CrossMark

Bayes procedures for adaptive inference in inverse problems for the white noise model

B. T. Knapik · B. T. Szabó · A. W. van der Vaart · J. H. van Zanten

Abstract We study empirical and hierarchical Bayes approaches to the problem of estimating an infinite-dimensional parameter in mildly ill-posed inverse problems. We consider a class of prior distributions indexed by a hyperparameter that quantifies regularity. We prove that both methods we consider succeed in automatically selecting this parameter optimally, resulting in optimal convergence rates for truths with Sobolev or analytic "smoothness", without using knowledge about this regularity. Both methods are illustrated by simulation examples.

Keywords Adaptation · Empirical Bayes · Hierarchical Bayes · Posterior distribution · Gaussian prior · Rate of convergence · Nonparametric inverse problems

B. T. Knapik (✉)
Department of Mathematics, VU University Amsterdam, Amsterdam, The Netherlands
e-mail: b.t.knapik@vu.nl

B. T. Szabó
Department of Stochastics, Budapest University of Technology and Economics, Budapest, Hungary
e-mail: bszabo@math.bme.hu

A. W. van der Vaart
Mathematical Institute, Leiden University, Leiden, The Netherlands
e-mail: avdvaart@math.leidenuniv.nl

J. H. van Zanten
Korteweg–de Vries Institute for Mathematics, University of Amsterdam,
Amsterdam, The Netherlands
e-mail: hvzanten@uva.nl

Mathematics Subject Classification 62G05 · 62C12 · 62G20

1 Introduction

In recent years, Bayesian approaches have become more and more common in dealing with nonparametric statistical inverse problems. Such problems arise in many fields of applied science, including geophysics, genomics, medical image analysis and astronomy, to mention but a few. In nonparametric inverse problems some form of regularization is usually needed in order to estimate the (typically functional) parameter of interest. One possible explanation of the increasing popularity of Bayesian methods is the fact that assigning a prior distribution to an unknown functional parameter is a natural way of specifying a degree of regularization. Probably at least as important is the fact that various computational methods exist to carry out the inference in practice, including MCMC methods and approximate methods like expectation propagation, Laplace approximations and approximate Bayesian computation. A third important aspect that appeals to users of Bayes methods is that an implementation of a Bayesian procedure typically produces not only an estimate of the unknown quantity of interest (usually a posterior mean or mode), but also a large number of samples from the whole posterior distribution. These can then be used to report a credible set, i.e. a set of parameter values that receives a large fixed fraction of the posterior mass, that serves as a quantification of the uncertainty in the estimate. Some examples of papers using Bayesian methods in nonparametric inverse problems in various applied settings include [3,16,24,27,28]. The paper [34] provides a nice overview and many additional references.

Work on the fundamental properties of Bayes procedures for nonparametric inverse problems, like consistency, (optimal) convergence rates, etcetera, has only started to appear recently. The few papers in this area include [1,14,22,23,30]. Other papers addressing frequentist properties of Bayes procedures for different, but related inverse problems include [21] and [15]. This is in sharp contrast with the work on frequentist methodology, which is quite well developed. See for instance the overviews given by Cavalier [8,9].

Our focus in this paper is on the ability of Bayesian methods to achieve adaptive, rate-optimal inference in so-called mildly ill-posed nonparametric inverse problems (in the terminology of, e.g., [8]). Nonparametric priors typically involve one or more tuning parameters, or hyper-parameters, that determine the degree of regularization. In practice there is widespread use of empirical Bayes and full, hierarchical Bayes methods to automatically select the appropriate values of such parameters. These methods are generally considered to be preferable to methods that use only a single, fixed value of the hyper-parameters. In the inverse problem setting it is known from the recent paper [22] that using a fixed prior can indeed be undesirable, since it can lead to convergence rates that are sub-optimal, unless by chance the statistician has selected a prior that captures the fine properties of the unknown parameter (like its degree of smoothness, if it is a function). Theoretical work that supports the preference for empirical or hierarchical Bayes methods does not exist at the present time however. It has until now been unknown whether these approaches can indeed robustify a

procedure against prior mismatch. In this paper we answer this question in the affirmative. We show that empirical and hierarchical Bayes methods can lead to adaptive, rate-optimal procedures in the context of nonparametric inverse problems, provided they are properly constructed.

We study this problem in the context of the canonical signal-in-white-noise model, or, equivalently, the infinite-dimensional normal mean model. Using singular value decompositions many nonparametric, linear inverse problems can be cast in this form (e.g. [9,22]). Specifically, we assume that we observe a sequence of noisy coefficients $Y = (Y_1, Y_2, \ldots)$ satisfying

$$Y_i = \kappa_i \mu_i + \frac{1}{\sqrt{n}} Z_i, \qquad i = 1, 2 \ldots, \tag{1.1}$$

where $Z_1, Z_2, \ldots$ are independent, standard normal random variables, $\mu = (\mu_1, \mu_2, \ldots) \in \ell_2$ is the infinite-dimensional parameter of interest, and (κ_i) is a known sequence that may converge to 0 as $i \to \infty$, which complicates the inference. We suppose the problem is mildly ill-posed of order $p \geq 0$, in the sense that

$$C^{-1} i^{-p} \leq \kappa_i \leq C i^{-p}, \qquad i = 1, 2 \ldots, \tag{1.2}$$

for some $C \geq 1$. Minimax lower bounds for the rate of convergence of estimators for μ are well known in this setting. For instance, the lower bound over Sobolev balls of regularity $\beta > 0$ is given by $n^{-\beta/(1+2\beta+2p)}$ and over certain "analytic balls" the lower bound is of the order $n^{-1/2} \log^{1/2+p} n$ (see [8]). There are several regularization methods which attain these rates, including classical Tikhonov regularization and Bayes procedures with Gaussian priors.

Many of the older existing methods for nonparametric inverse problems are not adaptive, in the sense that they rely on knowledge of the regularity (e.g. in Sobolev sense) of the unknown parameter of interest to select the appropriate regularization. This also holds for the Bayesian approach with fixed Gaussian priors. Early papers on the direct problem, i.e. the case $p = 0$ in (1.2), include [33,41]. The more recent papers [22] and [1] study the inverse problem case, but also obtain non-adaptive results only. In the last decade however, several methods have been developed in frequentist literature that achieve the minimax convergence rate without knowledge of the regularity of the truth. This development parallels the earlier work on adaptive methods for the direct nonparametric problem to some extent, although the inverse case is technically usually more demanding. The adaptive methods typically involve a data-driven choice of a tuning parameter in order to automatically achieve an optimal bias-variance trade-off, as in Lepski's method for instance.

For nonparametric inverse problems, the construction of an adaptive estimator based on a properly penalized blockwise Stein's rule has been studied in [12], cf. also [6]. This estimator is adaptive both over Sobolev and analytic scales. In [10] the data-driven choice of the regularizing parameters is based on unbiased risk estimation. The authors consider projection estimators and derive the corresponding oracle inequalities. For μ in the Sobolev scale they obtain asymptotically sharp adaptation in a minimax sense, whereas for μ in analytic scale, their rate is optimal up to a logarithmic term. Yet

another approach to adaptation in inverse problems is the risk hull method studied in [11]. In this paper the authors consider spectral cut-off estimators and provide oracle inequalities. An extension of their approach is presented in [25]. The link between the penalized blockwise Stein's rule and the risk hull method is presented in [26].

Adaptation properties of Bayes procedures for mildly ill-posed nonparametric inverse problems have until now not been studied in the literature, with an exception of [15] in a different setting. Results in our setting are only available for the direct problem, i.e. the case that $\kappa_i = 1$ for every i, or, equivalently, $p = 0$ in (1.2). In the paper [5] it is shown that in this case adaptive Bayesian inference is possible using a hierarchical, conditionally Gaussian prior, while in [35] partially adaptation is shown using Gaussian priors with scale parameter determined by an empirical Bayes method. Other recent papers also exhibit priors that yield rate-adaptive procedures in the direct signal-in-white-noise problem (see for instance [2,13,32,38]), but it is important to note that these papers use general theorems on contraction rates for posterior distributions (as given in [18] for instance) that are *not* suitable to deal with the truly ill-posed case in which $k_i \to 0$ as $i \to \infty$. The reason is that if these general theorems are applied in the inverse case, we only obtain convergence rates relative to the (squared) norm $\mu \mapsto \sum \kappa_i^2 \mu_i^2$, which is not very interesting. Obtaining rates relative to the ℓ_2-norm is much more involved and requires a different approach. Extending the testing approach of [17,18] would be one possibility, cf. the recent work of [30], although it seems difficult to obtain sharp results in this manner. In this paper we follow a more pragmatic approach, relying on partly explicit computations in a relatively tractable setting.

To obtain rate-adaptive Bayes procedures for the model (1.1) we consider a family $(\Pi_\alpha : \alpha > 0)$ of Gaussian priors for the parameter μ. These priors are indexed by a parameter $\alpha > 0$ which quantifies the "regularity" of the prior Π_α (details in Sect. 2). Instead of choosing a fixed value for α (which is the approach studied in [22]) we view it as a tuning-, or hyper-parameter and consider two different methods for selecting it in a data-driven manner. The approach typically preferred by Bayesian statisticians is to endow the hyper-parameter with a prior distribution itself. This results in a full, hierarchical Bayes procedure. The paper [5] follows the same approach in the direct problem. We prove that under a mild assumption on the hyper-prior on α, we obtain an adaptive procedure for the inverse problem using the hierarchical prior. Optimal convergence rates are obtained (up to lower order factors), uniformly over Sobolev and analytic scales. For tractability, the priors Π_α that we use put independent, Gaussian prior weights on the coefficients μ_i in (1.1). Extensions to more general priors, including non-Gaussian densities or priors that are not exactly diagonal (as in [30] for instance) should be possible, but would require considerable additional technical work.

A second approach we study consists in first "estimating" α from the data and then substituting the estimator $\hat{\alpha}_n$ for α in the posterior distribution for μ corresponding to the prior Π_α. This empirical Bayes procedure is not really Bayesian in the strict sense of the word. However, for computational reasons empirical Bayes methods of this type are widely used in practice, making it relevant to study their theoretical performance. Rigorous results about the asymptotic behavior of empirical Bayes selectors of hyper-parameters in infinite-dimensional problems only exist for a limited number of special

problems, see e.g. [4,19,20,35,40]. In this paper we prove that the likelihood-based empirical Bayes method that we propose has the same desirable adaptation and rate-optimality properties in nonparametric inverse problems as the hierarchical Bayes approach.

The estimator $\hat\alpha_n$ for α that we propose is the commonly used likelihood-based empirical Bayes estimator for the hyper-parameter. Concretely, it is the maximum likelihood estimator for α in the model in which the data Y is generated by first drawing μ from Π_α and then generating $Y = (Y_1, Y_2, \ldots)$ according to (1.1), i.e.

$$\mu|\alpha \sim \Pi_\alpha, \quad \text{and} \quad Y|(\mu, \alpha) \sim \bigotimes_{i=1}^{\infty} N\left(\kappa_i \mu_i, \frac{1}{n}\right). \tag{1.3}$$

A crucial element in the proof of the adaptation properties of both procedures we consider is understanding the asymptotic behavior of $\hat\alpha_n$. In contrast to the typical situation in parametric models (see [29]) this turns out to be rather delicate, since the likelihood for α can have complicated behavior. We are able however to derive deterministic asymptotic lower and upper bounds for $\hat\alpha_n$. In general these depend on the true parameter μ_0 in a complicated way. It appears that in general the difference between these bounds does not become asymptotically negligible, but it can be shown that any value between the bounds gives the correct bias-variance trade-off for the class containing the particular μ_0, whence adaptive minimaxity arises.

In the special case that the true parameter has regular behavior of the form $\mu_{0,i} \asymp i^{-1/2-\beta}$ for some $\beta > 0$, both bounds tends to β and hence $\hat\alpha_n$ is essentially a consistent estimator for β (see Lemma 1). This means that in this case the estimator $\hat\alpha_n$ correctly "estimates the regularity" of the true parameter (see [4] for work in a similar direction). Since the typical models used to define "minimax adaptation" only impose upper bounds on the parameters (e.g. $\mu_{0,i} \lesssim i^{-1/2-\beta}$ or an integrated version of this), in general the "regularity" of a parameter is an ill-defined concept. The value $\hat\alpha_n$ may then have complicated behaviour, but it still gives minimaxity over the class.

Our priors Π_α model the coordinates μ_i as independent $N(0, i^{-1-2\alpha})$ variables. This is flexible enough to adapt to the full scale of Sobolev spaces, and also to models of supersmooth parameters (up to logarithmic factors). In [35] it was shown (only for the direct problem) that priors of the form $N(0, \tau^2 i^{-1-2\alpha})$ for a fixed exponent α and adaptation to scale τ achieves adaptive minimaxity over Sobolev classes only in a limited range, dependent on α.

The remainder of the paper is organized as follows. In Sect. 2 we first describe the empirical and hierarchical Bayes procedures in detail. Then we present a theorem on the asymptotic behavior of estimator $\hat\alpha_n$ for the hyper-parameter, followed by two results on the adaptation and rate of contraction of the empirical and hierarchical Bayes posteriors over Sobolev and analytic scales. These results all concern global ℓ_2-loss. In Sect. 2.3 we briefly comment on rates relative to other losses. Specifically we discuss contraction rates of marginal posteriors for linear functionals of the parameter μ. We conjecture that the procedures that we prove to be adaptive and rate-optimal for global ℓ_2-loss, will be sub-optimal for estimating certain unbounded linear functionals. A detailed study of this issue is outside the scope of the present paper. The empirical and

hierarchical Bayes approaches are illustrated numerically in Sect. 3. We apply them to simulated data from an inverse signal-in-white-noise problem, where the problem is to recover a signal from a noisy observation of its primitive and also another example with a smaller degree of ill-posedness. Proofs of the main results are presented in Sects. 4–7. Some auxiliary lemmas are collected in Sect. 8.

1.1 Notation

For $\beta, \gamma \geq 0$, the Sobolev norm $\|\mu\|_\beta$, the analytic norm $\|\mu\|_{A^\gamma}$ and the ℓ_2-norm $\|\mu\|$ of an element $\mu \in \ell_2$ are defined by

$$\|\mu\|_\beta^2 = \sum_{i=1}^\infty i^{2\beta} \mu_i^2, \qquad \|\mu\|_{A^\gamma}^2 = \sum_{i=1}^\infty e^{2\gamma i} \mu_i^2, \qquad \|\mu\|^2 = \sum_{i=1}^\infty \mu_i^2,$$

and the corresponding Sobolev space by $S^\beta = \{\mu \in \ell_2 : \|\mu\|_\beta < \infty\}$, and the analytic space by $A^\gamma = \{\mu \in \ell_2 : \|\mu\|_{A^\gamma} < \infty\}$.

For two sequences (a_n) and (b_n) of numbers, $a_n \asymp b_n$ means that $|a_n/b_n|$ is bounded away from zero and infinity as $n \to \infty$, $a_n \lesssim b_n$ means that a_n/b_n is bounded, $a_n \sim b_n$ means that $a_n/b_n \to 1$ as $n \to \infty$, and $a_n \ll b_n$ means that $a_n/b_n \to 0$ as $n \to \infty$. For two real numbers a and b, we denote by $a \vee b$ their maximum, and by $a \wedge b$ their minimum.

2 Main results

2.1 Description of the empirical and hierarchical Bayes procedures

We assume that we observe the sequence of noisy coefficients $Y = (Y_1, Y_2, \ldots)$ satisfying (1.1), for $Z_1, Z_2, \ldots$ independent, standard normal random variables, $\mu = (\mu_1, \mu_2, \ldots) \in \ell_2$, and a known sequence (κ_i) satisfying (1.2) for some $p \geq 0$ and $C \geq 1$. We denote the distribution of the sequence Y corresponding to the "true" parameter μ_0 by P_0, and the corresponding expectation by E_0.

For $\alpha > 0$, consider the product prior Π_α on ℓ_2 given by

$$\Pi_\alpha = \bigotimes_{i=1}^\infty N\big(0, i^{-1-2\alpha}\big). \tag{2.1}$$

It is easy to see that this prior is "α-regular", in the sense that for every $\alpha' < \alpha$, it assigns mass 1 to the Sobolev space $S^{\alpha'}$. In [22] it was proved that if for the true parameter μ_0 we have $\mu_0 \in S^\beta$ for $\beta > 0$, then the posterior distribution corresponding to the Gaussian prior Π_α contracts around μ_0 at the optimal rate $n^{-\beta/(1+2\beta+2p)}$ if $\alpha = \beta$. If $\alpha \neq \beta$, only sub-optimal rates are attained in general (cf. [7]). In other words, when using a Gaussian prior with a fixed regularity, optimal convergence rates are obtained if and only if the regularity of the prior and the truth are matched. Since the latter is unknown however, choosing the prior that is optimal from the point of view of convergence rates is typically not possible in practice.

However, the results in [22] indicate that a regular enough prior ($\beta \leq 1 + 2\alpha + 2p$) can be appropriately scaled to attain the optimal rate. This observation in the direct case $p = 0$, led to the study of a data-driven selection of the scaling parameter τ_n in [35] with priors of the form $N(0, \tau^2 i^{-1-2\alpha})$. Already in the direct case ($p = 0$), the performance of the empirical Bayes procedure cuts the range $\beta \leq 1 + 2\alpha$ where the optimal deterministic scaling is possible, into two subregimes. If $\beta < 1/2 + \alpha$, the empirical Bayes leads to the optimal rate. Otherwise, that is when $1/2 + \alpha \leq \beta \leq 1 + 2\alpha$, the performance of the empirical Bayes procedure is strictly worse than the optimal procedure. Therefore, the procedure is suboptimal not only over a wide range of Sobolev classes, but also over certain "analytic balls", e.g., A^γ for all $\gamma > 0$. The same conclusions hold for the hierarchical Bayes procedure.

Therefore, in this paper we fix $\tau \equiv 1$ and consider two data-driven methods for selecting the regularity α of the prior.

The first is a likelihood-based empirical Bayes method, which attempts to estimate the appropriate value of the hyper-parameter α from the data. In the Bayesian setting described by the conditional distributions (1.3), it holds that

$$Y|\alpha \sim \bigotimes_{i=1}^{\infty} N\left(0, i^{-1-2\alpha}\kappa_i^2 + \frac{1}{n}\right).$$

The corresponding log-likelihood for α (relative to an infinite product of $N(0, 1/n)$-distributions) is easily seen to be given by

$$\ell_n(\alpha) = -\frac{1}{2}\sum_{i=1}^{\infty}\left(\log\left(1 + \frac{n}{i^{1+2\alpha}\kappa_i^{-2}}\right) - \frac{n^2}{i^{1+2\alpha}\kappa_i^{-2} + n}Y_i^2\right). \qquad (2.2)$$

The idea is to "estimate" α by the maximizer of ℓ_n. The results ahead (Lemma 1 and Theorem 1) imply that with P_0-probability tending to one, ℓ_n has a global maximum on $[0, \log n]$ if $\mu_{0,i} \neq 0$ for some $i \geq 2$. (In fact, the cited results imply the maximum is attained on the slightly smaller interval $[0, (\log n)/(2\log 2) - 1/2 - p]$). If the latter condition is not satisfied (if $\mu_0 = 0$ for instance), ℓ_n may attain its maximum only at ∞. Therefore, we truncate the maximizer at $\log n$ and define

$$\hat{\alpha}_n = \operatorname*{argmax}_{\alpha \in [0, \log n]} \ell_n(\alpha).$$

The continuity of ℓ_n ensures the argmax exists. If it is not unique, any value may be chosen. We will always assume at least that μ_0 has Sobolev regularity of some order $\beta > 0$. Lemma 1 and Theorem 1 imply that in this case $\hat{\alpha}_n > 0$ with probability tending to 1. An alternative to the truncation of the argmax of ℓ_n at $\log n$ could be to extend the definition of the priors Π_α to include the case $\alpha = \infty$. The prior Π_∞ should then be defined as the product $N(0, 1) \otimes \delta_0 \otimes \delta_0 \otimes \ldots$, with δ_0 the Dirac measure concentrated at 0. However, from a practical perspective it is more convenient to define $\hat{\alpha}_n$ as above.

The empirical Bayes procedure consists in computing the posterior distribution of μ corresponding to a fixed prior Π_α and then substituting $\hat{\alpha}_n$ for α. Under the model described above and the prior (2.1) the coordinates $(\mu_{0,i}, Y_i)$ of the vector (μ_0, Y) are independent, and hence the conditional distribution of μ_0 given Y factorizes over the coordinates as well. The computation of the posterior distribution reduces to countably many posterior computations in conjugate normal models. Therefore (see also [22]) the posterior distribution corresponding to the prior Π_α is given by

$$\Pi_\alpha(\cdot|Y) = \bigotimes_{i=1}^\infty N\left(\frac{n\kappa_i^{-1}}{i^{1+2\alpha}\kappa_i^{-2} + n} Y_i, \frac{\kappa_i^{-2}}{i^{1+2\alpha}\kappa_i^{-2} + n}\right). \tag{2.3}$$

Then the *empirical Bayes posterior* is the random measure $\Pi_{\hat{\alpha}_n}(\cdot|Y)$ defined by

$$\Pi_{\hat{\alpha}_n}(B|Y) = \Pi_\alpha(B|Y)\Big|_{\alpha=\hat{\alpha}_n} \tag{2.4}$$

for measurable subsets $B \subset \ell_2$. Note that the construction of the empirical Bayes posterior does not use information about the regularity of the true parameter. In Theorem 2 below we prove that it contracts around the truth at an optimal rate (up to lower order factors), uniformly over Sobolev and analytic scales.

The second method we consider is a full, hierarchical Bayes approach where we put a prior distribution on the hyper-parameter α. We use a prior on α with a positive Lebesgue density λ on $(0, \infty)$. The full, hierarchical prior for μ is then given by

$$\Pi = \int_0^\infty \lambda(\alpha)\Pi_\alpha \, d\alpha. \tag{2.5}$$

In Theorem 3 below we prove that under mild assumptions on the prior density λ, the corresponding posterior distribution $\Pi(\cdot|Y)$ has the same desirable asymptotic properties as the empirical Bayes posterior (2.4).

2.2 Adaptation and contraction rates for the full parameter

Understanding of the asymptotic behavior of the maximum likelihood estimator $\hat{\alpha}_n$ is a crucial element in our proofs of the contraction rate results for the empirical and hierarchical Bayes procedures. The estimator somehow "estimates" the regularity of the true parameter μ_0, but in a rather indirect and involved manner in general. Our first theorem gives deterministic upper and lower bounds for $\hat{\alpha}_n$, whose construction involves the function $h_n : (0, \infty) \to [0, \infty)$ defined by

$$h_n(\alpha) = \frac{1 + 2\alpha + 2p}{n^{1/(1+2\alpha+2p)} \log n} \sum_{i=1}^\infty \frac{n^2 i^{1+2\alpha} \mu_{0,i}^2 \log i}{(i^{1+2\alpha}\kappa_i^{-2} + n)^2}. \tag{2.6}$$

For positive constants $0 < l < L$ we define the lower and upper bounds as

$$\underline{\alpha}_n = \inf\{\alpha > 0 : h_n(\alpha) > l\} \wedge \sqrt{\log n}, \tag{2.7}$$

$$\overline{\alpha}_n = \inf\{\alpha > 0 : h_n(\alpha) > L(\log n)^2\}, \tag{2.8}$$

and the infimum of the empty set is considered ∞.

One can see that the function h_n and hence the lower and upper bounds $\underline{\alpha}_n$ and $\overline{\alpha}_n$ depend on the true μ_0. We show in Theorem 1 that the maximum likelihood estimator $\hat{\alpha}_n$ is between these bounds with probability tending to one. In general the true μ_0 can have very complicated tail behavior, which makes it difficult to understand the behavior of the upper and lower bounds. If μ_0 has regular tails however, we can get some insight in the nature of the bounds. We have the following lemma, proved in Sect. 4.

Lemma 1 *For any $l, L > 0$ in the definitions (2.7)–(2.8) the following statements hold.*

(i) *For all $\beta, R > 0$, there exists $c_0 > 0$ such that*

$$\inf_{\|\mu_0\|_\beta \leq R} \underline{\alpha}_n \geq \beta - \frac{c_0}{\log n}$$

for n large enough.

(ii) *For all $\gamma, R > 0$,*

$$\inf_{\|\mu_0\|_{A^\gamma} \leq R} \underline{\alpha}_n \geq \frac{\sqrt{\log n}}{\log \log n}$$

for n large enough.

(iii) *If $\mu_{0,i} \geq ci^{-\gamma-1/2}$ for some $c, \gamma > 0$, then for a constant $C_0 > 0$ only depending on c and γ, we have $\overline{\alpha}_n \leq \gamma + C_0(\log \log n)/\log n$ for all n large enough.*

(iv) *If $\mu_{0,i} \neq 0$ for some $i \geq 2$, then $\overline{\alpha}_n \leq (\log n)/(2\log 2) - 1/2 - p$ for n large enough.*

We note that items (i) and (iii) of the lemma imply that if $\mu_{0,i} \asymp i^{-1/2-\beta}$, then the interval $[\underline{\alpha}_n, \overline{\alpha}_n]$ concentrates around the value β asymptotically. In combination with Theorem 1 this shows that at least in this regular case, $\hat{\alpha}_n$ correctly estimates the regularity of the truth. A parameter μ_0 in an analytic class A^γ could be viewed as being infinitely regular. By item (ii) of the lemma, which shows that $\underline{\alpha}_n \to \infty$ in this case, the procedure correctly detects this infinite regularity (although of course it does not reveal the value of γ).

Item (iv) implies that if $\mu_{0,i} \neq 0$ for some $i \geq 2$, then $\overline{\alpha}_n < \log n < \infty$ for large n. Conversely, the definitions of h_n and $\overline{\alpha}_n$ show that if $\mu_{0,i} = 0$ for all $i \geq 2$, then $h_n \equiv 0$ and hence $\overline{\alpha}_n = \infty$. This justifies the choice of the truncated $\hat{\alpha}_n$ in the definition of the empirical Bayes posterior.

The following theorem asserts that the point(s) where ℓ_n is maximal is (are) asymptotically between the bounds just defined, uniformly over Sobolev and analytic scales. The proof is given in Sect. 5.

Theorem 1 *For every $R > 0$ the constants l and L in (2.7) and (2.8) can be chosen such that*

$$\inf_{\mu_0 \in \mathscr{B}(R)} P_0 \left(\operatorname*{argmax}_{\alpha \in [0, \log n]} \ell_n(\alpha) \in [\underline{\alpha}_n, \overline{\alpha}_n] \right) \to 1,$$

where $\mathscr{B}(R) = \{\mu_0 \in \ell_2 : \|\mu_0\|_\beta \le R\}$ or $\mathscr{B}(R) = \{\mu_0 \in \ell_2 : \|\mu_0\|_{A^\gamma} \le R\}$.

With the help of Theorem 1 we can prove the following theorem, which states that the empirical Bayes posterior distribution (2.4) achieves optimal minimax contraction rates up to a slowly varying factor, uniformly over Sobolev and analytic scales. Careful inspection of the proof of Theorem 1 indicates that $\hat{\alpha}_n$ is contained with probability tending to 1 in a slightly smaller interval obtained by raising or lowering the bounds by a suitable multiple of $1/\log n$, but this does not help to improve the main results of the paper presented below. We also note that posterior contraction at a rate ε_n implies the existence of estimators, based on the posterior, that converge at the same rate. See for instance the construction in Sect. 4 of [5].

Theorem 2 *For every $\beta, \gamma, R > 0$ and $M_n \to \infty$ we have*

$$\sup_{\|\mu_0\|_\beta \le R} \mathrm{E}_0 \Pi_{\hat{\alpha}_n} \left(\|\mu - \mu_0\| \ge M_n L_n n^{-\beta/(1+2\beta+2p)} \,\big|\, Y \right) \to 0$$

and

$$\sup_{\|\mu_0\|_{A^\gamma} \le R} \mathrm{E}_0 \Pi_{\hat{\alpha}_n} \left(\|\mu - \mu_0\| \ge M_n L_n (\log n)^{1/2+p} n^{-1/2} \,\big|\, Y \right) \to 0,$$

where (L_n) is a slowly varying sequence.

So indeed we see that both in the Sobolev and analytic cases, we obtain the optimal minimax rates up to a slowly varying factor. The proofs of the statements (given in Sect. 6) show that in the first case we can take $L_n = (\log n)^2 (\log \log n)^{1/2}$ and in the second case $L_n = (\log n)^{(1/2+p)\sqrt{\log n}/2 + 1 - p} (\log \log n)^{1/2}$. These sequences converge to infinity but they are slowly varying, hence they converge slower than any power of n.

The full Bayes procedure using the hierarchical prior (2.5) achieves the same results as the empirical Bayes method, under mild assumptions on the prior density λ for α.

Assumption 1 *Assume that for every $c_1 > 0$ there exist $c_2 \ge 0$, $c_3 \in \mathbb{R}$, with $c_3 > 1$ if $c_2 = 0$, and $c_4 > 0$ such that*

$$c_4^{-1} \alpha^{-c_3} \exp(-c_2 \alpha) \le \lambda(\alpha) \le c_4 \alpha^{-c_3} \exp(-c_2 \alpha)$$

for $\alpha \ge c_1$.

One can see that a many distributions satisfy this assumption, for instance the exponential, gamma and inverse gamma distributions. Careful inspection of the proof of the following theorem, given in Sect. 7, can lead to weaker assumptions, although these will be less attractive to formulate. Recall the notation $\Pi(\cdot | Y)$ for the posterior corresponding to the hierarchical prior (2.5).

Theorem 3 *Suppose the prior density* λ *satisfies Assumption* 1. *Then for every* $\beta, \gamma, R > 0$ *and* $M_n \to \infty$ *we have*

$$\sup_{\|\mu_0\|_\beta \leq R} \mathrm{E}_0 \Pi\big(\|\mu - \mu_0\| \geq M_n L_n n^{-\beta/(1+2\beta+2p)} \,\big|\, Y\big) \to 0$$

and

$$\sup_{\|\mu_0\|_{A\gamma} \leq R} \mathrm{E}_0 \Pi\big(\|\mu - \mu_0\| \geq M_n L_n (\log n)^{1/2+p} n^{-1/2} \,\big|\, Y\big) \to 0,$$

where (L_n) *is a slowly varying sequence.*

The hierarchical Bayes method thus yields exactly the same rates as the empirical method, and therefore the interpretation of this theorem is the same as before. We note that already in the direct case $p = 0$ this theorem is an interesting extension of the existing results of [5]. In particular we find that using hierarchical Bayes we can adapt to a continuous range of Sobolev regularities while incurring only a logarithmic correction of the optimal rate.

2.3 Discussion on linear functionals

It is known already in the non-adaptive situation that for attaining optimal rates relative to losses other than the ℓ_2-norm, it may be necessary to set the hyperparameter to a value different from the optimal choice for ℓ_2-recovery of the full parameter μ. If we are for instance interested in optimal estimation of the (possibly unbounded) linear functional

$$L\mu = \sum l_i \mu_i, \tag{2.9}$$

where $l_i \asymp i^{-q-1/2}$ for some $q < p$, then if $\mu_0 \in S^\beta$ for $\beta > -q$ the optimal Gaussian prior (2.1) is not Π_β, but rather $\Pi_{\beta-1/2}$. The resulting, optimal rate is of the order $n^{-(\beta+q)/(2\beta+2p)}$ (see [22], Sect. 5).

An example of this phenomenon occurs when considering global L_2-loss estimation of a function versus pointwise estimation. If for instance the μ_i are the Fourier coefficients of a smooth function of interest $f \in L^2[0, 1]$ relative to the standard Fourier basis e_i and for a fixed $t \in [0, 1]$, $l_i = e_i(t)$, then estimating μ relative to ℓ_2-loss corresponds to estimating f relative to L_2-loss and estimating the functional $L\mu$ in (2.9) corresponds to pointwise estimation of f in the point t (in this case $q = -1/2$).

Theorems 2 and 3 show that the empirical and hierarchical Bayes procedures automatically achieve a bias-variance-posterior spread trade-off that is optimal for the recovery of the full parameter μ_0 relative to the global ℓ_2-norm. As conjectured in a similar setting in [22] this suggests that the adaptive approaches might be sub-optimal outside the ℓ_2-setting. In view of the findings in the non-adaptive case we might expect however that we can slightly alter the procedures to deal with linear functionals. For instance, it is natural to expect that for the linear functional (2.9), the empirical Bayes posterior $\Pi_{\hat\alpha_n - 1/2}(\cdot|Y)$ yields optimal rates.

Matters seem to be more delicate however. A combination of elements of the proof of Theorem 5.1 of [22] and new results on the coverage of credible sets from the paper

[36] lead us to conjecture that for linear functionals L with coefficients $l_i \asymp i^{-q-1/2}$ for some $q < p$ and $\beta > -q$ there exists a $\mu_0 \in S^\beta$ such that along a subsequence n_j,

$$E_0 \Pi_{\hat{\alpha}_{n_j}-1/2} \left(\mu : |L\mu_0 - L\mu| \geq m n_j^{-(\beta+q)/(1+2\beta+2p)} | Y \right) \to 1,$$

as $j \to \infty$ for a positive, small enough constant $m > 0$. Since $n_j^{-(\beta+q)/(1+2\beta+2p)}$ tends to zero at a slower rate than the minimax rate $n^{-(\beta+q)/(2\beta+2p)}$ for S^β, this means that there exist "bad truths" for which the adjusted empirical Bayes procedure does not concentrate at the optimal rate along a subsequence. For linear functionals (2.9) the empirical Bayes posterior $\Pi_{\hat{\alpha}_n-1/2}(\cdot|Y)$ seems only to contract at an optimal rate for "sufficiently nice" truths, for instance of the form $\mu_{0,i} \asymp i^{-1/2-\beta}$, or the more general *polished-tail* sequences considered in [36].

Similar statements are expected to hold for hierarchical Bayes procedures. This adds to the list of remarkable behaviours of marginal posteriors for linear functionals, cf. also [31], for instance. Further research is necessary to shed more light on these matters.

3 Numerical illustration

Consider the inverse signal-in-white-noise problem where we observe the process $(Y_t : t \in [0, 1])$ given by

$$Y_t = \int_0^t \int_0^s \mu(u) \, du \, ds + \frac{1}{\sqrt{n}} W_t,$$

with W a standard Brownian motion, and the aim is to recover the function μ. If, slightly abusing notation, we define $Y_i = \int_0^1 e_i(t) \, dY_t$, for e_i the orthonormal basis functions given by $e_i(t) = \sqrt{2} \cos((i - 1/2)\pi t)$, then it is easily verified that the observations Y_i satisfy (1.1), with $\kappa_i^2 = ((i - 1/2)^2 \pi^2)^{-1}$, i.e. $p = 1$ in (1.2), and μ_i the Fourier coefficients of μ relative to the basis e_i.

We first consider simulated data from this model for μ_0 the function with Fourier coefficients $\mu_{0,i} = i^{-3/2} \sin(i)$, so we have a truth which essentially has regularity 1. In the following figure we plot the true function μ_0 (black dashed curve) and the empirical Bayes posterior mean (red curve) in the left panels, and the corresponding normalized likelihood $\exp(\ell_n)/\max(\exp(\ell_n))$ in the right panels (we truncated the sum in (2.2) at a high level). Figure 1 shows the results for the empirical Bayes procedure with simulated data for $n = 10^3, 10^5, 10^7, 10^9$, and 10^{11}, from top to bottom. The figure shows that the estimator $\hat{\alpha}_n$ does a good job in this case at estimating the regularity level 1, at least for large enough n. We also see however that due to the ill-posedness of the problem, a large signal-to-noise ratio n is necessary for accurate recovery of the function μ.

We applied the hierarchical Bayes method to the simulated data as well. We chose a standard exponential prior distribution on α, which satisfies Assumption 1. Since the posterior can not be computed explicitly, we implemented an MCMC algorithm that generates (approximate) draws from the posterior distribution of the pair (α, μ). More precisely, we fixed a large index $J \in \mathbb{N}$ and defined the vector $\mu^J = (\mu_1, \ldots, \mu_J)$
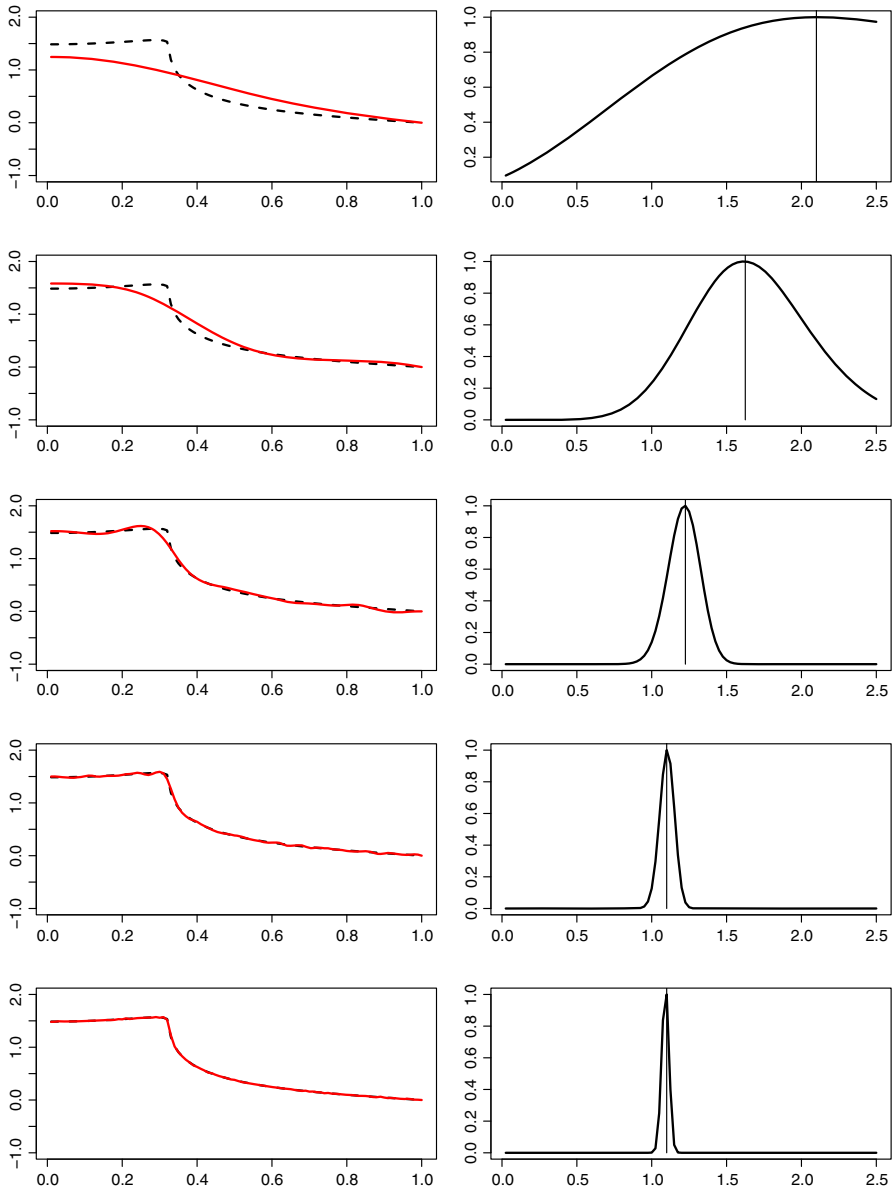
Fig. 1 The degree of ill-posedness $p = 1$. *Left panels* the empirical Bayes posterior mean (*red*) and the true curve (*black, dashed*). *Right panels* corresponding normalized likelihood for α. We have $n = 10^3, 10^5, 10^7, 10^9$, and 10^{11}, from *top to bottom* (color figure online)

consisting of the first J coefficients of μ. (If μ has positive Sobolev regularity, then taking J at least of the order $n^{1/(1+2p)}$ ensures that the approximation error $\|\mu^J - \mu\|$ is of lower order than the estimation rate.) Then we devised a Metropolis-within-Gibbs algorithm for sampling from the posterior distribution of (α, μ^J) (e.g. [37]). The algorithm alternates between draws from the conditional distribution $\mu^J | \alpha, Y$ and the conditional

distribution $\alpha|\mu^J, Y$. The former is explicitly given by (2.3). To sample from $\alpha|\mu^J, Y$ we used a standard Metropolis-Hastings step. It is easily verified that the Metropolis-Hastings acceptance probability for a move from (α, μ) to (α', μ) is given by

$$1 \wedge \frac{q(\alpha'|\alpha)p(\mu^J|\alpha')\lambda(\alpha')}{q(\alpha|\alpha')p(\mu^J|\alpha)\lambda(\alpha)},$$

where $p(\cdot|\alpha)$ is the density of μ^J if $\mu \sim \Pi_\alpha$, i.e.

$$p(\mu^J|\alpha) \propto \prod_{j=1}^{J} j^{1/2+\alpha} e^{-\frac{1}{2}j^{1+2\alpha}\mu_j^2},$$

and q is the transition kernel of the proposal chain. We used a proposal chain that, if it is currently at location α, moves to a new $N(\alpha, \sigma^2)$-distributed location provided the latter is positive. We omit further details, the implementation is straightforward.

The results for the hierarchical Bayes procedure are given in Fig. 2. The figure shows the results for simulated data with $n = 10^3, 10^5, 10^7, 10^9$ and 10^{11}, from top to bottom. Every time we see the posterior mean (in blue) and the true curve (black, dashed) on the left and a histogram for the posterior of α on the right. The results are comparable to what we found for the empirical Bayes procedure.

To illustrate the impact of ill-posedness on the quality of the empirical Bayes procedure we also considered simulated data from the model (1.1) with $\kappa_i = i^{-0.1}$. Recall that $\mu_{0,i} = i^{-3/2}\sin(i)$ are the coefficients of μ_0 relative to the basis e_i as before. Figure 3 shows the results for the empirical Bayes procedure with simulated data for $n = 10, 10^2, 10^3, 10^4$, and 10^5, from top to bottom. Again, in the left panels we plot the true function μ_0 (black dashed curve) and the empirical Bayes posterior mean (red curve), and the corresponding normalized likelihood $\exp(\ell_n)/\max(\exp(\ell_n))$ in the right panels. In this case the estimator $\hat{\alpha}_n$ does a good job at estimating the regularity level 1 already for $n = 100$. Moreover, the true function μ_0 is accurately recovered for moderate values of n.

We also considered simulated data from the original model with $\kappa_i \asymp i^{-1}$ for μ_0 with Fourier coefficients $\mu_{0,i} = (-1)^{i+1}\exp(-2i)$. This function μ_0 is essentially of infinite regularity. Figure 4 shows the results of the empirical Bayes procedure with simulated data for $n = 10^2, 10^3, 10^4, 10^5$, and 10^6, from top to bottom. We can observe that the empirical posterior Bayes mean recovers the function μ well for $n = 10^5$ or 10^6. We also note that the estimated value $\hat{\alpha}_n$ is rather large and unstable. This is not surprising in this case: item (ii) of Lemma 1 shows that the lower bound for $\hat{\alpha}_n$ diverges to infinity. However, large values of α are good enough to capture the infinite regularity of the truth in the empirical Bayes posterior.

4 Proof of Lemma 1

In the proofs we assume for brevity that we have the exact equality $\kappa_i = i^{-p}$. Dealing with the general case (1.2) is straightforward, but makes the proofs somewhat lengthier.
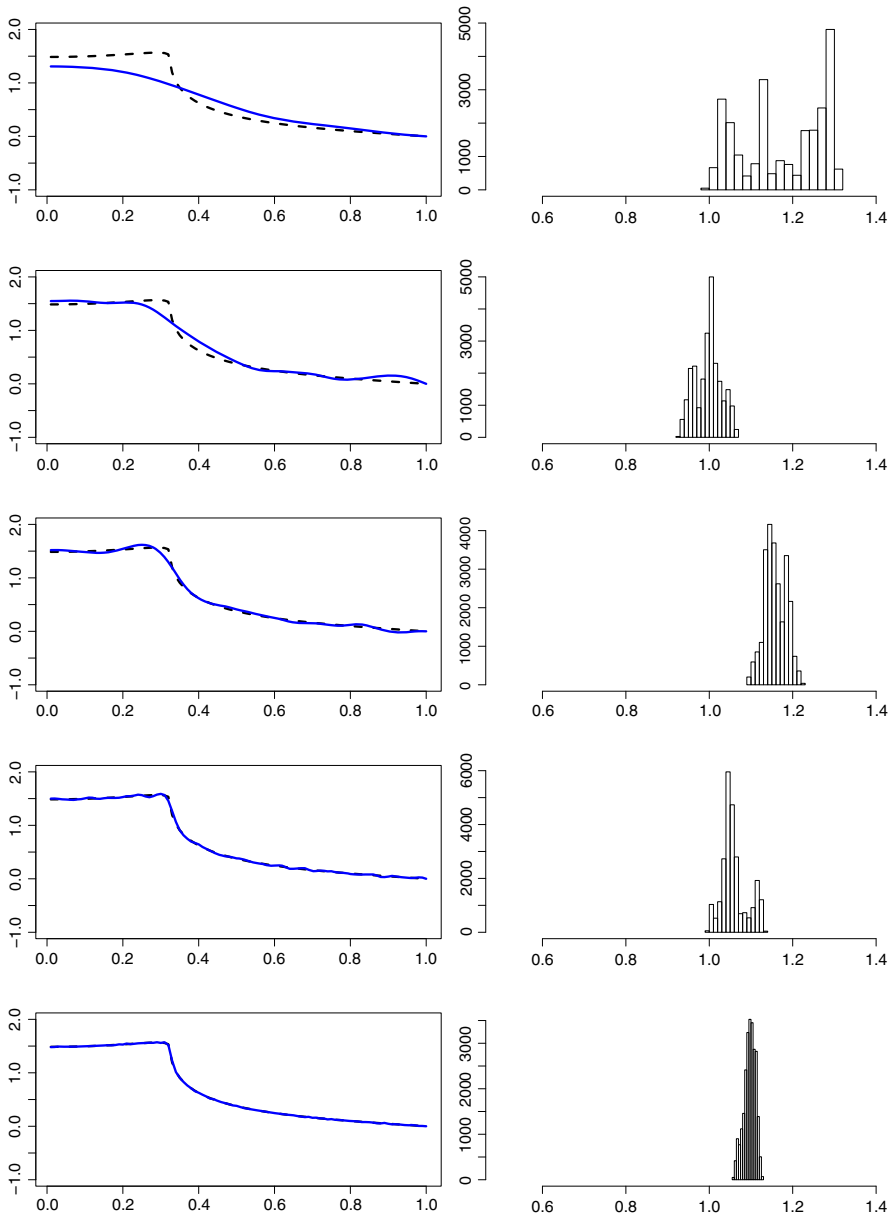
Fig. 2 The degree of ill-posedness $p = 1$. *Left panels* the hierarchical Bayes posterior mean (*blue*) and the true curve (*black*). *Right panels* histograms of posterior for α. We have $n = 10^3, 10^5, 10^7, 10^9$, and 10^{11} from *top to bottom* (color figure online)

(i) We show that for all $\alpha \leq \beta - c_0 / \log n$, for some large enough constant $c_0 > 0$ that only depends on $l, \beta, \|\mu_0\|_\beta$ and p, it holds that $h_n(\alpha) \leq l$, where l is the given positive constant in the definition of $\underline{\alpha}_n$.
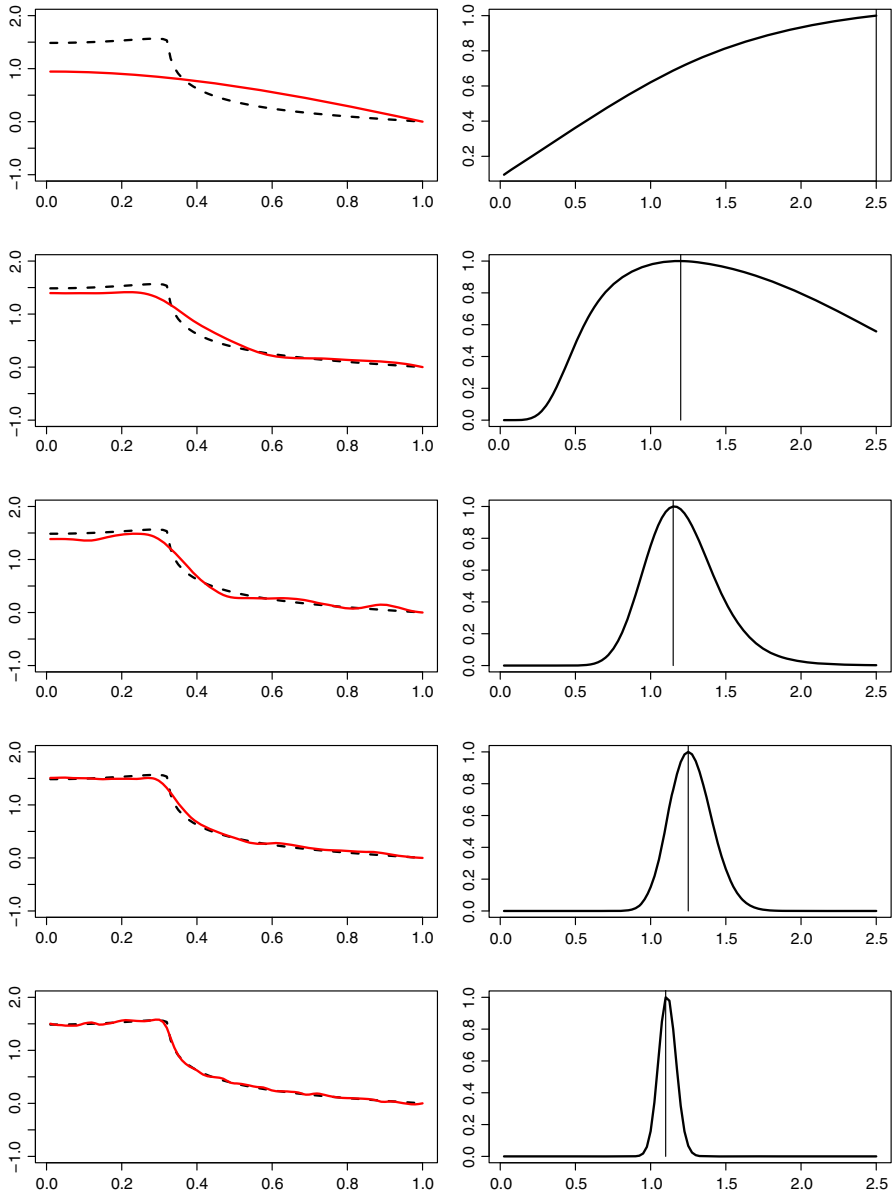
Fig. 3 The degree of ill-posedness $p = 0.1$. *Left panels* the empirical Bayes posterior mean (*red*) and the true curve (*black*). *Right panels* corresponding normalized likelihood for α. We have $n = 10, 10^2, 10^3, 10^4$, and 10^5, from *top to bottom* (color figure online)

The sum in the Definition (2.6) of h_n can be split into two sums, one over indices $i \leq n^{1/(1+2\alpha+2p)}$ and one over indices $i > n^{1/(1+2\alpha+2p)}$. The second sum is bounded by
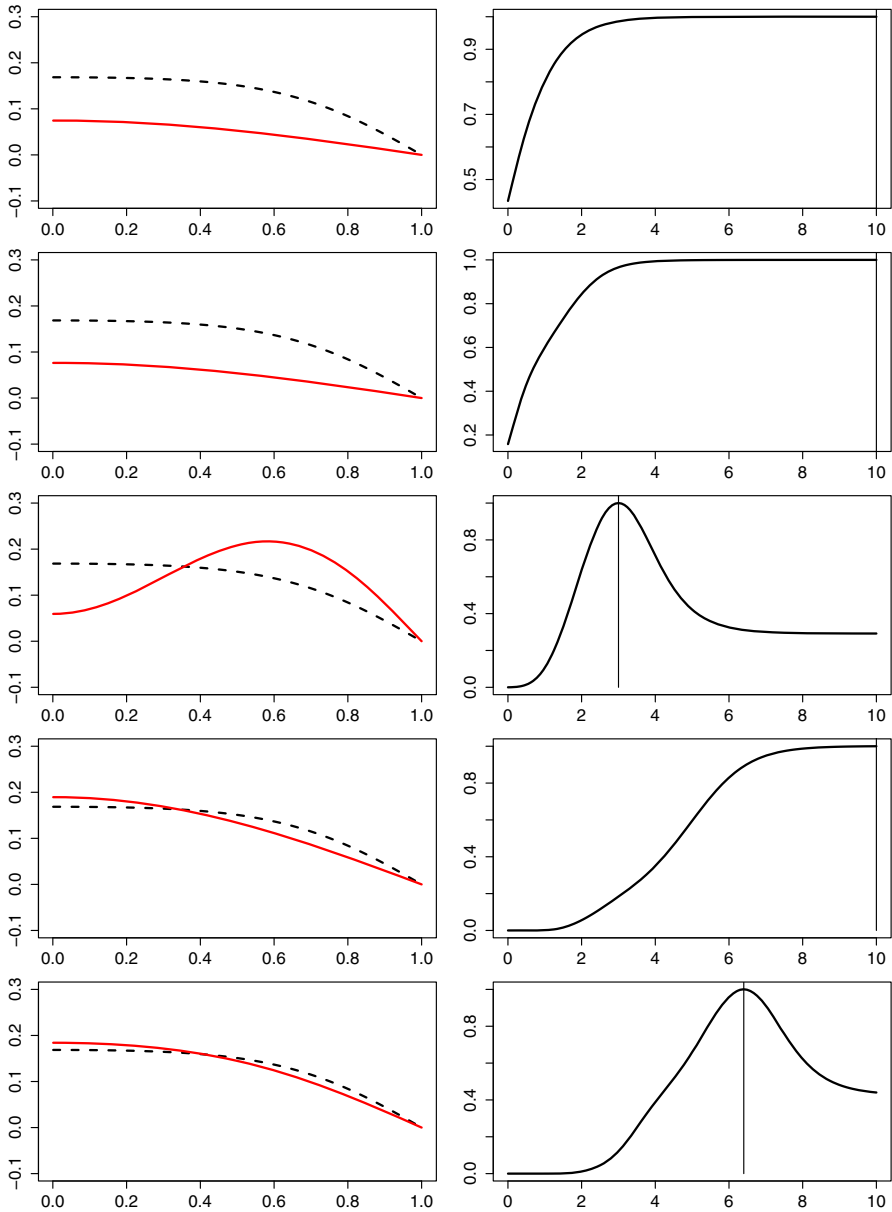
Fig. 4 The degree of ill-posedness $p = 1$. *Left panels* the empirical Bayes posterior mean (*red*) and the true curve (*black*). *Right panels* corresponding normalized likelihood for α. We have $n = 10^2, 10^3, 10^4, 10^5$, and 10^6, from *top to bottom* (color figure online)

$$n^2 \sum_{i \geq n^{1/(1+2\alpha+2p)}} i^{-1-2\alpha-4p-2\beta} (\log i) i^{2\beta} \mu_{0,i}^2.$$

Since the function $x \mapsto x^{-\gamma} \log x$ is decreasing on $[e^{1/\gamma}, \infty)$, this is further bounded by

$$\frac{\|\mu_0\|_\beta^2}{1 + 2\alpha + 2p} n^{\frac{1+2\alpha-2\beta}{1+2\alpha+2p}} \log n.$$

The sum over $i \leq n^{1/(1+2\alpha+2p)}$ is upper bounded by

$$\sum_{i \leq n^{1/(1+2\alpha+2p)}} i^{1+2\alpha-2\beta} i^{2\beta} \mu_{0,i}^2 \log i.$$

Since the logarithm is increasing we can take $(\log n)/(1 + 2\alpha + 2p)$ outside the sum and then bound $i^{1+2\alpha-2\beta}$ above by $n^{(1+2\alpha-2\beta)/(1+2\alpha+2p)\vee 0}$ to arrive at the subsequent bound

$$\frac{\|\mu_0\|_\beta^2}{1 + 2\alpha + 2p} n^{0 \vee \frac{1+2\alpha-2\beta}{1+2\alpha+2p}} \log n.$$

Combining the bounds for the two sums we obtain the upper bound

$$h_n(\alpha) \leq \|\mu_0\|_\beta^2 n^{-\frac{1\wedge 2(\beta-\alpha)}{1+2\alpha+2p}},$$

valid for all $\alpha > 0$. Now suppose that $\alpha \leq \beta - c_0/\log n$. Then for n large enough, the power of n on the right-hand side is bounded by

$$n^{-\frac{1\wedge 2(c_0/\log n)}{1+2\beta+2p}} = e^{-\frac{2c_0}{1+2\beta+2p}}.$$

Hence given $l > 0$ we can choose c_0 so large, only depending on $l, \beta, \|\mu_0\|_\beta$ and p, that $h_n(\alpha) \leq l$ for $\alpha \leq \beta - c_0/\log n$.

(ii) We show that in this case we have $h_n(\alpha) \leq l$ for $\alpha \leq \sqrt{\log n}/(\log \log n)$ and $n \geq n_0$, where n_0 only depends on $\|\mu_0\|_{A^\gamma}$. Again we give an upper bound for h_n by splitting the sum in its definition into two smaller sums. The one over indices $i > n^{1/(1+2\alpha+2p)}$ is bounded by

$$n^2 \sum_{i > n^{1/(1+2\alpha+2p)}} i^{-1-2\alpha-4p} e^{-2\gamma i} (\log i) e^{2\gamma i} \mu_{0,i}^2.$$

Using the fact that for $\delta > 0$ the function $x \mapsto x^{-\delta} e^{-2\gamma x} \log x$ is decreasing on $[e^{1/\delta}, \infty)$ we can see that this is further bounded by

$$\frac{\|\mu_0\|_{A^\gamma}^2}{1 + 2\alpha + 2p} e^{-2\gamma n^{1/(1+2\alpha+2p)}} n^{\frac{1+2\alpha}{1+2\alpha+2p}} \log n.$$

The sum over indices $i \leq n^{1/(1+2\alpha+2p)}$ is bounded by

$$\frac{\log n}{1 + 2\alpha + 2p} \sum_{i \leq n^{1/(1+2\alpha+2p)}} i^{1+2\alpha} e^{-2\gamma i} e^{2\gamma i} \mu_{0,i}^2.$$

Since the maximum on $(0, \infty)$ of the function $x \mapsto x^{1+2\alpha} \exp(-2\gamma x)$ equals $\exp((1 + 2\alpha)(\log((1 + 2\alpha)/2\gamma) - 1))$, we have the subsequent bound

$$\frac{\|\mu_0\|_{A^\gamma}^2}{1 + 2\alpha + 2p} e^{(1+2\alpha) \log((1+2\alpha)/2\gamma)} \log n.$$

Combining the two bounds we find that

$$h_n(\alpha) \le \|\mu_0\|_{A^\gamma}^2 \left(n^{\frac{2\alpha}{1+2\alpha+2p}} e^{-2\gamma n^{\frac{1}{1+2\alpha+2p}}} + n^{-\frac{1}{1+2\alpha+2p}} e^{(1+2\alpha) \log \frac{1+2\alpha}{2\gamma}} \right)$$

for all $\alpha > 0$. It is then easily verified that for the given constant $l > 0$, we have $h_n(\alpha) \le l$ for $n \ge n_0$ if $\alpha \le \sqrt{\log n}/\log\log n$, where n_0 only depends on $\|\mu_0\|_{A^\gamma}$.

(iii) Let $\gamma_n = \gamma + C_0(\log\log n)/(\log n)$. We will show that for n large enough, $h_n(\gamma_n) \ge L(\log n)^2$, provided C_0 is large enough. Note that

$$\sum_{i=1}^\infty \frac{n^2 i^{1+2\gamma_n} \mu_{0,i}^2 \log i}{(i^{1+2\gamma_n+2p} + n)^2} \ge \frac{c^2}{4} \sum_{i \le n^{1/(1+2\gamma_n+2p)}} i^{2(\gamma_n - \gamma)} \log i.$$

By monotonicity and the fact that $\lfloor x \rfloor \ge x/2$ for x large, the sum on the right is bounded from below by the integral

$$\int_0^{n^{1/(1+2\gamma_n+2p)}/2} x^{2\gamma_n - 2\gamma} \log x \, dx.$$

This integral can be computed explicitly and is for large n bounded from below by a constant times

$$\frac{\log n}{1 + 2\gamma_n + 2p} n^{\frac{2\gamma_n - 2\gamma + 1}{1+2\gamma_n+2p}}.$$

It follows that, for large enough n, $h_n(\gamma_n)$ is bounded from below by a constant times $c^2 n^{2(\gamma_n - \gamma)/(1+2\gamma_n+2p)}$. Since $(\log\log n)/(\log n) \le 1/4$ for n large enough, we obtain

$$n^{2(\gamma_n - \gamma)/(1+2\gamma_n+2p)} \ge n^{\frac{1}{\log n}(\log\log n)\frac{2C_0}{1+2\gamma+C_0/2+2p}} = (\log n)^{2C_0/(1+2\gamma+C_0/2+2p)}.$$

Hence for C_0 large enough, only depending on c and γ, we indeed have that and $h_n(\gamma_n) \ge L(\log n)^2$ for large n.

(iv) If $\mu_{0,i} \ne 0$ for $i \ge 2$, then

$$h_n(\alpha) \gtrsim \frac{1 + 2\alpha + 2p}{n^{1/(1+2\alpha+2p)} \log n} \frac{n^2 i^{1+2\alpha}}{(i^{1+2\alpha+2p} + n)^2}.$$

Now define α_n such that $i^{1+2\alpha_n+2p} = n$. Then by construction we have $h_n(\alpha_n) \gtrsim n^{1-(1+2p)/(1+2\alpha_n+2p)}$. Since $\alpha_n \to \infty$ the right side is larger than $L \log^2 n$ for n large enough, irrespective of the value of L, hence $\bar\alpha_n \le \alpha_n \le (\log n)/(2\log 2) - 1/2 - p$.

5 Proof of Theorem 1

With the help of the dominated convergence theorem one can see that the random function ℓ_n is $(P_0 - a.s.)$ differentiable and its derivative, which we denote by $\mathbb{M}_n$, is given by

$$\mathbb{M}_n(\alpha) = \sum_{i=1}^{\infty} \frac{n \log i}{i^{1+2\alpha} \kappa_i^{-2} + n} - \sum_{i=1}^{\infty} \frac{n^2 i^{1+2\alpha} \kappa_i^{-2} \log i}{(i^{1+2\alpha} \kappa_i^{-2} + n)^2} Y_i^2.$$

We will show that on the interval $(0, \underline{\alpha}_n + 1/\log n]$ the random function $\mathbb{M}_n$ is positive and bounded away from 0 with probability tending to one, hence ℓ_n has no local maximum in this interval. Next we distinguish two cases according to the value of $\overline{\alpha}_n$. If $\overline{\alpha}_n > \log n$, then the inequality $\hat{\alpha}_n \leq \overline{\alpha}_n$ trivially holds. In the case $\overline{\alpha}_n \leq \log n$ we show that for a constant $C_1 > 0$ we a.s. have

$$\ell_n(\alpha) - \ell_n(\overline{\alpha}_n) = \int_{\overline{\alpha}_n}^{\alpha} \mathbb{M}_n(\gamma) \, d\gamma \leq C_1 \frac{n^{1/(1+2\overline{\alpha}_n + 2p)} (\log n)^2}{1 + 2\overline{\alpha}_n + 2p} \tag{5.1}$$

for all $\alpha \geq \overline{\alpha}_n$. Then we prove that for any given $C_2 > 0$, the constant L can be set such that for $\gamma \in [\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n]$ we have

$$\mathbb{M}_n(\gamma) \leq -C_2 \frac{n^{1/(1+2\overline{\alpha}_n + 2p)} (\log n)^3}{1 + 2\overline{\alpha}_n + 2p}$$

with probability tending to one uniformly. Together with (5.1) this means that on the interval $[\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n]$ the function ℓ_n decreases more than it can possibly increase on the interval $[\overline{\alpha}_n, \infty)$. Therefore, it holds with probability tending to one that ℓ_n has no global maximum on $(\overline{\alpha}_n - 1/\log n, \infty)$.

Recall that $\mathscr{B}(R) = \{\mu_0 \in \ell_2 : \|\mu_0\|_\beta \leq R\}$ or $\mathscr{B}(R) = \{\mu_0 \in \ell_2 : \|\mu_0\|_{A^\gamma} \leq R\}$. Again for simplicity we assume $\kappa_i = i^{-p}$ in the proof.

5.1 $\mathbb{M}_n(\alpha)$ on $[\overline{\alpha}_n, \infty)$

In this section we give a deterministic upper bound for the integral of $\mathbb{M}_n(\alpha)$ on the interval $[\overline{\alpha}_n, \infty)$.

We have the trivial bound

$$\mathbb{M}_n(\alpha) \leq \sum_{i=1}^{\infty} \frac{n \log i}{i^{1+2\alpha+2p} + n}.$$

An application of Lemma 7(i) with $r = 1 + 2\alpha + 2p$ and $c = \beta + 2p$ shows that for $\beta/2 < \alpha \leq \log n$,

$$\mathbb{M}_n(\alpha) \lesssim \frac{1}{1 + 2\alpha + 2p} n^{1/(1+2\alpha+2p)} \log n.$$

For $\alpha \geq \log n$ we apply Lemma 7(ii), and see that $\mathbb{M}_n(\alpha) \lesssim n2^{-1-2\alpha-2p}$. Using the fact that $x \mapsto 2^{-x}x^3$ is decreasing for large x, it is easily seen that $n2^{-1-2\alpha-2p} \lesssim (\log n)^3/(1+2\alpha+2p)^3$ for $\alpha \geq \log n$, hence

$$\mathbb{M}_n(\alpha) \lesssim \frac{(\log n)^3}{(1+2\alpha+2p)^3}.$$

By Lemma 1 we have $\beta/2 < \overline{\alpha}_n$ for large enough n, both for the case that $\mu_0 \in S^\beta$ and $\mu_0 \in A^\gamma$, since for any $\beta > 0$ we have $\sqrt{\log n}/\log\log n \geq \beta/2$ for large enough n. It follows that the integral we want to bound is bounded by a constant times

$$n^{1/(1+2\overline{\alpha}_n+2p)} \log n \int_{\overline{\alpha}_n}^{\log n} \frac{1}{1+2\alpha+2p} \, d\alpha + \log^3 n \int_{\log n}^{\infty} \frac{1}{(1+2\alpha+2p)^3} \, d\alpha.$$

This quantity is bounded by a constant times

$$\frac{n^{1/(1+2\overline{\alpha}_n+2p)}(\log n)^2}{1+2\overline{\alpha}_n+2p}.$$

5.2 $\mathbb{M}_n(\alpha)$ on $\alpha \in [\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n]$

In this section we show that the process $\mathbb{M}_n(\alpha)$ is with probability going to one smaller than a negative, arbitrary large constant times $n^{1/(1+2\overline{\alpha}_n+2p)}(\log n)^3/(1+2\overline{\alpha}_n+2p)$ uniformly on the interval $[\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n]$. More precisely, we show that for every $\beta, R, M > 0$, the constant $L > 0$ in the definition of $\overline{\alpha}_n$ can be chosen such that

$$\limsup_{n\to\infty} \sup_{\mu_0\in\mathscr{B}(R)} \sup_{\alpha\in[\overline{\alpha}_n-1/\log n,\overline{\alpha}_n]} \mathrm{E}_0 \frac{(1+2\alpha+2p)\mathbb{M}_n(\alpha)}{n^{1/(1+2\alpha+2p)}(\log n)^3} < -M \qquad (5.2)$$

$$\sup_{\mu_0\in\mathscr{B}(R)} \mathrm{E}_0 \sup_{\alpha\in[\overline{\alpha}_n-1/\log n,\overline{\alpha}_n]} \frac{(1+2\alpha+2p)|\mathbb{M}_n(\alpha) - \mathrm{E}_0\mathbb{M}_n(\alpha)|}{n^{1/(1+2\alpha+2p)}(\log n)^3} \to 0. \quad (5.3)$$

The expected value of the normalized version of the process $\mathbb{M}_n$ given on the left-hand side of (5.2) is equal to

$$\frac{1+2\alpha+2p}{n^{1/(1+2\alpha+2p)}(\log n)^3} \left(\sum_{i=1}^{\infty} \frac{n^2 \log i}{(i^{1+2\alpha+2p}+n)^2} - \sum_{i=1}^{\infty} \frac{n^2 i^{1+2\alpha} \mu_{0,i}^2 \log i}{(i^{1+2\alpha+2p}+n)^2} \right). \quad (5.4)$$

We write this as the sum of two terms and bound the first term by

$$\frac{1+2\alpha+2p}{n^{1/(1+2\alpha+2p)}(\log n)^3} \sum_{i=1}^{\infty} \frac{n \log i}{i^{1+2\alpha+2p}+n}.$$

We want to bound this quantity for $\alpha \in [\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n]$. By Lemma 1, $\beta/4 < \overline{\alpha}_n - 1/\log n$ for large enough n, both for the case that $\mu_0 \in S^\beta$ and $\mu_0 \in A^\gamma$, so this

interval is included in $(\beta/4, \infty)$. Taking $c = \beta/2 + 2p$ in Lemma 7(i) then shows that the first term is bounded by a multiple of $1/(\log n)^2$ and hence tends to zero, uniformly over $[\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n]$. We now consider the second term in (5.4), which is equal to $h_n(\alpha)/(\log n)^2$. By Lemma 2 for any $\mu_0 \in \ell_2$ and $n \geq e^4$ we have

$$\frac{h_n(\alpha)}{(\log n)^2} \gtrsim \frac{1}{(\log n)^2} h_n(\overline{\alpha}_n) = L,$$

where the last equality holds by the definition of $\overline{\alpha}_n$. This concludes the proof of (5.2).

To verify (5.3) it suffices, by Corollary 2.2.5 in [39] (applied with $\psi(x) = x^2$), to show that

$$\sup_{\mu_0 \in \mathscr{B}(R)} \sup_{\alpha \in [\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n]} \mathrm{var}_0 \frac{(1 + 2\alpha + 2p)\mathbb{M}_n(\alpha)}{n^{1/(1+2\alpha+2p)}(\log n)^3} \to 0, \qquad (5.5)$$

and

$$\sup_{\mu_0 \in \mathscr{B}(R)} \int_0^{\mathrm{diam}_n} \sqrt{N(\varepsilon, [\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n], d_n)} \, d\varepsilon \to 0,$$

where d_n is the semimetric defined by

$$d_n^2(\alpha_1, \alpha_2) = \mathrm{var}_0 \left(\frac{(1 + 2\alpha_1 + 2p)\mathbb{M}_n(\alpha_1)}{n^{1/(1+2\alpha_1+2p)}(\log n)^3} - \frac{(1 + 2\alpha_2 + 2p)\mathbb{M}_n(\alpha_2)}{n^{1/(1+2\alpha_2+2p)}(\log n)^3} \right),$$

diam_n is the diameter of $[\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n]$ relative do d_n, and $N(\varepsilon, B, d)$ is the minimal number of d-balls of radius ε needed to cover the set B.

By Lemma 3

$$\mathrm{var}_0 \frac{(1 + 2\alpha + 2p)\mathbb{M}_n(\alpha)}{n^{1/(1+2\alpha+2p)}(\log n)^3} \lesssim \frac{n^{-1/(1+2\alpha+2p)}}{(\log n)^4} \left(1 + h_n(\alpha) \right), \qquad (5.6)$$

(with an implicit constant that does not depend on μ_0 and α). By the definition of $\overline{\alpha}_n$ the function $h_n(\alpha)$ is bounded above by $L(\log n)^2$ on the interval $[\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n]$. Together with (5.6) it proves (5.5).

The last bound also shows that the d_n-diameter of the set $[\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n]$ is bounded above by a constant times $(\log n)^{-1}$, with a constant that does not depend on μ_0 and α. By Lemma 4 and the fact that $h_n(\alpha) \leq L(\log n)^2$ for $\alpha \in [\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n]$, we get the upper bound, $\alpha_1, \alpha_2 \in [\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n]$,

$$d_n(\alpha_1, \alpha_2) \lesssim |\alpha_1 - \alpha_2|,$$

with a constant that does not depend on μ_0. Therefore $N(\varepsilon, [\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n], d_n) \lesssim 1/(\varepsilon \log n)$ and hence

$$\sup_{\mu_0 \in \mathscr{B}(R)} \int_0^{\mathrm{diam}_n} \sqrt{N(\varepsilon, [\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n], d_n)} \, d\varepsilon \lesssim \frac{1}{\log n} \to 0.$$

5.3 $\mathbb{M}_n(\alpha)$ on $(0, \underline{\alpha}_n + 1/\log n]$

In this subsection we prove that if the constant l in the definition of $\underline{\alpha}_n$ is small enough, then

$$\liminf_{n \to \infty} \inf_{\mu_0 \in \ell_2} \inf_{\alpha \in (0, \underline{\alpha}_n + 1/\log n]} E_0 \frac{(1 + 2\alpha + 2p)\mathbb{M}_n(\alpha)}{n^{1/(1+2\alpha+2p)} \log n} > 0 \qquad (5.7)$$

$$\sup_{\mu_0 \in \ell_2} E_0 \sup_{\alpha \in (0, \underline{\alpha}_n + 1/\log n]} \frac{(1 + 2\alpha + 2p)|\mathbb{M}_n(\alpha) - E_0\mathbb{M}_n(\alpha)|}{n^{1/(1+2\alpha+2p)} \log n} \to 0. \qquad (5.8)$$

This shows that $\mathbb{M}_n$ is positive throughout $(0, \underline{\alpha}_n + 1/\log n]$ with probability tending to one uniformly over ℓ_2.

Since $E_0 Y_i^2 = \kappa_i^2 \mu_{0,i}^2 + 1/n$, the expected value on the left-hand side of (5.7) is equal to

$$\frac{1 + 2\alpha + 2p}{n^{1/(1+2\alpha+2p)} \log n} \sum_{i=1}^{\infty} \frac{n^2 \log i}{(i^{1+2\alpha+2p} + n)^2} - h_n(\alpha). \qquad (5.9)$$

We first find a lower bound for the first term. Since $\underline{\alpha}_n \leq \sqrt{\log n}$ by definition, we have $\alpha \ll \log n$ for all $\alpha \in (0, \underline{\alpha}_n + 1/\log n]$. Then it follows from Lemma 9 that for n large enough, the first term in (5.9) is bounded from below by $1/12$ for all $\alpha \in (0, \underline{\alpha}_n + 1/\log n]$. Next note that by definition of h_n and Lemma 2, we have

$$\sup_{\alpha \in (0, \underline{\alpha}_n + 1/\log n]} h_n(\alpha) \leq Kl,$$

where $K > 0$ is a constant independent of μ_0. So by choosing $l > 0$ small enough, we can indeed ensure that (5.7) is true.

To verify (5.8) it suffices again, by Corollary 2.2.5 in [39] applied with $\psi(x) = x^2$, to show that

$$\sup_{\mu_0 \in \ell_2} \sup_{\alpha \in (0, \overline{\alpha}_n + 1/\log n]} \text{var}_0 \frac{(1 + 2\alpha + 2p)\mathbb{M}_n(\alpha)}{n^{1/(1+2\alpha+2p)} \log n} \to 0, \qquad (5.10)$$

and

$$\sup_{\mu_0 \in \ell_2} \int_0^{\text{diam}_n} \sqrt{N(\varepsilon, (0, \underline{\alpha}_n + 1/\log n], d_n)} \, d\varepsilon \to 0,$$

where d_n is the semimetric defined by

$$d_n^2(\alpha_1, \alpha_2) = \text{var}_0 \left(\frac{(1 + 2\alpha_1 + 2p)\mathbb{M}_n(\alpha_1)}{n^{1/(1+2\alpha_1+2p)} \log n} - \frac{(1 + 2\alpha_2 + 2p)\mathbb{M}_n(\alpha_2)}{n^{1/(1+2\alpha_2+2p)} \log n} \right),$$

diam_n is the diameter of $(0, \underline{\alpha}_n + 1/\log n]$ relative to d_n, and $N(\varepsilon, B, d)$ is the minimal number of d-balls of radius ε needed to cover the set B.

By Lemma 3

$$\text{var}_0 \, \frac{(1 + 2\alpha + 2p)\mathbb{M}_n(\alpha)}{n^{1/(1+2\alpha+2p)} \log n} \lesssim n^{-1/(1+2\alpha+2p)} \left(1 + h_n(\alpha)\right), \qquad (5.11)$$

with a constant that does not depend on μ_0 and α. We have seen that on the interval $(0, \underline{\alpha}_n + 1/\log n]$ the function h_n is bounded by a constant times l, hence the variance in (5.10) is bounded by a multiple of $n^{-1/(1+2\underline{\alpha}_n+2/\log n+2p)} \leq e^{-(1/3)\sqrt{\log n}} \to 0$, which proves (5.10).

The variance bound above also imply that the d_n-diameter of the set $(0, \underline{\alpha}_n + 1/\log n]$ is bounded by a multiple of $e^{-(1/6)\sqrt{\log n}}$. By Lemma 4, the definition of $\underline{\alpha}_n$ and Lemma 2,

$$d_n(\alpha_1, \alpha_2) \lesssim |\alpha_1 - \alpha_2|(\log n)\sqrt{n^{-1/(1+2\underline{\alpha}_n+2/\log n+2p)}} \lesssim |\alpha_1 - \alpha_2|,$$

with constants that do not depend on μ_0. Hence for the covering number of $(0, \underline{\alpha}_n + 1/\log n] \subset (0, 2\sqrt{\log n})$ we have

$$N(\varepsilon, (0, \underline{\alpha}_n + 1/\log n], d_n) \lesssim \frac{\sqrt{\log n}}{\varepsilon},$$

and therefore

$$\sup_{\mu_0 \in \ell_2} \int_0^{\text{diam}_n} \sqrt{N(\varepsilon, (0, \underline{\alpha}_n + 1/\log n], d_n)} \, d\varepsilon \lesssim (\log n)^{1/4} e^{-(1/12)\sqrt{\log n}} \to 0.$$

5.4 Bounds on $h_n(\alpha)$, variances and distances

In this section we prove a number of auxiliary lemmas used in the preceding. The first one is about the behavior of the function h_n in a neighborhood of $\underline{\alpha}_n$ and $\overline{\alpha}_n$.

Lemma 2 *The function h_n satisfies the following bounds:*

$$h_n(\alpha) \gtrsim h_n(\overline{\alpha}_n), \quad \text{for } \alpha \in \left[\overline{\alpha}_n - \frac{1}{\log n}, \overline{\alpha}_n\right] \text{ and } n \geq e^4,$$

$$h_n(\alpha) \lesssim h_n(\underline{\alpha}_n), \quad \text{for } \alpha \in \left[\underline{\alpha}_n, \underline{\alpha}_n + \frac{1}{\log n}\right] \text{ and } n \geq e^2.$$

Proof We provide a detailed proof of the first inequality, the second one can be proved using similar arguments.

Let

$$S_n(\alpha) = \sum_{i=1}^{\infty} \frac{n^2 i^{1+2\alpha} \mu_{0,i}^2 \log i}{(i^{1+2\alpha+2p} + n)^2}$$

be the sum in the definition of h_n. Splitting the sum into two parts we get, for $\alpha \in [\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n]$,

$$4S_n(\alpha) \geq \sum_{i \leq n^{1/(1+2\alpha+2p)}} i^{1+2\overline{\alpha}_n - 2/\log n} \mu_{0,i}^2 \log i$$
$$+ n^2 \sum_{i > n^{1/(1+2\alpha+2p)}} i^{-1-2\overline{\alpha}_n - 4p} \mu_{0,i}^2 \log i.$$

In the first sum $i^{-2/\log n}$ can be bounded below by $\exp(-2)$. Furthermore, for $i \in [n^{1/(1+2\overline{\alpha}_n+2p)}, n^{1/(1+2\alpha+2p)}]$, we have the inequality

$$i^{1+2\overline{\alpha}_n} \mu_{0,i}^2 \log i \geq n^2 i^{-1-2\overline{\alpha}_n - 4p} \mu_{0,i}^2 \log i.$$

Therefore $S_n(\alpha)$ can be bounded from below by a constant times

$$\sum_{i \leq n^{1/(1+2\overline{\alpha}_n+2p)}} i^{1+2\overline{\alpha}_n} \mu_{0,i}^2 \log i + n^2 \sum_{i > n^{1/(1+2\overline{\alpha}_n+2p)}} i^{-1-2\overline{\alpha}_n - 4p} \mu_{0,i}^2 \log i$$

$$\geq \sum_{i \leq n^{1/(1+2\overline{\alpha}_n+2p)}} \frac{n^2 i^{1+2\overline{\alpha}_n} \mu_{0,i}^2 \log i}{(i^{1+2\overline{\alpha}_n+2p} + n)^2} + \sum_{i > n^{1/(1+2\overline{\alpha}_n+2p)}} \frac{n^2 i^{1+2\overline{\alpha}_n} \mu_{0,i}^2 \log i}{(i^{1+2\overline{\alpha}_n+2p} + n)^2}.$$

Hence, we have $S_n(\alpha) \gtrsim S_n(\overline{\alpha}_n)$ for $\alpha \in [\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n]$.

Next note that for $n \geq e^4$ we have $2(1 + 2\overline{\alpha}_n - 2/\log n + 2p) \geq 1 + 2\overline{\alpha}_n + 2p$. Moreover, $n^{-1/(1+2\overline{\alpha}_n - 2/\log n+2p)} \gtrsim n^{-1/(1+2\overline{\alpha}_n+2p)}$. Therefore

$$\frac{1 + 2\alpha + 2p}{n^{1/(1+2\alpha+2p)} \log n} \gtrsim \frac{1 + 2\overline{\alpha}_n + 2p}{n^{1/(1+2\overline{\alpha}_n+2p)} \log n}$$

for $\alpha \in [\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n]$ and for $n \geq e^4$. Combining this with the inequality for $S_n(\alpha)$ yields the desired result.

Next we present two results on variances involving the random function $\mathbb{M}_n$.

Lemma 3 *For any $\alpha > 0$,*

$$\text{var}_0 \frac{(1 + 2\alpha + 2p)\mathbb{M}_n(\alpha)}{n^{1/(1+2\alpha+2p)}} \lesssim n^{-1/(1+2\alpha+2p)}(\log n)^2 \left(1 + h_n(\alpha)\right).$$

Proof The random variables Y_i^2 are independent and $\text{var}_0 Y_i^2 = 2/n^2 + 4\kappa_i^2 \mu_{0,i}^2/n$, hence the variance in the statement of the lemma is equal to

$$\frac{2n^2(1 + 2\alpha + 2p)^2}{n^{2/(1+2\alpha+2p)}} \sum_{i=1}^{\infty} \frac{i^{2+4\alpha+4p}(\log i)^2}{(i^{1+2\alpha+2p} + n)^4}$$

$$+ \frac{4n^3(1 + 2\alpha + 2p)^2}{n^{2/(1+2\alpha+2p)}} \sum_{i=1}^{\infty} \frac{i^{2+4\alpha+2p}(\log i)^2 \mu_{0,i}^2}{(i^{1+2\alpha+2p} + n)^4}. \tag{5.12}$$

By Lemma 10 the first term is bounded by

$$
\frac{2n(1+2\alpha+2p)\log n}{n^{2/(1+2\alpha+2p)}} \sum_{i=1}^{\infty} \frac{i^{1+2\alpha+2p}\log i}{(i^{1+2\alpha+2p}+n)^2}
$$

$$
\leq \frac{2(1+2\alpha+2p)\log n}{n^{2/(1+2\alpha+2p)}} \sum_{i=1}^{\infty} \frac{n\log i}{i^{1+2\alpha+2p}+n}.
$$

Lemma 7(i) further bounds the right hand side of the above display by a multiple of $n^{-1/(1+2\alpha+2p)}(\log n)^2$ uniformly for $\alpha > c$, where $c > 0$ is an arbitrary constant. For $\alpha \leq c$ we get the same bound by applying Lemma 8 (with $m = 2$, $l = 4$, $r = 1 + 2\alpha + 2p$, $r_0 = 1 + 2c + 2p$, and $s = 2r$) to the first term in (5.12). By Lemma 10, the second term in (5.12) is bounded by

$$
4n^{-2/(1+2\alpha+2p)}(1+2\alpha+2p)(\log n) \sum_{i=1}^{\infty} \frac{n^2 i^{1+2\alpha}\mu_{0,i}^2 \log i}{(i^{1+2\alpha}\kappa_i^{-2}+n)^2}
$$

$$
= 4n^{-1/(1+2\alpha+2p)}(\log n)^2 h_n(\alpha).
$$

Combining the upper bounds for the two terms we arrive at the assertion of the lemma.

Lemma 4 *For any $0 < \alpha_1 < \alpha_2 < \infty$ we have that*

$$
\mathrm{var}_0 \left(\frac{(1+2\alpha_1+2p)\mathbb{M}_n(\alpha_1)}{n^{1/(1+2\alpha_1+2p)}} - \frac{(1+2\alpha_2+2p)\mathbb{M}_n(\alpha_2)}{n^{1/(1+2\alpha_2+2p)}} \right)
$$
$$
\lesssim (\alpha_1 - \alpha_2)^2 (\log n)^4 \sup_{\alpha \in [\alpha_1,\alpha_2]} n^{-1/(1+2\alpha+2p)}(1+h_n(\alpha)),
$$

with a constant that does not depend on α and μ_0.

Proof The variance we have to bound can be written as

$$
n^4 \sum_{i=1}^{\infty} (f_i(\alpha_1) - f_i(\alpha_2))^2 (\log i)^2 \, \mathrm{var}_0 \, Y_i^2,
$$

where $f_i(\alpha) = (1 + 2\alpha + 2p)i^{1+2\alpha+2p}n^{-1/(1+2\alpha+2p)}(i^{1+2\alpha+2p}+n)^{-2}$. For the derivative of f_i we have $f_1'(\alpha) = 2f_1(\alpha)(1/(1+2\alpha+2p) + \log n/(1+2\alpha+2p)^2)$ and for $i \geq 2$,

$$
|f_i'(\alpha)| = \left| 2f_i(\alpha) \left(\frac{1}{1+2\alpha+2p} + \log i + \frac{\log n}{(1+2\alpha+2p)^2} - \frac{2i^{1+2\alpha+2p}\log i}{i^{1+2\alpha+2p}+n} \right) \right|
$$
$$
\leq 8f_i(\alpha) \left(\log i + (\log n)/(1+2\alpha+2p)^2 \right).
$$

It follows that the variance is bounded by a constant times

$$(\alpha_1 - \alpha_2)^2 n^4 \sup_{\alpha \in [\alpha_1, \alpha_2]} (1 + 2\alpha + 2p)^2$$

$$\times \left(\sum_{i=1}^{\infty} \frac{i^{2+4\alpha+4p}(\log i)^2 \left(1 \vee \log i + (\log n)/(1 + 2\alpha + 2p)^2\right)^2}{n^{2/(1+2\alpha+2p)}(i^{1+2\alpha+2p} + n)^4} \operatorname{var}_0 Y_i^2 \right).$$

Since $\operatorname{var}_0 Y_i^2 = 2/n^2 + 4\kappa_i^2 \mu_{0,i}^2/n$, it suffices to show that both

$$n^2 \sup_{\alpha \in [\alpha_1, \alpha_2]} (1 + 2\alpha + 2p)^2$$

$$\times \left(\sum_{i=1}^{\infty} \frac{i^{2+4\alpha+4p}(\log i)^2 \left(1 \vee \log i + (\log n)/(1 + 2\alpha + 2p)^2\right)^2}{n^{2/(1+2\alpha+2p)}(i^{1+2\alpha+2p} + n)^4} \right) \quad (5.13)$$

and

$$n^3 \sup_{\alpha \in [\alpha_1, \alpha_2]} (1 + 2\alpha + 2p)^2$$

$$\times \left(\sum_{i=1}^{\infty} \frac{i^{2+4\alpha+2p}(\log i)^2 \mu_{0,i}^2 \left(1 \vee \log i + (\log n)/(1 + 2\alpha + 2p)^2\right)^2}{n^{2/(1+2\alpha+2p)}(i^{1+2\alpha+2p} + n)^4} \right) \quad (5.14)$$

are bounded by a constant times $(\log n)^4 \sup_{\alpha \in [\alpha_1, \alpha_2]} n^{-1/(1+2\alpha+2p)} (1 + h_n(\alpha))$.

By applying Lemma 10 twice (once the first statement with $r = 1 + 2\alpha + 2p$ and $m = 1$ and once the second one with the same r and $m = 3$ and $\xi = 1$) the expression in (5.14) is seen to be bounded above by a constant times

$$(\log n)^3 \sup_{\alpha \in [\alpha_1, \alpha_2]} \left(n^{-2/(1+2\alpha+2p)} (1 + 2\alpha + 2p) \sum_{i=1}^{\infty} \frac{n^2 i^{1+2\alpha} \mu_{0,i}^2 \log i}{(i^{1+2\alpha+2p} + n)^2} \right).$$

The expression in the parentheses equals $h_n(\alpha) n^{-1/(1+2\alpha+2p)} \log n$. Now fix $c > 0$. Again, applying Lemma 10 twice implies that we get that (5.13) is bounded above by

$$(\log n)^3 \sup_{\alpha \in [\alpha_1, \alpha_2]} \left(\frac{2n^{-2/(1+2\alpha+2p)}}{1 + 2\alpha + 2p} \sum_{i=1}^{\infty} \frac{n i^{1+2\alpha+2p} \log i}{(i^{1+2\alpha+2p} + n)^2} \right).$$

Using the inequality $x/(x + y) \le 1$ and Lemma 7(i), the expression in the parenthesis can be bounded by a constant times $n^{-1/(1+2\alpha+2p)} \log n$ for $\alpha > c$. For $\alpha \le c$, Lemma 8 (with $m = 2$ or $m = 4$, $l = 4$, $r = 1 + 2\alpha + 2p$, $r_0 = 1 + 2c + 2p$, and $s = 2r$) gives the same bound (or even a better one) for (5.13). The proof is completed by combining the obtained bounds.

6 Proof of Theorem 2

We only present the details of the proof for the Sobolev case $\mu_0 \in S^\beta$. The analytic case differs from the Sobolev case mainly in the upper bound for $n^{-2\underline{\alpha}_n/(1+2\underline{\alpha}_n+2p)}$, see also Sect. 6.5. Again, we assume the exact equality $\kappa_i = i^{-p}$ for simplicity.

By Markov's inequality and Theorem 1,

$$
\begin{aligned}
\sup_{\|\mu_0\|_\beta \leq R} & \mathrm{E}_0 \Pi_{\hat{\alpha}_n} \big(\|\mu - \mu_0\| \geq M_n \varepsilon_n \,\big|\, Y \big) \\
&\leq \frac{1}{M_n^2 \varepsilon_n^2} \sup_{\|\mu_0\|_\beta \leq R} \mathrm{E}_0 \sup_{\alpha \in [\underline{\alpha}_n, \overline{\alpha}_n \wedge \log n]} R_n(\alpha) + o(1),
\end{aligned}
\tag{6.1}
$$

where

$$
R_n(\alpha) = \int \|\mu - \mu_0\|^2 \, \Pi_\alpha(d\mu|Y)
$$

is the posterior risk. We will show in the subsequent subsections that for $\varepsilon_n = n^{-\beta/(1+2\beta+2p)}(\log n)^2(\log\log n)^{1/2}$ and arbitrary $M_n \to \infty$, the first term on the right of (6.1) vanishes as $n \to \infty$. Note that by the explicit posterior computation (2.3), we have

$$
R_n(\alpha) = \sum_{i=1}^\infty (\hat{\mu}_{\alpha,i} - \mu_{0,i})^2 + \sum_{i=1}^\infty \frac{i^{2p}}{i^{1+2\alpha+2p} + n},
\tag{6.2}
$$

where $\hat{\mu}_{\alpha,i} = n i^p (i^{1+2\alpha+2p} + n)^{-1} Y_i$ is the ith coefficient of the posterior mean. We divide the Sobolev-ball $\|\mu_0\|_\beta \leq R$ into two subsets

$$
\begin{aligned}
P_n &= \{\mu_0 : \|\mu_0\|_\beta \leq R, \ \overline{\alpha}_n \leq (\log n)/\log 2 - 1/2 - p\}, \\
Q_n &= \{\mu_0 : \|\mu_0\|_\beta \leq R, \ \overline{\alpha}_n > (\log n)/\log 2 - 1/2 - p\},
\end{aligned}
$$

and show that on both subsets the posterior risks are of the order ε_n^2.

6.1 Bound for the expected posterior risk over P_n

In this section we prove that

$$
\sup_{\mu_0 \in P_n} \sup_{\alpha \in [\underline{\alpha}_n, \overline{\alpha}_n]} \mathrm{E}_0 R_n(\alpha) = O(\varepsilon_n^2).
\tag{6.3}
$$

The second term of (6.2) is deterministic. The expectation of the first term can be split into square bias and variance terms. We find that the expectation of (6.2) is given by

$$
\sum_{i=1}^\infty \frac{i^{2+4\alpha+4p} \mu_{0,i}^2}{(i^{1+2\alpha+2p} + n)^2} + n \sum_{i=1}^\infty \frac{i^{2p}}{(i^{1+2\alpha+2p} + n)^2} + \sum_{i=1}^\infty \frac{i^{2p}}{i^{1+2\alpha+2p} + n}.
\tag{6.4}
$$

Note that the second and third terms in (6.4) are independent of μ_0, and that the second is bounded by the third. By Lemma 8 (with $m = 0$, $l = 1$, $r = 1 + 2\alpha + 2p$ and $s = 2p$) the latter is for $\alpha \geq \underline{\alpha}_n$ further bounded by

$$n^{-\frac{2\alpha}{1+2\alpha+2p}} \leq n^{-\frac{2\underline{\alpha}_n}{1+2\underline{\alpha}_n+2p}}.$$

In view of Lemma 1 (i), the right-hand side is bounded by a constant times $n^{-2\beta/(1+2\beta+2p)}$ for large n.

It remains to consider the first sum in (6.4), which we divide into three parts and show that each of the parts has the stated order. First we note that

$$\sum_{i>n^{1/(1+2\beta+2p)}} \frac{i^{2+4\alpha+4p}\mu_{0,i}^2}{(i^{1+2\alpha+2p}+n)^2} \leq \sum_{i>n^{1/(1+2\beta+2p)}} \mu_{0,i}^2 \leq \|\mu_0\|_\beta^2 n^{-2\beta/(1+2\beta+2p)}. \quad (6.5)$$

Next, observe that elementary calculus shows that for $\alpha > 0$ and $n \geq e$, the maximum of the function $i \mapsto i^{1+2\alpha+4p}/\log i$ over the interval $[2, n^{1/(1+2\alpha+2p)}]$ is attained at $i = n^{1/(1+2\alpha+2p)}$, for $\alpha \leq \log n/(2\log 2) - 1/2 - p$. It follows that for $\alpha > 0$,

$$\sum_{i\leq n^{1/(1+2\alpha+2p)}} \frac{i^{2+4\alpha+4p}\mu_{0,i}^2}{(i^{1+2\alpha+2p}+n)^2}$$

$$= \frac{\mu_{0,1}^2}{(1+n)^2} + \frac{1}{n^2} \sum_{2\leq i\leq n^{1/(1+2\alpha+2p)}} \frac{((i^{1+2\alpha+4p})/\log i)n^2 i^{1+2\alpha}\mu_{0,i}^2 \log i}{(i^{1+2\alpha+2p}+n)^2}$$

$$\leq \frac{\mu_{0,1}^2}{(1+n)^2} + n^{-\frac{2\alpha}{1+2\alpha+2p}} h_n(\alpha).$$

We note that for $\alpha > \log n/(2\log 2) - 1/2 - p$ the second term on the right hand side of the preceding display disappears and for $\mu_0 \in P_n$ we have that $\overline{\alpha}_n$ is finite. Since $n^{1/(1+2\overline{\alpha}_n+2p)} \leq n^{1/(1+2\alpha+2p)}$ for $\alpha \leq \overline{\alpha}_n$, the preceding implies that

$$\sup_{\mu_0\in P_n} \sup_{\alpha\in[\underline{\alpha}_n,\overline{\alpha}_n]} \sum_{i\leq n^{1/(1+2\overline{\alpha}_n+2p)}} \frac{i^{2+4\alpha+4p}\mu_{0,i}^2}{(i^{1+2\alpha+2p}+n)^2} \lesssim \frac{R^2}{n^2} + Ln^{-\frac{2\overline{\alpha}_n}{1+2\overline{\alpha}_n+2p}}\log^2 n.$$

By Lemma 1, $\underline{\alpha}_n \geq \beta - c_0/\log n$ for a constant $c_0 > 0$ (only depending on β, R, p). Hence, using that $x \mapsto x/(c+x)$ is increasing for every $c > 0$ the right-hand side is bounded by a constant times $n^{-2\beta/(1+2\beta+2p)}\log^2 n$.

To complete the proof we deal with the terms between $n^{1/(1+2\overline{\alpha}_n+2p)}$ and $n^{1/(1+2\beta+2p)}$. Let $J = J(n)$ be the smallest integer such that $\overline{\alpha}_n/(1+1/\log n)^J \leq \beta$. One can see that J is bounded above by a multiple of $(\log n)(\log\log n)$ for any positive β. We partition the summation range under consideration into J pieces using the auxiliary numbers

$$b_j = 1 + 2\frac{\overline{\alpha}_n}{(1+1/\log n)^j} + 2p, \quad j = 0, \ldots, J.$$

Note that the sequence b_j is decreasing. Now we have

$$\sum_{i=n^{1/(1+2\overline{\alpha}_n+2p)}}^{n^{1/(1+2\beta+2p)}} \frac{i^{2+4\alpha+4p}\mu_{0,i}^2}{(i^{1+2\alpha+2p}+n)^2} \leq \sum_{j=0}^{J-1}\sum_{i=n^{1/b_j}}^{n^{1/b_{j+1}}} \mu_{0,i}^2 \leq 4\sum_{j=0}^{J-1}\sum_{i=n^{1/b_j}}^{n^{1/b_{j+1}}} \frac{ni^{b_j}\mu_{0,i}^2}{(i^{b_{j+1}}+n)^2},$$

and the upper bound is uniform in α. Since $(b_j - b_{j+1})\log n = b_{j+1} - 1 - 2p$, it holds for $n^{1/b_j} \leq i \leq n^{1/b_{j+1}}$ that $i^{b_j-b_{j+1}} \leq n^{1/\log n} = e$. On the same interval i^{2p} is bounded by $n^{2p/b_{j+1}}$. Therefore the right hand side of the preceding display is further bounded by a constant times

$$\sum_{j=0}^{J-1}\sum_{i=n^{1/b_j}}^{n^{1/b_{j+1}}} \frac{ni^{b_{j+1}}\mu_{0,i}^2 \log i}{(i^{b_{j+1}}+n)^2} \leq \sum_{j=0}^{J-1} n^{2p/b_{j+1}-1} \sum_{i=n^{1/b_j}}^{n^{1/b_{j+1}}} \frac{n^2 i^{b_{j+1}-2p}\mu_{0,i}^2 \log i}{(i^{b_{j+1}}+n)^2}$$

$$\leq \sum_{j=0}^{J-1} n^{2p/b_{j+1}-1} h_n\left(\frac{\overline{\alpha}_n}{(1+1/\log n)^{j+1}}\right) n^{1/b_{j+1}}\frac{\log n}{b_{j+1}}$$

$$\leq (\log n) \sum_{j=0}^{J-1} n^{(1+2p-b_{j+1})/b_{j+1}} h_n(b_{j+1}/2 - 1/2 - p)$$

$$\leq (\log n) n^{-\frac{2\beta/(1+1/\log n)}{1+2\beta/(1+1/\log n)+2p}} \sum_{j=0}^{J-1} h_n(b_{j+1}/2 - 1/2 - p).$$

In the last step we used the fact that by construction, $b_j/2-1/2-p \geq \beta/(1+1/\log n)$ for $j \leq J$. Because $b_j/2 - 1/2 - p \leq \overline{\alpha}_n$ for every $j \geq 0$, it follows from the definition of $\overline{\alpha}_n$ that $h_n(b_j/2-1/2-p)$ is bounded above by $L(\log n)^2$, and we recall that $J = J(n)$ is bounded above by a multiple of $(\log n)(\log\log n)$. Finally we note that

$$n^{-\frac{2\beta/(1+1/\log n)}{1+2\beta/(1+1/\log n)+2p}} \leq en^{-2\beta/(1+2\beta+2p)}.$$

Therefore the first sum in (6.4) over the range $[n^{1/(1+2\overline{\alpha}_n+2p)}, n^{1/(1+2\beta+2p)}]$ is bounded above by a multiple of $n^{-2\beta/(1+2\beta+2p)}(\log n)^4(\log\log n)$, in the appropriate uniform sense over P_n. Putting the bounds above together we conclude (6.3).

6.2 Bound for the centered posterior risk over P_n

We show in this section that for the set P_n we also have

$$\sup_{\mu_0 \in P_n} \mathrm{E}_0 \sup_{\alpha \in [\underline{\alpha}_n, \overline{\alpha}_n]} \left| \sum_{i=1}^{\infty} \left(\hat{\mu}_{\alpha,i} - \mu_{0,i}\right)^2 - \mathrm{E}_0 \sum_{i=1}^{\infty} \left(\hat{\mu}_{\alpha,i} - \mu_{0,i}\right)^2 \right| = O(\varepsilon_n^2),$$

for $\varepsilon_n = n^{-\beta/(1+2\beta+2p)}(\log n)^2(\log\log n)^{1/2}$. Using the explicit expression for the posterior mean $\hat{\mu}_{\alpha,i}$ we see that the random variable in the supremum is the absolute value of $\mathbb{V}(\alpha)/n - 2\mathbb{W}(\alpha)/\sqrt{n}$, where

$$\mathbb{V}(\alpha) = \sum_{i=1}^{\infty} \frac{n^2\kappa_i^{-2}}{(i^{1+2\alpha}\kappa_i^{-2}+n)^2}(Z_i^2-1), \qquad \mathbb{W}(\alpha) = \sum_{i=1}^{\infty} \frac{ni^{1+2\alpha}\kappa_i^{-3}\mu_{0,i}}{(i^{1+2\alpha}\kappa_i^{-2}+n)^2}Z_i.$$

We deal with the two processes separately.

For the process $\mathbb{V}$, Corollary 2.2.5 in [39] implies that

$$E_0 \sup_{\alpha\in[\underline{\alpha}_n,\infty)} |\mathbb{V}(\alpha)| \lesssim \sup_{\alpha\in[\underline{\alpha}_n,\infty)} \sqrt{\text{var}_0\,\mathbb{V}(\alpha)} + \int_0^{\text{diam}_n} \sqrt{N(\varepsilon,[\underline{\alpha}_n,\infty),d_n)}\,d\alpha,$$

where $d_n^2(\alpha_1,\alpha_2) = \text{var}_0(\mathbb{V}(\alpha_1)-\mathbb{V}(\alpha_2))$ and diam_n is the d_n-diameter of $[\underline{\alpha}_n,\infty)$. Now the variance of $\mathbb{V}(\alpha)$ is equal to

$$\text{var}_0\,\mathbb{V}(\alpha) = 2n^4\sum_{i=1}^{\infty} \frac{i^{4p}}{(i^{1+2\alpha+2p}+n)^4},$$

since $\text{var}_0\,Z_i^2 = 2$. Using Lemma 8 (with $m = 0$, $l = 4$, $r = 1+2\alpha+2p$ and $s = 4p$), we can conclude that the variance of $\mathbb{V}(\alpha)$ is bounded above by a multiple of $n^{(1+4p)/(1+2\alpha+2p)}$. It follows that the diameter of the interval $\text{diam}_n \lesssim n^{(1+4p)/(1+2\underline{\alpha}_n+2p)}$. To compute the covering number of the interval $[\underline{\alpha}_n,\infty)$ we first note that for $0 < \alpha_1 < \alpha_2$,

$$\text{var}_0\,(\mathbb{V}(\alpha_1)-\mathbb{V}(\alpha_2)) = \sum_{i=2}^{\infty}\left(\frac{n^2i^{2p}}{(i^{1+2\alpha_1+2p}+n)^2} - \frac{n^2i^{2p}}{(i^{1+2\alpha_2+2p}+n)^2}\right)^2 \text{var}\,Z_i^2$$

$$\leq 2\sum_{i=2}^{\infty} \frac{n^4i^{4p}}{(i^{1+2\alpha_1+2p}+n)^4} \leq 2n^4\sum_{i=2}^{\infty} i^{-4-8\alpha_1-4p} \lesssim n^42^{-8\alpha_1}.$$

Hence for $\varepsilon > 0$, a single ε-ball covers the whole interval $[K\log(n/\varepsilon),\infty)$ for some constant $K > 0$. By Lemma 5, the distance $d_n(\alpha_1,\alpha_2)$ is bounded above by a multiple of $|\alpha_1 - \alpha_2|n^{(1+4p)/(2+4\underline{\alpha}_n+4p)}(\log n)$. Therefore the covering number of the interval $[\underline{\alpha}_n, K\log(n/\varepsilon)]$ relative to the metric d_n is bounded above by a multiple of $(\log n)n^{(1+4p)/(2+4\underline{\alpha}_n+4p)}(\log(n/\varepsilon))/\varepsilon$. Combining everything we see that

$$E_0 \sup_{\alpha\in[\underline{\alpha}_n,\infty)} |\mathbb{V}(\alpha)| \lesssim n^{\frac{1+4p}{2+4\underline{\alpha}_n+4p}}(\log n).$$

By the fact that $x \mapsto x/(x+c)$ is increasing and Lemma 1 (i), the right-hand side divided by n is bounded by

$$n^{-\frac{2\underline{\alpha}_n}{1+2\underline{\alpha}_n+2p}}(\log n) \lesssim n^{-2\beta/(1+2\beta+2p)}(\log n).$$

It remains to deal with the process $\mathbb{W}$. The basic line of reasoning is the same as followed above for $\mathbb{V}$. An essential difference however is the derivation of a bound for the variance of $\mathbb{W}$, of which we provide the details. The rest of the proof is left to the reader. The variance $\mathbb{W}(\alpha)/\sqrt{n}$ is given by

$$\text{var}_0\left(\frac{\mathbb{W}(\alpha)}{\sqrt{n}}\right) = \sum_{i=1}^{\infty} \frac{ni^{2+4\alpha+6p}\mu_{0,i}^2}{(i^{1+2\alpha+2p}+n)^4}.$$

We show that uniformly for $\alpha \in [\underline{\alpha}_n, \overline{\alpha}_n]$, this variance is bounded above by a constant (which depends only on $\|\mu_0\|_\beta$) times $n^{-(1+4\beta)/(1+2\beta+2p)}(\log n)^2$. We note that on the set P_n the upper bound $\overline{\alpha}_n \leq \log n/\log 2 - 1/2 - p$ is finite.

For the sum over $i \leq n^{1/(1+2\alpha+2p)}$ we have

$$\sum_{i \leq n^{1/(1+2\alpha+2p)}} \frac{ni^{2+4\alpha+6p}\mu_{0,i}^2}{(i^{1+2\alpha+2p}+n)^4}$$

$$\leq \frac{\mu_{0,1}^2}{n^3} + \frac{1}{n^3} \sum_{2 \leq i \leq n^{1/(1+2\alpha+2p)}} \frac{n^2 i^{1+2\alpha+6p}(\log i)^{-1}i^{1+2\alpha}\mu_{0,i}^2\log i}{(i^{1+2\alpha+2p}+n)^2}$$

$$\leq \frac{\|\mu_0\|_\beta^2}{n^3} + (1+2\alpha+2p)\frac{n^{4p/(1+2\alpha+2p)}}{(\log n)n^2} \sum_{i \leq n^{1/(1+2\alpha+2p)}} \frac{n^2 i^{1+2\alpha}\mu_{0,i}^2\log i}{(i^{1+2\alpha+2p}+n)^2}$$

$$\leq \frac{\|\mu_0\|_\beta^2}{n^3} + n^{-\frac{1+4\alpha}{1+2\alpha+2p}}h_n(\alpha). \tag{6.6}$$

We note that the second term on the right hand side of the preceding display disappears for $\alpha > \log n/(2\log 2) - 1/2 - p$. We have used again the fact that on the range $i \leq n^{1/(1+2\alpha+2p)}$, the quantity $i^{1+2\alpha+6p}(\log i)^{-1}$ is maximal for the largest i. Now the function $x \mapsto -(1+2x)/(x+c)$ is decreasing on $(0,\infty)$ for any $c > 1/2$. Moreover $h_n(\alpha) \leq L(\log n)^2$ for any $\alpha \leq \overline{\alpha}_n$, thus the preceding display is bounded above by a multiple of $n^{-(1+4\alpha_n)/(1+2\alpha_n+2p)}(\log n)^2$. Using Lemma 1(i) this is further bounded by a constant times $n^{-(1+4\beta)/(1+2\beta+2p)}(\log n)^2$.

Next we consider sum over the range $i > n^{1/(1+2\alpha+2p)}$. We distinguish two cases according to the value of α. First suppose that $1+2\alpha \geq 2p$. Then $i^{-1-2\alpha+2p}(\log i)^{-1}$ is decreasing in i, hence

$$\sum_{i > n^{1/(1+2\alpha+2p)}} \frac{ni^{2+4\alpha+6p}\mu_{0,i}^2}{(i^{1+2\alpha+2p}+n)^4}$$

$$\leq \frac{1}{n} \sum_{i > n^{1/(1+2\alpha+2p)}} \frac{n^2 i^{-1-2\alpha+2p}(\log i)^{-1}i^{1+2\alpha}\mu_{0,i}^2\log i}{(i^{1+2\alpha+2p}+n)^2}$$

$$\leq \frac{1+2\alpha+2p}{n^{(2+4\alpha)/(1+2\alpha+2p)}\log n} \sum_{i > n^{1/(1+2\alpha+2p)}} \frac{n^2 i^{1+2\alpha}\mu_{0,i}^2\log i}{(i^{1+2\alpha+2p}+n)^2}$$

$$\leq n^{-\frac{1+4\alpha}{1+2\alpha+2p}}h_n(\alpha).$$

As above, this is further bounded by a constant times the desired rate $n^{-(1+4\beta)/(1+2\beta+2p)}$ $(\log n)^2$. If $1 + 2\alpha < 2p$, then

$$\sum_{i>n^{1/(1+2\alpha+2p)}} \frac{ni^{2+4\alpha+6p}\mu_{0,i}^2}{(i^{1+2\alpha+2p}+n)^4} \le n \sum_{i>n^{1/(1+2\alpha+2p)}} i^{-2-4\alpha-2p-2\beta}i^{2\beta}\mu_{0,i}^2$$

$$\le \|\mu_0\|_\beta^2 n^{\frac{2p-2\beta}{1+2\alpha+2p}-1}.$$

Since $\underline{\alpha}_n \ge \beta - c_0/\log n$, we have $1 + 2\alpha > 2\beta$ for large enough n, for any $\alpha \in [\underline{\alpha}_n, \overline{\alpha}_n]$. Since we have assumed $1 + 2\alpha < 2p$, this implies that $2p > 2\beta$. Therefore the right hand side of the preceding display attains its maximum at $\alpha = \underline{\alpha}_n$. Using again that $\underline{\alpha}_n \ge \beta - c_0/\log n$, it is straightforward to show that for $\alpha \in [\underline{\alpha}_n, \overline{\alpha}_n]$,

$$n^{\frac{2p-2\beta}{1+2\alpha+2p}-1} \le n^{\frac{2p-2\beta}{1+2\underline{\alpha}_n+2p}-1} \le e^{4c_0}n^{-\frac{1+4\beta}{1+2\beta+2p}}.$$

6.3 Bound for the expected and centered posterior risk over Q_n

To complete the proof of Theorem 2 we show that similar results to Sects. 6.1 and 6.2 hold over the set Q_n as well:

$$\sup_{\mu_0 \in Q_n} \sup_{\alpha \in [\underline{\alpha}_n, \infty)} E_0 R_n(\alpha) = O(\varepsilon_n^2), \tag{6.7}$$

$$\sup_{\mu_0 \in Q_n} E_0 \sup_{\alpha \in [\underline{\alpha}_n, \infty)} \left| \sum_{i=1}^\infty (\hat{\mu}_{\alpha,i} - \mu_{0,i})^2 - E_0 \sum_{i=1}^\infty (\hat{\mu}_{\alpha,i} - \mu_{0,i})^2 \right| = O(\varepsilon_n^2). \tag{6.8}$$

For the first statement (6.7) we follow the same line of reasoning as in Sect. 6.1. The second and third terms in (6.4) are free of μ_0, and hence the same upper bound as in Sect. 6.1 apply. The first term in (6.4) is also treated exactly as in Sect. 6.1, except that $n^{1/(1+2\overline{\alpha}_n+2p)} \le 2$ if $\mu_0 \in Q_n$ and hence the sum over the terms $i < n^{1/(1+2\overline{\alpha}_n+2p}$ need not be treated, and we can proceed by replacing $\overline{\alpha}_n$ by $\log n/(2\log 2) - 1/2 - p$ in the definition of J and the sequence b_j.

To bound the centered posterior risk (6.8) we follow the proof given in Sect. 6.2. There the process $\mathbb{V}(\alpha)$ is already bounded uniformly over $[\underline{\alpha}_n, \infty)$, whence it remains to deal with the process $\mathbb{W}(\alpha)$. The only essential difference is the upper bound for the variance of the process $\mathbb{W}(\alpha)/\sqrt{n}$. In Sect. 6.2 this was shown to be bounded above by a multiple of the desired rate $(\log n)^2 n^{-(1+4\beta)/(1+2\beta+2p)}$ for $\alpha \in [\underline{\alpha}_n, \overline{\alpha}_n \wedge (\log n/\log 2 - 1/2 - p)]$, which is $\alpha \in [\underline{\alpha}_n, \log n/\log 2 - 1/2 - p]$ on the set Q_n. Finally, for $\alpha \ge \log n/\log 2 - 1/2 - p$ we have

$$\sum_{i=1}^\infty \frac{ni^{2+4\alpha+6p}\mu_{0,i}^2}{(i^{1+2\alpha+2p}+n)^4} \le \frac{\mu_{0,1}^2}{n^3} + \sum_{i=2}^\infty \frac{ni^{-1-2\alpha}\mu_{0,i}^2}{i^{1+2\alpha+2p}+n}$$

$$\le \frac{\|\mu_0\|_\beta^2}{n^3} + \sum_{i=2}^\infty i^{-1-2\alpha-2\beta}i^{2\beta}\mu_{0,i}^2$$

$$\leq \frac{\|\mu_0\|_\beta^2}{n^3} + 2^{-1-2\alpha}\|\mu_0\|_\beta^2 \leq \frac{\|\mu_0\|_\beta^2}{n^3} + 2^{2p}\frac{\|\mu_0\|_\beta^2}{n^2}$$
$$\lesssim n^{-2}. \tag{6.9}$$

This completes the proof.

6.4 Bounds for the semimetrics associated to $\mathbb{V}$ and $\mathbb{W}$

The following lemma is used in Sect. 6.2.

Lemma 5 *For any $\underline{\alpha}_n \leq \alpha_1 < \alpha_2 \leq \overline{\alpha}_n$ the following inequalities hold:*

$$\mathrm{var}_0\left(\mathbb{V}(\alpha_1) - \mathbb{V}(\alpha_2)\right) \lesssim (\alpha_1 - \alpha_2)^2 n^{(1+4p)/(1+2\underline{\alpha}_n+2p)}(\log n)^2,$$
$$\mathrm{var}_0\left(\frac{\mathbb{W}(\alpha_1)}{\sqrt{n}} - \frac{\mathbb{W}(\alpha_2)}{\sqrt{n}}\right) \lesssim (\alpha_1 - \alpha_2)^2 n^{-\frac{1+4\overline{\alpha}_n}{1+2\overline{\alpha}_n+2p}}(\log n)^4,$$

with a constant that does not depend on α and μ_0.

Proof The left-hand side of the first inequality is equal to

$$n^4 \sum_{i=1}^{\infty}(f_i(\alpha_1) - f_i(\alpha_2))^2 i^{4p}\, \mathrm{var}\, Z_i^2,$$

where $f_i(\alpha) = (i^{1+2\alpha+2p} + n)^{-2}$. The derivative of f_i is given by $f_i'(\alpha) = -4i^{1+2\alpha+2p}(\log i)/(i^{1+2\alpha+2p} + n)^3$, hence the preceding display is bounded above by a multiple of

$$(\alpha_1 - \alpha_2)^2 n^4 \sup_{\alpha\in[\alpha_1,\alpha_2]} \sum_{i=1}^{\infty} \frac{i^{2+4\alpha+8p}(\log i)^2}{(i^{1+2\alpha+2p} + n)^6}$$
$$\leq (\alpha_1 - \alpha_2)^2 n^3 (\log n)^2 \sup_{\alpha\in[\alpha_1,\alpha_2]} \frac{1}{(1+2\alpha+2p)^2} \sum_{i=1}^{\infty} \frac{i^{1+2\alpha+6p}}{(i^{1+2\alpha+2p} + n)^4}$$
$$\lesssim (\alpha_1 - \alpha_2)^2 (\log n)^2 \sup_{\alpha\in[\alpha_1,\alpha_2]} n^{(1+4p)/(1+2\alpha+2p)},$$

with the help of Lemma 10 (with $r = 1 + 2\alpha + 2p$, and $m = 2$), and Lemma 8 (with $m = 0, l = 4, r = 1 + 2\alpha + 2p$, and $s = r + 4p$). Since $\alpha \geq \underline{\alpha}_n$, we get the first assertion of the lemma.

We next consider $\mathbb{W}/\sqrt{n}$. The left-hand side of the second inequality in the statement of the lemma is equal to

$$\sum_{i=1}^{\infty}(f_i(\alpha_1) - f_i(\alpha_2))^2 n\mu_{0,i}^2\, \mathrm{var}\, Z_i,$$

where now $f_i(\alpha) = i^{1+2\alpha+3p}/(i^{1+2\alpha+2p} + n)^2$. The derivative of this f_i satisfies $|f_i'(\alpha)| \le 2(\log i) f_i(\alpha)$, hence we get the upper bound

$$4(\alpha_2 - \alpha_1)^2 \sup_{\alpha \in [\alpha_1, \alpha_2]} \sum_{i=1}^{\infty} \frac{ni^{2+4\alpha+6p} \mu_{0,i}^2 \log^2 i}{(i^{1+2\alpha+2p} + n)^4}.$$

The proof is completed by arguing as in (6.6) or (6.9).

6.5 Proof of Theorem 2 in the case of the analytic truth

The assertion of Theorem 2 in the case of the analytic truth $\mu_0 \in A^\gamma$ can be proven along the lines of the proof presented above. In view of Lemma 1.(ii), $\sqrt{\log n}/(\log \log n) < \underline{\alpha}_n$, and whence

$$n^{-\frac{2\underline{\alpha}_n}{1+2\underline{\alpha}_n+2p}} \le n^{\frac{2\sqrt{\log n}/(\log \log n)}{1+2\sqrt{\log n}/(\log \log n)+2p}}$$

$$= n^{-1}n^{\frac{1+2p}{1+2\sqrt{\log n}/(\log \log n)+2p}} \le n^{-1}(\log n)^{(1/2+p)\sqrt{\log n}}.$$

We note that the computations in Sect. 6.1 go through for the analytic case by replacing β and $\|\mu_0\|_\beta$ by $\sqrt{\log n}/\log \log n$ and $\|\mu_0\|_{A^\gamma}$, respectively. Furthermore in Sect. 6.2 it is sufficient to consider the case $1 + 2\alpha \ge 2p$ following from Lemma 1 (ii).

7 Proof of Theorem 3

Let $\mathcal{B}(R)$ denote a Sobolev or analytic ball of radius R, and $\varepsilon_{n,\mathcal{B}}$ the corresponding contraction rate. Let A_n be the event that $\hat{\alpha}_n \in [\underline{\alpha}_n, \overline{\alpha}_n]$. Then with $\alpha \mapsto \lambda_n(\alpha|Y)$ denoting the posterior Lebesgue density of α, we have

$$\sup_{\mu_0 \in \mathcal{B}(R)} E_0 \Pi(\|\mu - \mu_0\| \ge M_n \varepsilon_{n,\mathcal{B}}|Y)$$

$$\le \sup_{\mu_0 \in \mathcal{B}(R)} P_0(A_n^c) + \sup_{\mu_0 \in \mathcal{B}(R)} E_0 \int_0^{\overline{\alpha}_n} \lambda_n(\alpha|Y) \, d\alpha \, 1_{A_n}$$

$$+ \sup_{\mu_0 \in \mathcal{B}(R)} E_0 \int_{\underline{\alpha}_n}^{\infty} \lambda_n(\alpha|Y) \Pi_\alpha(\|\mu - \mu_0\| \ge M_n \varepsilon_{n,\mathcal{B}}|Y) \, d\alpha \, 1_{A_n}. \quad (7.1)$$

By Theorem 1 the first term on the right vanishes as $n \to \infty$, provided l and L in the definitions of $\underline{\alpha}_n$ and $\overline{\alpha}_n$ are chosen small and large enough, respectively. We will show that the other terms tend to 0 as well.

Observe that $\lambda_n(\alpha|Y) \propto L_n(\alpha)\lambda(\alpha)$, where $L_n(\alpha) = \exp(\ell_n(\alpha))$, for ℓ_n the random function defined by (2.2). In Sect. 5.3 we have shown that on the interval $(0, \underline{\alpha}_n + 1/\log n]$

$$\ell'_n(\alpha) = \mathbb{M}_n(\alpha) \gtrsim \frac{n^{1/(1+2\alpha+2p)} \log n}{1 + 2\alpha + 2p},$$

on the event A_n. Therefore on the interval $(0, \underline{\alpha}_n]$ we have

$$\ell_n(\alpha) < \ell_n(\underline{\alpha}_n) \leq \ell_n \left(\underline{\alpha}_n + \frac{1}{2 \log n} \right) - \frac{K n^{1/(1+2\underline{\alpha}_n+2p)}}{1 + 2\underline{\alpha}_n + 2p}$$

for some $K > 0$, since for $n > e$ and $\alpha \in [\underline{\alpha}_n, \underline{\alpha}_n + 1/(2 \log n)])$, we have $n^{1/(1+2\alpha+2)}/(1 + 2\alpha + 2p) \gtrsim n^{1/(1+2\underline{\alpha}_n+2p)}/(1 + 2\underline{\alpha}_n + 2p)$, and on the interval $[\underline{\alpha}_n + 1/(2 \log n), \underline{\alpha}_n + 1/\log n]$,

$$\ell_n(\alpha) \geq \ell_n \left(\underline{\alpha}_n + \frac{1}{2 \log n} \right).$$

For the likelihood L_n we have the corresponding bounds

$$L_n(\alpha) < \exp \left(-\frac{K n^{1/(1+2\underline{\alpha}_n+2p)}}{1 + 2\underline{\alpha}_n + 2p} \right) L_n \left(\underline{\alpha}_n + \frac{1}{2 \log n} \right)$$

for $\alpha \in (0, \underline{\alpha}_n]$ and

$$L_n(\alpha) \geq L_n \left(\underline{\alpha}_n + \frac{1}{2 \log n} \right)$$

for $\alpha \in [\underline{\alpha}_n + 1/(2 \log n), \underline{\alpha}_n + 1/\log n]$ on the event A_n. Using these estimates for L_n we obtain the following upper bound for the second term on the right-hand side of (7.1):

$$\sup_{\mu_0 \in \mathscr{B}(R)} \mathrm{E}_0 \frac{\int_0^{\underline{\alpha}_n} \lambda(\alpha) L_n(\alpha) \, d\alpha}{\int_0^\infty \lambda(\alpha) L_n(\alpha) \, d\alpha}$$

$$\leq \sup_{\mu_0 \in \mathscr{B}(R)} \mathrm{E}_0 \exp \left(-\frac{K n^{1/(1+2\underline{\alpha}_n+2p)}}{1 + 2\underline{\alpha}_n + 2p} \right) \frac{L_n \left(\underline{\alpha}_n + \frac{1}{2 \log n} \right) \int_0^{\underline{\alpha}_n} \lambda(\alpha) \, d\alpha}{L_n \left(\underline{\alpha}_n + \frac{1}{2 \log n} \right) \int_{\underline{\alpha}_n+1/(2 \log n)}^{\underline{\alpha}_n+1/\log n} \lambda(\alpha) \, d\alpha}$$

$$\leq \sup_{\mu_0 \in \mathscr{B}(R)} \exp \left(-\frac{K n^{1/(1+2\underline{\alpha}_n+2p)}}{1 + 2\underline{\alpha}_n + 2p} \right) \left(\int_{\underline{\alpha}_n+1/(2 \log n)}^{\underline{\alpha}_n+1/\log n} \lambda(\alpha) \, d\alpha \right)^{-1}. \tag{7.2}$$

From Lemma 1 we know that $\underline{\alpha}_n \geq \beta/2$ for large enough n, hence by Assumption 1, Lemma 6, and the definition of $\underline{\alpha}_n$,

$$\int_{\underline{\alpha}_n+1/(2 \log n)}^{\underline{\alpha}_n+1/\log n} \lambda(\alpha) \, d\alpha \geq C_1 (2 \log n)^{-C_2} \exp \left(-C_3 \exp(\sqrt{\log n}/3) \right)$$

for some $C_1, C_2, C_3 > 0$. Therefore the right hand side of (7.2) is bounded above by a constant times

$$\exp\left(-\frac{Kn^{1/(1+2\sqrt{\log n}+2p)}}{1+2\sqrt{\log n}+2p}\right)(\log n)^{C_2}\exp\left(C_3\exp\left(\frac{\sqrt{\log n}}{3}\right)\right).$$

It is easy to see that this quantity tends to 0 as $n \to \infty$.

In bounding the third term on the right hand side of (7.1) we replace the supremum over $\mathscr{B}(R)$ by the suprema over the sets P_n and Q_n defined in the beginning of Sect. 6. The supremum over Q_n is bounded above by

$$\sup_{\mu_0 \in Q_n} E_0 \sup_{\alpha \in [\underline{\alpha}_n, \infty)} \Pi_\alpha(\|\mu - \mu_0\| \geq M_n \varepsilon_{n,\mathscr{B}}|Y).$$

This goes to zero, as follows from Sect. 6.3 and Markov's inequality. The supremum over P_n we write as

$$\sup_{\mu_0 \in P_n} E_0 \left(\int_{\underline{\alpha}_n}^{\overline{\alpha}_n} \lambda_n(\alpha|Y) \Pi_\alpha(\|\mu - \mu_0\| \geq M_n \varepsilon_{n,\mathscr{B}}|Y) \, d\alpha \right.$$
$$\left. + \int_{\overline{\alpha}_n}^\infty \lambda_n(\alpha|Y) \Pi_\alpha(\|\mu - \mu_0\| \geq M_n \varepsilon_{n,\mathscr{B}}|Y) \, d\alpha \right) 1_{A_n}. \qquad (7.3)$$

The first term in (7.3) is bounded above by

$$\sup_{\mu_0 \in P_n} E_0 \sup_{\alpha \in [\underline{\alpha}_n, \overline{\alpha}_n]} \Pi_\alpha(\|\mu - \mu_0\| \geq M_n \varepsilon_{n,\mathscr{B}}|Y).$$

This goes to zero, following from Sects. 6.1 and 6.2 and Markov's inequality. In Sect. 5.1 we have shown that the differentiated log-likelihood function $\mathbb{M}_n$ on the interval $[\overline{\alpha}_n, \infty)$ can increase maximally by a multiple of

$$\frac{n^{1/(1+2\overline{\alpha}_n+2p)}(\log n)^2}{1+2\overline{\alpha}_n+2p}.$$

Moreover, in Sect. 5.2 we have shown that for $\alpha \in [\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n]$,

$$\ell_n'(\alpha) = \mathbb{M}_n(\alpha) < -M\frac{n^{1/(1+2\overline{\alpha}_n+2p)}(\log n)^3}{1+2\overline{\alpha}_n+2p}$$

on the event A_n, and M can be made arbitrarily large by increasing the constant L in the definition of $\overline{\alpha}_n$. Therefore the integral of $\mathbb{M}_n(\alpha)$ on $[\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n - 1/(2\log n)]$ is bounded above by

$$-\frac{M}{2}\frac{n^{1/(1+2\overline{\alpha}_n+2p)}(\log n)^2}{1+2\overline{\alpha}_n+2p},$$

and by choosing a large enough constant L in the definition of $\overline{\alpha}_n$ it holds that for some $N > 0$,

$$\ell_n(\alpha) \leq \ell_n\left(\overline{\alpha}_n - \frac{1}{2\log n}\right) - N\frac{n^{1/(1+2\overline{\alpha}_n+2p)}(\log n)^2}{1 + 2\overline{\alpha}_n + 2p}$$

for $\alpha \in [\overline{\alpha}_n, \infty)$, and

$$\ell_n(\alpha) \geq \ell_n\left(\overline{\alpha}_n - \frac{1}{2\log n}\right)$$

for $\alpha \in [\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n - 1/(2\log n)]$. These bounds lead to the following bounds for the likelihood:

$$L_n(\alpha) \leq L_n\left(\overline{\alpha}_n - \frac{1}{2\log n}\right)\exp\left(-N\frac{n^{1/(1+2\overline{\alpha}_n+2p)}(\log n)^2}{1 + 2\overline{\alpha}_n + 2p}\right)$$

for $\alpha \in [\overline{\alpha}_n, \infty)$, and

$$L_n(\alpha) \geq L_n\left(\overline{\alpha}_n - \frac{1}{2\log n}\right)$$

for $\alpha \in [\overline{\alpha}_n - 1/\log n, \overline{\alpha}_n - 1/(2\log n)]$. Similarly to the upper bound for the second term of (7.1) we now write

$$\sup_{\mu_0 \in P_n} \mathrm{E}_0 \int_{\overline{\alpha}_n}^{\infty} \lambda_n(\alpha|Y)\,d\alpha \leq \sup_{\mu_0 \in P_n} \mathrm{E}_0 \frac{\int_{\overline{\alpha}_n}^{\infty} \lambda(\alpha)L_n(\alpha)\,d\alpha}{\int_0^{\infty} \lambda(\alpha)L_n(\alpha)\,d\alpha}$$

$$\leq \sup_{\mu_0 \in P_n} \exp\left(-N\frac{n^{1/(1+2\overline{\alpha}_n+2p)}(\log n)^2}{1 + 2\overline{\alpha}_n + 2p}\right)\frac{\int_{\overline{\alpha}_n}^{\infty} \lambda(\alpha)\,d\alpha}{\int_{\overline{\alpha}_n - 1/\log n}^{\overline{\alpha}_n - 1/(2\log n)} \lambda(\alpha)\,d\alpha}.$$

Since $\overline{\alpha}_n \geq \underline{\alpha}_n \geq \beta/2$ for n large enough, Assumption 1 and Lemma 6 imply that

$$\frac{\int_{\overline{\alpha}_n}^{\infty} \lambda(\alpha)\,d\alpha}{\int_{\overline{\alpha}_n - 1/\log n}^{\overline{\alpha}_n - 1/(2\log n)} \lambda(\alpha)\,d\alpha} \leq C_4(\log n)^{C_5}\exp\left(C_6\overline{\alpha}_n^{C_7}\right).$$

Since $\overline{\alpha}_n \leq \log n/(2\log 2) - 1/2 - p$ for $\mu_0 \in P_n$, the right-hand side of the preceding display is bounded above by

$$C_4\exp\left(-2C_9(\log 2)(\log n)\right)(\log n)^{C_5}\exp\left(C_6\left(\frac{\log n}{2\log 2} - \frac{1}{2} - p\right)^{C_7}\right),$$

which tends to zero for any fixed constant C_7 smaller than 1.

Lemma 6 *Suppose that the prior density λ satisfies Assumption 1 for some $c_1 > 0$. Then there exist positive constants $C_1, \ldots, C_6$ and $C_7 < 1$ depending on c_1 only such that for all $x \geq c_1$, every $\delta_n \to 0$, and n large enough*

$$\int_{x+\delta_n}^{x+2\delta_n} \lambda(\alpha)\,d\alpha \geq C_1\delta_n^{C_2} \exp\left(-C_3 \exp\left(\frac{x}{3}\right)\right)$$

and

$$\frac{\int_x^\infty \lambda(\alpha)\,d\alpha}{\int_{x-2\delta_n}^{x-\delta_n} \lambda(\alpha)\,d\alpha} \leq C_4\delta_n^{-C_5} \exp(C_6 x^{C_7}).$$

Proof The proof only involves straightforward calculus.

8 Auxiliary lemmas

In this section we collect several lemmas that we use throughout the proofs to upper and lower bound certain sums.

Lemma 7 *Let $c > 0$ and $r \geq 1 + c$.*

(i) *For $n \geq 1$*

$$\sum_{i=1}^\infty \frac{n\log i}{i^r + n} \leq \left(2 + \frac{2}{c} + \frac{2}{c^2 \log 2}\right)\frac{n^{1/r}\log n}{r}.$$

(ii) *If $r > (\log n)/(\log 2)$, then for $n \geq 1$*

$$\sum_{i=1}^\infty \frac{n\log i}{i^r + n} \leq \left(1 + \frac{2}{c} + \frac{2}{c^2 \log 2}\right)(\log 2)n2^{-r}.$$

Proof First consider $r \leq (\log n)/(\log 2)$, which implies that $n^{1/r} \geq 2$. We split the series in two parts, and bound the denominator $i^r + n$ by n or i^r. Since $\log i$ is increasing, we see that

$$\sum_{i=1}^{\lfloor n^{1/r}\rfloor} \log i \leq \frac{n^{1/r}\log n}{r}.$$

Since $f(x) = x^{-\gamma}\log x$ is decreasing for $x \geq e^{1/\gamma}$, we see that $i^{-r}\log i$ is decreasing on interval $\left[\lceil n^{1/r}\rceil, \infty\right)$ for $n \geq e$. Therefore

$$\sum_{i=\lceil n^{1/r}\rceil}^\infty \frac{n\log i}{i^r} \leq n\frac{\log\lceil n^{1/r}\rceil}{\lceil n^{1/r}\rceil^r} + n\int_{\lceil n^{1/r}\rceil}^\infty \frac{\log x}{x^r}\,dx.$$

Since $\lceil x \rceil / x \leq 2$ for $x \geq 1$, and $n^{1/r} \geq 2$,

$$n \frac{\log \lceil n^{1/r} \rceil}{\lceil n^{1/r} \rceil^r} \leq 2 \log n^{1/r} \leq \frac{n^{1/r} \log n}{r}.$$

Moreover

$$\int_{\lceil n^{1/r} \rceil}^{\infty} \frac{\log x}{x^r} \, dx \leq \int_{n^{1/r}}^{\infty} \frac{\log x}{x^r} \, dx = n^{-1+1/r} \frac{(r-1) \log n^{1/r} + 1}{(r-1)^2}.$$

Since $r \geq 1 + c$, we have

$$\frac{\log n^{1/r}}{r-1} \leq \frac{1}{c} \cdot \frac{\log n}{r}, \qquad \frac{1}{(r-1)^2} \leq \frac{\log n^{1/r}}{(r-1)^2 \log 2} \leq \frac{1}{c^2 \log 2} \cdot \frac{\log n}{r}.$$

This proves (i) for the case $r \leq (\log n)/(\log 2)$.

We now consider $r > (\log n)/(\log 2)$, which implies that $n^{1/r} < 2$. We have

$$\sum_{i=2}^{\infty} \frac{n \log i}{i^r + n} \leq n \sum_{i=2}^{\infty} \frac{\log i}{i^r} \leq n2^{-r} \log 2 + n \int_2^{\infty} x^{-r} \log x \, dx,$$

by monotonicity of the function f defined above (with $\gamma = r$). We have

$$\int_2^{\infty} x^{-r} \log x \, dx = 2^{1-r} \frac{(r-1) \log 2 + 1}{(r-1)^2},$$

and since $r \geq 1 + c$

$$\frac{\log 2}{r-1} \leq \frac{\log 2}{c}, \qquad \frac{1}{(r-1)^2} \leq \frac{1}{c^2},$$

which finishes the proof of (ii).

To complete the proof of (i), we consider the function $f(x) = 2^{-x} x$ and note that it is decreasing for $x > 1/\log 2$. Therefore $n2^{-r} = (n2^{-r}r)/r \leq (\log n)/(r \log 2)$, for $n \geq 3$. Since $1 \leq n^{1/r}$, we get the desired result.

Lemma 8 *For any $m > 0$, $l \geq 1$, $r_0 > 0$, $r \in (0, r_0]$, $s \in (0, rl - 2]$, and $n \geq e^{2mr_0}$*

$$\sum_{i=1}^{\infty} \frac{i^s (\log i)^m}{(i^r + n)^l} \leq 4n^{(1+s-lr)/r} \frac{(\log n)^m}{r^m}.$$

The same upper bound holds for $m = 0$, $r \in (0, \infty)$, $s \in (0, rl - 1)$, and $n \geq 1$.

Proof We deal with this sum by splitting the sum in the parts $i \leq n^{1/r}$ and $i > n^{1/r}$. In the first range we bound the sum by

$$\sum_{i=1}^{n^{1/r}} n^{-l} i^s (\log i)^m \leq n^{1/r} n^{-l+s/r} \frac{(\log n)^m}{r^m},$$

by monotonicity of the function $f(x) = x^s (\log x)^m$.

Suppose that $m > 0$. The derivative of the function $f(x) = x^{-1/2} (\log x)^m$ is $f'(x) = x^{-3/2} (\log x)^{m-1} (m - (\log x)/2)$, hence it is monotone decreasing for $x \geq e^{2m}$. Since $n^{1/r} \geq n^{1/r_0}$ and $n > e^{2mr_0}$, the function f is decreasing on interval $[n^{1/r}, \infty)$. Therefore we bound the sum over the second range by

$$\sum_{i=n^{1/r}}^{\infty} i^{s-rl} (\log i)^m \leq n^{-1/(2r)} \frac{(\log n)^m}{r^m} \sum_{i=n^{1/r}}^{\infty} i^{1/2+s-rl}.$$

Since $s \leq rl - 2$, $i^{1/2+s-rl}$ is decreasing and $rd - s - 3/2 \geq 1/2$. We get

$$\sum_{i=n^{1/r}}^{\infty} i^{1/2+s-rl} \leq n^{(1/2+s-rl)/r} + \int_{n^{1/r}}^{\infty} x^{1/2+s-rl} \, dx$$

$$= n^{(1/2+s-rl)/r} + \frac{1}{-3/2 - s + rl} n^{(3/2+s-rl)/r}$$

$$\leq 3 n^{(3/2+s-rl)/r}.$$

In the case $m = 0$, we use monotonicity of i^{s-rl} for all $i \geq 1$.

Lemma 9 *For any $r \in (1, (\log n)/(2 \log(3e/2))]$, and $\gamma > 0$,*

$$\sum_{i=1}^{\infty} \frac{n^\gamma \log i}{(i^r + n)^\gamma} \geq \frac{1}{3 \cdot 2^\gamma r} n^{1/r} \log n.$$

Proof In the range $i \leq n^{1/r}$ we have $i^r + n \leq 2n$, thus

$$\sum_{i=1}^{\infty} \frac{n^\gamma \log i}{(i^r + n)^\gamma} \geq \frac{1}{2^\gamma} \sum_{i=1}^{\lfloor n^{1/r} \rfloor} \log i \geq \frac{1}{2^\gamma} \int_1^{\lfloor n^{1/r} \rfloor} \log x \, dx \geq \frac{1}{2^\gamma} \int_1^{(2/3)n^{1/r}} \log x \, dx,$$

since $n^{1/r} \geq 2$ and $\lfloor x \rfloor \geq 2x/3$ for $x \geq 2$. The latter integral equals $(2/3)n^{1/r} (\log((2/3)n^{1/r}) - 1) + 1$. Since $\log n \geq 2 \log(3e/2) r$ implies that $(\log n)/(2r) \geq \log(3e/2)$, we have

$$\frac{2}{3} n^{1/r} \left(\log \left(\frac{2}{3} n^{1/r} \right) - 1 \right) = \frac{2}{3} n^{1/r} \left(\frac{1}{r} \log n - \log \frac{3e}{2} \right) \geq \frac{1}{3r} n^{1/r} \log n.$$

Lemma 10 *Let m, i, r, and ξ be positive reals. Then for* $n \geq e^m$

$$\frac{ni^r \, (r \log i)^m}{(i^r + n)^2} \leq (\log n)^m, \quad \text{and} \quad \frac{n^\xi \, (r \log i)^{\xi m}}{(i^r + n)^\xi} \leq (\log n)^{\xi m}.$$

Proof Assume first that $i \leq n^{1/r}$, then the left hand side of the first inequality is bounded above by

$$\frac{n^2 \, \left(r \log n^{1/r}\right)^m}{n^2} = (\log n)^m.$$

Next assume that $i > n^{1/r}$. The derivative of the function $f(x) = x^{-c} (\log x)^m$ is $f'(x) = x^{-c-1} (\log x)^{m-1} \left(-c(\log x) + m\right)$, hence $f(x)$ is monotone decreasing for $x \geq e^{m/c}$. Therefore the function $i^{-r} (\log i)^m$ is monotone decreasing for $i \geq e^{m/r}$ and since by assumption $i > n^{1/r}$, we get that for $n \geq e^m$ the function $f(i) = i^{-r} (\log i)^m$ takes its maximum at $i = n^{1/r}$. Hence the left hand side of the inequality is bounded above by

$$n \, (r \log i)^m \, i^{-r} \leq nr^m \left(\log n^{1/r}\right)^m n^{-1} = (\log n)^m.$$

The second inequality can be proven similarly. □

References

1. Agapiou, S., Larsson, S., Stuart, A.M.: Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems. Stoch. Process. Appl. **123**(10), 3828–3860 (2013)
2. Arbel, J., Gayraud, G., Rousseau, J.: Bayesian optimal adaptive estimation using a sieve prior. Scand. J. Stat. **40**(3), 549–570 (2013)
3. Bailer-Jones, C.A.L.: Bayesian inference of stellar parameters and interstellar extinction using parallaxes and multiband photometry. Mon. Not. R. Astron. Soc. **411**(1), 435–452 (2011)
4. Belitser, E., Enikeeva, F.: Empirical Bayesian test of the smoothness. Math. Methods Stat. **17**(1), 1–18 (2008)
5. Belitser, E., Ghosal, S.: Adaptive Bayesian inference on the mean of an infinite-dimensional normal distribution. Ann. Stat. **31**(2), 536–559 (2003)
6. Cai, T.T.: On adaptive wavelet estimation of a derivative and other related linear inverse problems. J. Stat. Plan. Inference **108**(1–2), 329–349 (2002)
7. Castillo, I.: Lower bounds for posterior rates with Gaussian process priors. Electron. J. Stat. **2**, 1281–1299 (2008)
8. Cavalier, L.: Nonparametric statistical inverse problems. Inverse Problems, **24**(3):034004, 19, (2008)
9. Cavalier, L.: Inverse problems in statistics. In: Inverse Problems and High-Dimensional Estimation: Stats in the Château Summer School, vol. 203. Lecture Notes in Statistics. Springer, pp. 3–96 (2011)
10. Cavalier, L., Golubev, G.K., Picard, D., Tsybakov, A.B.: Oracle inequalities for inverse problems. Ann. Stat. **30**(3), 843–874 (2002)
11. Cavalier, L., Golubev, Y.: Risk hull method and regularization by projections of ill-posed inverse problems. Ann. Stat. **34**(4), 1653–1677 (2006)

12. Cavalier, L., Tsybakov, A.: Sharp adaptation for inverse problems with random noise. Probab. Theory Relat. Fields **123**(3), 323–354 (2002)
13. de Jonge, R., van Zanten, J.H.: Adaptive nonparametric Bayesian inference using location-scale mixture priors. Ann. Stat. **38**(6), 3300–3320 (2010)
14. Florens, J., Simoni, A.: Regularized posteriors in linear ill-posed inverse problems. Scand. J. Stat. **39**(2), 214–235 (2012)
15. Florens, J., Simoni, A.: Regularized priors for linear inverse problems. IDEI Working Paper 767–2013 (2013)
16. Gao, P., Honkela, A., Rattail, M., Lawrence, N.D.: Gaussian process modelling of latent chemical species: applications to inferring transcription factor activities. Bioinformatics **24**, 70–75 (2008)
17. Ghosal, S., Ghosh, J.K., van der Vaart, A.W.: Convergence rates of posterior distributions. Ann. Stat. **28**(2), 500–531 (2000)
18. Ghosal, S., van der Vaart, A.: Convergence rates of posterior distributions for non-i.i.d. observations. Ann. Stat. **35**(1), 192–223 (2007)
19. Johnstone, I.M., Silverman, B.W.: Needles and straw in haystacks: empirical Bayes estimates of possibly sparse sequences. Ann. Stat. **32**(4), 1594–1649 (2004)
20. Johnstone, I.M., Silverman, B.W.: Empirical Bayes selection of wavelet thresholds. Ann. Stat. **33**(4), 1700–1752 (2005)
21. Kato, K.: Quasi-Bayesian analysis of nonparametric instrumental variables models. Ann. Stat. **41**(5), 2359–2390 (2013)
22. Knapik, B.T., van der Vaart, A.W., van Zanten, J.H.: Bayesian inverse problems with Gaussian priors. Ann. Stat. **39**(5), 2626–2657 (2011)
23. Knapik, B.T., van der Vaart, A.W., van Zanten, J.H.: Bayesian recovery of the initial condition for the heat equation. Commun. Stat. Theory Methods **42**(7), 1294–1313 (2013)
24. Lashkari, D., Sridharan, R., Vul, E., Hsieh, P.-J., Kanwisher, N., Golland, P.: Search for patterns of functional specificity in the brain: a nonparametric hierarchical Bayesian model for group fMRI data. NeuroImage **59**(2), 1348–1368 (2012)
25. Marteau, C.: Risk hull method for spectral regularization in linear statistical inverse problems. ESAIM Probab. Stat. **14**, 409–434 (2010)
26. Marteau, C.: The Stein hull. J. Nonparametr. Stat. **22**(5–6), 685–702 (2010)
27. Oh, S.-H., Kwon, B.-D.: Geostatistical approach to Bayesian inversion of geophysical data: Markov chain Monte Carlo method. Earth Planets Space **53**, 777–791 (2001)
28. Orbanz, P., Buhmann, J.M.: Nonparametric Bayesian image segmentation. Int. J. Comput. Vis. **77**, 25–45 (2008)
29. Petrone, S., Rousseau, J., Scricciolo, C.: Bayes and empirical Bayes: do they merge? Biometrika **101**(2), 285–302. doi:10.1093/biomet/ast067 (2014)
30. Ray, K.: Bayesian inverse problems with non-conjugate priors. Electron. J. Stat. **7**, 2516–2549 (2013)
31. Rivoirard, V., Rousseau, J.: Bernstein-von Mises theorem for linear functionals of the density. Ann. Stat. **40**(3), 1489–1523 (2012)
32. Shen, W., Ghosal, S.: MCMC-free adaptive Bayesian procedures usingrandom series prior. arXiv:1204.4238 (2012)
33. Shen, X., Wasserman, L.: Rates of convergence of posterior distributions. Ann. Stat. **29**(3), 687–714 (2001)
34. Stuart, A.M.: Inverse problems: a Bayesian perspective. Acta Numer. **19**, 451–559 (2010)
35. Szabó, B.T., van der Vaart, A.W., van Zanten, J.H.: Empirical Bayes scaling of Gaussian priors in the white noise model. Electron. J. Stat. **7**, 991–1018 (2013)
36. Szabó, B.T., van der Vaart, A.W., van Zanten, J.H.: Frequentist coverage of adaptive nonparametric Bayesian credible sets. To appear in Ann. Statist. arXiv:1310.4489 (2013)
37. Tierney, L.: Markov chains for exploring posterior distributions. Ann. Stat. **22**(4), 1701–1762 (1994)
38. van der Vaart, A.W., van Zanten, J.H.: Adaptive Bayesian estimation using a Gaussian random field with inverse gamma bandwidth. Ann. Stat. **37**(5B), 2655–2675 (2009)
39. van der Vaart, A.W., Wellner, J.A.: Weak Convergence and Empirical Processes. Springer, New York (1996)
40. Zhang, C.-H.: General empirical Bayes wavelet methods and exactly adaptive minimax estimation. Ann. Stat. **33**(1), 54–100 (2005)
41. Zhao, L.H.: Bayesian aspects of some nonparametric problems. Ann. Stat. **28**(2), 532–552 (2000)