Sparsity-promoting and edge-preserving maximum a posteriori estimators in non-parametric Bayesian inverse problems — **Source link** ⬈

Sergios Agapiou, Martin Burger, Masoumeh Dashti, Tapio Helin

Institutions: University of Cyprus, University of Münster, University of Sussex, University of Helsinki

Related papers:

- An equivalence between high dimensional Bayes optimal inference and M-estimation

- Revisiting maximum-a-posteriori estimation in log-concave models

- Revisiting maximum-a-posteriori estimation in log-concave models: from differential geometry to decision theory

- On bayesian estimation and proximity operators

- Partially Linear Bayesian Estimation with Application to Sparse Approximations

Sparsity-promoting and edge-preserving maximum a posteriori estimators in non-parametric Bayesian inverse problems

SPARSITY-PROMOTING AND EDGE-PRESERVING MAXIMUM A POSTERIORI ESTIMATORS IN NON-PARAMETRIC BAYESIAN INVERSE PROBLEMS

By Sergios Agapiou* Martin Burger† Masoumeh Dashti‡ and Tapio Helin§

University of Cyprus *, University of Münster†, University of Sussex ‡ and University of Helsinki§

We consider the inverse problem of recovering an unknown functional parameter u in a separable Banach space, from a noisy observation vector y of its image through a known possibly non-linear map $\mathcal{G}$. We adopt a Bayesian approach to the problem and consider Besov space priors (see [47]), which are well-known for their edge-preserving and sparsity-promoting properties and have recently attracted wide attention especially in the medical imaging community.

Our key result is to show that in this non-parametric setup the maximum a posteriori (MAP) estimates are characterized by the minimizers of a generalized Onsager–Machlup functional of the posterior. This is done independently for the so-called weak and strong MAP estimates, which as we show coincide in our context. In addition, we prove a form of weak consistency for the MAP estimators in the infinitely informative data limit. Our results are remarkable for two reasons: first, the prior distribution is non-Gaussian and does not meet the smoothness conditions required in previous research on non-parametric MAP estimates. Second, the result analytically justifies existing uses of the MAP estimate in finite but high dimensional discretizations of Bayesian inverse problems with the considered Besov priors.

1. Introduction. We consider the inverse problem of recovering an unknown functional parameter $u \in X$ from a noisy and indirect observation $y \in Y$. We work in a framework in which X is an infinite-dimensional separable Banach space, while $Y = \mathbb{R}^J$. In particular, we consider the additive noise model

$$(1) \qquad\qquad y = \mathcal{G}(u) + \xi,$$

where ξ is mean zero Gaussian observational noise, with a positive definite covariance matrix $\Sigma \in \mathbb{R}^{J \times J}$, $\xi \sim N(0, \Sigma)$. Here the possibly nonlinear operator $\mathcal{G} : X \to Y$ describes the system response, connecting the observation y to the unknown parameter u. More specifically, $\mathcal{G}$ captures both the forward model and the observation mechanism and is assumed to be known.

Inverse problems are mathematically characterized by being ill-posed: the lack of sufficient information in the observation prohibits the unique or stable reconstruction of the unknown. This can either be due to an inherent loss of information in the forward model, or due to the incomplete and noisy observation. To address ill-posedness, regularization techniques are employed in which the available information in the observation is augmented using *a priori* available knowledge on the properties of candidate solutions.

We adopt a Bayesian approach to the regularization of inverse problems, which has in recent years attracted enormous attention in the inverse problems and imaging literature: see the early work [29], the books [69, 37] and the more recent works [66, 16, 46] and references therein. In this approach, prior information is encoded in the *prior* distribution μ_0 on the unknown u, and the Bayesian methodology is used to formally obtain the *posterior* distribution μ^y on $u|y$, in the form

$$(2) \qquad\qquad \mu^y(du|y) \propto \exp\big(-\Phi(u;y)\big)\mu_0(du),$$

where

$$\Phi(u;y) = \frac{1}{2}\left\|\Sigma^{-\frac{1}{2}}\big(y - \mathcal{G}(u)\big)\right\|_Y^2.$$

Our work is driven by two types of a priori information: on the one hand, we aim to recover unknown functions with a blocky structure, as is typically the case in image processing [23, 55] and medical imaging applications [22]. On the other hand, we are interested in prior models which promote sparse solutions, that is high-dimensional solutions that can be represented by a small number of coefficients in an appropriate expansion. To achieve these effects, we utilize so-called Besov space priors [47, 17], which are well-known for their edge-preserving and sparsity-promoting properties, see e.g. [44, 58, 50, 35, 9]. In the core of these priors is the fact that they employ a wavelet basis and use ℓ^1-type regularization on the corresponding coefficients, an idea rooted in the extensive statistical literature developed by Donoho, Johnstone, Elad, Candes and others, see for example [36, 14, 25, 24, 20, 13, 52, 64], which is also the central idea in the field of compressive sensing. With these goals in mind, our work relates to classical L^1-type regularization methods such as penalized least squares with Total Variation penalty [60, 65]. Other Bayesian approaches which promote sparse and blocky solutions include the hierarchical methods extensively studied by Calvetti and Somersalo, see for example [10, 11, 4, 12].

In practice, the implementation of the Bayesian approach to an inverse problem typically requires a high-dimensional discretization of the unknown and large computational resources. From a computational perspective, it is hence imperative that the probability models and estimators related to finite-dimensional discretizations of the problem scale well with respect to refining discretization. In particular, there is a fundamental need to understand whether specialized prior information used frequently in applications, for example of the type described above, leads to well-defined non-parametric probability models and whether the related finite-dimensional estimators have well-behaving limits; this motivates the study of Bayesian inverse problems in the infinite-dimensional function-space setting. A body of literature on this topic has emerged during the last years, centered around two theories:

a) the *discretization invariance* theory, see [48, 47, 46] and references therein, which aims to ensure that when investing more resources in increasing the discretization level, the prior remains faithful to the intended information on the unknown function and the posterior converges to a well defined limit which is an improved representation of the reality;

b) the *well-posedness* of the posterior theory found in [66, 16], which secures that the posterior is well defined in the infinite-dimensional limit and robust with respect to perturbations in the data as well as approximations of the forward model.

Our work studies Bayesian inversion in function spaces especially from the perspective of point estimators. Namely, we study and give a rigorous meaning to *maximum a posteriori* (MAP) estimates for Bayesian inverse problems with certain Besov priors in the infinite-dimensional function-space

setting. A MAP estimate is understood here as the mode of the posterior probability measure, hence our results require a careful definition of a mode in the infinite-dimensional setting. The main challenge is then to establish a connection between the topological definition of a MAP estimate (mode of the posterior) and an explicit variational problem. Succeeding in doing so, opens up the possibility of studying the behaviour of MAP estimators in certain situations. In particular, we are able to prove a weak form of consistency of the MAP estimator in the infinitely informative data limit.

1.1. *Non-gaussian prior information and the need for MAP estimators.* A major challenge in Bayesian statistics is the extraction of information from the posterior distribution. In Gaussian-conjugate settings, such as linear inverse problems, on the one hand there are explicit formulae for the posterior mean which can be used as an estimator of the unknown, and on the other hand one can (in principle) sample directly from a discretized version of the posterior, [53, 49, 1, 56, 54]. Nevertheless, draws from Gaussian priors do not vary sufficiently sharply on their domain to have a blocky structure, neither do they give rise to sparse estimators.

For this purpose, the so-called TV prior has been used widely in applied literature, e.g. in medical imaging [63, 45]. Drawing intuition from the classical regularization literature, the TV prior has a formal density of the form

$$\pi(u) \propto \exp(-\alpha \|u\|_{BV}),$$

where $\alpha > 0$ is a (hyper)parameter. Here, the norm of bounded variation, $\|u\|_{BV}$, can be formally thought of as the L^1-norm of the derivative of u. However, the numerical implementation of Bayesian inversion with a TV prior miss-behaves as the discretization level of the unknown increases, and in particular the TV prior is not discretization invariant, [48, 47, 44]. For example, depending on the choice of the parameter α as a function of the discretization level, the posterior mean either diverges or converges to an estimate corresponding to a Brownian bridge prior. For alternative types of approaches to edge-preserving non-parametric Bayesian inversion, see [26, 33, 31].

We consider the family of Besov-priors which were proposed and shown to be discretization invariant in [47]. A well-posedness theory of the posterior was developed in [17]. These priors are defined by a wavelet expansion with random coefficients, motivated by a formal density of the form

$$(3) \qquad\qquad \pi(u) \propto \exp(-\|u\|_{B_p^s}^p),$$

where $\|\cdot\|_{B_p^s} = \|\cdot\|_{B_{pp}^s}$ is the Besov space norm with regularity parameter s and integrability parameters p. For $p = 2$ the Besov space B_2^s corresponds to the Sobolev space of functions with s square-integrable derivatives, H^s, and the corresponding family of priors are Gaussian with Sobolev-type smoothness parametrized by s. We are especially interested in the case $p = 1$, which is highly interesting for edge-preserving and sparsity-promoting Bayesian inversion, [44]. For the case $s = 1$ this is due to the close resemblance between the way $\|u\|_{B_1^1}$ and $\|u\|_{BV}$ work, both dealing with the L^1-norm of a (generalized) derivative of u, see [44, Section 2]. We define rigorously the class of Besov priors for $p = 1$ and $s \in \mathbb{R}$, called B_1^s-Besov priors, in section 3 below.

To probe the posterior in the non-conjugate context of linear or nonlinear inverse problems with Besov priors, one typically resorts to Markov chain Monte Carlo (MCMC) methods. Unfortunately, in practice standard MCMC algorithms become prohibitively expensive for large scale inverse problems, consider e.g. photo-acoustic tomography [57, 70]. This is due to the heavy computational effort

required for solving the forward problem which is needed for computing the acceptance probability at each step of the chain.

In such situations maximum a posteriori estimates are computationally attractive as they only require solving a single optimization problem. Furthermore, it was shown in [44] that for the B_1^1-Besov prior defined via wavelet bases, at finite discretization levels the resulting MAP estimators are sparse and preserve the locations of the edges in the unknown. This remains true as discretization is refined, while our work validates that, unlike in the TV prior case, the discretized MAP estimators converge to the MAP estimators of the limiting infinite dimensional posterior distribution.

1.2. *Review of the main results.* Our results draw inspiration from previous papers by the authors [18, 32] (see also [26]), where concepts that we will call *strong* and *weak MAP estimates* were coined. We quote both definitions in section 2 below.

As mentioned earlier, we can formally think of a MAP estimator as a mode of the posterior. In finite dimensional contexts, especially when working with continuous probability distributions, the definition of a mode is straightforward as a maximizer of the probability density function. If the prior has probability density function of the form

$$\pi(u) \propto \exp(-W(u)),$$

for a suitable positive function $W : X \to [0, \infty)$, the density of the posterior is

$$\pi^y(u) \propto \exp(-I(u; y)),$$

where

$$(4) \qquad\qquad I(u; y) = \Phi(u; y) + W(u).$$

In this case, a MAP estimator can be interpreted as a classical estimator of the unknown, arising from Tikhonov regularization with penalty term given by the negative log-density of the prior, [28].

In the infinite dimensional setting things are less straightforward due to the lack of a uniform reference measure. An intuitive approach to define a mode in a function space X is as follows: compute the measure of balls with any center $u \in X$ and a fixed radius $\epsilon > 0$ and proceed by letting ϵ tend to zero. A mode $\hat{u}$ is a center point maximizing these small ball probabilities asymptotically (as ϵ decreases) in a specific sense. What distinguishes a strong mode from a weak mode is exactly how 'asymptotic maximality' is perceived:

a) in the strong mode case, we look for the maximum probability among all centres in X;
b) in the weak mode case, we look for a centre of a ball with the property that it has maximum probability among all shifts of the ball by elements of a dense subspace $E \subset X$.

Natural choices of E turn out to be spaces of zero probability, thus giving one interpretation for the term 'weak'. Note that weak and strong modes coincide for $E = X$. Each of the two notions of mode gives rise to a notion of MAP estimator termed strong and weak MAP estimators, respectively.

Let $B_\epsilon(z) \subset X$ denote the open ball of radius ϵ, centered at $z \in X$. If we can find a functional I defined on an appropriate dense subspace $F \subset X$, such that

$$(5) \qquad\qquad \lim_{\epsilon \to 0} \frac{\mu(B_\epsilon(z_2))}{\mu(B_\epsilon(z_1))} = \exp(I(z_1) - I(z_2)),$$

when $\mu = \mu^y$, then for any fixed $z_1 \in F$, a $z_2 \in F$ maximising the limit or equivalently minimising I is a potential MAP estimator. For weak MAP estimators, if the space $E \subset F$ over which we shift the ball centres, can be chosen sufficiently regular so that

$$(6) \qquad \lim_{\epsilon \to 0} \frac{\mu(B_\epsilon(u - h))}{\mu(B_\epsilon(u))} \quad \text{is continuous in } u \in X \text{ for any fixed } h \in E,$$

again for $\mu = \mu^y$, then it is straightforward to establish the equivalence of weak MAP estimators and the minimisers of I. In the case of strong MAP estimators one needs to work considerably more, the difficulty stemming from the 'smallness' of the subspace $F \subset X$ with respect to the prior hence also the posterior.

If a functional I satisfying (5) exists, it is called the (generalised) Onsager-Machlup functional, [27, 34, 18, 32]. As for the limit in (6), it is identified with R_h^μ, the Radon-Nikodym derivative of the shift of μ by $h \in E$ with respect to μ itself, provided this derivative has a continuous representative, see [32, lemma 2] (quoted in lemma 2.3 below). In finite dimensions, both the existence of the Onsager-Machlup functional and the statement in (6) follow from the Lebesgue differentiation theorem under very mild conditions on Φ and the density of the prior. In the Banach-space setting, establishing (5) and (6) for the posterior boils down to showing similar results for the prior given that the posterior is obtained by a suitably regular transformation of the prior.

Strong MAP estimators were defined and studied in [18], in the context of nonlinear Bayesian inverse problems with Gaussian priors. In this case, an expression for the limit in (5) was readily available from the Gaussian literature for F being the Cameron–Martin space of μ_0. The identification of minimisers of the resulting Onsager–Machlup functional with strong MAP estimators, however, required a considerable amount of work, in particular many new estimates involving small ball probabilities under the prior.

Weak MAP estimators were defined and studied in [32], in the context of linear Bayesian inverse problems with a general class of priors. The authors used the tools from the differentiation and quasi-invariance theory of measures, developed by Fomin and Skorohod [6], to connect the zero points of β_h^μ, the logarithmic derivative of a measure μ in the direction h, to the minimisers of the Onsager-Machlup functional. An essential assumption that makes this possible is the continuity of β_h^μ over X for sufficiently regular h. The authors considered as examples Besov priors with integrability parameter $p > 1$ and a conditionally Gaussian hierarchical prior with a hyper-prior on the mean.

In both [18] and [32], the Onsager-Machlup functional was shown to be a Tikhonov-type functional as in (4), where W is the formal negative log-density of the prior, hence MAP estimators are identified with the corresponding Tikhonov approximations in the studied contexts.

In this work, we are interested in defining and studying both the strong and the weak MAP estimates for generally nonlinear inverse problems with B_1^s-Besov priors, that is for Besov priors with integrability parameter $p = 1$. Since the prior has formal density as in (3), we expect and indeed show that MAP estimates are identified with minimizers of the Tikhonov-type functional

$$(7) \qquad\qquad I(u; y) = \Phi(u; y) + \|u\|_{B_1^s}.$$

In particular, the weak and strong MAP estimates coincide. For the considered prior, the general theory developed in [32] for weak MAP estimators does not apply, due to the fact that the logarithmic derivative of the B_1^s-Besov priors is inherently discontinuous. We will show that the continuity

of the Radon-Nikodym derivative R_h^μ, for h in a suitable subspace $E \subset X$, is sufficient to get the result for weak MAP estimators. For the B_1^s-Besov priors, we prove the continuity of R_h^μ for shifts $h \in E \subseteq B_1^r(\mathbb{T}^d)$ for any $r > s$, establish appropriate small ball probability ratio asymptotics, and show the validity of (5) over $F = B_1^s(\mathbb{T}^d)$. It is then straightforward to prove the identification of weak MAP estimators by the minimisers of the functional I in (7). Furthermore, we generalise the program of [18] to the assumed non-Gaussian case, developing some properties of maximizers of small ball probabilities along the way, and show that strong MAP estimators are identified by the minimisers of I in (7). We also consider the theory of local weak modes separately, relying on the Fomin literature.

One natural question arising from our work is under what conditions strong and weak MAP estimators coincide. This question is investigated in the very recent work [51], where the authors give general conditions for the equivalence of the two notions. The proofs of our results as presented below, aptly illustrate that, in practice, it is preferred to work with weak MAP estimators.

1.3. *Consistency of MAP estimators.* Under the frequentist assumption of data generated from a fixed underlying true $u^\dagger$, it is desirable to verify that in the infinitely informative data limit, the Bayesian posterior distribution μ^y contracts optimally to a Dirac distribution centered on $u^\dagger$. In recent years, there have been many studies on the rates of posterior contraction in the context of Bayesian inverse problems. The case of linear inverse problems with Gaussian and conditionally Gaussian priors is now well understood [42, 1, 43, 2, 39, 68, 41], and a theory for linear inverse problems with non-Gaussian priors is also being developed [59, 40]. For nonlinear inverse problems the asymptotic performance of the posterior is not yet fully understood, with some partial contributions being [56, 71, 72].

Note that for general nonlinear problems, especially with finite dimensional data as in the present paper, one cannot expect to recover the underlying truth $u^\dagger$ in the infinitely informative data limit. Instead the aim is to recover a $u^* \in X$ such that $\mathcal{G}(u^*) = \mathcal{G}(u^\dagger)$. Posterior contraction rates for Bayesian inverse problems with Besov priors are studied in ongoing work of a subset of the authors. A form of weak consistency of the strong MAP estimator in the presence of repeated independent observations, for general nonlinear inverse problems with Gaussian priors, was shown in [18]. In the present paper, we prove a similar result for the strong (hence the weak) MAP estimator obtained using the Besov prior for $p = 1$.

1.4. *Notation.* Throughout the paper we assume that X is a separable Banach space equipped with the Borel σ-algebra. All probability measures are assumed to be Borel measures. The Euclidean norm in $\mathbb{R}^J$ is denoted by $|\cdot|$ to distuinguish it from the norm of any general X denoted by $\|\cdot\|_X$. We write $f \propto g$ for two functions $f, g : X \to \mathbb{R}$ if there exists a universal constant $c \in \mathbb{R}$ such that $f = cg$ as functions.

DEFINITION 1.1. *A measure μ is called* quasi-invariant *along h, if the translated measure $\mu_h(\cdot) := \mu(\cdot - h)$ is absolutely continuous with respect to μ. We define*

$$Q(\mu) = \{h \,|\, \mu \text{ is quasi-invariant along } h\},$$

which is readily verified to be a linear subspace.

NOTATION 1.2. *Let $h \in Q(\mu)$. We denote the Radon–Nikodym derivative of μ_h with respect to μ by $R_h^\mu \in L^1(\mu)$.*

1.5. *Organization of the paper.* This paper is organized as follows: in section 2 we discuss the definition of modes for probability measures on separable Banach spaces. We introduce novel localized versions of the modes studied in previous work and discuss briefly how different modes can be characterized for log- or quasi-concave measures. The Besov priors are introduced and discussed in section 3 including the key results relating to the Radon–Nikodym derivative of the Besov prior in section 3.2. Section 4 covers the Bayesian inverse problem setup and our main results related to identification of weak and strong MAP estimates as the minimizers of certain variational problem. Moreover, the weak consistency result is given in section 4.2. In section 5 we discuss the logarithmic derivative of the posterior and its use in characterizing the MAP estimates. Finally, all proofs are postponed to section 6.

2. Modes of measures on Banach spaces. In subsection 2.1 we introduce the two existing notions of maximum a posteriori estimator (modes of the posterior measure) proposed in [18, 32] in the context of measures on infinite-dimensional spaces. We also define two new notions of local modes and hence local MAP estimates. In subsection 2.2, we focus on log-concave measures, study the structure of the set of modes, and give conditions for local modes to be global.

2.1. *Weak and strong, global and local.* The following definition of a mode, introduced in [18], grows out of the idea that highest small ball probabilities are obtained asymptotically at the mode.

DEFINITION 2.1. *Let $M^\epsilon = \sup_{u \in X} \mu(B_\epsilon(u))$. We call a point $\hat{u} \in X$ a* mode *of the measure μ, if it satisfies*

$$\lim_{\epsilon \to 0} \frac{\mu(B_\epsilon(\hat{u}))}{M^\epsilon} = 1.$$

A mode of the posterior measure μ^y in (2), is called a maximum a posteriori *(MAP) estimate.*

Below we occasionally use the terms *strong mode* and *strong MAP* estimator for the concepts introduced in definition 2.1 in order to distinguish them from the following weaker notion of a mode (similarly, the *weak mode* or *weak MAP*) introduced in [32].

DEFINITION 2.2. *Let E be a dense subspace of X. We call a point $\hat{u} \in X$, $\hat{u} \in \mathrm{supp}(\mu)$, a* weak mode *of μ if*

$$(8) \qquad \lim_{\epsilon \to 0} \frac{\mu(B_\epsilon(\hat{u} - h))}{\mu(B_\epsilon(\hat{u}))} \leq 1,$$

for all $h \in E$. A weak mode of the posterior measure μ^y in (2), is called a weak maximum a posteriori *(wMAP) estimate.*

Notice that the definition of a weak mode is dependent on the choice of the subspace E. Therefore, in some contexts it may be more appropriate to discuss E-weak modes. In the following, however, we will suppress this dependence since the key question to our study is whether such a space exists.

The notions of weak and strong mode are related as follows: any strong mode is a weak mode for the choice $E = X$ [32, lemma 3], which is straightforward to see by simply estimating $\mu(B_\epsilon(\hat{u} - h)) \leq M^\epsilon$. The key motivation to study the weak definition, is the case when small ball asymptotics are not available explicitly or only available in some subspace of translations h. It is then of interest

to choose E so that an expression for the limit on the left hand side of (8) exists pointwise. This typically leads to choices of E which have zero probability with respect to μ. The following lemma, which is an immediate generalization of [32, lemma 2], provides further guidance for this choice.

LEMMA 2.3.   *Assume that μ is quasi-invariant along the vector h. Let $A \in \mathcal{B}(X)$ be convex, bounded and symmetric and define $A^\epsilon := \epsilon A$. Suppose R_h^μ has a continuous representative $\tilde{R}_h^\mu \in C(X)$, i.e., $R_h^\mu - \tilde{R}_h^\mu = 0$ in $L^1(\mu)$. Then it holds that*

$$\lim_{\epsilon \to 0} \frac{\mu_h(A^\epsilon + u)}{\mu(A^\epsilon + u)} = \tilde{R}_h^\mu(u)$$

for any $u \in X$.

REMARK 2.4.   *According to lemma 2.3, it is desirable to consider a subspace E, such that R_h^μ is continuous for $h \in E$. A sufficient condition for the continuity of R_h^μ was given in [32], namely the so-called logarithmic derivative of μ along h needed to be continuous and exponentially integrable with respect to μ. For the sparsity promoting measure that we will consider in this paper, the logarithmic derivative is inherently discontinuous (section 5). We are however able to show the continuity of R_h^μ over an appropriate subspace E in subsection 3.2 by using its explicit expression.*

Both strong and weak mode can also be associated with a natural localization described by the following definitions:

DEFINITION 2.5 (Local modes).   *Let $\hat{u} \in X$ be such that $\hat{u} \in \mathrm{supp}(\mu)$.*

(1) We call $\hat{u}$ a local mode *of the measure μ, if there exists a $\delta > 0$ such that the quantity $M_\delta^\epsilon = \sup_{u \in B_\delta(\hat{u})} \mu(B_\epsilon(u))$ satisfies*

$$\lim_{\epsilon \to 0} \frac{\mu(B_\epsilon(\hat{u}))}{M_\delta^\epsilon} = 1.$$

(2) We call $\hat{u}$ a local weak mode *of μ if there exists $\delta > 0$ such that*

$$(9) \qquad \lim_{\epsilon \to 0} \frac{\mu(B_\epsilon(\hat{u} - h))}{\mu(B_\epsilon(\hat{u}))} \leq 1$$

for all $h \in B_\delta^X(0) \cap E$.

Local modes represent an analogue to local maxima of the probability density function in finite dimensions. In the setting of [32] the local wMAP coincides with the zero points of the logarithmic derivative of the posterior. Especially, regularization techniques in non-linear inverse problems are often known to give birth to local maxima and, therefore, the local wMAP can give statistical interpretation for these points.

2.2. *Modes for log-concave measures.*   This work studies the Besov prior which is a prototypical example of a log-concave measure. Below we show some general properties regarding the modes of such class of measures. Most of these ideas naturally extend to a larger class called the quasi-concave measures.

DEFINITION 2.6. *A probability measure μ in $(X, \mathcal{B}(X))$ is called* logarithmically-concave *or* log-concave, *if*

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

for all $A, B \in \mathcal{B}(X)$. Moreover, μ is called quasi-concave *if*

$$\mu(\lambda A + (1 - \lambda)B) \geq \min\{\mu(A), \mu(B)\},$$

for all $A, B \in \mathcal{B}(X)$.

It is straightforward to verify that any log-concave measure is also quasi-concave. An immediate result of quasi-concavity is the well-known Anderson inequality [3, 8].

PROPOSITION 2.7. *Let μ be a symmetric quasi-concave measure on X. For any symmetric and convex set $A \subset X$ we have*

$$\mu(A + x) \leq \mu(A), \quad \text{for any } x \in X.$$

The next result follows from the Anderson inequality.

PROPOSITION 2.8. *Suppose that the measure μ on X is symmetric around u and quasi-concave. Then u is a strong mode of μ.*

Let us next consider briefly the structure of the set of strong modes of a quasi-concave measure μ. When working in finite dimensions, $X = \mathbb{R}^d$, a probability density function f with respect to the Lebesgue measure is called quasi-concave if for all $x, y \in \mathbb{R}^d$ and all $\lambda \in [0, 1]$ we have

$$f((1 - \lambda)x + \lambda y) \geq \min\{f(x), f(y)\}.$$

Clearly a quasi-concave probability density function has a convex set of global modes. For a reference on convexity and unimodality in finite dimensions see [21]. We show that a similar result holds in infinite dimensions for our definition of strong mode. The weak mode case is covered in [32].

PROPOSITION 2.9. *Suppose that the measure μ in X is quasi-concave (but not necessarily symmetric). Then the set of strong modes is convex.*

It turns out log-concavity is a sufficient condition for the global and local modes to coincide.

THEOREM 2.10. *Suppose μ is log-concave and $\hat{u}$ is a local mode. Then $\hat{u}$ is also a global mode. Similarly, if $\hat{u}$ is a local weak mode then it is also a global weak mode.*

3. Besov priors with $p = 1$. The family of Besov priors has been introduced in [47] and studied in [17, 32]. In subsection 3.1 we recall the definition and some useful properties of Besov priors with integrability parameter $p = 1$ and regularity parameter $s > 0$, termed B_1^s-Besov priors, on which we focus in this work. We also present some straightforward convexity properties of B_1^s-Besov priors. The main results of this section are listed in subsection 3.2, where we compute the Radon-Nikodym derivative R_h^μ for B_1^s-Besov priors, determine the space E in which h needs to live in order for the corresponding Radon-Nikodym derivative R_h^μ to be continuous, and finally show that $I_0(u) = \|u\|_{B_1^s}$ is the Onsager-Machlup functional for B_1^s-Besov measures.

3.1. *Definition and basic properties.* We work with periodic functions on a d-dimensional torus, $\mathbb{T}^d$. We first define the periodic Besov spaces $B_{pq}^s(\mathbb{T}^d)$, where $s \in \mathbb{R}$ parametrises smoothness and $p, q \geq 1$ are integrability parameters. We concentrate on the case $p = q$ and write $B_p^s = B_{pp}^s$. To define the Besov spaces, we let $\{\psi_\ell\}_{\ell=1}^\infty$ be an orthonormal wavelet basis for $L^2(\mathbb{T}^d)$, where we have utilized a global indexing. We can then characterise $B_p^s(\mathbb{T}^d)$ using the given basis in the following way: the function $f : \mathbb{T}^d \to \mathbb{R}$ defined by the series expansion

$$(10) \qquad f(x) = \sum_{\ell=1}^\infty c_\ell \psi_\ell(x)$$

belongs to $B_p^s(\mathbb{T}^d)$, if and only if the norm

$$(11) \qquad \|f\|_{B_p^s(\mathbb{T}^d)} = \left(\sum_{\ell=1}^\infty \ell^{p(\frac{s}{d}+\frac{1}{2})-1} |c_\ell|^p \right)^{\frac{1}{p}},$$

is finite. Throughout, we assume that the basis is r-regular for r large enough in order to consist a basis for a Besov space with smoothness s, [19].

We now follow the construction in [47] to define periodic Besov priors corresponding to $p = 1$, using series expansions in the above wavelet basis with random coefficients. Notice also the work [30] on defining Besov priors for functions on the full space $\mathbb{R}^d$.

DEFINITION 3.1. *Let $(X_\ell)_{\ell=1}^\infty$ be independent identically distributed real-valued random variables with the probability density function*

$$(12) \qquad \pi_X(x) = \frac{1}{2} \exp(-|x|).$$

Let U be the random function

$$U(x) = \sum_{\ell=1}^\infty \ell^{-(\frac{s}{d}-\frac{1}{2})} X_\ell \psi_\ell(x), \quad x \in \mathbb{T}^d.$$

Then we say that U is distributed according to a B_1^s-Besov prior.

The next lemma determines the smoothness of functions drawn from the B_1^s-Besov prior and shows the existence of certain exponential moments.

LEMMA 3.2. *[47, lemma 2] Let U be as in Definition 3.1 and let $t < s - d$. Then it holds that*

(i) $\|U\|_{B_1^t} < \infty$, almost surely, and
(ii) $\mathbb{E} \exp(\frac{1}{2} \|U\|_{B_1^t}) < \infty$.

NOTATION 3.3. *We denote by ρ_ℓ the probability measure of the random variable $\ell^{-s/d-1/2} X_\ell$ on $\mathbb{R}$. We identify the random function U in definition 3.1 with the product measure of coefficients $(\ell^{-s/d-1/2} X_\ell)_\ell$ in $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$, which we denote by $\lambda = \bigotimes_{\ell=1}^\infty \rho_\ell$.*

We next consider the convexity of the B_1^s-Besov prior.

LEMMA 3.4.   *For any $s > 0$, the B_1^s-Besov measure λ is logarithmically concave.*

An immediate consequence of the last lemma, is that by proposition 2.7, the B_1^s-Besov prior satisfies Anderson's inequality. Notice that lemma 3.4 (and hence proposition 2.7) also holds for B_p^s-Besov measures with $p > 1$, as defined in [47].

3.2. *Radon–Nikodym derivative R_h^μ and small ball probabilities.*   Recall the definitions of quasi-invariance for a measure μ and of the subspace of directions in which μ is quasi-invariant, $Q(\mu)$. In general the structure of Q is not known and there are even examples of measures for which Q fails to be locally convex [6, exercise 5.5.2]. The space Q is known to be a Hilbert space for certain families of measures, for example for α-stable measures with $\alpha \geq 1$ and for countable products of a single distribution with finite Fisher information, see [6, theorem 5.2.1] and [62] (note that the B_1^s-Besov prior, λ, is not an α-stable measure). Using similar techniques to [62], namely the Kakutani–Hellinger theory, we now show that Q is a Hilbert space also for the B_1^s-Besov measure λ and calculate the Radon-Nikodym derivative R_h^λ between λ and the shifted measure λ_h for $h \in Q(\lambda)$.

LEMMA 3.5.   *For the B_1^s-Besov measure we have $Q(\lambda) = B_2^{s-\frac{d}{2}}(\mathbb{T}^d)$. For $h \in Q(\lambda)$ we have*

$$\frac{d\lambda_h}{d\lambda}(u) = \lim_{N \to \infty} \exp \sum_{\ell=1}^N \big(-\alpha_\ell |h_\ell - u_\ell| + \alpha_\ell |u_\ell| \big)$$

in $L^1(\mathbb{R}^\infty, \lambda)$, where $\alpha_\ell = \ell^{s/d-1/2}$.

We next provide a more detailed view of spaces of shifts h for which R_h^λ has a continuous representative, which we denote by $\tilde{R}_h^\lambda(u)$. This is a crucial result for our study of weak MAP estimators, see lemma 2.3, Remark 2.4 and the discussion in subsection 1.2.

LEMMA 3.6.   *Let $h = \sum_{\ell \in \mathbb{N}} h_\ell \psi_\ell \in B_1^r(\mathbb{T}^d)$ with $r > s$. Then $\tilde{R}_h^\lambda(u) = \exp \sum_{\ell=1}^\infty \big(-\alpha_\ell |h_\ell - u_\ell| + \alpha_\ell |u_\ell| \big)$ is continuous with respect to $u = \sum_{\ell \in \mathbb{N}} u_\ell \psi_\ell \in B_1^t(\mathbb{T}^d)$ for any $t < s - d$.*

Note that in the expression for the Radon-Nikodym derivative $\tilde{R}_h^\lambda(u)$, $h \in E$ and the less regular $u \in X$ are coupled component-wise (in the wavelet basis defining the Besov measure) and hence establishing the continuity of $\tilde{R}_h^\lambda(u)$ with respect to u is not straightforward. See section 6 for the proof of above lemma.

We record the following immediate corollary of the last lemma and lemma 2.3.

COROLLARY 3.7.   *Let $h \in B_1^r(\mathbb{T}^d)$, $r > s$. Then it holds that*

$$(13) \qquad \lim_{\epsilon \to 0} \frac{\lambda_h(B_\epsilon(u))}{\lambda(B_\epsilon(u))} = \exp \sum_{\ell=1}^\infty (-\alpha_\ell |h_\ell - u_\ell| + \alpha_\ell |u_\ell|),$$

for any $u \in B_1^t$.

REMARK 3.8.   *Let $X = B_1^t(\mathbb{T}^d)$ for $t < s - d$. By remark 2.4 and noting that the Besov space $B_1^r(\mathbb{T}^d)$ is dense in $B_1^t(\mathbb{T}^d)$ for any $r > t$, the last corollary shows that it is natural to choose $E = B_1^r(\mathbb{T}^d)$ for any $r > s$ in the definition of wMAP estimate. It also shows that the origin is the unique weak mode and, therefore, by proposition 2.8 also the unique strong mode.*

We also remark that each term of the series in the right-hand side of (13) consists of the difference of two terms, whose respective series are not convergent in general for $u \in B_1^t(\mathbb{T}^d)$ and $h \in B_1^r(\mathbb{T}^d)$. However, they coincide with the difference of the norms in $B_1^s(\mathbb{T}^d)$ if h and u are both elements of $B_1^s(\mathbb{T}^d)$. Based on this view we close this section with an important building block for the study of the MAP estimate. It extends the last corollary to $h \in B_1^s(\mathbb{T}^d)$, for balls centered at the origin.

THEOREM 3.9.   *Suppose that $h \in B_1^s(\mathbb{T}^d)$ and $t < s - d$. Let $A \in \mathcal{B}(B_1^t(\mathbb{T}^d))$ denote a convex and zero-centered symmetric and bounded set. Then*

$$\lim_{\epsilon \to 0} \frac{\lambda_h(\epsilon A)}{\lambda(\epsilon A)} = \exp(-\|h\|_{B_1^s}).$$

It follows immediately from the above theorem that for $z_1, z_2 \in B_1^s(\mathbb{T}^d)$,

$$(14) \qquad\qquad \lim_{\epsilon \to 0} \frac{\lambda(B_\epsilon(z_1))}{\lambda(B_\epsilon(z_2))} = \exp(-\|z_1\|_{B_1^s} + \|z_2\|_{B_1^s}),$$

giving the Onsager–Machlup functional of λ. The space $B_1^s(\mathbb{T}^d)$ here, is the largest space on which the Onsager–Machlup functional is defined. This is the space F of the discussion in subsection 1.2 in the case of Besov priors.

It is also worth noting the difference between (13) and (14). The centres of the balls in (13) are in $X = B_1^t(\mathbb{T}^d)$, the continuity of the right-hand side in u is due to sufficient regularity of the shift h. In (14), the centres of the balls are more regular than (13), but the shift is less regular in general, i.e. only in $B_1^s(\mathbb{T}^d)$.

4. Characterization of MAP estimates and their weak consistency.   In this section, building on what we have shown on properties of B_1^s-Besov priors, we first show the existence of weak and strong MAP estimates in Bayesian inverse problems with such priors, and identify both with the minimisers of the Onsager–Machlup functional. Using this characterization, we then prove a weak consistency result for MAP estimators.

4.1. *Identification of MAP estimates for Bayesian inverse problems.*   We consider the inverse problem of estimating a function $u \in X = B_1^t(\mathbb{T}^d)$ from a noisy and indirect observation $y \in \mathbb{R}^J$, modelled as

$$(15) \qquad\qquad\qquad\qquad y = \mathcal{G}(u) + \xi.$$

Here $\mathcal{G} : B_1^t(\mathbb{T}^d) \to \mathbb{R}^J$ is a locally Lipschitz continuous, possibly non-linear operator and ξ is Gaussian observational noise in $\mathbb{R}^J$, $\xi \sim N(0, \Sigma)$ for a positive definite covariance matrix $\Sigma \in \mathbb{R}^{J \times J}$.

We assume that u is distributed according to the B_1^s-Besov measure λ defined in Section 3, so that $\lambda(B_1^t(\mathbb{T}^d)) = 1$ for $t < s - d$. Under the assumption of local Lipschitz continuity of $\mathcal{G}$, it follows that almost surely with respect to y the posterior distribution μ^y on $u|y$, has the following Radon–Nikodym derivative with respect to the prior λ:

$$(16) \qquad\qquad\qquad\qquad \frac{d\mu^y}{d\lambda}(u) = \frac{1}{Z(y)} \exp(-\Phi(u; y)),$$

where

$$(17) \qquad \Phi(\cdot; y) = \frac{1}{2} \left| \Sigma^{-\frac{1}{2}} (y - \mathcal{G}(\cdot)) \right|^2$$

and $Z(y)$ is the normalization constant. Indeed, local Lipschitz continuity of $\mathcal{G}$ implies measurability of $\Phi(\cdot, y)$ with respect to λ; it also, together with non-negativity of Φ, gives the finiteness and non-singularity of Z. For details see [16].

Note that the majority of the results presented below hold for more general data models. Indeed, the explicit form of the potential Φ is used only for the consistency results contained in section 4.2.

Following the intuition described in subsection 1.2, we define the Tikhonov-type functional

$$(18) \qquad I(u; y) := \begin{cases} \Phi(u; y) + \|u\|_{B_1^s}, & u \in B_1^s(\mathbb{T}^d), \\ \infty & \text{otherwise.} \end{cases}$$

The existence of minimizers for I is classical, however, we include the proof for completeness.

LEMMA 4.1. *The functional $I(\cdot; y)$ in equation (18) has a minimizer $\hat{u} \in B_1^s(\mathbb{T}^d)$.*

The underpinning of our main results in this section, theorems 4.3 and 4.6, are lemma 3.6 and theorem 3.9. In lemma 3.6, for a B_1^s-Besov prior, we showed the existence of a subspace over which the limit of the translated small ball probability ratios has a continuous representative, and in theorem 3.9 we established the Onsager-Machlup functional for such prior measures. We now note that, by the local Lipschitz continuity of Φ, it follows directly that the posterior inherits the above properties of the prior:

PROPOSITION 4.2. *Let $A \in \mathcal{B}(B_1^t(\mathbb{T}^d))$ be convex, bounded and symmetric and define $A^\epsilon := \epsilon A$.*

i) For any $h \in B_1^r(\mathbb{T}^d)$ with $r > s$ the mapping

$$u \mapsto \lim_{\epsilon \to 0} \frac{\mu_h^y(A^\epsilon + u)}{\mu^y(A^\epsilon + u)}$$

is a continuous function of $u \in B_1^t(\mathbb{T}^d)$.

ii) The Tikhonov-type functional defined in (18) is the generalized Onsager–Machlup functional for the posterior μ^y in (16). That is, for any $z_1, z_2 \in B_1^s(\mathbb{T}^d)$,

$$\lim_{\epsilon \to 0} \frac{\mu^y(A^\epsilon + z_2)}{\mu^y(A^\epsilon + z_1)} = \exp(I(z_1; y) - I(z_2; y)).$$

The theorems below show that the weak and strong MAP estimates of the posterior μ^y in (16), are identified with minimisers of functional I. In particular, the weak and strong MAP estimates coincide for the inverse problems considered here.

THEOREM 4.3. *An element $u \in B_1^s(\mathbb{T}^d)$ minimizes $I(\cdot; y)$ if and only if it is a weak MAP estimate for the posterior measure μ^y in (16).*

We note that the last result implies the existence of weak MAP estimates by lemma 4.1. We next show the existence of strong MAP estimators, and that any strong MAP estimate is a minimiser of I. The following result also shows that the modes can be approximated arbitrarily closely in B_1^t by the centres of balls of fixed sufficiently small radius with maximal probability.

PROPOSITION 4.4.   *Consider the measure μ^y given by (16) and (17) with the B_1^s-Besov prior λ and $\mathcal{G} : B_1^t(\mathbb{T}^d) \to \mathbb{R}^J$ locally Lipschitz for $t < s - d$.*

   i) *For any $\delta > 0$ there exists $z^\delta \in B_1^t(\mathbb{T}^d)$ satisfying $z^\delta = \arg\max_{z \in X} \mu^y(A^\delta + z)$, where $A^\delta := \delta A$ with A a convex, symmetric and bounded set in $B_1^t(\mathbb{T}^d)$.*
   ii) *There is a $\bar{z} \in B_1^s(\mathbb{T}^d)$ and a subsequence of $\{z^\delta\}_{\delta > 0}$ which converges to $\bar{z}$ strongly in $B_1^t(\mathbb{T}^d)$.*
   iii) *The limit $\bar{z}$ is a strong MAP estimator and a minimizer of $I(u; y)$ in equation (18).*

   COROLLARY 4.5.   *Under the conditions of Proposition 4.4, the mapping $u \mapsto \mu^y(A^\delta + u)$ is continuous in $B_1^t(\mathbb{T}^d)$.*

In the following theorem we prove that any minimizer of I is a strong MAP estimate and, hence, show the identification of strong MAP estimates and minimizers of I using part (iii) of proposition 4.4.

   THEOREM 4.6.   *Suppose that conditions of proposition 4.4 hold. Then the strong MAP estimators of μ^y are characterised by the minimizers of the Onsager–Machlup functional I given in (18).*

The proof of theorem 4.3 concerning weak MAP estimates is relatively straightforward and relies on lemma 3.6, i.e., the ability to consider the subspace $B_1^r(\mathbb{T}^d)$, where the Radon–Nikodym derivative $R_h^{\mu^y}$ has a continuous representative. The proof of proposition 4.4 related to strong MAP estimates is more involved and requires a series of results developing asymptotic estimates for the small ball probability ratios (lemma 6.2 to 6.4 in section 6.3).

The difference in difficulty of the proofs related to the two notions of MAP estimates highlights the flexibility of weak MAP estimators. It seems that explicit calculations are typically required for the proof in the case of strong MAP estimates. For practical purposes, it is very interesting to find general conditions under which the two MAP estimate concepts coincide.

   REMARK 4.7.   *Proposition 4.3.8 in [6] shows that if Φ is convex in u, then since by lemma 3.4 the B_1^s-Besov prior λ is logarithmically-concave, the posterior μ^y is also logarithmically-concave and hence quasi-concave. In that case proposition 2.9 shows that the set of modes is convex. The convexity of Φ depends on the forward operator $\mathcal{G}$: if for example $\mathcal{G}$ is linear then Φ is convex, however in the general nonlinear case Φ may be non-convex.*

4.2. *Weak consistency of the strong MAP.*   We consider the frequentist setup, in which

$$y_j \overset{i.i.d.}{\sim} N(\mathcal{G}(u^\dagger), \Sigma), \;\; j = 1, \ldots, n,$$

for a fixed underlying value of the unknown functional parameter $u^\dagger \in X$. As before, we assume that $\mathcal{G} : X \to \mathbb{R}^J$ is locally Lipschitz and $\Sigma \in \mathbb{R}^{J \times J}$ is a positive definite matrix.

For this set of data and with a B_1^s-Besov prior, λ, the posterior measure satisfies

$$(19) \qquad \frac{d\mu^{y_1,y_2,\ldots,y_n}}{d\lambda}(u) \propto \exp\Big(-\frac{1}{2}\sum_{j=1}^{n}\big|\Sigma^{-\frac{1}{2}}(y_j - \mathcal{G}(u))\big|^2\Big).$$

Here also the Lipschitz continuity of $\mathcal{G}$ implies the well-definedness $\mu^{y_1,y_2,\ldots,y_n}$ [16]. Proposition 4.4 then implies that the strong MAP estimator of the above posterior measure is a minimizer of

$$(20) \qquad I_n(u) := \|u\|_{B_1^s} + \frac{1}{2}\sum_{j=1}^{n}\big|\Sigma^{-\frac{1}{2}}(y_j - \mathcal{G}(u))\big|^2.$$

THEOREM 4.8. *Suppose that $\mathcal{G} : B_1^t(\mathbb{T}^d) \to \mathbb{R}^J$ is locally Lipschitz and $u^\dagger \in B_1^s(\mathbb{T}^d)$. For each $n \in N$, let u_n denote a minimizer of I_n given in (20). Then there exists $u^* \in B_1^s(\mathbb{T}^d)$ and a subsequence of $\{u_n\}$ such that $u_n \to u^*$ in $B_1^{\tilde{s}}(\mathbb{T}^d)$ almost surely for any $\tilde{s} < s$. For any such u^* we have $\mathcal{G}(u^*) = \mathcal{G}(u^\dagger)$.*

If $u^\dagger$ lives only in $X = B_1^t(\mathbb{T}^d)$, and not necessarily in $B_1^s(\mathbb{T}^d)$, we can only get the convergence of $\{\mathcal{G}(u_n)\}$:

COROLLARY 4.9. *Let $\mathcal{G}$ and u_n, $n \in \mathbb{N}$, satisfy the assumptions of theorem 4.8 and suppose that $u^\dagger \in B_1^t(\mathbb{T}^d)$. Then $\{\mathcal{G}(u_n)\}_{n\in\mathbb{N}}$ converges to $\mathcal{G}(u^\dagger)$ in probability.*

REMARK 4.10. *Theorem 4.8 states that the true solution is identified in the range of $\mathcal{G}$, which is the natural objective also in regularization theory [28]. We remark the following:*

  i) *The full identification of $u^\dagger$ is dependent on further properties of $\mathcal{G}$, e.g., injectivity of $\mathcal{G}$ would immediately yield $u^* = u^\dagger$.*
  ii) *An inspection of the proof of Theorem 4.8 shows that for an injective operator $\mathcal{G}$, we have convergence in probability of the full sequence $\{u_n\}_{n\in\mathbb{N}}$ to $u^\dagger$. Indeed, for such a $\mathcal{G}$, assuming that the full sequence does not converge in probability contradicts equation (41).*
  iii) *It is an immediate consequence of the Corollary 4.9 that there exists a subsequence of $\{\mathcal{G}(u_n)\}_{n\in\mathbb{N}}$ converging to $\mathcal{G}(u^\dagger)$ almost surely.*

5. Connections to logarithmic derivative.   In this section we discuss the logarithmic derivative of the posterior measure. We mainly revisit known results (see e.g. [6]) and also derive the logarithmic derivative of the posterior μ^y given in (16). The intuition behind logarithmic derivative is that it roughly corresponds to the Gâteaux derivative of the posterior potential I. If the logarithmic derivative is smooth, then its zero points can determine the weak MAP estimates as shown in [32]. The Besov B_1^s-prior does not meet this criteria due to the discontinuity of its logarithmic derivative at the origin as we show in theorem 5.8. The logarithmic derivative also determines the Radon–Nikodym derivative as recorded in proposition 5.3 below. Moreover, it can be used as a basis of Newton-type algorithms to estimate the weak MAP in case an explicit form of the potential I is not easily accessible, see e.g. Cauchy priors in [67].

DEFINITION 5.1. *A measure μ on X is called Fomin differentiable along the vector h if, for every set $\mathcal{A} \in \mathcal{B}(X)$, there exists a finite limit*

$$(21) \qquad d_h\mu(\mathcal{A}) = \lim_{t\to 0}\frac{\mu(\mathcal{A}+th) - \mu(\mathcal{A})}{t}$$

It is well-known that if μ is Fomin differentiable along h then the limit $d_h\mu$ is a countably additive signed measure on $\mathcal{B}(X)$ and has bounded variation [6]. Moreover, $d_h\mu$ is absolutely continuous with respect to μ.

We denote the domain of differentiability by

$$(22) \qquad\qquad D(\mu) = \{h \in X \mid \mu \text{ is Fomin differentiable along } h\}$$

DEFINITION 5.2.   *The Radon–Nikodym density of the measure $d_h\mu$ with respect to μ is denoted by β_h^μ and is called the logarithmic derivative of μ along h.*

PROPOSITION 5.3.   *[6, prop. 6.4.1] Suppose μ is a Radon measure on a locally convex space X and is Fomin differentiable along a vector $h \in X$. If it holds that $\exp(\epsilon|\beta_h^\mu(\cdot)|) \in L^1(\mu)$ for some $\epsilon > 0$, then μ is quasi-invariant along h and the Radon–Nikodym density R_h^μ of μ_h with respect to μ satisfies the equality*

$$(23) \qquad\qquad R_h^\mu(u) = \exp\left(\int_0^1 \beta_h^\mu(u - sh)ds\right) \quad \text{in } L^1(\mu).$$

REMARK 5.4.   *Recall that as discussed in Remark 2.4, it is desirable to choose E in the definition of wMAP to be a subspace $E \subset X$ such that R_h^μ, $h \in E$, has a continuous representative $\tilde{R}_h^\mu$. Therefore, the integral $\int_0^1 \beta_h^\mu(u-sh)ds$ in (23) has a measurable representative, which is continuous outside the set $\{u \in X \mid \tilde{R}_h^\mu(u) = 0\}$. Moreover, a weak mode $\hat{u}$ of μ can be equivalently defined by condition*

$$\int_0^1 \beta_h^\mu(u - sh)ds \le 0$$

for all $h \in E$.

The construction of the Besov prior in definition 3.1 is a prototypical example of a product measure. By setting $\lambda_\ell = \frac{1}{a_\ell}\pi_X\left(\frac{x}{a_\ell}\right)dx$ for $a_n = \ell^{-\left(\frac{s}{d}-\frac{1}{2}\right)}$ we can define the probability law of the Besov prior on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ by $\lambda = \otimes_{\ell=1}^\infty \lambda_\ell$. For product measures, the Fomin differentiability calculus reduces to finite dimensional projections in a straightforward manner.

LEMMA 5.5.   *[6, prop. 3.4.1 (iii)] Let μ be a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then μ is Fomin differentiable along $h \ne 0$ if and only if has an absolutely continuous density π_μ whose derivative satisfies $\pi_\mu' \in L^1(\mathbb{R})$. In this case, $d_1\mu = \pi_\mu'dx$.*

The Fomin differentiability of a product measure $\mu = \otimes_{n=1}^\infty \mu_n$ on the space $X = \prod_{n=1}^\infty X_n$ assigned with the product topology is characterized by the following theorem:

PROPOSITION 5.6.   *[6, prop. 4.1.1.] Suppose that $\beta_{h_n}^{\mu_n}$ is the logarithmic derivative of μ_n in the direction $h_n \in X_n$. The following claims are equivalent:*

(i) μ is differentiable along $h = (h_j)_{j=1}^\infty \in X$,
(ii) the series $\sum_{n=1}^\infty \beta_{h_n}^{\mu_n}$ converges in the norm of $L^1(\mu)$ and

$$\beta_h^\mu = \lim_{N\to\infty} \sum_{n=1}^N \beta_{h_n}^{\mu_n}.$$

Let us introduce the following subspace

$$H(\mu) = \{h \in D(\mu) \mid \beta_h^\mu \in L^2(\mu)\} \subset D(\mu),$$

which has a natural Hilbert space structure [6, section 5]. Surprisingly, for a large class of product measures $H(\mu)$ coincides with $D(\mu)$. The following proposition follows from [7, cor. 2] and [6, ex. 5.2.3].

PROPOSITION 5.7. *Suppose $\mu = \pi(x)dx$ is a Borel probability measure on the real line such that*

$$\int_{\mathbb{R}} \frac{\pi'(t)^2}{\pi(t)} dt < \infty.$$

If we set $\mu_n(A) = \mu(A/a_n)$, where $a_n > 0$, and $\mu = \otimes_{n=1}^\infty \mu_n$, then it follows that

$$D(\mu) = H(\mu) = \left\{ h \in \mathbb{R}^\infty \ \middle| \ \sum_{n=1}^\infty a_n^{-2} h_n^2 < \infty \right\}.$$

Let us record the following direct consequence of proposition 5.3: if $t > 0$ then

(24) $$R_{th}(u) = \exp\left(\int_0^1 \beta_{th}^\mu (u - s \cdot th) ds \right) = \exp\left(\int_0^t \beta_h^\mu (u - s' \cdot h) ds' \right)$$

in $L^1(\mu)$. It is rather easy to see that λ_1 (and consequently λ_ℓ for any $\ell \in \mathbb{N}$) is Fomin differentiable since

$$\left| \lim_{t \to 0} \frac{\lambda_1(\mathcal{A} + t) - \lambda_1(\mathcal{A})}{t} \right| = \lim_{t \to 0} \frac{1}{2t} \left| \int_{\mathcal{A} \cap \mathbb{R}_-} (e^{x+t} - e^x) dx + \int_{\mathcal{A} \cap \mathbb{R}_+} (e^{-x-t} - e^{-x}) dx \right| < 1$$

for any $\mathcal{A} \in \mathcal{B}(\mathbb{R})$. In more generality, this follows from lemma 5.5 since the density function is absolutely continuous.

THEOREM 5.8. *Let λ be the B_1^s-Besov measure given in definition 3.1. The set of differentiability is given by $D(\lambda) = B_2^{s-\frac{d}{2}}(\mathbb{T}^d)$ and for any $h = \sum_{\ell=1}^\infty h_\ell \phi_\ell \in D(\lambda)$ we have*

$$\beta_h^\lambda(u) = \lim_{N \to \infty} \sum_{\ell=1}^N \beta_{h_\ell}^{\lambda_\ell}(u_\ell) \quad \text{in } L^1(\lambda),$$

where

$$\beta_{h_\ell}^{\lambda_\ell}(u_\ell) = -\ell^{\frac{s}{d}-\frac{1}{2}} \operatorname{sign}(u_\ell) h_\ell$$

is the logarithmic derivative of ρ_ℓ (see notation 3.3).

Notice that the previous theorem directly states that for any $h \in B_1^s(\mathbb{T}^d)$ the logarithmic derivative is bounded $|\beta_h^\lambda(u)| \leq C \|h\|_{B_1^s}$ λ-almost surely.

For the posterior distribution μ^y we can solve the logarithmic derivative by using properties of the prior and the functional Φ. The following result follows directly from [6, prop. 3.3.12] (see also [26, thm. 5.7]).

THEOREM 5.9. *Suppose that $\Phi : B_1^t(\mathbb{T}^d) \to \mathbb{R}$, with $t < s - d$, is bounded from below and possesses a uniformly bounded derivative. Then we have $\beta_h^{\mu^y} = -\partial_h \Phi(u) - \beta_h^\lambda(u)$ for any $h \in D(\lambda)$.*

6. Proofs.

6.1. *Proofs of results in section 2.* *Proof of proposition 2.8.* Without loss of generality assume that μ is symmetric around the origin and show that the origin is a strong mode. For any $u \in X$ the Anderson inequality (proposition 2.7) implies $\mu(B_\epsilon(u)) \leq \mu(B_\epsilon(0))$, and so $\mu(B_\epsilon(0)) = \sup_{u \in X} \mu(B_\epsilon(u))$ and the origin is a strong mode of μ. $\square$

Proof of proposition 2.9. Suppose u_1, u_2 are strong modes. For $\kappa \in (0,1)$ we show that $\hat{u} = \kappa u_1 + (1-\kappa)u_2$ is also a strong mode. For $\epsilon > 0$ define $M_\epsilon = \sup_{u \in X} \mu(B_\epsilon(u))$. By quasi-concavity and the identity

$$(25) \qquad\qquad B_\epsilon(\kappa u_1 + (1-\kappa)u_2) = \kappa B_\epsilon(u_1) + (1-\kappa)B_\epsilon(u_2),$$

we have that

$$\frac{\mu(B_\epsilon(\hat{u}))}{M_\epsilon} \geq \frac{\min\{\mu(B_\epsilon(u_1)), \mu(B_\epsilon(u_2))\}}{M_\epsilon},$$

so that since u_1, u_2 are strong modes we get

$$\liminf_{\epsilon \to 0} \frac{\mu(B_\epsilon(\hat{u}))}{M_\epsilon} \geq 1.$$

Since for all $\epsilon > 0$ we have $\mu(B_\epsilon(\hat{u}))/M_\epsilon \leq 1$, we get that $\hat{u}$ is a strong mode. $\square$

Proof of theorem 2.10. Let us consider the identity (25) with values $u_1 = \hat{u} - h$ and $u_2 = \hat{u}$. By applying log-concavity we have

$$\mu(B_\epsilon(\hat{u} - \kappa h)) \geq \mu(B_\epsilon(\hat{u} - h))^\kappa \mu(B_\epsilon(\hat{u}))^{1-\kappa}$$

and, consequently,

$$(26) \qquad\qquad \frac{\mu(B_\epsilon(\hat{u} - h))}{\mu(B_\epsilon(\hat{u}))} \leq \left(\frac{\mu(B_\epsilon(\hat{u} - \kappa h))}{\mu(B_\epsilon(\hat{u}))} \right)^{1/\kappa}.$$

for any $0 \leq \kappa \leq 1$.

Now suppose $\hat{u} \in X$ is local mode in a neighborhood $B_\delta(\hat{u})$ but not a global mode. Then there exists $\delta > 0$ such that $\hat{u}$ is a local mode in the neighborhood $B_\delta(\hat{u})$ but not in $B_{\delta+1}(\hat{u})$. That is, in the larger neighborhood $B_{\delta+1}(\hat{u})$ we have for some $\eta > 0$ that there exists such a subsequence $\{\epsilon_j\}_{j=1}^\infty$ that

$$\lim_{j \to \infty} \frac{\mu(B_{\epsilon_j}(\hat{u}))}{M_{\delta+1}^{\epsilon_j}} < 1 - \eta.$$

Now let us choose a sequence $\{u_j\}_{j=1}^\infty \subset B_{\delta+1}(\hat{u})$ that

$$M_{\delta+1}^{\epsilon_j} \leq \mu(B_{\epsilon_j}(u_j)) + \frac{M_{\delta+1}^{\epsilon_j}}{j}.$$

Then it follows that

$$\limsup_{j \to \infty} \frac{\mu(B_{\epsilon_j}(\hat{u}))}{\mu(B_{\epsilon_j}(u_j))} \leq \lim_{j \to \infty} \frac{\mu(B_{\epsilon_j}(\hat{u}))}{(1 - \frac{1}{j})M_{\delta+1}^{\epsilon_j}} \leq 1 - \eta.$$

Since $\hat{u}$ is a local mode there exists $\tilde{\epsilon} > 0$ such that for any $\epsilon < \tilde{\epsilon}$ we have

$$\frac{\mu(B_\epsilon(\hat{u}))}{M_\delta^\epsilon} \geq \left(1 - \frac{\eta}{2}\right)^{\tilde{\delta}}$$

for $\tilde{\delta} = \frac{\delta}{2(1+\delta)}$ and some $\eta > 0$. In fact, for this choice of $\tilde{\delta}$ we have $\hat{u} - \tilde{\delta}(\hat{u} - u_j) \in B_\delta(\hat{u})$ and it follows by (26) for any $\epsilon_j < \tilde{\epsilon}$ that

$$
\begin{aligned}
1 - \frac{\eta}{2} &\leq \left(\frac{\mu(B_{\epsilon_j}(\hat{u}))}{M_\delta^{\epsilon_j}}\right)^{1/\tilde{\delta}} \\
&\leq \left(\frac{\mu(B_{\epsilon_j}(\hat{u}))}{\mu(B_{\epsilon_j}(\hat{u} - \tilde{\delta}(\hat{u} - u_{\epsilon_j})))}\right)^{1/\tilde{\delta}} \\
&\leq \frac{\mu(B_{\epsilon_j}(\hat{u}))}{\mu(B_{\epsilon_j}(u_{\epsilon_j}))} \\
&\leq 1 - \eta.
\end{aligned}
$$

This yields a contradiction and proves the claim for strong MAP estimates.

Suppose now that $\hat{u} \in X$ is a local wMAP but not a global wMAP. Assume like above that $B_\delta(\hat{u})$ is the maximal neighborhood, where $\hat{u}$ is a local wMAP. Then there exists an element $h \in E$ and $h \notin B_\delta(\hat{u})$ such that

$$\lim_{\epsilon \to 0} \frac{\mu(B_\epsilon(\hat{u} - \tilde{\delta}h))}{\mu(B_\epsilon(\hat{u}))} \leq 1 \quad \text{but} \quad \lim_{\epsilon \to 0} \frac{\mu(B_\epsilon(\hat{u} - h))}{\mu(B_\epsilon(\hat{u}))} > 1 + \eta$$

for $\tilde{\delta} = \frac{\delta}{2\|h\|_X}$, since $\hat{u} - \tilde{\delta}h \in B_\delta(\hat{u})$. Again we see that the inequality (26) yields a contradiction. This completes the proof. $\qquad \square$

6.2. *Proofs of results in section 3.* *Proof of lemma 3.4.* Let $\lambda^N = \bigotimes_{\ell=1}^N \rho_\ell$. Then it is straightforward to check that λ^N converges weakly to λ as $N \to \infty$. By [8, theorem 2.2], for λ to be logarithmically concave, it suffices to show that the measures λ^N in $\mathbb{R}^N$ are logarithmically concave. Note that the measures λ^N have density, denoted π_N, with respect to the Lebesgue measure in $\mathbb{R}^N$, given by

$$\pi_N(x_1, .., x_N) = \left(\prod_{\ell=1}^N \frac{\alpha_\ell}{2}\right) \exp\left(-\sum_{\ell=1}^N \alpha_\ell |x_\ell|\right),$$

for all $x = (x_1, .., x_N) \in \mathbb{R}^N$, where $\alpha_\ell = \ell^{\frac{s}{d} - \frac{1}{2}}$ are the coefficients in the expansion defining the B_1^s-Besov measure. Since the density π_N is a logarithmically concave function, by [5, theorem 1.8.4] λ^N is logarithmically concave and the result follows. $\qquad \square$

Proof of lemma 3.5. We use the Kakutani-Hellinger theory [15, Chapter 2]. We start by calculating

the Hellinger integrals $H(\rho_{h,\ell}, \rho_\ell)$, where $\rho_{h,\ell}(\cdot) := \rho_\ell(\cdot - h_\ell)$:

$$
\begin{aligned}
H(\rho_{h,\ell}, \rho_\ell) &= \int_{-\infty}^\infty \sqrt{\frac{d\rho_{h,\ell}}{dx}\frac{d\rho_\ell}{dx}}\, dx \\
&= \int_{-\infty}^\infty \frac{\alpha_\ell}{2} \mathrm{e}^{-\frac{\alpha_\ell}{2}(|x-h_\ell|+|x|)}\, dx \\
&= \mathrm{e}^{-\frac{\alpha_\ell}{2}|h_\ell|}\left(1 + \frac{\alpha_\ell}{2}|h_\ell|\right).
\end{aligned}
$$

By [15, lemma 2.5], we have

$$
H(\lambda_h, \lambda) = \lim_{N\to\infty} H_N,
$$

where $H_N = \prod_{\ell=1}^N \mathrm{e}^{-\frac{\alpha_\ell}{2}|h_\ell|}\left(1 + \frac{\alpha_\ell}{2}|h_\ell|\right) \in (0,1]$. By taking the negative logarithm we get

$$
-\log\left(H(\lambda_h, \lambda)\right) = -\log\left(\lim_{N\to\infty} H_N\right) = -\lim_{N\to\infty}\log(H_N) = \lim_{N\to\infty}\sum_{\ell=1}^N\left(\frac{\alpha_\ell}{2}|h_\ell| - \log\left(1 + \frac{\alpha_\ell}{2}|h_\ell|\right)\right).
$$

By [15, theorem 2.7] the set $Q(\lambda)$ coincides with the set of h such that $-\log\left(H(\lambda_h, \lambda)\right) < \infty$. The Taylor theorem implies the upper and lower bounds

$$
x - \frac{x^2}{2} \leq \log(1 + x) \leq x - \frac{x^2}{2(1+x)^2}.
$$

Using the lower bound, we get that

$$
-\log\left(H(\lambda_h, \lambda)\right) \leq \lim_{N\to\infty}\sum_{\ell=1}^N \frac{1}{2}\left(\frac{\alpha_\ell}{2}|h_\ell|\right)^2 = \frac{1}{8}\sum_{\ell=1}^\infty \alpha_\ell^2 |h_\ell|^2,
$$

which implies that a sufficient condition for the equivalence of λ_h and λ is

$$
\text{(27)} \qquad \sum_{\ell=1}^\infty \ell^{\frac{2s}{d}-1} h_\ell^2 < \infty.
$$

Using the upper bound, and letting $x_\ell = \frac{\alpha_\ell}{2}|h_\ell|$, we get that

$$
-\log\left(H(\lambda_h, \lambda)\right) \geq \sum_{\ell=1}^\infty \frac{x_\ell^2}{2(1+x_\ell)^2}.
$$

If x_ℓ is unbounded, then the sum on the right hand side is infinite and we have that λ_h and λ are singular (note that if x_ℓ is unbounded then obviously condition (27) does not hold). If x_ℓ is bounded, $x_\ell \leq M$, then

$$
\sum_{\ell=1}^\infty \frac{x_\ell^2}{2(1+x_\ell)^2} \geq \frac{1}{2(1+M)^2}\sum_{\ell=1}^\infty x_\ell^2,
$$

therefore condition (27) is also necessary for the equivalence of the measures λ_h and λ.

Note, that condition (27) is equivalent to $h \in B_2^{s-\frac{d}{2}}(\mathbb{T}^d)$. Observing that

$$
\frac{d\rho_{h,\ell}}{d\rho_\ell}(x) = \mathrm{e}^{-\alpha_\ell|x-h_\ell|+\alpha_\ell|x|}
$$

for all $x \in \mathbb{R}$ and $\ell \in \mathbb{N}$, the claimed expression for $\frac{d\lambda_h}{d\lambda}$ follows from [15, theorem 2.7]. $\qquad \square$

Proof of lemma 3.6. We have

(28)
$$-|h_\ell - u_\ell| + |u_\ell| = -|h_\ell| + \big(|u_\ell| + \operatorname{sign}(h_\ell)u_\ell\big)$$

if $|u_\ell| \leq |h_\ell|$ and

(29)
$$-|h_\ell - u_\ell| + |u_\ell| = \operatorname{sign}(u_\ell)h_\ell$$

if $|u_\ell| > |h_\ell|$. Fix $\eta > 0$ and consider $v \in B_1^t(\mathbb{T}^d)$ such that $\|v - u\|_{B_1^t} < \eta$. Let us next define index sets

$$
\begin{aligned}
A_1 &= \{\ell \in \mathbb{N} : |u_\ell| \leq |h_\ell|, |v_\ell| \leq |h_\ell|\}, \\
A_2 &= \{\ell \in \mathbb{N} : |u_\ell| \leq |h_\ell|, |v_\ell| > |h_\ell|\}, \\
A_3 &= \{\ell \in \mathbb{N} : |u_\ell| > |h_\ell|, |v_\ell| \leq |h_\ell|\} \quad \text{and} \\
A_4 &= \{\ell \in \mathbb{N} : |u_\ell| > |h_\ell|, |v_\ell| > |h_\ell|\}.
\end{aligned}
$$

Clearly A_j are disjoint and $\mathbb{N} = \cup_{j=1}^4 A_j$. We utilise the index sets A_j to rewrite

(30)
$$
\begin{aligned}
|\tilde{R}_h^\lambda(v) - \tilde{R}_h^\lambda(u)| &= \left| \tilde{R}_h^\lambda(u) \left(\frac{\tilde{R}_h^\lambda(v)}{\tilde{R}_h^\lambda(u)} - 1 \right) \right| \\
&\leq e^{\|h\|_{B_1^s}} \left| \frac{\tilde{R}_h^\lambda(v)}{\tilde{R}_h^\lambda(u)} - 1 \right| \leq e^{\|h\|_{B_1^s}} \left| \exp(\mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3 + \mathcal{I}_4) - 1 \right|,
\end{aligned}
$$

where

$$\mathcal{I}_j := \sum_{\ell \in A_j} \alpha_\ell \left(-|h_\ell - v_\ell| + |v_\ell| - \left(-|h_\ell - u_\ell| + |u_\ell| \right) \right).$$

We continue by studying the terms $\mathcal{I}_j$ separately. Let $0 < \delta < 1$ be a value, which we later fix. For $\mathcal{I}_1$ we have by the Hölder inequality

$$
\begin{aligned}
|\mathcal{I}_1| &= \left| \sum_{\ell \in A_1} \alpha_\ell \left(-|h_\ell| + \big(|v_\ell| + \operatorname{sign}(h_\ell)v_\ell\big) - \big(-|h_\ell| + \big(|u_\ell| + \operatorname{sign}(h_\ell)u_\ell\big)\big) \right) \right| \\
&\leq \sum_{\ell \in A_1} \alpha_\ell \big(\big||v_\ell| - |u_\ell|\big| + |v_\ell - u_\ell|\big) \\
&\leq 2 \sum_{\ell \in A_1} \alpha_\ell |v_\ell - u_\ell| \\
&= 2 \sum_{\ell \in A_1} \ell^{\delta(\frac{t}{d} - \frac{1}{2})} |v_\ell - u_\ell|^\delta \ell^{\frac{s - \delta t}{d} - (1-\delta)\frac{1}{2}} |v_\ell - u_\ell|^{1-\delta} \\
&\leq \left(\sum_{\ell \in A_1} 2\ell^{\frac{t}{d} - \frac{1}{2}} |v_\ell - u_\ell| \right)^\delta \left(\sum_{\ell \in A_1} \ell^{\frac{s - \delta t}{d(1-\delta)} - \frac{1}{2}} |v_\ell - u_\ell| \right)^{1-\delta} \\
&\leq 2\eta^\delta \|h\|_{B_1^{s + \delta_1(s-t)}}^{1-\delta}
\end{aligned}
$$

for $\delta_1 = \frac{\delta}{1-\delta}$, where to bound the second parenthesis in the second to last line we have used that $|v_\ell - u_\ell| \le 2|h_\ell|$ for any $\ell \in A_1$. For $\mathcal{I}_2$ we have

$$\mathcal{I}_2 = \sum_{\ell \in A_2} \alpha_\ell \, \mathrm{sign}(v_\ell) h_\ell - \Big(- \alpha_\ell |h_\ell| + \alpha_\ell \big(|u_\ell| + \mathrm{sign}(h_\ell) u_\ell \big) \Big)$$

$$= \sum_{I \in A_2} \alpha_\ell \Big(\mathrm{sign}(v_\ell) h_\ell - \mathrm{sign}(h_\ell) u_\ell \Big) + \alpha_\ell \big(|h_\ell| - |u_\ell| \big).$$

We note that since $|v_\ell| > |h_\ell| \ge |u_\ell|$, the following two inequalities hold:

$$|h_\ell| - |u_\ell| \;\le\; \min(|h_\ell|, |v_\ell - u_\ell|) \quad \text{and}$$
$$|\mathrm{sign}(v_\ell)|h_\ell| - u_\ell| \;\le\; \min(2|h_\ell|, |v_\ell - u_\ell|)$$

As a direct consequence we have

$$\max(|h_\ell| - |u_\ell|, |\mathrm{sign}(v_\ell)|h_\ell| - u_\ell|) \le c |h_\ell|^{1-\delta} |v_\ell - u_\ell|^\delta$$

for any $0 \le \delta \le 1$. Moreover, since

$$|\mathrm{sign}(v_\ell) h_\ell - \mathrm{sign}(h_\ell) u_\ell| = |\mathrm{sign}(v_\ell)|h_\ell| - u_\ell|$$

we obtain

$$\mathcal{I}_2 \le c \sum_{\ell \in A_2} \ell^{\delta(\frac{t}{d} - \frac{1}{2})} \ell^{\frac{s - \delta t}{d} - (1-\delta)\frac{1}{2}} |h_\ell|^{1-\delta} \, |v_\ell - u_\ell|^\delta \;\le\; c \eta^\delta \|h\|_{B_1^{s+\delta_1(s-t)}}^{1-\delta},$$

where as before $\delta_1 = \frac{\delta}{1-\delta}$. The term $\mathcal{I}_3$ is very similar to $\mathcal{I}_2$ (only the role of u_ℓ and v_ℓ is swapped), and we have

$$|\mathcal{I}_3| \;\le\; c \eta^\delta \, \|h\|_{B_1^{s+\delta_1(s-t)}}^{1-\delta}.$$

For $\mathcal{I}_4$ we first note that since $|v_\ell| > |h_\ell|$ and $|u_\ell| > |h_\ell|$ for $\ell \in A_4$, if $|h_\ell| > \frac{\eta}{2\beta_\ell}$, with $\beta_\ell = \ell^{\frac{t}{d} - \frac{1}{2}}$, we must have $\mathrm{sign}(u_\ell) = \mathrm{sign}(v_\ell)$. Now it follows that

$$
\begin{aligned}
|\mathcal{I}_4| \;&=\; \left| \sum_{\ell \in A_4} \alpha_\ell \left(\mathrm{sign}(v_\ell) - \mathrm{sign}(u_\ell) \right) h_\ell \right| \\
&\le\; 2 \sum_{A_4 \cap \{ l \,;\, |h_\ell| < \frac{\eta}{2\beta_\ell} \}} \alpha_\ell \, |h_\ell| \\
&\le\; 2^{1-\delta} \eta^\delta \sum_{\ell=1}^{\infty} \ell^{\frac{s}{d} - \frac{1}{2} - \delta(\frac{t}{d} - \frac{1}{2})} |h_\ell|^{1-\delta} \\
&=\; 2^{1-\delta} \eta^\delta \sum_{\ell=1}^{\infty} \ell^{-\delta - \delta^2} \ell^{\frac{s}{d} - \frac{1}{2} - \delta(\frac{t}{d} - \frac{3}{2} - \delta)} |h_\ell|^{1-\delta} \\
&\le\; 2^{1-\delta} \eta^\delta \left(\sum_{\ell=1}^{\infty} \ell^{-1-\delta} \right)^\delta \left(\sum_{\ell=1}^{\infty} \ell^{(\frac{s}{d} - \frac{1}{2} - \tilde\delta)\frac{1}{1-\delta}} |h_\ell| \right)^{1-\delta} \\
&\le\; 2^{1-\delta} c_\delta \eta^\delta \|h\|_{B_1^{s+\delta_2}}^{1-\delta},
\end{aligned}
$$

where

$$\tilde{\delta} = \delta \left(\frac{t}{d} - \frac{3}{2} - \delta \right) \quad \text{and} \quad \delta_2 = \frac{\delta}{1 - \delta} \left(s - \frac{d}{2} \right) - \frac{d\tilde{\delta}}{1 - \delta}.$$

Combining, since for any $r > s$ there exists $\delta > 0$ small enough so that $h \in B_1^{s+\delta_1(s-t)}(\mathbb{T}^d) \cap B_1^{s+\delta_2}(\mathbb{T}^d)$, we get that as $\eta \to 0$, $\tilde{R}_h^\lambda(v) - \tilde{R}_h^\lambda(u) \to 0$. This proves the claim. $\qquad \square$

Proof of theorem 3.9. Let $A^\epsilon := \epsilon A$. We first note that by lemma 3.5 we can write

$$\begin{aligned}
\lambda_h(A^\epsilon) &= \int_{A^\epsilon} \lim_{N\to\infty} e^{\sum_{\ell=1}^N -\alpha_\ell |h_\ell - u_\ell| + \alpha_\ell |u_\ell|} \lambda(du) \\
&\geq \int_{A^\epsilon} \lim_{N\to\infty} e^{\sum_{\ell=1}^N -\alpha_\ell |h_\ell|} \lambda(du) \\
&= e^{-\|h\|_{B_1^s}} \int_{A^\epsilon} \lambda(du) \\
&= e^{-\|h\|_{B_1^s}} \lambda(A^\epsilon).
\end{aligned}$$

This implies that

(31)
$$\liminf_{\epsilon\to 0} \frac{\lambda_h(A^\epsilon)}{\lambda(A^\epsilon)} \geq e^{-\|h\|_{B_1^s}}.$$

Now consider $\{h^j\}_{j=1}^\infty \subset B_1^{s+1}(\mathbb{T}^d)$ with $h^j \to h$ in $B_1^s(\mathbb{T}^d)$. We have

$$\begin{aligned}
\lambda_h(A^\epsilon) &= \int_{A^\epsilon} \lim_{N\to\infty} e^{\sum_{\ell=1}^N -\alpha_\ell |h_\ell - u_\ell| + \alpha_\ell |u_\ell|} \lambda(du) \\
&= e^{\|h-h^j\|_{B_1^s}} \int_{A^\epsilon} \lim_{N\to\infty} e^{\sum_{\ell=1}^N -\alpha_\ell |h_\ell - u_\ell| + \alpha_\ell |u_\ell| - \alpha_\ell |h_\ell^j - h_\ell|} \lambda(du) \\
&\leq e^{\|h-h^j\|_{B_1^s}} \int_{A^\epsilon} \lim_{N\to\infty} e^{\sum_{\ell=1}^N -\alpha_\ell |h_\ell^j - u_\ell| + \alpha_\ell |u_\ell|} \lambda(du) \\
&= e^{\|h-h^j\|_{B_1^s}} \lambda_{h^j}(A^\epsilon).
\end{aligned}$$

This, using lemmas 3.6 and 2.3 implies that

$$\limsup_{\epsilon\to 0} \frac{\lambda_h(A^\epsilon)}{\lambda(A^\epsilon)} \leq e^{\|h-h^j\|_{B_1^s}} e^{-\|h^j\|_{B_1^s}},$$

and letting $j \to \infty$ in the right-hand side gives

$$\limsup_{\epsilon\to 0} \frac{\lambda_h(A^\epsilon)}{\lambda(A^\epsilon)} \leq e^{-\|h\|_{B_1^s}}.$$

The above inequality together with (31) give the result. $\qquad \square$

6.3. *Proofs of results in section 4.* In some of the proofs below we consider the sequences that converge in the weak*-topology of $B_1^t(\mathbb{T}^d)$. We note that the Banach space $B_1^t(\mathbb{T}^d)$ is isomorphic to a weighted l^1 space, and hence its pre-dual is the space of functions $\{v \in B_\infty^{-t}(\mathbb{T}^d) : \lim_{j\to\infty} \langle v, \psi_\ell \rangle = 0\}$.

Proof of lemma 4.1. Suppose $\{u_j\}_{j=1}^\infty \subset B_1^t(\mathbb{T}^d)$ is a minimizing sequence of functional I. Clearly, we can assume $\{I(u_j, y)\}_{j=1}^\infty$, and therefore also $\|u_j\|_{B_1^s}$ to be bounded. By the Banach–Alaoglu theorem there exists a subsequence that converges to some $\hat{u} \in B_1^s(\mathbb{T}^d)$ in the weak*-topology. Notice that the norm of $B_1^s(\mathbb{T}^d)$ is lower semicontinuous in the weak*-topology and consequently $\hat{u} \in B_1^s(\mathbb{T}^d)$. We now show the strong convergence of the above subsequence in $B_1^{\tilde{s}}(\mathbb{T}^d)$ for any $\tilde{s} < s$. We have

$$
\begin{aligned}
\|u_j - \hat{u}\|_{B_1^{\tilde{s}}} &= \sum_{\ell=1}^\infty \ell^{\frac{\tilde{s}}{d} - \frac{1}{2}} |\langle u_j - \hat{u}, \psi_\ell \rangle| \\
&\leq \sum_{\ell=1}^N \ell^{\frac{\tilde{s}}{d} - \frac{1}{2}} |\langle u_j - \hat{u}, \psi_\ell \rangle| + N^{-\frac{s-\tilde{s}}{d}} \sum_{\ell=N+1}^\infty \ell^{\frac{s}{d} - \frac{1}{2}} |\langle u_j - \hat{u}, \psi_\ell \rangle|.
\end{aligned}
$$

Noting that $\|\hat{u}\|_{B_1^s} + \|u_j\|_{B_1^s} \leq C < \infty$, given any $\epsilon > 0$, N can be chosen large enough, independently of j, such that the second term in the last line of the above inequality is bounded by $\epsilon/2$. Having convergence coefficient-wise, there is $M \in \mathbb{N}$ large enough so that for $j > M$ the first term is bounded by $\epsilon/2$ as well. We therefore conclude that $u_j \to \hat{u}$ in $B_1^{\tilde{s}}(\mathbb{T}^d)$ for any $\tilde{s} < s$ and hence in particular for $\tilde{s} = t < s - d$. By the continuity assumption on Φ, it now follows that

$$
\begin{aligned}
\Phi(\hat{u}; y) + \|\hat{u}\|_{B_1^s} &\leq \lim_{k\to\infty} \Phi(u_{j_k}; y) + \liminf_{k\to\infty} \|u_{j_k}\|_{B_1^s} \\
&\leq \liminf_{k\to\infty} \left(\Phi(u_{j_k}; y) + \|u_{j_k}\|_{B_1^s} \right).
\end{aligned}
$$

Therefore, $\hat{u}$ must be a minimizer. $\qquad\square$

Proof of theorem 4.3. Assume that $u_{min} \in B_1^s(\mathbb{T}^d)$ is a minimizer of $I(\cdot; y)$. By lemma 3.6 and Lipschitz continuity of $\mathcal{G}$ we know that $R_h^{\mu^y} \in C(B_1^t(\mathbb{T}^d))$ for any $h \in B_1^r(\mathbb{T}^d)$ with $r > s$. Since $u_{min} \in B_1^s(\mathbb{T}^d)$, we can study $R_h^{\mu^y}$ pointwise and obtain

$$
\log(R_h^{\mu^y}(u_{min})) = I(u_{min}) - I(u_{min} - h) \leq 0
$$

for any $h \in B_1^r(\mathbb{T}^d)$ due to the minimizing property of u_{min}. Therefore, u_{min} is a weak MAP.

Consider the reversed claim and assume that $\hat{u}$ is a weak MAP to the posterior μ^y in (16). Let us also assume that $\hat{u} \in B_1^t(\mathbb{T}^d) \setminus B_1^s(\mathbb{T}^d)$. Due to the continuity of Φ and lemma 3.6 we have

$$
(32) \qquad \log R_h^{\mu^y}(\hat{u}) = -\Phi(\hat{u} - h) + \Phi(\hat{u}) + \sum_{\ell=1}^\infty \alpha_\ell(-|h_\ell - u_\ell| + |u_\ell|) \leq 0
$$

for any $h \in B_1^r(\mathbb{T}^d)$, $r > s$. Let us construct a particular function $h^N = \sum_{\ell=1}^\infty h_\ell^N \psi_\ell \in B_1^r(\mathbb{T}^d)$ by defining its coefficient vector according to

$$
(h_\ell^N)_{\ell=1}^\infty = \begin{cases} \epsilon \hat{u}_\ell, & l \leq N, \\ 0, & l > N, \end{cases}
$$

for some small $\epsilon > 0$. It follows by inequality (32) and continuity of Φ that

$$
(33) \qquad \epsilon \sum_{\ell=1}^N \alpha_\ell |u_\ell| \leq |\Phi(\hat{u} - h^N) - \Phi(\hat{u})| \leq C\epsilon,
$$

where $C > 0$ is the local Lipschitz constant on the neighbourhood of $\hat{u}$. However, N was chosen arbitrarily and by our assumption on the smoothness of $\hat{u}$ the sum on the left hand side of (33) does not stay bounded when N increases. Therefore, inequality (33) leads to a contradiction and we must have $\hat{u} \in B_1^s(\mathbb{T}^d)$.

Assuming now that the weak MAP $\hat{u} \in B_1^s(\mathbb{T}^d)$, we can separate the sum in (32) and obtain

$$I(\hat{u} - h) - I(\hat{u}) = \log R_h^{\mu^y}(\hat{u}) \leq 0$$

for any $h \in B_1^r(\mathbb{T}^d)$. By continuity of I and density of $B_1^r(\mathbb{T}^d)$ in $B_1^s(\mathbb{T}^d)$, we find that $\hat{u}$ minimizes I. □

Proof of proposition 4.4 relies on the following four lemmas giving some properties of the Besov prior measure we have here. We list these lemmas and their proofs first.

LEMMA 6.1. *Let X be a separable Banach space, and B an open and convex set in $\mathcal{B}(X)$. For any non-degenerate measure μ with full support we have $\mu(\partial B) = 0$.*

PROOF. For any $\epsilon > 0$, there exists a cylindrical set B_ϵ with $B_\epsilon \supset B$ satisfying $\mu(B_\epsilon) - \mu(B) \leq \epsilon$ [5, lemma 2.1.6]. By definition of cylindrical sets, there exists some $n \in \mathbb{N}$ and $B_\epsilon^0 \in \mathcal{B}(\mathbb{R}^n)$ such that

$$B_\epsilon = \{x \in X : (l_1(x), \ldots, l_n(x)) \in B_\epsilon^0\}.$$

Let $h = (l_1, \ldots, l_n)$. Without loss of generality we can assume that $l_j \in X^*$, $j = 1, \ldots, n$, are linearly independent and, therefore, h is surjective (see discussion in [5, Sec. 2.1.]). Now we have $B \subset h^{-1}(hB) \subset B_\epsilon$. Notice carefully that hB is an open set, since h is an open map by the open mapping theorem. Hence we obtain

$$\begin{aligned}
\mu(\partial B) &\leq \mu(h^{-1}(hB) \setminus B) + \mu(\partial h^{-1}(hB)) \\
&= \mu(h^{-1}(hB) \setminus B) + \mu(\overline{h^{-1}(hB)} \cap (h^{-1}(hB))^c) \\
&= \mu(h^{-1}(hB) \setminus B) + \mu(h^{-1}(\overline{hB}) \cap h^{-1}((hB)^c)) \\
&= \mu(h^{-1}(hB) \setminus B) + \mu(h^{-1}(\overline{hB} \cap (hB)^c)) \\
&= \mu(h^{-1}(hB) \setminus B) + \mu_n(\partial(hB)).
\end{aligned}$$

By our assumption on non-degeneracy of μ, we have that μ_n is absolutely continuous with respect to the Lebesgue measure in $\mathbb{R}^n$, and hence we have $\mu_n(\partial(hB)) = 0$. Noting that $\mu(h^{-1}(hB) \setminus B) \leq \mu(B_\epsilon) - \mu(B) \leq \epsilon$, the result follows. □ □

In the following λ stands for a centered B_1^s-Besov prior on $B_1^t(\mathbb{T}^d)$ with $t < s - d$. Also, we frequently consider projections of λ to a subspace $\text{span}\{\psi_1, \cdots, \psi_n\} \subset B_1^t(\mathbb{T}^d)$. We write $P_n : B_1^t(\mathbb{T}^d) \to \mathbb{R}^n$ as

$$(34) \qquad\qquad P_n u = (\langle \psi_\ell, u \rangle)_{\ell=1}^n$$

and define $\lambda_n(A) := (\lambda \circ P_n^{-1})(P_n A)$ for any $A \in \mathcal{B}(B_1^t)$.

LEMMA 6.2. *Let $A \subset \mathcal{B}(B_1^t(\mathbb{T}^d))$ be any convex, symmetric and bounded set with diameter $\delta = \sup_{u,v \in A} \|u - v\|_{B_1^t} > 0$. For any $z \in B_1^t(\mathbb{T}^d)$, with $t < s - d$, we have*

$$\frac{\lambda(A + z)}{\lambda(A)} \leq e^{-\frac{1}{2}\|z\|_{B_1^t} + \delta}.$$

PROOF. First consider the subspace $\text{span}\{\psi_1, \cdots, \psi_n\} \subset B_1^t(\mathbb{T}^d)$ and let $\tilde{\alpha}_\ell = \ell^{\frac{t}{d} - \frac{1}{2}}$. Now recall $\alpha_\ell \geq \tilde{\alpha}_\ell$ since $s > t + d$. We have

$$
\begin{aligned}
\frac{\lambda_n(A + z)}{\lambda_n(A)} &= \frac{\int_{P_n A + P_n z} \exp\left(-\sum_{\ell=1}^n \alpha_\ell |u_\ell|\right) du}{\int_{P_n A} \exp\left(-\sum_{\ell=1}^n \alpha_\ell |u_\ell|\right) du} \\[2mm]
&\leq \frac{e^{-\frac{1}{2}\left(\|z\|_{B_1^t} - \delta\right)} \int_{P_n A + P_n z} \exp\left(-\sum_{\ell=1}^n (\alpha_\ell - \frac{1}{2}\tilde{\alpha}_\ell)|u_\ell|\right) du}{e^{-\frac{\delta}{2}} \int_{P_n A} \exp\left(-\sum_{\ell=1}^n (\alpha_\ell - \frac{1}{2}\tilde{\alpha}_\ell)|u_\ell|\right) du} \\[2mm]
&= e^{-\frac{1}{2}\|z\|_{B_1^t} + \delta} \frac{\tilde{\lambda}_n(P_n A + P_n z)}{\tilde{\lambda}_n(P_n A)},
\end{aligned}
$$

where $\tilde{\lambda}_n = \prod_{\ell=1}^n \tilde{\rho}_\ell$ with $\tilde{\rho}_\ell \sim \tilde{c}_\ell \exp\left((\alpha_\ell - \frac{1}{2}\tilde{\alpha}_\ell)|u_\ell|\right)$. Since $\tilde{\lambda}_n$ is logarithmically concave, by theorem 6.1 of [8] (see the proof of proposition 2.7 above) we have that $\tilde{\lambda}_n(A + z) \leq \tilde{\lambda}_n(A)$.

To consider the limiting case, note that $P_n^{-1}(P_n A)$ is convex and symmetric. Therefore, we can use the first part of the proof above, and write

$$(35) \qquad \lambda(A + z) \leq \lambda_n(A + z) \leq e^{-\frac{1}{2}\|P_n z\|_{B_1^t} + \delta} \lambda_n(A) \leq e^{-\frac{1}{2}\|z\|_{B_1^t} + \delta}(\lambda(A) + \epsilon),$$

with $\epsilon \to 0$ as $n \to \infty$ due to weak convergence of λ_n to λ and since A is a continuity set by lemma 6.1. $\qquad\square \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

LEMMA 6.3.   *Suppose that $\bar{z} \notin B_1^s(\mathbb{T}^d)$, $\{z^\delta\}_{\delta > 0} \subset B_1^t(\mathbb{T}^d)$, $t < s - d$, and z^δ converges to $\bar{z}$ in the weak*-topology of $B_1^t(\mathbb{T}^d)$ as $\delta \to 0$. Then for any $\epsilon > 0$ there exists δ small enough such that*

$$\frac{\lambda(A^\delta + z^\delta)}{\lambda(A^\delta)} \leq \epsilon,$$

for $A^\delta = \delta A$ where A any convex, symmetric and bounded set in $\mathcal{B}(B_1^t(\mathbb{T}^d))$.

PROOF. Below we write $u_\ell = \langle u, \psi_\ell \rangle$ for any $u \in B_1^t(\mathbb{T}^d)$ and without losing any generality assume that A has diameter 1. Since $\bar{z} \notin B_1^s(\mathbb{T}^d)$, for any $M > 0$ there is an N large enough such that

$$\sum_{\ell=1}^N \alpha_\ell |\bar{z}_\ell| > 4M.$$

Let $\delta_0 < 2MN^{\frac{t-s}{d}}$. Since z^δ converges to $\bar{z}$ in weak*-topology as $\delta \to 0$, we have $\langle \psi_\ell, z^\delta \rangle \to \langle \psi_\ell, \bar{z} \rangle$ for all j, and therefore $\delta_1 < \delta_0$ can be chosen small enough such that

$$\sum_{\ell=1}^N \alpha_\ell |z_\ell^\delta - \bar{z}_\ell| < M.$$

Thus, for any $z \in A^{\delta_1} + z^{\delta_1}$, we can write

$$\sum_{\ell=1}^N \alpha_\ell |z_\ell| \geq \sum_{\ell=1}^N \alpha_\ell (|\bar{z}_\ell| - |z_\ell^{\delta_1} - \bar{z}_\ell| - |z_\ell^{\delta_1} - z_\ell|) > 4M - M - \delta_1 N^{\frac{s-t}{d}} > M.$$

Let $\delta \leq \delta_1$ be sufficiently small so that

$$\inf_{z \in A} \exp\left(-\frac{1}{2}\sum_{\ell=1}^{N} \alpha_\ell |z_\ell|\right) \geq \frac{1}{2}$$

For λ_n we have that for any $M > 0$ there exist $N > 0$ and $\delta_1 > 0$ such that for $n \geq N$ and $\delta < \delta_1$ it follows

$$\frac{\lambda_n(A^\delta + z^\delta)}{\lambda_n(A)} = \frac{\int_{P_n A^\delta + P_n z^\delta} \exp\left(-\sum_{\ell=1}^{n} \alpha_\ell |u_\ell|\right) du}{\int_{P_n A^\delta} \exp\left(-\sum_{\ell=1}^{n} \alpha_\ell |u_\ell|\right) du}$$

$$\leq 2\mathrm{e}^{-M/2} \frac{\int_{P_n A^\delta + P_n z^\delta} \exp\left(-\sum_{\ell=1}^{N} \frac{\alpha_\ell}{2} |u_\ell| - \sum_{\ell=N+1}^{n} \alpha_\ell |u_\ell|\right) du}{\int_{P_n A^\delta} \exp\left(-\sum_{\ell=1}^{N} \frac{\alpha_\ell}{2} |u_\ell| - \sum_{\ell=N+1}^{n} \alpha_\ell |u_\ell|\right) du}$$

$$\leq 2\mathrm{e}^{-M/2},$$

where in the last line we have used the fact that the density in the integrals of the third line is log concave and A is absolutely convex. Similar inequality as (35) generalizes the result for λ. $\square$ $\square$

LEMMA 6.4. *Suppose that $\{z^\delta\}_{\delta>0} \subset B_1^t(\mathbb{T}^d)$ converges in weak*-topology and not strongly in $B_1^t(\mathbb{T}^d)$ to 0 as $\delta \to 0$. Then for any $\epsilon > 0$ there exists δ small enough such that*

$$\frac{\lambda(A^\delta + z^\delta)}{\lambda(A^\delta)} \leq \epsilon$$

for $A^\delta = \delta A$ where A is any convex, bounded and symmetric set in $\mathcal{B}(B_1^t(\mathbb{T}^d))$.

PROOF. Let $\tilde{\alpha}_\ell = \ell^{\frac{t}{d}-\frac{1}{2}}$ and without loss of generality assume that A has diameter 1. For any $j \in \mathbb{N}$ we have $\langle \psi_\ell, z^\delta \rangle \to 0$, as $\delta \to 0$. There exists a subsequence which we relabel $\{z^\delta\}$ for which elements there exists $\kappa > 0$ such that

(36) $$\|z^\delta\|_{B_1^t} > \kappa.$$

Let $M > 0$ be arbitrary and choose N large enough so that $\alpha_\ell > M\tilde{\alpha}_\ell$ for any $j > N$. By the weak* convergence there exists δ small enough such that

$$\sum_{\ell=1}^{N} \tilde{\alpha}_\ell |z_\ell^\delta| < \frac{\kappa}{4}$$

where $z_\ell^\delta = \langle \psi_\ell, z^\delta \rangle$. This then, using (36), means that there exists $n > N$ such that

$$\sum_{\ell=N+1}^{n} \tilde{\alpha}_\ell |z_\ell^\delta| > \frac{\kappa}{4}.$$

Now one can show that

$$\sum_{\ell=N+1}^{n} \tilde{\alpha}_\ell |x_\ell| \geq M\left(\frac{\kappa}{4} - \delta\right) \quad \text{for any } x \in A^\delta + z^\delta$$

and
$$\sum_{\ell=N+1}^{n} \tilde{\alpha}_\ell |x_\ell| \le M\delta \quad \text{for any } x \in A^\delta.$$

Having these bounds we obtain

$$
\begin{aligned}
\frac{\lambda_n(A^\delta + z^\delta)}{\lambda_n(A^\delta)}
&= \frac{\int_{P_n A^\delta + P_n z^\delta} \exp\left(-\sum_{\ell=1}^n \alpha_\ell |u_\ell|\right) du}{\int_{P_n A^\delta} \exp\left(-\sum_{\ell=1}^n \alpha_\ell |u_\ell|\right) du} \\
&\le \frac{e^{-M(\frac{\kappa}{4}-\delta)} \int_{P_n A^\delta + P_n z^\delta} \exp\left(-\sum_{\ell=1}^N \alpha_\ell |u_\ell| - \sum_{\ell=N+1}^n (\alpha_\ell - M\tilde{\alpha}_\ell)|u_\ell|\right) du}{e^{-M\delta} \int_{P_n A^\delta} \exp\left(-\sum_{\ell=1}^N \alpha_\ell |u_\ell| - \sum_{\ell=N+1}^n (\alpha_\ell - M\tilde{\alpha}_\ell)|u_\ell|\right) du} \\
&\le e^{-M(\frac{\kappa}{4}-2\delta)},
\end{aligned}
$$

where we used the log-concavity of the integrands on the last line. Since κ is fixed, M is arbitrary and δ decreases, the result follows for λ_n. In a same way as the previous two lemmas an inequality similar to (35) yields the result. $\qquad\square$ $\qquad\square$

Proof of proposition 4.4.

i) Without losing any generality we assume that the diameter of A is 1. The function $z \mapsto \mu^y(A^\delta + z)$ for fixed $\delta > 0$ is bounded from $B_1^t(\mathbb{T}^d)$ to $[0,1]$. Let $\{z_j\}_{j=1}^\infty \subset B_1^t(\mathbb{T}^d)$ denote the maximizing sequence such that

$$(37) \qquad \sup_{z \in X} \mu^y(A^\delta + z) = \lim_{j \to \infty} \mu^y(A^\delta + z_j).$$

Suppose the sequence z_j is unbounded. Since $\Phi \ge 0$ we have for any $z \in B_1^t(\mathbb{T}^d)$ that

$$\mu^y(A^\delta + z) = \frac{1}{Z} \int_{A^\delta+z} e^{-\Phi(u)} d\lambda(u) \le \frac{1}{Z} \int_{A^\delta+z} d\lambda(u) \le \frac{1}{Z} e^{-\frac{1}{2}\|z\|_X + \delta} \mu^y(A^\delta)$$

by lemma 6.2. Now as $\|z_j\|_{B_1^t} \to \infty$, we have $\mu^y(A^\delta + z_j) \to 0$ which yields a contradiction. Therefore, the sequence z_j must be bounded.

Next, by the Banach–Alaoglu theorem there exists a limit $w \in B_1^t(\mathbb{T}^d)$ in the weak-* topology. In particular, we have $P_n z_j \to P_n w$ in $\mathbb{R}^n$ for any $n > 0$. Let us write $\mu_n^y(A) = \mu^y \circ P_n^{-1}(P_n A)$ for any $A \in \mathcal{B}(B_1^t(\mathbb{T}^d))$. By weak convergence of μ_n^y to μ^y for any $\epsilon > 0$ there exists $N > 0$ such that for $n > N$ we have

$$\mu_n^y(A^\delta + w) = \mu_n^y(A^\delta + w) \le \mu^y(A^\delta + w) + \epsilon,$$

since A^δ is a set of continuity for μ^y by lemma 6.1, i.e. $\mu^y(\partial A^\delta) = 0$. In consequence, we have

$$
\begin{aligned}
\mu^y(A^\delta &+ z_j) - \mu^y(A^\delta + w) \\
&\le \mu_n^y(A^\delta + z_j) - \mu_n^y(A^\delta + w) + \epsilon \\
&\longrightarrow \epsilon \quad \text{as } j \to \infty
\end{aligned}
$$

(38)

where the convergence to ϵ in (38) appears since λ_n is non-degenerate. Since $\epsilon > 0$ was arbitrary, we obtain

$$\lim_{j \to \infty} \mu^y(A^\delta + z_j) \le \mu^y(A^\delta + w)$$

and by (37) the supremum must be attained at $z^\delta = w \in B_1^t(\mathbb{T}^d)$.

ii) Since $\Phi(u) \geq 0$ and $\mathcal{G}$ is locally Lipschitz continuous, for any $\delta \leq 1$ we have

$$(39) \qquad 1 \leq \frac{\mu(B^\delta(z^\delta))}{\mu(B^\delta(0))} \leq \frac{\lambda(B^\delta(z^\delta))}{\mathrm{e}^{-L}\lambda(B^\delta(0))}$$

with a constant L independent of δ. Suppose that $\{z^\delta\}$ is not bounded in X. Then by lemma 6.2, for any $\epsilon > 0$, there exists δ small enough such that

$$\frac{\mu(B^\delta(z^\delta))}{\mu(B^\delta(0))} \leq \frac{\lambda(B^\delta(z^\delta))}{\mathrm{e}^{-L}\lambda(B^\delta(0))} < \epsilon.$$

This contradicts the first inequality in (39). Hence $\{z^\delta\}$ is bounded in $B_1^t(\mathbb{T}^d)$ for any $t < s - d$ and there exists $\bar{z} \in B_1^t(\mathbb{T}^d)$ and a subsequence (also denoted by) $\{z^\delta\} \subset B_1^t(\mathbb{T}^d)$ such that z^δ converges to $\bar{z}$ in the weak*-topology. Now lemma 6.3 and 6.4 together with (39) imply that $\bar{z} \in B_1^s(\mathbb{T}^d)$ and $z^\delta \to \bar{z}$ in $B_1^t(\mathbb{T}^d)$.

iii) We first note that by local Lipschitz continuity of Φ, there exists a constant L depending on $\|\bar{z}\|_{B_1^t}$ such that

$$\frac{\mu^y(A^\delta + z^\delta)}{\mu^y(A^\delta + \bar{z})} \leq \mathrm{e}^{L\delta}\mathrm{e}^{-\Phi(z^\delta) + \Phi(\bar{z})}\frac{\lambda(A^\delta + z^\delta)}{\lambda(A^\delta + \bar{z})},$$

and therefore, since Φ is continuous on $B_1^t(\mathbb{T}^d)$ and $z^\delta \to \bar{z}$ in $B_1^t(\mathbb{T}^d)$, we have

$$\limsup_{\delta \to 0} \frac{\mu^y(A^\delta + z^\delta)}{\mu^y(A^\delta + \bar{z})} \leq \limsup_{\delta \to 0} \frac{\lambda(A^\delta + z^\delta)}{\lambda(A^\delta + \bar{z})}.$$

Now consider a sequence $\{w^j\}_{j=1}^\infty \subset B_1^{s+1}(\mathbb{T}^d)$ with $w^j \to \bar{z}$ in $B_1^s(\mathbb{T}^d)$ as $j \to \infty$. Then we have

$$
\begin{aligned}
\frac{\lambda(A^\delta + z^\delta)}{\lambda(A^\delta + \bar{z})} &= \frac{\int_{A^\delta + z^\delta - w^j} \lim_{N \to \infty} \exp\left(\sum_{\ell=1}^N -\alpha_\ell |w_\ell^j - u_\ell| + \alpha_\ell |u_\ell|\right) \lambda(du)}{\exp(-\|\bar{z}\|_{B_1^s})\int_{A^\delta} \lim_{N \to \infty} \exp\left(\sum_{\ell=1}^N -\alpha_\ell |\bar{z}_\ell - u_\ell| + \alpha_\ell |u_\ell| + \alpha_\ell |\bar{z}_\ell|\right) \lambda(du)} \\
&\leq \frac{\sup_{u \in A^\delta + z^\delta - w^j} \exp\left(\sum_{\ell=1}^\infty -\alpha_\ell |w_\ell^j - u_\ell| + \alpha_\ell |u_\ell|\right)}{\exp(-\|\bar{z}\|_{B_1^s})} \cdot \frac{\lambda(A^\delta + z^\delta - w^j)}{\lambda(A^\delta)} \\
&\leq \frac{\sup_{u \in A^\delta + z^\delta - w^j} \exp\left(\sum_{\ell=1}^\infty -\alpha_\ell |w_\ell^j - u_\ell| + \alpha_\ell |u_\ell|\right)}{\exp(-\|\bar{z}\|_{B_1^s})}.
\end{aligned}
$$

This by lemma 3.6 implies that

$$\limsup_{\delta \to 0} \frac{\lambda(A^\delta + z^\delta)}{\lambda(A^\delta + \bar{z})} \leq \frac{\mathrm{e}^{-\|\bar{z}\|_{B_1^s} + \|\bar{z} - w^j\|_{B_1^s}}}{\mathrm{e}^{-\|\bar{z}\|_{B_1^s}}} = \mathrm{e}^{\|\bar{z} - w^j\|_{B_1^s}}$$

and then letting $j \to \infty$ in the right-hand side we get

$$\limsup_{\delta \to 0} \frac{\lambda(A^\delta + z^\delta)}{\lambda(A^\delta + \bar{z})} \leq 1, \quad \text{hence} \quad \limsup_{\delta \to 0} \frac{\mu^y(A^\delta + z^\delta)}{\mu^y(A^\delta + \bar{z})} \leq 1.$$

Since by definition of z^δ we have that $\mu^y(A^\delta + z^\delta) \geq \mu^y(A^\delta + \bar{z})$, we get that

$$\liminf_{\delta \to 0} \frac{\mu^y(A^\delta + z^\delta)}{\mu^y(A^\delta + \bar{z})} \geq 1.$$

It follows that

$$\tag{40} \lim_{\delta \to 0} \frac{\mu^y(A^\delta + z^\delta)}{\mu^y(A^\delta + \bar{z})} = 1,$$

and therefore $\bar{z}$ is a strong MAP estimator.

It remains to show that $\bar{z}$ is a minimizer of I. Let $z^* := \arg\min_{u \in B_1^s} I(u)$. Suppose that $\bar{z}$ is not a minimizer so that $I(\bar{z}) > I(z^*)$. We first note that by local Lipschitz continuity of Φ, as before we have that there is an L depending on $\|\bar{z}\|_{B_1^t}$ and $\|z^*\|_{B_1^t}$ such that

$$\frac{\mu^y(A^\delta + \bar{z})}{\mu^y(A^\delta + z^*)} \leq e^{L\delta} e^{-\Phi(\bar{z}) + \Phi(z^*)} \frac{\lambda(A^\delta + \bar{z})}{\lambda(A^\delta + z^*)}.$$

Now consider a sequence $\{w^j\}_{j=1}^\infty \subset B_1^{s+1}(\mathbb{T}^d)$ with $w^j \to \bar{z}$ in $B_1^s(\mathbb{T}^d)$. Then, similar to what we did above, we can show that

$$\limsup_{\delta \to 0} \frac{\mu^y(A^\delta + \bar{z})}{\mu^y(A^\delta + z^*)} \leq e^{-\Phi(\bar{z}) + \Phi(z^*)} \frac{e^{-\|\bar{z}\|_{B_1^s} + \|\bar{z} - w^j\|_{B_1^s}}}{e^{-\|z^*\|_{B_1^s}}} \to e^{-I(\bar{z}) + I(z^*)}$$

as $j \to \infty$. We now note that

$$\frac{\mu^y(A^\delta + z^\delta)}{\mu^y(A^\delta + z^*)} = \frac{\mu^y(A^\delta + z^\delta)}{\mu^y(A^\delta + \bar{z})} \frac{\mu^y(A^\delta + \bar{z})}{\mu^y(A^\delta + z^*)}$$

which by (40) gives

$$\limsup_{\delta \to 0} \frac{\mu^y(A^\delta + z^\delta)}{\mu^y(A^\delta + z^*)} \leq e^{-I(\bar{z}) + I(z^*)} < 1.$$

This contradicts the fact that by definition of z^δ, $\mu^y(A^\delta + z^\delta) \geq \mu^y(A^\delta + z^*)$. $\qquad\square$

Proof of corollary 4.5. This follows from the proof of part (i) of proposition 4.4. $\qquad\square$

Proof of theorem 4.6. Suppose that $\tilde{z}$ is a strong MAP estimator. Any strong MAP estimate is a weak MAP estimate and hence, by theorem 4.3, $\tilde{z}$ is a minimizer of I.

Now let z^* be a minimizer of I. By proposition 4.4 we know that there exists a strong MAP estimate $\bar{z} \in B_1^s(\mathbb{T}^d)$ which also minimises I. Therefore, by proposition 4.2, we have

$$\lim_{\epsilon \to 0} \frac{\mu^y(A^\epsilon + \bar{z})}{\mu^y(A^\epsilon + z^*)} = 1.$$

Let $z^\epsilon = \arg\min_{z \in X} \mu^y(A^\epsilon + z)$. By Definition 2.1, We can write

$$\lim_{\epsilon \to 0} \frac{\mu^y(A^\epsilon + z^*)}{\mu^y(A^\epsilon + z^\epsilon)} = \lim_{\epsilon \to 0} \frac{\mu^y(A^\epsilon + \bar{z})}{\mu^y(A^\epsilon + z^\epsilon)} \lim_{\epsilon \to 0} \frac{\mu^y(A^\epsilon + z^*)}{\mu^y(A^\epsilon + \bar{z})} = 1.$$

The result follows.   □

Proof of theorem 4.8. Substituting y_j by $\mathcal{G}(u^\dagger) + \xi_j$, we have

$$\arg\min I_n(u) = \arg\min \left\{ \frac{1}{n}\|u\|_{B_1^s} + \left|\Sigma^{-\frac{1}{2}}(\mathcal{G}(u^\dagger) - \mathcal{G}(u))\right|^2 + \frac{2}{n}\sum_{j=1}^n \left\langle \Sigma^{-\frac{1}{2}}(\mathcal{G}(u^\dagger) - \mathcal{G}(u)), \Sigma^{-\frac{1}{2}}\xi_j \right\rangle \right\}.$$

Since u_n is a minimizer of I_n, we can write

$$\frac{1}{n}\|u_n\|_{B_1^s} + \left|\Sigma^{-\frac{1}{2}}(\mathcal{G}(u^\dagger) - \mathcal{G}(u_n))\right|^2$$

$$\leq \frac{1}{n}\|u^\dagger\|_{B_1^s} + \frac{2}{n}\left|\Sigma^{-\frac{1}{2}}(\mathcal{G}(u^\dagger) - \mathcal{G}(u_n))\right| \left|\sum_{j=1}^n \Sigma^{-\frac{1}{2}}\xi_j\right|$$

$$\leq \frac{1}{n}\|u^\dagger\|_{B_1^s} + \frac{1}{2}\left|\Sigma^{-\frac{1}{2}}(\mathcal{G}(u^\dagger) - \mathcal{G}(u_n))\right|^2 + \frac{2}{n^2}\left|\sum_{j=1}^n \Sigma^{-\frac{1}{2}}\xi_j\right|^2,$$

using Young's inequality. Taking the expectation and using the independence of $\{\xi_j\}$, we obtain

(41) $$\mathbb{E}\left|\Sigma^{-\frac{1}{2}}(\mathcal{G}(u^\dagger) - \mathcal{G}(u_n))\right|^2 \to 0 \quad \text{as} \quad n \to \infty.$$

and, by application of the Jensen inequality,

(42) $$\mathbb{E}\|u_n\|_{B_1^s} \leq \|u^\dagger\|_{B_1^s} + \mathbb{E}\left|\Sigma^{-\frac{1}{2}}\xi_1\right|^2 < \infty.$$

First, by (41), $\mathcal{G}(u_n) \to \mathcal{G}(u^\dagger)$ in probability as $n \to \infty$. Therefore there exists a subsequence which satisfies (after labelling by n again)

(43) $$\mathcal{G}(u_n) \to \mathcal{G}(u^\dagger) \quad \text{almost surely.}$$

Now let $\{\psi_\ell\}_{\ell=1}^\infty$ be an r-regular orthonormal wavelet basis for $L^2(\mathbb{R}^d)$ with $r \geq s$. We then, by (42), have

$$\mathbb{E}|\langle u_n, \psi_\ell\rangle| \leq C_\ell \mathbb{E}\|u_n\|_{B_1^s} \leq C_\ell(\|u^\dagger\|_{B_1^s} + K), \quad \text{for all } \ell \in \mathbb{N}.$$

For $\ell = 1$, the above bound implies the existence of $\{u_{n_1(k)}\}_{k\in\mathbb{N}} \subset \{u_n\}_{n\in\mathbb{N}}$ and $\eta_1 \in \mathbb{R}$ such that $\mathbb{E}|\langle u_{n_1(k)}, \psi_1\rangle| \to \eta_1$ as $k \to \infty$. Considering $\ell = 2, 3, \ldots$ successively one can similarly show the existence of $\{u_{n_1(k)}\}_{k\in\mathbb{N}} \supset \{u_{n_2(k)}\}_{k\in\mathbb{N}} \supset \ldots \{u_{n_\ell(k)}\}_{k\in\mathbb{N}} \supset \ldots$ and $\{\eta_\ell\} \in \mathbb{R}^\infty$ such that $\mathbb{E}|\langle u_{n_\ell(k)}, \psi_\ell\rangle| \to \eta_\ell$ for any $\ell \in \mathbb{N}$ as $k \to \infty$. The subsequence $\{u_{n_k(k)}\}_{k\in\mathbb{N}}$ hence satisfies

(44) $$\mathbb{E}|\langle u_{n_k(k)}, \psi_\ell\rangle| \to \eta_\ell, \quad \text{as } k \to \infty \text{ for any } \ell \in \mathbb{N}.$$

We relabel the above subsequence by $\{u_n\}_{n\in\mathbb{N}}$. Let $u^* := \sum_{\ell=1}^\infty \eta_\ell \psi_\ell$. We have, by Jensen's inequality,

$$\begin{aligned}
\sum_{\ell=1}^N \ell^{\frac{s}{d}-\frac{1}{2}}|\eta_\ell| &\leq \lim_{n\to\infty} \mathbb{E}\sum_{\ell=1}^N \ell^{\frac{s}{d}-\frac{1}{2}}|\langle u_n, \psi_\ell\rangle| \\
&\leq \lim_{n\to\infty} \mathbb{E}\|u_n\|_{B_1^s} \\
&\leq \|u^\dagger\|_{B_1^s} + K.
\end{aligned}$$

Hence $u^* \in B_1^s$. Let us now consider the strong convergence of the distribution in a larger space $B_1^{\tilde{s}}$ for $\tilde{s} < s$. Take $v \in B_\infty^{-\tilde{s}}(\mathbb{T}^d)$ and write

$$\begin{aligned}
\mathbb{E}\|u_n - u^*\|_{B_1^{\tilde{s}}} &= \sum_{\ell=1}^\infty \ell^{\frac{\tilde{s}}{d} - \frac{1}{2}} \mathbb{E}|\langle u_n - u^*, \psi_\ell\rangle| \\
&\leq \sum_{\ell=1}^N \ell^{\frac{\tilde{s}}{d} - \frac{1}{2}} \mathbb{E}|\langle u_n - u^*, \psi_\ell\rangle| + N^{-\frac{s-\tilde{s}}{d}} \sum_{\ell=N+1}^\infty \ell^{\frac{s}{d} - \frac{1}{2}} \mathbb{E}|\langle u_n - u^*, \psi_\ell\rangle|,
\end{aligned}$$

since $\mathbb{E}\|u_n\|_{B_1^s}$ is uniformly bounded, and as $\sum_{\ell=1}^\infty \ell^{\frac{\tilde{s}}{d} - \frac{1}{2}} \mathbb{E}|\langle u_n - u^*, \psi_\ell\rangle| < \infty$ we can pass the expectation inside [61, theorem 1.38]. Noting that $\|u^*\|_{B_1^s} + \mathbb{E}\|u_n\|_{B_1^s} \leq C < \infty$, given any $\epsilon > 0$, N can be chosen large enough, independently of n, such that the second term in the last line of the above inequality is bounded by $\epsilon/2$. Having convergence coefficient-wise, there is $M \in \mathbb{N}$ large enough so that for $n > M$ the first term is bounded by $\epsilon/2$ as well. We therefore conclude that u_n converges strongly to u^* in $B_1^{\tilde{s}}(\mathbb{T}^d)$ in probability. This then implies ([38, Lemma 4.2]) the existence of a subsequence, which we label $\{u_n\}_{n=1}^\infty$ again, such that

$$u_n \to u^* \text{ in } B_1^{\tilde{s}}(\mathbb{T}^d) \text{ almost surely and for all } \tilde{s} < s.$$

This is true in particular for $t < s - d$. By continuity of $\mathcal{G}$ in $B_1^t(\mathbb{T}^d)$, we conclude that

$$\mathcal{G}(u_n) \to \mathcal{G}(u^*) \quad \text{in } B_1^t(\mathbb{T}^d) \text{ a.s.}$$

This together with (43) give $\mathcal{G}(u^\dagger) = \mathcal{G}(u^*)$.   □

Proof of corollary 4.9. Since $B_1^s(\mathbb{T}^d)$ is dense in $B_1^t(\mathbb{T}^d)$, for any $\epsilon > 0$, there exists $v \in B_1^s(\mathbb{T}^d)$ such that $\|u^\dagger - v\|_X \leq \epsilon$. As u_n is a minimizer of I_n we have

$$\begin{aligned}
&\frac{1}{n}\|u_n\|_{B_1^s} + \left|\Sigma^{-\frac{1}{2}}(\mathcal{G}(u^\dagger) - \mathcal{G}(u_n))\right|^2 + \frac{2}{n}\sum_{j=1}^n \left\langle \Sigma^{-\frac{1}{2}}(\mathcal{G}(u^\dagger) - \mathcal{G}(u_n)), \Sigma^{-\frac{1}{2}}\xi_j\right\rangle \\
&\leq \frac{1}{n}\|v\|_{B_1^s} + \left|\Sigma^{-\frac{1}{2}}(\mathcal{G}(u^\dagger) - \mathcal{G}(v))\right|^2 + \frac{2}{n}\sum_{j=1}^n \left\langle \Sigma^{-\frac{1}{2}}(\mathcal{G}(u^\dagger) - \mathcal{G}(v)), \Sigma^{-\frac{1}{2}}\xi_j\right\rangle.
\end{aligned}$$

Rearranging and using Young's inequality, similar to the previous proof, then gives

$$\frac{1}{2}\left|\Sigma^{-\frac{1}{2}}(\mathcal{G}(u^\dagger) - \mathcal{G}(u_n))\right|^2 \leq 2\left|\Sigma^{-\frac{1}{2}}(\mathcal{G}(u^\dagger) - \mathcal{G}(v))\right|^2 + \frac{1}{n}\|v\|_{B_1^s} + \frac{3}{n^2}\left(\sum_{j=1}^n \Sigma^{-\frac{1}{2}}\xi_j\right)^2$$

Noting that $\|v\|_{B_1^s} \leq \|u^\dagger\|_{B_1^s} + \epsilon$ and $\left|\Sigma^{-\frac{1}{2}}(\mathcal{G}(u^\dagger) - \mathcal{G}(v))\right| \leq C\epsilon$ by local Lipschitz continuity of $\mathcal{G}$, we have

$$\mathbb{E}\left|\Sigma^{-\frac{1}{2}}(\mathcal{G}(u^\dagger) - \mathcal{G}(u_n))\right|^2 \leq 4C^2\epsilon^2 + \frac{2}{n}\left(\|u^\dagger\|_{B_1^s} + \epsilon + 3\mathbb{E}|\Sigma^{-\frac{1}{2}}\xi_j|^2\right).$$

This implies that

$$\limsup_{n\to\infty} \mathbb{E}\left|\Sigma^{-\frac{1}{2}}(\mathcal{G}(u^\dagger) - \mathcal{G}(u_n))\right|^2 \leq 4C^2\epsilon^2.$$

Since ϵ was arbitrary, we conclude that $\lim_{n\to\infty} \mathbb{E}|\Sigma^{-\frac{1}{2}}(\mathcal{G}(u^\dagger) - \mathcal{G}(u_n))|^2 = 0$, and therefore, $|\Sigma^{-\frac{1}{2}}(\mathcal{G}(u^\dagger) - \mathcal{G}(u_n))| \to 0$ in probability. Hence, there exists a subsequence of $\{\mathcal{G}(u_n)\}$ which converges to $\mathcal{G}(u^\dagger)$ almost surely, giving the result.   □

6.4. *Proofs of results in section 5.* *Proof of theorem 5.8.* The probability distribution π_X on $\mathbb{R}$ has finite Fisher information since

$$(45) \qquad \int_{\mathbb{R}} \frac{\pi'_X(t)^2}{\pi_X(t)} dt = \int_{\mathbb{R}} \pi_X(t) dt = 1.$$

As pointed out above we have $\lambda_\ell(A) = \lambda_1(\alpha_\ell A)$, for any $A \in \mathcal{B}(\mathbb{R})$, where $\alpha_\ell = \ell^{s/d-1/2}$. By proposition 5.7 we have

$$D(\lambda) = \left\{ y \in \mathbb{R}^\infty \mid \sum_{\ell=1}^\infty \alpha_\ell^2 y_\ell^2 < \infty \right\} = \left\{ y \in \mathbb{R}^\infty \mid \sum_{\ell=1}^\infty \ell^{\frac{2s}{d}-1} y_\ell^2 < \infty \right\}.$$

and, consequently, $D(\lambda) = B_2^{s'}(\mathbb{T}^d)$ for $s' = s - \frac{d}{2}$. The rest follows from theorem 5.6. $\qquad \square$

REFERENCES

[1] S. Agapiou, S. Larsson, and A. M. Stuart. Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems. *Stochastic Processes and their Applications*, 2013.

[2] S. Agapiou, A. M. Stuart, and Y. X. Zhang. Bayesian posterior contraction rates for linear severely ill-posed inverse problems. *J. Inverse Ill-Posed Probl.*, 22(3):297–321, 2014.

[3] T. W. Anderson. The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proceedings of the American Mathematical Society*, 6(2):170–176, 1955.

[4] J. M. Bardsley, D. Calvetti, and E. Somersalo. Hierarchical regularization for edge-preserving reconstruction of pet images. *Inverse Problems*, 26(3):035010, 2010.

[5] V. I. Bogachev. *Gaussian measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1998.

[6] V. I. Bogachev. *Differentiable measures and the Malliavin calculus*, volume 164 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2010.

[7] V. I. Bogachev and O. G. Smolyanov. Analytic properties of infinite-dimensional distributions. *Uspekhi Mat. Nauk*, 45(3(273)):3–83, 221, 1990.

[8] C. Borell. Convex measures on locally convex spaces. *Arkiv för Matematik*, 12(1):239–252, 1974.

[9] T. Bui-Thanh and O. Ghattas. A scalable algorithm for MAP estimators in Bayesian inverse problems with Besov priors. *Inverse Problems and Imaging*, 9(1):27–53, 2015.

[10] D. Calvetti and E. Somersalo. A gaussian hypermodel to recover blocky objects. *Inverse problems*, 23(2):733, 2007.

[11] D. Calvetti and E. Somersalo. Hypermodels in the bayesian imaging framework. *Inverse Problems*, 24(3):034013, 2008.

[12] D. Calvetti, E. Somersalo, and R. Spies. Variable order smoothness priors for ill-posed inverse problems. *Mathematics of Computation*, 84(294):1753–1773, 2015.

[13] E. Candes and J. Romberg. l1-magic: Recovery of sparse signals via convex programming. *URL: www. acm. caltech. edu/l1magic/downloads/l1magic. pdf*, 4:14, 2005.

[14] E. J. Candes and D. L. Donoho. Curvelets: A surprisingly effective nonadaptive representation for objects with edges. Technical report, Stanford Univ Ca Dept of Statistics, 2000.

[15] G. Da Prato. *An introduction to infinite-dimensional analysis*. Springer Science & Business Media, 2006.

[16] M. Dashti and A.M.Stuart. The Bayesian Approach to Inverse Problems. In R. Ghanem, D. Higdon, and H. Owhadi, editors, *Handbook of Uncertainty Quantification*. Springer, 2015.

[17] M. Dashti, S. Harris, and A. Stuart. Besov priors for Bayesian inverse problems. *Inverse Probl. Imaging*, 6(2):183–200, 2012.

[18] M. Dashti, K. J. H. Law, A. M. Stuart, and J. Voss. MAP estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Problems*, 29(9):095017, 27, 2013.

[19] I. Daubechies. *Ten lectures on wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[20] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on pure and applied mathematics*, 57(11):1413–1457, 2004.

[21] S. Dharmadhikari and K. Joag-Dev. *Unimodality, convexity, and applications*. Elsevier, 1988.

[22] D. C. Dobson and F. Santosa. An image-enhancement technique for electrical impedance tomography. *Inverse problems*, 10(2):317, 1994.

[23] D. C. Dobson and F. Santosa. Recovery of blocky images from noisy and blurred data. *SIAM Journal on Applied Mathematics*, 56(4):1181–1198, 1996.

[24] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via 1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003.

[25] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47(7):2845–2862, 2001.

[26] M. Dunlop and A. Stuart. MAP estimators for piecewise continuous inversion. *Inverse Problems*, 32:105003, 2016.

[27] D. Dürr and A. Bach. The Onsager-Machlup function as Lagrangian for the most probable path of a diffusion process. *Communications in Mathematical Physics*, 60(2):153–170, 1978.

[28] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

[29] J. N. Franklin. Well-posed stochastic extensions of ill-posed linear problems. *Journal of Mathematical Analysis and Applications*, 31(3):682–716, 1970.

[30] D. Gerth and R. Ramlau. A stochastic convergence analysis for Tikhonov regularization with sparsity constraints. *Inverse Problems*, 30(5):055009, 24, 2014.

[31] T. Helin. On infinite-dimensional hierarchical probability models in statistical inverse problems. *Inverse Probl. Imaging*, 3(4):567–597, 2009.

[32] T. Helin and M. Burger. Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems. *Inverse Problems*, 31(8):085009, 2015.

[33] T. Helin and M. Lassas. Hierarchical models in statistical inverse problems and the Mumford-Shah functional. *Inverse Problems*, 27(1):015008, 32, 2011.

[34] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*, volume 24. Elsevier, 2014.

[35] J. Jia, J. Peng, and J. Gao. Bayesian approach to inverse problems for functions with a variable-index Besov prior. *Inverse Problems*, 32(8):085006, 2016.

[36] I. M. Johnstone. Minimax bayes, asymptotic minimax and sparse wavelet priors. In *Statistical Decision Theory and Related Topics V*, pages 303–326. Springer, 1994.

[37] J. Kaipio and E. Somersalo. *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media, 2006.

[38] O. Kallenberg. *Probability and Its Applications*. Springer-Verlag, New York, second edition, 2002.

[39] H. Kekkonen, M. Lassas, and S. Siltanen. Posterior consistency and convergence rates for bayesian inversion with hypoelliptic operators. *Inverse Problems*, 32:085005, 2016.

[40] B. T. Knapik and J.-B. Salomond. A general approach to posterior contraction in nonparametric inverse problems. *arXiv preprint arXiv:1407.0335*, 2014.

[41] B. T. Knapik, B. T. Szabó, A. W. Vaart, and J. H. Zanten. Bayes procedures for adaptive inference in inverse problems for the white noise model. *Probability Theory and Related Fields*, 164(3-4):771–813, 2016.

[42] B. T. Knapik, A. W. van Der Vaart, and J. H. van Zanten. Bayesian inverse problems with Gaussian priors. *The Annals of Statistics*, 39(5):2626–2657, 2011.

[43] B. T. Knapik, A. W. van der Vaart, and J. H. van Zanten. Bayesian recovery of the initial condition for the heat equation. *Communications in Statistics-Theory and Methods*, 42(7):1294–1313, 2013.

[44] V. Kolehmainen, M. Lassas, K. Niinimäki, and S. Siltanen. Sparsity-promoting Bayesian inversion. *Inverse Problems*, 28(2):025005, 2012.

[45] V. Kolehmainen, S. Siltanen, S. Järvenpää, J. Kaipio, P. Koistinen, M. Lassas, J. Pirttilä, and E. Somersalo. Statistical inversion for medical X-ray tomography with few radiographs: II. Application to dental radiology. *Physics in Medicine and Biology*, 48(10):1465, 2003.

[46] S. Lasanen. Non-gaussian statistical inverse problems. part i: Posterior distributions. *Inverse Problems & Imaging*, 6(2), 2012.

[47] M. Lassas, E. Saksman, and S. Siltanen. Discretization-invariant Bayesian inversion and Besov space priors. *Inverse Probl. Imaging*, 3(1):87–122, 2009.

[48] M. Lassas and S. Siltanen. Can one use total variation prior for edge-preserving Bayesian inversion? *Inverse Problems*, 20(5):1537–1563, 2004.

[49] M. S. Lehtinen, L. Paivarinta, and E. Somersalo. Linear inverse problems for generalised random variables. *Inverse Problems*, 5(4):599, 1989.

[50] D. Leporini and J.-C. Pesquet. Bayesian wavelet denoising: Besov priors and non-Gaussian noises. *Signal processing*, 81(1):55–67, 2001.

[51] H. C. Lie and T. J. Sullivan. Equivalence of weak and strong modes of measures on topological vector spaces. *arXiv preprint arXiv:1708.02516*, 2017.

[52] I. Loris, G. Nolet, I. Daubechies, and F. Dahlen. Tomographic inversion using 1-norm regularization of wavelet coefficients. *Geophysical Journal International*, 170(1):359–370, 2007.

[53] A. Mandelbaum. Linear estimators and measurable linear transformations on a Hilbert space. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 65(3):385–397, 1984.

[54] O. Papaspiliopoulos, Y. Pokern, G. O. Roberts, and A. M. Stuart. Nonparametric estimation of diffusions: a differential equations approach. *Biometrika*, 99(3):511–531, 2012.

[55] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on pattern analysis and machine intelligence*, 12(7):629–639, 1990.

[56] Y. Pokern, A. M. Stuart, and J. H. van Zanten. Posterior consistency via precision operators for Bayesian nonparametric drift estimation in SDEs. *Stochastic Process. Appl.*, 123(2):603–628, 2013.

[57] A. Pulkkinen, B. T. Cox, S. R. Arridge, J. P. Kaipio, and T. Tarvainen. Bayesian parameter estimation in spectral quantitative photoacoustic tomography. *Proc. SPIE*, 9708:97081G–97081G–8, 2016.

[58] M. Rantala, S. Vanska, S. Jarvenpaa, M. Kalke, M. Lassas, J. Moberg, and S. Siltanen. Wavelet-based reconstruction for limited-angle X-ray tomography. *IEEE transactions on medical imaging*, 25(2):210–217, 2006.

[59] K. Ray. Bayesian inverse problems with non-conjugate priors. *Electronic Journal of Statistics*, 7:2516–2549, 2013.

[60] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[61] W. Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.

[62] L. Shepp. Distinguishing a sequence of random variables from a translate of itself. *The Annals of Mathematical Statistics*, 36(4):1107–1112, 1965.

[63] S. Siltanen, V. Kolehmainen, S. Järvenpää, J. Kaipio, P. Koistinen, M. Lassas, J. Pirttilä, and E. Somersalo. Statistical inversion for medical X-ray tomography with few radiographs: I. General theory. *Physics in medicine and biology*, 48(10):1437, 2003.

[64] A. S. Stern, D. L. Donoho, and J. C. Hoch. Nmr data processing using iterative thresholding and minimum l 1-norm reconstruction. *Journal of magnetic resonance*, 188(2):295–300, 2007.

[65] D. Strong and T. Chan. Edge-preserving and scale-dependent properties of total variation regularization. *Inverse problems*, 19(6):S165, 2003.

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

[67] T. Sullivan. Well-posed Bayesian inverse problems and heavy-tailed stable quasi-Banach space priors. *arXiv preprint arXiv:1605.05898*, 2016.

[68] B. T. Szabó, A. W. van der Vaart, and J. H. van Zanten. Empirical Bayes scaling of Gaussian priors in the white noise model. *Electronic Journal of Statistics*, 7:991–1018, 2013.

[69] A. Tarantola. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005.

[70] J. Tick, A. Pulkkinen, and T. Tarvainen. Image reconstruction with uncertainty quantification in photoacoustic tomography. *The Journal of the Acoustical Society of America*, 139(4):1951–1961, 2016.

[71] J. van Waaij and H. van Zanten. Gaussian process methods for one-dimensional diffusions: optimal rates and adaptation. *Electronic Journal of Statistics*, 10(1):628–645, 2016.

[72] S. J. Vollmer. Posterior consistency for Bayesian inverse problems through stability and regression results. *Inverse Problems*, 29(12):125011, 2013.

S. Agapiou
Department of Mathematics and Statistics
1 University Avenue
2109 Nicosia
Cyprus
E-mail: agapiou.sergios@ucy.ac.cy

M. Dashti
Department of Mathematics
University of Sussex
BN1 5DJ Brighton
United Kingdom
E-mail: m.dashti@sussex.ac.uk

M. Burger
Institute for Computational
and Applied Mathematics
Westfälische Wilhelms-Universität Münster
and Cells in Motion Cluster of Excellence
Münster
Germany
E-mail: martin.burger@wwu.de

T. Helin
Department of Mathematics and Statistics
University of Helsinki
Gustaf Hällströmin katu 2b
FI-00014 Helsinki
Finland
E-mail: tapio.helin@helsinki.fi