

PAPER • OPEN ACCESS

Strong maximum *a posteriori* estimation in Banach spaces with Gaussian priors

To cite this article: Hefin Lambley 2023 *Inverse Problems* **39** 125010

View the [article online](#) for updates and enhancements.

You may also like

- [Bayesian inverse problems in measure spaces with application to Burgers and Hamilton–Jacobi equations with white noise forcing](#)
Viet Ha Hoang
- [Bayesian inversion for electromyography using low-rank tensor formats](#)
Anna Rörich, Tim A. Werthmann, Dominik Göttsche et al.
- [Sparsity-promoting and edge-preserving maximum *a posteriori* estimators in non-parametric Bayesian inverse problems](#)
Sergios Agapiou, Martin Burger, Masoumeh Dashti et al.

Strong maximum *a posteriori* estimation in Banach spaces with Gaussian priors

Hefin Lambley 

Mathematics Institute, University of Warwick, Coventry CV4 7AL, United Kingdom

E-mail: hefin.lambley@warwick.ac.uk

Received 6 July 2023; revised 3 October 2023

Accepted for publication 27 October 2023

Published 7 November 2023



CrossMark

Abstract

This article shows that a large class of posterior measures that are absolutely continuous with respect to a Gaussian prior have strong maximum *a posteriori* estimators in the sense of Dashti *et al* (2013 *Inverse Problems* 29 095017). This result holds in any separable Banach space and applies in particular to nonparametric Bayesian inverse problems with additive noise. When applied to Bayesian inverse problems, this significantly extends existing results on maximum *a posteriori* estimators by relaxing the conditions on the log-likelihood and on the space in which the inverse problem is set.

Keywords: Bayesian inverse problems, maximum *a posteriori* estimation, modes of probability measures, Bayesian nonparametrics

1. Introduction

Nonparametric Bayesian models—which have infinite-dimensional parameters such as functions—are increasingly popular in modern statistical practice. For inverse problems, the need for prior information to overcome ill-posedness motivates the use of a Bayesian approach, and the desire for algorithms consistent at every resolution makes the nonparametric approach (as advocated by Stuart 2010) very appealing. One challenge in nonparametric Bayesian inference is that the posterior is a probability distribution on an infinite-dimensional space, making it difficult to analyse and interpret.

This article studies *maximum a posteriori (MAP) estimation* in nonparametric Bayesian inverse problems. A MAP estimator is a *mode* of the posterior: a summary by a ‘most likely’ point under the measure. The usual definition of a MAP estimator as a maximiser of the



Original Content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Lebesgue density is not available when the posterior is a measure on an infinite-dimensional parameter space, so it is common to define modes as the centres of metric balls with asymptotically maximal probability as proposed by Dashti *et al* (2013).

This definition allows modes of probability measures to be studied in very general settings, but we restrict attention to posterior measures arising in nonparametric Bayesian inverse problems. In particular, we study the case that μ^y is a posterior measure on a separable Banach space X which is absolutely continuous with respect to a Gaussian prior μ_0 and has Radon–Nikodym derivative

$$\mu^y(dx) = \exp(-\Phi(x)) \mu_0(dx). \quad (1.1)$$

The *potential* $\Phi: X \rightarrow \mathbb{R}$ is determined by the structure of the problem of interest and is essentially the negative log-likelihood of the statistical model. Any map Φ satisfying mild regularity conditions (theorem 2.8) yields a well-defined probability measure μ^y .

A classical example giving rise to such a posterior is the nonlinear inverse problem of inferring a parameter $x \in X$, which is typically a function, from a noisy observation $y \in Y$ given a Gaussian prior μ_0 for x , with

$$y = \mathcal{G}(x) + \xi. \quad (1.2)$$

The *observation operator* $\mathcal{G}: X \rightarrow Y$ is a measurable map relating the unknown $x \in X$ with the idealised observation $\mathcal{G}(x) \in Y$, which is often assumed to have finite dimension. This observation is then corrupted by the additive random noise ξ taking values in Y . Under appropriate regularity conditions on \mathcal{G} and ξ , the posterior μ^y for the conditional distribution $x | y$ is given by Bayes' rule and has the form (1.1) for some potential $\Phi(\cdot; y): X \rightarrow \mathbb{R}$ (theorem 2.7).

In the Bayesian inverse problems literature, Dashti *et al* (2013) developed the notion of a *strong mode* (definition 2.1) to define MAP estimators in the nonparametric setting and proved that strong modes exist when X is a separable Hilbert space under mild assumptions on the potential Φ . They also showed that strong modes coincide with minimisers of an Onsager–Machlup (OM) functional (definition 2.4) for the posterior in this setting, connecting their approach with previous work on most-likely paths of diffusion processes. For Bayesian inverse problems with additive noise as in (1.2), the OM functional can be viewed as a Tikhonov-regularised misfit functional, so the variational solution to an inverse problem—that is, the minimiser of the Tikhonov functional—can be viewed as a MAP estimator for a fully Bayesian approach. This connection is a significant driver for the development of the nonparametric mode theory described here.

As pointed out by Klebanov and Wacker (2023), although Dashti *et al* (2013) stated their results in the Banach setting, technical complications limit their proof strategy to Hilbert spaces, and some additional results are needed to complete the proof even in the Hilbert case, as described by Kretschmann (2019).

Recent work by Klebanov and Wacker extended the existence result to the sequence spaces $X = \ell^p(\mathbb{N}; \mathbb{R})$, $1 \leq p < \infty$, for Gaussian priors with *diagonal* covariance structure with respect to the canonical basis, i.e. $\mu_0 = \bigotimes_{n \in \mathbb{N}} N(0, \sigma_n^2)$ for some $(\sigma_n)_{n \in \mathbb{N}} \in \ell^p(\mathbb{N}; \mathbb{R})$.

This article proves the existence of strong modes for posteriors of the form (1.1) defined on a separable Banach space with any Gaussian prior, as originally claimed by Dashti *et al* (2013), theorem 3.5, and shows that strong modes are equivalent in this setting to the other types of small-ball modes present in the literature: the weak mode (definition 2.1) and the generalised strong mode (definition 2.3).

Theorem 1.1. *Let X be a separable Banach space equipped with a centred nondegenerate Gaussian prior μ_0 . Let μ^ν be the corresponding Bayesian posterior of the form (1.1) for some continuous potential $\Phi : X \rightarrow \mathbb{R}$, and suppose that, for each $\eta > 0$, there exists $K(\eta) \in \mathbb{R}$ such that*

$$\Phi(x) \geq K(\eta) - \eta \|x\|_X^2 \text{ for all } x \in X. \quad (1.3)$$

Then:

- (a) μ^ν has a strong mode, i.e. a strong MAP estimator, and any strong mode lies in the Cameron–Martin space of μ_0 ;
- (b) strong modes, generalised strong modes, weak modes and minimisers of an OM functional for μ^ν coincide.

The conditions imposed on the potential are weaker than those used by Dashti *et al* (2013) and Klebanov and Wacker (2023), who assumed that the potential was globally bounded below and locally Lipschitz. As pointed out by Kretschmann (2019, 2023), a global lower bound on Φ excludes Bayesian inverse problems with observations corrupted by additive white noise or Laplacian noise, and the proof of Dashti *et al* (2013) can be extended to handle these cases using a less restrictive lower bound. These cases can also be treated under the yet weaker conditions used here, which are similar to the assumptions used by Stuart (2010), assumption 2.6 in developing a well-posedness theory for nonparametric Bayesian inverse problems.

1.1. Outline

Section 2 defines the small-ball modes used in this paper in the general setting of a metric space and states some essential results, including the strong–weak dichotomy (lemma 2.2) which appears to be new to the literature. This section also recalls properties of Gaussian measures used throughout the article and briefly outlines the motivating application of Bayesian inverse problems.

Section 3 states the main estimate (proposition 3.1) needed to prove theorem 1.1, which can be viewed as an analogue of the explicit Anderson inequality of Dashti *et al* (2013), lemma 3.6. This is then used to establish the M -property for Gaussian measures on a separable Banach space (corollary 3.3), which was until now known rigorously only for special cases such as separable Hilbert spaces and ℓ^p spaces equipped with diagonal Gaussian measures.

Section 4 uses the tools developed in the previous section to study MAP estimators for Bayesian posteriors of the form (1.1). First, it states a short proof for the existence of weak modes using the M -property. Then, the bound in proposition 3.1 is used to show that any asymptotic maximising family (definition 4.2) for the posterior has a limit point (lemma 4.4). Lemma 4.5 shows that such a point must be a strong mode, extending a previous proof of Klebanov and Wacker (2023) to the Banach case, and this completes the proof of theorem 1.1.

Section 5 studies consistency theory for MAP estimators of Bayesian inverse problems of the type (1.2). Using theorem 1.1, the consistency results of Dashti *et al* (2013) are extended to apply in any separable Banach space X (theorems 5.1 and 5.2).

Section 6 gives some concluding remarks and suggests directions for future research.

2. Preliminaries and related work

For most of the paper, X will be a separable real Banach space, although some definitions and preliminary results in this section will be given in the more general case that X is a metric

space. In any metric space, the closed ball of radius r will be denoted $B_r(x)$. We consider only Borel measures and denote the set of Borel probability measures on X by $\mathcal{P}(X)$. When X is separable, the topological support

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

is nonempty (Aliprantis and Border 2006); this ensures that the quantity M_r defined in (2.2) is strictly positive for all $r > 0$.

2.1. Mode theory

As mentioned in the introduction, the small-ball mode theory has been developed largely in the Bayesian inverse problems literature. Strong modes were proposed by Dashti *et al* (2013), and weak modes were later suggested by Helin and Burger (2015) as a more convenient definition when connecting MAP estimators with variational solutions to inverse problems. Following Ayanbayev *et al* (2022a), we consider only *global* weak modes in this article.

Definition 2.1. Let X be a metric space and let $\mu \in \mathcal{P}(X)$. A *weak mode* of μ is any point $x^* \in \text{supp}(\mu)$ such that, for all $x \in X$,

$$\limsup_{r \rightarrow 0} \frac{\mu(B_r(x))}{\mu(B_r(x^*))} \leq 1. \quad (2.1)$$

Suppose also that X is separable. Then a *strong mode* of μ is any point $x^* \in \text{supp}(\mu)$ such that

$$\lim_{r \rightarrow 0} \frac{\mu(B_r(x^*))}{M_r} = 1, \quad M_r := \sup_{x \in X} \mu(B_r(x)). \quad (2.2)$$

The modes of a posterior measure μ^y will also be called *MAP estimators*. The difference between the two definitions (2.1) and (2.2) amounts to the order in which the supremum is taken: a weak mode must have asymptotically greater mass when compared to every other point individually, whereas a strong mode must asymptotically have the supremal ball mass. All strong modes are weak modes, because if x^* is a strong mode and $x \in X$, then

$$\limsup_{r \rightarrow 0} \frac{\mu(B_r(x))}{\mu(B_r(x^*))} \leq \lim_{r \rightarrow 0} \frac{M_r}{\mu(B_r(x^*))} = 1.$$

Lie and Sullivan (2018) proved that the converse may be false: there exist measures which have only weak modes and no strong modes. While the literature on modes largely treats ‘strong’ or ‘weak’ as a property of the mode itself, one should really think of ‘strong’ or ‘weak’ as a global regularity condition on the measure, because either all modes of a measure are strong or none of them are strong, as the following result shows.

Lemma 2.2 (Strong–weak dichotomy for modes). *Let X be a separable metric space. If $\mu \in \mathcal{P}(X)$ has a strong mode, then all weak modes of μ are strong modes.*

Proof. Suppose that x^* is a strong mode and y^* is a weak mode. As both x^* and y^* are weak modes, the definitions imply that

$$1 \leq \left(\limsup_{r \rightarrow 0} \frac{\mu(B_r(y^*))}{\mu(B_r(x^*))} \right)^{-1} = \liminf_{r \rightarrow 0} \frac{\mu(B_r(x^*))}{\mu(B_r(y^*))} \leq \limsup_{r \rightarrow 0} \frac{\mu(B_r(x^*))}{\mu(B_r(y^*))} \leq 1.$$

An application of the product rule for limits shows that y^* must also be a strong mode:

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

□

Clason *et al* (2019) proposed the *generalised strong mode*, motivated by inverse problems with hard parameter constraints (in the spirit of Ivanov regularisation) which lead to a posterior assigning zero mass outside of some feasible set.

Definition 2.3. Let X be a separable metric space. A *generalised strong mode* of $\mu \in \mathcal{P}(X)$ is any point $x^* \in X$ such that, for each sequence $(r_n)_{n \in \mathbb{N}} \rightarrow 0$, there exists $(x_n)_{n \in \mathbb{N}} \rightarrow x^*$ with

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

Taking the constant sequence $x_n = x^*$ in the definition shows that a strong mode x^* is also a generalised strong mode. Unlike strong and weak modes, generalised strong modes need not lie in the support of the measure. Furthermore, there is no strong–generalised strong dichotomy or weak–generalised strong dichotomy analogous to lemma 2.2: for the measure on \mathbb{R} with Lebesgue density $\rho(x) = \mathbb{1}\{x \in [0, 1]\}$, any $x \in (0, 1)$ is a strong mode (and hence a weak mode), but the points $x = 0$ and $x = 1$ are only generalised strong modes and are neither strong modes nor weak modes.

An alternative approach to find ‘most likely’ points is to minimise an OM functional associated with the measure of interest. This arises from the study of most-probable paths of diffusion processes (Dürr and Bach 1978).

Definition 2.4. Let X be a metric space and let $\mu \in \mathcal{P}(X)$. Suppose that $\emptyset \neq E \subseteq \text{supp}(\mu)$. A function $I: E \rightarrow \mathbb{R}$ is called an *OM functional* for μ if, for all $x, x' \in E$,

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

OM functionals are unique up to additive constants and can be interpreted heuristically as the negative logarithm of the Lebesgue density—but this cannot be taken literally for measures on an infinite-dimensional space, where there is no Lebesgue measure. For example, an OM functional for a Gaussian measure on an infinite-dimensional Banach space can be defined only on a small subspace called the *Cameron–Martin space* (see (2.6)). As an OM functional need not be defined on the entire space X , it is not immediate that an OM minimiser is in any sense ‘most likely’ under the measure μ , and this is the motivation to study small-ball modes as in definition 2.1 instead. A weak mode is always a minimiser of any OM functional for μ , however, and the *M-property* of Ayanbayev *et al* (2022a) gives a sufficient condition to ensure that an OM minimiser is a weak mode.

Definition 2.5 (M-property). Let X be a metric space and let $\mu \in \mathcal{P}(X)$. *Property M*(μ, E) holds for the set $\emptyset \neq E \subseteq \text{supp}(\mu)$ if there exists $x^* \in E$ such that

$$\lim_{r \rightarrow 0} \frac{\mu(B_r(x))}{\mu(B_r(x^*))} = 0 \text{ for all } x \notin E.$$

The next result states this equivalence between OM minimisers and weak modes under the *M-property* and shows that the *M-property* is inherited by a posterior of the form (1.1) from the prior. This generalises proposition 4.1 and lemma B.8 of Ayanbayev *et al* (2022a) to potentials

that are merely continuous rather than locally uniformly continuous. In the specific case that μ_0 is a Gaussian measure on a separable Banach space X , the claim (a) generalises theorem 3.2 of Dashti *et al* (2013), which requires that the potential is locally bounded and Lipschitz.

Proposition 2.6. *Let X be a metric space and suppose that $\mu_0 \in \mathcal{P}(X)$ has OM functional $I_0 : E \rightarrow \mathbb{R}$. Suppose that property $M(\mu_0, E)$ holds and that μ^y is a probability measure on X of the form (1.1) for some continuous potential $\Phi : X \rightarrow \mathbb{R}$. Then:*

- (a) μ^y has OM functional $F : E \rightarrow \mathbb{R}$ given by $F(u) = I_0(u) + \Phi(u)$ and property $M(\mu^y, E)$ holds;
- (b) $x^* \in X$ is a weak mode for μ^y , i.e. a weak MAP estimator, if and only if $x^* \in E$ and x^* minimises F .

Proof. Let $x \in X$ and $x' \in \text{supp}(\mu_0)$. As the density $\exp(-\Phi)$ is strictly positive, $x' \in \text{supp}(\mu^y)$ and thus $\mu^y(B_r(x')) > 0$ for all $r > 0$. By the continuity of Φ , for each $\varepsilon > 0$ there exists $\delta > 0$ such that $\|u - x\|_X < \delta \implies |\Phi(u) - \Phi(x)| < \varepsilon$ and $\|u - x'\|_X < \delta \implies |\Phi(u) - \Phi(x')| < \varepsilon$. Hence, for $r < \delta$, it follows that

$$\begin{aligned} \frac{\mu^y(B_r(x))}{\mu^y(B_r(x'))} &= \frac{\exp(-\Phi(x)) \int_{B_r(x)} \exp(\Phi(x) - \Phi(u)) \mu_0(du)}{\exp(-\Phi(x')) \int_{B_r(x')} \exp(\Phi(x') - \Phi(u)) \mu_0(du)} \\ &< \exp(\Phi(x') - \Phi(x) + 2\varepsilon) \frac{\mu_0(B_r(x))}{\mu_0(B_r(x'))}. \end{aligned} \quad (2.3)$$

Property $M(\mu^y, E)$ follows immediately from (2.3) by choosing $x \notin E$, $x' \in E$ and taking the limsup as $r \rightarrow 0$. To obtain the OM functional F , suppose instead that $x, x' \in E$; then by (2.3) and using the OM functional I_0 for μ_0 ,

$$\begin{aligned} \limsup_{r \rightarrow 0} \frac{\mu^y(B_r(x))}{\mu^y(B_r(x'))} &\leq \exp(\Phi(x') - \Phi(x) + 2\varepsilon) \limsup_{r \rightarrow 0} \frac{\mu_0(B_r(x))}{\mu_0(B_r(x'))} \\ &= \exp(\Phi(x') - \Phi(x) + 2\varepsilon + I_0(x') - I_0(x)). \end{aligned}$$

By deriving a lower bound analogous to (2.3) using the continuity of Φ and taking the liminf as $r \rightarrow 0$, we obtain the inequality

$$\begin{aligned} \exp(\Phi(x') - \Phi(x) + I_0(x') - I_0(x) - 2\varepsilon) &= \exp(\Phi(x') - \Phi(x) - 2\varepsilon) \liminf_{r \rightarrow 0} \frac{\mu_0(B_r(x))}{\mu_0(B_r(x'))} \\ &\leq \liminf_{r \rightarrow 0} \frac{\mu^y(B_r(x))}{\mu^y(B_r(x'))}. \end{aligned}$$

As $\varepsilon > 0$ is arbitrary this proves that $F(u) = I_0(u) + \Phi(u)$.

The claim in (b) is an immediate consequence of Ayanbayev *et al* (2022a), proposition 4.1. \square

Thus, when property $M(\mu^y, E)$ holds, one can view F as an extended-real-valued function with value $+\infty$ outside E . This interpretation is not valid if the M -property does not hold and one can say very little about the behaviour of μ^y on balls centred outside of E using an OM functional in this case.

While this article considers only Gaussian priors, MAP estimators have also been studied for Bayesian inverse problems with Besov and Cauchy priors (Agapiou *et al* 2018,

Ayanbayev *et al* (2022b). Besov and Cauchy priors are typically constructed as product measures placing full mass on a Banach subspace of \mathbb{R}^∞ , and the product structure of \mathbb{R}^∞ makes finite-dimensional approximation arguments possible. As an arbitrary Banach space need not have such product structure, we instead exploit the fact that a Gaussian measure is fully determined by its behaviour on a Hilbert subspace (the Cameron–Martin space) whose geometry is much more convenient to work with.

2.2. Gaussian measures

This section summarises the properties of Gaussian measures used in the article; see the monograph of Bogachev (1998) for a thorough introduction to Gaussian measures. If X is a separable Banach space, a measure $\gamma \in \mathcal{P}(X)$ is *Gaussian* if the pushforward $\gamma \circ f^{-1}$ is a Gaussian measure on \mathbb{R} for every f lying in the topological dual X^* . The measure γ is centred if it has mean zero and nondegenerate if it has full support, i.e. $\text{supp}(\gamma) = X$; we assume that γ is always centred and nondegenerate in the remainder of the article.

The *reproducing-kernel Hilbert space* (RKHS) X_γ^* of γ is the $L^2(\gamma)$ -closure of X^* , and the covariance operator $R_\gamma : X_\gamma^* \rightarrow (X^*)'$, taking values in the algebraic dual $(X^*)'$ of X^* , is given by

$$R_\gamma(f)(g) := \int_X f(x)g(x) \gamma(\mathrm{d}x).$$

As X is separable, the measure γ is Radon and thus $R_\gamma(f)$ is representable by an element of X for any $f \in X_\gamma^*$ (Bogachev 1998, theorem 3.2.3). The image of R_γ in X is called the *Cameron–Martin space* $E \subset X$. It is a separable Hilbert space under the Cameron–Martin inner product $\langle h, k \rangle_E := \langle R_\gamma^{-1}h, R_\gamma^{-1}k \rangle_{L^2(\gamma)}$, which induces the norm $\|h\|_E := \|R_\gamma^{-1}h\|_{L^2(\gamma)}$. The Cameron–Martin space of a Radon Gaussian measure γ is compactly embedded in X , i.e. there exists $C > 0$ such that

$$\|x\|_X \leq C\|x\|_E \quad (2.4)$$

and the inclusion $\iota : E \rightarrow X$ is a compact operator (Bogachev 1998, corollary 3.2.4). In particular, any E -weakly convergent sequence is mapped by ι to an X -strongly convergent sequence.

The covariance operator $R_\gamma : X_\gamma^* \rightarrow E$ is a Hilbert isometric isomorphism between the RKHS (equipped with the $L^2(\gamma)$ -inner product) and the Cameron–Martin space (equipped with the Cameron–Martin inner product).

The Cameron–Martin space for γ is precisely the set of all directions $h \in X$ for which the shifted measure $\gamma_h(\cdot) := \gamma(\cdot - h)$ is absolutely continuous with respect to γ . The space E has γ -measure zero, but if γ is nondegenerate then E is dense in X . When $h \in E$, the density of the shifted measure γ_h with respect to γ is given by the *Cameron–Martin formula* (Bogachev 1998, corollary 2.4.3),

$$\gamma_h(\mathrm{d}x) = \exp\left(\left(R_\gamma^{-1}h\right)(x) - \frac{1}{2}\|h\|_E^2\right) \gamma(\mathrm{d}x). \quad (2.5)$$

If $h \notin E$, then the measures γ and γ_h are mutually singular by the Feldman–Hájek theorem (Bogachev 1998, theorem 2.7.2).

A centred Gaussian measure γ has OM functional

$$I : E \rightarrow \mathbb{R}, \quad I(u) = \frac{1}{2}\|u\|_E^2, \quad (2.6)$$

which is defined only on the Cameron–Martin space E . Property $M(\gamma, E)$ is known to hold when X is a separable Hilbert space, as proven by Dashti *et al* (2013), corollary 3.8 and

Ayanbayev *et al* (2022a), corollary 5.2, and when $X = \ell^p$, $1 \leq p < \infty$, provided that γ has diagonal covariance structure (Klebanov and Wacker 2023, lemma 4.5). The measure γ also satisfies *Anderson's inequality* (Bogachev 1998, theorem 2.8.10):

$$\gamma(B_r(x)) \leq \gamma(B_r(0)) \quad \text{for any } x \in X \text{ and } r > 0. \quad (2.7)$$

Gaussian measures do not charge the boundaries of metric balls, i.e. $\gamma(\partial B_r(x)) = 0$ (see e.g. Agapiou *et al* 2018, lemma 6.1), so it would be equivalent to use open balls in any of the results in this article.

The tail behaviour of a Gaussian measure is described by Fernique's theorem (Fernique 1970), and this is the chief reason for the lower bound (1.3) on the potential needed in theorem 1.1. Fernique's theorem states that for any Gaussian measure γ on a separable Banach space X , there exists $\eta > 0$ such that

$$\int_X \exp(\eta \|x\|_X^2) \gamma(dx) < \infty.$$

In the rest of the article, γ will denote a centred nondegenerate Gaussian measure and the prior measure μ_0 will always be a centred nondegenerate Gaussian; in either case, E will denote the corresponding Cameron–Martin space.

2.3. Bayesian inverse problems

Ill-posed inverse problems are challenging to solve and require the use of prior information about the solution x to restore the well-posedness of the problem. The motivating example in this article is the nonlinear inverse problem of recovering an infinite-dimensional parameter (e.g. a function) $x \in X$ from a noisy observation of the finite-dimensional quantity $y = \mathcal{G}(x)$, as discussed in the introduction.

Well-posedness is essential to allow for numerical solution of inverse problems, and the classical approach to restoring well-posedness uses regularisation (see e.g. Benning and Burger 2018): a variational solution to the inverse problem (1.2) is a minimiser of the *Tikhonov functional*

$$F(x) = \|\mathcal{G}(x) - y\| + \alpha \|x\|',$$

where $\|\cdot\|'$ is some norm penalising undesirable properties of the solution x , e.g. the total-variation norm of the function x (Rudin *et al* 1992).

In contrast, the Bayesian approach incorporates prior information using a prior measure on the solution space. As stated in the next theorem, under mild conditions on the prior μ_0 and on the potential Φ arising from the observation operator \mathcal{G} , an analogue of Bayes' rule gives an expression for the posterior for $x | y$ on the infinite-dimensional parameter space.

Theorem 2.7 (Dashti and Stuart 2017, theorem 14). *Let X and Y be separable Banach spaces and suppose that $\mathcal{G}: X \rightarrow Y$ is measurable. Suppose that x has prior distribution $\mu_0 \in \mathcal{P}(X)$ and*

$$y = \mathcal{G}(x) + \xi,$$

where ξ is random noise with distribution $\tau_0 \in \mathcal{P}(Y)$, which is assumed to be independent of x . Suppose that the translated measure $\tau_{\mathcal{G}(x)}(\cdot) := \tau_0(\cdot - \mathcal{G}(x))$ is absolutely continuous with respect to τ_0 for μ_0 -almost all $x \in X$ and define the potential

$$\Phi(x; y) := -\log \frac{d\tau_{\mathcal{G}(x)}}{d\tau_0}(y).$$

Suppose further that $\Phi : X \times Y \rightarrow \mathbb{R}$ is measurable with respect to the product measure $\mu_0 \otimes \tau_0$, and that for τ_0 -almost all $y \in Y$,

$$Z(y) := \int_X \exp(-\Phi(x; y)) \mu_0(dx) > 0. \quad (2.8)$$

Then the conditional distribution μ^y of $x \mid y$ exists, is absolutely continuous with respect to μ_0 , and

$$\frac{d\mu^y}{d\mu_0}(x) = \frac{1}{Z(y)} \exp(-\Phi(x; y)). \quad (2.9)$$

The data $y \in Y$ will be considered fixed and we suppress the explicit dependence on y ; thus, the potential is a map $\Phi : X \rightarrow \mathbb{R}$. When y has finite dimension and τ_0 is absolutely continuous with respect to the Lebesgue measure, $\tau_{\mathcal{G}(x)}$ is absolutely continuous with respect to τ_0 and $\Phi(x)$ can typically be interpreted as a misfit functional: when ξ has mean-zero Gaussian distribution $\xi \sim N(0, \Sigma)$ on $Y = \mathbb{R}^d$, for example, one can take

$$\Phi(x) \propto \|\mathcal{G}(x) - y\|_{\Sigma}^2, \quad \|\cdot\|_{\Sigma} = \|\Sigma^{-1/2} \cdot\|. \quad (2.10)$$

By absorbing the normalisation factor $\frac{1}{Z(y)}$ into Φ , the posterior (2.9) can be expressed in the form (1.1) discussed in the introduction.

To ensure that the posterior measure is normalisable for a given potential $\Phi : X \rightarrow \mathbb{R}$, i.e. is a probability measure, we impose mild conditions on the form of the potential Φ in theorem 1.1. If the measure μ^y does indeed arise from an inverse problem as in theorem 2.7, the following result is merely a sufficient condition to ensure that $Z(y) > 0$ in (2.8).

Theorem 2.8 (Stuart 2010, theorem 4.1). *Suppose that the potential $\Phi : X \rightarrow \mathbb{R}$ is continuous and that for each $\eta > 0$, there exists a constant $K(\eta) \in \mathbb{R}$ such that*

$$\Phi(x) \geq K(\eta) - \eta \|x\|_X^2 \text{ for all } x \in X. \quad (2.11)$$

Then the posterior measure μ^y given by (1.1) can be normalised to yield a probability measure.

Proof. Given the unnormalised density $\exp(-\Phi)$, one can normalise to obtain a probability measure with density $\exp(-\Phi')$ by setting $\Phi' = \Phi - \log Z$ with the finite normalisation constant

$$Z := \int_X \exp(-\Phi(x)) \mu_0(dx) \leq \int_X \exp(-K(\eta) + \eta \|x\|_X^2) \mu_0(dx) < \infty,$$

where the upper bound follows by applying (2.11) with an appropriate $\eta > 0$ such that the integral is finite by Fernique's theorem. \square

As discussed, a significant reason for studying MAP estimators is that they connect the Bayesian and variational approaches to inverse problems. When the M -property holds, the weak MAP estimators of a Bayesian inverse problem coincide with minimisers of an OM functional, and when a Gaussian prior is used, an OM functional for the posterior has the form of a Tikhonov functional (see e.g. Dashti *et al* 2013). This correspondence depends on the M -property, which until now has been shown only for Gaussian measures on separable Hilbert spaces and for diagonal Gaussian measures on $X = \ell^p$, $1 \leq p < \infty$. This article therefore extends the connection between Bayesian and variational approaches to Banach spaces.

3. Small-ball probabilities for Gaussian measures in Banach spaces

The main technical result required for the proof of theorem 1.1 is the following bound on the ratio of the measures of small balls under a Gaussian measure stated in proposition 3.1. This bound is similar in spirit to the explicit Anderson inequality of Dashti *et al* (2013), lemma 3.6, which takes the form

$$\frac{\gamma(B_r(x))}{\gamma(B_r(0))} \leq \exp\left(a\left(r^2 - (\|x\|_X - r)^2\right)\right) \quad (3.1)$$

when γ is a centred nondegenerate Gaussian measure on the separable Banach space X , $a = a(\gamma) > 0$, $x \in X$ and $r > 0$. Both (3.1) and the bound we prove in proposition 3.1 may be thought of as quantitative analogues of the Anderson inequality (2.7). In contrast to the inequality (3.1), which is written in terms of the ambient norm of the Banach space, the result here is written in terms of the *decentring function* (Ghosal and van der Vaart 2017) given by

$$\psi_x(r) := \inf_{h \in E \cap B_r(x)} \frac{1}{2} \|h\|_E^2.$$

We will show in proposition 3.1 that the infimum in the decentring function is attained by some point $h^* \in E$, justifying the use of a minimum instead.

When X is a separable Hilbert space, the Cameron–Martin norm can be viewed as a reweighting of the norm of X and the bound (3.1) in X -norm suffices to prove the desired results on MAP estimators. In a Banach space, however, this is no longer true—thus, writing the bound in terms of the Cameron–Martin norm is a natural generalisation, with the compact embedding (2.4) providing the means to relate the two norms.

Proposition 3.1 (Explicit Anderson inequality in Cameron–Martin norm). *Let X be a separable Banach space equipped with a centred nondegenerate Gaussian measure γ . For any $x \in X$ and $r > 0$,*

$$\frac{\gamma(B_r(x))}{\gamma(B_r(0))} \leq \exp\left(-\min_{h \in E \cap B_r(x)} \frac{1}{2} \|h\|_E^2\right).$$

Proof. This is an immediate corollary of Ghosal and van der Vaart (2017), proposition 11.19, and we give a version of the proof here. The set $E \cap B_r(x)$ is nonempty (as γ is nondegenerate), E -closed (as it is the preimage of $B_r(x)$ under the continuous embedding $\iota: E \rightarrow X$) and convex. This implies that $E \cap B_r(x)$ is E -weakly closed. Hence, the E -weakly lower semicontinuous map $h \mapsto \|h\|_E^2$ defined on $E \cap B_r(x)$ attains its minimum on some $h^* = R_\gamma g^* \in E$. The Cameron–Martin formula (2.5) gives the equality

$$\gamma(B_r(x)) = \gamma_{-h^*}(B_r(x - h^*)) = \exp\left(-\frac{1}{2} \|h^*\|_E^2\right) \int_{B_r(x - h^*)} \exp(-g^*(u)) \gamma(du), \quad (3.2)$$

and we now show that $g^*(u) \geq 0$ for γ -almost all $u \in B_r(x - h^*)$. As $E \cap B_r(x)$ is convex and h^* minimises the E -norm on $E \cap B_r(x)$, it follows that

$$\|(1 - \lambda)h^* + \lambda h\|_E^2 \geq \|h^*\|_E^2 \text{ for any } h \in E \cap B_r(x) \text{ and } \lambda \in [0, 1].$$

Rearranging and taking limits as $\lambda \rightarrow 0$ shows that

$$\langle h, h^* \rangle_E \geq \|h^*\|_E^2 \text{ for any } h \in E \cap B_r(x). \quad (3.3)$$

Now, let $(h_n)_{n \in \mathbb{N}}$ be any orthonormal basis of E . As the covariance operator R_γ is a Hilbert isomorphism and $\sum_{k=1}^n h_k \langle h_k, h^* \rangle_E \rightarrow h^*$ in E as $n \rightarrow \infty$, it follows that

$$g_n := \sum_{k=1}^n (R_\gamma^{-1} h_k) \langle h_k, h^* \rangle_E = R_\gamma^{-1} \left[\sum_{k=1}^n h_k \langle h_k, h^* \rangle_E \right] \rightarrow g^* \text{ in } L^2(\gamma) \text{ as } n \rightarrow \infty.$$

Hence, there is a subsequence $(g_{n_k})_{k \in \mathbb{N}}$ converging pointwise γ -almost everywhere to g^* . By Bogachev (1998), theorem 3.5.1, γ -almost all elements $u \in B_r(x - h^*)$ may be written as

$$u = \sum_{k \in \mathbb{N}} (R_\gamma^{-1} h_k)(u) h_k,$$

where the convergence of the series is in the norm of X . Hence, for all n sufficiently large and $u \in B_r(x - h^*)$,

$$\sum_{k=1}^n (R_\gamma^{-1} h_k)(u) h_k + h^* \in B_r(x).$$

Using (3.3), we observe that

$$g_n(u) + \|h^*\|_E^2 = \left\langle \sum_{k=1}^n (R_\gamma^{-1} h_k) h_k + h^*, h^* \right\rangle_E \geq \|h^*\|_E^2,$$

so it immediately follows that $g_n(u) \geq 0$. As $g_n(u) \rightarrow g^*(u)$ as $n \rightarrow \infty$ γ -almost everywhere, we obtain the claimed lower bound $g^*(u) = \lim_{n \rightarrow \infty} g_n(u) \geq 0$ for γ -almost all $u \in B_r(x - h^*)$. The result follows by bounding the integrand in (3.2) and using Anderson's inequality (2.7):

$$\begin{aligned} \gamma(B_r(x)) &= \exp\left(-\frac{1}{2}\|h^*\|_E^2\right) \int_{B_r(x-h^*)} \exp(-g^*(u)) \gamma(du) \\ &\leq \exp\left(-\frac{1}{2}\|h^*\|_E^2\right) \gamma(B_r(x-h^*)) \\ &\leq \exp\left(-\frac{1}{2}\|h^*\|_E^2\right) \gamma(B_r(0)). \end{aligned}$$

□

Though we shall not make use of (3.1), it can be proven easily from proposition 3.1 by applying the compact embedding (2.4).

The following corollary on the measure of balls with centres converging to some $x^* \in E$ is slightly weaker than the corresponding results of Kretschmann (2019), lemma 4.14 and Klebanov and Wacker (2023), lemma A.2, but it is sufficient for our purposes. The proof stated here takes advantage of the bound developed in proposition 3.1.

Corollary 3.2. *Let X be a separable Banach space equipped with a centred nondegenerate Gaussian measure γ . Suppose that $(x_n, r_n)_{n \in \mathbb{N}} \subset X \times [0, \infty)$ converges to $(x^*, 0)$. After passing to a subsequence without relabelling,*

$$\limsup_{n \rightarrow \infty} \frac{\gamma(B_{r_n}(x_n))}{\gamma(B_{r_n}(x^*))} \leq 1.$$

Proof. Construct the sequence $(h_n)_{n \in \mathbb{N}} \subset E$ by selecting a minimiser (which exists as argued in the proof of proposition 3.1) of $h \mapsto \|h\|_E^2$ from $E \cap B_{r_n}(x_n)$. Using the OM functional I defined by (2.6) for γ , which satisfies $I(0) = 0$, and by applying the upper bound from proposition 3.1, we may write

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{\gamma(B_{r_n}(x_n))}{\gamma(B_{r_n}(x^*))} &= \limsup_{n \rightarrow \infty} \frac{\gamma(B_{r_n}(x_n))}{\gamma(B_{r_n}(0))} \lim_{n \rightarrow \infty} \frac{\gamma(B_{r_n}(0))}{\gamma(B_{r_n}(x^*))} \\ &\leq \limsup_{n \rightarrow \infty} \exp\left(-\frac{1}{2}\|h_n\|_E^2 + I(x^*)\right). \end{aligned}$$

If $(h_n)_{n \in \mathbb{N}}$ has no E -bounded subsequence, then the claim follows immediately as the limit on the right-hand side is zero. Otherwise, pass to an E -bounded subsequence and, by reflexivity of E , pass to a further E -weakly convergent subsequence which we do not relabel. Since $(x_n)_{n \in \mathbb{N}} \rightarrow x^*$, it follows that $(h_n)_{n \in \mathbb{N}} \rightarrow x^*$ in X as $\|h_n - x_n\|_X \leq r_n$, and by the compact embedding of E in X , the E -weak limit of $(h_n)_{n \in \mathbb{N}}$ must agree with the X -strong limit. Hence, $(h_n)_{n \in \mathbb{N}} \rightharpoonup x^*$ weakly in E , and as the Cameron–Martin norm is E -weakly lower semi-continuous,

$$\limsup_{n \rightarrow \infty} \exp\left(-\frac{1}{2}\|h_n\|_E^2\right) \leq \exp\left(-\frac{1}{2}\|x^*\|_E^2\right) = \exp(-I(x^*)).$$

□

The next result establishes a technical approximation condition for sequences in X by elements of E , which is useful in combination with proposition 3.1, and applies it to establish property $M(\gamma, E)$ for Gaussian measures on Banach spaces. As discussed in section 2, this extends previous results which establish the M -property when X is a separable Hilbert space or when $X = \ell^p$, $1 \leq p < \infty$, and γ is a diagonal Gaussian measure. In particular, this is a natural analogue for Banach spaces of corollary 3.8 of Dashti *et al* (2013), which proves the M -property in separable Hilbert spaces.

Corollary 3.3. *Let X be a separable Banach space equipped with a centred nondegenerate Gaussian measure γ .*

(a) *Let $(r_n)_{n \in \mathbb{N}} \rightarrow 0$ and $(x_n)_{n \in \mathbb{N}} \subset X$. If*

$$\liminf_{n \rightarrow \infty} \min_{h \in E \cap B_{r_n}(x_n)} \frac{1}{2}\|h\|_E^2 < \infty,$$

then $(x_n)_{n \in \mathbb{N}}$ has an X -strong limit point which lies in E .

(b) *Property $M(\gamma, E)$ holds.*

Proof. (a) By hypothesis, there must exist a subsequence $(r_{n_k})_{k \in \mathbb{N}}$ and a sequence $(h_{n_k})_{k \in \mathbb{N}} \subset E$ that is uniformly bounded in E such that $\|h_{n_k} - x_{n_k}\|_X \leq r_{n_k}$. Pass to an E -weakly convergent subsequence of $(h_{n_k})_{k \in \mathbb{N}}$ with limit $h^* \in E$; the compact embedding of E in X implies that $(h_{n_k})_{k \in \mathbb{N}}$ converges strongly in X to h^* . As $\|h_{n_k} - x_{n_k}\|_X \leq r_{n_k}$, this implies that $(x_{n_k})_{k \in \mathbb{N}} \rightarrow h^*$ strongly in X .

(b) Let $x \notin E$. The constant sequence $(x)_{n \in \mathbb{N}}$ cannot have a limit point in E , so by (a), for any sequence $(r_n)_{n \in \mathbb{N}} \rightarrow 0$,

$$\lim_{n \rightarrow \infty} \min_{h \in E \cap B_{r_n}(x)} \frac{1}{2}\|h\|_E^2 = \infty.$$

Thus, by proposition 3.1, the M -property holds because

$$\limsup_{r \rightarrow 0} \frac{\gamma(B_r(x))}{\gamma(B_r(0))} \leq \limsup_{r \rightarrow 0} \exp \left(- \min_{h \in E \cap B_r(x)} \frac{1}{2} \|h\|_E^2 \right) = 0.$$

□

4. Existence of MAP estimators

4.1. Weak MAP estimators

With the M -property established for Gaussian measures on a separable Banach space, it is now possible to provide a short proof of the existence of weak MAP estimators for Bayesian posteriors of the form (1.1). One could prove the existence of strong MAP estimators directly, as in Dashti *et al* (2013), and use the fact that all strong modes are weak modes, but it is instructive to prove the existence of weak MAP estimators separately. Though weak modes were not proposed until the work of Helin and Burger (2015), corollary 3.8 of Dashti *et al* (2013) already proved what is now called the M -property for Gaussian priors on Hilbert spaces, taking an important step towards showing the existence of weak modes.

By proposition 2.6, it is sufficient to minimise the posterior OM functional F , and it is well known that F does indeed have a minimiser (see e.g. Stuart 2010, theorem 5.4) under mild conditions.

In particular, we only require *coercivity* of F in E to obtain weak modes rather than the lower bound on Φ needed in theorem 1.1. It is important to note that without the lower bound on Φ , it may not be possible to normalise the measure defined in (1.1) as theorem 2.8 need not hold; the following result considers only measures which *can* be normalised.

Observe also that the hypotheses of theorem 1.1 always imply E -coercivity of F : using the compact embedding (2.4) of E in X and the lower bound (1.3) gives

$$F(u) := \Phi(u) + \frac{1}{2} \|u\|_E^2 \geq K(\eta) + \left(\frac{1}{2} - C^2\eta \right) \|u\|_E^2 \text{ for all } u \in E,$$

and selecting $\eta > 0$ sufficiently small ensures that $\frac{1}{2} - C^2\eta > 0$.

It is not clear whether coercivity is sufficient to obtain a strong mode, and this question is left to future work.

Proposition 4.1 (Weak MAP estimators for Bayesian posteriors with Gaussian priors).

Let X be a separable Banach space and let μ_0 be a centred nondegenerate Gaussian measure. Suppose that μ^γ is a probability measure of the form (1.1) for some continuous potential $\Phi : X \rightarrow \mathbb{R}$. Suppose also that the posterior OM functional $F(u) := \Phi(u) + \frac{1}{2} \|u\|_E^2$ is E -coercive, i.e. there exists $A \in \mathbb{R}$ and $c > 0$ such that

$$A + F(u) \geq c \|u\|_E^2 \text{ for all } u \in E,$$

or equivalently, using the definition of F ,

$$A + \Phi(u) > -\frac{1}{2} \|u\|_E^2 \text{ for all } u \in E.$$

Then μ^γ has a weak mode.

Proof. The prior μ_0 has OM functional $I_0(u) = \frac{1}{2} \|u\|_E^2$ as described in (2.6) and corollary 3.3 proves that property $M(\mu_0, E)$ holds. Hence, by proposition 2.6, the posterior has OM functional $F(u) = \frac{1}{2} \|u\|_E^2 + \Phi(u)$ and property $M(\mu^\gamma, E)$ holds, and furthermore weak modes coincide with minimisers of F . It remains to show that F does have a minimiser.

First, note that Φ is E -weakly continuous: if $u_n \rightharpoonup u$ weakly in E , then by the compact embedding $u_n \rightarrow u$ strongly in X and thus $\Phi(u_n) \rightarrow \Phi(u)$ by strong continuity of Φ in X . As the E -norm is also clearly weakly lower semicontinuous, the OM functional F must be E -weakly lower semicontinuous. As F is also coercive, it has a minimiser in E by the direct method of the calculus of variations: take a sequence $(h_n)_{n \in \mathbb{N}} \subset E$ with $F(h_n) < \inf_{u \in E} F(u) + \frac{1}{n}$, and observe that it is E -bounded by coercivity; passing to an E -weakly convergent subsequence with limit h^* and using the weak lower semicontinuity of F proves that h^* is a minimiser of F . This minimiser is a weak mode by proposition 2.6. \square

4.2. Strong MAP estimators

We now prove the main theorem on the existence of strong MAP estimators. The strategy of the proof is similar in spirit to the prior work of Dashti *et al* (2013), Kretschmann (2019, 2023) and Klebanov and Wacker (2023).

In the proof of Dashti *et al* (2013), the explicit Anderson inequality (3.1) is first used to show that any family $(x_r^*)_{r>0}$ of maximisers of the posterior radius- r ball mass $x \mapsto \mu^y(B_r(x))$ must be bounded in X under some regularity assumptions on Φ . Next, a weakly convergent subsequence is extracted, and lemmas 3.7 and 3.9 of Dashti *et al* (2013) can be used to show that if the limit is not in E or the convergence is not strong, then

$$\frac{\gamma(B_r(x_r^*))}{\gamma(B_r(0))} \rightarrow 0 \quad \text{as } r \rightarrow 0.$$

This yields a contradiction because the assumptions on Φ mean this ratio cannot converge to zero, showing that the limit point lies in E and convergence is strong in X . Finally, this limit point is shown to be both a strong MAP estimator and an OM minimiser.

Klebanov and Wacker (2023) point out that it is not obvious that the radius- r maximisers exist and show that the proof can be adapted to use a family $(x_r)_{r>0}$ of ‘approximate maximisers’ nearly attaining the supremal radius- r mass instead. Klebanov and Wacker (2023) call such a family an *asymptotic maximising family*.

Definition 4.2. Let X be a metric space. An *asymptotic maximising family* (AMF) for $\mu \in \mathcal{P}(X)$ is a net $(x_r)_{r>0} \subset X$ such that, for some increasing function $\varepsilon: [0, \infty) \rightarrow [0, 1)$ with $\lim_{r \rightarrow 0} \varepsilon(r) = 0$,

$$\mu(B_r(x_r)) \geq (1 - \varepsilon(r))M_r.$$

Every measure has at least one AMF, though in general there may not exist any point $x_r^* \in X$ such that $\mu(B_r(x_r^*)) = M_r$. If X is a Hilbert space, then the radius- r maximisers x_r^* do always exist (Lambley and Sullivan 2023, corollary A.9), but we will use AMFs to avoid further discussion about these maximisers. The next result summarises the connection between AMFs and small-ball modes, which is explored in greater detail by Lambley and Sullivan (2023), theorem 4.11.

Proposition 4.3. Let X be a separable Banach space and suppose that $\mu \in \mathcal{P}(X)$.

- (a) Suppose that x^* is a generalised strong mode for μ . Then x^* is a limit point of some AMF $(x_r)_{r>0} \subset X$.
- (b) Suppose that $(x_r)_{r>0} \subset X$ is an AMF for μ which converges to x^* along every subsequence. Then x^* is a generalised strong mode.

Proof. Pick any sequence $(r_n)_{n \in \mathbb{N}} \rightarrow 0$ and choose a corresponding sequence $(x_{r_n})_{n \in \mathbb{N}} \rightarrow x^*$ from the definition of a generalised strong mode (definition 2.3). Selecting any AMF $(x_r)_{r>0}$ with this subsequence $(x_{r_n})_{n \in \mathbb{N}}$ proves the first claim. For the second claim, let $\varepsilon: [0, \infty) \rightarrow [0, 1)$ denote the function corresponding to the AMF $(x_r)_{r>0}$; for any sequence $(r_n)_{n \in \mathbb{N}} \rightarrow 0$, it follows by definition that

$$\lim_{n \rightarrow \infty} \frac{\mu(B_{r_n}(x_{r_n}))}{M_{r_n}} = \lim_{n \rightarrow \infty} 1 - \varepsilon(r_n) = 1,$$

proving that x^* is a generalised strong mode. \square

Aside from the issues associated with radius- r maximisers, the proof of Dashti *et al* (2013) omits some technical results which were later proved by Kretschmann (2019). Klebanov and Wacker (2023) argue that the proof also relies on several properties that do not hold in an arbitrary separable Banach space X . To give just one example, the step passing from a bounded sequence to a weakly convergent subsequence requires additional hypotheses, e.g. reflexivity of X .

To resolve this, Klebanov and Wacker (2023) first establish the proof when X is a separable Hilbert space. In this setting, any Gaussian measure γ is characterised by its mean $m \in X$ and covariance operator $\mathcal{C}: X \rightarrow X$, so by working in an eigenbasis of \mathcal{C} , one can reduce to the case $X = \ell^2(\mathbb{N}; \mathbb{R})$ with $\mu_0 = \bigotimes_{n \in \mathbb{N}} N(0, \sigma_n^2)$, with the Cameron–Martin norm given by a simple reweighting of the ℓ^2 -norm. Klebanov and Wacker (2023) then extend to the case $X = \ell^p(\mathbb{N}; \mathbb{R})$, $1 \leq p < \infty$, with $\mu_0 = \bigotimes_{n \in \mathbb{N}} N(0, \sigma_n^2)$; unlike in the Hilbert case, not all Gaussian measures on ℓ^p can be expressed in this product form. Even this generalisation is nontrivial since the Cameron–Martin norm can no longer be expressed as a reweighting of the X -norm. This motivates a technical convexification argument to bridge the gap between the two norms, making use of the diagonal structure of the prior to write the E -norm in terms of the canonical sequence-space basis. It is challenging to generalise this approach further given the heavy dependence on the diagonal structure.

We overcome this difficulty by using the explicit Anderson inequality of proposition 3.1. As discussed in section 3, this is more natural than the bound (3.1) used in prior work because the behaviour of γ is fully determined by its Cameron–Martin space, and the Cameron–Martin space has more favourable topological properties. proposition 3.1 first allows us to show that any AMF is bounded in X , and corollary 3.3 shows that any AMF is closely approximated in X by a sequence bounded in E . This sequence has an E -weakly convergent subsequence regardless of the choice of X , and applying the compact embedding of E in X yields strong convergence of this subsequence in X .

This approach avoids the need to explicitly prove lemmas 3.7 and 3.9 of Dashti *et al* (2013), since the necessary claims can be derived directly from proposition 3.1 and corollary 3.3.

We will later show in lemma 4.5 that a limit point of an AMF is a strong mode for the Bayesian posterior μ^y ; combining this result with the existence of limit points proven in the following result completes the proof of theorem 1.1.

Lemma 4.4 (Limit points of AMFs for Bayesian posteriors). *Under the assumptions of theorem 1.1, if $(x_r)_{r>0} \subset X$ is an AMF for μ^y , then:*

- (a) any limit point of $(x_r)_{r>0}$ lies in E ;
- (b) the net $(x_r)_{r>0}$ has at least one limit point.

Proof. Fix any decreasing sequence $(r_n)_{n \in \mathbb{N}} \rightarrow 0$. By proposition 3.1 and the compact embedding (2.4), we have

$$\frac{\mu_0(B_{r_n}(x_{r_n}))}{\mu_0(B_{r_n}(0))} \leq \exp\left(-\frac{1}{2C^2} \min_{h \in E \cap B_{r_n}(x_{r_n})} \|h\|_X^2\right) \leq \exp\left(-\frac{1}{2C^2} \inf_{x \in B_{r_n}(x_{r_n})} \|x\|_X^2\right). \quad (4.1)$$

On the other hand, let ε be the function corresponding to the AMF $(x_r)_{r>0}$; using the lower bound (1.3) on Φ and picking $\delta > 0$ from the definition of continuity such that $|\Phi(x) - \Phi(0)| < 1$ for $|x| < \delta$, we see that for all n such that $r_n < \delta$, the following lower bound holds:

$$\begin{aligned} \exp\left(-K(\eta) + \eta \sup_{x \in B_{r_n}(x_{r_n})} \|x\|_X^2\right) \mu_0(B_{r_n}(x_{r_n})) &\geq \mu^y(B_{r_n}(x_{r_n})) \\ &\geq (1 - \varepsilon(r_n)) M_{r_n} \\ &\geq (1 - \varepsilon(r_n)) \int_{B_{r_n}(0)} \exp(-\Phi(x)) \mu_0(dx) \\ &\geq (1 - \varepsilon(r_n)) \exp(-\Phi(0) - 1) \mu_0(B_{r_n}(0)). \end{aligned}$$

This inequality gives

$$\frac{\mu_0(B_{r_n}(x_{r_n}))}{\mu_0(B_{r_n}(0))} \geq (1 - \varepsilon(r_n)) \exp\left(K(\eta) - \Phi(0) - 1 - \eta \sup_{x \in B_{r_n}(x_{r_n})} \|x\|_X^2\right), \quad (4.2)$$

and combining this bound with (4.1) yields

$$(1 - \varepsilon(r_n)) \exp(K(\eta) - \Phi(0) - 1) \leq \exp\left(\eta \sup_{x \in B_{r_n}(x_{r_n})} \|x\|_X^2 - \frac{1}{2C^2} \inf_{x \in B_{r_n}(x_{r_n})} \|x\|_X^2\right).$$

This implies that the sequence $(x_{r_n})_{n \in \mathbb{N}}$ is bounded: if it were not, then setting $\eta < \frac{1}{2C^2}$ would give the contradiction

$$\begin{aligned} 0 &< \exp(K(\eta) - \Phi(0) - 1) \\ &\leq \liminf_{n \rightarrow \infty} \exp\left(\eta \sup_{x \in B_{r_n}(x_{r_n})} \|x\|_X^2 - \frac{1}{2C^2} \inf_{x \in B_{r_n}(x_{r_n})} \|x\|_X^2\right) \\ &\leq \liminf_{n \rightarrow \infty} \exp\left(\eta \left(\inf_{x \in B_{r_n}(x_{r_n})} \|x\|_X + 2r_n\right)^2 - \frac{1}{2C^2} \inf_{x \in B_{r_n}(x_{r_n})} \|x\|_X^2\right) = 0. \end{aligned}$$

As $(x_{r_n})_{n \in \mathbb{N}}$ is bounded, (4.2) implies that there is a constant $L > 0$ such that

$$0 < L(1 - \varepsilon(r_n)) \leq \frac{\mu_0(B_{r_n}(x_{r_n}))}{\mu_0(B_{r_n}(0))}.$$

Thus, again using the upper bound provided by proposition 3.1, we see that

$$0 < \liminf_{n \rightarrow \infty} L(1 - \varepsilon(r_n)) \leq \liminf_{n \rightarrow \infty} \frac{\mu_0(B_{r_n}(x_{r_n}))}{\mu_0(B_{r_n}(0))} \leq \liminf_{n \rightarrow \infty} \exp\left(-\min_{h \in E \cap B_{r_n}(x_{r_n})} \frac{1}{2} \|h\|_E^2\right).$$

It then follows from corollary 3.3 that $(x_{r_n})_{n \in \mathbb{N}}$ has a further subsequence converging to some point $x^* \in E$. In particular, if $(x_{r_n})_{n \in \mathbb{N}}$ is a convergent sequence, then the limit must lie in E . \square

As discussed, theorem 2.8 of Klebanov and Wacker (2023) proves that a limit point of an AMF for the posterior is a strong mode in the sequence-space setting. The following lemma generalises this result to any separable Banach space and slightly weakens the hypotheses required on the potential to be merely continuous rather than locally Lipschitz.

Lemma 4.5 (Limit points of AMFs are strong modes). *Under the assumptions of theorem 1.1, any X -strong limit point $x^* \in E$ of an AMF $(x_r)_{r>0} \subset X$ for μ^y is a strong mode.*

Proof. Let $x^* \in E$ be some limit point of $(x_r)_{r>0}$. To show that x^* is a strong mode, it suffices to check that for any $(r_n)_{n \in \mathbb{N}} \rightarrow 0$,

$$\lim_{n \rightarrow \infty} \frac{\mu^y(B_{r_n}(x^*))}{M_{r_n}} \geq 1. \tag{4.3}$$

Indeed, it would be enough to show that any $(r_n)_{n \in \mathbb{N}} \rightarrow 0$ has a further subsequence such that (4.3) holds along that subsequence: this follows from the fact that if $(u_n)_{n \in \mathbb{N}}$ is an arbitrary real sequence and any subsequence of $(u_n)_{n \in \mathbb{N}}$ has a further subsequence converging to u , then $u_n \rightarrow u$.

Hence, take any sequence $(r_n)_{n \in \mathbb{N}} \rightarrow 0$; by lemma 4.4 we may pass to a subsequence of $(r_n)_{n \in \mathbb{N}}$, which will not be relabelled, such that $(x_{r_n})_{n \in \mathbb{N}}$ converges to some $y^* \in E$. As Φ is continuous, for any $\varepsilon > 0$ there exists $\delta > 0$ such that for any $x \in B_\delta(y^*)$, it follows that $|\Phi(y^*) - \Phi(x)| < \varepsilon$. Since $(r_n)_{n \in \mathbb{N}} \rightarrow 0$ and $(x_{r_n})_{n \in \mathbb{N}} \rightarrow y^*$, there exists $N \in \mathbb{N}$ such that $|r_n| < \frac{\delta}{2}$ and $\|x_{r_n} - y^*\|_X < \frac{\delta}{2}$ for $n \geq N$. Hence for such n and any $x \in B_{r_n}(x_{r_n})$, we have $\Phi(x_{r_n}) - \Phi(x) < 2\varepsilon$. This implies that

$$\begin{aligned} \frac{\mu^y(B_{r_n}(x_{r_n}))}{\mu^y(B_{r_n}(y^*))} &= \exp(\Phi(y^*) - \Phi(x_{r_n})) \frac{\int_{B_{r_n}(x_{r_n})} \exp(\Phi(x_{r_n}) - \Phi(x)) \mu_0(dx)}{\int_{B_{r_n}(y^*)} \exp(\Phi(y^*) - \Phi(x)) \mu_0(dx)} \\ &< \exp(4\varepsilon) \frac{\mu_0(B_{r_n}(x_{r_n}))}{\mu_0(B_{r_n}(y^*))}. \end{aligned} \tag{4.4}$$

By corollary 3.2, we may pass to a further subsequence of $(x_{r_n})_{n \in \mathbb{N}}$ without relabelling such that

$$\limsup_{n \rightarrow \infty} \frac{\mu_0(B_{r_n}(x_{r_n}))}{\mu_0(B_{r_n}(y^*))} \leq 1,$$

and thus taking the limsup in (4.4) and infimising over $\varepsilon > 0$ yields

$$\left(\liminf_{n \rightarrow \infty} \frac{\mu^y(B_{r_n}(y^*))}{\mu^y(B_{r_n}(x_{r_n}))} \right)^{-1} = \limsup_{n \rightarrow \infty} \frac{\mu^y(B_{r_n}(x_{r_n}))}{\mu^y(B_{r_n}(y^*))} \leq \limsup_{n \rightarrow \infty} \frac{\mu_0(B_{r_n}(x_{r_n}))}{\mu_0(B_{r_n}(y^*))} \leq 1.$$

Since $(x_{r_n})_{n \in \mathbb{N}}$ is a subsequence of an AMF, the previous equation implies that

$$1 \leq \lim_{n \rightarrow \infty} \frac{\mu^y(B_{r_n}(x_{r_n}))}{M_{r_n}} \liminf_{n \rightarrow \infty} \frac{\mu^y(B_{r_n}(y^*))}{\mu^y(B_{r_n}(x_{r_n}))} = \lim_{n \rightarrow \infty} \frac{\mu^y(B_{r_n}(y^*))}{M_{r_n}} \leq 1.$$

In particular, the point x^* fixed at the start of the proof is a limit point of the AMF $(x_r)_{r>0}$, i.e. there exists $(s_n)_{n \in \mathbb{N}} \rightarrow 0$ such that $(x_{s_n})_{n \in \mathbb{N}} \rightarrow x^*$, so the above argument implies the existence of a subsequence such that

$$\lim_{n \rightarrow \infty} \frac{\mu^y(B_{s_n}(x^*))}{M_{s_n}} = 1. \tag{4.5}$$

This does not yet hold for *every* sequence $(r_n)_{n \in \mathbb{N}} \rightarrow 0$, only the specific sequence $(s_n)_{n \in \mathbb{N}}$. To complete the proof, fix an arbitrary $(r_n)_{n \in \mathbb{N}} \rightarrow 0$ and $y^* \in E$ as above. As μ^y has an OM functional F^y defined on E ,

$$\lim_{n \rightarrow \infty} \frac{\mu^y(B_{r_n}(x^*))}{\mu^y(B_{r_n}(y^*))} = \lim_{n \rightarrow \infty} \frac{\mu^y(B_{s_n}(x^*))}{\mu^y(B_{s_n}(y^*))} = \lim_{r \rightarrow 0} \frac{\mu^y(B_r(x^*))}{\mu^y(B_r(y^*))} = F^y(y^*) - F^y(x^*). \quad (4.6)$$

Hence, using (4.5) and (4.6), it follows that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\mu^y(B_{r_n}(x^*))}{M_{r_n}} &= \lim_{n \rightarrow \infty} \frac{\mu^y(B_{r_n}(y^*))}{M_{r_n}} \lim_{n \rightarrow \infty} \frac{\mu^y(B_{r_n}(x^*))}{\mu^y(B_{r_n}(y^*))} \\ &= \lim_{n \rightarrow \infty} \frac{\mu^y(B_{s_n}(x^*))}{\mu^y(B_{s_n}(y^*))} \geq \lim_{n \rightarrow \infty} \frac{\mu^y(B_{s_n}(x^*))}{M_{s_n}} = 1. \end{aligned}$$

Hence, for any $(r_n)_{n \in \mathbb{N}} \rightarrow 0$, there is a further subsequence for which (4.3) holds, and thus x^* is a strong mode. \square

Proof of theorem 1.1. (a) By lemma 4.4, any AMF $(x_r)_{r>0}$ has an X -strong limit point $x^* \in E$, and this limit point is a strong mode by lemma 4.5.
 (b) Lemma 2.2 proves that strong and weak modes coincide as a strong mode exists, and proposition 2.6 shows that weak modes coincide with minimisers of the OM functional. As any generalised strong mode must be the limit point of an AMF (proposition 4.3) and lemma 4.4 implies that such a point lies in E , lemma 4.5 implies that the generalised strong mode is also a strong mode. \square

5. Consistency of MAP estimators

We return to the additive-noise Bayesian inverse problem (1.2) discussed in the introduction. Suppose that X is a separable Banach space, $Y := \mathbb{R}^d$ and $\mathcal{G}: X \rightarrow Y$. For simplicity, we restrict attention to the case of mean-zero Gaussian noise ξ and Gaussian prior μ_0 , giving the model

$$y = \mathcal{G}(x) + \xi, \quad \xi \sim N(0, \Sigma), \quad x \sim \mu_0. \quad (5.1)$$

Under the frequentist assumption that there is a fixed true parameter $x^\dagger \in X$, consistency theory studies the behaviour of the posterior and point estimators—which depend on the random observations y_1, \dots, y_N —in the infinite-data or small-noise limit.

Classically, a sequence of posterior measures is consistent at x^\dagger if, for any neighbourhood U of x^\dagger , the posterior measure of U^c converges to zero in probability (Ghosal and van der Vaart 2017). For Bayesian inverse problems, this notion is often too restrictive: the parameter x^\dagger need not even be identifiable from the model, because there may exist $x \neq x^\dagger$ such that $\mathcal{G}(x) = \mathcal{G}(x^\dagger)$. Indeed, if \mathcal{G} is a bounded linear operator, then $\mathcal{G}(x^\dagger) = \mathcal{G}(x^\dagger + z)$ for any $z \in \ker \mathcal{G} \neq \emptyset$, so x^\dagger is never identifiable, and thus one cannot expect posterior consistency to hold. Posterior consistency is often possible to show if \mathcal{G} is known more explicitly, e.g. if it is the solution operator for a partial differential equation (see Knapik *et al* 2011, Agapiou *et al* 2013, Vollmer 2013), but we focus on the general case of a possibly nonlinear operator $\mathcal{G}: X \rightarrow Y$.

In a similar vein, one cannot expect a sequence of MAP estimators to be consistent estimators of x^\dagger , i.e. the MAP estimators need not converge in probability to x^\dagger . It is instead typical to study a weaker notion of consistency for MAP estimators, where one identifies a limit point x^* of any sequence of MAP estimators and shows that $\mathcal{G}(x^\dagger) = \mathcal{G}(x^*)$ (Dashti *et al* 2013, Agapiou *et al* 2018, Dunlop 2019). The results of Dashti *et al* (2013), section 4 on the consistency of

MAP estimators in the setting of (5.1) depend on the correspondence between strong MAP estimators and OM minimisers, and on the existence of strong MAP estimators. Thus, theorem 1.1 can be used to extend the applicability of these consistency results from separable Hilbert spaces to arbitrary separable Banach spaces.

5.1. Infinite-data limit

Assume that data y_1, \dots, y_N, \dots are repeated observations from the model (5.1) assuming the fixed parameter x^\dagger , i.e.

$$y_n = \mathcal{G}(x^\dagger) + \xi_n, \quad \xi_n \stackrel{\text{i.i.d.}}{\sim} N(0, \Sigma). \quad (5.2)$$

Using Bayes' rule (theorem 2.7) with the concatenated observation vector $\mathbf{y} = (y_1, \dots, y_N) \in \mathbb{R}^{Nd}$ and the appropriate product noise distribution, we see that the posterior is absolutely continuous with respect to μ_0 and has density

$$\frac{d\mu^{y_1, \dots, y_N}}{d\mu_0}(x) \propto \exp(-\Phi^{y_1, \dots, y_N}(x)), \quad \Phi^{y_1, \dots, y_N}(x) = \frac{1}{2} \sum_{n=1}^N \|\mathcal{G}(x) - y_n\|_\Sigma^2, \quad (5.3)$$

with $\|\cdot\|_\Sigma := \|\Sigma^{-1/2} \cdot\|$, analogous to the potential derived in (2.10). As discussed in section 2.3, the normalisation constant for the density can be absorbed into Φ^{y_1, \dots, y_N} by adding a constant depending on y_1, \dots, y_N to the potential. Assuming that \mathcal{G} is continuous, the conclusions of theorem 1.1 hold because Φ^{y_1, \dots, y_N} is continuous and bounded below, so for each N there exists at least one strong MAP estimator x_N of μ^{y_1, \dots, y_N} , which can be obtained by minimising the OM functional

$$P^{y_1, \dots, y_N}(x) = \frac{1}{2} \|x\|_E^2 + \Phi^{y_1, \dots, y_N}(x)$$

over all $x \in E$. Following the proof of Dashti *et al* (2013), theorem 4.1 and corollary 4.3, we obtain a weak consistency result for MAP estimators which applies more generally in any separable Banach space with any continuous observation operator \mathcal{G} .

Theorem 5.1. *Let X be a separable Banach space and let $\mathcal{G}: X \rightarrow \mathbb{R}^d$ be continuous. Let $x^\dagger \in X$ be arbitrary. Suppose that the data y_1, \dots, y_N, \dots are generated from the observation model (5.2) and μ^{y_1, \dots, y_N} is the corresponding posterior. For each $N \in \mathbb{N}$, let x_N be any strong MAP estimator of μ^{y_1, \dots, y_N} . Then:*

- (a) *there is a subsequence of $(\mathcal{G}(x_N))_{N \in \mathbb{N}}$ converging to $\mathcal{G}(x^\dagger)$ almost surely;*
- (b) *if $x^\dagger \in E$, there is a subsequence of $(x_N)_{N \in \mathbb{N}}$ converging to some $x^* \in E$ weakly in E almost surely, and $\mathcal{G}(x^*) = \mathcal{G}(x^\dagger)$.*

Since the proof is a straightforward adaptation of that of Dashti *et al* (2013), it is omitted.

5.2. Small-noise limit

Assume that y_1, \dots, y_N, \dots are observations from the sequence of models

$$y_n = \mathcal{G}(x^\dagger) + \frac{1}{n} \xi_n, \quad \xi_n \stackrel{\text{i.i.d.}}{\sim} N(0, \Sigma). \quad (5.4)$$

Unlike in the infinite-data scenario, we consider the sequence of posteriors μ^{y_N} obtained from just a single observation; again applying Bayes' rule we obtain that μ^{y_N} is absolutely continuous with respect to μ_0 and has density

$$\frac{d\mu^{y_N}}{d\mu_0}(x) \propto \exp(-\Phi^{y_N}(x)), \quad \Phi^{y_N}(x) = \frac{N^2}{2} \|\mathcal{G}(x) - y_N\|_{\Sigma}^2.$$

As before, the potential satisfies the hypotheses of theorem 1.1, and so at least one strong MAP estimator exists for each posterior μ^{y_N} . Following the proof of Dashti *et al* (2013), theorem 4.4 and corollary 4.5, we also obtain weak consistency for MAP estimators in the small-noise setting.

Theorem 5.2. *Let X be a separable Banach space and let $\mathcal{G}: X \rightarrow \mathbb{R}^d$ be continuous. Let $x^\dagger \in X$ be arbitrary. Suppose that the data y_1, \dots, y_N, \dots are generated from the observation model (5.4) and μ^{y_N} is the corresponding posterior. For each $N \in \mathbb{N}$, let x_N be any strong MAP estimator of μ^{y_N} . Then:*

- (a) *there is a subsequence of $(\mathcal{G}(x_N))_{N \in \mathbb{N}}$ converging to $\mathcal{G}(x^\dagger)$ almost surely;*
- (b) *if $x^\dagger \in E$, there is a subsequence of $(x_N)_{N \in \mathbb{N}}$ converging to some $x^* \in E$ weakly in E almost surely, and $\mathcal{G}(x^*) = \mathcal{G}(x^\dagger)$.*

6. Closing remarks

MAP estimators provide a simple summary of the posterior distribution, but in the nonparametric setting it is not straightforward even to verify that MAP estimators exist. This article has shown that Bayesian inverse problems defined on any separable Banach space with a Gaussian prior have well-defined strong MAP estimators under very mild conditions on the forward problem. The fact that MAP estimators correspond with minimisers of a Tikhonov functional is an important justification for the Bayesian approach, and this article has also extended the connection between MAP estimators and variational minimisers to the Banach setting. As a corollary of theorem 1.1 on the existence of strong MAP estimators, this article also extends results on the consistency of MAP estimators to additive-noise Bayesian inverse problems set in any separable Banach space.

The strategy adopted here depends on two essential points: the statistical structure of the Bayesian inverse problem (1.2), which ensures the posterior is absolutely continuous with respect to the prior, and the topological structure provided by the Gaussian prior through the compactly embedded Cameron–Martin space.

In more general settings, such as those where the observed quantity has infinite dimension, it need not be the case that the posterior is absolutely continuous with respect to the prior (Stuart 2010, remark 3.8). While the small-ball theory for modes (section 2.1) does not depend on this absolute continuity, new techniques are needed to translate statements from prior to posterior without a density to relate the two.

Though this article has restricted attention to the case that X is a separable Banach space, the results on Gaussian measures used in this article hold more generally for Radon Gaussian measures on a locally convex space X , and this would form a natural extension of this work.

Another possible extension is to other priors with similar structure used in nonparametric Bayesian inverse problems, such as the p -exponential priors of Agapiou *et al* (2021). As pointed out by Agapiou *et al* (2021), remark 2.14, a bound analogous to proposition 3.1 is more challenging for non-Gaussian p -exponential measures (i.e. $p < 2$) because the appropriate analogue of the Cameron–Martin space is not a Hilbert space; on the other hand, p -exponential

measures are defined on subspaces of the countable product space \mathbb{R}^∞ , which provides a useful topological structure not present in an arbitrary Banach space.

It would also be interesting to know whether the hypothesis of coercivity in proposition 4.1 — which was sufficient to prove the existence of weak modes—would also suffice for proving the existence of strong modes.

Data availability statement

No new data were created or analysed in this study.

Acknowledgments

The author thanks Ilja Klebanov and Tim Sullivan for helpful feedback and comments. The author is supported by the Warwick Mathematics Institute Centre for Doctoral Training and gratefully acknowledges funding from the University of Warwick and the UK Engineering and Physical Sciences Research Council (Grant Number: EP/W524645/1). For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

ORCID iD

Hefin Lambley  <https://orcid.org/0000-0001-8337-5411>

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
- Agapiou S, Dashti M and Helin T 2021 Rates of contraction of posterior distributions based on *p*-exponential priors *Bernoulli* **27** 1616–42
- Agapiou S, Larsson S and Stuart A M 2013 Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems *Stoch. Proc. Appl.* **123** 3828–60
- Aliprantis C D and Border K C 2006 *Infinite Dimensional Analysis: A Hitchhiker's Guide* 3rd edn (Springer)
- Ayanbayev B, Klebanov I, Lie H C and Sullivan T J 2022a Γ -convergence of Onsager–Machlup functionals: I. With applications to maximum *a posteriori* estimation in Bayesian inverse problems *Inverse Problems* **38** 025005
- Ayanbayev B, Klebanov I, Lie H C and Sullivan T J 2022b Γ -convergence of Onsager–Machlup functionals: II. Infinite product measures on Banach spaces *Inverse Problems* **38** 025006
- Benning M and Burger M 2018 Modern regularization methods for inverse problems *Acta Numer.* **27** 1–111
- Bogachev V I 1998 *Gaussian Measures (Mathematical Surveys and Monographs vol 62)* (American Mathematical Society)
- Clason C, Helin T, Kretschmann R and Piiroinen P 2019 Generalized modes in Bayesian inverse problems *SIAM/ASA J. Uncertain. Quantification* **7** 652–84
- 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
- Dashti M and Stuart A M 2017 The Bayesian approach to inverse problems *Handbook of Uncertainty Quantification* vol 1, 2, 3 (Springer) ch 7, pp 311–428
- Dunlop M M 2019 Multiplicative noise in Bayesian inverse problems: well-posedness and consistency of MAP estimators arXiv:1910.14632
- 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

- Fernique X 1970 Intégrabilité des vecteurs gaussiens *C. R. Acad. Sci., Paris A–B* **270** A1698–9
- Ghosal S and van der Vaart A 2017 *Fundamentals of Nonparametric Bayesian Inference* 1st edn (Cambridge University Press)
- Helin T and Burger M 2015 Maximum *a posteriori* probability estimates in infinite-dimensional Bayesian inverse problems *Inverse Problems* **31** 085009
- Klebanov I and Wacker P 2023 Maximum *a posteriori* estimators in ℓ^p are well-defined for diagonal Gaussian priors *Inverse Problems* **39** 065009
- Knapik B T, van der Vaart A W and van Zanten J H 2011 Bayesian inverse problems with Gaussian priors *Ann. Stat.* **39** 2626–57
- Kretschmann R 2019 Nonparametric Bayesian inverse problems with Laplacian noise *PhD Thesis* Universität Duisburg-Essen
- Kretschmann R 2023 Are minimizers of the Onsager–Machlup functional strong posterior modes? *SIAM/ASA J. Uncertain. Quantification* **11** 1105–38
- Lambley H and Sullivan T J 2023 An order-theoretic perspective on modes and maximum *a posteriori* estimation in Bayesian inverse problems *SIAM/ASA J. Uncertain. Quantification* **11** 1195–224
- Lie H C and Sullivan T J 2018 Equivalence of weak and strong modes of measures on topological vector spaces *Inverse Problems* **34** 115013
- Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica D* **60** 259–68
- Stuart A M 2010 Inverse problems: a Bayesian perspective *Acta Numer.* **19** 451–559
- Vollmer S J 2013 Posterior consistency for Bayesian inverse problems through stability and regression results *Inverse Problems* **29** 125011