"Adaptive Bayesian estimation in indirect Gaussian sequence space models"

Johannes, Jan ; Simoni, Anna ; Schenk, Rudolf

Abstract

In an indirect Gaussian sequence space model lower and upper bounds are derived for the concentration rate of the posterior distribution of the parameter of interest shrinking to the parameter value θ^0 that generates the data. While this establishes posterior consistency, however, the concentration rate depends on both θ^0 and a tuning parameter which enters the prior distribution. We first provide an oracle optimal choice of the tuning parameter, i.e., optimized for each θ^0 separately. The optimal choice of the prior distribution allows us to derive an oracle optimal concentration rate of the associated posterior distribution. Moreover, for a given class of parameters and a suitable choice of the tuning parameter, we show that the resulting uniform concentration rate over the given class is optimal in a minimax sense. Finally, we construct a hierarchical prior that is adaptive. This means that, given a parameter θ^0 or a class of parameters, respectively, the posterior distrib...

Document type : *Document de travail (Working Paper)*

Référence bibliographique

Johannes, Jan ; Simoni, Anna ; Schenk, Rudolf. *Adaptive Bayesian estimation in indirect Gaussian sequence space models.* ISBA Discussion Paper ; 2015/03 (2015) 35 pages

I N S T I T U T D E S T A T I S T I Q U E

B I O S T A T I S T I Q U E E T

S C I E N C E S A C T U A R I E L L E S

(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N

P A P E R

2015/03

Adaptive Bayesian estimation in indirect Gaussian sequence space models

JOHANNES, J., SIMONI, A. and R. SCHENK

Adaptive Bayesian estimation in indirect Gaussian sequence space models

Jan Johannes*[1]
CREST-Ensai and Université catholique de Louvain

Anna Simoni[2]
CNRS and CREST

Rudolf Schenk[3]
Université catholique de Louvain

Abstract

In an indirect Gaussian sequence space model lower and upper bounds are derived for the concentration rate of the posterior distribution of the parameter of interest shrinking to the parameter value θ° that generates the data. While this establishes posterior consistency, however, the concentration rate depends on both θ° and a tuning parameter which enters the prior distribution. We first provide an oracle optimal choice of the tuning parameter, i.e., optimized for each θ° separately. The optimal choice of the prior distribution allows us to derive an oracle optimal concentration rate of the associated posterior distribution. Moreover, for a given class of parameters and a suitable choice of the tuning parameter, we show that the resulting uniform concentration rate over the given class is optimal in a minimax sense. Finally, we construct a hierarchical prior that is adaptive. This means that, given a parameter θ° or a class of parameters, respectively, the posterior distribution contracts at the oracle rate or at the minimax rate over the class. Notably, the hierarchical prior does not depend neither on θ° nor on the given class. Moreover, convergence of the fully data-driven Bayes estimator at the oracle or at the minimax rate is established.

Keywords: Bayesian nonparametrics, Sieve prior, hierarchical Bayes, exact concentration rates, oracle optimality, minimax theory, adaptation.

AMS 2000 subject classifications: Primary 62C10; secondary 62G05, 62G20.

*Corresponding author.

[1]Ensai, Campus de Ker Lann, Rue Blaise Pascal - BP 37203, 35172 Bruz Cedex, France and ISBA/CORE, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium, e-mail: jan.johannes@ensai.fr

[2]CREST, 15 Boulevard Gabriel Péri, 92240 Malakoff, France, e-mail: simoni.anna@gmail.com

[3]Rudolf is deceased. This manuscript is based on a joint work while he was a Ph.D. student at ISBA, Université catholique de Louvain.

1 Introduction

Accounting for the fact that inverse problems are widely used in many fields of science, there has been over the last decades a growing interest in statistical inverse problems (see, *e.g.*, Korostelev and Tsybakov [1993], Mair and Ruymgaart [1996], Evans and Stark [2002], Kaipio and Somersalo [2005], Bissantz et al. [2007] and references therein). Mathematical statistics has paid special attention to oracle or minimax optimal nonparametric estimation and adaptation in the framework of inverse problems (see Efromovich and Koltchinskii [2001], Cavalier et al. [2003], Cavalier [2008] and Hoffmann and Reiß [2008], to name but a few). Nonparametric estimation in general requires to choose a tuning parameter which is challenging in practise. Oracle and minimax estimation is achieved, respectively, if the tuning parameter is set to an optimal value which relies either on a knowledge of the unknown parameter of interest or of certain characteristics of it (such as smoothness). Since both the parameter and its smoothness are unknown, it is necessary to design a feasible procedure to select the tuning parameter that adapts to the unknown underlying function or to its regularity and achieves the oracle or minimax rate. Among the most prominent approaches stand without doubts model selection (cf. Barron et al. [1999] and its exhaustive discussion in Massart [2007]), Stein's unbiased risk estimation and its extensions (cf. Cavalier et al. [2002], Cavalier et al. [2002] or Cavalier and Hengartner [2005]), Lepski's method (see, e.g., Lepskij [1990], Birgé [2001], Efromovich and Koltchinskii [2001] or Mathé [2006]) or combinations of the aforementioned strategies (cf. Goldenshluger and Lepski [2011] and Comte and Johannes [2012]). On the other hand side, it seems natural to adopt a Bayesian point of view where the tuning parameter can be endowed with a prior. As the theory for a general inverse problem – with a possibly unknown or noisy operator – is technically highly involved, we consider in this paper as a starting point an indirect Gaussian regression which is well known to be equivalent to an indirect Gaussian sequence space model (in a Le Cam [1964] sense, see, e.g., Brown and Low [1996] for the direct case and Meister [2011] for the indirect case).

Let ℓ_2 be the Hilbert space of square summable real valued sequences endowed with the usual inner product $\langle\cdot,\cdot\rangle_{\ell_2}$ and associated norm $\|\cdot\|_{\ell_2}$. In an indirect Gaussian sequence space model (iGSSM) one aim is to recover a parameter sequence $\theta = \big(\theta_j\big)_{j\geqslant 1} \in \ell_2$ from a transformed version $(\lambda_j\theta_j)_{j\geqslant 1}$ that is blurred by a Gaussian white noise. Precisely, an observable sequence of random variables $(\mathrm{Y})_{j\geqslant 1}$, Y for short, obeys an indirect Gaussian sequence space model, if

$$\mathrm{Y}_j = \lambda_j\theta_j + \sqrt{\varepsilon}\xi_j, \qquad j \in \mathbb{N}, \tag{1.1}$$

where $\{\xi_j\}_{j\geqslant 1}$ are unobservable error terms, which are independent and standard normally distributed, and $0 < \varepsilon < 1$ is the noise level. The sequence $\lambda = \big(\lambda_j\big)_{j\geqslant 1}$ represents the operator that transforms the signal θ. In the particular case of a constant sequence λ the sequence space model is called direct while it is called an indirect sequence space model if the sequence λ tends to zero. We assume throughout the paper that the sequence is bounded.

In this paper we adopt a Bayesian approach, where the parameter sequence of interest $\theta = (\theta_j)_{j\geqslant 1}$ itself is a realisation of a random variable $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_j)_{j\geqslant 1}$ and the observable random variable $\mathrm{Y} = (\mathrm{Y}_j)_{j\geqslant 1}$ satisfies

$$\mathrm{Y}_j = \lambda_j\,\boldsymbol{\vartheta}_j + \sqrt{\varepsilon}\xi_j, \quad j \in \mathbb{N} \tag{1.2}$$

with independent and standard normally distributed error terms $\{\xi_j\}_{j\geqslant 1}$ and noise level $0 < \varepsilon < 1$. Throughout the paper we assume that random parameters $\{\boldsymbol{\vartheta}_j\}_{j\geqslant 1}$ and the error terms $\{\xi_j\}_{j\geqslant 1}$ are independent. Consequently, (1.2) and a specification of the prior distribution $P_{\boldsymbol{\vartheta}}$ of $\boldsymbol{\vartheta}$ determine completely the joint distribution of Y and $\boldsymbol{\vartheta}$. For a broader overview on Bayesian procedures we refer the reader to the monograph by Robert [2007].

Typical prior specifications studied in the direct sequence space model literature are compound priors, also known as Sieve priors (see, e.g., Zhao [2000], Shen and Wasserman [2001] or Arbel et al. [2013], Gaussian series priors (cf. Freedman [1999], Cox [1993] or Castillo [2008]), block priors (cf. Gao and Zhou [2014]), countable mixture of normal priors (cf. Belitser and Ghosal [2003]) and finite mixtures of normal and Dirac priors (*e.g.* Abramovich et al. [1998]). In the context of an iGSSM, Knapik et al. [2011] and Knapik et al. [2014] consider Gaussian series priors and continuous mixture of Gaussian series priors, respectively.

By considering an iGSSM we derive in this paper theoretical properties of a Bayes procedure with a Sieve prior specification from a frequentist point of view, meaning that there exists a true parameter value $\theta^\circ = (\theta^\circ_j)_{j\geqslant 1}$ associated with the data generating process of $\big(\mathrm{Y}_j\big)_{j\geqslant 1}$. A broader overview of frequentist asymptotic properties of nonparametric Bayes procedures can be found, for example, in Ghosh and Ramamoorthi [2003], while direct and indirect models, respectively, are considered by e.g., Zhao [2000], Belitser and Ghosal [2003], Castillo [2008] and Gao and Zhou [2014], and, e.g., Knapik et al. [2011] and Knapik et al. [2014]. Bayesian procedures in the context of slightly different Gaussian inverse problems and their asymptotic properties are studied in, e.g., Agapiou et al. [2013] and Florens and Simoni [2014]. However, our special attention is given to posterior consistency and optimal posterior concentration in an oracle or minimax sense, which we elaborate in the following.

In this paper we consider a sieve prior family $\{P_{\boldsymbol{\vartheta}^m}\}_m$ where the prior distribution $P_{\boldsymbol{\vartheta}^m}$ of the random parameter sequence $\boldsymbol{\vartheta}^m = (\boldsymbol{\vartheta}^m_j)_{j\geqslant 1}$ is Gaussian and degenerated for all $j > m$. More precisely, the first m coordinates $\{\boldsymbol{\vartheta}^m_j\}_{j=1}^m$ are independent and normally distributed random variables while the remaining coordinates $\{\boldsymbol{\vartheta}^m_j\}_{j>m}$ are degenerated at a point. Note that the dimension parameter m plays the role of a tuning parameter. Assuming an observation $\mathrm{Y} = (\mathrm{Y}_j)_{j\geqslant 1}$ satisfying $\mathrm{Y}_j = \boldsymbol{\vartheta}^m_j + \sqrt{\varepsilon}\xi_j$, we denote by $P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}$ the corresponding posterior distribution of $\boldsymbol{\vartheta}^m$ given Y. Given a prior sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon}}\}_{m_\varepsilon}$ in dependence of the noise level ε, our objective is the study of frequentist properties of the associated posterior sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon}\,|\,\mathrm{Y}}\}_{m_\varepsilon}$. To be more precise, let θ° be the realization of the random parameter $\boldsymbol{\vartheta}$ associated with the data-generating distribution and denote by $\mathbb{E}_{\theta^\circ}$ the corresponding expectation. A quantity Φ_ε which is up to a constant a lower and an upper bound of the concentration of the posterior sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon}\,|\,\mathrm{Y}}\}_{m_\varepsilon}$, i.e.,

$$\lim_{\varepsilon\to 0}\mathbb{E}_{\theta^\circ}P_{\boldsymbol{\vartheta}^{m_\varepsilon}\,|\,\mathrm{Y}}((K)^{-1}\,\Phi_\varepsilon \leqslant \|\boldsymbol{\vartheta}^{m_\varepsilon} - \theta^\circ\|^2_{\ell_2} \leqslant K\,\Phi_\varepsilon) = 1 \quad \text{with } 1 \leqslant K < \infty, \tag{1.3}$$

is called exact posterior concentration (see, e.g., Barron et al. [1999], Ghosal et al. [2000] or Castillo [2008] for a broader discussion of the concept of posterior concentration). We shall emphasise that the derivation of the posterior concentration relies strongly on tail bounds for non-central χ^2 distributions established in Birgé [2001]. Moreover, if $\Phi_\varepsilon \to 0$ as $\varepsilon \to 0$ then the lower and upper bound given in (1.3) establish posterior consistency and Φ_ε is called exact posterior concentration rate. Obviously, the exact rate depends on the prior sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon}}\}_{m_\varepsilon}$ as well as on the unknown parameter θ°.

In the spirit of a frequentist oracle approach, given a parameter θ° we derive in this paper a prior sub-family $\{P_{\boldsymbol{\vartheta}^{m^\circ_\varepsilon}}\}_{m^\circ_\varepsilon}$ with smallest possible exact posterior concentration rate Φ°_ε which we call, respectively, an oracle prior sub-family and an oracle posterior concentration rate. On the other hand side, following a minimax approach, Johannes and Schwarz [2013], for example, derive the minimax rate of convergence $\Phi^\star_\varepsilon$ of the maximal mean integrated squared error (MISE) over a given class $\Theta_{\mathfrak{a}}$ of parameters (introduced below). We construct a sub-family $\{P_{\boldsymbol{\vartheta}^{m^\star_\varepsilon}}\}_{m^\star_\varepsilon}$ of prior distributions with exact posterior concentration rate $\Phi^\star_\varepsilon$ uniformly over $\Theta_{\mathfrak{a}}$ which does not depend on the true parameter θ° but only on the set of possible parameters $\Theta_{\mathfrak{a}}$. It is interesting to note that in a direct GSSM Castillo [2008] establishes up to a constant the minimax-rate as an upper bound of the posterior concentration, while the derived lower bound features a logarithmic factor compared to the minimax rate. Arbel et al. [2013], for example, in a direct GSSM and Knapik et al. [2014] in an indirect GSSM provide only upper bounds of the posterior concentration rate which differ up to a logarithmic factor from the minimax rate. We shall emphasize, that the prior specifications we propose in this paper lead to exact posterior concentration rates that are optimal in an oracle or minimax sense over certain classes of parameters not only in the direct model but also in the more general indirect model. However, both oracle and minimax sieve prior are unfeasible in practise since they rely on the knowledge of either θ° itself or its smoothness.

Our main contribution in this paper is the construction of a hierarchical prior $P_{\boldsymbol{\vartheta}^{\mathrm{M}}}$ that is adaptive. Meaning that, given a parameter $\theta^\circ \in \ell_2$ or a classes $\Theta_{\mathfrak{a}} \subset \ell_2$ of parameters, the posterior distribution $P_{\boldsymbol{\vartheta}^{\mathrm{M}}\,|\,Y}$ contracts, respectively, at the oracle rate or the minimax rate over $\Theta_{\mathfrak{a}}$ while the hierarchical prior $P_{\boldsymbol{\vartheta}^{\mathrm{M}}}$ does not rely neither on the knowledge of θ° nor the class $\Theta_{\mathfrak{a}}$. Let us briefly elaborate on the hierarchical structure of the prior which induces an additional prior on the tuning parameter m, i.e., m itself is a realisation of a random variable M. We construct a prior for M such that the marginal posterior for $\boldsymbol{\vartheta}^{\mathrm{M}}$ (obtained by integrating out M with respect to its posterior) contracts exactly at the oracle concentration rate. This is possible for every θ° whose components differ from the components of the prior mean infinitely many times. In addition, for every θ° in the class $\Theta_{\mathfrak{a}}$ we show that the posterior distribution $P_{\boldsymbol{\vartheta}^{\mathrm{M}}\,|\,Y}$ contracts at least at the minimax rate $\Phi_\varepsilon^\star$ and that the corresponding Bayes estimate is minimax-optimal. Thereby, the proposed Bayesian procedure is *minimax adaptive* over the class $\Theta_{\mathfrak{a}}$.

Although adaptation has attracted remarkable interest in the frequentist literature, only few contributions are available in the Bayesian literature on Gaussian sequence space models. In a direct model Belitser and Ghosal [2003], Szabó et al. [2013], Arbel et al. [2013] and Gao and Zhou [2014] derive Bayesian methods that achieve minimax adaptation while in an indirect Gaussian sequence space model, to the best of our knowledge, only Knapik et al. [2014] has derived an adaptive Bayesian procedure. In this paper, we extend previous results on adaptation obtained through sieve priors to the indirect Gaussian sequence space model. This requires a specification of the prior on the tuning parameter M different from the one used by, e.g., Zhao [2000] and Arbel et al. [2013]. Interestingly, our novel prior specification on M improves the general results of Arbel et al. [2013] since it allows to obtain adaptation without a rate loss (given by a logarithmic factor) even in the direct model. Compared to Knapik et al. [2014] our procedure relies on a sieve prior while they use a family of Gaussian prior for $\boldsymbol{\vartheta}$ that is not degenerate in any component of $\boldsymbol{\vartheta}$ and where the hyper-parameter is represented by the smoothness of the prior variance. Their procedure is minimax-adaptive up to a logarithmic deterioration of the minimax rate on certain smoothness classes for θ° which is, instead, avoided by our procedure.

The rest of the paper is organised as follows. The prior scheme is specified in Section 2. In Section 3 we derive the lower and upper bound of the posterior concentration, the oracle posterior concentration rate and the minimax rate. In Section 4 we introduce a prior distribution P_{M} for the random dimension M and we prove adaptation of the hierarchical Bayes procedure. The proofs are given in the appendix.

2 Basic model assumptions

Let us consider a Gaussian prior distribution for the parameter $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_j)_{j\geqslant 1}$, that is, $\{\boldsymbol{\vartheta}_j\}_{j\geqslant 1}$ are independent, normally distributed with prior means $(\theta_j^\times)_{j\geqslant 1}$ and prior variances $(\varsigma_j)_{j\geqslant 1}$. Standard calculus shows that the posterior distribution of $\boldsymbol{\vartheta}$ given $\mathrm{Y} = (\mathrm{Y}_j)_{j\geqslant 1}$ is Gaussian, that is, given Y, $\{\boldsymbol{\vartheta}_j\}_{j\geqslant 1}$ are conditionally independent, normally distributed random variables with posterior variance $\sigma_j := \mathbb{V}\mathrm{ar}(\boldsymbol{\vartheta}_j \,|\, \mathrm{Y}) = (\lambda_j^2\varepsilon^{-1}+\varsigma_j^{-1})^{-1}$ and posterior mean $\theta_j^{\mathrm{Y}} := \mathbb{E}[\boldsymbol{\vartheta}_j \,|\, \mathrm{Y}] = \sigma_j(\varsigma_j^{-1}\theta_j^\times + \lambda_j\varepsilon^{-1}\,\mathrm{Y}_j)$, for all $j \in \mathbb{N}$. Taking this as a starting point, we construct a sequence of hierarchical Sieve prior distributions. To be more precise, let us denote by δ_x the Dirac measure in the point x. Given $m \in \mathbb{N}$, we consider the independent random variables $\big\{\boldsymbol{\vartheta}_j^m\big\}_{j\geqslant 1}$ with marginal distributions

$$\boldsymbol{\vartheta}_j^m \sim \mathcal{N}(\theta_j^\times, \varsigma_j),\ 1 \leqslant j \leqslant m \text{ and } \boldsymbol{\vartheta}_j^m \sim \delta_{\theta_j^\times},\ m < j, \tag{2.1}$$

resulting in the degenerate prior distribution $P_{\boldsymbol{\vartheta}^m}$. Here, we use the notation $\boldsymbol{\vartheta}^m = \big(\boldsymbol{\vartheta}_j^m\big)_{j\geqslant 1}$. Consequently, $\{\boldsymbol{\vartheta}_j^m\}_{j\geqslant 1}$ are conditionally independent given Y and their posterior distribution $P_{\boldsymbol{\vartheta}_j^m \,|\, \mathrm{Y}}$ is Gaussian with mean θ_j^{Y} and variance σ_j for $1 \leqslant j \leqslant m$ while being degenerate on $\theta_j^\times$ for $j > m$.

Let $\mathbb{1}_A$ denote the indicator function which takes the value one if the condition A holds true, and the value zero otherwise. We consider the posterior mean $\widehat{\theta}^m = \big(\widehat{\theta}_j^m\big)_{j\geqslant 1} := \mathbb{E}[\boldsymbol{\vartheta}^m \,|\, \mathrm{Y}]$ given for $j \geqslant 1$ by $\widehat{\theta}_j^m := \theta_j^{\mathrm{Y}}\, \mathbb{1}_{\{j\,\leqslant\, m\}} + \theta_j^\times\, \mathbb{1}_{\{j\,>\,m\}}$ as Bayes estimator of θ. We shall emphasize an improper specification of the prior, that is, $\theta^\times = \big(\theta_j^\times\big)_{j\geqslant 1} \equiv 0$ and $\varsigma = \big(\varsigma_j\big)_{j\geqslant 1} \equiv \infty$. Obviously, in this situation $\theta^{\mathrm{Y}} = Y/\lambda = \big(Y_j/\lambda_j\big)_{j\geqslant 1}$ and $\sigma = \varepsilon/\lambda^2 = \big(\varepsilon/\lambda_j^2\big)_{j\geqslant 1}$ are the posterior mean and variance sequences, respectively. Consequently, under the improper prior specification, for each $m \in \mathbb{N}$ the posterior mean $\widehat{\theta}^m = \mathbb{E}[\boldsymbol{\vartheta}^m \,|\, \mathrm{Y}]$ of $\boldsymbol{\vartheta}^m$ corresponds to an orthogonal projection estimator, i.e., $\widehat{\theta}^m = (Y/\lambda)^m$ with $(Y/\lambda)_j^m = Y_j/\lambda_j\, \mathbb{1}_{\{1\,\leqslant\, j\,\leqslant\, m\}}$.

From a Bayesian point of view the thresholding parameter m is a hyper-parameter and hence, we may complete the prior specification by introducing a prior distribution on it. Consider a random thresholding parameter M taking its values in $\{1, \dots, G_\varepsilon\}$ for some $G_\varepsilon \in \mathbb{N}$ with prior distribution P_{M}. Both G_ε and P_{M} will be specified in Section 4. Moreover, the distribution of the random variables $\{\mathrm{Y}_j\}_{j\geqslant 1}$ and $\{\boldsymbol{\vartheta}_j^{\mathrm{M}}\}_{j\geqslant 1}$ conditionally on M are determined by

$$\mathrm{Y}_j = \lambda_j\, \boldsymbol{\vartheta}^{\mathrm{M}} + \sqrt{\varepsilon}\xi_j \quad \text{ and } \quad \boldsymbol{\vartheta}_j^{\mathrm{M}} = \theta_j^\times + \sqrt{\varsigma_j}\eta_j\, \mathbb{1}_{\{1\,\leqslant\, j\,\leqslant\, \mathrm{M}\}}$$

where $\{\xi_j, \eta_j\}_{j\geqslant 1}$ are iid. standard normal random variables independent of M. Furthermore, the posterior mean $\widehat{\theta} := \mathbb{E}[\boldsymbol{\vartheta}^{\mathrm{M}} \,|\, \mathrm{Y}]$ satisfies $\widehat{\theta}_j = \theta_j^\times$ for $j > G_\varepsilon$ and $\widehat{\theta}_j = \theta_j^\times\, P(1 \leqslant \mathrm{M} < j|\,\mathrm{Y}) + \theta_j^{\mathrm{Y}}\, P(j \leqslant \mathrm{M} \leqslant G_\varepsilon|\,\mathrm{Y})$ for all $1 \leqslant j \leqslant G_\varepsilon$. It is important to note, that the marginal posterior distribution $P_{\boldsymbol{\vartheta}^{\mathrm{M}}\,|\,\mathrm{Y}}$ of $\boldsymbol{\vartheta}^{\mathrm{M}} = \big(\boldsymbol{\vartheta}_j^{\mathrm{M}}\big)_{j\geqslant 1}$ given the observation Y does depend on the prior specification and the observation only, and hence it is fully data-driven. Revisiting the improper prior specification introduced above, the data-driven Bayes estimator equals a shrunk orthogonal projection estimator. More precisely, we have $\widehat{\theta}_j = P(j \leqslant \mathrm{M} \leqslant G_\varepsilon|\,\mathrm{Y}) \times Y_j/\lambda_j\, \mathbb{1}_{\{1\,\leqslant\, j\,\leqslant\, G_\varepsilon\}}$. Interestingly, rather than using the data to select the dimension parameter m in the set of possible values $\{1, \dots, G_\varepsilon\}$, the Bayes estimator uses all components, up to G_ε, shrunk by a weight decreasing with the index.

3 Optimal concentration rate

3.1 Consistency

Note that conditional on Y the random variables $\{\boldsymbol{\vartheta}_j^m - \theta_j^\circ\}_{j=1}^m$ are independent and normally distributed with conditional mean $\theta_j^{\mathrm{Y}} - \theta_j^\circ$ and conditional variance σ_j. The next assertion presents a version of tail bounds for sums of independent squared Gaussian random variables. It is shown in the appendix using a result due to Birgé [2001] which can be shown along the lines of the proof of Lemma 1 in Laurent et al. [2012].

Lemma 3.1. *Let $\{X_j\}_{j\geqslant 1}$ be independent and normally distributed r.v. with mean $\alpha_j \in \mathbb{R}$ and standard deviation $\beta_j \geqslant 0$, $j \in \mathbb{N}$. For $m \in \mathbb{N}$ set $S_m := \sum_{j=1}^m X_j^2$ and consider $v_m \geqslant \sum_{j=1}^m \beta_j^2$, $t_m \geqslant \max_{1\leqslant j\leqslant m} \beta_j^2$ and $r_m \geqslant \sum_{j=1}^m \alpha_j^2$. Then for all $c \geqslant 0$ we have*

$$\sup_{m\geqslant 1} \exp\Big(\frac{c(c\wedge 1)(v_m + 2r_m)}{4t_m}\Big) P\big(S_m - \mathbb{E}S_m \leqslant -c(v_m + 2r_m)\big) \leqslant 1; \tag{3.1}$$

$$\sup_{m\geqslant 1} \exp\Big(\frac{c(c\wedge 1)(v_m + 2r_m)}{4t_m}\Big) P\big(S_m - \mathbb{E}S_m \geqslant \frac{3c}{2}(v_m + 2r_m)\big) \leqslant 1. \tag{3.2}$$

A major step towards establishing a concentration rate of the posterior distribution consists in finding a finite sample bound for a fixed $m \in \mathbb{N}$. We express these bounds in terms of

$$\mathfrak{b}_m := \sum_{j>m} (\theta_j^\circ - \theta_j^\times)^2, \quad m\overline{\sigma}_m := \sum_{j=1}^m \sigma_j \quad \text{with } \sigma_j = (\lambda_j^2 \varepsilon^{-1} + \varsigma_j^{-1})^{-1};$$

$$\sigma_{(m)} := \max_{1\leqslant j\leqslant m} \sigma_j \quad \text{and} \quad \mathfrak{r}_m := \sum_{j=1}^m (\mathbb{E}_{\theta^\circ}[\theta_j^{\mathrm{Y}}] - \theta_j^\circ)^2 = \sum_{j=1}^m \sigma_j^2 \varsigma_j^{-2} (\theta_j^\times - \theta_j^\circ)^2.$$

Proposition 3.2. *For all $m \in \mathbb{N}$, for all $\varepsilon > 0$ and for all $0 < c < 1/5$ we have*

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m \,|\, \mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|_{\ell_2}^2 > \mathfrak{b}_m + 3m\overline{\sigma}_m + 3\, m\, \sigma_{(m)}/2 + 4\mathfrak{r}_m) \leqslant 2\exp(-m/36); \quad (3.3)$$

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m \,|\, \mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|_{\ell_2}^2 < \mathfrak{b}_m + m\overline{\sigma}_m - 4\, c\, (m\, \sigma_{(m)} + \mathfrak{r}_m)) \leqslant 2\exp(-c^2 m/2). \quad (3.4)$$

The desired convergence to zero of all the aforementioned sequences necessitates to consider an appropriate sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon}}\}_{m_\varepsilon}$ in dependence of the noise level ε, notably introducing consequently sub-sequences $(m_\varepsilon\overline{\sigma}_{m_\varepsilon})_{m_\varepsilon \geqslant 1}$, $(\sigma_{(m_\varepsilon)})_{m_\varepsilon \geqslant 1}$ and $(\mathfrak{r}_{m_\varepsilon})_{m_\varepsilon \geqslant 1}$.

Assumption A.1. *There exist constants $0 < \varepsilon_\circ := \varepsilon_\circ(\theta^\circ, \lambda, \theta^\times, \varsigma) < 1$ and $1 \leqslant K := K(\theta^\circ, \lambda, \theta^\times, \varsigma) < \infty$ such that the Sieve sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon}}\}_{m_\varepsilon}$ of prior distributions satisfies the condition $\sup_{0<\varepsilon<\varepsilon_\circ}(\mathfrak{r}_{m_\varepsilon} \vee m_\varepsilon \sigma_{(m_\varepsilon)})/(\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon\overline{\sigma}_{m_\varepsilon}) \leqslant K$.*

The following corollary can be immediately deduced from Proposition 3.2 and we omit its proof.

Corollary 3.3. *Under Assumption A.1 for all $0 < \varepsilon < \varepsilon_\circ$ and $0 < c < 1/(8K)$ hold*

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon} \,|\, \mathrm{Y}}(\|\boldsymbol{\vartheta}^{m_\varepsilon} - \theta^\circ\|_{\ell_2}^2 > (4 + (11/2)K)[\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon\overline{\sigma}_{m_\varepsilon}]) \leqslant 2\exp(-\frac{m_\varepsilon}{36}); \quad (3.5)$$

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon} \,|\, \mathrm{Y}}(\|\boldsymbol{\vartheta}^{m_\varepsilon} - \theta^\circ\|_{\ell_2}^2 < (1 - 8\, c\, K)[\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon\overline{\sigma}_{m_\varepsilon}]) \leqslant 2\exp(-c^2 m_\varepsilon/2). \quad (3.6)$$

Note that the sequence $(\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon\overline{\sigma}_{m_\varepsilon})_{m_\varepsilon \geqslant 1}$ generally does not converge to zero. However, supposing that $m_\varepsilon \to \infty$ as $\varepsilon \to 0$ then it follows from the dominated convergence theorem that $\mathfrak{b}_{m_\varepsilon} = o(1)$. Hence, assuming additionally that $m_\varepsilon\overline{\sigma}_{m_\varepsilon} = o(1)$ holds true is sufficient to ensure that $(\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon\overline{\sigma}_{m_\varepsilon})_{m_\varepsilon \geqslant 1}$ converges to zero and it is indeed a posterior concentration rate. The next assertion summarises this result and we omit its elementary proof.

Proposition 3.4 (Posterior consistency). *Let Assumption A.1 be satisfied. If $m_\varepsilon \to \infty$ and $m_\varepsilon\overline{\sigma}_{m_\varepsilon} = o(1)$ as $\varepsilon \to 0$, then*

$$\lim_{\varepsilon \to 0} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon} \,|\, \mathrm{Y}}\big((10K)^{-1}[\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon\overline{\sigma}_{m_\varepsilon}] \leqslant \|\boldsymbol{\vartheta}^{m_\varepsilon} - \theta^\circ\|_{\ell_2}^2 \leqslant 10K[\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon\overline{\sigma}_{m_\varepsilon}]\big) = 1.$$

The last assertion shows that $(\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon\overline{\sigma}_{m_\varepsilon})_{m_\varepsilon \geqslant 1}$ is up to a constant a lower and upper bound of the concentration rate associated with the Sieve sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon}}\}_{m_\varepsilon}$ of prior distributions. It is easily shown that it also provides an upper bound of the frequentist risk of the associated Bayes estimator.

Proposition 3.5 (Bayes estimator consistency). *Let the assumptions of Proposition 3.4 be satisfied. Consider the Bayes estimator $\widehat{\theta}^{m_\varepsilon} := \mathbb{E}[\boldsymbol{\vartheta}^{m_\varepsilon} \,|\, \mathrm{Y}]$ then*

$$\mathbb{E}_{\theta^\circ}\|\widehat{\theta}^{m_\varepsilon} - \theta^\circ\|_{\ell_2}^2 \leqslant (2 + K)[\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon\overline{\sigma}_{m_\varepsilon}]$$

and consequently $\mathbb{E}_{\theta^\circ}\|\widehat{\theta}^{m_\varepsilon} - \theta^\circ\|_{\ell_2}^2 = o(1)$ as $\varepsilon \to 0$.

The previous results are obtained under Assumption A.1. However, it may be difficult to verify whether a given sub-family of priors $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon}}\}_{m_\varepsilon}$ satisfies such an assumption. Therefore, we now introduce an assumption which states a more precise requirement on the prior variance and that can be more easily verified. Define for $j, m \in \mathbb{N}$

$$\Lambda_j := \lambda_j^{-2}, \quad \Lambda_{(m)} := \max_{1 \leqslant j \leqslant m} \Lambda_j, \quad \overline{\Lambda}_m := m^{-1} \sum_{j=1}^{m} \Lambda_j \quad \text{and} \quad \Phi_\varepsilon^m := [\mathfrak{b}_m \vee \varepsilon\, m\, \overline{\Lambda}_m].$$

Assumption A.2. *Let $G_\varepsilon := \max\{1 \leqslant m \leqslant \lfloor \varepsilon^{-1} \rfloor : \varepsilon \Lambda_{(m)} \leqslant \Lambda_1\}$. There exists a finite constant $d > 0$ such that $\varsigma_j \geqslant d[\varepsilon^{1/2}\Lambda_j^{1/2} \vee \varepsilon \Lambda_j]$ for all $1 \leqslant j \leqslant G_\varepsilon$ and for all $\varepsilon \in (0, 1)$.*

Note that in the last Assumption the defining set of G_ε is not empty, since $\varepsilon \Lambda_{(1)} \leqslant \Lambda_1$ for all $\varepsilon \leqslant 1$. Moreover, under Assumption A.2, by some elementary algebra, it is readily verified for all $1 \leqslant j \leqslant G_\varepsilon$ that

$$1 \leqslant \varepsilon \Lambda_j / \sigma_j \leqslant (1 + 1/d) \quad \text{and} \quad \sigma_j / \varsigma_j \leqslant (1 \wedge d^{-1} \varepsilon^{1/2} \Lambda_j^{1/2})$$

which in turn implies for all $1 \leqslant m \leqslant G_\varepsilon$ that

$$\mathfrak{r}_m \leqslant d^{-2} \|\theta^\times - \theta^\circ\|_{\ell_2}^2 \varepsilon\, \Lambda_{(m)},\ 1 \leqslant \varepsilon\, m\, \Lambda_{(m)} (m \sigma_{(m)})^{-1} \text{ and } 1 \leqslant \varepsilon\, m \overline{\Lambda}_m\, (m \overline{\sigma}_m)^{-1} \leqslant (1 + 1/d).$$

We will use these elementary bounds in the sequel without further reference. Returning to the Sieve sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon}}\}_{m_\varepsilon}$ of prior distributions, if in addition to Assumption A.2 there exists a constant $1 \leqslant L := L(\theta^\circ, \lambda, \theta^\times) < \infty$ such that

$$\sup_{0 < \varepsilon < 1} \varepsilon\, m_\varepsilon\, \Lambda_{(m_\varepsilon)} (\Phi_\varepsilon^{m_\varepsilon})^{-1} \leqslant L \tag{3.7}$$

and $\Phi_\varepsilon^{m_\varepsilon} = o(1)$ as $\varepsilon \to 0$ hold true, then the sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon}}\}_{m_\varepsilon}$ satisfies Assumption A.1 with $K := ((1 + d^{-1}) \vee d^{-2} \|\theta^\circ - \theta^\times\|_{\ell_2}^2) L$. Indeed, if $\Phi_\varepsilon^{m_\varepsilon} = o(1)$ and, hence $\Phi_\varepsilon^{m_\varepsilon} \leqslant \Lambda_1 / L$ for all $\varepsilon \in (0, \varepsilon_\circ)$, then $m_\varepsilon \leqslant G_\varepsilon$ holds true for all $\varepsilon \in (0, \varepsilon_\circ)$ since $\varepsilon m_\varepsilon \Lambda_1 \leqslant \varepsilon m_\varepsilon \Lambda_{(m_\varepsilon)} \leqslant L \Phi_\varepsilon^{m_\varepsilon} \leqslant \Lambda_1$ and thus $m_\varepsilon \leqslant \lfloor \varepsilon^{-1} \rfloor$ and $\varepsilon \Lambda_{(m_\varepsilon)} \leqslant \Lambda_1$. In other words, for all $\varepsilon \in (0, \varepsilon_\circ)$ we can apply Assumption A.2 and the claim follows taking into account the aforementioned elementary bounds. Note further that the constant K does not depend on the prior variances ς but only on the constant d given by Assumption A.2. The next assertion follows immediately from Corollary 3.3 and we omit its proof.

Corollary 3.6. *Under Assumption A.2 consider a sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon}}\}_{m_\varepsilon}$ such that (3.7) and $\Phi_\varepsilon^{m_\varepsilon} = o(1)$ as $\varepsilon \to 0$ are satisfied, then there exists $\varepsilon_\circ \in (0, 1)$ such that for all $0 < \varepsilon < \varepsilon_\circ$ and $0 < c < 1/(8K)$ with $K = ((1 + d^{-1}) \vee d^{-2} \|\theta^\circ - \theta^\times\|_{\ell_2}^2) L$ hold*

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon} \,|\, Y} \big(\|\boldsymbol{\vartheta}^{m_\varepsilon} - \theta^\circ\|_{\ell_2}^2 > (4 + (11/2)K) \Phi_\varepsilon^{m_\varepsilon} \big) \leqslant 2 \exp(-\frac{m_\varepsilon}{36}); \tag{3.8}$$

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon} \,|\, Y} \big(\|\boldsymbol{\vartheta}^{m_\varepsilon} - \theta^\circ\|_{\ell_2}^2 < (1 - 8\, c\, K)(1 + d^{-1})^{-1} \Phi_\varepsilon^{m_\varepsilon} \big) \leqslant 2 \exp(-c^2 m_\varepsilon / 2). \tag{3.9}$$

The result implies consistency if $m_\varepsilon \to \infty$ as $\varepsilon \to 0$ but it does not answer the question of an optimal rate in a satisfactory way.

3.2 Oracle concentration rate

Considering the Sieve family $\{P_{\boldsymbol{\vartheta}^m}\}_m$ of prior distributions, the sequence $(\Phi_\varepsilon^{m_\varepsilon})_{m_\varepsilon\geqslant 1}$ provides up to constants a lower and upper bound for the posterior concentration rate for each sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon}}\}_{m_\varepsilon}$ satisfying the conditions of Corollary 3.6. Observe that the term $\mathfrak{b}_{m_\varepsilon}$ and hence the rate depends on the parameter of interest θ°. Let us minimise the rate for each θ° separately. For a sequence $(a_m)_{m\geqslant 1}$ with minimal value in A we set $\arg\min_{m\in A}\{a_m\} := \min\{m : a_m \leqslant a_k, \forall k \in A\}$ and define for all $\varepsilon > 0$

$$m_\varepsilon^\circ := m_\varepsilon^\circ(\theta^\circ,\theta^\times,\lambda) := \arg\min_{m\geqslant 1}\{\Phi_\varepsilon^m\} \text{ and}$$
$$\Phi_\varepsilon^\circ := \Phi_\varepsilon^\circ(\theta^\circ,\theta^\times,\lambda) := \Phi_\varepsilon^{m_\varepsilon^\circ} = \min_{m\geqslant 1}\Phi_\varepsilon^m \quad . \qquad (3.10)$$

We may emphasise that $\Phi_\varepsilon^\circ = o(1)$ as $\varepsilon \to 0$. Indeed, for all $\delta > 0$ there exists a dimension m_δ and a noise level ε_δ such that $\Phi_\varepsilon^\circ \leqslant [\mathfrak{b}_{m_\delta} \vee \varepsilon_\delta\, m_\delta\,\overline{\Lambda}_{m_\delta}] \leqslant \delta$ for all $0 < \varepsilon \leqslant \varepsilon_\delta$. Obviously, given $\theta^\circ \in \Theta$ the rate Φ_ε° is a lower bound for all posterior concentration rates $\Phi_\varepsilon^{m_\varepsilon}$ associated with a prior sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon}}\}_{m_\varepsilon}$ satisfying the conditions of Corollary 3.6. Moreover, the next assertion establishes Φ_ε° up to constants as upper and lower bound for the concentration rate associated with the sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon^\circ}}\}_{m_\varepsilon^\circ}$. Consequently, Φ_ε° is called oracle posterior concentration rate and $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon^\circ}}\}_{m_\varepsilon^\circ}$ oracle prior sub-family. The assertion follows again from Corollary 3.3 (with $c = 1/(9K)$) and we omit its proof.

Theorem 3.7 (Oracle posterior concentration rate). *Suppose that Assumption A.2 holds true and that there exists a constant $1 \leqslant L^\circ := L^\circ(\theta^\circ,\lambda,\theta^\times) < \infty$ such that*

$$\sup_{0<\varepsilon<1} \varepsilon\, m_\varepsilon^\circ\, \Lambda_{(m_\varepsilon^\circ)}(\Phi_\varepsilon^\circ)^{-1} \leqslant L^\circ. \qquad (3.11)$$

If in addition $m_\varepsilon^\circ \to \infty$ as $\varepsilon \to 0$ and $K^\circ := 10((1+d^{-1}) \vee d^{-2}\|\theta^\circ - \theta^\times\|_{\ell_2}^2)L^\circ$, then

$$\lim_{\varepsilon\to 0}\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon^\circ}\,|\,\mathrm{Y}}((K^\circ)^{-1}\Phi_\varepsilon^\circ \leqslant \|\boldsymbol{\vartheta}^{m_\varepsilon^\circ} - \theta^\circ\|_{\ell_2}^2 \leqslant K^\circ\Phi_\varepsilon^\circ) = 1.$$

Note that $m_\varepsilon^\circ \to \infty$ as $\varepsilon \to 0$ if and only if $\mathfrak{b}_m > 0$ for all $m \geqslant 1$. Roughly speaking, the last assertion establishes Φ_ε° as oracle posterior concentration rate for all parameter of interest θ° with components differing from the components of the prior mean $\theta^\times$ infinitely many times. However, we do not need this additional assumption to prove the next assertion which establishes Φ_ε° as oracle rate for the family $\{\widehat{\theta}^m\}_m$ of Bayes estimator and that $\widehat{\theta}^{m_\varepsilon^\circ}$ is an oracle Bayes estimator.

Theorem 3.8 (Oracle Bayes estimator). *Consider the family $\{\widehat{\theta}^m\}_m$ of Bayes estimators. Under Assumption A.2 we have* (i) *$\mathbb{E}_{\theta^\circ}\|\widehat{\theta}^{m_\varepsilon^\circ} - \theta^\circ\|_{\ell_2}^2 \leqslant (2 + d^{-2}\|\theta^\circ - \theta^\times\|_{\ell_2}^2)\Phi_\varepsilon^\circ$ and* (ii) *$\inf_{m\geqslant 1}\mathbb{E}_{\theta^\circ}\|\widehat{\theta}^m - \theta^\circ\|_{\ell_2}^2 \geqslant (1+1/d)^{-2}\Phi_\varepsilon^\circ$ for all $\varepsilon \in (0,\varepsilon_o)$.*

Note that, the oracle choice m_ε° depends on the parameter of interest θ° and thus the oracle Bayes estimator $\widehat{\theta}^{m_\varepsilon^\circ}$ as well as the associated oracle sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon^\circ}}\}_{m_\varepsilon^\circ}$ of prior distributions are generally not feasible.

3.3 Minimax concentration rate

In the spirit of a minimax theory we are interested in the following in a uniform rate over a class of parameters rather than optimising the rate for each θ° separately. Given a strictly positive and non-increasing sequence $\mathfrak{a} = \big(\mathfrak{a}_j\big)_{j\geqslant 1}$ with $\mathfrak{a}_1 = 1$ and $\lim_{j\to\infty} \mathfrak{a}_j = 0$ consider for $\theta \in \ell_2$ its weighted norm $\|\theta\|_{\mathfrak{a}}^2 := \sum_{j\geqslant 1} \theta_j^2/\mathfrak{a}_j$. We define $\ell_2^{\mathfrak{a}}$ as the completion of ℓ_2 with respect to $\|\cdot\|_{\mathfrak{a}}$. In order to formulate the optimality of the posterior concentration rate let us define

$$m_\varepsilon^\star := m_\varepsilon^\star(\mathfrak{a}, \lambda) := \underset{m\geqslant 1}{\arg\min} \left\{\mathfrak{a}_m \vee \varepsilon\, m\, \overline{\Lambda}_m\right\} \text{ and}$$
$$\Phi_\varepsilon^\star := \Phi_\varepsilon^\star(\mathfrak{a}, \lambda) := [\mathfrak{a}_{m_\varepsilon^\star} \vee \varepsilon\, m_\varepsilon^\star\, \overline{\Lambda}_{m_\varepsilon^\star}] \quad \text{for all } \varepsilon > 0. \quad (3.12)$$

We remark that $\Phi_\varepsilon^\star = o(1)$ and $m_\varepsilon^\star \to \infty$ as $\varepsilon \to 0$ since $\mathfrak{a}$ is strictly positive and tends monotonically to zero. We assume in the following that the parameter θ° belongs to the ellipsoid $\Theta_{\mathfrak{a}}^r := \{\theta \in \ell_2^{\mathfrak{a}} : \|\theta - \theta^\times\|_{\mathfrak{a}}^2 \leqslant r\}$ and therefore, $\mathfrak{b}_m(\theta^\circ) \leqslant \mathfrak{a}_m r$. Note that $\Phi_\varepsilon^\circ = \min_{m\geqslant 1}[\mathfrak{b}_m \vee \varepsilon\, m\, \overline{\Lambda}_m] \leqslant (1\vee r) \min_{m\geqslant 1}[\mathfrak{a}_m \vee \varepsilon\, m\, \overline{\Lambda}_m] = (1\vee r)\Phi_\varepsilon^\star$ and $\|\theta^\circ - \theta^\times\|_{\ell_2}^2 \leqslant r$, and hence from Theorem 3.8 it follows $\mathbb{E}_{\theta^\circ}\|\widehat{\theta}^{m_\varepsilon^\circ} - \theta^\circ\|_{\ell_2}^2 \leqslant (2 + r/d^2)(1 \vee r)\Phi_\varepsilon^\star$. On the other hand side, given an estimator $\widehat{\theta}$ of θ let $\sup_{\theta\in\Theta_{\mathfrak{a}}^r} \mathbb{E}_\theta\|\widehat{\theta} - \theta\|_{\ell_2}^2$ denote the maximal mean integrated squared error over the class $\Theta_{\mathfrak{a}}^r$. It has been shown in Johannes and Schwarz [2013] that $\Phi_\varepsilon^\star$ provides up to a constant a lower bound for the maximal MISE over the class $\Theta_{\mathfrak{a}}^r$ (assuming a prior mean $\theta^\times = 0$) if the next assumption is satisfied.

Assumption A.3. *Let $\mathfrak{a}$ and λ be sequences such that*

$$0 < \kappa^\star := \kappa^\star(\mathfrak{a}, \lambda) := \inf_{0<\varepsilon<\varepsilon_o} \left\{(\Phi_\varepsilon^\star)^{-1}[\mathfrak{a}_{m_\varepsilon^\star} \wedge \varepsilon\, m_\varepsilon^\star\, \overline{\Lambda}_{m_\varepsilon^\star}]\right\} \leqslant 1. \quad (3.13)$$

We may emphasise that under Assumption A.3 the rate $\Phi_\varepsilon^\star = \Phi_\varepsilon^\star(\mathfrak{a}, \lambda)$ is optimal in a minimax sense and the Bayes estimate $\widehat{\theta}^{m_\varepsilon^\circ}$ attains the minimax rate up to a constant. However, the dimension parameter m_ε° depends still on the parameter of interest θ°. Therefore, let us consider the Bayes estimate $\widehat{\theta}^{m_\varepsilon^\star}$ and the sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon^\star}}\}_{m_\varepsilon^\star}$ of prior distributions which do not depend anymore on the parameter of interest θ° but only on the set of possible parameters $\Theta_{\mathfrak{a}}^r$ characterised by the weight sequence $\mathfrak{a}$. The next assertion can be shown along the lines of the proof of Theorem 3.8, and, hence we omit its proof.

Theorem 3.9 (Minimax optimal Bayes estimator). *Let Assumption A.2 be satisfied. Considering the Bayes estimator* $\widehat{\theta}^{m_\varepsilon^\star} := \mathbb{E}[\boldsymbol{\vartheta}^{m_\varepsilon^\star} \,|\, Y]$ *we have*

$$\sup_{\theta^\circ \in \Theta_{\mathfrak{a}}^r} \mathbb{E}_{\theta^\circ} \|\widehat{\theta}^{m_\varepsilon^\star} - \theta^\circ\|_{\ell_2}^2 \leqslant (2 + r/d^2)(1 \vee r)\Phi_\varepsilon^\star \quad \text{for all } \varepsilon \in (0, \varepsilon_o).$$

The last assertion establishes the minimax optimality of the Bayes estimate $\widehat{\theta}^{m_\varepsilon^\star}$ over the class $\Theta_{\mathfrak{a}}^r$. Moreover, the minimax rate $\Phi_\varepsilon^\star$ provides up to a constant a lower and an upper bound for the posterior concentration rate associated with the prior sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon^\star}}\}_{m_\varepsilon^\star}$, which is summarised in the next assertion.

Theorem 3.10 (Minimax optimal posterior concentration rate). *Let Assumption A.2 and A.3 hold true. If there exists a constant* $1 \leqslant L^\star := L^\star(\mathfrak{a}, \lambda) < \infty$ *such that*

$$\sup_{0<\varepsilon<\varepsilon_o} \varepsilon\, m_\varepsilon^\star\, \Lambda_{(m_\varepsilon^\star)} (\Phi_\varepsilon^\star)^{-1} \leqslant L^\star \tag{3.14}$$

and $K^\star := K^\star(r, \mathfrak{a}, \lambda, d, \kappa) := 10((1 + 1/d) \vee r/d^2)(1 \vee r)(L^\star/\kappa^\star)$*, then*

$$\lim_{\varepsilon\to 0} \inf_{\theta^\circ \in \Theta_{\mathfrak{a}}^r} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon^\star} \,|\, Y}((K^\star)^{-1}\Phi_\varepsilon^\star \leqslant \|\boldsymbol{\vartheta}^{m_\varepsilon^\star} - \theta^\circ\|_{\ell_2}^2 \leqslant K^\star \Phi_\varepsilon^\star) = 1.$$

Comparing the last result with the result of Theorem 3.7 and keeping in mind that $(1 \vee r)\Phi_\varepsilon^\star \geqslant \Phi_\varepsilon^\circ$, the posterior concentration rate associated with the prior sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon^\star}}\}_{m_\varepsilon^\star}$ is of order of the minimax rate $\Phi_\varepsilon^\star$ uniformly for all parameter of interest $\theta^\circ \in \Theta_{\mathfrak{a}}^r$. However, for certain parameter θ° the minimax rate $\Phi_\varepsilon^\star$ may be far slower than the oracle rate Φ_ε°. For example, as shown in case **[P-P]** in the following illustration the minimax rate $\Phi_\varepsilon^\star$ is of order $O(\varepsilon^{2p/(2a+2p+1)})$ while it is not hard to see, that for all parameter θ° with $\mathfrak{b}_m \asymp \exp(-m^{2p})$ the oracle rate is of order $O(\varepsilon|\log\varepsilon|^{(2a+1)/(2p)})$ (see case **[E-P]**). Moreover, the optimal choice $m_\varepsilon^\star$ of the dimension parameter still depends on the class $\Theta_{\mathfrak{a}}^r$, which might be unknown in practise, therefore we will consider in the next section a fully data-driven choice using a hierarchical specification of the prior distribution.

Illustration 1. *We illustrate the last assumptions and the minimax rate for typical choices of the sequences* $\mathfrak{a}$ *and* λ*. For two strictly positive sequences* $(a_j)_{j\geqslant 1}$ *and* $(b_j)_{j\geqslant 1}$ *we write* $a_j \asymp b_j$*, if* $(a_j/b_j)_{j\geqslant 1}$ *is bounded away from 0 and infinity.*

[P-P] *Consider* $\mathfrak{a}_j \asymp j^{-2p}$ *and* $\lambda_j^2 \asymp j^{-2a}$ *with* $p > 0$ *and* $a > 0$ *then* $m_\varepsilon^\star \asymp \varepsilon^{-1/(2p+2a+1)}$ *and* $\Phi_\varepsilon^\star \asymp \varepsilon^{2p/(2a+2p+1)}$.

[E-P] *Consider* $\mathfrak{a}_j \asymp \exp(-j^{2p} + 1)$ *and* $\lambda_j^2 \asymp j^{-2a}$ *with* $p > 0$ *and* $a > 0$ *then* $m_\varepsilon^\star \asymp |\log\varepsilon - \frac{2a+1}{2p}(\log|\log\varepsilon|)|^{1/(2p)}$ *and* $\Phi_\varepsilon^\star \asymp \varepsilon|\log\varepsilon|^{(2a+1)/(2p)}$.

[P-E] *Consider* $\mathfrak{a}_j \asymp j^{-2p}$ *and* $\lambda_j^2 \asymp \exp(-j^{2a}+1)$, *with* $p>0$ *and* $a>0$ *then* $m_\varepsilon^\star \asymp |\log\varepsilon - \frac{2p+(2a-1)_+}{2a}(\log|\log\varepsilon|)|^{1/(2a)}$ *and* $\Phi_\varepsilon^\star \asymp |\log\varepsilon|^{-p/a}$.

In all three cases Assumption A.3 and (3.14) hold true. □

4 Data-driven Bayesian estimation

We will derive in this section a concentration rate given the aforementioned hierarchical prior distribution. For this purpose we impose additional conditions on the behaviour of the sequence $\lambda = (\lambda_j)_{j\geqslant 1}$.

Assumption A.4. *There exist finite constants* $C_\lambda \geqslant 1$ *and* $L_\lambda \geqslant 1$ *such that for all* $k,l \in \mathbb{N}$ *hold* (i) $\max_{j>k}\lambda_j^2 \leqslant C_\lambda \min_{1\leqslant j\leqslant k}\lambda_j^2 = C_\lambda \Lambda_{(k)}^{-1}$; (ii) $\Lambda_{(kl)} \leqslant \Lambda_{(k)}\Lambda_{(l)}$; (iii) $1 \leqslant \Lambda_{(k)}/\overline{\Lambda}_k \leqslant L_\lambda$.

We may emphasise that Assumption A.4 (i) holds trivially with $C_\lambda = 1$ if the sequence λ is monotonically decreasing. Moreover, considering the typical choices of the sequence λ presented in Illustration 1, Assumption A.4 (ii) and (iii) hold only true in case of a polynomial decay, i.e., **[P-P]** and **[E-P]**. In other words, Assumption A.4 excludes an exponential decay of λ, i.e., **[P-E]**.

Assumption A.5. *Let* $\theta^\times$, θ° *and* λ *be sequences such that*

$$0 < \kappa^\circ := \kappa^\circ(\theta^\times,\theta^\circ,\lambda) := \inf_{0<\varepsilon<\varepsilon_o}\left\{(\Phi_\varepsilon^\circ)^{-1}[\mathfrak{b}_{m_\varepsilon^\circ} \wedge \varepsilon\, m_\varepsilon^\circ\, \overline{\Lambda}_{m_\varepsilon^\circ}]\right\} \leqslant 1. \tag{4.1}$$

Observe that $\mathfrak{b}_{m_\varepsilon^\circ} \geqslant \kappa^\circ\Phi_\varepsilon^\circ > 0$ due to Assumption A.5 which in turn implies $\mathfrak{b}_k > 0$ for all $k \in \mathbb{N}$ and, hence $m_\varepsilon^\circ \to \infty$ as $\varepsilon \to 0$. Indeed, if there exists $K \in \mathbb{N}$ such that $\mathfrak{b}_K = 0$ and $\mathfrak{b}_{K-1} > 0$ then there exists $\varepsilon_\circ \in (0,1)$ with $\varepsilon_\circ K\overline{\Lambda}_K < \mathfrak{b}_{K-1}$ and for all $\varepsilon \in (0,\varepsilon_\circ)$ it is easily seen that $m_\varepsilon^\circ = K$ and hence $\mathfrak{b}_{m_\varepsilon^\circ} = 0$. Moreover, due to Assumption A.4 (iii) there exists a constant L_λ depending only on λ such that $\varepsilon\, m_\varepsilon^\circ \Lambda_{(m_\varepsilon^\circ)}(\Phi_\varepsilon^\circ)^{-1} \leqslant \Lambda_{(m_\varepsilon^\circ)}(\overline{\Lambda}_{m_\varepsilon^\circ})^{-1} \leqslant L_\lambda$, i.e., condition (3.7) holds true uniformly for all parameters $\theta \in \ell_2$. If we suppose in addition to Assumption A.4 and A.5 that the sequence of prior variances meets Assumption A.2 and that $m_\varepsilon^\circ \to \infty$ as $\varepsilon \to 0$, then the assumptions of Theorem 3.7 are satisfied and Φ_ε° provides up to a constant an upper and lower bound of the posterior concentration rate associated with the oracle prior sub-family $\{P_{m_\varepsilon^\circ}\}_{m_\varepsilon^\circ}$.

Let us specify the prior distribution P_{M} of the thresholding parameter M taking its values in $\{1,\dots,G_\varepsilon\}$ with G_ε as in Assumption A.2, and for $1 \leqslant m \leqslant G_\varepsilon$

$$p_{\mathrm{M}}(m) := P_{\mathrm{M}}(\mathrm{M}=m) = \frac{\exp(-3C_\lambda m/2)\prod_{j=1}^m(\varsigma_j/\sigma_j)^{1/2}}{\sum_{k=1}^{G_\varepsilon}\exp(-3C_\lambda k/2)\prod_{j=1}^k(\varsigma_j/\sigma_j)^{1/2}}. \tag{4.2}$$

Keeping in mind the sequences $\theta^{\mathrm{Y}} = (\theta_j^{\mathrm{Y}})_{j\geqslant 1}$ and $\sigma = (\sigma_j)_{j\geqslant 1}$ of conditional means and variances, respectively, given by $\theta_j^{\mathrm{Y}} = \sigma_j(\lambda_j\varepsilon^{-1}Y_j + \varsigma_j^{-1}\theta_j^\times)$ and $\sigma_j = (\varsigma_j^{-1} + \lambda_j^2\varepsilon^{-1})^{-1}$, for each $m \in \mathbb{N}$ the sequence $\widehat{\theta}^m = \big(\widehat{\theta}^m_j\big)_{j\geqslant 1} = \mathbb{E}[\boldsymbol{\vartheta}^m \,|\, \mathrm{Y}]$ of posterior means of $\boldsymbol{\vartheta}^m$ satisfies $\widehat{\theta}^m_j = \theta^{\mathrm{Y}}_j \,\mathbb{1}_{\{1\leqslant j\leqslant m\}} + \theta^{\times}_j \,\mathbb{1}_{\{j>m\}}$. Introducing further the weighted norm $\|\theta\|^2_\sigma := \sum_{j\geqslant 1}\theta_j^2/\sigma_j$ for $\theta \in \ell_2$ the posterior distribution $P_{\mathrm{M}\,|\,\mathrm{Y}}$ of the thresholding parameter M is given by

$$p_{\mathrm{M}\,|\,\mathrm{Y}}(m) = P_{\mathrm{M}\,|\,\mathrm{Y}}(\mathrm{M}=m) = \frac{\exp(-\frac{1}{2}\{-\|\widehat{\theta}^m - \theta^\times\|^2_\sigma + 3C_\lambda m\})}{\sum_{k=1}^{G_\varepsilon}\exp(-\frac{1}{2}\{-\|\widehat{\theta}^k - \theta^\times\|^2_\sigma + 3C_\lambda k\})} \tag{4.3}$$

Interestingly, the posterior distribution $P_{\mathrm{M}\,|\,\mathrm{Y}}$ of the thresholding parameter M is concentrating around the oracle dimension parameter m°_ε as ε tends to zero. To be more precise, there exists $\varepsilon_\circ \in (0,1)$ such that $m^\circ_\varepsilon \leqslant G_\varepsilon$ for all $\varepsilon \in (0,\varepsilon_\circ)$ since $\Phi^\circ_\varepsilon = o(1)$ for $\varepsilon \to 0$. Let us further define for all $\varepsilon \in (0,\varepsilon_\circ)$

$$G^-_\varepsilon := \min\{m \in \{1,\dots,m^\circ_\varepsilon\} : \mathfrak{b}_m \leqslant 8L_\lambda C_\lambda(1+1/d)\Phi^\circ_\varepsilon\} \quad \text{and}$$
$$G^+_\varepsilon := \max\big\{m \in \{m^\circ_\varepsilon,\dots,G_\varepsilon\} : m \leqslant 5L_\lambda(\varepsilon\Lambda_{(m^\circ_\varepsilon)})^{-1}\Phi^\circ_\varepsilon\big\} \tag{4.4}$$

where the defining sets are not empty under Assumption A.4 since $8L_\lambda C_\lambda(1+1/d)\Phi^\circ_\varepsilon \geqslant 8L_\lambda C_\lambda(1+1/d)\mathfrak{b}_{m^\circ_\varepsilon} \geqslant \mathfrak{b}_{m^\circ_\varepsilon}$ and $5L_\lambda(\varepsilon\Lambda_{(m^\circ_\varepsilon)})^{-1}\Phi^\circ_\varepsilon \geqslant 5m^\circ_\varepsilon \geqslant m^\circ_\varepsilon$. Moreover, under Assumption A.5 it is easily verified that $G^-_\varepsilon \to \infty$ as $\varepsilon \to 0$.

Lemma 4.1. *If Assumptions A.2 and A.4 hold true then for all $\varepsilon \in (0,\varepsilon_\circ)$*

(i) $\mathbb{E}_{\theta^\circ}P_{\mathrm{M}\,|\,\mathrm{Y}}(1 \leqslant \mathrm{M} < G^-_\varepsilon) \leqslant 2\exp\big(-\frac{7C_\lambda}{32}m^\circ_\varepsilon + \log G_\varepsilon\big) \leqslant 2\exp\big(-\frac{C_\lambda}{5}m^\circ_\varepsilon + \log G_\varepsilon\big)$;

(ii) $\mathbb{E}_{\theta^\circ}P_{\mathrm{M}\,|\,\mathrm{Y}}(G^+_\varepsilon < \mathrm{M} \leqslant G_\varepsilon) \leqslant 2\exp\big(-\frac{4C_\lambda}{9}m^\circ_\varepsilon + \log G_\varepsilon\big) \leqslant 2\exp\big(-\frac{C_\lambda}{5}m^\circ_\varepsilon + \log G_\varepsilon\big)$.

Recall that $m^\circ_\varepsilon \to \infty$ as $\varepsilon \to 0$ under Assumption A.5. If in addition $m^\circ_\varepsilon/(\log G_\varepsilon) \to \infty$ as $\varepsilon \to 0$ then Lemma 4.1 states that the posterior distribution of the thresholding parameter M is vanishing outside the set $\{G^-_\varepsilon,\dots,G^+_\varepsilon\}$ as $\varepsilon \to 0$. On the other hand side, the posterior distribution $P_{\boldsymbol{\vartheta}^{\mathrm{M}}\,|\,\mathrm{Y}}$ of $\boldsymbol{\vartheta}^{\mathrm{M}} = \big(\boldsymbol{\vartheta}^{\mathrm{M}}_j\big)_{j\geqslant 1}$ associated with the hierarchical prior is a weighted mixture of the posterior distributions $\{P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}\}_{m=1}^{G_\varepsilon}$ studied in section 3, that is, $P_{\boldsymbol{\vartheta}^{\mathrm{M}}\,|\,\mathrm{Y}} = \sum_{m=1}^{G_\varepsilon} p_{\mathrm{M}\,|\,\mathrm{Y}}(m)P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}$. The next assertion shows that considering posterior distributions $\{P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}\}_{m=G^-_\varepsilon}^{G^+_\varepsilon}$ associated with thresholding parameters belonging to $\{G^-_\varepsilon,\dots,G^+_\varepsilon\}$ only, then their concentration rate equals Φ°_ε up to a constant.

Lemma 4.2. *If Assumptions A.2, A.4 and A.5 hold true then for all $\varepsilon \in (0,\varepsilon_\circ)$*

(i) $\sum_{G^-_\varepsilon\leqslant m\leqslant G^+_\varepsilon}\mathbb{E}_{\theta^\circ}P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}\big(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2_{\ell_2} > K^o\Phi^\circ_\varepsilon\big) \leqslant 74\exp(-G^-_\varepsilon/36)$;

(ii) $\sum_{G^-_\varepsilon\leqslant m\leqslant G^+_\varepsilon}\mathbb{E}_{\theta^\circ}P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}\big(\|\boldsymbol{\vartheta}^m - \theta^\circ\|^2_{\ell_2} < (K^o)^{-1}\Phi^\circ_\varepsilon\big) \leqslant 4(K^\circ)^2\exp(-G^-_\varepsilon/(K^\circ)^2)$,

where $K^\circ := 10((1+1/d) \vee \|\theta^\circ - \theta^\times\|^2_{\ell_2}/d^2)L^2_\lambda(8C_\lambda(1+1/d) \vee D^\circ\Lambda_{(D^\circ)})$ *with* $D^\circ := D^\circ(\theta^\times,\theta^\circ,\lambda) := \lceil 5L_\lambda/\kappa^\circ\rceil$.

From Lemma 4.1 and 4.2 we derive next upper and lower bounds for the concentration rate of the posterior distribution $P_{\boldsymbol{\vartheta}^{\mathrm{M}}\,|\,\mathrm{Y}}$ by decomposing the weighted mixture into three parts with respect to G_ε^- and G_ε^+ which we bound separately.

Theorem 4.3 (Oracle posterior concentration rate). *Let Assumptions A.2, A.4 and A.5 hold true. If in addition $(\log G_\varepsilon)/m_\varepsilon^\circ \to 0$ as $\varepsilon \to 0$, then*

$$\lim_{\varepsilon\to 0} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{\mathrm{M}}\,|\,\mathrm{Y}}((K^\circ)^{-1}\Phi_\varepsilon^\circ \leqslant \|\boldsymbol{\vartheta}^{\mathrm{M}} - \theta^\circ\|_{\ell_2}^2 \leqslant K^\circ \Phi_\varepsilon^\circ) = 1$$

where K° is given in Lemma 4.2.

We shall emphasise that the Bayes estimator $\widehat{\theta} := \big(\widehat{\theta}_j\big)_{j\geqslant 1} := \mathbb{E}[\boldsymbol{\vartheta}^{\mathrm{M}}\,|\,\mathrm{Y}]$ associated with the hierarchical prior and given by $\widehat{\theta}_j = \theta_j^\times$ for $j > G_\varepsilon$ and $\widehat{\theta}_j = \theta_j^\times P(1 \leqslant \mathrm{M} < j|\,\mathrm{Y}) + \theta_j^{\mathrm{Y}} P(j \leqslant \mathrm{M} \leqslant G_\varepsilon|\,\mathrm{Y})$ for all $1 \leqslant j \leqslant G_\varepsilon$, does not take into account any prior information related to the parameter of interest, and hence it is fully data-driven. The next assertion provides an upper bound of its MISE.

Theorem 4.4 (Oracle optimal Bayes estimator). *Under Assumptions A.2, A.4 and A.5 consider the Bayes estimator $\widehat{\theta} := \mathbb{E}[\boldsymbol{\vartheta}^{\mathrm{M}}\,|\,\mathrm{Y}]$. If in addition $\log(G_\varepsilon/\Phi_\varepsilon^\circ)/m_\varepsilon^\circ \to 0$ as $\varepsilon \to 0$, then there exists a constant $K^\circ := K^\circ(\theta^\circ, \theta^\times, \lambda, d, L) < \infty$ such that $\mathbb{E}_{\theta^\circ}\|\widehat{\theta} - \theta^\circ\|_{\ell_2}^2 \leqslant K^\circ \Phi_\varepsilon^\circ$ for all $\varepsilon \in (0, \varepsilon_\circ)$.*

Both Theorems, 4.3 and 4.4 hold true only under Assumption A.5, which we have seen before imposes an additional restriction on the parameter of interest θ°, i.e., its components differ from the components of the prior mean $\theta^\times$ infinitely many times. However, for all parameters of interest satisfying Assumption A.5, the hierarchical prior sequence allows to recover the oracle posterior concentration rate and the fully data driven Bayes estimator attains the oracle rate. In the last part of this section we show that for all $\theta^\circ \in \Theta_{\mathfrak{a}}^r$ the posterior concentration rate and the MISE of the Bayes estimator associated with the hierarchical prior are bounded from above by the minimax rate $\Phi_\varepsilon^\star$ up to a constant. In other words, the fully data-driven hierarchical prior and the associated Bayes estimator are minimax-rate optimal.

Recall the definition (3.12) of $m_\varepsilon^\star$ and $\Phi_\varepsilon^\star$. Consider the prior distribution P_{M} of the thresholding parameter M, and observe that there exists $\varepsilon_\star$ such that $m_\varepsilon^\star \leqslant G_\varepsilon$ for all $\varepsilon \in (0, \varepsilon_\star)$ since $\Phi_\varepsilon^\star = o(1)$ as $\varepsilon \to 0$. Remark that $\varepsilon\, m_\varepsilon^\star \Lambda_{(m_\varepsilon^\star)} (\Phi_\varepsilon^\star)^{-1} \leqslant \Lambda_{(m_\varepsilon^\star)} (\overline{\Lambda}_{m_\varepsilon^\star})^{-1} \leqslant L_\lambda$ with L_λ depending only on λ due to Assumption A.4 (iii), i.e., condition (3.14) holds true uniformly for all parameters $\theta \in \ell_2$. If we assume in addition that the sequence of prior variances satisfies Assumption A.2 and that Assumption A.3 holds true, then the conditions of Theorem 3.10 are satisfied and $\Phi_\varepsilon^\star$ provides up to a constant an upper and lower bound of the posterior concentration rate associated with the minimax prior sub-family $\{P_{m_\varepsilon^\star}\}_{m_\varepsilon^\star}$. On the other hand side, the posterior distribution $P_{\mathrm{M}\,|\,\mathrm{Y}}$ of the thresholding parameter M is concentrating around the minimax-optimal dimension parameter $m^\star_\varepsilon$ as ε tends to zero. To be more precise, for $\varepsilon \in (0, \varepsilon_\star)$ let us define

$$G^{\star-}_\varepsilon := \min\{m \in \{1, \dots, m^\star_\varepsilon\} : \mathfrak{b}_m \leqslant 8L_\lambda C_\lambda(1 + 1/d)(1 \vee r)\Phi^\star_\varepsilon\} \quad \text{and}$$
$$G^{\star+}_\varepsilon := \max\big\{m \in \{m^\star_\varepsilon, \dots, G_\varepsilon\} : m \leqslant 5L_\lambda(\varepsilon\Lambda_{(m^\star_\varepsilon)})^{-1}(1 \vee r)\Phi^\star_\varepsilon\big\} \quad (4.5)$$

where the defining sets are not empty under Assumption A.4 since $8L_\lambda C_\lambda(1 + 1/d)(1 \vee r)\Phi^\star_\varepsilon \geqslant 8L_\lambda C_\lambda(1+1/d)r\mathfrak{a}_{m^\star_\varepsilon} \geqslant 8L_\lambda C_\lambda(1+1/d)\mathfrak{b}_{m^\star_\varepsilon} \geqslant \mathfrak{b}_{m^\star_\varepsilon}$ and $5L_\lambda(\varepsilon\Lambda_{(m^\star_\varepsilon)})^{-1}(1\vee r)\Phi^\star_\varepsilon \geqslant 5m^\star_\varepsilon \geqslant m^\star_\varepsilon$. Moreover, it is again straightforward to see that $G^{\star-}_\varepsilon \to \infty$ as $\varepsilon \to 0$.

Lemma 4.5. *If Assumption A.2 and A.4 hold true then for all $\theta^\circ \in \Theta^r_\mathfrak{a}$ and $\varepsilon \in (0, \varepsilon_\star)$*

(i) $\mathbb{E}_{\theta^\circ} P_{\mathrm{M}\,|\,\mathrm{Y}}(1 \leqslant \mathrm{M} < G^{\star-}_\varepsilon) \leqslant 2\exp\big(-\frac{C_\lambda(1\vee r)}{5}m^\star_\varepsilon + \log G_\varepsilon\big)$;

(ii) $\mathbb{E}_{\theta^\circ} P_{\mathrm{M}\,|\,\mathrm{Y}}(G^{\star+}_\varepsilon < \mathrm{M} \leqslant G_\varepsilon) \leqslant 2\exp\big(-\frac{C_\lambda(1\vee r)}{5}m^\star_\varepsilon + \log G_\varepsilon\big)$.

By employing Lemma 4.5 we show next for each $\theta^\circ \in \Theta^r_\mathfrak{a}$ that the minimax rate $\Phi^\star_\varepsilon$ provides up to a constant an upper bound for the posterior concentration rate associated with the fully data-driven hierarchical prior distribution $P_{\boldsymbol{\vartheta}^{\mathrm{M}}}$.

Theorem 4.6 (Minimax optimal posterior concentration rate). *Let Assumption A.2, A.3 and A.4 hold true. If in addition $(\log G_\varepsilon)/m^\star_\varepsilon \to 0$ as $\varepsilon \to 0$, then*

(i) *for all $\theta^\circ \in \Theta^r_\mathfrak{a}$ we have*

$$\lim_{\varepsilon\to 0} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{\mathrm{M}}\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^{\mathrm{M}} - \theta^\circ\|^2_{\ell_2} \leqslant K^\star\Phi^\star_\varepsilon) = 1$$

where $K^\star := 16((1 + 1/d) \vee r/d^2)L^2_\lambda(8C_\lambda(1 + 1/d) \vee D^\star\Lambda_{(D^\star)})(1 \vee r)$ with $D^\star := D^\star(\mathfrak{a}, \lambda) := \lceil 5L_\lambda/\kappa^\star\rceil$;

(ii) *for any monotonically increasing and unbounded sequence $(K_\varepsilon)_\varepsilon$ holds*

$$\lim_{\varepsilon\to 0} \inf_{\theta^\circ\in\Theta^r_\mathfrak{a}} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{\mathrm{M}}\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^{\mathrm{M}} - \theta^\circ\|^2_{\ell_2} \leqslant K_\varepsilon\Phi^\star_\varepsilon) = 1.$$

We shall emphasise that due to Theorem 4.3 for all $\theta^\circ \in \Theta^r_\mathfrak{a}$ satisfying Assumption A.5 the posterior concentration rate associated with the hierarchical prior attains the oracle rate Φ°_ε which might be far smaller than the minimax-rate $\Phi^\star_\varepsilon$. Consequently, the minimax rate cannot provide an uniform lower bound over $\Theta^r_\mathfrak{a}$ for the posterior concentration rate associated with the hierarchical prior. However, due to Theorem 4.6 the posterior concentration rate is for all $\theta^\circ \in \Theta^r_\mathfrak{a}$, independently that Assumption A.5 holds, at least of the order of the minimax rate $\Phi^\star_\varepsilon$. The next assertion establishes the minimax-rate optimality of the fully data-driven Bayes estimator.

Theorem 4.7 (Minimax optimal Bayes estimate). *Under Assumption A.2, A.3 and A.4 consider the Bayes estimator $\widehat{\theta} := \mathbb{E}[\boldsymbol{\vartheta}^{\mathrm{M}} \,|\, \mathrm{Y}]$. If in addition $\log(G_\varepsilon/\Phi_\varepsilon^\star)/m_\varepsilon^\star \to 0$ as $\varepsilon \to 0$, then there exists $K^\star := K^\star(\Theta_\mathfrak{a}^r, \lambda, d) < \infty$ such that $\sup_{\theta^\circ \in \Theta_\mathfrak{a}^r} \mathbb{E}_{\theta^\circ} \|\widehat{\theta} - \theta^\circ\|_{\ell_2}^2 \leqslant K^\star \Phi_\varepsilon^\star$ for all $\varepsilon \in (0, \varepsilon_\star)$.*

Let us briefly comment on the last assertion by considering again the improper specification of the prior family $\{P_{\boldsymbol{\vartheta}^m}\}_m$ introduced in Section 2. Recall that in this situation for each $m \in \mathbb{N}$ the Bayes estimator $\widehat{\theta}^m = \mathbb{E}[\boldsymbol{\vartheta}^m \,|\, \mathrm{Y}]$ of $\boldsymbol{\vartheta}^m$ equals an orthogonal projection estimator, i.e., $\widehat{\theta}^m = (Y/\lambda)^m$. Moreover, the posterior probability of the thresholding parameter M taking a value $m \in \{1, \dots, G_\varepsilon\}$ is proportional to $\exp(-\frac{1}{2}\{-\|(Y/\lambda)^m\|_{\varepsilon\Lambda}^2 + 3C_\lambda m\})$, and hence the data-driven Bayes estimator $\widehat{\theta} = \big(\widehat{\theta}_j\big)_{j\geqslant 1} = \mathbb{E}[\boldsymbol{\vartheta}^{\mathrm{M}} \,|\, \mathrm{Y}]$ equals the shrinked orthogonal projection estimator given by

$$\widehat{\theta}_j = \frac{\sum_{m=j}^{G_\varepsilon} \exp(-\frac{1}{2}\{-\|(Y/\lambda)^m\|_{\varepsilon\Lambda}^2 + 3C_\lambda m\})}{\sum_{m=1}^{G_\varepsilon} \exp(-\frac{1}{2}\{-\|(Y/\lambda)^m\|_{\varepsilon\Lambda}^2 + 3C_\lambda m\})} \times \frac{Y_j}{\lambda_j} \mathbb{1}_{\{1 \leqslant j \leqslant G_\varepsilon\}}\,.$$

From Theorem 4.7 it follows now, that the fully data-driven shrinkage estimator $\widehat{\theta}$ is minimax-optimal up to a constant for a wide variety of parameter spaces $\Theta_\mathfrak{a}^r$ provided Assumptions A.3 and A.4 hold true. Interestingly, identifying $\Upsilon(\widehat{\theta}^m) := -(1/2)\|(Y/\lambda)^m\|_{\varepsilon\Lambda}^2$ as a contrast and $\mathrm{pen}_m := 3/2 C_\lambda m$ as a penalty term the j-th shrinkage weight is proportional to $\sum_{m=j}^{G_\varepsilon} \exp(-\{\Upsilon(\widehat{\theta}^m) + \mathrm{pen}_m\})$. Roughly speaking, in comparison to a classical model selection approach where a data-driven estimator $\widehat{\theta}^{\widehat{m}} = (Y/\lambda)^{\widehat{m}}$ is obtained by selecting the dimension parameter $\widehat{m}$ as minimum of a penalised contrast criterion over a class of admissible models $\{1, \dots, G_\varepsilon\}$, i.e., $\widehat{m} = \arg\min_{1 \leqslant m \leqslant G_\varepsilon}\{\Upsilon(\widehat{\theta}^m) + \mathrm{pen}_m\}$, following the Bayesian approach each of the G_ε components of the data-driven Bayes estimator is shrunk proportional to the associated values of the penalised contrast criterion.

Conclusions and perspectives. In this paper we have presented a hierarchical prior leading to a fully-data driven Bayes estimator that is minimax-optimal in an indirect sequence space model. Obviously, the concentration rate based on a hierarchical prior in an indirect sequence space model with additional noise in the eigenvalues is only one amongst the many interesting questions for further research and we are currently exploring this topic. Moreover, inspired by the specific form of the fully-data driven Bayes estimator, as discussed in the last section, we are currently studying the effect of different choices for the contrast and the penalty term on the properties of the estimator.

Acknowledgements. The work of Jan Johannes and Rudolf Schenk was supported by the IAP research network no. P7/06 of the Belgian Government (Belgian Science Policy), by the “Fonds Spéciaux de Recherche” from the Université catholique de Louvain and by the ARC contract 12/17-045 of the "Communauté française de Belgique", granted by the Académie universitaire Louvain. Anna Simoni thanks financial support from ANR-13-BSH1-0004 (IPANEMA) and from the Collaborative Research Center 884.

A Appendix: Proofs of Section 3

Proof of Lemma 3.1. Let $X_j = \beta_j Z_j + \alpha_j$ with independent and standard normally distributed random variables $\{Z_j\}_{j=1}^m$. We start our proof with the observation that $\mathbb{E}(S_m) = \sum_{j=1}^m \{\beta_j^2 + \alpha_j^2\}$ and define $\Sigma_m := \frac{1}{2}\sum_{j=1}^m \mathbb{V}\mathrm{ar}\left((\beta_j Z_j + \alpha_j)^2\right) = \sum_{j=1}^m \beta_j^2(\beta_j^2 + 2\alpha_j^2)$. Let $t_m := \max_{1\leqslant j\leqslant m}\beta_j^2$ and by using that $v_m \geqslant \sum_{j=1}^m \beta_j^2$ and $r_m \geqslant \sum_{j=1}^m \alpha_j^2$ we have $\mathbb{E}(S_m) \leqslant v_m + r_m$ and $\Sigma_m \leqslant t_m\,(v_m + 2r_m)$. These bounds are used below without further reference. There exist several results of tail bound for sums of independent squared Gaussian random variables and we present next a version which is due to Birgé [2001] and can be shown following the lines of the proof of Lemma 1 in Laurent et al. [2012]. For all $x > 0$ we have

$$P(S_m - \mathbb{E}S_m \geqslant 2\sqrt{\Sigma_m x} + 2t_m x) \leqslant \exp(-x) \qquad \text{and}$$
$$P(S_m - \mathbb{E}S_m \leqslant -2\sqrt{\Sigma_m x}) \leqslant \exp(-x). \quad \text{(A.1)}$$

Consider (3.2). Keeping in mind that for all $c \geqslant 0$, $(3/2)c(v_m + 2r_m) \geqslant c(v_m + 2r_m) + 2t_m c(c\wedge 1)(v_m + 2r_m)/(4t_m)$ and $(c\vee 1)t_m(v_m + 2r_m) \geqslant \Sigma_m$ we conclude for $x := c(c\wedge 1)(v_m + 2r_m)/(4t_m)$ that $(3/2)c(v_m + 2r_m) \geqslant 2\sqrt{\Sigma_m x} + 2t_m x$ and hence by employing the first exponential bound in (A.1) we obtain (3.2). On the other hand side, since $c(v_m + 2r_m) \geqslant 2\sqrt{\Sigma_m x}$ for all $c \geqslant 0$ assertion (3.1) follows by employing the second exponential bound in (A.1), which completes the proof. □

Proof of Proposition 3.2. We intend to apply the technical Lemma 3.1. Consider first the assertion (3.3). Let s_m and c_1 be positive constants (to be specified below). Keeping in mind that the posterior distribution of $\boldsymbol{\vartheta}_j^m$ given Y_j is degenerated on $\theta_j^\times$ for $j > m$ and that $\mathfrak{b}_m = \sum_{j>m}(\theta_j^\circ - \theta_j^\times)^2$ we have

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}\left(\|\boldsymbol{\vartheta}^m - \theta^\circ\|_{\ell_2}^2 > \mathfrak{b}_m + m\overline{\sigma}_m + \frac{3c_1}{2}m\sigma_{(m)} + (3c_1+1)s_m\right)$$
$$= \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}\left(\sum_{j=1}^m(\boldsymbol{\vartheta}_j^m - \theta_j^\circ)^2 > m\overline{\sigma}_m + \frac{3c_1}{2}m\sigma_{(m)} + (3c_1+1)s_m\right).$$

Define $S_m^{\boldsymbol{\vartheta}^m} := \sum_{j=1}^m(\boldsymbol{\vartheta}_j^m - \theta_j^\circ)^2$ where conditional on Y the random variables $\{\boldsymbol{\vartheta}_j^m - \theta_j^\circ\}_{j=1}^m$ are independent and normally distributed with conditional mean $\theta_j^{\mathrm{Y}} - \theta_j^\circ$ and conditional variance σ_j. Observe that $m\overline{\sigma}_m = \sum_{j=1}^m \sigma_j$ and $\mathbb{E}_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}[S_m^{\boldsymbol{\vartheta}^m}] = m\overline{\sigma}_m + \sum_{j=1}^m(\theta_j^{\mathrm{Y}} - \theta_j^\circ)^2$.

Introduce the event $\Omega_m := \Big\{\sum_{j=1}^m(\theta_j^{\mathrm{Y}} - \theta_j^\circ)^2 \leqslant s_m\Big\}$ where obviously $\mathbb{1}_{\Omega_m}\, \mathbb{E}_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}[S_m^{\boldsymbol{\vartheta}^m}] \leqslant m\overline{\sigma}_m + s_m$ and hence,

$$\mathbb{E}_{\theta^\circ}\, \mathbb{1}_{\Omega_m}\, P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}\Big(S_m^{\boldsymbol{\vartheta}^m} > m\overline{\sigma}_m + \frac{3c_1}{2} m\sigma_{(m)} + (3c_1+1)s_m\Big) \leqslant \mathbb{E}_{\theta^\circ}\, \mathbb{1}_{\Omega_m}\, P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}\Big(S_m^{\boldsymbol{\vartheta}^m} - \mathbb{E}_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}[S_m^{\boldsymbol{\vartheta}^m}] > \frac{3c_1}{2}(m\sigma_{(m)} + 2s_m)\Big).$$

Employing (3.2) in Lemma 3.1 we bound the left hand side in the last display and we obtain

$$\mathbb{E}_{\theta^\circ}\, \mathbb{1}_{\Omega_m}\, P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}\Big(S_m^{\boldsymbol{\vartheta}^m} > m\overline{\sigma}_m + \frac{3c_1}{2} m\sigma_{(m)} + (3c_1+1)s_m\Big) \leqslant \exp\big(-\frac{c_1(c_1\wedge 1)(m\sigma_{(m)} + 2s_m)}{4\sigma_{(m)}}\big)$$

where we used that $m\sigma_{(m)} \geqslant \sum_{j=1}^m \sigma_j$ for $\sigma_{(m)} = \max_{1\leqslant j\leqslant m}\sigma_j$. As a consequence,

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|_{\ell_2}^2 > \mathfrak{b}_m + m\overline{\sigma}_m + \frac{3c_1}{2} m\sigma_{(m)} + (3c_1+1)s_m) \leqslant \exp\big(-\frac{c_1(c_1\wedge 1)(m\sigma_{(m)} + 2s_m)}{4\sigma_{(m)}}\big) + P_{\theta^\circ}(\Omega_m^c). \quad \text{(A.2)}$$

In the following, we bound the remainder probability of the event $\Omega_m^c = \big\{S_m^{\mathrm{Y}} > s_m\big\}$ for $S_m^{\mathrm{Y}} := \sum_{j=1}^m(\theta_j^{\mathrm{Y}} - \theta_j^\circ)^2$ where the random variables $\{\theta_j^{\mathrm{Y}} - \theta_j^\circ\}_{j=1}^m$ are independent and normally distributed with mean $\mathbb{E}_{\theta^\circ}[\theta_j^{\mathrm{Y}}] - \theta_j^\circ$ and standard deviation $\beta_j := \varepsilon^{1/2}\lambda_j^{-1}\mu_j$ for $\mu_j := (\varepsilon\lambda_j^{-2}\varsigma_j^{-1}+1)^{-1}$. Since $\sigma_j = \varepsilon\lambda_j^{-2}\mu_j$ and $\mu_j \leqslant 1$ if follows that $m\overline{\sigma}_m \geqslant \sum_{j=1}^m \beta_j^2$ and $\sigma_{(m)} \geqslant \max_{1\leqslant j\leqslant m}\beta_j^2$. Moreover, $\mathfrak{r}_m = \sum_{j=1}^m(\mathbb{E}_{\theta^\circ}[\theta_j^{\mathrm{Y}}] - \theta_j^\circ)^2$ and hence $\mathbb{E}_{\theta^\circ}[S_m^{\mathrm{Y}}] \leqslant m\overline{\sigma}_m + \mathfrak{r}_m$. Denote $s_m := m\overline{\sigma}_m + \frac{3c_2}{2} m\sigma_{(m)} + (3c_2+1)\mathfrak{r}_m$ which allows us to write

$$P_{\theta^\circ}(\Omega_m^c) = P_{\theta^\circ}\Big(S_m^{\mathrm{Y}} > m\overline{\sigma}_m + \frac{3c_2}{2} m\sigma_{(m)} + (3c_2+1)\mathfrak{r}_m\Big) \leqslant P_{\theta^\circ}\Big(S_m^{\mathrm{Y}} - \mathbb{E}_{\theta^\circ}[S_m^{\mathrm{Y}}] > \frac{3c_2}{2}(m\sigma_{(m)} + 2\mathfrak{r}_m)\Big)$$

The right hand side in the last display is bounded by employing (3.2) in Lemma 3.1, and hence

$$P_{\theta^\circ}(\Omega_m^c) \leqslant \exp\big(-\frac{c_2(c_2\wedge 1)(m\sigma_{(m)} + 2\mathfrak{r}_m)}{4\sigma_{(m)}}\big). \quad \text{(A.3)}$$

By combination of (A.2), (A.3) and $s_m = m\overline{\sigma}_m + \frac{3c_2}{2} m\sigma_{(m)} + (3c_2+1)\mathfrak{r}_m$ it follows that

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}\big(\|\boldsymbol{\vartheta}^m - \theta^\circ\|_{\ell_2}^2 > \mathfrak{b}_m + m\overline{\sigma}_m + \frac{3c_1}{2} m\sigma_{(m)} + (3c_1+1)[m\overline{\sigma}_m + \frac{3c_2}{2} m\sigma_{(m)} + (3c_2+1)\mathfrak{r}_m]\big) \leqslant \exp\big(-\frac{c_1(c_1\wedge 1)(3c_2+1)(m\sigma_{(m)} + 2\mathfrak{r}_m)}{4\sigma_{(m)}}\big) + \exp\big(-\frac{c_2(c_2\wedge 1)(m\sigma_{(m)} + 2\mathfrak{r}_m)}{4\sigma_{(m)}}\big)$$

The assertion (3.3) follows now by taking $c_1 = 1/3 = c_2$. The proof of the assertion (3.4) follows along the lines of the proof of (3.3). Let c_3 be a positive constant (to be specified below). Since $\mathbb{E}_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}[S_m^{\boldsymbol{\vartheta}^m}] \geqslant m\overline{\sigma}_m$ it trivially follows from (3.1) in Lemma 3.1 that

$$\mathbb{E}_{\theta^\circ}\, \mathbb{1}_{\Omega_m}\, P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}\Big(S_m^{\boldsymbol{\vartheta}^m} < m\overline{\sigma}_m - c_3 m\sigma_{(m)} - 2c_3 s_m\Big)$$
$$\leqslant \mathbb{E}_{\theta^\circ}\, \mathbb{1}_{\Omega_m}\, P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}\Big(S_m^{\boldsymbol{\vartheta}^m} - \mathbb{E}_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}[S_m^{\boldsymbol{\vartheta}^m}] < -c_3(m\sigma_{(m)} + 2s_m)\Big)$$
$$\leqslant \exp\big(-\frac{c_3(c_3\wedge 1)(m\sigma_{(m)} + 2s_m)}{4\sigma_{(m)}}\big)$$

Combining the last bound, the estimate (A.3) and $\mathfrak{b}_m = \sum_{j>m}(\theta_j^\circ - \theta_j^\times)^2$ it follows that

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}\Big(\|\boldsymbol{\vartheta}^m - \theta^\circ\|_{\ell_2}^2 < \mathfrak{b}_m + m\overline{\sigma}_m - c_3 m\sigma_{(m)} - 2c_3[m\overline{\sigma}_m + \frac{3c_2}{2}m\sigma_{(m)} + (3c_2+1)\mathfrak{r}_m]\Big)$$
$$\leqslant \mathbb{E}_{\theta^\circ}\, \mathbb{1}_{\Omega_m}\, P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}(S_m^{\boldsymbol{\vartheta}^m} < m\overline{\sigma}_m - c_3 m\sigma_{(m)} - 2c_3 s_m) + P_{\theta^\circ}(\Omega_m^c)$$
$$\leqslant \exp\big(-\frac{c_3(c_3\wedge 1)(3c_2+1)(m\sigma_{(m)} + 2\mathfrak{r}_m)}{4\sigma_{(m)}}\big) + \exp\big(-\frac{c_2(c_2\wedge 1)(m\sigma_{(m)} + 2\mathfrak{r}_m)}{4\sigma_{(m)}}\big)$$

The assertion (3.4) follows now by taking $c_2 = 1/3$ which completes the proof. □

Proof of Proposition 3.5. Keeping in mind the notations and findings used in the proof of Proposition 3.2 we have

$$\mathbb{E}_{\theta^\circ}\|\widehat{\theta}^{m_\varepsilon} - \theta^\circ\|_{\ell_2}^2 = \mathbb{E}_{\theta^\circ}\sum_{j=1}^{m_\varepsilon}(\theta_j^{\mathrm{Y}} - \theta_j^\circ)^2 + \sum_{j>m_\varepsilon}(\theta_j^\times - \theta_j^\circ)^2$$
$$= \sum_{j=1}^{m_\varepsilon}\sigma_j(\sigma_j\lambda_j^2\varepsilon^{-1}) + \mathfrak{r}_{m_\varepsilon} + \mathfrak{b}_{m_\varepsilon}, \quad \text{(A.4)}$$

which together with $\sigma_j\lambda_j^2\varepsilon^{-1} \leqslant 1$ implies $\mathbb{E}_{\theta^\circ}\|\widehat{\theta}^{m_\varepsilon} - \theta^\circ\|_{\ell_2}^2 \leqslant \mathfrak{b}_{m_\varepsilon} + m_\varepsilon\overline{\sigma}_{m_\varepsilon} + \mathfrak{r}_{m_\varepsilon}$. Exploiting the Assumption A.1, that is, $\mathfrak{r}_{m_\varepsilon} \leqslant K[\mathfrak{b}_{m_\varepsilon} \vee m_\varepsilon\overline{\sigma}_{m_\varepsilon}]$, we obtain the assertion. □

Proof of Theorem 3.8. The assertion follows from (A.4) given in the proof of Proposition 3.5. Indeed, (i) follows by combination of (A.4), $\sum_{j=1}^m \sigma_j(\sigma_j\lambda_j^2\varepsilon^{-1}) \leqslant \varepsilon m\overline{\Lambda}_m$ and $\mathfrak{r}_m \leqslant d^{-2}\|\theta^\circ - \theta^\times\|_{\ell_2}^2\varepsilon\Lambda_{(m)}$ while (A.4), $\sum_{j=1}^m \sigma_j(\sigma_j\lambda_j^2\varepsilon^{-1}) \geqslant (1+1/d)^{-2}\varepsilon m\overline{\Lambda}_m$ and $\mathfrak{r}_m \geqslant 0$ imply together (ii). Note that these elementary bounds hold due to Assumption A.2 for all $\varepsilon \in (0, \varepsilon_o)$ since $\Phi_\varepsilon^\circ = o(1)$ as $\varepsilon \to 0$, which completes the proof. □

Proof of Theorem 3.10. We start the proof with the observation that due to Assumption A.3 and (3.14) the sub-family $\{P_{\boldsymbol{\vartheta}^{m_\varepsilon^\star}}\}_{m_\varepsilon^\star}$ satisfies the condition (3.7) uniformly for all $\theta^\circ \in \Theta_{\mathfrak{a}}^r$ with $L = L^\star/\kappa^\star$. Moreover, we have $\Phi_\varepsilon^\star = o(1)$, as $\varepsilon \to 0$ and we suppose that Assumption A.2 holds true. Thereby, the assumptions of Corollary 3.6 are satisfied. From $((1+1/d)\vee r/d^2)(L^\star/\kappa^\star) \geqslant ((1+1/d)\vee d^{-2}\|\theta^\circ - \theta^\times\|_{\ell_2}^2)L = K$ and the definition of $K^\star$ it follows further that $K^\star \geqslant (4+(11/2)K)(1\vee r)$ and $(K^\star)^{-1} \leqslant (1/9)(1+1/d)^{-1}\kappa^\star$ for all $\theta^\circ \in \Theta_{\mathfrak{a}}^r$. Moreover, for all $0 < \varepsilon < \varepsilon_o$ we have $(1\vee r)\,\Phi_\varepsilon^\star \geqslant \Phi_\varepsilon^{m_\varepsilon^\star} = [\mathfrak{b}_{m_\varepsilon^\star} \vee \varepsilon\, m_\varepsilon^\star\, \overline{\Lambda}_{m_\varepsilon^\star}] \geqslant \kappa^\star\,\Phi_\varepsilon^\star$. By combining these elementary inequalities and Corollary 3.6 with $c := 1/(9K)$ and $c \geqslant 1/K^\star$ uniformly for all $\theta^\circ \in \Theta_{\mathfrak{a}}^r$ we obtain for all $\varepsilon \in (0, \varepsilon_o)$

$$
\begin{aligned}
&\sup_{\theta^\circ\in\Theta_{\mathfrak{a}}^r} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon^\star}\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^{m_\varepsilon^\star} - \theta^\circ\|_{\ell_2}^2 > K^\star\Phi_\varepsilon^\star)\\
&\leqslant \sup_{\theta^\circ\in\Theta_{\mathfrak{a}}^r} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon^\star}\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^{m_\varepsilon^\star} - \theta^\circ\|_{\ell_2}^2 > (4+(11/2)K)\Phi_\varepsilon^{m_\varepsilon^\star})\\
&\hspace{20em}\leqslant 2\exp(-m_\varepsilon^\star/36); \quad \text{(A.5)}
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{\theta^\circ\in\Theta_{\mathfrak{a}}^r} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon^\star}\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^{m_\varepsilon^\star} - \theta^\circ\|_{\ell_2}^2 < (K^\star)^{-1}\Phi_\varepsilon^\star)\\
&\leqslant \sup_{\theta^\circ\in\Theta_{\mathfrak{a}}^r} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{m_\varepsilon^\star}\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^{m_\varepsilon^\star} - \theta^\circ\|_{\ell_2}^2 < (1-8\,c\,K)\{(1+1/d)\}^{-1}\Phi_\varepsilon^{m_\varepsilon^\star})\\
&\hspace{20em}\leqslant 2\exp(-m_\varepsilon^\star/[2(K^\star)^2]). \quad \text{(A.6)}
\end{aligned}
$$

By combining (A.5) and (A.6) we obtain the assertion of the theorem since $m_\varepsilon^\star \to \infty$, which completes the proof. □

B Appendix: Proofs of Section 4

B.1 Proof of Theorem 4.3

PROOF OF LEMMA 4.1. Consider ***(i)***. The claim holds trivially true in case $G_\varepsilon^- = 1$, thus suppose $G_\varepsilon^- > 1$ and let $1 \leqslant m < G_\varepsilon^- \leqslant m_\varepsilon^\circ$. Define $S_m := \|\widehat{\theta}^{m_\varepsilon^\circ} - \theta^\times\|_\sigma^2 - \|\widehat{\theta}^m - \theta^\times\|_\sigma^2$. Given an event $\mathcal{A}_m$ and its complement $\mathcal{A}_m^c$ (to be specified below) it follows

$$
\begin{aligned}
p_{\mathrm{M}\,|\,\mathrm{Y}}(m) &= \frac{\exp(\frac{1}{2}\{\|\widehat{\theta}^m - \theta^\times\|_\sigma^2 - 3C_\lambda m\})}{\sum_{k=1}^{G_\varepsilon}\exp(\frac{1}{2}\{\|\widehat{\theta}^k - \theta^\times\|_\sigma^2 - 3C_\lambda k\})}\\
&= \exp\left(\frac{1}{2}\big\{-S_m + 3C_\lambda[m_\varepsilon^\circ - m]\big\}\right)\mathbb{1}_{\mathcal{A}_m} + \mathbb{1}_{\mathcal{A}_m^c} \quad \text{(B.1)}
\end{aligned}
$$

Moreover, elementary algebra shows

$$
S_m = \sum_{j=m+1}^{m_\varepsilon^\circ} \frac{\lambda_j^2\sigma_j}{\varepsilon^2}(Y_j - \lambda_j\theta_j^\times)^2
$$

where the random variables $\{\lambda_j\sigma_j^{1/2}\varepsilon^{-1}(Y_j - \lambda_j\theta_j^{\times})\}_{j\geqslant 1}$ are independent and normally distributed with standard deviation $\beta_j = \lambda_j\sigma_j^{1/2}\varepsilon^{-1/2}$ and mean $\alpha_j = \beta_j\varepsilon^{-1/2}\lambda_j(\theta_j^{\circ} - \theta_j^{\times})$. Keeping in mind the notations used in Lemma 3.1 define $v_m := \sum_{j=m+1}^{m_\varepsilon^\circ}\beta_j^2$ and $r_m := \sum_{j=m+1}^{m_\varepsilon^\circ}\alpha_j^2$. We observe that Assumption A.2 implies that $1 \geqslant \beta_j^2 \geqslant (1+1/d)^{-1}$ and hence it follows by employing $\min_{m<j\leqslant m_\varepsilon^\circ}\lambda_j^2 \geqslant \min_{1\leqslant j\leqslant m_\varepsilon^\circ}\lambda_j^2 = \Lambda_{(m_\varepsilon^\circ)}^{-1}$ and Assumption A.4 (iii) that

$$\begin{aligned}&L_\lambda(\varepsilon\Lambda_{(m_\varepsilon^\circ)})^{-1}\Phi_\varepsilon^\circ \geqslant L_\lambda(\varepsilon\Lambda_{(m_\varepsilon^\circ)})^{-1}\varepsilon m_\varepsilon^\circ\overline{\Lambda}_{m_\varepsilon^\circ} \geqslant m_\varepsilon^\circ \quad \text{and}\\ &(1+1/d)^{-1}(\varepsilon\Lambda_{(m_\varepsilon^\circ)})^{-1}[\mathfrak{b}_m - \Phi_\varepsilon^\circ] \leqslant (1+1/d)^{-1}(\varepsilon\Lambda_{(m_\varepsilon^\circ)})^{-1}[\mathfrak{b}_m - \mathfrak{b}_{m_\varepsilon^\circ}] \leqslant r_m. \quad \text{(B.2)}\end{aligned}$$

Moreover, we set $t_m := 1 \geqslant \max_{m<j\leqslant m_\varepsilon^\circ}\beta_j^2$ and $\mu_m := \mathbb{E}S_m = v_m + r_m$. Introduce the event $\mathcal{A}_m := \{S_m - \mu_m \geqslant -(1/4)(v_m + 2r_m)\}$ and its complement $\mathcal{A}_m^c := \{S_m - \mu_m < -(1/4)(v_m + 2r_m)\}$. By employing successively Lemma 3.1, (B.2) and $\mathfrak{b}_{m_\varepsilon^\circ} \leqslant \Phi_\varepsilon^\circ$ it follows now from (B.1) that

$$\begin{aligned}\mathbb{E}_{\theta^\circ}p_{\mathrm{M}\,|\,\mathrm{Y}}(m) &\leqslant \mathbb{E}_{\theta^\circ}\exp\big(\{-(S_m - \mu_m) - \mu_m + 3C_\lambda[m_\varepsilon^\circ - m]\}/2\big)\,\mathbb{1}_{\mathcal{A}_m} + \mathbb{E}_{\theta^\circ}\,\mathbb{1}_{\mathcal{A}_m^c}\\ &\leqslant \exp\big(\{-3v_m/4 - r_m/2 + 3C_\lambda[m_\varepsilon^\circ - m]\}/2\big) + \exp\big(-(1/64)(v_m + 2r_m)\big)\\ &\leqslant \exp\big(-r_m/4 + 3C_\lambda m_\varepsilon^\circ/2\big) + \exp\big(-r_m/32\big)\\ &\leqslant \exp\Big(-\frac{[\mathfrak{b}_m - \Phi_\varepsilon^\circ]}{4(1+1/d)\varepsilon\Lambda_{(m_\varepsilon^\circ)}} + \frac{3C_\lambda L_\lambda\Phi_\varepsilon^\circ}{2\varepsilon\Lambda_{(m_\varepsilon^\circ)}}\}\Big) + \exp\Big(-\frac{[\mathfrak{b}_m - \Phi_\varepsilon^\circ]}{32(1+1/d)\varepsilon\Lambda_{(m_\varepsilon^\circ)}})\Big)\\ &\leqslant \exp\Big(-\frac{\mathfrak{b}_m}{4(1+1/d)\varepsilon\Lambda_{(m_\varepsilon^\circ)}} + \frac{2C_\lambda L_\lambda\Phi_\varepsilon^\circ}{\varepsilon\Lambda_{(m_\varepsilon^\circ)}}\Big) \times \exp\Big(-\frac{L_\lambda C_\lambda\Phi_\varepsilon^\circ}{4\varepsilon\Lambda_{(m_\varepsilon^\circ)}}\Big)\\ &\qquad\qquad + \exp\Big(-\frac{[\mathfrak{b}_m - \Phi_\varepsilon^\circ]}{32(1+1/d)\varepsilon\Lambda_{(m_\varepsilon^\circ)}}\Big)\end{aligned}$$

Taking into account the definition (4.4) of G_ε^-, i.e., $\mathfrak{b}_m > 8L_\lambda C_\lambda(1+1/d)\Phi_\varepsilon^\circ$ for all $1 \leqslant m < G_\varepsilon^-$, and $L_\lambda\Phi_\varepsilon^\circ(\varepsilon\Lambda_{(m_\varepsilon^\circ)})^{-1} \geqslant m_\varepsilon^\circ$ due to Assumption A.4 (iii), we obtain

$$\mathbb{E}_{\theta^\circ}p_{\mathrm{M}\,|\,\mathrm{Y}}(m) \leqslant \exp\Big(-\frac{L_\lambda C_\lambda\Phi_\varepsilon^\circ}{4\varepsilon\Lambda_{(m_\varepsilon^\circ)}}\Big) + \exp\Big(-\frac{7L_\lambda C_\lambda\Phi_\varepsilon^\circ}{32\varepsilon\Lambda_{(m_\varepsilon^\circ)}}\Big) \leqslant 2\exp\Big(-\frac{7C_\lambda}{32}m_\varepsilon^\circ\Big).$$

Thereby, $\mathbb{E}_{\theta^\circ}P_{\mathrm{M}\,|\,\mathrm{Y}}(1 \leqslant \mathrm{M} < G_\varepsilon^-) = \sum_{m=1}^{G_\varepsilon^- - 1}\mathbb{E}_{\theta^\circ}p_{\mathrm{M}\,|\,\mathrm{Y}}(m) \leqslant 2\exp\big(-\frac{7C_\lambda}{32}m_\varepsilon^\circ + \log G_\varepsilon\big)$ using that $G_\varepsilon \geqslant G_\varepsilon^-$ which proves the assertion ***(i)***. Consider now ***(ii)***. The claim holds trivially true in case $G_\varepsilon^+ = G_\varepsilon$, thus suppose $G_\varepsilon^+ < G_\varepsilon$ and let $G_\varepsilon \geqslant m > G_\varepsilon^+ \geqslant m_\varepsilon^\circ$. Consider again the upper bound given in (B.1) where now

$$-S_m = \sum_{j=m_\varepsilon^\circ+1}^{m}\frac{\lambda_j^2\sigma_j}{\varepsilon^2}(Y_j - \lambda_j\theta_j^{\times})^2.$$

Employing the notations α_j and β_j introduced in the proof of ***(i)*** and keeping in mind Lemma 3.1 we define $v_m := \sum_{j=m_\varepsilon^\circ+1}^{m}\beta_j^2$ and $r_m := \sum_{j=m_\varepsilon^\circ+1}^{m}\alpha_j^2$ where $1 \geqslant$

$\beta_j^2 \geqslant (1+1/d)^{-1}$ due to Assumption A.2. Moreover, from Assumption A.4 (i) follows $\max_{m_\varepsilon^\circ<j\leqslant m}\lambda_j^2 \geqslant \max_{m_\varepsilon^\circ<j}\lambda_j^2 \leqslant C_\lambda \min_{1\leqslant j\leqslant m_\varepsilon^\circ}\lambda_j^2 = C_\lambda\Lambda_{(m_\varepsilon^\circ)}^{-1}$ and taking into account in addition Assumption A.4 (iii) that

$$L_\lambda(\varepsilon\Lambda_{(m_\varepsilon^\circ)})^{-1}\Phi_\varepsilon^\circ \geqslant m_\varepsilon^\circ, \quad v_m \leqslant m - m_\varepsilon^\circ \quad \text{and} \quad C_\lambda(\varepsilon\Lambda_{(m_\varepsilon^\circ)})^{-1}\Phi_\varepsilon^\circ \geqslant C_\lambda(\varepsilon\Lambda_{(m_\varepsilon^\circ)})^{-1}[\mathfrak{b}_{m_\varepsilon^\circ} - \mathfrak{b}_m] \geqslant r_m. \tag{B.3}$$

Moreover, we set $t_m := 1 \geqslant \max_{m_\varepsilon^\circ<j\leqslant m}\beta_j^2$ and $\mu_m := C_\lambda[m-m_\varepsilon^\circ] + C_\lambda(\Lambda_{(m_\varepsilon^\circ)}\varepsilon)^{-1}[\mathfrak{b}_{m_\varepsilon^\circ} - \mathfrak{b}_m] \geqslant \mathbb{E}S_m = v_m + r_m$. Consider now the event $\mathcal{A}_m := \{-S_m - \mu_m \leqslant (C_\lambda[m-m_\varepsilon^\circ] + 2C_\lambda(\Lambda_{(m_\varepsilon^\circ)}\varepsilon)^{-1}[\mathfrak{b}_{m_\varepsilon^\circ} - \mathfrak{b}_m])\}$ and its complement $\mathcal{A}_m^c := \{-S_m - \mu_m > (C_\lambda[m-m_\varepsilon^\circ] + 2C_\lambda(\Lambda_{(m_\varepsilon^\circ)}\varepsilon)^{-1}[\mathfrak{b}_{m_\varepsilon^\circ} - \mathfrak{b}_m])\}$. By employing successively Lemma 3.1, (B.3) and $\mathfrak{b}_{m_\varepsilon^\circ} \leqslant \Phi_\varepsilon^\circ$ it follows now from (B.1) that

$$\begin{aligned}
\mathbb{E}_{\theta^\circ} p_{\mathrm{M}\,|\,\mathrm{Y}}(m) &\leqslant \mathbb{E}_{\theta^\circ} \exp\big(\{(-S_m-\mu_m)+\mu_m+3C_\lambda[m_\varepsilon^\circ-m]\}/2\big)\,\mathbb{1}_{\mathcal{A}_m} + \mathbb{E}_{\theta^\circ}\mathbb{1}_{\mathcal{A}_m^c}\\
&\leqslant \exp\big(\{2C_\lambda[m-m_\varepsilon^\circ] + 3C_\lambda(\Lambda_{(m_\varepsilon^\circ)}\varepsilon)^{-1}[\mathfrak{b}_{m_\varepsilon^\circ}-\mathfrak{b}_m] + 3C_\lambda[m_\varepsilon^\circ - m]\}/2\big)\\
&\quad + \exp\big(-\{C_\lambda[m-m_\varepsilon^\circ] + 2C_\lambda(\Lambda_{(m_\varepsilon^\circ)}\varepsilon)^{-1}[\mathfrak{b}_{m_\varepsilon^\circ}-\mathfrak{b}_m]\}/9\big)\\
&\leqslant \exp\big(\{C_\lambda[m_\varepsilon^\circ-m] + 3C_\lambda(\Lambda_{(m_\varepsilon^\circ)}\varepsilon)^{-1}\Phi_\varepsilon^\circ\}/2\big) + \exp\big(-C_\lambda[m-m_\varepsilon^\circ]/9\big)\\
&\leqslant \exp\big(C_\lambda\{-m + 3(\Lambda_{(m_\varepsilon^\circ)}\varepsilon)^{-1}\Phi_\varepsilon^\circ + L_\lambda(\varepsilon\Lambda_{(m_\varepsilon^\circ)})^{-1}\Phi_\varepsilon^\circ\}/2\big)\\
&\quad + \exp\big(-C_\lambda(m - L_\lambda(\varepsilon\Lambda_{(m_\varepsilon^\circ)})^{-1}\Phi_\varepsilon^\circ)/9\big)\\
&\leqslant \exp\big(C_\lambda\{-m + 5L_\lambda(\Lambda_{(m_\varepsilon^\circ)}\varepsilon)^{-1}\Phi_\varepsilon^\circ\}/2\big) \times \exp\Big(-\frac{C_\lambda L_\lambda\Phi_\varepsilon^\circ}{2\Lambda_{(m_\varepsilon^\circ)}\varepsilon}\Big)\\
&\qquad + \exp\big(-C_\lambda(m - L_\lambda(\varepsilon\Lambda_{(m_\varepsilon^\circ)})^{-1}\Phi_\varepsilon^\circ)/9\big)
\end{aligned}$$

Taking into account the definition (4.4) of G_ε^+, i.e., $m > 5L_\lambda(\varepsilon\Lambda_{(m_\varepsilon^\circ)})^{-1}\Phi_\varepsilon^\circ$ for all $G_\varepsilon \geqslant m > G_\varepsilon^+$, and $L_\lambda\Phi_\varepsilon^\circ(\varepsilon\Lambda_{(m_\varepsilon^\circ)})^{-1} \geqslant m_\varepsilon^\circ$ due to Assumption A.4 (iii), we obtain

$$\mathbb{E}_{\theta^\circ} p_{\mathrm{M}\,|\,\mathrm{Y}}(m) \leqslant \exp\Big(-\frac{L_\lambda C_\lambda\Phi_\varepsilon^\circ}{2\varepsilon\Lambda_{(m_\varepsilon^\circ)}}\Big) + \exp\Big(-\frac{4L_\lambda C_\lambda\Phi_\varepsilon^\circ}{9\varepsilon\Lambda_{(m_\varepsilon^\circ)}}\Big) \leqslant 2\exp\Big(-\frac{4C_\lambda}{9}m_\varepsilon^\circ\Big).$$

Thereby, $\mathbb{E}_{\theta^\circ} P_{\mathrm{M}\,|\,\mathrm{Y}}(G_\varepsilon^+ < \mathrm{M} \leqslant G_\varepsilon) = \sum_{m=G_\varepsilon^+ +1}^{G_\varepsilon}\mathbb{E}_{\theta^\circ} p_{\mathrm{M}\,|\,\mathrm{Y}}(m) \leqslant 2\exp\big(-\frac{4C_\lambda}{9}m_\varepsilon^\circ + \log G_\varepsilon\big)$ which shows the assertion ***(ii)*** and completes the proof. □

Proof of Lemma 4.2. Consider ***(i)***. We start the proof with the observation that due to Assumption A.4 (iii) the condition (3.7) holds true with $L = L_\lambda$ uniformly for all $m \in \mathbb{N}$ and $\varepsilon \in (0,1)$, and hence imposing Assumption A.2 the conditions of Corollary 3.6 are satisfied, which in turn setting $c := 1/(9K)$ with $K := ((1+d^{-1}) \vee d^{-2}\|\theta^\circ - \theta^\times\|_{\ell_2}^2)L_\lambda$ implies for all $1 \leqslant m \leqslant G_\varepsilon$ and $\varepsilon \in (0, \varepsilon_\circ)$ that

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|_{\ell_2}^2 > (4+(11/2)K)[\mathfrak{b}_m \vee \varepsilon m\overline{\Lambda}_m]) \leqslant 2\exp(-m/36); \tag{B.4}$$

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|_{\ell_2}^2 < \{9(1+1/d)\}^{-1}[\mathfrak{b}_m \vee \varepsilon m\overline{\Lambda}_m]) \leqslant 2\exp(-m/(162K^2)). \tag{B.5}$$

On the other hand side, taking into account the definition (4.4) of G_ε^+ and G_ε^-, and the monotonicity of $(\mathfrak{b}_m)_{m\geqslant 1}$ and $(\varepsilon m\overline{\Lambda}_m)_{m\geqslant 1}$ we have for all $G_\varepsilon^- \leqslant m \leqslant m_\varepsilon^\circ$ that

$$\varepsilon m\overline{\Lambda}_m \leqslant \varepsilon m_\varepsilon^\circ \overline{\Lambda}_{m_\varepsilon^\circ} \leqslant \Phi_\varepsilon^\circ \quad \text{and} \quad \mathfrak{b}_m \leqslant 8L_\lambda C_\lambda(1+1/d)\Phi_\varepsilon^\circ$$

while for all $G_\varepsilon^+ \geqslant m \geqslant m_\varepsilon^\circ$ (keeping in mind Assumption A.5) hold

$$m \leqslant 5L_\lambda(\varepsilon\Lambda_{(m_\varepsilon^\circ)})^{-1}\Phi_\varepsilon^\circ \leqslant 5L_\lambda(\varepsilon\Lambda_{(m_\varepsilon^\circ)})^{-1}(\kappa^o)^{-1}\varepsilon m_\varepsilon^\circ\overline{\Lambda}_{m_\varepsilon^\circ} \leqslant (5L_\lambda/\kappa^\circ)m_\varepsilon^\circ \leqslant D^\circ m_\varepsilon^\circ \quad \text{and}$$
$$\mathfrak{b}_m \leqslant \mathfrak{b}_{m_\varepsilon^\circ} \leqslant \Phi_\varepsilon^\circ$$

where $D^\circ := D^\circ(\theta^\times, \theta^\circ, \lambda) := \lceil 5L_\lambda/\kappa^\circ\rceil$. Due to Assumption A.4 (ii) and (iii) it follows from $m \leqslant D^\circ m_\varepsilon^\circ$ that $\Lambda_{(m)} \leqslant \Lambda_{(D^\circ m_\varepsilon^\circ)} \leqslant \Lambda_{(D^\circ)}\Lambda_{(m_\varepsilon^\circ)}$ and $\overline{\Lambda}_m \leqslant \Lambda_{(m)} \leqslant \Lambda_{(D^\circ)}\Lambda_{(m_\varepsilon^\circ)} \leqslant \Lambda_{(D^\circ)}L_\lambda\overline{\Lambda}_{m_\varepsilon^\circ}$ which in turn implies $\varepsilon m\overline{\Lambda}_m \leqslant L_\lambda D^o\Lambda_{(D^\circ)}\varepsilon m_\varepsilon^\circ\overline{\Lambda}_{m_\varepsilon^\circ} \leqslant L_\lambda D^\circ\Lambda_{(D^\circ)}\Phi_\varepsilon^\circ$ for all $m \leqslant G_\varepsilon^+$. Combining the upper bounds we have $(4+11K/2)[\mathfrak{b}_m \vee \varepsilon m\overline{\Lambda}_m] \leqslant K^\circ\Phi_\varepsilon^\circ$ for all $G_\varepsilon^- \leqslant m \leqslant G_\varepsilon^+$ since $K^\circ \geqslant (4+11K/2)(8L_\lambda C_\lambda(1+1/d) \vee L_\lambda D^\circ\Lambda_{(D^\circ)})$, and together with (B.4) follows

$$\begin{aligned}\sum_{m=G_\varepsilon^-}^{G_\varepsilon^+} &\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m\,|\,Y}\big(\|\boldsymbol{\vartheta}^m - \theta^\circ\|_{\ell_2}^2 > K^o\Phi_\varepsilon^\circ\big)\\ &\leqslant \sum_{m=G_\varepsilon^-}^{G_\varepsilon^+} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m\,|\,Y}\big(\|\boldsymbol{\vartheta}^m - \theta^\circ\|_{\ell_2}^2 > (4+(11/2)K)[\mathfrak{b}_m \vee \varepsilon m\overline{\Lambda}_m]\big)\\ &\qquad\qquad \leqslant 2\sum_{m=G_\varepsilon^-}^{G_\varepsilon^+} \exp(-m/36) \leqslant 74\exp(-G_\varepsilon^-/36)\end{aligned}$$

which proves the assertion ***(i)***. Consider now ***(ii)***. We observe that by definition (3.10) of Φ_ε° for all $m \in \mathbb{N}$ holds $\Phi_\varepsilon^\circ \leqslant [\varepsilon m\overline{\Lambda}_m \vee \mathfrak{b}_m]$, and hence $\{9(1+1/d)\}^{-1}[\mathfrak{b}_m \vee \varepsilon m\overline{\Lambda}_m] \geqslant (K^\circ)^{-1}\Phi_\varepsilon^\circ$ since $K^\circ \geqslant 9(1+1/d)$. Combining the last estimate, (B.5) and $K^\circ \geqslant 10K$ it follows that

$$\begin{aligned}\sum_{m=G_\varepsilon^-}^{G_\varepsilon^+} &\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m\,|\,Y}\big(\|\boldsymbol{\vartheta}^m - \theta^\circ\|_{\ell_2}^2 < (K^\circ)^{-1}\Phi_\varepsilon^\circ\big)\\ &\leqslant \sum_{m=G_\varepsilon^-}^{G_\varepsilon^+} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m\,|\,Y}\big(\|\boldsymbol{\vartheta}^m - \theta^\circ\|_{\ell_2}^2 < \{9(1+1/d)\}^{-1}[\mathfrak{b}_m \vee \varepsilon m\overline{\Lambda}_m]\big)\\ &\qquad\qquad \leqslant 2\sum_{m=G_\varepsilon^-}^{G_\varepsilon^+} \exp(-m/(K^\circ)^2) \leqslant 4(K^\circ)^2\exp(-G_\varepsilon^-/(K^\circ)^2)\end{aligned}$$

which shows the assertion ***(ii)*** and completes the proof. □

Proof of Theorem 4.3. We start the proof with the observation that Lemma 4.1 together with Lemma 4.2 *(i)* imply

$$\begin{aligned}\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{\mathrm{M}}\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^{\mathrm{M}} - \theta^\circ\|_{\ell_2}^2 > K^o \Phi_\varepsilon^\circ) &= \mathbb{E}_{\theta^\circ} \sum_{m=1}^{G_\varepsilon} p_{\mathrm{M}\,|\,\mathrm{Y}}(m) P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|_{\ell_2}^2 > K^o \Phi_\varepsilon^\circ) \\ \leqslant \mathbb{E}_{\theta^\circ} P_{\mathrm{M}\,|\,\mathrm{Y}}(1 \leqslant \mathrm{M} < G_\varepsilon^-) + \mathbb{E}_{\theta^\circ} P_{\mathrm{M}\,|\,\mathrm{Y}}(G_\varepsilon^+ < \mathrm{M} \leqslant G_\varepsilon) & \\ + \sum_{m=G_\varepsilon^-}^{G_\varepsilon^+} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|_{\ell_2}^2 > K^o \Phi_\varepsilon^\circ) & \\ \leqslant 4 \exp\big(-m_\varepsilon^\circ\{C_\lambda/5 - \log G_\varepsilon/m_\varepsilon^\circ\}\big) + 74 \exp(-G_\varepsilon^-/36) & \qquad \text{(B.6)}\end{aligned}$$

On the other hand side, from Lemma 4.1 together with Lemma 4.2 *(ii)* also follows that

$$\begin{aligned}\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{\mathrm{M}}\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^{\mathrm{M}} - \theta^\circ\|_{\ell_2}^2 < (K^o)^{-1} \Phi_\varepsilon^\circ) &\leqslant \mathbb{E}_{\theta^\circ} P_{\mathrm{M}\,|\,\mathrm{Y}}(1 \leqslant \mathrm{M} < G_\varepsilon^-) \\ + \mathbb{E}_{\theta^\circ} P_{\mathrm{M}\,|\,\mathrm{Y}}(G_\varepsilon^+ < \mathrm{M} \leqslant G_\varepsilon) + \mathbb{E}_{\theta^\circ} \sum_{m=G_\varepsilon^-}^{G_\varepsilon^+} & p_{\mathrm{M}\,|\,\mathrm{Y}}(m) P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|_{\ell_2}^2 < (K^o)^{-1} \Phi_\varepsilon^\circ) \\ \leqslant 4 \exp\big(-m_\varepsilon^\circ\{C_\lambda/5 - \log G_\varepsilon/m_\varepsilon^\circ\}\big) + 4(K^\circ)^2 & \exp(-G_\varepsilon^-/(K^\circ)^2) \qquad \text{(B.7)}\end{aligned}$$

By combining (B.6) and (B.7) we obtain the assertion of the theorem since $G_\varepsilon^-, m_\varepsilon^\circ \to \infty$ and $\log G_\varepsilon/m_\varepsilon^\circ = o(1)$ as $\varepsilon \to 0$ which completes the proof. □

B.2 Proof of Theorem 4.4

The next assertion presents a concentration inequality for Gaussian random variables.

Lemma B.1. *Let the assumptions of Lemma 3.1 be satisfied. For all $c \geqslant 0$ we have*

$$\sup_{m\geqslant 1} (6t_m)^{-1} \exp\left(\frac{c(v_m + 2r_m)}{4t_m}\right) \mathbb{E}\left(S_m - \mathbb{E}S_m - \frac{3}{2}c(v_m + 2r_m)\right)_+ \leqslant 1 \qquad \text{(B.8)}$$

where $(a)_+ := (a \vee 0)$.

Proof of Lemma B.1. The assertion follows from Lemma 3.1 (keeping in mind that

$c \geqslant 1$), indeed

$$\mathbb{E}\Big(S_m - \mathbb{E}S_m - \frac{3}{2}c(v_m + 2r_m)\Big)_+ = \int_0^\infty P(S_m - \mathbb{E}S_m \geqslant x + \frac{3}{2}c(v_m + 2r_m))dx$$
$$= \int_0^\infty P(S_m - \mathbb{E}S_m \geqslant \frac{3}{2}(2x/(3(v_m + 2r_m)) + c)(v_m + 2r_m))dx$$
$$\leqslant \int_0^\infty \exp\Big(- \frac{(2x/(3(v_m + 2r_m)) + c)(v_m + 2r_m)}{4t_m}\Big)dx$$
$$= \int_0^\infty \exp\Big(- \frac{2x/3 + c(v_m + 2r_m)}{4t_m}\Big)dx$$
$$= \exp\Big(- \frac{c(v_m + 2r_m)}{4t_m}\Big)\int_0^\infty \exp\Big(- \frac{x}{6t_m}\Big)dx = \exp\Big(- \frac{c(v_m + 2r_m)}{4t_m}\Big)(6t_m)$$

□

Lemma B.2. *If Assumption A.2 and A.4 hold true then for all $\varepsilon \in (0, \varepsilon_\circ)$*

(i) $\sum_{j=1}^{G_\varepsilon} \sigma_j^2\lambda_j^2\varepsilon^{-2}\mathbb{E}_{\theta^\circ}\{(\mathrm{Y}_j - \lambda_j\theta_j^\circ)P_{\mathrm{M}\,|\,\mathrm{Y}}(j \leqslant \mathrm{M} \leqslant G_\varepsilon)\}^2$
$\leqslant \varepsilon G_\varepsilon^+\overline{\Lambda}_{G_\varepsilon^+} + 10\Lambda_1 \exp\big(- m_\varepsilon^\circ/5 + 2\log G_\varepsilon\big)$;

(ii) $\sum_{j=1}^{G_\varepsilon}(\theta_j^\times - \theta_j^\circ)^2\mathbb{E}_{\theta^\circ}\mathbb{E}_{\mathrm{M}\,|\,\mathrm{Y}}\{\mathbb{1}_{\{1 \leqslant \mathrm{M} < j\}} + (\sigma_j/\varsigma_j)^2\,\mathbb{1}_{\{j \leqslant \mathrm{M} \leqslant G_\varepsilon\}}\} + \sum_{j > G_\varepsilon}(\theta_j^\times - \theta_j^\circ)^2$
$\leqslant \mathfrak{b}_{G_\varepsilon^-} + \|\theta^\times - \theta^\circ\|_{\ell_2}^2\{d^{-2}\varepsilon\Lambda_{(G_\varepsilon^-)} + 2\exp\big(- m_\varepsilon^\circ/5 + \log G_\varepsilon\big)\}$.

Proof of Lemma B.2. Consider *(i)*. We start with the observation that the random variables $\{\xi_j := \varepsilon^{-1/2}(\mathrm{Y}_j - \lambda_j\theta_j^\circ)\}_{j \geqslant 1}$ are independent and standard normally distributed. Moreover, applying Jensen's inequality we have

$$\{\xi_j P_{\mathrm{M}\,|\,\mathrm{Y}}(j \leqslant \mathrm{M} \leqslant G_\varepsilon)\}^2 = \{\mathbb{E}_{\mathrm{M}\,|\,\mathrm{Y}}\xi_j\,\mathbb{1}_{\{j \leqslant \mathrm{M} \leqslant G_\varepsilon\}}\}^2 \leqslant \mathbb{E}_{\mathrm{M}\,|\,\mathrm{Y}}\xi_j^2\,\mathbb{1}_{\{j \leqslant \mathrm{M} \leqslant G_\varepsilon\}}$$
$$= \xi_j^2 P_{\mathrm{M}\,|\,\mathrm{Y}}(j \leqslant \mathrm{M} \leqslant G_\varepsilon).$$

We split the sum into two parts which we bound separately. Precisely,

$$\sum_{j=1}^{G_\varepsilon}\sigma_j^2\lambda_j^2\varepsilon^{-2}\{(\mathrm{Y}_j - \lambda_j\theta_j^\circ)P_{\mathrm{M}\,|\,\mathrm{Y}}(j \leqslant \mathrm{M} \leqslant G_\varepsilon)\}^2$$
$$\leqslant \sum_{j=1}^{G_\varepsilon^+}\varepsilon\Lambda_j\xi_j^2 + \sum_{j=1}^{G_\varepsilon}\varepsilon\Lambda_j\xi_j^2 P_{\mathrm{M}\,|\,\mathrm{Y}}(G_\varepsilon^+ < \mathrm{M} \leqslant G_\varepsilon) \quad \text{(B.9)}$$

where we used that $\sigma_j \leqslant \varepsilon\Lambda_j$. Keeping in mind the notations used in Lemma B.1 let $S_{G_\varepsilon} := \sum_{j=1}^{G_\varepsilon}\varepsilon\Lambda_j\xi_j^2$ and observe that $\alpha_j = 0$ and $\beta_j^2 = \varepsilon\Lambda_j$, and hence $r_{G_\varepsilon} = 0$. Keeping in mind that $G_\varepsilon := \max\{1 \leqslant m \leqslant \lfloor\varepsilon^{-1}\rfloor : \varepsilon\Lambda_{(m)} \leqslant \Lambda_1\}$ we set $t_{G_\varepsilon} := \Lambda_1 \geqslant \varepsilon\Lambda_{(G_\varepsilon)} = \max_{1 \leqslant j \leqslant G_\varepsilon}\beta_j^2$ and $v_{G_\varepsilon} := \Lambda_1 G_\varepsilon = G_\varepsilon t_{G_\varepsilon} \geqslant \sum_{j=1}^{G_\varepsilon}\beta_j^2$, where $\mathbb{E}_{\theta^\circ}S_{G_\varepsilon} \leqslant v_{G_\varepsilon}$. From

Lemma B.1 with $c = 2/3$ follows that $\mathbb{E}_{\theta^\circ}(S_{G_\varepsilon} - 2\Lambda_1 G_\varepsilon)_+ \leqslant (6t_{G_\varepsilon})\exp(-v_{G_\varepsilon}/(6t_{G_\varepsilon})) = (6\Lambda_1)\exp(-G_\varepsilon/6)$, and hence

$$\begin{aligned}&\sum_{j=1}^{G_\varepsilon}\varepsilon\Lambda_j\mathbb{E}_{\theta^\circ}\{\xi_j^2 P_{\mathrm{M}\,|\,\mathrm{Y}}(G_\varepsilon^+ < \mathrm{M} \leqslant G_\varepsilon)\}\\ &\leqslant \mathbb{E}\big(S_{G_\varepsilon} - 2\Lambda_1 G_\varepsilon\big)_+ + 2\Lambda_1 G_\varepsilon \mathbb{E}_{\theta^\circ} P_{\mathrm{M}\,|\,\mathrm{Y}}(G_\varepsilon^+ < \mathrm{M} \leqslant G_\varepsilon)\\ &\qquad\leqslant 6\Lambda_1\exp(-G_\varepsilon/6) + 2\Lambda_1 G_\varepsilon \mathbb{E}_{\theta^\circ} P_{\mathrm{M}\,|\,\mathrm{Y}}(G_\varepsilon^+ < \mathrm{M} \leqslant G_\varepsilon). \quad \text{(B.10)}\end{aligned}$$

We distinguish two cases. First, if $G_\varepsilon^+ = G_\varepsilon$, then assertion ***(i)*** follows by combining (B.9) and $\mathbb{E}_{\theta^\circ}\xi_j^2 = 1$. Second, if $G_\varepsilon^+ < G_\varepsilon$, then the definition (4.4) of G_ε^+ implies $G_\varepsilon > 5m_\varepsilon^\circ$ which in turn implies the assertion ***(i)*** by combining (B.9), $\mathbb{E}_{\theta^\circ}\xi_j^2 = 1$, (B.10) and Lemma 4.1 ***(ii)***. Consider ***(ii)***. Due to Assumption A.2 we have $(\sigma_j/\varsigma_j)^2 \leqslant (1 \wedge d^{-2}\varepsilon\Lambda_j)$ which we will use without further reference. Splitting the first sum into two parts we obtain

$$\begin{aligned}&\sum_{j=1}^{G_\varepsilon}(\theta_j^\times - \theta_j^\circ)^2\mathbb{E}_{\theta^\circ}\{\mathbb{1}_{\{1\leqslant \mathrm{M} < j\}} + (\sigma_j/\varsigma_j)^2\,\mathbb{1}_{\{j\leqslant \mathrm{M}\leqslant G_\varepsilon\}}\} + \sum_{j>G_\varepsilon}(\theta_j^\times - \theta_j^\circ)^2\\ &\qquad\leqslant \sum_{j=1}^{G_\varepsilon^-}(\theta_j^\times - \theta_j^\circ)^2\mathbb{E}_{\theta^\circ}\{\mathbb{1}_{\{1\leqslant \mathrm{M} < j\}} + d^{-2}\varepsilon\Lambda_j\}\\ &+ \sum_{j=G_\varepsilon^-+1}^{G_\varepsilon}(\theta_j^\times - \theta_j^\circ)^2\mathbb{E}_{\theta^\circ}\{\mathbb{1}_{\{1\leqslant \mathrm{M} < j\}} + \mathbb{1}_{\{j\leqslant \mathrm{M}\leqslant G_\varepsilon\}}\} + \sum_{j>G_\varepsilon}(\theta_j^\times - \theta_j^\circ)^2\\ &\qquad\qquad\leqslant \|\theta^\times - \theta^\circ\|_{\ell_2}^2\{\mathbb{E}_{\theta^\circ}P_{\mathrm{M}\,|\,\mathrm{Y}}(1\leqslant \mathrm{M} < G_\varepsilon^-) + d^{-2}\varepsilon\Lambda_{(G_\varepsilon^-)}\} + \sum_{j>G_\varepsilon^-}(\theta_j^\times - \theta_j^\circ)^2\end{aligned}$$

The assertion ***(ii)*** follows now by combining the last estimate and Lemma 4.1 ***(i)***, which completes the proof. □

Proof of Theorem 4.4. We start the proof with the observation that $\widehat{\theta}_j - \theta_j^\circ = (\theta_j^{\mathrm{Y}} - \theta_j^\circ)P_{\mathrm{M}\,|\,\mathrm{Y}}(j\leqslant \mathrm{M}\leqslant G_\varepsilon) + (\theta_j^\times - \theta_j^\circ)P_{\mathrm{M}\,|\,\mathrm{Y}}(1\leqslant \mathrm{M} < j)\}$ for all $1\leqslant j\leqslant G_\varepsilon$ and $\widehat{\theta}_j - \theta_j^\circ = \theta_j^\times - \theta_j^\circ$ for all $j > G_\varepsilon$. From the identity $\theta_j^{\mathrm{Y}} - \theta_j^\circ = (\sigma_j/\varsigma_j)(\theta_j^\times - \theta_j^\circ) + (\sigma_j\lambda_j\varepsilon^{-1})(\mathrm{Y}_j - \lambda_j\theta_j^\circ)$ and Lemma B.2 follows that

$$\begin{aligned}&\mathbb{E}_{\theta^\circ}\|\widehat{\theta} - \theta^\circ\|_{\ell_2}^2 \leqslant \sum_{j=1}^{G_\varepsilon}2\sigma^2\lambda_j^2\varepsilon^{-2}\mathbb{E}_{\theta^\circ}\{(\mathrm{Y}_j - \lambda_j\theta_j^\circ)P_{\mathrm{M}\,|\,\mathrm{Y}}(j\leqslant \mathrm{M}\leqslant G_\varepsilon)\\ &+\sum_{j=1}^{G_\varepsilon}2(\theta_j^\times - \theta_j^\circ)^2\mathbb{E}_{\theta^\circ}\{(\sigma_j/\varsigma_j)P_{\mathrm{M}\,|\,\mathrm{Y}}(j\leqslant \mathrm{M}\leqslant G_\varepsilon) + P_{\mathrm{M}\,|\,\mathrm{Y}}(1\leqslant \mathrm{M} < j)\}^2 + \sum_{j>G_\varepsilon}(\theta_j^\times - \theta_j^\circ)^2\\ &\leqslant 2\{\varepsilon G_\varepsilon^+\overline{\Lambda}_{G_\varepsilon^+} + 10\Lambda_1\exp\big(-m_\varepsilon^\circ/5 + 2\log G_\varepsilon\big)\}\\ &\qquad\qquad + 2\{\mathfrak{b}_{G_\varepsilon^-} + \|\theta^\times - \theta^\circ\|_{\ell_2}^2\{d^{-2}\varepsilon\Lambda_{(G_\varepsilon^-)} + 2\exp\big(-m_\varepsilon^\circ/5 + \log G_\varepsilon\big)\}\}.\end{aligned}$$

On the other hand side, taking into account the definition (4.4) of G_ε^- and G_ε^+, we have show in the proof of Lemma 4.2 that $\mathfrak{b}_{G_\varepsilon^-} \leqslant 8L_\lambda C_\lambda(1+1/d)\Phi_\varepsilon^\circ$ and $\varepsilon G_\varepsilon^+ \overline{\Lambda}_{G_\varepsilon^+} \leqslant L_\lambda D^o \Lambda_{(D^o)}\Phi_\varepsilon^\circ$, while trivially $\varepsilon\Lambda_{(G_\varepsilon^-)} \leqslant \varepsilon\Lambda_{(m_\varepsilon^\circ)} \leqslant \Phi_\varepsilon^\circ$. By combination of these estimates we obtain

$$\mathbb{E}_{\theta^\circ}\|\widehat{\theta}-\theta^\circ\|_{\ell_2}^2 \leqslant \{2L_\lambda D^o\Lambda_{(D^o)} + 16L_\lambda C_\lambda(1+1/d) + 2d^{-2}\|\theta^\times-\theta^\circ\|_{\ell_2}^2\}\Phi_\varepsilon^\circ \\ + (20\Lambda_1 + 4\|\theta^\times-\theta^\circ\|_{\ell_2}^2)\exp\big(-m_\varepsilon^\circ/5 + 2\log G_\varepsilon - \log\Phi_\varepsilon^\circ\big)\}\Phi_\varepsilon^\circ$$

From the last bound follows the assertion of the theorem since $(2\log G_\varepsilon - \log\Phi_\varepsilon^\circ)/m_\varepsilon^\circ \to 0$ as $\varepsilon\to 0$ which completes the proof. □

B.3 Proof of Theorem 4.6

Proof of Lemma 4.5. The proof follows along the lines of the proof of Lemma 4.1, where we replace G_ε^-, G_ε^+, m_ε° and Φ_ε° by its counterpart $G_\varepsilon^{\star-}$, $G_\varepsilon^{\star+}$, $m_\varepsilon^\star$ and $\Phi_\varepsilon^\star$, respectively. Moreover, we will use without further reference, that for all $\theta^\circ\in\Theta_{\mathfrak{a}}^r$ the bias is bound by $\mathfrak{b}_m \leqslant r\mathfrak{a}_m$, for all $m\in\mathbb{N}$, and hence $\mathfrak{b}_{m_\varepsilon^\star} \leqslant (1\vee r)\Phi_\varepsilon^\star$.

Consider *(i)*. The claim holds trivially true in case $G_\varepsilon^{\star-}=1$, thus suppose $G_\varepsilon^{\star-}>1$ and let $1\leqslant m < G_\varepsilon^{\star-} \leqslant m_\varepsilon^\star$. Define $S_m := \|\widehat{\theta}^{m_\varepsilon^\star}-\theta^\times\|_\sigma^2 - \|\widehat{\theta}^m-\theta^\times\|_\sigma^2$. Let $\mathcal{A}_m$ and $\mathcal{A}_m^c$, respectively, be an event and its complement defined as in the Proof of Lemma 4.1, then it follows

$$p_{\mathrm{M}\,|\,\mathrm{Y}}(m) \leqslant \exp\left(\frac{1}{2}\big\{-S_m + 3C_\lambda[m_\varepsilon^\star - m]\big\}\right)\mathbb{1}_{\mathcal{A}_m} + \mathbb{1}_{\mathcal{A}_m^c} \tag{B.11}$$

where $S_m = \sum_{j=m+1}^{m_\varepsilon^\star}\frac{\lambda_j^2\sigma_j}{\varepsilon^2}(Y_j-\lambda_j\theta_j^\times)^2$. We use the notation introduced in Lemma 4.1, where again $1\geqslant\beta_j^2\geqslant(1+1/d)^{-1}$ due to Assumption A.2 and by employing $\min_{m<j\leqslant m_\varepsilon^\star}\lambda_j^2 \geqslant \Lambda_{(m_\varepsilon^\star)}^{-1}$ together with Assumption A.4 (iii)

$$L_\lambda(\varepsilon\Lambda_{(m_\varepsilon^\star)})^{-1}(1\vee r)\Phi_\varepsilon^\star \geqslant L_\lambda(\varepsilon\Lambda_{(m_\varepsilon^\star)})^{-1}\varepsilon m_\varepsilon^\star\overline{\Lambda}_{m_\varepsilon^\star} \geqslant m_\varepsilon^\star \quad \text{and} \\ (1+1/d)^{-1}(\varepsilon\Lambda_{(m_\varepsilon^\star)})^{-1}[\mathfrak{b}_m - (1\vee r)\Phi_\varepsilon^\star] \leqslant (1+1/d)^{-1}(\varepsilon\Lambda_{(m_\varepsilon^\star)})^{-1}[\mathfrak{b}_m - \mathfrak{b}_{m_\varepsilon^\star}] \leqslant r_m. \tag{B.12}$$

By employing successively Lemma 3.1, (B.12) and $\mathfrak{b}_{m_\varepsilon^\star} \leqslant (1\vee r)\Phi_\varepsilon^\star$ for all $\theta^\circ\in\Theta_{\mathfrak{a}}^r$ it follows now from (B.11) that

$$\mathbb{E}_{\theta^\circ}p_{\mathrm{M}\,|\,\mathrm{Y}}(m) \leqslant \exp\big(-\frac{\mathfrak{b}_m}{4(1+1/d)\varepsilon\Lambda_{(m_\varepsilon^\star)}} + \frac{2C_\lambda L_\lambda(1\vee r)\Phi_\varepsilon^\star}{\varepsilon\Lambda_{(m_\varepsilon^\star)}}\big)\times\exp\big(-\frac{L_\lambda C_\lambda(1\vee r)\Phi_\varepsilon^\star}{4\varepsilon\Lambda_{(m_\varepsilon^\star)}}\big) \\ + \exp\big(-\frac{[\mathfrak{b}_m - (1\vee r)\Phi_\varepsilon^\star]}{32(1+1/d)\varepsilon\Lambda_{(m_\varepsilon^\star)}}\big).$$

Taking into account the definition (4.5) of $G_\varepsilon^{\star-}$, i.e., $\mathfrak{b}_m > 8L_\lambda C_\lambda(1+1/d)(1\vee r)\Phi_\varepsilon^\star$ for all $1\leqslant m < G_\varepsilon^{\star-}$, and $L_\lambda\Phi_\varepsilon^\star(\varepsilon\Lambda_{(m_\varepsilon^\star)})^{-1}\geqslant m_\varepsilon^\star$ due to Assumption A.4 (iii), we obtain

$$\mathbb{E}_{\theta^\circ}p_{\mathrm{M}\,|\,\mathrm{Y}}(m)\leqslant \exp\big(-\frac{L_\lambda C_\lambda(1\vee r)\Phi_\varepsilon^\star}{4\varepsilon\Lambda_{(m_\varepsilon^\star)}}\big)+\exp\big(-\frac{7L_\lambda C_\lambda(1\vee r)\Phi_\varepsilon^\star}{32\varepsilon\Lambda_{(m_\varepsilon^\star)}}\big)\leqslant 2\exp\big(-\frac{7C_\lambda(1\vee r)}{32}m_\varepsilon^\star\big).$$

Thereby, $\mathbb{E}_{\theta^\circ}P_{\mathrm{M}\,|\,\mathrm{Y}}(1\leqslant \mathrm{M} < G_\varepsilon^{\star-})=\sum_{m=1}^{G_\varepsilon^{\star-}-1}\mathbb{E}_{\theta^\circ}p_{\mathrm{M}\,|\,\mathrm{Y}}(m)\leqslant 2\exp\big(-\frac{7C_\lambda(1\vee r)}{32}m_\varepsilon^\star+\log G_\varepsilon\big)$ using that $G_\varepsilon\geqslant G_\varepsilon^{\star-}$ which proves the assertion ***(i)***. Consider now ***(ii)***. The claim holds trivially true in case $G_\varepsilon^{\star+}=G_\varepsilon$, thus suppose $G_\varepsilon^{\star+}<G_\varepsilon$ and let $G_\varepsilon\geqslant m > G_\varepsilon^{\star+}\geqslant m_\varepsilon^\star$. Consider the upper bound (B.11) where $-S_m=\sum_{j=m_\varepsilon^\star+1}^m\frac{\lambda_j^2\sigma_j}{\varepsilon^2}(Y_j-\lambda_j\theta_j^\times)^2$. Employing the notations introduced in the Proof of Lemma 4.1 where we had $1\geqslant \beta_j^2\geqslant(1+1/d)^{-1}$ due to Assumption A.2, we obtain from Assumption A.4 (i) that $\max_{m_\varepsilon^\star<j\leqslant m}\lambda_j^2\leqslant C_\lambda\Lambda_{(m_\varepsilon^\star)}^{-1}$ and taking into account in addition Assumption A.4 (iii) that

$$\begin{aligned}&L_\lambda(\varepsilon\Lambda_{(m_\varepsilon^\star)})^{-1}(1\vee r)\Phi_\varepsilon^\star\geqslant m_\varepsilon^\star,\quad v_m\leqslant m-m_\varepsilon^\star\quad\text{and}\\ &\qquad C_\lambda(\varepsilon\Lambda_{(m_\varepsilon^\star)})^{-1}(1\vee r)\Phi_\varepsilon^\star\geqslant C_\lambda(\varepsilon\Lambda_{(m_\varepsilon^\star)})^{-1}[\mathfrak{b}_{m_\varepsilon^\star}-\mathfrak{b}_m]\geqslant r_m.\end{aligned}\tag{B.13}$$

By employing successively Lemma 3.1, (B.13) and $\mathfrak{b}_{m_\varepsilon^\star}\leqslant(1\vee r)\Phi_\varepsilon^\circ$ it follows now from (B.11) that

$$\begin{aligned}\mathbb{E}_{\theta^\circ}p_{\mathrm{M}\,|\,\mathrm{Y}}(m)\leqslant \exp\big(C_\lambda\{-m+5L_\lambda(\Lambda_{(m_\varepsilon^\star)}\varepsilon)^{-1}(1\vee r)\Phi_\varepsilon^\star\}/2\big)\times\exp\big(-\frac{C_\lambda L_\lambda(1\vee r)\Phi_\varepsilon^\star}{2\Lambda_{(m_\varepsilon^\star)}\varepsilon}\big)\\ +\exp\big(-C_\lambda(m-L_\lambda(\varepsilon\Lambda_{(m_\varepsilon^\star)})^{-1}(1\vee r)\Phi_\varepsilon^\star)/9\big)\end{aligned}$$

Taking into account the definition (4.5) of $G_\varepsilon^{\star+}$, i.e., $m>5L_\lambda(\varepsilon\Lambda_{(m_\varepsilon^\star)})^{-1}(1\vee r)\Phi_\varepsilon^\star$ for all $G_\varepsilon\geqslant m > G_\varepsilon^{\star+}$, and $L_\lambda(1\vee r)\Phi_\varepsilon^\star(\varepsilon\Lambda_{(m_\varepsilon^\star)})^{-1}\geqslant(1\vee r)m_\varepsilon^\star$ due to Assumption A.4 (iii), we obtain

$$\mathbb{E}_{\theta^\circ}p_{\mathrm{M}\,|\,\mathrm{Y}}(m)\leqslant \exp\big(-\frac{L_\lambda C_\lambda(1\vee r)\Phi_\varepsilon^\star}{2\varepsilon\Lambda_{(m_\varepsilon^\star)}}\big)+\exp\big(-\frac{4L_\lambda C_\lambda(1\vee r)\Phi_\varepsilon^\star}{9\varepsilon\Lambda_{(m_\varepsilon^\star)}}\big)\leqslant 2\exp\big(-\frac{4C_\lambda(1\vee r)}{9}m_\varepsilon^\star\big).$$

Thereby, $\mathbb{E}_{\theta^\circ}P_{\mathrm{M}\,|\,\mathrm{Y}}(G_\varepsilon^{\star+}<\mathrm{M}\leqslant G_\varepsilon)=\sum_{m=G_\varepsilon^{\star+}+1}^{G_\varepsilon}\mathbb{E}_{\theta^\circ}p_{\mathrm{M}\,|\,\mathrm{Y}}(m)\leqslant 2\exp\big(-\frac{4C_\lambda(1\vee r)}{9}m_\varepsilon^\star+\log G_\varepsilon\big)$ which shows the assertion ***(ii)*** and completes the proof. □

Proof of Theorem 4.6. We start the proof with the observation that due to Assumption A.4 (iii) the condition (3.7) holds true with $L=L_\lambda$ uniformly for all $m\in\mathbb{N}$ and $\varepsilon\in(0,1)$, and hence imposing Assumption A.2 the conditions of Corollary 3.6 (3.8) are satisfied, which in turn implies, by setting $K:=((1+1/d)\vee r/d^2)L_\lambda\geqslant((1+d^{-1})\vee d^{-2}\|\theta^\circ-\theta^\times\|_{\ell_2}^2)L_\lambda$, that for all $1\leqslant m\leqslant G_\varepsilon$ and $\varepsilon\in(0,\varepsilon_\star)$

$$\mathbb{E}_{\theta^\circ}P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^m-\theta^\circ\|_{\ell_2}^2>(4+(11/2)K)[\mathfrak{b}_m\vee\varepsilon m\overline{\Lambda}_m])\leqslant 2\exp(-m/36).\tag{B.14}$$

Moreover, exploiting the inequality below (A.3) with $c_1 = 1/3$ and $c_2 \geqslant 1$, it is possible to prove a slightly modified version of Corollary 3.6 (3.8) which implies for all $c_2 \geqslant 1$

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m \,|\, \mathrm{Y}}(\|\boldsymbol{\vartheta}^m - \theta^\circ\|_{\ell_2}^2 > 16 c_2 K [\mathfrak{b}_m \vee \varepsilon m \overline{\Lambda}_m]) \leqslant 2 \exp(-c_2 m/12). \tag{B.15}$$

Consider *(i)*. Following line by line the proof of Lemma 4.2 *(i)*, using (B.14) rather than (B.4) and exploiting $[\mathfrak{b}_m \vee \varepsilon m \overline{\Lambda}_m] \leqslant 8 L_\lambda C_\lambda (1+1/d)(1 \vee r) \Phi_\varepsilon^\star$ for all $G_\varepsilon^{\star -} \leqslant m \leqslant m_\varepsilon^\star$ and $[\mathfrak{b}_m \vee \varepsilon m \overline{\Lambda}_m] \leqslant L_\lambda D^\star \Lambda_{(D^\star)} (1 \vee r) \Phi_\varepsilon^\star$ with $D^\star := \lceil 5 L_\lambda / \kappa^\star \rceil$ for all $m_\varepsilon^\star \leqslant m \leqslant G_\varepsilon^{\star +}$ (keep in mind that $m \leqslant D^\star m_\varepsilon^\star$) , we obtain

$$\begin{aligned}
&\sum_{m=G_\varepsilon^{\star -}}^{G_\varepsilon^{\star +}} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m \,|\, \mathrm{Y}}\big(\|\boldsymbol{\vartheta}^m - \theta^\circ\|_{\ell_2}^2 > K^\star \Phi_\varepsilon^\star\big) \\
&\leqslant \sum_{m=G_\varepsilon^{\star -}}^{G_\varepsilon^{\star +}} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m \,|\, \mathrm{Y}}\big(\|\boldsymbol{\vartheta}^m - \theta^\circ\|_{\ell_2}^2 > (4 + (11/2)K)[\mathfrak{b}_m \vee \varepsilon m \overline{\Lambda}_m]\big) \\
&\qquad\qquad \leqslant 2 \sum_{m=G_\varepsilon^{\star -}}^{G_\varepsilon^{\star +}} \exp(-m/36) \leqslant 74 \exp(-G_\varepsilon^{\star -}/36).
\end{aligned}$$

Combining the last estimate, Lemma 4.5 and the decomposition (B.6) used in the proof of Theorem 4.3 (with G_ε^- and G_ε^+ replaced by $G_\varepsilon^{\star -}$, $G_\varepsilon^{\star +}$) it follows that

$$\begin{aligned}
&\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{\mathrm{M}} \,|\, \mathrm{Y}}(\|\boldsymbol{\vartheta}^{\mathrm{M}} - \theta^\circ\|_{\ell_2}^2 > K^\star \Phi_\varepsilon^\star) \\
&\qquad\qquad \leqslant 4 \exp\big(- m_\varepsilon^\star \{C_\lambda/5 - \log G_\varepsilon / m_\varepsilon^\star\}\big) + 74 \exp(-G_\varepsilon^{\star -}/36)
\end{aligned} \tag{B.16}$$

Taking into account that $m_\varepsilon^\star \to \infty$ and $\log G_\varepsilon / m_\varepsilon^\star = o(1)$ as $\varepsilon \to 0$, we obtain the assertion *(i)* of the Theorem for any $\theta^\circ \in \Theta_{\mathfrak{a}}^r$ such that $G_\varepsilon^{\star -} \to \infty$ as $\varepsilon \to 0$. On the other hand side, if $\theta^\circ \in \Theta_{\mathfrak{a}}^r$ such that $G_\varepsilon^{\star -} \not\to \infty$, i.e., $\sup_\varepsilon G_\varepsilon^{\star -} < \infty$, then there exists $\varepsilon_\circ \in (0,1)$ such that $G_{\varepsilon_\circ}^{\star -} = G_\varepsilon^{\star -}$ for all $\varepsilon \in (0, \varepsilon_\circ)$ (keep in mind that $(G_\varepsilon^{\star -})_\varepsilon$ is an integer-valued monotonically increasing sequence). Moreover, by construction $\mathfrak{b}_{G_{\varepsilon_\circ}^{\star -}} \leqslant 8 L_\lambda C_\lambda (1 + 1/d)(1 \vee r) \Phi_\varepsilon^\star$ for all $\varepsilon \in (0, \varepsilon_\circ)$ which in turn implies $\mathfrak{b}_m \leqslant \mathfrak{b}_{G_{\varepsilon_\circ}^{\star -}} = 0$ for all $m \geqslant G_{\varepsilon_\circ}^{\star -}$, since $\Phi_\varepsilon^\star = o(1)$ as $\varepsilon \to 0$. Thereby, for all $m \geqslant G_{\varepsilon_\circ}^{\star -}$ follows $\Phi_\varepsilon^\star / [\mathfrak{b}_m \vee \varepsilon m \overline{\Lambda}_m] = \Phi_\varepsilon^\star / [\varepsilon m \overline{\Lambda}_m] \geqslant [\varepsilon m_\varepsilon^\star \overline{\Lambda}_{m_\varepsilon^\star}] / [\varepsilon m \overline{\Lambda}_m] \geqslant m_\varepsilon^\star / [L_\lambda m]$ using that $L_\lambda \overline{\Lambda}_{m_\varepsilon^\star} \geqslant \Lambda_{(m_\varepsilon^\star)} \geqslant \Lambda_{(m)} \geqslant \overline{\Lambda}_m$ due to Assumption A.4 (iii), which in turn together with $K^\star \Phi_\varepsilon^\star / [\mathfrak{b}_m \vee \varepsilon m \overline{\Lambda}_m] \geqslant K^\star m_\varepsilon^\star / [L_\lambda m] = 16 c_2 K$, $c_2 := (8 C_\lambda (1 + 1/d) \vee D^\star \Lambda_{(D^\star)})(1 \vee r) m_\varepsilon^\star / m \geqslant 1$ and (B.15) implies

$$\begin{aligned}
&\sum_{m=G_\varepsilon^{\star -}}^{G_\varepsilon^{\star +}} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m \,|\, \mathrm{Y}}\big(\|\boldsymbol{\vartheta}^m - \theta^\circ\|_{\ell_2}^2 > K^\star \Phi_\varepsilon^\star\big) \\
&\leqslant 2 \exp(-(8 C_\lambda (1 + 1/d) \vee D^\star \Lambda_{(D^\star)})(1 \vee r) m_\varepsilon^\star / 12 + \log G_\varepsilon) \\
&\qquad\qquad \leqslant 2 \exp(-C_\lambda m_\varepsilon^\star / 5 + \log G_\varepsilon).
\end{aligned}$$

Consequently, we have

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{\mathrm{M}}\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^{\mathrm{M}} - \theta^\circ\|_{\ell_2}^2 > K^\star \Phi_\varepsilon^\star) \leqslant 6 \exp\big(-m_\varepsilon^\star\{C_\lambda/5 - \log G_\varepsilon/m_\varepsilon^\star\}\big)$$

which shows that assertion ***(i)*** holds for any $\theta^\circ \in \Theta_{\mathfrak{a}}^r$ since $m_\varepsilon^\star \to \infty$ and $\log G_\varepsilon/m_\varepsilon^\star = o(1)$ as $\varepsilon \to 0$. Consider ***(ii)***. Employing that for all $\theta^\circ \in \Theta_{\mathfrak{a}}^r$ it holds $K^\star \Phi_\varepsilon^\star \geqslant 16K[\mathfrak{b}_m \vee \varepsilon m \overline{\Lambda}_m]$ for all $G_\varepsilon^{\star-} \leqslant m \leqslant G_\varepsilon^{\star+}$ it follows that $K_\varepsilon \Phi_\varepsilon^\star/[\mathfrak{b}_m \vee \varepsilon m \overline{\Lambda}_m] \geqslant 16 c_2 K$ where $c_2 := K_\varepsilon/K^\star \geqslant 12$ for all $\varepsilon \in (0, \widetilde{\varepsilon}_\star)$ since $K_\varepsilon \to \infty$ as $\varepsilon \to 0$. Therefore, by applying (B.15) we have

$$\sum_{m=G_\varepsilon^{\star-}}^{G_\varepsilon^{\star+}} \mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^m\,|\,\mathrm{Y}}\big(\|\boldsymbol{\vartheta}^m - \theta^\circ\|_{\ell_2}^2 > K_\varepsilon \Phi_\varepsilon^\star\big) \leqslant 4\exp(-K_\varepsilon/[12K^\star]).$$

and hence from Lemma 4.5 follows for all $\varepsilon \leqslant (\widetilde{\varepsilon}_\star \wedge \varepsilon_\star)$

$$\mathbb{E}_{\theta^\circ} P_{\boldsymbol{\vartheta}^{\mathrm{M}}\,|\,\mathrm{Y}}(\|\boldsymbol{\vartheta}^{\mathrm{M}} - \theta^\circ\|_{\ell_2}^2 > K_\varepsilon \Phi_\varepsilon^\star) \\ \leqslant 4 \exp\big(-m_\varepsilon^\star\{C_\lambda/5 - \log G_\varepsilon/m_\varepsilon^\star\}\big) + 4\exp(-K_\varepsilon/[12K^\star]).$$

Observe, that $(\widetilde{\varepsilon}_\star \wedge \varepsilon_\star)$ depends only on the class $\Theta_{\mathfrak{a}}^r$ and thus the upper bound given in the last display holds true uniformly for all $\theta^\circ \in \Theta_{\mathfrak{a}}^r$, which implies the assertion ***(ii)*** by using that $K_\varepsilon \to \infty$, $m_\varepsilon^\star \to \infty$ and $\log G_\varepsilon/m_\varepsilon^\star = o(1)$ as $\varepsilon \to 0$, and completes the proof. □

B.4 Proof of Theorem 4.7

LEMMA B.3. *If Assumption A.2 and A.4 hold true then for all $\theta^\circ \in \Theta_{\mathfrak{a}}^r$ and $\varepsilon \in (0, \varepsilon_o)$*

(i) $\sum_{j=1}^{G_\varepsilon} \sigma_j^2 \lambda_j^2 \varepsilon^{-2} \mathbb{E}_{\theta^\circ}\{(\mathrm{Y}_j - \lambda_j \theta_j^\circ) P_{\mathrm{M}\,|\,\mathrm{Y}}(j \leqslant \mathrm{M} \leqslant G_\varepsilon)\}^2 \leqslant \varepsilon G_\varepsilon^{\star+} \overline{\Lambda}_{G_\varepsilon^{\star+}} + 10\Lambda_1 \exp\big(-m_\varepsilon^\star/5 + 2\log G_\varepsilon\big);$

(ii) $\sum_{j=1}^{G_\varepsilon} (\theta_j^\times - \theta_j^\circ)^2 \mathbb{E}_{\theta^\circ} \mathbb{E}_{\mathrm{M}\,|\,\mathrm{Y}}\{\mathbb{1}_{\{1 \leqslant \mathrm{M} < j\}} + (\sigma_j^2 \varsigma_j^{-1})^2\, \mathbb{1}_{\{j \leqslant \mathrm{M} \leqslant G_\varepsilon\}}\} + \sum_{j > G_\varepsilon} (\theta_j^\times - \theta_j^\circ)^2 \leqslant \mathfrak{b}_{G_\varepsilon^{\star-}} + \|\theta^\times - \theta^\circ\|_{\ell_2}^2 \{d^{-2}\varepsilon \Lambda_{(G_\varepsilon^{\star-})} + 2\exp\big(-\frac{C_\lambda(1 \vee r)}{5} m_\varepsilon^\star + \log G_\varepsilon\big)\}.$

PROOF OF LEMMA B.3. The proof follows along the lines of the proof of Lemma B.2, where we replace G_ε^-, G_ε^+, m_ε° and Φ_ε° by its counterpart $G_\varepsilon^{\star-}$, $G_\varepsilon^{\star+}$, $m_\varepsilon^\star$ and $\Phi_\varepsilon^\star$, respectively.

Consider ***(i)***. Following the proof of (B.9) it is straightforward to see that

$$\sum_{j=1}^{G_\varepsilon} \sigma_j^2 \lambda_j^2 \varepsilon^{-2} \{(\mathrm{Y}_j - \lambda_j \theta_j^\circ) P_{\mathrm{M}\,|\,\mathrm{Y}}(j \leqslant \mathrm{M} \leqslant G_\varepsilon)\}^2 \\ \leqslant \sum_{j=1}^{G_\varepsilon^{\star+}} \varepsilon \Lambda_j \xi_j^2 + \sum_{j=1}^{G_\varepsilon} \varepsilon \Lambda_j \xi_j^2 P_{\mathrm{M}\,|\,\mathrm{Y}}(G_\varepsilon^{\star+} < \mathrm{M} \leqslant G_\varepsilon) \quad \text{(B.17)}$$

and following line by line the proof of (B.10) we conclude

$$\sum_{j=1}^{G_\varepsilon} \varepsilon\Lambda_j \mathbb{E}_{\theta^\circ}\{\xi_j^2 P_{\mathrm{M}\,|\,\mathrm{Y}}(G_\varepsilon^{\star+} < \mathrm{M} \leqslant G_\varepsilon)\} \\ \leqslant 6\Lambda_1 \exp(-G_\varepsilon/6) + 2\Lambda_1 G_\varepsilon \mathbb{E}_{\theta^\circ} P_{\mathrm{M}\,|\,\mathrm{Y}}(G_\varepsilon^{\star+} < \mathrm{M} \leqslant G_\varepsilon). \quad \text{(B.18)}$$

We distinguish two cases. First, if $G_\varepsilon^{\star+} = G_\varepsilon$, then assertion ***(i)*** follows by combining (B.17) and $\mathbb{E}_{\theta^\circ}\xi_j^2 = 1$. Second, if $G_\varepsilon^{\star+} < G_\varepsilon$, then the definition (4.5) of $G_\varepsilon^{\star+}$ implies $G_\varepsilon > 5m_\varepsilon^\star$ which in turn implies the assertion ***(i)*** by combining (B.17), $\mathbb{E}_{\theta^\circ}\xi_j^2 = 1$, (B.18) and Lemma 4.5 ***(i)***. Consider ***(ii)***. Following the proof of Lemma B.2 ***(ii)*** we obtain

$$\sum_{j=1}^{G_\varepsilon}(\theta_j^\times - \theta_j^\circ)^2 \mathbb{E}_{\theta^\circ}\{\mathbb{1}_{\{1 \leqslant \mathrm{M} < j\}} + (\sigma_j/\varsigma_j)^2\, \mathbb{1}_{\{j \leqslant \mathrm{M} \leqslant G_\varepsilon\}}\} + \sum_{j>G_\varepsilon}(\theta_j^\times - \theta_j^\circ)^2 \\ \leqslant \|\theta^\times - \theta^\circ\|_{\ell_2}^2\{\mathbb{E}_{\theta^\circ} P_{\mathrm{M}\,|\,\mathrm{Y}}(1 \leqslant \mathrm{M} < G_\varepsilon^{\star-}) + d^{-2}\varepsilon\Lambda_{(G_\varepsilon^{\star-})}\} + \sum_{j>G_\varepsilon^{\star-}}(\theta_j^\times - \theta_j^\circ)^2$$

The assertion ***(ii)*** follows now by combining the last estimate and Lemma 4.5 ***(ii)***, which completes the proof. □

Proof of Theorem 4.7. The proof follows line by line the proof of Theorem 4.4 using Lemma B.3 rather than Lemma B.2, more precisely from Lemma B.3 follows

$$\mathbb{E}_{\theta^\circ}\|\widehat{\theta} - \theta^\circ\|_{\ell_2}^2 \leqslant \sum_{j=1}^{G_\varepsilon} 2\sigma^2\lambda_j^2\varepsilon^{-2}\mathbb{E}_{\theta^\circ}\{(\mathrm{Y}_j - \lambda_j\theta_j^\circ)P_{\mathrm{M}\,|\,\mathrm{Y}}(j \leqslant \mathrm{M} \leqslant G_\varepsilon)\\ + \sum_{j=1}^{G_\varepsilon} 2(\theta_j^\times - \theta_j^\circ)^2\mathbb{E}_{\theta^\circ}\{(\sigma_j/\varsigma_j)P_{\mathrm{M}\,|\,\mathrm{Y}}(j \leqslant \mathrm{M} \leqslant G_\varepsilon) + P_{\mathrm{M}\,|\,\mathrm{Y}}(1 \leqslant \mathrm{M} < j)\}^2 + \sum_{j>G_\varepsilon}(\theta_j^\times - \theta_j^\circ)^2 \\ \leqslant 2\{\varepsilon G_\varepsilon^{\star+}\overline{\Lambda}_{G_\varepsilon^{\star+}} + 10\Lambda_1\exp\big(-\frac{C_\lambda(1\vee r)}{5}m_\varepsilon^\star + 2\log G_\varepsilon\big)\} \\ + 2\{\mathfrak{b}_{G_\varepsilon^{\star-}} + \|\theta^\times - \theta^\circ\|_{\ell_2}^2\{d^{-2}\varepsilon\Lambda_{(G_\varepsilon^{\star-})} + 2\exp\big(-\frac{C_\lambda(1\vee r)}{5}m_\varepsilon^\star + \log G_\varepsilon\big)\}\}.$$

Taking further into account the definition (4.5) of $G_\varepsilon^{\star-}$ and $G_\varepsilon^{\star+}$, we have $\mathfrak{b}_{G_\varepsilon^{\star-}} \leqslant 8L_\lambda C_\lambda(1+1/d)(1\vee r)\Phi_\varepsilon^\star$ and (keeping in mind Assumption A.3) $G_\varepsilon^{\star+} \leqslant D^\star m_\varepsilon^\star$ with $D^\star := D^\star(\Theta_{\mathfrak{a}}^r, \lambda) := \lceil 5L_\lambda(1\vee r)/\kappa\rceil$, which in turn implies $\varepsilon G_\varepsilon^{\star+}\overline{\Lambda}_{G_\varepsilon^{\star+}} \leqslant L_\lambda D^\star\Lambda_{(D^\star)}\Phi_\varepsilon^\star$, while trivially $\varepsilon\Lambda_{(G_\varepsilon^{\star-})} \leqslant \varepsilon\Lambda_{(m_\varepsilon^\star)} \leqslant \Phi_\varepsilon^\star$ and $\|\theta^\times - \theta^\circ\|_{\ell_2}^2 \leqslant r$. By combination of these estimates we obtain uniformly for all $\theta^\circ \in \Theta_{\mathfrak{a}}^r$ that

$$\mathbb{E}_{\theta^\circ}\|\widehat{\theta} - \theta^\circ\|_{\ell_2}^2 \leqslant \{2L_\lambda D^\star\Lambda_{(D^\star)} + 16L_\lambda C_\lambda(1+1/d)(1\vee r) + 2d^{-2}r\}\Phi_\varepsilon^\star \\ + (20\Lambda_1 + 4r)\exp\big(-\frac{C_\lambda(1\vee r)}{5}m_\varepsilon^\star + 2\log G_\varepsilon - \log\Phi_\varepsilon^\star\big)\}\Phi_\varepsilon^\star.$$

Note that in the last display the multiplicative factors of $\Phi_\varepsilon^\star$ depend only on the class $\Theta_\mathfrak{a}^r$, the constant d and the sequence λ. Thereby, the assertion of the theorem follows from $\log(G_\varepsilon/\Phi_\varepsilon^\star)/m_\varepsilon^\star \to 0$ as $\varepsilon \to 0$ which completes the proof. □

References

F. Abramovich, T. Sapatinas, and B. W. Silverman. Wavelet thresholding via a bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(4):725–749, 1998.

S. Agapiou, S. Larsson, and A. Stuart. Posterior consistency of the bayesian approach to linear ill-posed inverse problems. *Stochastic Processes and Applications*, 123:3828–3860, 2013.

J. Arbel, G. Gayraud, and J. Rousseau. Bayesian optimal adaptive estimation using a sieve prior. *Scandinavian J. Statist.*, 40:549–570, 2013.

A. Barron, L. Birgé, and P. Massart. Risk bounds for model selection via penalization. *Probability Theory and Related Fields*, 113(3):301–413, 1999.

E. Belitser and S. Ghosal. Adaptive bayesian inference on the mean of an infinite-dimensional normal distribution. *Ann. Statist.*, 31:536–559, 2003.

L. Birgé. An alternative point of view on lepski's method. In Monogr., editor, *State of the art in probability and statistics*, volume 36 of *IMS Lecture Notes*, pages 113–133. (Leiden 1999), 2001.

N. Bissantz, T. Hohage, A. Munk, and F. H. Ruymgaart. Convergence rates of general regularization methods for statistical inverse problems and applications. *SIAM Journal on Numerical Analysis*, 45(6):2610–2636, 2007.

L. D. Brown and M. G. Low. Asymptotic equivalence of nonparametric regression and white noise. *The Annals of Statistics*, 24:2384–2398, 1996.

I. Castillo. Lower bounds for posterior rates with gaussian process priors. *Electron. J. Statist.*, 2:1281–1299, 2008.

L. Cavalier. Nonparametric statistical inverse problems. *Inverse Problems*, 24:1–19, 2008.

L. Cavalier and N. W. Hengartner. Adaptive estimation for inverse problems with noisy operators. *Inverse Problems*, 21:1345–1361, 2005.

L. Cavalier, G. Golubev, D. Picard, and A. Tsybakov. Oracle inequalities for inverse problems. *The Annals of Statistics*, 30(3):843–874, 2002.

L. Cavalier, Y. Golubev, O. Lepski, and A. Tsybakov. Block thresholding and sharp adaptive estimation in severely ill-posed inverse problems. *Theory of Probability and its Applications*, 48(3):426–446, 2003.

F. Comte and J. Johannes. Adaptive functional linear regression. *The Annals of Statistics*, 40(6):2765–2797, 2012.

D. D. Cox. An analysis of bayesian inference for nonparametric regression. *The Annals of Statistics*, 21:903–923, 1993.

S. Efromovich and V. Koltchinskii. On inverse problems with unknown operators. *IEEE Transactions on Information Theory*, 47(7):2876–2894, 2001.

S. N. Evans and P. B. Stark. Inverse problems as statistics. *Inverse Problems*, 18(4): R55, 2002.

J.-P. Florens and A. Simoni. Regularizing priors for linear inverse problems. *Econometric Theory*, FirstView:1–51, 2014.

D. Freedman. Wald lecture: On the bernstein-von mises theorem with infinite-dimensional parameters. *The Annals of Statistics*, 27(4):1119–1141, 1999.

C. Gao and H. H. Zhou. Adaptive bayesian estimation via block prior. *Preprint*, 2014.

S. Ghosal, J. K. Ghosh, and A. W. Van Der Vaart. Convergence rates of posterior distributions. *The Annals of Statistics*, pages 500–531, 2000.

J. Ghosh and R. Ramamoorthi. *Bayesian Nonparametrics*. Springer Series in Statistics. Springer-Verlag, New York, 2003.

A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39: 1608–1632, 2011.

M. Hoffmann and M. Reiß. Nonlinear estimation for linear inverse problems with error in the operator. *The Annals of Statistics*, 36(1):310–336, 2008.

J. Johannes and M. Schwarz. Adaptive gaussian inverse regression with partially unknown operator. *Communications in Statistics - Theory and Methods*, 42(7):1343–1362, 2013.

J. Kaipio and E. Somersalo. *Statistical and computational inverse problems.* New York, NY: Springer, 2005.

B. Knapik, A. Van der Vaart, and J. Van Zanten. Bayesian inverse problems with gaussian priors. *The Annals of Statistics*, 39:2626–2657, 2011.

B. Knapik, B. Szabó, A. Van der Vaart, and J. Van Zanten. Bayes procedures for adaptive inference in inverse problems for the white noise model. *Preprint, arXiv:1209.3628v2*, 2014.

A. P. Korostelev and A. B. Tsybakov. *Minimax theory of image reconstruction*, volume 82 of *Lecture Notes in Statistics.* Springer New York, 1993.

B. Laurent, J. Loubes, and C. Marteau. Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. *Electronic Journal of Statistics*, 6: 91–122, 2012.

L. Le Cam. Sufficiency and approximate sufficiency. *The Annals of Mathematical Statistics*, 35:1419–1455, 1964.

O. Lepskij. On a problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatn. Primen.*, 35(3):459–470, 1990.

B. A. Mair and F. H. Ruymgaart. Statistical inverse estimation in Hilbert scales. *SIAM Journal on Applied Mathematics*, 56(5):1424–1444, 1996.

P. Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics.* Springer, Berlin, 2007.

P. Mathé. The Lepskiĭ principle revisited. *Inverse Problems*, 22(3):11–15, 2006.

A. Meister. Asymptotic equivalence of functional linear regression and a white noise inverse problem. *The Annals of Statistics*, 39(3):1471–1495, 2011.

C. Robert. *The Bayesian Choice*, volume 2nd edition of *Springer Texts in Statistics.* Springer, New York, 2007.

X. Shen and L. Wasserman. Rates of convergence of posterior distributions. *Ann. Statist.*, 29:687–714, 2001.

B. Szabó, A. Van der Vaart, and J. Van Zanten. Empirical bayes scaling of gaussian priors in the white noise model. *Electronic Journal of Statistics*, 7:991–1018, 2013.

L. H. Zhao. Bayesian aspects of some nonparametric problems. *The Annals of Statistics*, 28:532–552, 2000.