k^{-\frac{\bar{s}}{d} + \frac{1}{2}} =: \overline{\gamma}_k, \quad k, n \in \mathbb{N},$$

and, for any $\theta \geqslant 0$ and $n \in \mathbb{N}$,

$$I_{\mu^{(n)}}^{-1}([-\infty, \theta]) := \left\{ u \in X \,\bigg|\, \sum_{k \in \mathbb{N}} \frac{|u_k|}{\gamma_k^{(n)}} \leqslant \theta \right\} \subseteq \prod_{k \in \mathbb{N}} [-\overline{\gamma}_k \theta, \overline{\gamma}_k \theta] =: K_\theta.$$

We will now show that $K_\theta \subseteq X$ is precompact. For this purpose, we define the operators

$$T, T_m : \ell^\infty \to X = \ell_{\delta^{(\infty)}}^1, \qquad T(x) = (\overline{\gamma}_k x_k)_{k \in \mathbb{N}},$$
$$T_m(x) = (\overline{\gamma}_k x_k)_{k=1,\dots,m}, \qquad m \in \mathbb{N}.$$

All $T_m$ are finite-rank operators that converge to $T$ in the operator norm:

$$\|T_m - T\| = \sup_{\|x\|_{\ell^\infty} \leqslant 1} \sum_{k > m} \frac{\overline{\gamma}_k}{\delta_k^{(\infty)}} |x_k| \leqslant \sum_{k > m} k^{-1 - \frac{\eta^{(\infty)}}{2}} \xrightarrow[m \to \infty]{} 0.$$

Therefore, $T$ is a compact operator and $K_\theta = \theta \, T B_1^{\ell^\infty}(0)$ is precompact, finalising the proof of equicoercivity. Note that for $\theta < 0$ there is nothing to prove, since $I_{\mu^{(n)}}^{-1}([-\infty, \theta]) = \varnothing$ for each $n \in \mathbb{N} \cup \{\infty\}$ in this case.

In order to prove the $\Gamma$-convergence statement, we will first show that $\|\gamma^{(n)} - \gamma^{(\infty)}\|_X \to 0$. Since the sequences $a^{(n)} := (k^{-1-\eta} |k^{\frac{s^{(\infty)} - s^{(n)}}{d}} - 1|)_{k \in \mathbb{N}}$ are uniformly bounded by the summable sequence $a = (2k^{-1 - \frac{\eta^{(\infty)}}{2}})_{k \in \mathbb{N}}$ (where we used $|s^{(n)} - s^{(\infty)}| \leqslant \frac{d\eta^{(\infty)}}{2}$), the reverse Fatou lemma implies $\|\gamma^{(n)} - \gamma^{(\infty)}\|_X \to 0$ via

$$\limsup_{n \to \infty} \|\gamma^{(n)} - \gamma^{(\infty)}\|_X = \limsup_{n \to \infty} \sum_{k \in \mathbb{N}} k^{-1-\eta} \left| k^{\frac{s^{(\infty)} - s^{(n)}}{d}} - 1 \right|$$

$$\leqslant \sum_{k \in \mathbb{N}} \limsup_{n \to \infty} k^{-1-\eta} \left| k^{\frac{s^{(\infty)} - s^{(n)}}{d}} - 1 \right|$$

$$= 0.$$

For the $\Gamma$-$\liminf$ inequality, it follows from $\|u^{(n)} - u^{(\infty)}\|_X \to 0$ and $\|\gamma^{(n)} - \gamma^{(\infty)}\|_X \to 0$ that $\frac{u_k^{(n)}}{\gamma_k^{(n)}} \to \frac{u_k^{(\infty)}}{\gamma_k^{(\infty)}}$ for all $k \in \mathbb{N}$. Thus, by Fatou's lemma,

$$I_{\mu^{(\infty)}}(u^{(\infty)}) = \|u^{(\infty)}\|_{\ell^1_{\gamma^{(\infty)}}}$$

$$= \sum_{k \in \mathbb{N}} \frac{|u_k^{(\infty)}|}{\gamma_k^{(\infty)}}$$

$$= \sum_{k \in \mathbb{N}} \liminf_{n \to \infty} \frac{|u_k^{(n)}|}{\gamma_k^{(n)}}$$

$$\leqslant \liminf_{n \to \infty} \sum_{k \in \mathbb{N}} \frac{|u_k^{(n)}|}{\gamma_k^{(n)}}$$

$$= \liminf_{n \to \infty} I_{\mu^{(n)}}(u^{(n)}).$$

For the $\Gamma$-lim sup inequality, note that, if $I_{\mu^{(\infty)}}(u^{(\infty)}) = \infty$, then there is nothing to prove. Therefore, let us assume that $I_{\mu^{(\infty)}}(u^{(\infty)}) < \infty$, and define $u^{(n)}$ by $u_k^{(n)} = \gamma_k^{(n)} \frac{u_k^{(\infty)}}{\gamma_k^{(\infty)}}$, $k \in \mathbb{N}$. Then

$$I_{\mu^{(n)}}(u^{(n)}) = \sum_{k \in \mathbb{N}} \frac{|u_k^{(n)}|}{\gamma_k^{(n)}} = \sum_{k \in \mathbb{N}} \frac{|u_k^{(\infty)}|}{\gamma_k^{(\infty)}} = I_{\mu^{(\infty)}}(u^{(\infty)}) < \infty,$$

and $\limsup_{n \to \infty} I_{\mu^{(n)}}(u^{(n)}) \leqslant I_{\mu^{(\infty)}}(u^{(\infty)})$. Additionally,

$$\|u^{(n)} - u^{(\infty)}\|_X = \sum_{k \in \mathbb{N}} \frac{|u_k^{(n)} - u_k^{(\infty)}|}{\delta_k^{(\infty)}}$$

$$= \sum_{k \in \mathbb{N}} \frac{|u_k^{(\infty)}|}{\gamma_k^{(\infty)}} \frac{\left|\gamma_k^{(n)} - \gamma_k^{(\infty)}\right|}{\delta_k^{(\infty)}}$$

$$\leqslant I_{\mu^{(\infty)}}(u^{(\infty)}) \|\gamma^{(n)} - \gamma^{(\infty)}\|_X \to 0 \quad \text{as } n \to \infty,$$

finalising the proof of the $\Gamma$-lim sup inequality. $\qquad\square$

The following corollary is now a direct consequence of theorems 4.2 and 5.13:

**Corollary 5.14.** *Let $X$, $(\mu^{(n)})_{n \in \mathbb{N} \cup \{\infty\}}$ be as in theorem* 5.13. *If, for each $n \in \mathbb{N}$, $u^{(n)}$ is a global weak mode of $\mu^{(n)}$, then every convergent subsequence of $(u^{(n)})_{n \in \mathbb{N}}$ has as its limit a global weak mode of $\mu^{(\infty)}$.*

## 6. Consequences for MAP estimation in BIPs

The $\Gamma$-convergence theory of OM functionals described in section 5 has important consequences for the stability of MAP estimators of BIPs, in particular those BIPs that use the probability measures considered above as prior distributions.

An inverse problem consists of the recovery of an unknown $u$ from related observational data $y$. In the Bayesian approach to inverse problems (Kaipio and Somersalo 2005, Stuart 2010), these two objects are treated as coupled random variables $\boldsymbol{u}$ and $\boldsymbol{y}$ that take values in spaces $X$ and $Y$ respectively. *A priori* knowledge about $\boldsymbol{u}$ is represented by a prior probability measure $\mu_0 \in \mathcal{P}(X)$ and one is given access to a realisation $y$ of $\boldsymbol{y}$. The *solution* of the BIP is, by definition, the posterior probability measure $\mu^y \in \mathcal{P}(X)$, i.e. the conditional distribution of $\boldsymbol{u}$

given that $\boldsymbol{y} = y$. For the sake of space, we omit here all technical discussion of the existence and regularity of this conditional distribution and focus exclusively on the case that $\mu^y$ has a Radon–Nikodym derivative with respect to $\mu_0$ of the form

$$\mu^y(\mathrm{d}u) \propto \exp(-\Phi(u; y))\mu_0(\mathrm{d}u),$$

for some $\Phi : X \times Y \to \mathbb{R}$. The function $\Phi$, often called the potential, encodes both the idealised relationship between the unknown and the data and statistical assumptions about any observational noise. The textbook example is that $X$ is a separable Hilbert or Banach space of functions, $Y = \mathbb{R}^J$ for some $J \in \mathbb{N}$, and that $\boldsymbol{y} = \mathcal{O}(\boldsymbol{u}) + \boldsymbol{\eta}$ for some deterministic observation map $\mathcal{O} : X \to Y$ and additive non-degenerate Gaussian noise $\boldsymbol{\eta} \sim \mathcal{N}(0, C_{\boldsymbol{\eta}})$ that is *a priori* independent of $\boldsymbol{u}$, in which case $\Phi$ is the familiar quadratic misfit

$$\Phi(u; y) = \frac{1}{2} \left\| C_{\boldsymbol{\eta}}^{-1/2}(y - \mathcal{O}(u)) \right\|^2.$$

One convenient point summary of $\mu^y$ is a MAP estimator, i.e. a point of maximum probability under $\mu^y$ in the sense of a maximiser of a small ball probability. Under the conditions laid out in section 4, these points (in the sense of global weak modes) are the minimisers of the OM functional of $\mu^y$. However, we note that there are many problems of interest for which a more generalised notion of MAP estimator and a correspondingly generalised OM functional are needed, particularly problems in which the prior may have bounded support or the potential may take the value $+\infty$ (Clason *et al* 2019).

Our interest lies in assessing the stability of $\mu^y$ (more precisely, the stability of the MAP estimators of $\mu^y$) in response to the following:

- Perturbations of the observed data $y$, to be reassured that the posterior is not unduly sensitive to observational errors;
- Perturbations of the potential[9] $\Phi$, for example to be reassured that the posterior is not unduly sensitive to numerical approximation of $\mathcal{O}$ by some $\mathcal{O}^{(n)}$ (e.g. using a finite element solver to solve a partial differential equation), or to examine the small-noise limit $C_{\boldsymbol{\eta}} \to 0$;
- Perturbations of the prior $\mu_0$, to be reassured that the posterior is not unduly sensitive to prior assumptions, e.g. relating to the regularity of $u$.

We propose to address this question using the $\Gamma$-convergence results of the previous section. The classes of measures for whose OM functionals explicit $\Gamma$-limits were computed in section 5 will serve here as Bayesian prior measures.

Our main result concerns the transfer of convergence properties of sequences of prior OM functionals and sequences of potentials to the convergence of posterior OM functionals.

**Theorem 6.1 (Transfer of property *M*, $\Gamma$-convergence, equicoercivity, and MAP estimators).**     *Let $X$ be a metric space. For each $n \in \mathbb{N} \cup \{\infty\}$, let $\mu_0^{(n)} \in \mathcal{P}(X)$ and let $\Phi^{(n)} : X \to \mathbb{R}$ be locally uniformly continuous. Suppose that, for each $n \in \mathbb{N} \cup \{\infty\}$, $Z^{(n)} := \int_X e^{-\Phi^{(n)}(x)} \mu_0^{(n)}(\mathrm{d}x) \in (0, \infty)$ and set*

$$\mu^{(n)}(\mathrm{d}x) := \frac{1}{Z^{(n)}} e^{-\Phi^{(n)}(x)} \mu_0^{(n)}(\mathrm{d}x).$$

---

[9] Of course, a perturbation of the data $y$ induces a perturbation of $\Phi(\cdot; y)$. Sometimes it is easier to consider data perturbations and potential perturbations separately, and sometimes, as we do in theorem 6.1, it is simpler to consider them both as perturbations of the potential.

*Suppose that each $\mu_0^{(n)}$ has an OM functional $I_0^{(n)} : E^{(n)} \to \mathbb{R}$. Then the following statements hold:*

(a) *Each $\mu^{(n)}$ has $I^{(n)} := \Phi^{(n)} + I_0^{(n)} : E^{(n)} \to \mathbb{R}$ as an OM functional.*

(b) *Suppose that property $M(\mu_0^{(n)}, E^{(n)})$ holds. Then property $M(\mu^{(n)}, E^{(n)})$ also holds, and the global weak modes of $\mu_0^{(n)}$ (resp. of $\mu^{(n)}$) are the global minimisers of the extended OM functional $I_0^{(n)} : X \to \overline{\mathbb{R}}$ (resp. of $I^{(n)} : X \to \overline{\mathbb{R}}$).*

(c) *Suppose that $I_0^{(n)} \xrightarrow{\Gamma} I_0^{(\infty)}$ and $\Phi^{(n)} \to \Phi^{(\infty)}$ continuously[10] as $n \to \infty$. Then the OM functionals $I^{(n)}$ satisfy*

$$\Gamma\text{-}\lim_{n\to\infty} I^{(n)} = I^{(\infty)},$$

(d) *Suppose that the sequence $(I_0^{(n)})_{n\in\mathbb{N}}$ is equicoercive and the functions $\Phi^{(n)} \geqslant M$ are uniformly bounded from below by some constant $M \in \mathbb{R}$. Then the sequence $(I^{(n)})_{n\in\mathbb{N}}$ is also equicoercive with respect to the same representatives of $I^{(n)}$ as for the $\Gamma$-convergence (cf remark 3.4).*

(e) *Suppose that the assumptions of parts (b)–(d) all hold. Then the cluster points as $n \to \infty$ of the global weak modes of the posteriors $\mu^{(n)}$ are the global weak modes of the limiting posterior $\mu^{(\infty)}$.*

**Proof.** Parts (a) and (b) follow from lemma B.8, and part (c) follows from proposition A.4 i.e. Dal Maso (1993), (proposition 6.20).

For part (d), let $(I_0^{(n)})_{n\in\mathbb{N}}$ be equicoercive and $\Phi^{(n)} \geqslant M$ be uniformly bounded from below. Then, for any $t \in \mathbb{R}$, there exists a compact $K_t \subseteq X$ such that, for all $n \in \mathbb{N}$, $(I_0^{(n)})^{-1}([-\infty, t]) \subseteq K_t$. Since $I^{(n)}(x) = I_0^{(n)}(x) + \Phi^{(n)}(x) \leqslant t$ implies $I_0^{(n)}(x) \leqslant t - M$, it follows that, for any $t \in \mathbb{R}$ and $n \in \mathbb{N}$, $(I^{(n)})^{-1}([-\infty, t]) \subseteq K_{t-M}$.

Finally, part (e) is just a restatement of theorem 4.2. $\qquad\square$

**Remark 6.2.** Loosely speaking, the hypothesis in theorem 6.1(d) that the potentials are uniformly bounded below corresponds to a likelihood model in which the observed data are (uniformly) finite dimensional. BIPs with infinite-dimensional data are known to involve potentials that are unbounded below. Such potentials cannot be interpreted as (non-negative) misfit functionals, as discussed by e.g. Stuart (2010), (remark 3.8) and Kasanický and Mandel (2017).

Note also that a standing assumption of Dashti *et al* (2013) is that $\Phi$ is locally Lipschitz continuous, which is stronger than the local uniform continuity assumed in theorem 6.1, and that boundedness of $\Phi$ from below is also assumed by Dashti *et al* (2013), (theorem 3.5), just as in the hypothesis of theorem 6.1(d).

**Corollary 6.3.** *Consider a BIP with prior $\mu_0 = \mu_0^{(\infty)}$, potential $\Phi = \Phi^{(\infty)}$, and observed data $y = y^{(\infty)}$, each of which may now be approximated. In addition to the assumptions of theorem 6.1, assume for simplicity that the OM functional of $\mu_0$ is lower semicontinuous, so that it equals its own $\Gamma$-limit (theorem A.5).*

(a) *If the data $y =: y^{(\infty)}$ are approximated by a sequence $(y^{(n)})_{n\in\mathbb{N}}$ and the potential $\Phi$ and prior $\mu_0$ are held constant, then continuous convergence of $\Phi^{(n)} := \Phi(\cdot; y^{(n)})$ to*

---

[10] See definition A.2 for the definition of continuous convergence. Note that proposition A.4 is agnostic as to which of the two summands converges continuously, and so theorem 6.1(c) also holds if $I_0^{(n)} \to I_0^{(\infty)}$ continuously and $\Phi^{(n)} \xrightarrow{\Gamma} \Phi^{(\infty)}$, which would be a weaker hypothesis on the potentials but a stronger one on the prior OM functionals. However, since we have not studied the continuous convergence of prior OM functionals, we do not stress this version of the theorem.

$\Phi^{(\infty)} := \Phi(\cdot; y^{(\infty)})$ *ensures* $\Gamma$*-convergent and equicoercive sequences of posterior OM functionals and convergent sequences of MAP estimators* (*in the sense of global weak modes, and up to subsequences*).

(*b*) *Similarly, if the data and potential are held constant and the prior* $\mu_0 =: \mu_0^{(\infty)}$ *is approximated by a sequence of priors* $(\mu_0^{(n)})_{n \in \mathbb{N}}$, *then* $\Gamma$*-convergence of prior OM functionals, i.e.* $I_{\mu_0^{(n)}} \xrightarrow{\Gamma} I_{\mu_0^{(\infty)}}$, *yields convergent sequences of MAP estimators.*

(*c*) *Finally, if the data and prior are held constant and the potential* $\Phi =: \Phi^{(\infty)}$ *is approximated by a sequence of potentials* $(\Phi^{(n)})_{n \in \mathbb{N}}$, *then continuous convergence* $\Phi^{(n)}(\cdot; y) \to \Phi^{(\infty)}(\cdot; y)$ *yields convergent sequences of MAP estimators; in particular, this holds when the approximate misfit/potential* $\Phi^{(n)}$ *arises through projection, e.g. Galerkin discretisation* (*lemma* B.9).

**Example 6.4 (Small-noise limits).** Regrettably, the analysis of MAP estimators of small-noise (infinite-precision) limits is not entirely trivial even under the $\Gamma$-convergence theory that we have outlined. Consider a BIP on $X$ with prior $\mu_0$ and potential $\Phi$. Assume that $\mu_0$ has OM functional $I_0 : E \to \mathbb{R}$ that satisfies $M(\mu_0, E)$, leading to a lower semi-continuous and coercive extended OM functional $I_0 : X \to \overline{\mathbb{R}}$. Assume also that $\Phi$ is locally uniformly continuous, is bounded below, and attains its lower bound—without loss of generality, take this minimal value to be 0. Now consider the posterior

$$\mu^{(n)}(\mathrm{d}x) := \frac{1}{Z^{(n)}} e^{-n\Phi(x)} \mu_0(\mathrm{d}x),$$

in the small-noise limit $n \to \infty$. By theorem 6.1, $\mu^{(n)}$ has OM functional $I^{(n)} = n\Phi + I_0$. It is easy to see that, pointwise,

$$\lim_{n \to \infty} I^{(n)}(x) = I^{(\infty)}(x) := \begin{cases} I_0(x), & \text{if } \Phi(x) = 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

It is natural to hope that $\Gamma\text{-}\lim_{n \to \infty} I^{(n)} = I^{(\infty)}$ as well, and hence that the MAP estimators of $\mu^{(n)}$ converge, in the small-noise limit $n \to \infty$, to the constrained minimisers of the prior OM functional $I_0$ among the global minima of $\Phi$. However, this $\Gamma$-convergence is not straightforward to establish.

- For the $\Gamma\text{-}\lim\sup$ inequality, choose any $x \in X$. Consider first the case that $\Phi(x) > 0$: for the recovery sequence $x_n \equiv x$,

$$I^{(n)}(x_n) = n\Phi(x) + I_0(x) \xrightarrow[n \to \infty]{} +\infty = I^{(\infty)}(x).$$

  Similarly, in the case $\Phi(x) = 0$, we may use the same recovery sequence to obtain $I^{(n)}(x_n) = I^{(n)}(x) = I_0(x) = I^{(\infty)}(x)$.

- For the $\Gamma\text{-}\lim\inf$ inequality, choose any $x \in X$ and any sequence $x_n \to x$. Taking $\omega_{\Phi,x}$ to be a local modulus of continuity for $\Phi$ near $x$, we have

$$|\Phi(x_n) - \Phi(x)| \leqslant \omega_{\Phi,x}(\|x_n - x\|),$$

and hence

$$
\begin{aligned}
\liminf_{n\to\infty} I^{(n)}(x_n) &= \liminf_{n\to\infty} (n\Phi(x_n) + I_0(x_n)) \\
&\geqslant \liminf_{n\to\infty} \left(n\Phi(x) - n\omega_{\Phi,x}(\|x_n - x\|) + I_0(x_n)\right) \\
&\geqslant \lim_{n\to\infty} n\,\Phi(x) - \liminf_{n\to\infty} n\omega_{\Phi,x}(\|x_n - x\|) + I_0(x),
\end{aligned}
$$

where the last inequality uses the lower semicontinuity of $I_0$. At this point we encounter a problem. For $x$ such that $\Phi(x) > 0$, the right-hand side of the above display is indeed $+\infty$, as required. However, for $x$ such that $\Phi(x) = 0$, the $\Gamma$-lim inf inequality only holds if $\liminf_{n\to\infty} n\omega_{\Phi,x}(\|x_n - x\|) = 0$, and this holds only if $x_n$ converges sufficiently rapidly to $x$, which is not at all guaranteed.

We close this section by repeating the observation made at the end of section 4, that the necessity of the continuous convergence/$\Gamma$-convergence assumptions, as opposed to simple pointwise convergence of densities or OM functionals, is shown by example 1.1(c) from the introduction, which can easily be interpreted as a pointwise but not continuously convergent sequence of likelihoods/potentials and a Gaussian prior.

## 7. Closing remarks

The purpose of this paper was to establish a convergence theory for modes of probability measures (in the BIP setting, MAP estimators of Bayesian posterior measures) in the sense of maximisers of small ball probabilities, by first characterising them as minimisers of OM functionals and then using the well-established notion of $\Gamma$-convergence from the calculus of variations. The correspondence between modes and OM minimisers was established rigorously for global weak modes under the abstract $M$-property, and counterexamples were given to show that an extension to strong modes and relaxation of the $M$-property would be non-trivial if not impossible. The general programme of studying $\Gamma$-limits of OM functionals of measures was illustrated via two explicit example classes that are frequently used in the inverse problems literature, namely Gaussian measures and Besov $B_p^s$ measures with integrability parameter $p = 1$.

The Gaussian and Besov-1 measures treated in this paper are merely simple examples of a general class of measures, namely countable products of scaled copies of a measure on $\mathbb{R}$ (the normal and Laplace distributions respectively). General Besov-$p$ measures and infinite-product Cauchy measures fall into this class. Part II of this paper (Ayanbayev *et al* 2021) treats this class in a high degree of generality, following the same programme of determining the OM functional, verifying the $M$-property, and showing $\Gamma$-convergence and equicoercivity. The advantage of having considered the Gaussian and Besov-1 measures separately in this paper is that the requisite calculations could be done in more-or-less closed form and with much less notational overhead than the general case.

This work has made extensive use of the hypothesis that some measure $\mu$ of interest actually possesses an OM functional (and moreover one that satisfies property $M(\mu, E)$ for a 'good enough' $E$), and that $\mu$ possesses a mode. However, there are examples, even in finite dimension, of $\mu$ that have no strong or global weak modes, only *generalised* modes in the sense of Clason *et al* (2019), which are associated with generalised OM functionals. A natural further

generalisation of this article would be to study the $\Gamma$-convergence properties of such generalised OM functionals, and hence the convergence of generalised strong modes.

It would be of great value in applications not only to know that some sequence of approximations to an ideal limiting MAP problem $\Gamma$-converges, but also to quantify how quickly those approximate MAP estimators converge. Unfortunately, this is not trivial, since the basic framework of $\Gamma$-convergence does not easily deliver convergence rates for minimisers, especially when the objective functions are non-smooth, as is the case for most of the OM functionals in our setting. Therefore, the interesting question of convergence rates for modes/MAP estimators must be deferred to future work.

## Acknowledgments

## Data availability statement

No new data were created or analysed in this study.

## Appendix A. $\Gamma$-convergence

We collect here the basic definitions and results related to $\Gamma$-convergence as used in the main text. Standard references on $\Gamma$-convergence include the books of Braides (2002, 2006) and Dal Maso (1993).

**Definition A.1.** Let $X$ be a metric space and suppose that $F_n, F : X \to \overline{\mathbb{R}}$. We say that $F_n$ **$\Gamma$-converges** to $F$, written $\Gamma\text{-}\lim_{n\to\infty} F_n = F$ or $F_n \xrightarrow[n\to\infty]{\Gamma} F$, if, for every $x \in X$,

(a) ($\Gamma$-lim inf inequality) for every sequence $(x_n)_{n\in\mathbb{N}}$ converging to $x$,

$$F(x) \leqslant \liminf_{n\to\infty} F_n(x_n);$$

(b) ($\Gamma$-lim sup inequality) and there exists a 'recovery sequence' $(x_n)_{n\in\mathbb{N}}$ converging to $x$ such that

$$F(x) \geqslant \limsup_{n\to\infty} F_n(x_n).$$

We say that $(F_n)_{n\in\mathbb{N}}$ is **equicoercive** if for all $t \in \mathbb{R}$, there exists a compact $K_t \subseteq X$ such that, for all $n \in \mathbb{N}$, $F_n^{-1}([-\infty, t]) \subseteq K_t$.

In general, $\Gamma$-convergence and pointwise convergence are independent of one another, although the following inequality always holds:

$$\left(F_n \xrightarrow[n\to\infty]{\Gamma} F \text{ and } F_n \xrightarrow[n\to\infty]{} G \text{ pointwise}\right) \implies F \leqslant G.$$

However, one can compare $\Gamma$-convergence with continuous convergence:

**Definition A.2.** Let $X$ be a metric space and suppose that $F_n, F : X \to \overline{\mathbb{R}}$. We say that $F_n$ **converges continuously** to $F$ if, for every $x \in X$ and every neighbourhood $V$ of $F(x)$ in $\overline{\mathbb{R}}$, there exists $N \in \mathbb{N}$ and a neighbourhood $U$ of $x$ such that

$$(n \geqslant N \text{ and } x' \in U) \implies F_n(x') \in V.$$

Continuous convergence implies both pointwise convergence and $\Gamma$-convergence and, in the case that $F$ is continuous, is implied by uniform convergence of $F_n$ to $F$ (Dal Maso 1993, chapters 4 and 5).

**Theorem A.3 (Fundamental theorem of $\Gamma$-convergence; Braides (2006), theorem 2.10).** *Let $X$ be a metric space and suppose that $F_n, F : X \to \overline{\mathbb{R}}$ are such that $\Gamma\text{-}\lim_{n\to\infty} F_n = F$ and $(F_n)_{n\in\mathbb{N}}$ is equicoercive. Then $F$ has a minimum value and $\min_X F = \lim_{n\to\infty} \inf_X F_n$. Moreover, if $(x_n)_{n\in\mathbb{N}}$ is a precompact sequence such that $\lim_{n\to\infty} F_n(x_n) = \min_X F$, then every limit of a convergent subsequence of $(x_n)_{n\in\mathbb{N}}$ is a minimiser of $F$. Thus, if each $F_n$ has a minimiser $x_n$, then every convergent subsequence of $(x_n)_{n\in\mathbb{N}}$ has as its limit a minimiser of $F$.*

**Proposition A.4 (Dal Maso 1993, proposition 6.20).** *Let $X$ be a metric space and suppose that $F_n, F : X \to \overline{\mathbb{R}}$ and $G_n, G : X \to \mathbb{R}$ are such that $F_n \xrightarrow{\Gamma} F$ on $X$ and $G_n \to G$ continuously on $X$ as $n \to \infty$. Then*

$$F_n + G_n \xrightarrow[n\to\infty]{\Gamma} F + G \quad \text{on } X.$$

**Theorem A.5 (Braides 2006, proposition 2.5).** *The $\Gamma$-limit of a constant sequence $(F)_{n\in\mathbb{N}}$ is the lower semicontinuous envelope $F^{\text{lsc}}$ of $F$, i.e. the greatest lower semicontinuous function bounded above by $F$:*

$$F^{lsc}(x) := \liminf_{x'\to x} F(x').$$

*In particular, $F = \Gamma\text{-}\lim_{n\to\infty} F$ if and only if $F$ is lower semicontinuous.*

## Appendix B. Technical supporting results

*B.1. Supporting results for section 3*

**Lemma B.1 (The *M*-property).** *Let $X$ be a metric space and let $\mu_0 \in \mathcal{P}(X)$. Suppose that $\mu_0$ has an OM functional $I : E \to \mathbb{R}$ on a nonempty subset $E \subseteq \text{supp}(\mu)$.*

  (a) *If $x^\star \in E$ satisfies (3.4), then any $x^{\star\star} \in E \setminus \{x^\star\}$ satisfies (3.4). In particular, property $M(\mu_0, E)$ does not depend on the choice of $x^\star$ in (3.4).*

(b) *Let $\Phi : X \to \mathbb{R}$ be measurable, such that $\Phi$ is bounded on bounded subsets of $X$.[11] Suppose that $Z := \int_X e^{-\Phi(x)} \mu_0(\mathrm{d}x) \in (0, \infty)$ and define $\mu \in \mathcal{P}(X)$ by $\mu(\mathrm{d}x) := Z^{-1} e^{-\Phi(x)} \mu_0(\mathrm{d}x)$. If property $M(\mu_0, E)$ holds, then so too does $M(\mu, E)$.*

(c) *If property $M(\mu_0, E)$ holds, then no point of $X \backslash E$ can be a global weak mode for $\mu_0$, and hence cannot be a strong mode for $\mu_0$.*

**Proof.** Suppose that $x^\star \in E$ satisfies (3.4). Let $x^{\star\star} \in E$ and $x \in X \backslash E$. If $x \notin \mathrm{supp}(\mu_0)$, then for sufficiently small $r$, $\mu_0(B_r(x)) = 0$, so it suffices to prove the claim for $x \in \mathrm{supp}(\mu_0) \backslash E$. Since $\mu_0$ has an OM functional $I : E \to \mathbb{R}$,

$$
\begin{aligned}
\lim_{r \searrow 0} \frac{\mu_0(B_r(x))}{\mu_0(B_r(x^{\star\star}))} &= \lim_{r \searrow 0} \frac{\mu_0(B_r(x^\star))}{\mu_0(B_r(x^{\star\star}))} \frac{\mu_0(B_r(x))}{\mu_0(B_r(x^\star))} \\
&= \lim_{r \searrow 0} \frac{\mu_0(B_r(x^\star))}{\mu_0(B_r(x^{\star\star}))} \lim_{r \searrow 0} \frac{\mu_0(B_r(x))}{\mu_0(B_r(x^\star))} \\
&= \exp(I(x^{\star\star}) - I(x^\star)) \lim_{r \searrow 0} \frac{\mu_0(B_r(x))}{\mu_0(B_r(x^\star))} \\
&= 0,
\end{aligned}
$$

where we used (3.3) and (3.4) in the penultimate and last equation. This proves (a).

For (b), suppose that property $M(\mu_0, E)$ holds with $x^\star \in E$ satisfying

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

Observe that, for any $x \in X$,

$$
\begin{aligned}
\frac{\mu(B_r(x))}{\mu(B_r(x^\star))} &= \frac{\int_{B_r(x)} \exp(-\Phi(y)) \, \mu_0(\mathrm{d}y)}{\int_{B_r(x^\star)} \exp(-\Phi(y)) \, \mu_0(\mathrm{d}y)} \\
&\leqslant \exp\left( -\inf_{B_r(x)} \Phi + \sup_{B_r(x^\star)} \Phi \right) \frac{\mu_0(B_r(x))}{\mu_0(B_r(x^\star))}.
\end{aligned}
$$

The exponential on the right-hand side is finite, by the assumption that $\Phi$ is bounded on bounded subsets of $X$. If $x \in X \backslash E$, then taking the limit as $r \searrow 0$ yields property $M(\mu, E)$, as claimed.

Suppose that $x \in X \backslash E$ is a global weak mode in the sense of definition 3.7. Then $x \in \mathrm{supp}(\mu)$ and

$$
1 \geqslant \limsup_{r \searrow 0} \frac{\mu_0(B_r(x^\star))}{\mu_0(B_r(x))} = \left( \liminf_{r \searrow 0} \frac{\mu_0(B_r(x))}{\mu_0(B_r(x^\star))} \right)^{-1}.
$$

Above, we used that $x^\star \in E \subseteq \mathrm{supp}(\mu)$ to ensure that for every $r > 0$, $\frac{\mu_0(B_r(x^\star))}{\mu_0(B_r(x))} > 0$. The inequality above implies that

$$
\liminf_{r \searrow 0} \frac{\mu_0(B_r(x))}{\mu_0(B_r(x^\star))} \geqslant 1,
$$

---

[11] The assumption that $\Phi$ is bounded on bounded subsets of $X$ is not restrictive. If $\Phi : X \to \mathbb{R}$ is continuous, then the boundedness assumption holds whenever the bounded subset is contained in a sufficiently small ball. Mild continuity assumptions on forward models and log-likelihoods are commonplace in the study of BIPs e.g. Stuart (2010), (assumption 2.6(i)–(ii)).

and hence $x^\star$ does not satisfy (3.4). Finally, if $x$ is not a global weak mode, then by lemma 3.9, it cannot be a strong mode. This proves (c). $\qquad\square$

**Example B.2 (Importance of limit in the *M*-property).** For $n \in \mathbb{N}$, let

$$a_n := 2^{-\frac{(n-1)(n+2)}{2}}, \qquad b_n := \frac{a_n}{2}, \qquad \alpha_n := 2^{-n}(a_n - a_{n+1}),$$

$$\beta_n := \frac{\alpha_n}{2}, \qquad \varepsilon_n := 2\alpha_n, \qquad \delta_n := \alpha_n.$$

Let $\mu \in \mathcal{P}(\mathbb{R})$ be given by its probability density

$$\rho \propto \sum_{n \in \mathbb{N}} 2^n \left( 1_{[-1+\alpha_n, -1+2\alpha_n]} + 1_{[1-2\beta_n, 1-\beta_n]} \right),$$

where $1_A$ denotes the indicator function of $A \subseteq \mathbb{R}$, as visualized in figure B.1. The definitions of $a_n$ and $\alpha_n$ imply that $(a_n)_{n\in\mathbb{N}}$ is strictly decreasing to 0 and, for every $n \in \mathbb{N}$,

$$\alpha_n = 2^{-\frac{n^2+3n-2}{2}} - 2^{-\frac{n^2+5n}{2}} > 0,$$

$$\frac{\alpha_n}{\alpha_{n+1}} > \frac{2^{-\frac{n^2+3n-2}{2}} - 2^{-\frac{n^2+5n}{2}}}{2^{-\frac{n^2+5n+2}{2}}} = 2^{n+2} - 2^{-1} > 2,$$

so $(\alpha_n)_{n\in\mathbb{N}}$ is strictly positive and strictly decreasing to 0. Hence, the intervals $([-1+\alpha_n, -1+2\alpha_n])_{n\in\mathbb{N}}$ are disjoint, and the intervals $([1-2\beta_n, 1-\beta_n])_{n\in\mathbb{N}}$ are also disjoint. The facts that $(\alpha_n)_{n\in\mathbb{N}}$ are strictly positive and decreasing imply that if $m \geqslant n$, then $(-1+\alpha_m, -1+2\alpha_m) \subseteq (-1-2\alpha_n, -1+2\alpha_n) = B_{\varepsilon_n}(-1)$. The inequality $\frac{\alpha_{n-1}}{\alpha_n} > 2$ implies that $-1+2\alpha_n < -1+\alpha_m$ and hence $(-1+\alpha_m, -1+2\alpha_m) \cap (-1-2\alpha_n, -1+2\alpha_n) = \varnothing$ for any $m < n$. This implies that $\mu(B_{\varepsilon_n}(-1)) = \sum_{k=n}^\infty 2^k \alpha_k = a_n$. Similar arguments yield

$$\frac{\mu(B_{\varepsilon_n}(-1))}{\mu(B_{\varepsilon_n}(1))} = \frac{\sum_{k=n}^\infty 2^k \alpha_k}{\sum_{k=n}^\infty 2^k \beta_k} = \frac{a_n}{b_n} = 2, \tag{B.1}$$

$$\frac{\mu(B_{\delta_n}(-1))}{\mu(B_{\delta_n}(1))} = \frac{\sum_{k=n+1}^\infty 2^k \alpha_k}{\sum_{k=n}^\infty 2^k \beta_k} = \frac{a_{n+1}}{b_n} = 2^{-n-2} \xrightarrow[n\to\infty]{} 0. \tag{B.2}$$

Thus, $\limsup_{\varepsilon \searrow 0} \frac{\mu(B_\varepsilon(-1))}{\mu(B_\varepsilon(1))} = 2$ and $\liminf_{\varepsilon \searrow 0} \frac{\mu(B_\varepsilon(-1))}{\mu(B_\varepsilon(1))} = 0$. For $x \in \text{supp}(\mu)\setminus\{-1,1\}$, it follows from the disjointness of $([-1+\alpha_n, -1+2\alpha_n])_{n\in\mathbb{N}}$ and $([1-2\beta_n, 1-\beta_n])_{n\in\mathbb{N}}$ that there exists a unique $m \in \mathbb{N}$ such that either $x \in [-1+\alpha_m, -1+2\alpha_m]$, or $x \in [1-2\beta_m, 1-\beta_m]$. In either case, for sufficiently small $\varepsilon$ it holds that $\mu(B_\varepsilon(x)) = (2\varepsilon)2^m$, if $x$ is in the interior of either interval. If $x$ is an endpoint of one of the intervals, then for sufficiently small $\varepsilon$, $\mu(B_\varepsilon(x)) = \varepsilon 2^m$, so it suffices to consider the case where $x$ is in the interior of one of the intervals. Since $\mu(B_{\varepsilon_n}(1)) = b_n$, it follows that

$$\frac{\mu(B_{\varepsilon_n}(x))}{\mu(B_{\varepsilon_n}(1))} = \frac{2^{m+1}2\alpha_n}{b_n} = \frac{2^{m+1-n}(a_n - a_{n+1})}{a_n} < 2^{m+1-n}\frac{a_n}{a_n} = 2^{m+1-n} \xrightarrow[n\to\infty]{} 0.$$

Then $\liminf_{\varepsilon \searrow 0} \frac{\mu(B_\varepsilon(x))}{\mu(B_\varepsilon(1))} = 0$ for any point. Hence, for $E = \{1\}$, the $\liminf$ part of property $M(\mu, E)$ is satisfied. Further, $u = 1$ is a minimiser of any OM functional on $E$ since $E = \{1\}$, but $u = 1$ is not a global weak mode due to (B.1).
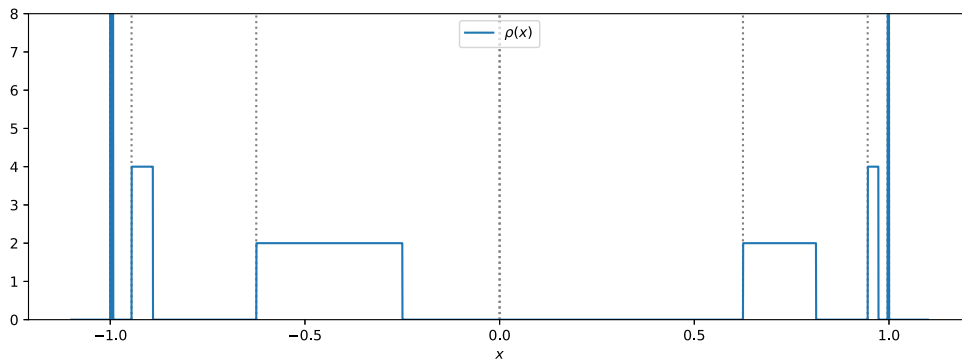
**Figure B.1.** Probability density $\rho$ (unnormalized) in example B.2. $u = 1$ is a minimiser of the OM functional $I_{\mu,E}$ on $E = \{1\}$, but not a global weak mode, even though the 'lim inf-only' version of property $M(\mu, E)$ is satisfied. The vertical lines are located at $-1 + \alpha_n$ and $1 - \alpha_n$, respectively.

Note that the above example can be modified to be even more extreme: if one sets $a_n := 2^{-n(n-1)}$ and $b_n := \frac{a_n}{2^n}$, then one obtains that $a_n/b_n = 2^n$ and $a_{n+1}/b_n = 2^{-n}$, and hence

$$\liminf_{\varepsilon \searrow 0} \frac{\mu(B_\varepsilon(-1))}{\mu(B_\varepsilon(1))} = 0, \qquad \limsup_{\varepsilon \searrow 0} \frac{\mu(B_\varepsilon(-1))}{\mu(B_\varepsilon(1))} = \infty.$$

**Proposition B.3 (Open v. closed balls).** *Let $X$ be a metric space, $\mu \in \mathcal{P}(X)$ a probability measure on $(X, \mathcal{B}(X))$ and $x_1, x_2 \in X$ with $x_2 \in \text{supp}(\mu)$. For $\varepsilon > 0$ define the ratios $R_\varepsilon := \frac{\mu(B_\varepsilon(x_1))}{\mu(B_\varepsilon(x_2))}$ and $\bar{R}_\varepsilon := \frac{\mu(\bar{B}_\varepsilon(x_1))}{\mu(\bar{B}_\varepsilon(x_2))}$, where $\bar{B}_r(x)$ denotes the closed ball in $X$ of radius $r$ centred on $x$. Then*

$$\limsup_{\varepsilon \searrow 0} \bar{R}_\varepsilon = \limsup_{\varepsilon \searrow 0} R_\varepsilon \quad and \quad \liminf_{\varepsilon \searrow 0} \bar{R}_\varepsilon = \liminf_{\varepsilon \searrow 0} R_\varepsilon.$$

*Hence, $\lim_{\varepsilon \searrow 0} \bar{R}_\varepsilon$ exists if and only if $\lim_{\varepsilon \searrow 0} R_\varepsilon$ exists, in which case these two values agree.*

**Proof.** First assume that $\limsup_{\varepsilon \searrow 0} \bar{R}_\varepsilon > \limsup_{\varepsilon \searrow 0} R_\varepsilon =: \mathring{s}$. Then there exists $\zeta > 0$ and a positive null sequence $(\varepsilon_n)_{n \in \mathbb{N}}$ such that $\bar{R}_{\varepsilon_n} \geqslant \mathring{s} + \zeta$. For each $n \in \mathbb{N}$ perform the following construction: since $\bigcap_{\delta > 0} B_{\varepsilon_n + \delta}(x) = \bar{B}_{\varepsilon_n}(x)$ for any $x \in X$ and using continuity of probability measures, we obtain

$$\lim_{\delta \searrow 0} R_{\varepsilon_n + \delta} = \frac{\lim_{\delta \searrow 0} \mu(B_{\varepsilon_n + \delta}(x_1))}{\lim_{\delta \searrow 0} \mu(B_{\varepsilon_n + \delta}(x_2))} = \frac{\mu(\bar{B}_{\varepsilon_n}(x_1))}{\mu(\bar{B}_{\varepsilon_n}(x_2))} = \bar{R}_{\varepsilon_n} \geqslant \mathring{s} + \zeta,$$

and there exists $0 < \delta_n < n^{-1}$ such that $R_{\varepsilon_n + \delta_n} \geqslant \mathring{s} + \zeta/2$. Hence, we have constructed a null sequence $(\tilde{\varepsilon}_n)_{n \in \mathbb{N}} := (\varepsilon_n + \delta_n)_{n \in \mathbb{N}}$ with

$$\mathring{s} = \limsup_{\varepsilon \searrow 0} R_\varepsilon \geqslant \limsup_{n \to \infty} R_{\tilde{\varepsilon}_n} \geqslant \mathring{s} + \zeta/2,$$

which is a contradiction. Therefore, our assumption was false and $\limsup_{\varepsilon \searrow 0} \bar{R}_\varepsilon \leqslant \limsup_{\varepsilon \searrow 0} R_\varepsilon$. The other inequality can be proven similarly using $\bigcup_{\delta > 0} \bar{B}_{\varepsilon_n - \delta}(x) = B_{\varepsilon_n}(x)$ and a similar argument works for the corresponding lim inf statement. $\qquad\square$
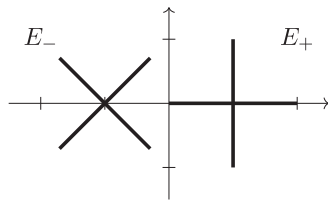
**Figure B.2.** Uniform length measure $\mu$ on the set $E = E_- \cup E_+ \subset \mathbb{R}^2$ shown here has a unique strong mode and global weak mode at $(1, 0)$ with respect to the one-norm on $\mathbb{R}^2$, but at $(-1, 0)$ with respect to the $\infty$-norm. The OM functionals for $\mu$ associated to these two norms are likewise distinct, as discussed in example B.4.

**Example B.4 (OM functionals and changes of metric).** Following Lie and Sullivan (2018), (example 5.6), let $\mu$ be the finite Borel measure on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ that is one-dimensional Hausdorff measure (i.e. uniform length measure) on the disjoint union $E$ of two right-angled crosses in the plane, with one cross, $E_+$, aligned with the coordinate axes and centred at $e_1 := (1, 0)$ and the other, $E_-$, aligned at $\pi/4$ to the axes and centred at $-e_1$, as illustrated in figure B.2. (Note that there is a slight error in Lie and Sullivan (2018), (example 5.6) concerning the side lengths of the cross $E_-$ and hence the total mass of $\mu$, but this error does not affect the final conclusion of that example or this one, since it is only the mass near $\pm e_1$ that is important.) With respect to the one-norm,

$$\mu(B_r^1(-e_1)) = 2\sqrt{2}r, \qquad \mu(B_r^1(e_1)) = 4r,$$

whereas, with respect to the $\infty$-norm, which in this setting is Lipschitz equivalent to the one-norm,

$$\mu(B_r^\infty(-e_1)) = 4\sqrt{2}r, \qquad \mu(B_r^\infty(e_1)) = 4r,$$

and, after considering the other points of $\mathbb{R}^2$, it follows that $e_1$ (resp. $-e_1$) is the unique strong and global weak mode of $\mu$ with respect to the one-norm (resp. $\infty$-norm). These same calculations, though, can be used to show that $\mu$ has an OM functional $I^1 : E \to \mathbb{R}$ (resp. $I^\infty : E \to \mathbb{R}$) with respect to the one-norm (resp. $\infty$-norm) and moreover

$$I^1(-e_1) = I^1(e_1) + \log \sqrt{2},$$
$$I^\infty(-e_1) = I^\infty(e_1) - \log \sqrt{2}.$$

Indeed, more generally, $I^1$ takes greater values on $E_-$ than on $E_+$, whereas $I^\infty$ takes greater values on $E_+$ than on $E_-$. This shows that these two OM functionals are distinct (i.e. differ by more than an additive constant).

### B.2. Supporting results for section 4

**Example B.5.** Let $X = \mathbb{R}$, $E = \mathbb{N}$ and $\mu \in \mathcal{P}(X)$ have Lebesgue density $\rho := \frac{24}{5\pi^2} \sum_{k \in \mathbb{N}} \rho_k$, shown in figure B.3, where

$$\rho_0(x) := \frac{1}{4} \left( |x|^{-1/2} - 2 \right) \mathbf{1}_{\left[-\frac{1}{4}, \frac{1}{4}\right] \setminus \{0\}}(x), \tag{B.3}$$

$$\rho_k(x) := \frac{\rho_0(x-k)}{k^2} + k^2\, 1_{\left[-\frac{1}{2k^4}, \frac{1}{2k^4}\right]}(x-k). \tag{B.4}$$

Since supp $\rho_0 = [-\frac{1}{4}, \frac{1}{4}]$ and supp $\rho_k = [k - \frac{1}{4}, k + \frac{1}{4}]$, it follows that if $k \ne \ell$ then supp $\rho_k \cap$ supp $\rho_\ell = \varnothing$. In addition, for $r \leqslant \frac{1}{4}$, $\int_{B_r(0)} \rho_0(x)\,dx = r^{1/2} - r$. This implies that $\int_X \rho_0 = \frac{1}{4}$. Hence $\int_X \rho_k = \frac{5}{4k^2}$ and $\int_X \rho = 1$. For every $k \in \mathbb{N}$ and $r \leqslant \min\{\frac{1}{4}, \frac{1}{2k^4}\}$, $\int_{B_r(k)} \rho_k(x)\,dx = k^{-2}(r^{1/2} - r) + 2rk^2$. Then, for $u = 1$ and any $k \in E = \mathbb{N}$,

$$\frac{\mu(B_r(u))}{\mu(B_r(k))} \xrightarrow{r \searrow 0} k^2,$$

which implies that $u = 1$ is a weak mode or weak MAP estimate in the sense of Helin and Burger (2015), (definition 4) and hence a global weak mode in the sense of definition 3.7. In addition, $I_{\mu,E}(k) = 2\log k$ defines an OM functional for $\mu$ on $E$, with unique minimiser $u = 1$. Next, we show that property $M(\mu, E)$ holds. If $x \notin \bigcup_{k \in \mathbb{N}} \text{supp}\, \rho_k$, then for sufficiently small $r > 0$, $\mu(B_r(x)) = 0$. Thus, it suffices to consider $x \in \bigcup_{k \in \mathbb{N}} \text{supp}\, \rho_k \setminus \mathbb{N}$. There exists a unique $m \in \mathbb{N} = E$ such that $x \in \text{supp}\, \rho_m \setminus \{m\}$. Assume that $x = m + \delta$ for $0 < \delta \leqslant \frac{1}{4}$. Then, for $r$ small enough,

$$\mu(B_r(x)) = \frac{1}{2m^2}\left(\sqrt{\delta + r} - \sqrt{\delta - r}\right) - r + 2m^2 r \approx \frac{1}{2m^2}\left(\frac{1}{2\delta^{1/2}}2r\right) + r(2m^2 - 1),$$

using the Taylor expansion of $y \mapsto \sqrt{y}$. Thus, $\mu(B_r(x))$ decreases to zero linearly in $r$, whereas $\mu(B_r(m))$ decreases to zero like $r^{1/2}$. Recall that property $M(\mu, E)$ holds if there exists some $x^\star \in E$ such that if $x \in X \setminus E$ then (3.4) holds, i.e. $\lim_{r \searrow 0} \frac{\mu(B_r(x))}{\mu(B_r(x^\star))} = 0$. Using that $x = m + \delta$ and $x^\star = m$ shows that property $M(\mu, E)$ holds.

Next, recall from (3.3) that if $I = I_{\mu,E} : E \to \mathbb{R}$ is an OM functional for $\mu$, then $I$ must satisfy $\lim_{r \searrow 0} \frac{\mu(B_r(x_1))}{\mu(B_r(x_2))} = \exp(I(x_2) - I(x_1))$ for all $x_1, x_2 \in E$. For $x_1 = 1$ and $x_2 = \frac{5}{4}$, the preceding calculations show that the ratio $\frac{\mu(B_r(x_1))}{\mu(B_r(x_2))}$ increases to $\infty$ as $r \searrow 0$. If, on the other hand, we set $x_1 = \frac{5}{4}$ and $x_2 = 1$, then the resulting limiting ratio equals $0$. This shows that the domain of the OM functional $I_{\mu,E}$ cannot be extended beyond $E = \mathbb{N}$. However, for $u = 1$, $n \in E = \mathbb{N}$ and $r_n = \frac{1}{2n^4}$,

$$\liminf_{r \searrow 0} \frac{\mu(B_r(u))}{M_r} \leqslant \liminf_{n \to \infty} \frac{\mu(B_{r_n}(u))}{\mu(B_{r_n}(n))} \leqslant \liminf_{n \to \infty} \frac{\frac{1}{\sqrt{2}n^2} + \frac{1}{n^4}}{\frac{1}{n^2}} = \frac{1}{\sqrt{2}}.$$

Thus, $u = 1$ cannot be a strong mode of $\mu$, even though it minimises $I_{\mu,E}$.

### B.3. Supporting results for section 5

**Lemma B.6.** *Let $a^{(n)} = (a_k^n)_{k \in \mathbb{N}} \in \ell^2$, $n \in \mathbb{N}$, define a bounded sequence in $\ell^2$, i.e. there exists a constant $M > 0$ such that $\|a^{(n)}\|_{\ell^2} \leqslant M$ for each $n \in \mathbb{N}$. Further, let $a_k^{(n)} \xrightarrow[n \to \infty]{} a_k \in \mathbb{R}$ for each $k \in \mathbb{N}$. Then $a := (a_k)_{k \in \mathbb{N}} \in \ell^2$ and $\|a\|_{\ell^2} \leqslant M$.*

**Proof.** Assume that there exists $K \in \mathbb{N}$ such that $\sum_{k=1}^K |a_k|^2 > M^2 + \varepsilon$ for some $0 < \varepsilon < 1$. Since $a_k^{(n)} \xrightarrow[n \to \infty]{} a_k$ for each $k \in \mathbb{N}$, there exists, for each $k = 1, \dots, K$, a number $N(k) \in \mathbb{N}$ such that, for all $n \geqslant N(k)$, $|a_k^{(n)} - a_k| < \frac{\varepsilon}{4K(|a_k|+1)}$ ($\leqslant 1$). Hence, for each $k = 1, \dots, K$ and $N := \max(N(1), \dots, N(K))$,
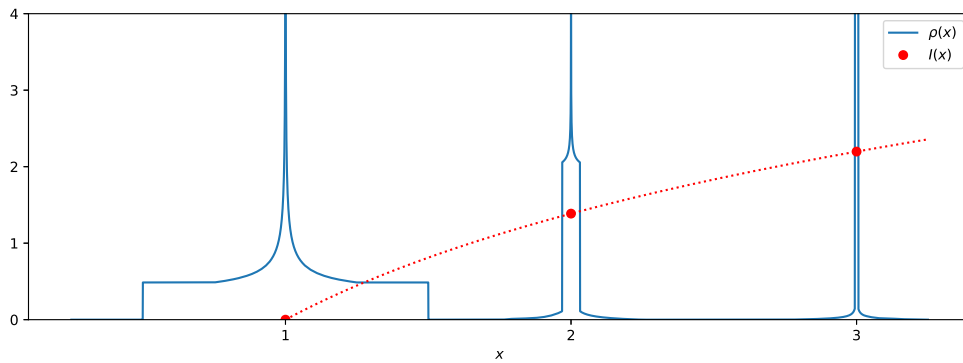
**Figure B.3.** Probability density $\rho$ in example B.5. $u = 1$ is a minimiser of the OM functional $I_{\mu,E}$ and an $E$-weak mode for $E = \mathbb{N}$, but not a strong mode.

$$|a_k|^2 \leqslant (|a_k^{(N)}| + |a_k^{(N)} - a_k|)^2$$

$$= |a_k^{(N)}|^2 + 2 \underbrace{|a_k^{(N)}|}_{\leqslant |a_k|+1} \underbrace{|a_k^{(N)} - a_k|}_{\leqslant \frac{\varepsilon}{4K(|a_k|+1)}} + \underbrace{|a_k^{(N)} - a_k|^2}_{\leqslant \frac{\varepsilon}{4K(|a_k|+1)}} \leqslant |a_k^{(N)}|^2 + \frac{\varepsilon}{K}.$$

Therefore, $\sum_{k=1}^{K} |a_k|^2 \leqslant M^2 + \varepsilon$, yielding a contradiction. Hence, our assumption was false and the lemma is proven. □

**Remark B.7.** Lemma B.6 does not state that $\|a^{(n)} - a\|_{\ell^2} \to 0$ and, in fact, this is not true in general. A counterexample is provided by $a = 0$ and $a^{(n)} = (\delta_{nk})_{k \in \mathbb{N}}$, where $\delta_{nk}$ denotes the Kronecker delta function.

### B.4. Supporting results for section 6

Recall that a function $f : X \to Y$ between metric spaces $X$ and $Y$ is **locally uniformly continuous** if, for every $x \in X$, there exists a function $\omega_{f,x} : [0, \infty) \to [0, \infty]$, a **local modulus of continuity** for $\Phi$ near $x$, such that

$$\text{for all } x' \in X, \quad d_Y(f(x'), f(x)) \leqslant \omega_{f,x}\left(d_X(x', x)\right), \tag{B.5}$$

$$\text{and} \quad \omega_{f,x}(r) \to 0 \quad \text{as } r \to 0. \tag{B.6}$$

In particular, (B.6) implies that, for each $x \in X$, there exists $r_x > 0$ such that $\omega_{f,x}(r_x)$ is finite for all $0 \leqslant r \leqslant r_x$. It is no loss of generality to assume that $\omega_{f,x}$ is an increasing function. Local uniform continuity is slightly but strictly stronger than $f$ being continuous: according to Izzo (1994), (theorem 1), on every infinite-dimensional separable normed space there exist bounded, continuous real-valued functions that are nowhere locally uniformly continuous; however, Izzo (1994), (theorem 4) also shows that every continuous real-valued function on a metric space can be approximated uniformly by locally uniformly continuous functions.

**Lemma B.8 (OM functionals for reweighted measures).** *Let $X$ be a metric space and let $\mu_0 \in \mathcal{P}(X)$ have an OM functional $I_0$ on $E \subseteq X$. Let $\Phi : X \to \mathbb{R}$ be locally uniformly continuous. Suppose that $Z := \int_X e^{-\Phi(x)} \mu_0(\mathrm{d}x) \in (0, \infty)$. Then $\mu \in \mathcal{P}(X)$ defined by $\mu(\mathrm{d}x) := Z^{-1} e^{-\Phi(x)} \mu_0(\mathrm{d}x)$ has $I := \Phi + I_0$ as an OM functional on $E$. If, furthermore, property*

$M(\mu_0, E)$ holds and $I_0$ is extended to $I_0 : X \to \overline{\mathbb{R}}$ by setting $I_0 \equiv +\infty$ on $X \backslash E$, then property $M(\mu, E)$ holds and $\Phi + I_0$ is an extended OM functional for $\mu$.

**Proof.** Let $u, v \in E$ and let $r_0$ be small enough that both $\omega_{\Phi,u}(r_0)$ and $\omega_{\Phi,v}(r_0)$ are finite; without loss of generality we henceforth consider only $0 < r < r_0$. By (B.5), on $B_r(u)$, $\Phi$ satisfies the bound

$$\Phi(u) - \omega_{\Phi,u}(r) \leqslant \Phi(\cdot) \leqslant \Phi(u) + \omega_{\Phi,u}(r).$$

Since $\mu(A) = Z^{-1} \int_A \exp(-\Phi(x))\mu_0(\mathrm{d}x)$ for each measurable set $A \in \mathcal{B}(X)$, it follows that

$$Z^{-1} \exp(-\Phi(u) - \omega_{\Phi,u}(r))\mu_0(B_r(u))$$
$$\leqslant \mu(B_r(u)) \leqslant Z^{-1} \exp(-\Phi(u) + \omega_{\Phi,u}(r))\mu_0(B_r(u)).$$

Similar arguments apply *mutatis mutandis* for $v$ in place of $u$. Hence,

$$\frac{\exp(-\Phi(u) - \omega_{\Phi,u}(r))\mu_0(B_r(u))}{\exp(-\Phi(v) + \omega_{\Phi,v}(r))\mu_0(B_r(v))} \leqslant \frac{\mu(B_r(u))}{\mu(B_r(v))} \leqslant \frac{\exp(-\Phi(u) + \omega_{\Phi,u}(r))\mu_0(B_r(u))}{\exp(-\Phi(v) - \omega_{\Phi,v}(r))\mu_0(B_r(v))},$$

and so, since $I_0$ is an OM functional for $\mu_0$ and both $\omega_{\Phi,u}(r)$ and $\omega_{\Phi,v}(r)$ tend to $0$ as $r \to 0$,

$$\exp(-\Phi(u) + \Phi(v) + I_0(v) - I_0(u)) \leqslant \lim_{r \searrow 0} \frac{\mu(B_r(u))}{\mu(B_r(v))} \leqslant \exp(-\Phi(u) + \Phi(v) + I_0(v) - I_0(u)).$$

which proves the first claim.

The second claim is an immediate consequence of the first part and lemma B.1(b). □

**Lemma B.9 (Continuous convergence of potentials via projection).** *Let $X$ be a separable Banach space and let $(X_n)_{n \in \mathbb{N}}$ be a sequence of (not necessarily nested) finite-dimensional subspaces with surjective uniformly bounded linear projection operators $P_n : X \to X_n$ such that*

$$\text{for all } x \in X, \quad \lim_{n \to \infty} \|P_n x - x\| = 0. \tag{B.7}$$

*Let $\Phi : X \to \mathbb{R}$ be locally uniformly continuous.*

(a) *For each $n \in \mathbb{N}$, $\Phi \circ P_n$ is locally uniformly continuous.*
(b) *$\Phi \circ P_n \to \Phi$ continuously as $n \to \infty$.*

**Proof.** Let $M \geqslant 1$ be a uniform upper bound for the operator norms $\|P_n\|$, $n \in \mathbb{N}$. (Note that, in the special case that $X$ is a separable Hilbert space with complete orthonormal system $\{\psi_n\}_{n \in \mathbb{N}}$ and $P_n$ is the orthogonal projection onto $\mathrm{span}\{\psi_1, \ldots, \psi_n\}$, we may take $M = 1$.)

To show (a), fix $n \in \mathbb{N}$ and $x \in X$ and let $\omega_{\Phi,P_n x} : [0, \infty) \to [0, \infty]$ be an increasing local modulus of continuity for $\Phi$ near $P_n x$. Then, for all $x' \in X$,

$$\left| (\Phi \circ P_n)(x') - (\Phi \circ P_n)(x) \right| \leqslant \omega_{\Phi, P_n x} \left( \| P_n x' - P_n x \| \right) \leqslant \omega_{\Phi, P_n x} \left( \| P_n \| \| x' - x \| \right).$$

Thus, $\omega_{\Phi \circ P_n, x}(r) := \omega_{\Phi, P_n x}(Mr)$ is a local modulus of continuity for $\Phi \circ P_n$ near $x$.

To establish (b), fix $x \in X$ and let $\omega_{\Phi, x} : [0, \infty) \to [0, \infty]$ be an increasing local modulus of continuity for $\Phi$ near $x$. Let $\varepsilon > 0$ be arbitrary and let $r_\varepsilon > 0$ be such that $\omega_{\Phi, x}(r_\varepsilon) < \varepsilon$. By (B.7), there exists $N \in \mathbb{N}$ such that, for all $n \geqslant N$, $\| P_n x - x \| < r_\varepsilon/2$. Then, for $n \geqslant N$ and $x' \in X$ with $\| x' - x \| < r_\varepsilon/2M$,

$$
\begin{aligned}
\left| (\Phi \circ P_n)(x') - \Phi(x) \right| &\leqslant \omega_{\Phi, x} \left( \| P_n x' - x \| \right) \\
&\leqslant \omega_{\Phi, x} \left( \| P_n x' - P_n x \| + \| P_n x - x \| \right) \\
&\leqslant \omega_{\Phi, x} \left( M \| x' - x \| + \| P_n x - x \| \right) \\
&\leqslant \omega_{\Phi, x}(r_\varepsilon) < \varepsilon,
\end{aligned}
$$

as required. □

## ORCID iDs

Birzhan Ayanbayev ⬤ https://orcid.org/0000-0002-0925-4045
Ilja Klebanov ⬤ https://orcid.org/0000-0002-1911-4281
Han Cheng Lie ⬤ https://orcid.org/0000-0002-6905-9903
T J Sullivan ⬤ https://orcid.org/0000-0003-4081-8301

## References

Agapiou S, Burger M, Dashti M and Helin T 2018 Sparsity-promoting and edge-preserving maximum *a posteriori* estimators in non-parametric Bayesian inverse problems *Inverse Problems* **34** 045002

Ayanbayev B, Klebanov I, Lie H C and Sullivan T J 2021 Γ-convergence of Onsager–Machlup functionals: II. Infinite product measures on Banach spaces *Inverse Problems* **38** 025006

Bogachev V I 1998 *Gaussian Measures* (*Mathematical Surveys and Monographs* vol 62) (Providence, RI: American Mathematical Society)

Braides A 2002 Γ-*Convergence for Beginners* (*Oxford Lecture Series in Mathematics and its Applications* vol 22) (Oxford: Oxford University Press)

Braides A 2006 A handbook of Γ-convergence *Handbook of Differential Equations: Stationary Partial Differential Equations* vol 3 (Amsterdam: Elsevier) pp 101–213

Clason C, Helin T, Kretschmann R and Piiroinen P 2019 Generalized modes in Bayesian inverse problems *SIAM/ASA J. Uncertain. Quantification* **7** 652–84

Dal Maso G 1993 *An Introduction to Γ-Convergence* (*Progress in Nonlinear Differential Equations and Their Applications* vol 8) (Boston, MA: Birkhäuser)

Dashti M, Harris S and Stuart A 2012 Besov priors for Bayesian inverse problems *Inverse Problems Imaging* **6** 183–200

Dashti M, Law K J H, Stuart A M and Voss J 2013 MAP estimators and their consistency in Bayesian nonparametric inverse problems *Inverse Problems* **29** 095017

De Giorgi E 2006 *Selected Papers* ed L Ambrosio, G Dal Maso, M Forti, M Miranda and S Spagnolo (Berlin: Springer)

Dembo A and Zeitouni O 1991 Onsager–Machlup functionals and maximum *a posteriori* estimation for a class of non-Gaussian random fields *J. Multivariate Anal.* **36** 243–62

Dunlop M M and Stuart A M 2016 MAP estimators for piecewise continuous inversion *Inverse Problems* **32** 105003

Dürr D and Bach A 1978 The Onsager–Machlup function as Lagrangian for the most probable path of a diffusion process *Commun. Math. Phys.* **60** 153–70

Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (*Mathematics and its Applications* vol 375) (Dordrecht: Kluwer)

Hegland M 2007 Approximate maximum *a posteriori* with Gaussian process priors *Constr. Approx.* **26** 205–24

Helin T and Burger M 2015 Maximum *a posteriori* probability estimates in infinite-dimensional Bayesian inverse problems *Inverse Problems* **31** 085009

Hosseini B 2017 Well-posed Bayesian inverse problems with infinitely divisible and heavy-tailed prior measures *SIAM/ASA J. Uncertain. Quantification* **5** 1024–60

Izzo A J 1994 Locally uniformly continuous functions *Proc. Am. Math. Soc.* **122** 1095–100

Kaipio J and Somersalo E 2005 *Statistical and Computational Inverse Problems* (*Applied Mathematical Sciences* vol 160) (Berlin: Springer)

Kasanický I and Mandel J 2017 On well-posedness of Bayesian data assimilation and inverse problems in Hilbert space (arXiv:1701.08298)

Kretschmann R 2019 Nonparametric Bayesian inverse problems with Laplacian noise *PhD Thesis* Universitaät Duisburg-Essen

Lasanen S 2012a Non-Gaussian statistical inverse problems: I. Posterior distributions *Inverse Problems Imaging* **6** 215–66

Lasanen S 2012b Non-Gaussian statistical inverse problems: II. Posterior convergence for approximated unknowns *Inverse Problems Imaging* **6** 267–87

Lassas M and Siltanen S 2004 Can one use total variation prior for edge-preserving Bayesian inversion? *Inverse Problems* **20** 1537–63

Lassas M, Saksman E and Siltanen S 2009 Discretization-invariant Bayesian inversion and Besov space priors *Inverse Problems Imaging* **3** 87–122

Latz J 2020 On the well-posedness of Bayesian inverse problems *SIAM/ASA J. Uncertain. Quantification* **8** 451–82

Lehtinen M S, Päivärinta L and Somersalo E 1989 Linear inverse problems for generalised random variables *Inverse Problems* **5** 599–612

Lie H C and Sullivan T J 2018 Equivalence of weak and strong modes of measures on topological vector spaces *Inverse Problems* **34** 115013

Sprungk B 2020 On the local Lipschitz stability of Bayesian inverse problems *Inverse Problems* **36** 055015

Stuart A M 2010 Inverse problems: a Bayesian perspective *Acta Numer.* **19** 451–559

Sullivan T J 2017 Well-posed Bayesian inverse problems and heavy-tailed stable quasi-Banach space priors *Inverse Problems Imaging* **11** 857–74

Wacker P 2020 MAP estimators for nonparametric Bayesian inverse problems in Banach spaces (arXiv:2007.12760)

Zeitouni O 1989 On the Onsager–Machlup functional of diffusion processes around non-$C^2$-curves *Ann. Probab.* **17** 1037–54