The Annals of Statistics
2017, Vol. 45, No. 3, 1124–1151
DOI: 10.1214/16-AOS1477

ON COVERAGE AND LOCAL RADIAL RATES OF CREDIBLE SETS

BY EDUARD BELITSER

VU Amsterdam

In the mildly ill-posed inverse signal-in-white-noise model, we construct confidence sets as credible balls with respect to the empirical Bayes posterior resulting from a certain two-level hierarchical prior. The quality of the posterior is characterized by the contraction rate which we allow to be local, that is, depending on the parameter. The issue of optimality of the constructed confidence sets is addressed via a trade-off between its "size" (the *local radial rate*) and its coverage probability. We introduce *excessive bias restriction* (EBR), more general than *self-similarity* and *polished tail condition* recently studied in the literature. Under EBR, we establish the confidence optimality of our credible set with some local (*oracle*) radial rate. We also derive the oracle estimation inequality and the oracle posterior contraction rate. The obtained local results are more powerful than global: adaptive minimax results for a number of smoothness scales follow as consequence, in particular, the ones considered by Szabó et al. [*Ann. Statist.* **43** (2015) 1391–1428].

1. Introduction. Let $\mathbb{N} = \{1, 2, \ldots\}$ and $\sigma = (\sigma_i, i \in \mathbb{N})$ be a positive nondecreasing sequence. We observe

$$X = X^{(\varepsilon)} = (X_i, i \in \mathbb{N}) \sim \mathrm{P}_\theta = \mathrm{P}_\theta^{(\varepsilon)} = \bigotimes_{i \in \mathbb{N}} N(\theta_i, \sigma_i^2), \tag{1}$$

that is, $X_i \overset{\text{ind}}{\sim} N(\theta_i, \sigma_i^2)$, $i \in \mathbb{N}$. Here, $\theta = (\theta_i, i \in \mathbb{N}) \in \Theta = \ell_2$ is an unknown parameter of interest. The general goal is make inference on θ by using a Bayesian approach. Without loss of generality, we set

$$\varepsilon^2 = \min_{i \in \mathbb{N}} \sigma_i^2 = \sigma_1^2 \quad \text{and} \quad \kappa_i = \frac{\sigma_i}{\varepsilon} \geq 1 \qquad \text{so that } \sigma_i^2 = \varepsilon^2 \kappa_i^2.$$

The parameter ε is the noise intensity describing the information increase in the data $X^{(\varepsilon)}$ as $\varepsilon \to 0$. The nondecreasing sequence $\{\kappa_i^2, i \in \mathbb{N}\}$ reflects the ill-posedness of the model. Later, we put certain conditions on this sequence, essentially making it of a polynomial type. To avoid overloaded notation, we often drop the dependence on ε; for example, $X = X^{(\varepsilon)}$ etc.

In this paper, we consider nonasymptotic results (for a fixed $\varepsilon > 0$), implying asymptotic assertions if needed and allowing to measure precisely the effect of

Received January 2016; revised May 2016.
MSC2010 subject classifications. Primary 62G15, 62C05; secondary 62G99.
Key words and phrases. Credible ball, excessive bias restriction, local radial rate.

the different quantities involved in the problem. However, the obtained nonasymptotic bounds are usually useful only for sufficiently small ε, making the setting essentially asymptotic as $\varepsilon \to 0$.

The model (1) is known to be the sequence version of the *inverse signal-in-white-noise model*. There is now a vast literature about this model (cf. [10, 11, 20]), especially for the direct case: $\kappa_i^2 = 1$, $i \in \mathbb{N}$ (cf. [5, 9, 25]). This model is of a canonical type and serves as a purified approximation to some other statistical models.

The main aim is to construct an optimal (to be defined later) confidence set for the parameter $\theta_0 \in \Theta$ on the basis of observation $X \sim \mathrm{P}_{\theta_0}$, with a prescribed coverage probability. We measure the size of a set by the smallest possible radius of an ℓ_2-ball containing that set. It is thus sufficient to consider only confidence balls as confidence sets. A general confidence ball for the parameter θ is of the form $B(\hat{\theta}, \hat{r}) = \{\theta \in \ell_2 : \|\theta - \hat{\theta}\| \le \hat{r}\}$, where $\|\theta\| = (\sum_{i=1}^{\infty} \theta_i^2)^{1/2}$ is the usual ℓ_2-norm, $\hat{\theta} = \hat{\theta}(X) = \hat{\theta}(X, \varepsilon) \in \ell_2$ is some *data dependent center* (DD-center) and $\hat{r} = \hat{r}(X) = \hat{r}(X, \varepsilon) \in \mathbb{R}_+ = \{a \in \mathbb{R} : a \ge 0\}$ is some *data dependent radius* (DD-radius). The quantities $\hat{\theta}$ and $\hat{r}$ are measurable functions of the data.

Let us specify the optimality framework for confidence sets. We would like to construct a confidence ball $B(\hat{\theta}, C\hat{r})$ such that for any $\alpha_1, \alpha_2 \in (0, 1]$ and some functional $r(\theta) = r_\varepsilon(\theta)$, $r_\varepsilon : \Theta \to \mathbb{R}_+$, there exist positive $C = C(\Theta_{\mathrm{cov}}, \alpha_1)$, $c = c(\Theta_{\mathrm{size}}, \alpha_2)$ such that for all $\varepsilon \in (0, \varepsilon_0]$ with some $\varepsilon_0 > 0$,

$$\sup_{\theta \in \Theta_{\mathrm{cov}}} \mathrm{P}_\theta\big(\theta \notin B(\hat{\theta}, C\hat{r})\big) \le \alpha_1, \qquad \sup_{\theta \in \Theta_{\mathrm{size}}} \mathrm{P}_\theta\big(\hat{r} \ge c r_\varepsilon(\theta)\big) \le \alpha_2, \tag{2}$$

where $\Theta_{\mathrm{cov}}, \Theta_{\mathrm{size}} \subseteq \Theta$. In some papers, a confidence set satisfying the first relation in (2) is called *honest* over Θ_{cov}. The quantity $r_\varepsilon(\theta)$ has the meaning of the effective radius of the confidence ball $B(\hat{\theta}, C\hat{r})$. We call $r_\varepsilon(\theta)$ *radial rate*. Clearly, there are many possible radial rates, but it is desirable to find the "fastest" (i.e., smallest) radial rate $r_\varepsilon(\theta)$, for which the relations (2) hold for "massive" $\Theta_{\mathrm{cov}}, \Theta_{\mathrm{size}} \subseteq \Theta$, ideally for $\Theta_{\mathrm{cov}} = \Theta_{\mathrm{size}} = \Theta$. The two relations in (2) are called *coverage* and *size properties*. An asymptotic formulation is also possible: $\limsup_{\varepsilon \to 0}$ should be taken, constants α_1, α_2, C, c (possibly also $\Theta_{\mathrm{cov}}, \Theta_{\mathrm{size}}$) can be allowed to depend on ε.

Thus, the following optimality aspects are involved in the framework (2): the coverage, the radial rate and the uniformity subsets $\Theta_{\mathrm{cov}}, \Theta_{\mathrm{size}}$. The optimality is basically a trade-off between these complementary aspects pushed to the utmost limits, when further improving upon one aspect leads to a deterioration in another aspect. For example, the smaller the local radial rate $r_\varepsilon(\theta)$ in (2), the better. But if it is too small, the size requirement in (2) may hold uniformly only over some "thin" set $\Theta_{\mathrm{size}} \subset \Theta$. On the other hand, if one insists on $\Theta_{\mathrm{cov}} = \Theta_{\mathrm{size}} = \Theta$, then it may be impossible to establish (2) for interesting (relatively small) radial rates $r_\varepsilon(\theta)$.

One approach to optimality is via the minimax estimation framework. It is assumed that $\theta \in \Theta_\beta \subseteq \Theta$ for some "smoothness" parameter $\beta \in \mathcal{B}$, which may be known or unknown (nonadaptive or adaptive formulation). The key notion here is the so-called *minimax rate* $R_\varepsilon(\Theta_\beta)$, the formal definition is given in the supplement [3]. The radial rate is taken to be $r_\varepsilon(\theta) = R_\varepsilon(\Theta_\beta)$, which is a global quantity as it is constant for all $\theta \in \Theta_\beta$.

An adaptation problem arises when, for a given family of models $\{\Theta_\beta, \beta \in \mathcal{B}\}$ (called *scale*), we only know that $\theta \in \Theta_\beta$ for some unknown $\beta \in \mathcal{B}$. In fact, $\theta \in \bigcup_{\beta\in\mathcal{B}} \Theta_\beta \subseteq \Theta$ and the problem becomes in general more difficult. For a $\Theta'_{\mathrm{cov}} \subseteq \Theta$, we want to construct a confidence ball $B(\hat\theta, C\hat r)$ such that

$$\begin{aligned} &\sup_{\theta\in\Theta'_{\mathrm{cov}}} \mathrm{P}_\theta\big(\theta \notin B(\hat\theta, C\hat r)\big) \le \alpha_1, \\ &\sup_{\theta\in\Theta_\beta} \mathrm{P}_\theta\big(\hat r \ge cR_\varepsilon(\Theta_\beta)\big) \le \alpha_2 \qquad \forall \beta \in \mathcal{B}, \end{aligned} \tag{3}$$

possibly in asymptotic setting: put $\limsup_{\varepsilon\to 0}$ in front of both sup in (3). Ideally, $\mathcal{B}$ is "massive" and $\Theta'_{\mathrm{cov}} \supseteq \Theta_\beta$. However, in general it is impossible to construct optimal (fully) adaptive confidence set in the minimax sense: the coverage requirement in (3) does not hold even for $\Theta'_{\mathrm{cov}} = \Theta_\beta$. For the classical many normal means model, there are negative results in [1, 8, 12, 13, 18, 21]; this is also discussed in [26]. A way to achieve adaptivity is to remove the so-called *deceptive parameters* (in [29] they are called *inconvenient truths*) from Θ, that is, consider a strictly smaller set $\Theta'_{\mathrm{cov}} \subset \Theta$. Examples are: $\Theta'_{\mathrm{cov}} = \Theta_{\mathrm{ss}}$, the so-called *self-similar* parameters (related to Sobolev/Besov scales) introduced in [24] and later studied in [6, 7, 23, 27, 29]; and $\Theta'_{\mathrm{cov}} = \Theta_{\mathrm{pt}}$, a more general class of *polished tail* parameters introduced in [29]. More literature on adaptive minimax confidence sets: [4, 6, 7, 14, 16, 17, 19, 22–24, 28, 29].

In all, the above mentioned papers global minimax radial rates $R_\varepsilon(\Theta_\beta)$ [as in (3)] were studied. In this paper, we allow local radial rates as in the framework (2). When applied appropriately, the local approach is more powerful than global. Namely, suppose that a local radial rate $r_\varepsilon(\theta)$ is such that, for some uniform $c > 0$,

$$r_\varepsilon(\theta) \le cR_\varepsilon(\Theta_\beta) \qquad \text{for all } \theta \in \Theta_\beta, \beta \in \mathcal{B}. \tag{4}$$

If in addition $\Theta'_{\mathrm{cov}} \subseteq \Theta_{\mathrm{cov}}$ and $\Theta_\beta \subseteq \Theta_{\mathrm{size}}$ for all $\beta \in \mathcal{B}$, then the results of type (2) imply the results of type (3), *simultaneously for all scales* $\{\Theta_\beta, \beta \in \mathcal{B}\}$ for which (4) is satisfied. We say that the local radial rate $r_\varepsilon(\theta)$ *covers* these scales; more details are in the supplement [3].

In this paper, we apply a Bayesian methodology: namely, we first construct an empirical Bayes posterior resulting from a certain two-level hierarchical prior, then construct a DD-center by using this posterior, and finally construct a credible ball (with respect to the posterior) around this DD-center as a confidence set. For this credible ball to be also a "good" confidence set, the posterior must possess certain frequentist properties. These are the upper and lower bounds on the posterior contraction rate in terms of a given local radial rate $r_\varepsilon(\theta_0)$. The upper bound result means that the posterior contracts at θ_0 with the local rate at least $r_\varepsilon(\theta_0)$, from the P_{θ_0}-perspective [then one can also construct a DD-center $\hat{\theta}$, an estimator of θ_0 with the rate $r_\varepsilon(\theta_0)$]. The lower bound result means that the posterior concentrates around the DD-center $\hat{\theta}$ at a rate that is not faster than $r_\varepsilon(\theta_0)$.

As is discussed in [2], the method of constructing confidence sets as credible balls can actually be extended to the so-called *data dependent measures* (DDMs). Namely, one can construct an appropriate DDM and then construct confidence sets as DDM-credible sets. Posteriors and empirical Bayes posteriors are particular (natural) examples of DDMs, but in general DDM does not have to originate from a Bayesian approach.

We first derive the upper bound local result, uniform in $\theta_0 \in \ell_2$. Namely, we establish that the posterior contracts, from the P_{θ_0}-perspective, to θ_0 with the local rate $r_\varepsilon(\theta_0)$, which is the best (fastest) contraction rate over some family of rates (therefore also called *oracle rate*). The local radial rate $r_\varepsilon(\theta_0)$ satisfies (4) for typical smoothness scales such as Sobolev and analytic ellipsoids, Sobolev hyperrectangles, tail classes, certain scales of Besov classes and ℓ_p-bodies. This means that we obtained, as a consequence of our local result, the adaptive minimax contraction rate results over all these scales. An accompanying result is that, by using our posterior, a DD-center $\hat{\theta}$ can be constructed that converges to θ_0 with the local rate $r_\varepsilon(\theta_0)$, thus also yielding the panorama of the minimax adaptive estimation results over all these scales simultaneously.

Although the upper bound results are of interest on their own, our main purpose is to construct an optimal [according to the framework (2)] confidence set. Recall that we use credible balls as confidence sets. To this end, the established upper bound results imply only the size relation for an appropriately constructed credible ball $B(\hat{\theta}, C\hat{r})$ in (2) with the local radial rate $r_\varepsilon(\theta_0)$, uniformly over $\Theta_{\mathrm{size}} = \ell_2$. For the coverage relation in (2) to hold, we also need the lower bound results. It turns out that the lower bound result can be established uniformly only over some $\Theta_{\mathrm{cov}} \subset \ell_2$, which forms an actual restriction. This is in accordance with the above mentioned fact that it is impossible to construct optimal (fully) adaptive confidence set in the minimax sense. We propose a set $\Theta_{\mathrm{cov}} = \Theta_{\mathrm{eb}}$ of (nondeceptive) parameters satisfying the so-called *excessive bias restriction* and derive the lower bound uniformly over this set. Combining the obtained upper and lower bounds, we establish the optimality (2) of the credible ball $B(\hat{\theta}, C\hat{r})$ with $\Theta_{\mathrm{cov}} = \Theta_{\mathrm{eb}}$, $\Theta_{\mathrm{size}} = \ell_2$ and the local radial rate $r_\varepsilon(\theta_0)$. The class Θ_{eb} is more general than the earlier mentioned self-similar and polished tail parameters, namely, $\Theta_{\mathrm{ss}} \subseteq \Theta_{\mathrm{pt}} \subseteq \Theta_{\mathrm{eb}}$ and $\Theta_{\mathrm{eb}} \nsubseteq \Theta_{\mathrm{pt}}$ (see Section 4.3 for exact definitions). Moreover, the established (local) optimality (2) implies the global optimality (3) in the sense of adaptive minimaxity over all scales for which (4) is fulfilled, in particular for the ones considered by in [29].

The paper is organized as follows. Section 2 provides preliminaries, Section 3 contains the main results, some discussion and concluding remarks are in Section 4, and all the proofs are collected in Section 5. The elaboration on some points and some background information related to the paper are provided in the supplement [3].

2. Preliminaries.

2.1. *Some notation and conditions on ill-posedness.* By default, all summations and products are over $\mathbb{N}$, unless otherwise specified, for example, $\sum_i = \sum_{i\in\mathbb{N}}$. Introduce some notation: for $a, b \in \mathbb{R}$, $\lfloor a \rfloor = \max\{z \in \mathbb{Z} : z \le a\}$, $\Sigma(a) = \sum_{i\le a} \sigma_i^2$, $a \vee b = \max\{a, b\}$, $a \wedge b = \min\{a, b\}$; $\varphi(x, \mu, \sigma^2)$ is the $N(\mu, \sigma^2)$-density at x, $N(\mu, 0)$ means a Dirac measure at μ; the indicator function $1\{E\} = 1$ if the event E occurs and is zero, otherwise. Let $\sum_{i=k}^{n} a_i = 0$ if $n < k$ and $\sum_{i=k}^{A} a_i = \sum_{i=k}^{\lfloor A\rfloor} a_i$ for $A > 0$. If random quantities appear in a relation, then this relation should be understood in P_{θ_0}-almost sure sense, for the "true" $\theta_0 \in \Theta$. For $a_\varepsilon, b_\varepsilon > 0$, $a_\varepsilon \asymp b_\varepsilon$ means that $\frac{a_\varepsilon}{b_\varepsilon}$ is bounded away from 0 and infinity as $\varepsilon \to 0$, $\triangleq$ means "equals by definition."

We complete this subsection with conditions on σ_i^2's (or, equivalently, on κ_i^2's): for any $\rho, \tau_0 \ge 1$, $\gamma > 0$, there exist some positive K_1, $K_2 = K_2(\rho)$, $K_3 = K_3(\gamma)$, $K_4 \in (0, 1)$, $\tau > 2$ and $K_5 = K_5(\tau_0)$ such that the relations

$$
\begin{aligned}
&\text{(i)} \quad n\sigma_n^2 \le K_1 \Sigma(n), \qquad \text{(ii)} \quad \Sigma(\rho n) \le K_2(\rho)\Sigma(n), \\
&\text{(iii)} \quad \sum_n e^{-\gamma n}\Sigma(n) \le K_3(\gamma)\sigma_1^2, \\
&\text{(iv)} \quad \Sigma\big(\lfloor m/\tau \rfloor\big) \le (1 - K_4)\Sigma(m), \\
&\text{(v)} \quad l\sigma^2_{\lfloor l/\tau_0\rfloor} \ge K_5(\tau_0) \sum_{i=\lfloor l/\tau_0\rfloor+1}^{l} \sigma_i^2,
\end{aligned}
\tag{5}
$$

hold for all $n \in \mathbb{N}$, all $m \ge \tau$ and all $l \ge \tau_0$.

Although there is in principle some freedom in choosing sequence κ_i describing the ill-posedness of the problem, to avoid unnecessary technical complications, from now on we assume the so-called *mildly ill-posed* case: $\kappa_i^2 = i^{2p}$, $i \in \mathbb{N}$, for some $p \ge 0$. The mildly ill-posed case $\kappa_i^2 = i^{2p}$ satisfies (5) with $K_1 = 2p + 1$, $K_2 = (\rho + 1)^{2p+1}$, $K_3 = \frac{4(8p+4)^{2p}}{(e\gamma)^{2p+1}(e^{\gamma/2}-1)}$ (a rough bound), $K_4 = \frac{1}{2}$, τ can be any number satisfying $\tau \ge 2^{1+1/(2p+1)}$ and $K_5 = (2\tau_0)^{-2p}$; see the supplement [3] for the calculations.

2.2. *Constructing an empirical Bayes posterior.* Introduce the two-level hierarchical prior Π: for some fixed $\alpha > 0$, $K \ge 1.87$ (see Theorem 1),

$$
\theta | (\mathcal{I} = I) \sim \Pi_{I,\mu(I)} = \bigotimes_i N\big(\mu_i(I), \tau_i^2(I)\big), \qquad \mathrm{P}(\mathcal{I} = I) = \lambda_I, \tag{6}
$$

where $\mu(I) = (\mu_i(I), i \in \mathbb{N})$ with $\mu_i(I) = \mu_{I,i} 1\{i \le I\}$,

$$\tau_i^2(I) = K\sigma_i^2 1\{i \le I\}, \qquad \lambda_I = C_\alpha e^{-\alpha I}, \qquad i, I \in \mathbb{N}, \tag{7}$$

$C_\alpha = e^\alpha - 1$ (so that $\sum_I \lambda_I = 1$). The idea of introducing the truncating parameter I in the prior is to model the "effective" dimension of θ. The model (1) and the prior (6) lead to the corresponding marginal

$$\begin{aligned} \mathrm{P}_{X,\mu}(X) &= \sum_I \lambda_I \mathrm{P}_{X,I,\mu(I)}(X) \\ &= \sum_I \lambda_I \prod_i \varphi\big(X_i, \mu_i(I), \tau_i^2(I) + \sigma_i^2\big) \end{aligned}$$

and the posterior $\Pi_\mu(\cdot|X) = \sum_I \Pi_\mu(\cdot|X, \mathcal{I} = I)\Pi_\mu(\mathcal{I} = I|X)$, where

$$\Pi_\mu(\cdot|X, \mathcal{I} = I) = \bigotimes_i N\left(\frac{\tau_i^2(I)X_i^2 + \sigma_i^2\mu_i(I)}{\sigma_i^2 + \tau_i^2(I)}, \frac{\sigma_i^2\tau_i^2(I)}{\sigma_i^2 + \tau_i^2(I)}\right),$$

$$\Pi_\mu(\mathcal{I} = I|X) = \frac{\lambda_I \prod_i \varphi(X_i, \mu_i(I), \tau_i^2(I) + \sigma_i^2)}{\sum_J \lambda_J \prod_i \varphi(X_i, \mu_i(J), \tau_i^2(J) + \sigma_i^2)}.$$

It is easy to obtain the estimator of the parameter $\mu = (\mu(I), I \in \mathbb{N})$ by maximizing the marginal $\mathrm{P}_{X,\mu}(X)$ with respect to μ: $\hat{\mu} = (\hat{\mu}(I), I \in \mathbb{N})$, $\hat{\mu}_i(I) = X_i 1\{i \le I\}$, $i \in \mathbb{N}$. Now we introduce $\mathrm{P}(\cdot|X) = \Pi_{\hat{\mu}}(\cdot|X)$, $\mathrm{P}_I(\cdot|X) = \Pi_{\hat{\mu}}(\cdot|X, \mathcal{I} = I)$ and $\mathrm{P}(\mathcal{I} = I|X) = \Pi_{\hat{\mu}}(\mathcal{I} = I|X)$, the empirical Bayes counterparts of the posteriors $\Pi_\mu(\cdot|X)$, $\Pi_\mu(\cdot|X, \mathcal{I} = I)$ and $\Pi_\mu(\mathcal{I} = I|X)$, respectively. Precisely,

$$\mathrm{P}(\cdot|X) = \mathrm{P}_{K,\alpha}(\cdot|X) = \sum_I \mathrm{P}_I(\cdot|X)\mathrm{P}(\mathcal{I} = I|X), \tag{8}$$

where

$$\mathrm{P}_I(\cdot|X) = \bigotimes_i N\big(X_i(I), L\sigma_i^2 1\{i \le I\}\big), \qquad L = \frac{K}{K+1}, \tag{9}$$

$$\mathrm{P}(\mathcal{I} = I|X) = \frac{\lambda_I \prod_i \varphi(X_i, X_i(I), \tau_i^2(I) + \sigma_i^2)}{\sum_J \lambda_J \prod_i \varphi(X_i, X_i(J), \tau_i^2(J) + \sigma_i^2)}, \tag{10}$$

$X_i(I) = X_i 1\{i \le I\}$, $i, I \in \mathbb{N}$, $\tau_i^2(I)$ and λ_I are defined by (7). The quantity (10) exists as P_{θ_0}-almost sure limit of

$$\mathrm{P}_n(\mathcal{I} = I|X) = \frac{\lambda_I \prod_{i=1}^n \varphi(X_i, X_i(I), \tau_i^2(I) + \sigma_i^2)}{\sum_J \lambda_J \prod_{i=1}^n \varphi(X_i, X_i(J), \tau_i^2(J) + \sigma_i^2)}.$$

3. Main results.

3.1. *Local contraction rate*: *Upper bound*. First, we introduce the notion of local posterior contraction rate. Notice that $\mathrm{P}(\cdot|X)$ defined in (8) is a random mixture over posteriors $\mathrm{P}_I(\cdot|X)$, $I \in \mathbb{N}$. From the P_{θ_0}-perspective, each $\mathrm{P}_I(\cdot|X)$ contracts to the true θ_0 with the local rate $r(I,\theta_0)$:

$$r^2(I,\theta_0) = r_\varepsilon^2(I,\theta_0) = \sum_{i\le I} \sigma_i^2 + \sum_{i>I} \theta_{0,i}^2, \qquad I \in \mathbb{N}. \tag{11}$$

Indeed, recalling that $X(I) = (X_i 1\{i \le I\}, i \in \mathbb{N})$, we evaluate

$$\begin{aligned} \mathrm{E}_{\theta_0}\mathrm{P}_I\big(\|\theta-\theta_0\| \ge Mr(I,\theta_0)|X\big) &\le \frac{\mathrm{E}_{\theta_0}[\|X(I)-\theta_0\|^2 + L\sum_{i\le I}\sigma_i^2]}{M^2 r^2(I,\theta_0)} \\ &= \frac{2\sum_{i\le I}\sigma_i^2 + \sum_{i>I}^\infty \theta_{0,i}^2}{M^2 r^2(I,\theta_0)} \le \frac{2}{M^2}. \end{aligned} \tag{12}$$

Thus, we have the family of local rates $\mathcal{P} = \mathcal{P}(\mathbb{N}) = \{r(I,\theta_0), I \in \mathbb{N}\}$. For each $\theta_0 \in \ell_2$, there is the best choice $I_o = I_o(\theta_0) = I_o(\theta_0,\varepsilon) = \min\{J : r(J,\theta_0) = \min_{I\in\mathbb{N}} r(I,\theta_0)\}$ of parameter I, called the *oracle*, corresponding to the smallest possible rate $r(I_o,\theta_0)$ called the *oracle rate* given by

$$r^2(\theta_0) = r^2(I_o,\theta_0) = \min_{I\in\mathbb{N}} r^2(I,\theta_0) = \sum_{i\le I_o}\sigma_i^2 + \sum_{i>I_o}\theta_{0,i}^2. \tag{13}$$

Notice $r^2(\theta_0) \ge \sigma_1^2 = \varepsilon^2$ and $I_o(\theta_0) \ge 1$ for any $\theta_0 \in \ell_2$, because we minimize over $\mathbb{N}$. This is not restrictive since if the minimum were taken over $I \in \mathbb{N}\cup\{0\}$, all the below results would hold only for the oracle rate with an additive penalty term, a multiple of ε^2. This would boil down to the same resulting local rate as (13).

The following theorem establishes a nonasymptotic local upper bound for the contraction rate of the empirical Bayes posterior (8), uniformly over ℓ_2-space.

THEOREM 1 (Upper bound). *Let the posterior* $\mathrm{P}(\cdot|X)$ *and the local rate* $r(\theta)$ *be defined by* (8) *and* (13), *respectively, with* $K \ge 1.87$, $\alpha > 0$. *Then there exists a constant* $C_{\mathrm{or}} = C_{\mathrm{or}}(K,\alpha)$ *such that, for any* $\theta_0 \in \ell_2$ *and* $M > 0$,

$$\mathrm{E}_{\theta_0}\mathrm{P}\big(\|\theta-\theta_0\| \ge Mr(\theta_0)|X\big) \le \frac{C_{\mathrm{or}}}{M^2}.$$

The proof of the theorem is in Section 5. Among other implications, this upper bound result ensures the size property in (2) for the confidence ball (20) with the radial rate $r(\theta_0)$ defined by (13) and $\Theta_{\mathrm{size}} = \ell_2$. We will come back to this when proving the main result, Theorem 4.

Besides being an essential ingredient for establishing the confidence optimality (2), the above theorem is of its own interest. The results on local contraction rates are intrinsically adaptive in the sense that the oracle contraction rate

$r(\theta_0)$ is fast for "smooth" θ_0's and slow for "rough" ones. This is a stronger and more refined property than being globally adaptive. Theorem 1 implies the whole panorama of the minimax adaptive results on posterior contraction rates, simultaneously over all scales for which (4) is fulfilled; in particular, the ones considered in [29]. An elaborate discussion on this issue is provided in Section 4.1.

We can use the posterior $\mathrm{P}(\cdot|X)$ defined by (8) also for estimating the parameter θ_0; namely, define the estimator

$$\hat{\theta} = \mathrm{E}(\theta|X) = \sum_I X(I)\mathrm{P}(\mathcal{I} = I|X), \qquad X(I) = \big(X_i 1\{i \le I\}, i \in \mathbb{N}\big), \tag{14}$$

which is just the $\mathrm{P}(\cdot|X)$-expectation. This estimator satisfies the following oracle inequality (the upper bound local result for the estimation problem).

THEOREM 2 (Oracle inequality). *Let the conditions of Theorem 1 be fulfilled, $\hat{\theta}$ be defined by* (14) *and the oracle rate $r(\theta_0)$ be defined by* (13). *Then there exists a constant $C_{\mathrm{est}} = C_{\mathrm{est}}(K, \alpha)$ such that, for any $\theta_0 \in \ell_2$,*

$$\mathrm{E}_{\theta_0}\|\hat{\theta} - \theta_0\|^2 \le C_{\mathrm{est}} r^2(\theta_0).$$

The proof of this theorem is essentially contained in the proof of Theorem 1, but it is provided for completeness in Section 5. Like Theorem 1, Theorem 2 is again of a local type (now for the estimation problem) and yields therefore the whole panorama of the global (minimax) adaptive estimation results in the mildly ill-posed inverse setting, simultaneously over all scales for which (4) is fulfilled; see Section 4.1. Besides, Theorem 2 (together with Corollary 1) is used for deriving the coverage property in (2).

The local rate $r(I, \theta_0)$ defined by (11) is also the ℓ_2-risk of the projection estimator $\check{\theta}(I) = X(I)$: $\mathrm{E}_{\theta_0}\|\check{\theta}(I) - \theta_0\|^2 = r^2(I, \theta_0)$. One can regard the oracle rate (13) as the smallest possible risk over the family of (projection) estimators $\check{\Theta}(\mathbb{N}) = \{\check{\theta}(I), I \in \mathbb{N}\}$, namely $r^2(\theta_0) = \inf_{I \in \mathbb{N}} \mathrm{E}_{\theta_0}\|\check{\theta}(I) - \theta_0\|^2$. Theorem 2 claims basically that the estimator $\hat{\theta}$ given by (14) *mimics the oracle estimator* $\check{\theta}(I_o)$, which is, strictly speaking, not an estimator as it depends on the true θ_0 through $I_o = I_o(\theta_0)$.

3.2. *Local contraction rate*: *Lower bound under EBR*. Recall that our main goal is to construct a confidence set as credible ball with respect to the posterior $\mathrm{P}(\cdot|X)$ defined by (8) and establish the size and coverage properties (2) for this set. The first ingredient for achieving this goal is the local upper bound (Theorems 1 and 2) for the posterior contraction rate. Basically, the upper bound implies the size property in (2), with $\Theta_{\mathrm{size}} = \ell_2$ and the local radial rate $r(\theta_0)$ given by (13). To establish the coverage property in (2), we also need the second ingredient which is a lower bound [in terms of the local rate $r(\theta_0)$] for posterior contraction rate around the DD-center $\hat{\theta}$.

First, we derive a lower bound for the contraction rate in terms of the so-called *surrogate oracle rate* $r(\bar{I}_o, \theta_0)$, where the local rate $r(I, \theta_0)$ is defined by (11) and the *surrogate oracle* $\bar{I}_o$ is defined as follows:

$$\text{(15)} \qquad \bar{I}_o = \arg\min_I R^2(I, \theta_0), \qquad R^2(I, \theta_0) = R^2_\sigma(I, \theta_0) = I\varepsilon^2 + \sum_{i>I} \frac{\theta^2_{0,i}}{\kappa_i^2},$$

for $I \in \mathbb{N}$. The surrogate oracle rate $R(\bar{I}_o, \theta_0)$ is in fact the oracle rate for the parameter $\bar{\theta}_0 = (\theta_{0,i}/\kappa_i, i \in \mathbb{N})$ in the "direct" model $\tilde{X} = (X_i/\kappa_i, i \in \mathbb{N}) \sim \bigotimes_i N(\bar{\theta}_{0,i}, \varepsilon^2)$. Note that in the direct case $\kappa_i^2 = 1$ the surrogate oracle coincides with the usual oracle: $\bar{I}_o = I_o$.

THEOREM 3 (Small ball posterior probability). *Let the DDM* $\mathrm{P}(\cdot|X) = \mathrm{P}_{K,\alpha}(\cdot|X)$ *be given by* (8), *with parameters* $K, \alpha > 0$ *such that*

$$\text{(16)} \qquad \alpha < a(K) \triangleq \frac{1}{4} - \frac{1}{2}\log\left(\frac{K+1}{2}\right).$$

Then there exists $C_{\mathrm{sb}} = C_{\mathrm{sb}}(K, \alpha) > 0$ *such that, for any* $\theta_0 \in \ell_2$, *any DD-center* $\tilde{\theta} = \tilde{\theta}(X)$ *and any* $\delta \in (0, \delta_{\mathrm{sb}}]$ *with* $\delta_{\mathrm{sb}} = 1 \wedge (\sqrt{\frac{K(2p+1)}{K+1}} (\frac{a(K)-\alpha}{4ea(K)})^{p+\frac{1}{2}})$,

$$\text{(17)} \qquad \mathrm{E}_{\theta_0}\mathrm{P}\big(\|\theta - \tilde{\theta}\| \le \delta \Sigma^{1/2}(\bar{I}_o)|X\big) \le C_{\mathrm{sb}}\delta\big[\log(\delta^{-1})\big]^{p+1/2},$$

where $\Sigma(\bar{I}_o) = \sum_{i \le \bar{I}_o} \sigma_i^2$ *and* $\bar{I}_o = \bar{I}_o(\theta_0)$ *is defined by* (15).

Notice that although the above lower bound holds uniformly in $\theta_0 \in \ell_2$ and the right-hand side of (17) does not depend on θ_0, the effective lower bound for the contraction rate is the quantity $\Sigma^{1/2}(\bar{I}_o)$ [the variance related term of the surrogate oracle rate $r(\bar{I}_o, \theta_0)$], and not the oracle rate $r(\theta_0)$ as we would like to have. We can formally apply this theorem with the oracle surrogate rate $r(\bar{I}_o, \theta_0)$ instead of $\Sigma^{1/2}(\bar{I}_o)$: for any $\delta \in (0, \delta_{\mathrm{sb}}(1 + t(\theta_0))^{-1/2}]$,

$$\mathrm{E}_{\theta_0}\mathrm{P}\big(\|\theta - \tilde{\theta}\| \le \delta r(\bar{I}_o, \theta_0)|X\big) \le C_{\mathrm{sb}}(1 + t(\theta_0))^{1/2}\delta\big[\log(\delta^{-1})\big]^{p+1/2},$$

where $t(\theta_0) = \frac{\sum_{i>\bar{I}_o}\theta^2_{0,i}}{\sum_{i\le\bar{I}_o}\sigma_i^2}$ is the ratio of the bias term of the surrogate oracle rate to the variance term. But then the right-hand side of the last relation is uniform only over the set on which the ratio $t(\theta_0)$ is bounded, and this is where a condition on parameter θ_0 comes into play.

This motivates introducing the *excessive bias restriction* (EBR): $\theta_0 \in \Theta_{\mathrm{eb}}(\tau)$ for $\tau > 0$, where, with $\bar{I}_o = \bar{I}_o(\theta)$ defined by (15),

$$\text{(18)} \qquad \Theta_{\mathrm{eb}} = \Theta_{\mathrm{eb}}(\tau) = \Theta_{\mathrm{eb}}(\tau, \varepsilon) = \Big\{\theta \in \ell_2 : \sum_{i>\bar{I}_o} \theta_i^2 \le \tau \sum_{i\le\bar{I}_o} \sigma_i^2\Big\}.$$

An elaborate discussion on EBR is provided in Section 4.3. Here, we just mention that EBR describes a more general set of nondeceptive parameters than the earlier mentioned sets of *self-similar* and *polished tail* parameters.

If $\theta_0 \in \Theta_{\mathrm{eb}}(\tau)$, then $r^2(\theta_0) = r^2(I_o, \theta_0) \leq r^2(\bar{I}_o, \theta_0) \leq (1+\tau)\Sigma(\bar{I}_o)$. This yields the corollary describing the lower bound for the posterior contraction rate, now in terms of the oracle rate $r(\theta_0)$ and uniformly over Θ_{eb}.

COROLLARY 1 (Lower bound under EBR). *Let the conditions of Theorem 3 be satisfied. Then for any DD-center $\tilde{\theta} = \tilde{\theta}(X)$, any $\tau > 0$ and any $\delta \in (0, \delta_{\mathrm{eb}}]$ with $\delta_{\mathrm{eb}} = \delta_{\mathrm{sb}}(1+\tau)^{-1/2}$,*

$$\sup_{\theta_0 \in \Theta_{\mathrm{eb}}(\tau)} \mathrm{E}_{\theta_0} \mathrm{P}\big(\|\theta - \tilde{\theta}\| \leq \delta r(\theta_0) | X\big) \leq C_{\mathrm{eb}} \delta \big[\log(\delta^{-1})\big]^{p+1/2},$$

where $C_{\mathrm{eb}} = C_{\mathrm{eb}}(K, \alpha, \tau) = C_{\mathrm{sb}}\sqrt{1+\tau}$, δ_{sb} and C_{sb} are from Theorem 3.

This lower bound, together with Theorem 2, ensures the coverage relation in (2) uniformly over $\Theta_{\mathrm{cov}} = \Theta_{\mathrm{eb}}$, see the next subsection.

3.3. *The main result*: *Confidence ball under EBR*. In this subsection, we establish the main result of the paper. Let the posterior $\mathrm{P}(\cdot|X) = \mathrm{P}_{K,\alpha}(\cdot|X)$ be given by (8), with some constants $K, \alpha > 0$ (fixed throughout this subsection) such that the conditions of Theorem 1 and 3 are fulfilled.

For some fixed $\kappa \in (0, 1)$ (e.g., fix $\kappa = \frac{1}{2}$) and the DD-center $\hat{\theta} = \hat{\theta}(X)$ given by (14), define the DD-radius

$$\hat{r} = \hat{r}(\kappa, X, \hat{\theta}) = \inf\big\{r : \mathrm{P}\big(\|\theta - \hat{\theta}\| \leq r | X\big) \geq 1 - \kappa\big\} \tag{19}$$

and then, for $M > 0$, construct the confidence ball

$$B(\hat{\theta}, M\hat{r}) = \big\{\theta \in \ell_2 : \|\theta - \hat{\theta}\| \leq M\hat{r}\big\}. \tag{20}$$

For $M = 1$, (20) is the smallest credible ball around $\hat{\theta}$ of level $1 - \kappa$.

The obtained upper and lower bounds for the posterior contraction rate, Theorems 1, 2 and Corollary 1, can now be used to establish the coverage and size properties (2) for the ball $B(\hat{\theta}, M\hat{r})$ defined by (20) with $\Theta_{\mathrm{cov}} = \Theta_{\mathrm{eb}}$, $\Theta_{\mathrm{size}} = \ell_2$ and the radial rate $r(\theta_0)$ defined by (13). The inflating factor M (not depending on θ_0) is intended to provide the coverage property. This is exactly what the next theorem, the main result of the paper, claims.

THEOREM 4 (Confidence optimality under EBR). *Let the confidence ball $B(\hat{\theta}, M\hat{r})$ be defined by (20) and the local radial rate $r(\theta_0)$ be defined by (13). Then for any $\tau > 0$ there exist $M_0 = M_0(\tau) > 0$ and $c_0 > 0$ such that*

$$\sup_{\theta_0 \in \Theta_{\mathrm{eb}}(\tau)} \mathrm{P}_{\theta_0}\big(\theta_0 \notin B(\hat{\theta}, M\hat{r})\big) \leq \varphi_1(M), \qquad \sup_{\theta_0 \in \ell_2} \mathrm{P}_{\theta_0}\big(\hat{r} \geq c r(\theta_0)\big) \leq \varphi_2(c),$$

where $\varphi_1(M)$ and $\varphi_2(c)$ are some monotonically decreasing to zero functions (given in the proof of the theorem) as $M, c \to \infty$, for $M \geq M_0$ and $c \geq c_0$.

For any $\alpha_1, \alpha_2 \in (0,1)$, by taking large enough $M_0 = M_0(\alpha_1, \tau)$ and $c_0 = c_0(\alpha_2)$, we can ensure that $\varphi_1(M) \le \alpha_1$ and $\varphi_2(c) \le \alpha_2$ for all $M \ge M_0$ and $c \ge c_0$.

4. Discussion and concluding remarks.

4.1. *Local results versus global ones.* Let us elucidate the potential strength of local results as compared to global ones.

We start with the posterior contraction rate. To characterize the quality of Bayesian procedures, the notion of posterior contraction rate was first introduced and studied in [15]. Typically in the literature, contraction rate is related to the (global) minimax rate $R(\Theta_\beta) = R_\varepsilon(\Theta_\beta)$ over a certain set Θ_β. The optimality of Bayesian procedures is then understood in the sense of adaptive minimax posterior contraction rate: given a prior (knowledge of β is not used in the prior), the resulting posterior contracts, from the P_{θ_0}-perspective, to the "true" $\theta_0 \in \Theta_\beta$ with the minimax rate $R(\Theta_\beta)$.

For a scale $\Theta(\mathcal{B}) = \{\Theta_\beta, \beta \in \mathcal{B}\}$, let $\{R(\Theta_\beta), \beta \in \mathcal{B}\}$ be the family of the pertaining minimax rates. Suppose (4) is fulfilled for the local rate $r(\theta_0)$ defined by (13) and $\{R(\Theta_\beta), \beta \in \mathcal{B}\}$. Then, in view of (4), Theorem 1 entails that the posterior $\mathrm{P}(\cdot|X)$ must also contract to θ_0 with (at least) the minimax rate $R(\Theta_\beta)$ uniformly in $\theta_0 \in \Theta_\beta$ for each $\beta \in \mathcal{B}$. Thus, the adaptive [over the scale $\Theta(\mathcal{B})$] minimax contraction rate result for $\mathrm{P}(\cdot|X)$ follows immediately. Foremost, Theorem 1 implies adaptive minimax results on the posterior contraction rates, *simultaneously for all scales* for which (4) is fulfilled. Theorem 2 does the same for the estimation problem, and Theorem 4 for the uncertainty quantification problem (3).

Let us consider general ellipsoids and hyperrectangles

$$\mathcal{E}(a) = \Big\{\theta \in \ell_2 : \sum_i \Big(\frac{\theta_i}{a_i}\Big)^2 \le 1\Big\}, \qquad \mathcal{H}(a) = \{\theta \in \ell_2 : |\theta_i| \le a_i, i \in \mathbb{N}\}, \tag{21}$$

where $a = (a_i, i \in \mathbb{N})$ is a nonincreasing sequence of numbers in $[0, +\infty]$ which converge to 0 as $i \to \infty$, $a_1 \ge c_1\varepsilon$ for some $c_1 > 0$. Here, we adopt the conventions $0/0 = 0$ and $x/(+\infty) = 0$ for $x \in \mathbb{R}$. Let $R^2(\Theta) = \inf_{\hat\theta} \sup_{\theta\in\Theta} \mathrm{E}_\theta \|\hat\theta - \theta\|^2$ denote the (quadratic) minimax rate over a set Θ, where the infimum is taken over all possible estimators $\hat\theta = \hat\theta(X)$, measurable functions of the data X. One can show (see the supplement [3]) that

$$\begin{aligned} \sup_{\theta_0 \in \mathcal{E}(a)} r^2(\theta_0) &\le \inf_I \Big\{\sum_{i\le I} \sigma_i^2 + a_{I+1}^2\Big\} \le (2\pi)^2 R^2(\mathcal{E}(a)), \\ \sup_{\theta_0 \in \mathcal{H}(a)} r^2(\theta_0) &\le \inf_I \Big\{\sum_{i\le I} \sigma_i^2 + \sum_{i>I} a_i^2\Big\} \le \frac{5}{2} R^2(\mathcal{H}(a)). \end{aligned} \tag{22}$$

In (22), one can put a tighter constant 4.44 instead of $(2\pi)^2$ in the direct case, which possibly holds for the ill-posed case as well; see the supplement [3].

In view of (22) and since $R^2(\mathcal{E}(a)) \le \inf_I \sup_{\theta\in\mathcal{E}(a)} \mathrm{E}_\theta \|X(I) - \theta\|^2 = \inf_I\{\sum_{i\le I}\sigma_i^2 + a_{I+1}^2\}$ and $R^2(\mathcal{H}(a)) \le \inf_I \sup_{\theta\in\mathcal{H}(a)} \mathrm{E}_\theta \|X(I) - \theta\|^2 = \inf_I\{\sum_{i\le I}\sigma_i^2 + \sum_{i>I} a_i^2\}$, we derive the minimax rates for $\mathcal{E}(a)$ and $\mathcal{H}(a)$:

$$R^2(\mathcal{E}(a)) \asymp \inf_I\Big\{\sum_{i\le I}\sigma_i^2 + a_{I+1}^2\Big\}, \qquad R^2(\mathcal{H}(a)) \asymp \inf_I\Big\{\sum_{i\le I}\sigma_i^2 + \sum_{i>I} a_i^2\Big\}. \tag{23}$$

Let the constants K, α be fixed and satisfy the conditions of Theorems 1 and 4. In view of (22), the ellipsoids $\mathcal{E}(a)$ and hyperrectangles $\mathcal{H}(a)$ are particular examples of scales for which (4) holds. Hence, Theorems 1, 2 and 4 imply that for some C_{or}, C_{est}, $C = C(\alpha_1, \tau)$ and $c = c(\alpha_2)$,

$$\sup_{\theta_0\in\Theta(a)} \mathrm{E}_{\theta_0}\mathrm{P}\big(\|\theta - \theta_0\| \ge MR\big(\Theta(a)\big)|X\big) \le \frac{C_{\mathrm{or}}}{M^2}, \tag{24}$$

$$\sup_{\theta_0\in\Theta(a)} \mathrm{E}_{\theta_0}\|\hat\theta - \theta_0\|^2 \le C_{\mathrm{est}} R^2\big(\Theta(a)\big), \tag{25}$$

$$\begin{aligned} &\sup_{\theta_0\in\Theta_{\mathrm{eb}}(\tau)} \mathrm{P}_{\theta_0}\big(\theta_0 \notin B(\hat\theta, C\hat r)\big) \le \alpha_1, \\ &\sup_{\theta_0\in\Theta(a)} \mathrm{P}_{\theta_0}\big(\hat r \ge cR_\varepsilon\big(\Theta(a)\big)\big) \le \alpha_2, \end{aligned} \tag{26}$$

where the DD-center $\hat\theta$ is defined by (14), the DD-radius $\hat r$ is defined by (19), and $\Theta(a)$ is either $\mathcal{E}(a)$ or $\mathcal{H}(a)$, for any *unknown* a.

In particular, we obtain the adaptive minimax results (for all the three problems: posterior contraction, estimation and uncertainty quantification) for the four particular scales considered in [29]: (for $Q, \beta, c, d > 0$, $N_0 \in \mathbb{N}$) Sobolev ellipsoids $\mathcal{E}_S = \mathcal{E}_S(\beta, Q)$, Sobolev hyperrectangles $\mathcal{H}_S = \mathcal{H}_S(\beta, Q)$, and the two so-called *supersmooth* scales of analytic ellipsoids $\mathcal{E}_A = \mathcal{E}_A(c, d, Q)$ and parametric hyperrectangles $\mathcal{H}_P = \mathcal{H}_P(N_0, Q)$. These scales are defined as follows: with $\mathcal{E}(a)$ and $\mathcal{H}(a)$ given by (21),

$$\begin{aligned} \mathcal{H}_S &= \mathcal{H}(a) \qquad \text{with } a_i^2 = Qi^{-(2\beta+1)}, \\ \mathcal{E}_S &= \mathcal{E}(a) \qquad \text{with } a_i^2 = Qi^{-2\beta}, \\ \mathcal{E}_A &= \mathcal{E}(a) \qquad \text{with } a_i^2 = Qe^{-ci^d}, \\ \mathcal{H}_P &= \mathcal{H}(a) \qquad \text{with } a_i^2 = Q1\{i \le N_0\}. \end{aligned} \tag{27}$$

By using (23), it is easy to compute the corresponding minimax rates under the asymptotic regime $\varepsilon \to 0$ (or, as in [29], $n \to \infty$ with $n = \varepsilon^{-2}$):

$$R^2(\mathcal{E}_S) \asymp \varepsilon^{\frac{4\beta}{2\beta+2p+1}} = n^{-\frac{2\beta}{2\beta+2p+1}}, \qquad R^2(\mathcal{H}_S) \asymp \varepsilon^{\frac{4\beta}{2\beta+2p+1}} = n^{-\frac{2\beta}{2\beta+2p+1}}, \tag{28}$$

$$R^2(\mathcal{E}_A) \asymp \varepsilon^2\big(\log\varepsilon^{-1}\big)^{\frac{2p+1}{d}} = n^{-1}(\log n)^{\frac{2p+1}{d}}, \qquad R^2(\mathcal{H}_P) \asymp \varepsilon^2 = n^{-1}. \tag{29}$$

The relation to the results of [29] is discussed in Section 4.2.

We emphasize that the scope of our local results, Theorems 1, 2 and 4, extends further than just these four specific scales, in fact even beyond general ellipsoids $\mathcal{E}(a)$ and hyperrectangles $\mathcal{H}(a)$. Basically, our local results imply adaptive minimax results simultaneously for all scales for which (4) is fulfilled. Besides ellipsoids and hyperrectangles, (4) is satisfied also for certain scales of Besov classes, ℓ_p-bodies, tail classes. In the supplement [3], we also consider the more general situation when the local oracle results over one family of rates imply the local oracle results over another family of rates.

4.2. *Relation to the minimax results of* [29]. For the mildly ill-posed inverse signal-in-white-noise model (1), an intriguing paper [29] deals with a certain Sobolev type family of priors, indexed by a smoothness parameter. A certain empirical Bayes posterior with respect to the smoothness parameter is proposed in [29]. This posterior is then used to construct a credible ball whose coverage and size properties are studied.

The main results of [29] are the asymptotic (in our notation: as $\varepsilon = n^{-1/2} \to 0$) versions of the minimax framework (3) with $\Theta'_{\text{cov}} = \Theta_{\text{pt}}$ (the *polished tail* class Θ_{pt} defined in Section 4.3), and the four scales: the two Sobolev type scales of ellipsoids $\mathcal{E}_S$ and hyperrectangles $\mathcal{H}_S$, and the two so-called *supersmooth* scales of analytic ellipsoids $\mathcal{E}_A$ and parametric hyperrectangles $\mathcal{H}_P$ defined by (27).

The posterior proposed in [29] is well suited to model Sobolev type scales as it delivers the optimal rates $R(\mathcal{E}_S)$ and $R(\mathcal{H}_S)$ [given by (28)] for Sobolev hyperrectangles and ellipsoids. But for the two supersmooth scales $\mathcal{E}_A$ and $\mathcal{H}_P$, only suboptimal rates $R_{\text{sub}}(\mathcal{E}_A)$ and $R_{\text{sub}}(\mathcal{H}_P)$ are derived in [29]:

$$R^2_{\text{sub}}(\mathcal{E}_A) = n^{-1}(\log n)^{(p+1/2)\sqrt{\log n}} \gg R^2(\mathcal{E}_A) = n^{-1}(\log n)^{(2p+1)/d},$$
$$R^2_{\text{sub}}(\mathcal{H}_P) = n^{-1}e^{(3p+3/2)\sqrt{\log N_0}\sqrt{\log n}} \gg R^2(\mathcal{H}_P) = n^{-1},$$

where $R^2(\mathcal{E}_A)$ and $R^2(\mathcal{H}_P)$ are given by (29).

If we relate the global results of [29] to our local results for the posterior $\mathrm{P}(\cdot|X)$ defined by (8), we see that, according to (27), the four above mentioned scales from [29] are particular examples of ellipsoids $\mathcal{E}(a)$ and hyperrectangles $\mathcal{H}(a)$, with specific choices of sequence a. Hence, the adaptive minimax results (for all the three problems: posterior contraction, estimation and uncertainty quantification) for all the four scales follow immediately from (24)–(26), by taking $\Theta(a)$ to be equal to $\mathcal{E}_S, \mathcal{H}_S, \mathcal{E}_A, \mathcal{H}_P$.

Notice that we improve on the results of [29] for the supersmooth scales $\mathcal{E}_A$ and $\mathcal{H}_P$ as we derive the optimal rates $R(\mathcal{E}_A)$ and $R(\mathcal{H}_P)$, in contrast to the suboptimal rates $R_{\text{sub}}(\mathcal{E}_A)$ and $R_{\text{sub}}(\mathcal{H}_P)$ obtained in [29]. Besides, the coverage relation in (26) is slightly stronger than the corresponding claim from [29] because $\Theta_{\text{pt}} \subseteq \Theta_{\text{eb}}$ as we show in the next subsection. Notice also that the parametric class $\mathcal{H}_P$ automatically satisfies EBR, that is, $\mathcal{H}_P \subseteq \Theta_{\text{eb}}$.

4.3. *Excessive bias restriction* (*EBR*). Unlike the size property, the coverage property in Theorem 4 does not hold uniformly over ℓ_2, but only over Θ_{eb}. This is in agreement with the fact mentioned in the Introduction that in general it is impossible to construct optimal fully adaptive confidence set with a prescribed high coverage probability. The intuition is that there are the so-called "deceptive" parameters θ_0 that "trick" the posterior $\mathrm{P}(\cdot|X)$ in the sense that the DD-radius of a $\mathrm{P}(\cdot|X)$-credible ball becomes overoptimistic, that is of a smaller order than the oracle radial rate $r(\theta_0)$. Then the resulting credible ball misses the true θ_0 with a high P_{θ_0}-probability, that is, the coverage probability is too small.

A way to fix this problem is to remove a set (preferably, minimal) of the deceptive parameters and derive the coverage relation in (2) for the remaining set Θ_{cov} of the nondeceptive parameters. In this paper, such a set $\Theta_{\mathrm{cov}} = \Theta_{\mathrm{eb}}$ defined by (18) emerged formally from the technical condition EBR for obtaining a uniform lower bound for the posterior contraction rate. An informal intuition behind the EBR can be as follows: the posterior $\mathrm{P}(\cdot|X)$ can extract information about the variance term of the oracle rate from the data, but not about the bias term, and the EBR allows to control the bias term via the variance term.

As is mentioned in the Introduction, the first example of nondeceptive parameters is the set Θ_{ss} of the so-called *self-similar* (SS) parameters, studied by many authors in various settings and models. A somewhat restrictive feature of the self-similarity property is that it is linked to the Sobolev (Besov) smoothness scale. In [29], a more general condition is introduced that is not linked to particular smoothness scales, the *polished tail* (PT) condition: for some $L_0 > 0$ ($L_0 \ge 1$ for Θ_{pt} to be not empty), $N_0 \in \mathbb{N}$ and $\rho_0 \ge 2$,

$$\Theta_{\mathrm{pt}} = \Theta_{\mathrm{pt}}(L_0, N_0, \rho_0) = \left\{\theta \in \ell_2 : \sum_{i=N}^{\infty} \theta_i^2 \le L_0 \sum_{i=N}^{\rho_0 N} \theta_i^2, \forall N \ge N_0\right\}.$$

In [29], it is shown that $\Theta_{\mathrm{ss}} \subseteq \Theta_{\mathrm{pt}}$, that is, PT is more general than SS.

Let us show that EBR is in turn more general than PT: $\Theta_{\mathrm{pt}} \subseteq \Theta_{\mathrm{eb}}$, which means that for any $L_0 \ge 1$, $N_0 \in \mathbb{N}$ and $\rho_0 \ge 2$, there exists a $\tau > 0$ such that $\Theta_{\mathrm{pt}}(L_0, N_0, \rho_0) \subseteq \Theta_{\mathrm{eb}}(\tau)$. From (15), it follows that for any $I > \bar{I}_o$, $\sum_{i=\bar{I}_o+1}^{I} \frac{\theta_i^2}{\sigma_i^2} \le I - \bar{I}_o$. Besides, by condition (i) in (5), $(n-l)\sigma_n^2 \le K_1 \sum_{i=1}^n \sigma_i^2$ for all $n \ge l \ge 1$. Using the last two relations and property (ii) from (5), we obtain for any $\theta \in \Theta_{\mathrm{pt}}(L_0, N_0, \rho_0)$ that

$$\begin{aligned}
\sum_{i>\bar{I}_o} \theta_i^2 &= \sum_{i=\bar{I}_o+1}^{N_0\bar{I}_o-1} \theta_i^2 + \sum_{i=N_0\bar{I}_o}^{\infty} \theta_i^2 \le \sum_{i=\bar{I}_o+1}^{N_0\bar{I}_o-1} \theta_i^2 + L_0 \sum_{i=N_0\bar{I}_o}^{\rho_0 N_0 \bar{I}_o} \theta_i^2 \\
&\le L_0 \sigma_{\rho_0 N_0 \bar{I}_o}^2 \sum_{i=\bar{I}_o+1}^{\rho_0 N_0 \bar{I}_o} \frac{\theta_i^2}{\sigma_i^2} \le L_0 \sigma_{\rho_0 N_0 \bar{I}_o}^2 (\rho_0 N_0 \bar{I}_o - \bar{I}_o) \\
&\le L_0 K_1 \sum_{i=1}^{\rho_0 N_0 \bar{I}_o} \sigma_i^2 \le L_0 K_1 K_2(\rho_0 N_0) \sum_{i=1}^{\bar{I}_o} \sigma_i^2,
\end{aligned}$$

so that $\Theta_{\rm pt}(L_0, N_0, \rho_0) \subseteq \Theta_{\rm eb}(L_0 K_1 K_2(\rho_0 N_0))$ for any $N_0 \geq 1$.

Note that in principle $\Theta_{\rm eb}$ also depends on ε. We can introduce the uniform (in ε) version of EBR:

$$\bar{\Theta}_{\rm eb}(\tau, \varepsilon_0) = \{\theta \in \Theta_{\rm eb}(\tau, \varepsilon) \text{ for all } \varepsilon \in (0, \varepsilon_0]\} = \bigcap_{\varepsilon \in (0,\varepsilon_0]} \Theta_{\rm eb}(\tau, \varepsilon).$$

This set is still large enough to contain $\Theta_{\rm pt}$ as $\Theta_{\rm pt}(L_0, N_0, \rho_0) \subseteq \Theta_{\rm eb}(L_0 K_1 K_2(\rho_0 N_0), \varepsilon)$ for any $\varepsilon > 0$. We do not consider $\bar{\Theta}_{\rm eb}(\tau, \varepsilon_0)$ and $\Theta_{\rm eb}(\tau, \varepsilon)$ separately and always use the latter notation $\Theta_{\rm eb}(\tau)$ for both, with the understanding that whenever one needs the uniform version, one can think of $\Theta_{\rm eb}(\tau)$ as $\bar{\Theta}_{\rm eb}(\tau, \varepsilon_0)$, as all the assertions in this paper hold also for the uniform version of EBR.

Summarizing the relations between three types of conditions describing the non-deceptive parameters introduced above, $\Theta_{\rm ss} \subseteq \Theta_{\rm pt} \subseteq \Theta_{\rm eb}$. Besides, it is easy to show that $\Theta_{\rm eb} \nsubseteq \Theta_{\rm pt}$, which (being the negation of $\Theta_{\rm eb} \subseteq \Theta_{\rm pt}$) exactly means that there exists a $\tau > 0$ such that for any $L_0 \geq 1$, $N_0 \in \mathbb{N}$ and $\rho_0 \geq 2$, $\Theta_{\rm eb}(\tau) \nsubseteq \Theta_{\rm pt}(L_0, N_0, \rho_0)$.

In fact, even a stronger property holds, namely,

$$\Theta_{\rm eb}(1) \nsubseteq \bigcup_{(L_0,N_0,\rho_0)\in\bar{S}} \Theta_{\rm pt}(L_0, N_0, \rho_0), \qquad \bar{S} = [1, +\infty) \times \mathbb{N} \times [2, +\infty).$$

Indeed, assume the direct case $\kappa_i = 1$, $i \in \mathbb{N}$, so that $\sigma_i^2 = \varepsilon^2$ and $\bar{I}_o = I_o$. Let $\varepsilon \in (0, 1]$. Next, let $(\rho_j, j \in \mathbb{N})$ and $(n_j, j \in \mathbb{N})$ be such that $1 \leq \rho_j \uparrow \infty$ (i.e. monotonically increasing to infinity as $j \to \infty$), $n_1 \geq 2$ and $n_{j+1} \geq \rho_j^2 n_j$ for all $j \in \mathbb{N}$. Consider $\bar{\theta} = (\bar{\theta}_i, i \in \mathbb{N})$, where $\bar{\theta}_1^2 = 2$, and for $i \geq 2$

$$\bar{\theta}_i^2 = \begin{cases} 0, & n_j \leq i \leq \rho_j n_j,\ j \in \mathbb{N}, \\ \varepsilon^2 2^{-i}, & \text{otherwise.} \end{cases}$$

The idea to insert expanding zero "gaps" in the sequence $\bar{\theta}$ is borrowed from the example of Theorem 3.1 in [29]. Notice that there is enough room for infinitely many nonzero coordinates $\bar{\theta}_i$'s: $\bar{\theta}_i^2 > 0$ for $\rho_j n_j < i < \rho_j^2 n_j$, $j \in \mathbb{N}$. Because of the expanding gaps, $\bar{\theta}$ does not satisfy the PT condition for any $(L_0, N_0, \rho_0) \in \bar{S}$. On the other hand, it is easy to see that $I_o(\bar{\theta}) = 1$ and $\sum_{i>I_o} \hat{\theta}_i^2 \leq \varepsilon^2 \sum_i 2^{-i} \leq \varepsilon^2 = I_o \varepsilon^2$, that is, $\bar{\theta} \in \Theta_{\rm eb}(1)$. It follows that $\Theta_{\rm eb}(1) \nsubseteq \bigcup_{(L_0,N_0,\rho_0)\in\bar{S}} \Theta_{\rm pt}(L_0, N_0, \rho_0)$.

Thus, the EBR is the most general condition among $\Theta_{\rm ss}$, $\Theta_{\rm pt}$ and $\Theta_{\rm eb}$. As to the question how big (or "typical") that set $\Theta_{\rm eb}$ is, [29] gives three types of arguments for the PT-parameters: topological, minimax and Bayesian. Since $\Theta_{\rm pt} \subseteq \Theta_{\rm eb}$, the same arguments certainly apply to $\Theta_{\rm eb}$; see [29] for more details on this.

4.4. *Concluding remarks*. *Data dependent measures*. We construct confidence sets as credible balls with respect to the obtained empirical Bayes posterior $\mathrm{P}(\cdot|X)$

defined by (8). However, we can look at this method from a broader perspective. Namely, we can construct a so-called *data dependent measure* (DDM), and then construct confidence sets as *DDM-credible* sets. This notion is properly introduced and discussed in [2]. The DDM-framework gives more modeling flexibility as one can use different ingredient in constructing DDMs. Different choices for $\mathrm{P}_I(\cdot|X)$ and $\mathrm{P}(\mathcal{I}=I|X)$ in (8) are possible, not necessarily coming from the (same) Bayesian approach.

For example, if in (8) instead of $\mathrm{P}(\mathcal{I}=I|X)$ given by (10) we use

$$\Pi'(\mathcal{I}=I|X)=\frac{\lambda_I\prod_i\varphi(X_i,0,\tau_i^2(I)+\sigma_i^2)}{\sum_J\lambda_J\prod_i\varphi(X_i,0,\tau_i^2(J)+\sigma_i^2)},\qquad I\in\mathbb{N},$$

the main results will still hold, with slightly different constants in the proof. More on this can be found in [2] and in the supplement [3].

Alternative empirical Bayes posterior. We can apply empirical Bayes approach to the parameter I, leading to yet another empirical Bayes posterior

$$\hat{\mathrm{P}}(\cdot|X)=\mathrm{P}_{\hat{I}}(\cdot|X)\qquad\text{with } \hat{I}=\min\Big\{\arg\max_{I\in\mathbb{N}}\mathrm{P}(\mathcal{I}=I|X)\Big\},\tag{30}$$

where $\mathrm{P}_I(\cdot|X)$ and $\mathrm{P}(\mathcal{I}=I|X)$ are defined by respectively (9) and (10).

For $\hat{\mathrm{P}}(\cdot|X)$ exactly the same results hold as for $\mathrm{P}(\cdot|X)$ defined by (8). Even the proofs are almost identical. Indeed, by the definition of $\hat{I}$, we derive that, for any $I, I_0\in\mathbb{N}$ and any $h\in[0,1]$,

$$\mathrm{P}_{\theta_0}(\hat{I}=I)\le\mathrm{P}_{\theta_0}\biggl(\frac{\mathrm{P}(\mathcal{I}=I|X)}{\mathrm{P}(\mathcal{I}=I_0|X)}\ge 1\biggr)\le\mathrm{E}_{\theta_0}\biggl[\frac{\mathrm{P}(\mathcal{I}=I|X)}{\mathrm{P}(\mathcal{I}=I_0|X)}\biggr]^h,$$

which yields the analogue of (31). From this point on, the proof of the properties of the posterior $\hat{\mathrm{P}}(\cdot|X)$ proceeds exactly in the same way as the proof for the posterior $\mathrm{P}(\cdot|X)$ defined by (8), with the only difference that everywhere (in the claims and in the proofs), $1\{\hat{I}=I\}$ is substituted instead of $\mathrm{P}(\mathcal{I}=I|X)$ and $\mathrm{P}_{\theta_0}(\hat{I}=I)$ is substituted instead of $\mathrm{E}_{\theta_0}\mathrm{P}(\mathcal{I}=I|X)$.

A connection of the posterior $\hat{\mathrm{P}}(\cdot|X)$ to *penalized estimators* is discussed in the supplement [3].

Range for constant K. The condition $\alpha<a(K)$ in Theorem 3 limits room for choosing constants $K,\alpha>0$, because $a(K)>0$ only for $K\in(0,2e^{1/2}-1)$. One can choose, for example, $K=2$ and $\alpha=0.04$. Theorem 1 requires the condition $K\ge 1.87$. Formally, the final range of allowable K's for the main result, Theorem 4, is $K\in[1.87,2.29]\subset[1.87,2e^{1/2}-1)$.

The condition $K\ge 1.87$ can slightly be relaxed, but some positive lower bound is unavoidable. One can interpret this as a requirement for sufficient prior variability in (7) of nonzero coordinates when putting a prior on θ. Interestingly, this lower bound requirement on K corresponds to the condition on the penalty constant in the penalization method; see the supplement [3].

The condition $\alpha<a(K)$ in Theorem 3 is, however, an artifact of the proof technique; actually, the results hold for any $\alpha>0$. We leave the results in their present form as the accurate proof for general $\alpha > 0$ will become significantly longer whereas we want to keep the proof as concise as possible.

Alternative DD-center and confidence ball. Instead of the DD-center $\tilde{\theta} = \tilde{\theta}(X)$ given by (14), we can actually use any other estimator that satisfy the oracle inequality in Theorem 2. For example, similar oracle inequality result has been obtained in [11] for the estimator based on the risk hull minimization method. In that paper, the oracle rate has an extra penalty term but the multiplicative constant is very tight.

Here, we construct yet another alternative DD-center by using the posterior $\mathrm{P}(\cdot|X)$. For a $\kappa^* \in (0, 1/2)$, define first

$$\hat{r}^* = \hat{r}^*(\kappa^*) = \inf\{r : \mathrm{P}(\|\theta - \theta'\| \le r|X) \ge 1 - \kappa^* \text{ for some } \theta' \in \Theta\}.$$

This is the smallest possible radius of credible ball of level $1 - \kappa^*$. Next, for some $\varsigma > 0$, take any (measurable function of data X) $\check{\theta} \in \Theta$ that satisfies

$$\mathrm{P}(\theta : \|\theta - \check{\theta}\| \le (1 + \varsigma)\hat{r}^*|X) \ge 1 - \kappa^*.$$

In words, $\check{\theta} = \check{\theta}(\kappa^*, \varsigma)$ is the center of the ball of nearly the smallest radius subject to the constraint that its $\mathrm{P}(\cdot|X)$-mass is at least $1 - \kappa^*$.

One can show that Theorem 4 holds also for the confidence ball $B(\check{\theta}, M\check{r})$, with this new DD-center $\check{\theta}$, where $\check{r} = \hat{r}(1/2, X, \check{\theta})$ and $\hat{r}$ is defined by (19).

5. Proofs of theorems.

5.1. *Proof of Theorem* 1. We prove Theorem 1 in several steps.

Step 1: *Bounds for* $\mathrm{E}_{\theta_0}\mathrm{P}(\mathcal{I} = I|X)$. For any $I, I_0 \in \mathbb{N}$ and any $h \in [0, 1]$, we have

$$\mathrm{E}_{\theta_0}\mathrm{P}(\mathcal{I} = I|X) \le \mathrm{E}_{\theta_0}\left[\frac{\lambda_I \prod_i \varphi(X_i, X_i(I), \tau_i^2(I) + \sigma_i^2)}{\lambda_{I_0} \prod_i \varphi(X_i, X_i(I_0), \tau_i^2(I_0) + \sigma_i^2)}\right]^h. \tag{31}$$

Recall the elementary identity: for $Y \sim N(\mu, \sigma^2)$ and $b > -\sigma^{-2}$,

$$\mathrm{E}(\exp\{-bY^2/2\}) = \exp\left\{-\frac{\mu^2 b}{2(1 + b\sigma^2)} - \frac{1}{2}\log(1 + b\sigma^2)\right\}. \tag{32}$$

Using (31) and (32) with $h = 1$, we derive that, for any $I, I_0 \in \mathbb{N}$ such that $I < I_0$,

$$\begin{aligned}\mathrm{E}_{\theta_0}\mathrm{P}(\mathcal{I} = I|X) &\le \mathrm{E}_{\theta_0}\frac{\lambda_I (K+1)^{-I/2}\exp\{-\sum_{i=I+1}^\infty \frac{X_i^2}{2\sigma_i^2}\}}{\lambda_{I_0}(K+1)^{-I_0/2}\exp\{-\sum_{i=I_0+1}^\infty \frac{X_i^2}{2\sigma_i^2}\}} \\ &= e^{\alpha(I_0 - I)}(K+1)^{(I_0 - I)/2}\mathrm{E}_{\theta_0}\exp\left\{-\frac{1}{2}\sum_{I+1}^{I_0}\frac{X_i^2}{\sigma_i^2}\right\} \qquad (33) \\ &= e^{-(\alpha + a_k)I}\exp\left\{(\alpha + a_K)I_0 - \frac{1}{4}\sum_{I+1}^{I_0}\frac{\theta_{0,i}^2}{\sigma_i^2}\right\},\end{aligned}$$

where $a_K = \frac{1}{2}\log(\frac{K+1}{2})$. Next, apply (31) and (32) to the case $I > I_0$: for any $h \in [0, 1)$,

$$\begin{aligned}
&\mathrm{E}_{\theta_0}\mathrm{P}(\mathcal{I} = I|X) \\
&\le e^{\alpha h(I_0-I)}(K+1)^{(I_0-I)h/2}\mathrm{E}_{\theta_0}\exp\left\{\frac{h}{2}\sum_{i=I_0+1}^{I}\frac{X_i^2}{\sigma_i^2}\right\} \qquad (34)\\
&= e^{-\alpha hI/2}\exp\left\{-\frac{\alpha hI}{2} + \alpha hI_0 - b_{K,h}(I - I_0) + \frac{h}{2(1-h)}\sum_{i=I_0+1}^{I}\frac{\theta_{0,i}^2}{\sigma_i^2}\right\},
\end{aligned}$$

where $b_{K,h} = \frac{h}{2}\log(K+1) + \frac{1}{2}\log(1-h)$. Clearly, $b_{K,h} > 0$ if $K > (1-h)^{-1/h} - 1$. Now take $h = 0.1$ in (34), then $b_{K,0.1} = \frac{1}{20}\log(K+1) + \frac{1}{2}\log(0.9) > 0$ since $K \ge 1.87 > (10/9)^{10} - 1$ by the condition of the theorem. Thus, for any $I, I_0 \in \mathbb{N}$ such that $I > I_0$, we obtain

$$\mathrm{E}_{\theta_0}\mathrm{P}(\mathcal{I} = I|X) \le e^{-\alpha I/20}\exp\left\{-\frac{\alpha}{20}\left(I - 2I_0 - \frac{10}{9\alpha}\sum_{i=I_0+1}^{I}\frac{\theta_{0,i}^2}{\sigma_i^2}\right)\right\}. \qquad (35)$$

Step 2: *A bound by the sum of three terms*. Recall $r^2(I,\theta_0) = \sum_{i\le I}\sigma_i^2 + \sum_{i>I}\theta_{0,i}^2$ and $r^2(\theta_0) = r^2(I_o,\theta_0) = \min_I r^2(I,\theta_0)$. Notice that

$$r^2(I,\theta_0) \le r^2(\theta_0) + 1\{I \le I_o\}\sum_{i=I+1}^{I_o}\theta_{0,i}^2 + 1\{I > I_o\}\sum_{i=I_o+1}^{I}\sigma_i^2. \qquad (36)$$

Next, as $\mathrm{P}_I(\cdot|X) = \bigotimes_i N(X_i 1\{i \le I\}, L\sigma_i^2 1\{i \le I\})$ with $L = \frac{K}{K+1} \le 1$, we obtain by applying the Markov inequality that

$$\begin{aligned}
\mathrm{P}_I\big(\|\theta - \theta_0\| \ge Mr(\theta_0)|X\big) &\le \frac{\mathrm{E}_I(\|\theta-\theta_0\|^2|X)}{M^2r^2(\theta_0)} \\
&= \frac{L\sum_{i\le I}\sigma_i^2 + \sum_{i>I}\theta_{0,i}^2 + \sum_{i\le I}(X_i - \theta_{0,i})^2}{M^2r^2(\theta_0)} \qquad (37)\\
&\le \frac{r^2(I,\theta_0) + \sum_{i\le I}\sigma_i^2\xi_i^2}{M^2r^2(\theta_0)} \triangleq \upsilon_I,
\end{aligned}$$

where $\xi_i = \sigma_i^{-1}(X_i - \theta_{0,i}) \overset{\text{ind}}{\sim} N(0,1)$ from the P_{θ_0}-perspective. Denote for brevity $p_I = \mathrm{P}(\mathcal{I} = I|X)$, so that $p_I \in [0,1]$ and $\sum_I p_I = 1$. In view of (8) and (37),

$$\mathrm{P}\big(\|\theta - \theta_0\| \ge Mr(\theta_0)|X\big) \le \sum_I \upsilon_I p_I = T_1 + T_2 + T_3, \qquad (38)$$

where $T_1 = \sum_{I\le I_o}\upsilon_I p_I$, $T_2 = \sum_{I_o<I\le\tau I_o}\upsilon_I p_I$, $T_3 = \sum_{I>\tau I_o}\upsilon_I p_I$, and $\tau > 2$ is from property (iv) of (5).

Step 3: *Handling the term* T_1. For $\tau_1 > 0$ to be chosen later, introduce the sets

$$\mathcal{O}^- = \mathcal{O}^-(\tau_1, \theta_0) = \left\{ I \in \mathbb{N} : I \le I_o, \sum_{i=I+1}^{I_o} \theta_{0,i}^2 \le \tau_1 \sum_{i=1}^{I_o} \sigma_i^2 \right\},$$

$$\mathcal{N}^- = \mathcal{N}^-(\tau_1, \theta_0) = \left\{ I \in \mathbb{N} : I \le I_o, \sum_{i=I+1}^{I_o} \theta_{0,i}^2 > \tau_1 \sum_{i=1}^{I_o} \sigma_i^2 \right\}.$$

By (36), $\max_{I \in \mathcal{O}^-} r^2(I, \theta_0) \le (1+\tau_1) r^2(\theta_0)$. This and (37) imply

$$\mathrm{E}_{\theta_0} \sum_{I \in \mathcal{O}^-} \upsilon_I p_I \le \mathrm{E}_{\theta_0} \max_{I \in \mathcal{O}^-} \upsilon_I \le \frac{1+\tau_1}{M^2} + \frac{\mathrm{E}_{\theta_0} \sum_{i \le I_o} \sigma_i^2 \xi_i^2}{M^2 r^2(\theta_0)} \le \frac{2+\tau_1}{M^2}. \tag{39}$$

The property (i) from (5) yields $I_o \le \frac{K_1}{\sigma_{I_o}^2} \sum_{i=1}^{I_o} \sigma_i^2$. Besides, for each $I \in \mathcal{N}^-$, $\sum_{i=1}^{I_o} \sigma_i^2 < \frac{1}{\tau_1} \sum_{i=I+1}^{I_o} \theta_{0,i}^2$. Set $\tau_1 = 8(\alpha + a_K) K_1$, with $a_K = \frac{1}{2} \log(\frac{K+1}{2}) > 0$. The last two relations and (33) imply that, for each $I \in \mathcal{N}^-$,

$$\begin{aligned}
\mathrm{E}_{\theta_0} p_I = \mathrm{E}_{\theta_0} \mathrm{P}(\mathcal{I} = I | X) &\le e^{-(\alpha+a_K)I} \exp\left\{ (\alpha + a_K) I_o - \frac{1}{4} \sum_{I+1}^{I_0} \frac{\theta_{0,i}^2}{\sigma_i^2} \right\} \\
&\le e^{-(\alpha+a_K)I} \exp\left\{ \frac{(\alpha+a_K)K_1}{\sigma_{I_o}^2} \sum_{i=1}^{I_o} \sigma_i^2 - \frac{1}{4\sigma_{I_o}^2} \sum_{i=I+1}^{I_o} \theta_{i,0}^2 \right\} \qquad (40) \\
&\le e^{-(\alpha+a_K)I} \exp\left\{ -\frac{1}{8\sigma_{I_o}^2} \sum_{i=I+1}^{I_o} \theta_{0,i}^2 \right\}.
\end{aligned}$$

Using (36), (37), (40) and the fact that $\max_{x \ge 0}\{x e^{-cx}\} \le (ce)^{-1}$ for $c > 0$, we obtain

$$\begin{aligned}
&\mathrm{E}_{\theta_0} \sum_{I \in \mathcal{N}^-} \upsilon_I p_I \\
&\quad \le \mathrm{E}_{\theta_0} \sum_{I \in \mathcal{N}^-} \frac{r^2(\theta_0) + \sum_{i=I+1}^{I_o} \theta_{0,i}^2 + \sum_{i \le I} \sigma_i^2 \xi_i^2}{M^2 r^2(\theta_0)} p_I \\
&\quad \le \frac{1}{M^2} + \frac{\mathrm{E}_{\theta_0} \sum_{i \le I_o} \sigma_i^2 \xi_i^2}{M^2 r^2(\theta_0)} + \sum_{I \in \mathcal{N}^-} \frac{(\sum_{i=I+1}^{I_o} \theta_{0,i}^2) \mathrm{E}_{\theta_0} p_I}{M^2 r^2(\theta_0)} \\
&\quad \le \frac{2}{M^2} + \sum_{I \in \mathcal{N}^-} \frac{(\sum_{i=I+1}^{I_o} \theta_{0,i}^2) \exp\{-(8\sigma_{I_o}^2)^{-1} \sum_{i=I+1}^{I_o} \theta_{0,i}^2\} e^{-(\alpha+a_K)I}}{M^2 r^2(\theta_0)} \\
&\quad \le \frac{2}{M^2} + \sum_{I \in \mathcal{N}^-} \frac{8e^{-1} \sigma_{I_o}^2 e^{-(\alpha+a_K)I}}{M^2 r^2(\theta_0)} \le \frac{2}{M^2} + \frac{8e^{-1}}{M^2} \sum_I e^{-(\alpha+a_K)I} = \frac{C_1}{M^2},
\end{aligned}$$

where $C_1 = 2 + \frac{8e^{-(1+\alpha+a_K)}}{1-e^{-(\alpha+a_K)}}$. The last relation and (39) give

$$\mathrm{E}_{\theta_0} T_1 = \mathrm{E}_{\theta_0} \sum_{I \in \mathcal{O}^-(\tau_1,\theta_0)} \upsilon_I p_I + \mathrm{E}_{\theta_0} \sum_{I \in \mathcal{N}^-(\tau_1,\theta_0)} \upsilon_I p_I \le \frac{C_2}{M^2}, \tag{41}$$

where $C_2 = 2 + \tau_1 + C_1 = 4 + 8(\alpha + a_K)K_1 + \frac{8e^{-(1+\alpha+a_K)}}{1-e^{-(\alpha+a_K)}}$.

Step 4: *Handling the term* T_2. Since $p_I \in [0,1]$ and $\sum_I p_I = 1$, $\mathrm{E}_{\theta_0} T_2 = \mathrm{E}_{\theta_0} \sum_{I_o < I \le \tau I_o} \upsilon_I p_I \le \mathrm{E}_{\theta_0}[\max_{I_o < I \le \tau I_o} \upsilon_I]$. Using this, (36), (37), (38) and property (ii) from (5), we get [with $\tau > 2$ from property (iv) of (5)]

$$\begin{aligned} \mathrm{E}_{\theta_0} T_2 &\le \mathrm{E}_{\theta_0} \max_{I_o < I \le \tau I_o} \upsilon_I \le \frac{\max_{I_o < I \le \tau I_o} r^2(I,\theta_0) + \mathrm{E}_{\theta_0} \sum_{i \le \tau I_o} \sigma_i^2 \xi_i^2}{M^2 r^2(\theta_0)} \\ &\le \frac{1}{M^2} + \frac{2\sum_{i \le \tau I_o} \sigma_i^2}{M^2 r^2(\theta_0)} \le \frac{1}{M^2} + \frac{2K_2(\tau)\sum_{i=1}^{I_o} \sigma_i^2}{M^2 r^2(\theta_0)} \le \frac{1 + 2K_2(\tau)}{M^2}. \end{aligned} \tag{42}$$

Step 5: *Handling the term* T_3. For some $\tau_2 > 0$ to be chosen later, introduce

$$\mathcal{O}^+ = \mathcal{O}^+(\tau,\tau_2,\theta_0) = \left\{ I \in \mathbb{N} : I > \tau I_o, \sum_{i=I_o+1}^{I} \sigma_i^2 \le \tau_2 \sum_{i > I_o} \theta_{0,i}^2 \right\},$$

$$\mathcal{N}^+ = \mathcal{N}^+(\tau,\tau_2,\theta_0) = \left\{ I \in \mathbb{N} : I > \tau I_o, \sum_{i=I_o+1}^{I} \sigma_i^2 > \tau_2 \sum_{i > I_o} \theta_{0,i}^2 \right\}.$$

By (36), $\max_{I \in \mathcal{O}^+} r^2(I,\theta_0) \le (1+\tau_2) r^2(\theta_0)$. Let $I^+ = \max\{\mathcal{O}^+\}$, then $\sum_{i \le I^+} \sigma_i^2 \le \sum_{i=1}^{I_o} \sigma_i^2 + \tau_2 \sum_{i > I_o} \theta_{0,i}^2 \le (1 \vee \tau_2) r^2(\theta_0)$. In view of (37), the last two relations entail that

$$\begin{aligned} \mathrm{E}_{\theta_0} \sum_{I \in \mathcal{O}^+} \upsilon_I p_I &\le \mathrm{E}_{\theta_0} \max_{I \in \mathcal{O}^+} \upsilon_I \le \frac{1+\tau_2}{M^2} + \frac{\mathrm{E}_{\theta_0} \sum_{i \le I^+} \sigma_i^2 \xi_i^2}{M^2 r^2(\theta_0)} \\ &\le \frac{2(1+\tau_2)}{M^2}. \end{aligned} \tag{43}$$

For K_4 and $\tau > 2$ from property (iv) of (5), we have that $\Sigma(m) - \Sigma(\lfloor m/\tau \rfloor) \ge K_4 \Sigma(m)$ for any $m \ge \tau$. This entails that, for each $I \in \mathcal{N}^+$,

$$\begin{aligned} \sum_{i=\lfloor I/\tau \rfloor+1}^{I} \sigma_i^2 &\ge K_4 \sum_{i=1}^{I} \sigma_i^2 \ge K_4 \sum_{i=I_o+1}^{I} \sigma_i^2 \ge K_4 \tau_2 \sum_{i=I_o+1}^{I} \theta_{0,i}^2 \\ &\ge K_4 \tau_2 \sum_{i=\lfloor I/\tau \rfloor+1}^{I} \theta_{0,i}^2. \end{aligned}$$

For each $I \in \mathcal{N}^+$, take $I_0 = I_0(I) = \lfloor I/\tau \rfloor$, then apply the property (v) of (5) and the last inequality with $\tau_2 = \frac{10\tau}{9\alpha(\tau-2)K_4K_5(\tau)}$ to derive

$$
\begin{aligned}
I - 2I_0 - \frac{10}{9\alpha}\sum_{i=I_0+1}^{I}\frac{\theta_{0,i}^2}{\sigma_i^2} &\geq \Big(1-\frac{2}{\tau}\Big)I - \frac{10}{9\alpha}\sum_{i=I_0+1}^{I}\frac{\theta_{0,i}^2}{\sigma_i^2} \\
&\geq \Big(1-\frac{2}{\tau}\Big)K_5\sum_{i=I_0+1}^{I}\frac{\sigma_i^2}{\sigma_{I_0}^2} - \frac{10}{9\alpha}\sum_{i=I_0+1}^{I}\frac{\theta_{0,i}^2}{\sigma_{I_0}^2} \\
&\geq \frac{(\tau-2)K_5(\tau)K_4\tau_2}{\tau}\sum_{i=I_0+1}^{I}\frac{\theta_{0,i}^2}{\sigma_{I_0}^2} - \frac{10}{9\alpha}\sum_{i=I_0+1}^{I}\frac{\theta_{0,i}^2}{\sigma_{I_0}^2} = 0.
\end{aligned}
$$

The last relation and the bound (35) with $I_0 = \lfloor I/\tau \rfloor$ imply that

$$
\text{(44)} \qquad \mathrm{E}_{\theta_0} p_I = \mathrm{E}_{\theta_0}\mathrm{P}(\mathcal{I} = I|X) \leq e^{-\gamma I}, \qquad I \in \mathcal{N}^+, \gamma = \frac{\alpha}{20}.
$$

Since $p_I \in [0,1]$ and $\mathrm{E}[\sum_{i=1}^m \sigma_i^2\xi_i^2]^2 \leq 3[\sum_{i=1}^m \sigma_i^2]^2$ for any $m \in \mathbb{N}$, we obtain by the Cauchy–Schwarz inequality that

$$
\text{(45)} \qquad \mathrm{E}_{\theta_0}\Big[p_I\sum_{i\leq I}\sigma_i^2\xi_i^2\Big] \leq \big(\mathrm{E}_{\theta_0}p_I^2\big)^{1/2}\sqrt{3}\sum_{i\leq I}\sigma_i^2 \leq \sqrt{3}(\mathrm{E}_{\theta_0}p_I)^{1/2}\sum_{i\leq I}\sigma_i^2.
$$

Combining (36), (37), (44), (45), the property (iii) of (5) and the fact that $\sigma_1^2 \leq r^2(\theta_0)$, we derive

$$
\begin{aligned}
\mathrm{E}_{\theta_0}\sum_{I\in\mathcal{N}^+} p_I\upsilon_I &= \sum_{I\in\mathcal{N}^+(\tau,\tau_2)}\frac{r^2(I,\theta_0)\mathrm{E}_{\theta_0}p_I + \mathrm{E}_{\theta_0}[p_I\sum_{i\leq I}\sigma_i^2\xi_i^2]}{M^2r^2(\theta_0)} \\
&\leq \frac{1}{M^2} + \sum_{I\in\mathcal{N}^+}\frac{(\sum_{i=I_o+1}^{I}\sigma_i^2)\mathrm{E}_{\theta_0}p_I + \sqrt{3}(\sum_{i\leq I}\sigma_i^2)(\mathrm{E}_{\theta_0}p_I)^{1/2}}{M^2r^2(\theta_0)} \\
&\leq \frac{1}{M^2} + \sum_{I\in\mathcal{N}^+}\frac{(\sum_{i=I_o+1}^{I}\sigma_i^2)e^{-\gamma I} + \sqrt{3}(\sum_{i\leq I}\sigma_i^2)e^{-\gamma I/2}}{M^2r^2(\theta_0)} \\
&\leq \frac{1 + K_3(\gamma) + \sqrt{3}K_3(\gamma/2)}{M^2}.
\end{aligned}
$$

Finally, the last relation and (43) entail the bound

$$
\text{(46)} \qquad \mathrm{E}_{\theta_0}T_3 = \mathrm{E}_{\theta_0}\sum_{I\in\mathcal{O}^+(\tau,\tau_2,\theta_0)}\upsilon_I p_I + \mathrm{E}_{\theta_0}\sum_{I\in\mathcal{N}^+(\tau,\tau_2,\theta_0)}\upsilon_I p_I \leq \frac{C_3}{M^2},
$$

where $C_3 = 2(1+\tau_2) + 1 + K_3(\gamma) + \sqrt{3}K_3(\gamma/2)$.

Step 6: *Finalizing the proof*. Piecing together the relations (38), (41), (42) and (46), we finally obtain

$$\mathrm{E}_{\theta_0}\mathrm{P}\big(\|\theta-\theta_0\|\ge Mr(\theta_0)|X\big)\le \mathrm{E}_{\theta_0}(T_1+T_2+T_3)\le \frac{C_{\mathrm{or}}}{M^2}.$$

The constant $C_{\mathrm{or}}=C_{\mathrm{or}}(K,\alpha)$ is as follows:

$$C_{\mathrm{or}}=C_2+1+2K_2(\tau)+2(1+\tau_2)+1+K_3(\gamma)+\sqrt{3}K_3(\gamma/2),$$

where $C_2=4+8(\alpha+a_K)K_1+\frac{8e^{-(1+\alpha+a_K)}}{1-e^{-(\alpha+a_K)}}$, $a_K=\frac{1}{2}\log(\frac{K+1}{2})$, $\tau_2=\frac{10\tau}{9\alpha(\tau-2)K_4K_5(\tau)}$, $\gamma=\frac{\alpha}{20}$, the constants τ, K_1, K_2, K_3, K_4, K_5 are from (5).

5.2. *Proof of Theorem* 2. The proof of this theorem is essentially contained in the proof of Theorem 1. Only a finishing argument is needed. According to (14), $\hat\theta=\mathrm{E}(\theta|X)=\sum_I X(I)p_I$, with $X(I)=\{X_i(I),i\in\mathbb{N}\}=\{X_i 1\{i\le I\},i\in\mathbb{N}\}$ and $p_I=\mathrm{P}(\mathcal{I}=I|X)$. By the Fubini theorem and the fact that $p_I^2\le p_I$, we derive

$$\begin{aligned}\mathrm{E}_{\theta_0}\|\hat\theta-\theta_0\|^2&=\mathrm{E}_{\theta_0}\sum_i\Big(\sum_I X_i(I)p_I-\theta_{0,i}\Big)^2\\&\le \mathrm{E}_{\theta_0}\sum_i\sum_I\big(X_i(I)-\theta_{0,i}\big)^2p_I=\mathrm{E}_{\theta_0}\sum_I\|X(I)-\theta_0\|^2p_I\\&=\mathrm{E}_{\theta_0}\sum_I\Big(\sum_{i\le I}\sigma_i^2\xi_i^2+\sum_{i>I}\theta_{0,i}^2\Big)p_I\le M^2r^2(\theta_0)\mathrm{E}_{\theta_0}(T_1+T_2+T_3),\end{aligned}$$

where T_1,T_2,T_3 are defined in (37) and (38). In the last step of the proof of Theorem 1, it is established that $\mathrm{E}_{\theta_0}(T_1+T_2+T_3)\le\frac{C_{\mathrm{or}}}{M^2}$. The theorem follows with the constant $C_{\mathrm{est}}=C_{\mathrm{or}}$.

5.3. *Proof of Theorem* 3. We prove Theorem 3 in several steps.

Step 1: *First technical lemma*.

LEMMA 1. *Let DDM* $\mathrm{P}_{K,\alpha}(\mathcal{I}=I|X)$ *be given by* (10) *with parameters* $K,\alpha>0$ *chosen in such a way that* $a(K)>\alpha$, *with* $a(K)$ *defined by* (16). *Let* $\varkappa_0=\varkappa_0(K,\alpha)=\frac{a(K)-\alpha}{a(K)}$. *Then for any* $\theta_0\in\ell_2$ *and any* $\varkappa\in[0,\varkappa_0)$

$$\mathrm{E}_{\theta_0}\mathrm{P}(\mathcal{I}\le\varkappa\bar I_o|X)\le C\exp\{-c\bar I_o\},\tag{47}$$

where $c=a(K)(1-\varkappa)-\alpha>0$, $C=C_\alpha^{-1}=(e^\alpha-1)^{-1}$, *and* $\bar I_o=\bar I_o(\theta_0)$ *is defined by* (15).

PROOF. By the definition (15) of the surrogate oracle, $R^2(I,\theta_0)\ge R^2(\bar I_o,\theta_0)$ for any $\theta_0\in\ell_2$. For $I<\bar I_o$, this implies that $\sum_{i=I+1}^{\bar I_o}\frac{\theta_{0,i}^2}{\sigma_i^2}\ge\bar I_o-I$. Using this, we obtain that for $I \le \varkappa \bar{I}_o$

$$\frac{1}{4}\sum_{i=I+1}^{\bar{I}_o}\frac{\theta_{0,i}^2}{\sigma_i^2} - \frac{1}{2}\log\left(\frac{K+1}{2}\right)(\bar{I}_o - I) \ge \left(\frac{1}{4} - \frac{1}{2}\log\left(\frac{K+1}{2}\right)\right)(\bar{I}_o - I)$$
$$= a(K)(\bar{I}_o - I) \ge a(K)(1-\varkappa)\bar{I}_o.$$

The lemma follows from the last relation, (33) and the fact that $\sum_I \lambda_I = 1$:

$$\begin{aligned}
&E_{\theta_0}P(\mathcal{I} \le \varkappa \bar{I}_o|X) \\
&\quad \le \sum_{I \le \varkappa I_o} \frac{\lambda_I}{\lambda_{\bar{I}_o}} \exp\left\{-\frac{1}{4}\sum_{i=I+1}^{\bar{I}_o}\frac{\theta_{0,i}^2}{\sigma_i^2} - \frac{1}{2}\log\left(\frac{K+1}{2}\right)(\bar{I}_o - I)\right\} \\
&\quad \le \sum_{I \le \varkappa \bar{I}_o} \frac{\lambda_I}{\lambda_{\bar{I}_o}} \exp\{-(a(K)(1-\varkappa))\bar{I}_o\} \\
&\quad \le C_\alpha^{-1}\exp\{-(a(K)(1-\varkappa)-\alpha)\bar{I}_o\}. \qquad \square
\end{aligned}$$

Step 2: *Second technical lemma.*

LEMMA 2. *Let* $\Lambda(S)$ *be the Lebesgue measure* (*or volume*) *of a bounded set* $S \subset \mathbb{R}^k$, $k \in \mathbb{N}$, *and* $B_k(r) = \{x \in \mathbb{R}^k : \|x\| \le r\}$ (*here* $\|\cdot\|$ *is the usual Euclidean norm in* $\mathbb{R}^k$) *be the Euclidean ball of radius* r *in space* $\mathbb{R}^k$. *Then*

$$\Lambda(B_k(r)) \le e\pi^{-1/2} r^k k^{-(k+1)/2}(2\pi e)^{k/2}.$$

PROOF. By using Stirling's approximation for the Gamma function $\Gamma(x) = \sqrt{2\pi}x^{x-1/2}e^{-x+\varsigma/(12x)}$ for all $x \ge 1$ and some $0 \le \varsigma \le C$, we derive

$$\begin{aligned}
\Gamma\left(1+\frac{k}{2}\right) &= \sqrt{2\pi}\left(1+\frac{k}{2}\right)^{\frac{k+1}{2}} e^{-1-\frac{k}{2}+\frac{\varsigma}{6k+12}} \\
&= \frac{(1+\frac{2}{k})^{(k+1)/2}\sqrt{\pi}}{e^{1-\varsigma/(6k+12)}} k^{\frac{k+1}{2}}(2e)^{-\frac{k}{2}} \\
&= c_k k^{(k+1)/2}(2e)^{-k/2} \ge e^{-1}\pi^{1/2}k^{(k+1)/2}(2e)^{-k/2},
\end{aligned}$$

since $c_k = \frac{(1+2/k)^{(k+1)/2}\sqrt{\pi}}{e^{1-\varsigma/(6k+12)}} > \frac{\sqrt{\pi}}{e}$. Combining the last relation with the fact that $\Lambda(B_k(r)) = r^k\Lambda(B_k(1)) = \frac{r^k\pi^{k/2}}{\Gamma(1+k/2)}$ completes the proof. $\square$

Step 3: *Small ball bound for* $P_I(\cdot|X)$. Recall that, with $L = K/(K+1)$,

$$P_I(\theta|X) = \bigotimes_i N(X_i 1\{i \le I\}, L\sigma_i^2 1\{i \le I\}), \qquad I \in \mathbb{N}.$$

We have that $\Sigma(I)=\sum_{i=1}^I\sigma_i^2\le\varepsilon^2\frac{(2I)^{2p+1}}{2p+1}$. By Stirling's bound, $\prod_{i=1}^I\kappa_i=(I!)^p\ge((I/e)^I\sqrt{2\pi I})^p$. Let $Z_1,\dots,Z_I$ be independent $N(0,1)$ random variables. Using these relations, Anderson's inequality and Lemma 2, we obtain that, P_{θ_0}-almost surely,

$$
\begin{aligned}
&\mathrm{P}_I\big(\|\theta-\tilde\theta\|\le\delta\Sigma^{1/2}(\bar I_o)|X\big)\\
&\quad=\mathrm{P}_I\big(\|\theta-\tilde\theta\|^2\le\delta^2\Sigma(\bar I_o)|X\big)\\
&\quad=\mathrm{P}\Big(\sum_{i\le I}(X_i+\sigma_i\sqrt{L}Z_i-\tilde\theta_i)^2+\sum_{i>I}\tilde\theta_i^2\le\delta^2\Sigma(\bar I_o)\Big|X\Big)\\
&\quad\le\mathrm{P}\Big(L\sum_{i\le I}\sigma_i^2Z_i^2\le\delta^2\Sigma(\bar I_o)\Big)=\mathrm{P}\Big(\sum_{i\le I}\sigma_i^2Z_i^2\le\frac{\delta^2\Sigma(\bar I_o)}{L}\Big)\\
&\quad\le\frac{\Lambda(B_I(\delta\sqrt{\Sigma(\bar I_o)/L}))}{\prod_{i=1}^I(2\pi\sigma_i^2)^{1/2}}\le\frac{(2\pi)^{-I/2}}{\prod_{i=1}^I\varepsilon\kappa_i}\frac{e}{\sqrt\pi}\Big(\frac{\delta^2\Sigma(\bar I_o)}{L}\Big)^{I/2}I^{-\frac{I+1}{2}}(2\pi e)^{I/2}\\
&\quad\le\frac{eI^{-(p+1)/2}}{(2\pi)^{p/2}\sqrt\pi}\Big[\Big(\frac{2e\bar I_o}{I}\Big)^{p+1/2}\Big(\frac{\delta}{\sqrt{L(2p+1)}}\Big)\Big]^I.
\end{aligned}
\tag{48}
$$

Step 4: *Applying Lemma* 1. Denote for brevity $\varrho=a(K)-\alpha$. By (16), $\varrho>0$. Applying Lemma 1 with $\varkappa=\frac{\varkappa_0}{2}=\frac{a(K)-\alpha}{2a(K)}$ [so that $a(K)(1-\varkappa)-\alpha=\frac{a(K)-\alpha}{2}=\frac{\varrho}{2}$], we obtain

$$
\mathrm{E}_{\theta_0}\mathrm{P}(\mathcal{I}<\varkappa\bar I_o|X)\le C_\alpha^{-1}e^{-\varrho\bar I_o/2}
\tag{49}
$$

for every $\theta_0\in\ell_2$. Consider the two cases: $e^{-\varrho\bar I_o/2}\le\delta$ and $e^{-\varrho\bar I_o/2}>\delta$.

Step 5: *The case* $e^{-\varrho\bar I_o/2}>\delta$. If $e^{-\varrho\bar I_o/2}>\delta$, then $\bar I_o<2\varrho^{-1}\log(\delta^{-1})$. By using this, (8), (48) and the notation $p_I=\mathrm{P}(\mathcal{I}=I|X)$, we derive that, for $e^{-\varrho\bar I_o/2}>\delta$,

$$
\begin{aligned}
&\mathrm{E}_{\theta_0}\mathrm{P}\big(\|\theta-\tilde\theta\|\le\delta\Sigma^{1/2}(\bar I_o)|X\big)\\
&\qquad=\mathrm{E}_{\theta_0}\sum_I\mathrm{P}_I\big(\|\theta-\tilde\theta\|^2\le\delta^2\Sigma(\bar I_o)|X\big)p_I\\
&\qquad\le\sum_I\frac{eI^{-(p+1)/2}}{(2\pi)^{p/2}\sqrt\pi}\Big[\Big(\frac{2e\bar I_o}{I}\Big)^{p+1/2}\Big(\frac{\delta}{\sqrt{L(2p+1)}}\Big)\Big]^I\mathrm{E}_{\theta_0}p_I\\
&\qquad\le C_2\delta\big[\log\big(\delta^{-1}\big)\big]^{p+1/2}\sum_I\frac{(C_1\delta[\log(\delta^{-1})]^{p+1/2})^{I-1}}{I^{I(p+1/2)+(p+1)/2}}\mathrm{E}_{\theta_0}p_I\\
&\qquad\le C_3\delta\big[\log\big(\delta^{-1}\big)\big]^{p+1/2},
\end{aligned}
\tag{50}
$$

with $C_1=C_1(K,\alpha)=\frac{(4e/\varrho)^{p+1/2}}{(L(2p+1))^{1/2}}$, $C_2=C_2(K,\alpha)=\frac{C_1e}{\pi^{1/2}(2\pi)^{p/2}}$, $\varrho=a(K)-\alpha$, $L=\frac{K}{K+1}$ and $C_3=C_3(K,\alpha)$. Let us evaluate C_3. Since $\max_{u>0}(\frac{c}{u})^u=e^{c/e}$ for

$c > 0$,

$$C_3 = C_2 \max_{I \geq 1} \frac{(C_1 a)^{I-1}}{I^{I(p+1/2)}} \leq \frac{C_2}{C_1 a}\Big[\max_{u>0}(b/u)^u\Big]^{p+1/2} = \frac{C_2}{C_1 a} e^{b(p+1/2)/e},$$

where $a = a(p) = \max_{0\leq\delta\leq 1}(\delta[\log(\delta^{-1})]^{p+1/2})$ and $b = (C_1 a)^{1/(p+1/2)}$.

Step 6: *The case* $e^{-\varrho \bar{I}_o/2} \leq \delta$. Now consider the case $e^{-\varrho \bar{I}_o/2} \leq \delta$. Clearly, $\sum_{I<\varkappa \bar{I}_o} p_I = \mathrm{P}(\mathcal{I} < \varkappa \bar{I}_o | X)$. In view of this, (48) and (49),

$$\begin{aligned}
&\mathrm{E}_{\theta_0}\mathrm{P}\big(\|\theta - \tilde{\theta}\| \leq \delta \Sigma^{1/2}(\bar{I}_o)|X\big) \\
&\quad = \mathrm{E}_{\theta_0} \sum_I \mathrm{P}_I\big(\|\theta - \tilde{\theta}\|^2 \leq \delta^2 \Sigma(\bar{I}_o)|X\big) p_I \\
&\quad \leq \sum_{I \geq \varkappa \bar{I}_o} \frac{e I^{-\frac{p+1}{2}}}{(2\pi)^{\frac{p}{2}}\sqrt{\pi}} \bigg[\bigg(\frac{2e\bar{I}_o}{I}\bigg)^{p+\frac{1}{2}} \bigg(\frac{\delta}{\sqrt{L(2p+1)}}\bigg)\bigg]^I \mathrm{E}_{\theta_0} p_I \\
&\qquad + \mathrm{E}_{\theta_0}\mathrm{P}(\mathcal{I} < \varkappa \bar{I}_o | X) \\
&\quad \leq C_4 \delta \sum_I \frac{[(\frac{2e}{\varkappa})^{p+\frac{1}{2}} \frac{\delta}{\sqrt{L(2p+1)}}]^{I-1}}{I^{(p+1)/2}} \mathrm{E}_{\theta_0} p_I + \frac{e^{-\varrho \bar{I}_o/2}}{C_\alpha} \leq (C_4 + C_\alpha^{-1})\delta
\end{aligned}$$

if $e^{-\varrho \bar{I}_o/2} \leq \delta$ and $(\frac{2e}{\varkappa})^{p+\frac{1}{2}} \frac{\delta}{\sqrt{L(2p+1)}} \leq 1$. Here, $C_4 = \frac{e}{(2\pi)^{\frac{p}{2}}\sqrt{\pi L(2p+1)}}(\frac{2e}{\varkappa})^{p+\frac{1}{2}}$.

Step 7: *Finalizing the proof of Theorem* 3. The last relation holds if $e^{-\varrho \bar{I}_o/2} \leq \delta \leq \sqrt{L(2p+1)}(\frac{\varkappa}{2e})^{p+1/2} = \sqrt{\frac{K(2p+1)}{K+1}}(\frac{\varkappa}{2e})^{p+1/2} = \bar{\delta}_{\mathrm{sb}}$ and the relation (50) holds if $e^{-\varrho \bar{I}_o/2} > \delta$. Combining these two relations concludes the proof of the theorem: for $0 < \delta \leq (1 \wedge \bar{\delta}_{\mathrm{sb}}) = \delta_{\mathrm{sb}}$,

$$\mathrm{E}_{\theta_0}\mathrm{P}\big(\|\theta - \tilde{\theta}\| \leq \delta \Sigma^{1/2}(\bar{I}_o)|X\big) \leq \max\{C_3, C_4 + C_\alpha^{-1}\}\delta\big[\log(\delta^{-1})\big]^{p+\frac{1}{2}}.$$

5.4. *Proof of Theorem* 4. First, we bound the coverage probability of the confidence ball (20). Corollary 1 yields that for any $\delta \in (0, \delta_{\mathrm{eb}}]$ with $\delta_{\mathrm{eb}} = \delta_{\mathrm{eb}}(\tau) = (1+\tau)^{-1/2}\delta_{\mathrm{sb}}$,

$$\sup_{\theta_0 \in \Theta_{\mathrm{eb}}(\tau)} \mathrm{E}_{\theta_0}\mathrm{P}\big(\|\theta - \hat{\theta}\| \leq \delta r(\theta_0)|X\big) \leq C_{\mathrm{eb}}\delta\big[\log(\delta^{-1})\big]^{p+1/2}, \tag{51}$$

where $C_{\mathrm{eb}} = C_{\mathrm{sb}}\sqrt{1+\tau}$, and δ_{sb} and C_{sb} are some absolute constants (since K, α, p are fixed) from Theorem 3. By using the Markov inequality, (19) with $\kappa = \frac{1}{2}$ and Theorem 2, we obtain

$$\begin{aligned}
&\mathrm{P}_{\theta_0}\big(\theta_0 \notin B(\hat{\theta}, M\hat{r})\big) \\
&\quad \leq \mathrm{P}_{\theta_0}\big(\|\theta_0 - \hat{\theta}\| > M\hat{r}, \hat{r} \geq \delta r(\theta_0)\big) + \mathrm{P}_{\theta_0}\big(\hat{r} < \delta r(\theta_0)\big)
\end{aligned}$$

$$\leq \mathrm{P}_{\theta_0}\big(\|\theta_0 - \hat{\theta}\| > M\delta r(\theta_0)\big) + \mathrm{P}_{\theta_0}\bigg(\mathrm{P}\big(\|\theta - \hat{\theta}\| \leq \delta r(\theta_0)|X\big) \geq \frac{1}{2}\bigg)$$

$$\leq \frac{C_{\mathrm{est}}}{M^2\delta^2} + 2\mathrm{E}_{\theta_0}\mathrm{P}\big(\|\theta - \hat{\theta}\| \leq \delta r(\theta_0)|X\big).$$

From the last relation and (51), it follows that, for any $M > 0$, $\delta \in (0, \delta_{\mathrm{eb}}]$,

$$\sup_{\theta_0 \in \Theta_{\mathrm{eb}}(\tau)} \mathrm{P}_{\theta_0}\big(\theta_0 \notin B(\hat{\theta}, M\hat{r})\big) \leq \frac{C_{\mathrm{est}}}{M^2\delta^2} + 2C_{\mathrm{eb}}\delta\big[\log(\delta^{-1})\big]^{p+1/2}. \tag{52}$$

Let $M_1 = M_1(\tau) \triangleq \delta_{\mathrm{eb}}^{-3/2}$. For any $M \geq M_1$, take $\delta = M^{-2/3}$ in (52) to get

$$\sup_{\theta_0 \in \Theta_{\mathrm{eb}}(\tau)} \mathrm{P}_{\theta_0}\big(\theta_0 \notin B(\hat{\theta}, M\hat{r})\big) \leq \frac{C_{\mathrm{est}} + 2C_{\mathrm{eb}}(\frac{2}{3}\log M)^{p+1/2}}{M^{2/3}} \triangleq \varphi_1(M).$$

The function $\varphi_1(M)$ is decreasing to zero for $M \geq M_2$, with some $M_2 > 0$. Then for all $M \geq M_0 = M_0(\tau) \triangleq M_1 \vee M_2$,

$$\sup_{\theta_0 \in \Theta_{\mathrm{eb}}(\tau)} \mathrm{P}_{\theta_0}\big(\theta_0 \notin B(\hat{\theta}, M\hat{r})\big) \leq \varphi_1(M). \tag{53}$$

Next, we verify the size property. By using the (conditional) Markov inequality, (19), Theorems 1 and 2, we derive that, for any $\theta_0 \in \ell_2$,

$$\begin{aligned}
&\mathrm{P}_{\theta_0}\big(\hat{r} \geq cr(\theta_0)\big)\\
&\quad\leq \mathrm{P}_{\theta_0}\bigg(\mathrm{P}\big(\|\theta - \hat{\theta}\| \leq cr(\theta_0)|X\big) \leq \frac{1}{2}\bigg)\\
&\quad= \mathrm{P}_{\theta_0}\big(\mathrm{P}\big(\|\theta - \hat{\theta}\| > cr(\theta_0)|X\big) > \gamma\big) \leq 2\mathrm{E}_{\theta_0}\big(\mathrm{P}\big(\|\theta - \hat{\theta}\| > cr(\theta_0)|X\big)\big)\\
&\quad\leq 2\mathrm{E}_{\theta_0}\bigg[\mathrm{P}\bigg(\|\theta - \theta_0\| \geq \frac{cr(\theta_0)}{2}\Big|X\bigg)\bigg] + 2\mathrm{E}_{\theta_0}\bigg[\mathrm{P}\bigg(\|\theta_0 - \hat{\theta}\| \geq \frac{cr(\theta_0)}{2}\Big|X\bigg)\bigg]\\
&\quad\leq \frac{8C_{\mathrm{or}}}{c^2} + \frac{8\mathrm{E}_{\theta_0}\|\theta_0 - \hat{\theta}\|^2}{c^2r^2(\theta_0)} \leq \frac{8(C_{\mathrm{or}} + C_{\mathrm{est}})}{c^2} \triangleq \varphi_2(c).
\end{aligned}$$

Thus,

$$\sup_{\theta_0 \in \ell_2} \mathrm{P}_{\theta_0}\big(\hat{r} \geq cr(\theta_0)\big) \leq \varphi_2(c). \tag{54}$$

The proof is complete as we established (53) and (54).

SUPPLEMENTARY MATERIAL

Supplement to "On coverage and local radial rates of credible sets" (DOI: 10.1214/16-AOS1477SUPP; .pdf). The elaboration on some points and some background information related to the paper are provided in the supplement [3].

REFERENCES

[1] BARAUD, Y. (2004). Confidence balls in Gaussian regression. *Ann. Statist.* **32** 528–551. MR2060168

[2] BELITSER, E. (2014). On coverage and local radial rates of DDM-credible sets. Available at arXiv:1407.5232.

[3] BELITSER, E. (2016). Supplement to "On coverage and local radial rates of credible sets." DOI:10.1214/16-AOS1477SUPP.

[4] BERAN, R. and DÜMBGEN, L. (1998). Modulation of estimators and confidence sets. *Ann. Statist.* **26** 1826–1856. MR1673280

[5] BIRGÉ, L. and MASSART, P. (2001). Gaussian model selection. *J. Eur. Math. Soc.* (*JEMS*) **3** 203–268. MR1848946

[6] BULL, A. D. (2012). Honest adaptive confidence bands and self-similar functions. *Electron. J. Stat.* **6** 1490–1516. MR2988456

[7] BULL, A. D. and NICKL, R. (2013). Adaptive confidence sets in L^2. *Probab. Theory Related Fields* **156** 889–919. MR3078289

[8] CAI, T. T. and LOW, M. G. (2006). Adaptive confidence balls. *Ann. Statist.* **34** 202–228. MR2275240

[9] CASTILLO, I. and NICKL, R. (2013). Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **41** 1999–2028. MR3127856

[10] CAVALIER, L. (2008). Nonparametric statistical inverse problems. *Inverse Probl.* **24** 034004, 19. MR2421941

[11] CAVALIER, L. and GOLUBEV, Y. (2006). Risk hull method and regularization by projections of ill-posed inverse problems. *Ann. Statist.* **34** 1653–1677. MR2283712

[12] COX, D. D. (1993). An analysis of Bayesian inference for nonparametric regression. *Ann. Statist.* **21** 903–923. MR1232525

[13] FREEDMAN, D. (1999). On the Bernstein–von Mises theorem with infinite-dimensional parameters. *Ann. Statist.* **27** 1119–1140. MR1740119

[14] GENOVESE, C. and WASSERMAN, L. (2008). Adaptive confidence bands. *Ann. Statist.* **36** 875–905. MR2396818

[15] GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531. MR1790007

[16] GINÉ, E. and NICKL, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122–1170. MR2604707

[17] HOFFMANN, M. and NICKL, R. (2011). On adaptive inference and confidence bands. *Ann. Statist.* **39** 2383–2409. MR2906872

[18] JOHNSTONE, I. M. (2010). High dimensional Bernstein–von Mises: Simple examples. In *Borrowing Strength*: *Theory Powering Applications—A Festschrift for Lawrence D. Brown*. *Inst. Math. Stat. Collect.* **6** 87–98. IMS, Beachwood, OH. MR2798513

[19] JUDITSKY, A. and LAMBERT-LACROIX, S. (2003). Nonparametric confidence set estimation. *Math. Methods Statist.* **12** 410–428. MR2054156

[20] KNAPIK, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2011). Bayesian inverse problems with Gaussian priors. *Ann. Statist.* **39** 2626–2657. MR2906881

[21] LI, K.-C. (1989). Honest confidence regions for nonparametric regression. *Ann. Statist.* **17** 1001–1008. MR1015135

[22] LOW, M. G. (1997). On nonparametric confidence intervals. *Ann. Statist.* **25** 2547–2554. MR1604412

[23] NICKL, R. and SZABÓ, B. T. (2014). A sharp adaptive confidence ball for self-similar functions. Available at arXiv:1406.3994.

[24] PICARD, D. and TRIBOULEY, K. (2000). Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.* **28** 298–335. MR1762913

[25] RAY, K. (2014). Adaptive Bernstein–von Mises theorems in Gaussian white noise. Available at arXiv:1407.3397.

[26] ROBINS, J. and VAN DER VAART, A. (2006). Adaptive nonparametric confidence sets. *Ann. Statist.* **34** 229–253. MR2275241

[27] SERRA, P. and KRIVOBOKOVA, T. (2014). Adaptive empirical Bayesian smoothing splines. Available at arXiv:1411.6860.

[28] SZABÓ, B., VAN DER VAART, A. and VAN ZANTEN, H. (2015). Honest Bayesian confidence sets for the L^2-norm. *J. Statist. Plann. Inference* **166** 36–51. MR3390132

[29] SZABÓ, B., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2015). Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.* **43** 1391–1428. MR3357861

DEPARTMENT OF MATHEMATICS
VU AMSTERDAM
1081 HV AMSTERDAM
THE NETHERLANDS
E-MAIL: e.n.belitser@vu.nl