Empirical Bayes analysis of spike and slab posterior distributions[*]

Ismaël Castillo and Romain Mismer

Sorbonne Unversité and Université Paris Diderot
Laboratoire Probabilités, Statistique et et Modélisation
UMR 8001 – 4, place Jussieu
75005 Paris, France
e-mail: ismael.castillo@upmc.fr*;* rmismer@free.fr

Abstract: In the sparse normal means model, convergence of the Bayesian posterior distribution associated to spike and slab prior distributions is considered. The key sparsity hyperparameter is calibrated via marginal maximum likelihood empirical Bayes. The plug-in posterior squared–L^2 norm is shown to converge at the minimax rate for the euclidean norm for appropriate choices of spike and slab distributions. Possible choices include standard spike and slab with heavy tailed slab, and the spike and slab LASSO of Ročková and George with heavy tailed slab. Surprisingly, the popular Laplace slab is shown to lead to a suboptimal rate for the empirical Bayes posterior itself. This provides a striking example where convergence of aspects of the empirical Bayes posterior such as the posterior mean or median does not entail convergence of the complete empirical Bayes posterior itself.

Contents

---

1. Introduction

In the sparse normal means model, one observes a sequence $X = (X_1, \ldots, X_n)$

$$X_i = \theta_i + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

with $\theta = (\theta_1, \ldots, \theta_n) \in \mathbb{R}^n$ and $\varepsilon_1, \ldots, \varepsilon_n$ i.i.d. $\mathcal{N}(0, 1)$. Given θ, the distribution of X is a product of Gaussians and is denoted by P_θ. Further, one assumes that the 'true' vector θ_0 belongs to

$$\ell_0[s_n] = \{\theta \in \mathbb{R}^n, \ \#\{i : \ \theta_i \neq 0\} \leq s_n\},$$

the set of vectors that have at most s_n nonzero coordinates, where $0 \leq s_n \leq n$. A typical *sparsity* assumption is that s_n is a sequence that may grow with n but is 'small' compared to n (e.g. in the asymptotics $n \to \infty$, one typically assumes $s_n/n = o(1)$ and $s_n \to \infty$). A natural problem is that of estimating θ with respect to the euclidean loss $\|\theta - \theta'\|^2 = \sum_{i=1}^n (\theta_i - \theta'_i)^2$. A benchmark is given by the minimax rate for this loss over the class of sparse vectors $\ell_0[s_n]$. Denoting

$$r_n := 2s_n \log(n/s_n),$$

[7] show that the minimax rate equals $(1 + o(1))r_n$ as $n \to \infty$.

Taking a Bayesian approach, one of the simplest and arguably most natural classes of prior distributions in this setting is given by so-called *spike and slab* distributions,

$$\theta \sim \bigotimes_{i=1}^n (1 - \alpha)\delta_0 + \alpha G,$$

where δ_0 denotes the Dirac mass at 0, the distribution G has density γ with respect to Lebesgue measure and α belongs to $[0, 1]$. These priors were introduced and advocated in a number of papers, including [15, 8, 9, 21]. One important point is the calibration of the tuning parameter α, which can be done in a number of ways, including: deterministic n-dependent choice, data-dependent choice based on a preliminary estimate $\hat{\alpha}$, fully Bayesian choice based on a prior distribution on α. Studying the behaviour of the posterior distributions in sparse settings is currently the object of a lot of activity. A brief (and by far not exhaustive) overview of recent works is given below. Given a prior distribution Π on θ, and interpreting P_θ as the law of X given θ, one forms the posterior distribution $\Pi[\cdot \,|\, X]$ which is the law of θ given X. The frequentist analysis of the posterior distribution consists in the study of the convergence of $\Pi[\cdot \,|\, X]$ in probability under P_{θ_0}, thus assuming that the data has actually been generated from some 'true' parameter θ_0.

In the present paper, we follow this path and are more particularly interested in obtaining a uniform bound on the posterior squared L^2-moment of the order of the optimal minimax rate, that is in proving, with C a large enough constant,

$$\sup_{\theta_0 \in \ell_0[s_n]} E_{\theta_0} \int \|\theta - \theta_0\|^2 d\Pi(\theta \,|\, X) \le C r_n \tag{2}$$

for Π a prior distribution constructed using a spike and slab approach, whose prior parameters may be calibrated using the data, that is following an empirical Bayes method. This is of interest for at least three reasons

- this provides adaptive convergence rates for the entire posterior distribution, using a fully data-driven procedure. This is more than obtaining convergence of aspects of the posterior such that posterior mean or mode, and in fact may require different conditions on the prior, as we shall see below.
- the inequality (2) automatically implies convergence of several commonly used point estimators derived from the posterior $\Pi[\cdot \,|\, X]$: it implies convergence at rate $C r_n$ of the posterior mean $\int \theta d\Pi(\theta \,|\, X)$ (using Jensen's inequality, see e.g. [6]), but also of the coordinatewise posterior median (see the supplement of [6] for details) and in fact of any fixed posterior coordinatewise quantile, for instance the quantile $1/4$ of $\Pi[\cdot \,|\, X]$. It also implies, using Tchebychev's inequality, convergence of the posterior distribution at rate $M_n r_n$ for $\| \cdot \|^2$ as in (3) below with $M = M_n$, for any $M_n \to \infty$.
- knowing (2) is a first step towards results for *uncertainty quantification*, in particular for the study of certain *credible sets*. Indeed, (2) suggests a natural way to build such a set, that is $\mathcal{C} \subset \mathbb{R}^n$ with $\Pi[\mathcal{C} \,|\, X] \ge 1 - \alpha$ for a given $\alpha \in (0, 1)$. Namely, define $\mathcal{C} = \{\theta : \|\theta - \bar{\theta}\|^2 \le r_X\}$, with $\bar{\theta}$ the posterior mean (or another suitable point estimate of θ) and r_X a large enough multiple of the $(1 - \alpha)$–quantile of $\int \|\theta - \bar{\theta}\|^2 d\Pi(\theta \,|\, X)$.

The present work is the first of a series of papers where we study aspects of inference using spike and slab prior distributions. In particular, based on the

present results, the behaviour of the previously mentioned credible sets is studied in the forthcoming paper [5].

Previous results on frequentist analysis of spike and slab type priors. In a seminal paper, Johnstone and Silverman [12] considered estimation of θ using spike and slab priors combined with an empirical Bayes method for choosing α. They chose $\alpha = \hat{\alpha}$ based on a marginal maximum likelihood approach to be described in more details below. Denoting $\hat{\theta}$ the associated posterior median (or posterior mean), [12] established that

$$\sup_{\theta_0 \in \ell_0[s_n]} E_{\theta_0} \|\hat{\theta} - \theta_0\|^2 \leq C r_n,$$

thereby proving minimaxity up to a constant of this estimator over $\ell_0[s_n]$. The estimator is adaptive, as the knowledge of s_n is not required in its construction.

In [6], convergence of the posterior distribution is studied in the case α is given a prior distribution. If $\alpha \sim \text{Beta}(1, n + 1)$, Π is the corresponding hierarchical prior, and $\Pi[\cdot \,|\, X]$ the associated posterior distribution, it is established in [6] that for large enough M, as $n \to \infty$,

$$\sup_{\theta_0 \in \ell_0[s_n]} E_{\theta_0} \Pi[\|\theta - \theta_0\|^2 \leq M r_n \,|\, X] \to 1. \tag{3}$$

In [14], Martin and Walker use a fractional likelihood approach to construct a certain empirical Bayes spike and slab prior, where the idea is to reweight the standard spike and slab prior by a power of the likelihood. They derive rate-optimal concentration results for the corresponding posterior distribution and posterior mean.

A related class of prior distributions recently put forward by Ročková [16] and Ročková and George [17], is given by

$$\theta \sim \bigotimes_{i=1}^{n} (1 - \alpha) G_0 + \alpha G_1,$$

where both distributions G_0, G_1 have densities with respect to Lebesgue measure. The authors in particular consider the choices $G_0 = \text{Lap}(\lambda_0)$ and $G_1 = \text{Lap}(\lambda_1)$, where $\text{Lap}(\lambda)$ denotes the Laplace (double-exponential) distribution. Taking λ_0 large enough enables one to mimic the spike of the standard spike and slab prior, and the fact that both G_0, G_1 are continuous distributions offers some computational advantages, especially when working with the posterior mode. One can also note that the posterior mode when $\alpha = 1$ leads to the standard LASSO estimator. For this reason, the authors in [16, 17] call this prior the *spike and slab LASSO* prior. It is shown in [16], Theorem 5.2 and corollaries, that a certain deterministic n-dependent choice of $\alpha, \lambda_0, \lambda_1$ (but independent on the unknown s_n) leads to posterior convergence at near-optimal rate $s_n \log n$, while putting a prior on α can yield ([16], Theorem 5.4) the minimax rate for the posterior, if a certain condition on the strength of the true non-zero coefficiencents of θ_0 is verified.

Other priors and related work. We briefly review other options to induce sparsity using a Bayesian approach. One option considered in [6] is first to draw a subset $S \subset \{1, \ldots, n\}$ at random and then to draw nonzero coordinates on this subset only. That is, sample first a dimension $k \in \{0, \ldots, n\}$ at random according to some prior π. Given k, sample S uniformly at random over subsets of size k and finally set

$$\theta_i \sim G \quad i \in S$$
$$\theta_i = 0 \quad i \notin S.$$

Under the assumption that the prior π on k is of the form, referred to as the complexity prior,

$$\pi(k) = ce^{-ak \log(nb/k)}, \tag{4}$$

[6] show that under this prior, both (3) and (2) are satisfied. However, such a 'strong' prior on the dimension is not necessary at least for (3) to hold: it can be checked for instance, for π the prior on dimension induced by the spike and slab prior on θ with $\alpha \sim \text{Beta}(1, n+1)$, that $\pi(s_n) \asymp \exp(-cs_n) \gg \exp(-cs_n \log(n/s_n))$. So in a sense the complexity prior 'penalises slightly more than necessary'.

Another popular way to induce sparsity is via the so-called *horseshoe* prior, which draws a θ from a continous distribution which is itself a mixture. As established in [18]–[19] the horseshoe yields the nearly-optimal rate $s_n \log n$ uniformly over the whole space $\ell_0[s_n]$, up again to the correct form of the logarithmic factor. In a different spirit but still without using Dirac masses at 0, the paper [11] shows that, remarkably, it is also possible to adopt an empirical Bayes approach on the entire unknown distribution function F of the vector θ, interpreting θ as sampled from a certain distribution, and the authors derive oracle results over $\ell^p, p > 0$, balls for the plug-in posterior mean (not including the case $p = 0$ though). We also note the interesting work [20] that investigates necessary and sufficient conditions for sparse continuous priors to be rate-optimal. However the latter is for a fixed regularity parameter s_n, while the results decribed in Section 2 (in particularity the suboptimality phenomenon, but also upper-bounds using the empirical Bayes approach) are related to adaptation.

Using complexity–type priors on the number of non-zero coordinates, Belitser and co-authors [1]–[2] consider Gaussian priors on non-zero coefficients, with a recentering of the posterior mean at the observation X_i– for those coordinates i that are selected– to adjust for overshrinkage. In [2], oracle results for the corresponding posterior are derived, that in particular imply convergence at the minimax rate up to constant over $\ell_0[s_n]$, and the authors also derive results on uncertainty quantification by studying the frequentist coverage of credible sets using their procedure.

For further references on the topic, in particular about relationships between spike and slab priors and absolutely continuous counterparts such as the horseshoe or the spike and slab LASSO, we refer to the paper [19] and its discussion by several authors of the previously mentioned works.

Overview of results and outline. This paper obtains the following results.

1. For the spike and slab prior, in Section 2.2 we establish lower bound results that show that the popular Laplace slab yields suboptimal rates when the complete empirical Bayes posterior is considered.
2. In Sections 2.3 and 2.6, we establish rate-optimal results for the posterior squared L^2–moment for the usual spike and slab with a Cauchy slab, when the prior hyperparameter is chosen via a marginal maximum likelihood method.
3. In Section 2.4, the spike and slab LASSO prior is considered and we provide a near-optimal adaptive rate for the corresponding complete empirical Bayes posterior distribution.

Section 2 introduces the framework, notation, and the main results, ending with a brief simulation study in Section 2.5 and discussion. Section 3 gathers the proofs of the lower-bound results as well as upper-bounds on the spike and slab prior. Technical lemmas for the spike and slab prior can be found in Section 4, while Sections 5–6 contain the proof of the result for the spike and slab LASSO prior.

For real-valued functions f, g, we write $f \lesssim g$ if there exists a universal constant C such that $f(x) \le Cg(x)$, and $f \gtrsim g$ is defined similarly. When x is a positive real number or an integer, we write $f(x) \asymp g(x)$ if there exists positive constants c, C, D such that for $x \ge D$, we have $cf(x) \le g(x) \le Cf(x)$. For reals a, b, one denotes $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$.

2. Framework and main results

2.1. Empirical Bayes estimation with spike and slab prior

In the setting of model (1), the spike and slab prior on θ with fixed parameter $\alpha \in [0, 1]$ is

$$\Pi_\alpha \sim \otimes_{i=1}^n (1 - \alpha)\delta_0 + \alpha G(\cdot), \tag{5}$$

where G is a given probability measure on $\mathbb{R}$. We consider the following choices

$$G = \begin{cases} & \text{Lap}(1) \\ or & \\ & \text{Cauchy}(1) \end{cases}$$

where $\text{Lap}(\lambda)$ denotes the Laplace (double exponential) distribution with parameter λ and $\text{Cauchy}(1)$ the standard Cauchy distribution. Different choices of parameters and prior distributions are possible (a brief discussion is included below) but for clarity of exposition we stick to these common distributions. In the sequel γ denotes the density of G with respect to Lebesgue measure.

By Bayes' formula the posterior distribution under (1) and (5) with fixed $\alpha \in [0, 1]$ is

$$\Pi_\alpha[\cdot \mid X] \sim \otimes_{i=1}^n (1 - a(X_i))\delta_0 + a(X_i)G_{X_i}(\cdot), \tag{6}$$

where, denoting by ϕ the standard normal density and $g(x) = \phi * G(x) = \int \phi(x - u)dG(u)$ the convolution of ϕ and G at point $x \in \mathbb{R}$, the posterior weight $a(X_i)$ is given by, for any i,

$$a(X_i) = a_\alpha(X_i) = \frac{\alpha g(X_i)}{(1 - \alpha)\phi(X_i) + \alpha g(X_i)}. \tag{7}$$

The distribution G_{X_i} has density

$$\gamma_{X_i}(\cdot) := \frac{\phi(X_i - \cdot)\gamma(\cdot)}{g(X_i)} \tag{8}$$

with respect to Lebesgue measure on $\mathbb{R}$. The behaviour of the posterior distribution $\Pi_\alpha[\cdot \,|\, X]$ heavily depends on the choices of the smoothing parameters α and γ. It turns out that some aspects of this distribution are thresholding-type estimators, as established in [12].

Posterior median and threshold $t(\alpha)$. The posterior median $\hat{\theta}_\alpha^{med}(X_i)$ of the ith coordinate has a thresholding property: there exists $t(\alpha) > 0$ such that $\hat{\theta}_\alpha^{med}(X_i) = 0$ if and only if $|X_i| \le t(\alpha)$. A default choice can be $\alpha = 1/n$; one can check that this leads to a posterior median behaving similarly as a hard thresholding estimator with threshold $\sqrt{2 \log n}$. One can significantly improve on this default choice by taking a well-chosen data-dependent α.

In order to choose α, in this paper we follow the empirical Bayes method proposed in [12]. The idea is to estimate α by maximising the marginal likelihood in α in the Bayesian model, which is the density of $\alpha \,|\, X$. The log-marginal likelihood in α can be written as

$$\ell(\alpha) = \ell_n(\alpha; X) = \sum_{i=1}^{n} \log((1 - \alpha)\phi(X_i) + \alpha g(X_i)). \tag{9}$$

Let $\hat{\alpha}$ be defined as the maximiser of the log-marginal likelihood

$$\hat{\alpha} = \underset{\alpha \in \mathcal{A}_n}{\operatorname{argmax}} \ \ell_n(\alpha; X), \tag{10}$$

where the maximisation is restricted to $\mathcal{A}_n = [\alpha_n, 1]$, with α_n defined by

$$t(\alpha_n) = \sqrt{2 \log n}.$$

The reason for this restriction is that one does not need to take α smaller than α_n, which would correspond to a choice of α 'more conservative' than hard-thresholding at threshold level $\sqrt{2 \log n}$.

In [12], Johnstone and Silverman prove that the posterior median $\hat{\alpha}^{med}(X_i)$ has remarkable optimality properties, for many choices of the slab density γ. For γ with tails 'at least as heavy as' the Laplace distribution, then this point estimator converges at the minimax rate over $\ell_0[s_n]$. More precisely, it follows from Theorem 1 in [12] that there exists constants C, c_0, c_1 such that if

$$c_1 \log^2 n \le s_n \le c_0 n, \tag{11}$$

then the posterior median $\hat{\theta}_{\hat{\alpha}}^{med} = (\hat{\theta}_{\hat{\alpha}}^{med}(X_i))_{1\le i \le n}$ is rate optimal

$$\sup_{\theta \in \ell_0[s_n]} E_\theta \|\hat{\theta}_{\hat{\alpha}}^{med} - \theta\|^2 \le C s_n \log(n/s_n). \tag{12}$$

One can actually remove the lower bound in condition (11) – see Theorem 2 in [12] – by a more complicated choice of $\hat{\alpha}$, for which $\hat{\alpha}$ in (10) is replaced by a smaller value if the empirical Bayes estimate is close to α_n given by $t(\alpha_n) = \sqrt{2\log n}$. In the present paper for simplicity of exposition we first work under the condition (11). In Section 2.6, we show that the lower bound part of the condition can be removed when working with the modified estimator as in [12].

Plug-in posterior distribution. The posterior we consider in this paper is $\Pi_{\hat{\alpha}}[\cdot \mid X]$, that is the distribution given by (6), where α has been replaced by its empirical Bayes (EB) estimate $\hat{\alpha}$ given by (10). This posterior is called complete EB posterior in the sequel. The value $\hat{\alpha}$ is easily found numerically, as implemented in the R package EbayesThresh, see [13]. As noted in [12], the posterior median $\hat{\alpha}^{med}(X_i)$ displays excellent behaviour in simulations. However, the entire posterior distribution $\Pi_{\hat{\alpha}}[\cdot \mid X]$ has not been studied so far. It turns out that the behaviour of the posterior median does not always reflect the behaviour of the complete posterior, as is seen in the next subsection.

2.2. Suboptimality of the Laplace slab for the complete EB posterior distribution

Theorem 1. *Let Π_α be the spike and slab prior distribution (5) with slab distribution G equal to the Laplace distribution $Lap(1)$. Let $\Pi_{\hat{\alpha}}[\cdot \mid X]$ be the corresponding plug-in posterior distribution given by (6), with $\hat{\alpha}$ chosen by the empirical Bayes procedure (10). There exist $D > 0$, $N_0 > 0$, and $c_0 > 0$ such that, for any $n \ge N_0$ and any s_n with $1 \le s_n \le c_0 n$, there exists $\theta_0 \in \ell_0[s_n]$ such that,*

$$E_{\theta_0} \int \|\theta - \theta_0\|^2 d\Pi_{\hat{\alpha}}[\theta \mid X] \ge D s_n e^{\sqrt{\log(n/s_n)}}.$$

Theorem 1 implies that taking a Laplace slab leads to a suboptimal convergence rate in terms of the posterior squared L^2–moment. This result is surprising at first, as we know by (12) that the posterior median converges at optimal rate r_n. The posterior mean also converges at rate r_n uniformly over $\ell_0[s_n]$, by Theorem 1 of [12]. So at first sight it would be quite natural to expect that so does the posterior second moment.

One can naturally ask whether the suboptimality result from Theorem 1 could come from considering an integrated L^2–moment, instead of simply asking for a posterior convergence result in probability, as is standard in the posterior rates literature following [10]. We now derive a stronger result than Theorem 1 under the mild condition $s_n \gtrsim \log^2 n$. The fact that the result is stronger follows from bounding from below the integral in the display of Theorem 1 by the integral restricted to the set where $\|\theta - \theta_0\|^2$ is larger than the target lower bound rate.

Theorem 2. *Under the same notation as in Theorem 1, if Π_α is a spike and slab distribution with as slab G the Laplace distribution, there exists $m > 0$ such that for any s_n with $s_n/n \to 0$ and $\log^2 n = O(s_n)$ as $n \to \infty$, there exists $\theta_0 \in \ell_0[s_n]$ such that, as $n \to \infty$,*

$$E_{\theta_0} \Pi_{\hat\alpha} \left[\, \|\theta - \theta_0\|^2 \leq m s_n e^{\sqrt{2 \log (n/s_n)}} \,\big|\, X \,\right] = o(1).$$

Theorem 2, by providing a lower bound in the spirit of [3], shows that the answer to the above question is negative, and for a Laplace slab, the plug-in posterior $\Pi_{\hat\alpha}[\cdot \,|\, X]$ does not converge at minimax rate uniformly over $\ell_0[s_n]$.

Note that the suboptimality occuring here does not result from an artificially constructed example (we work under exactly the same framework as [12]) and that this has important (negative) consequences for construction of credible sets. Due to the rate-suboptimality of the EB Laplace-posterior, typical credible sets derived from it (such as, e.g., taking quantiles of a recentered posterior second moment) will inherit the suboptimality in terms of their diameter, and therefore will not be of optimal size. Fortunately, it is still possible to achieve optimal rates for certain spike and slab EB posteriors: the previous phenomenon indeed disappears if the tails of the slab in the prior distribution are heavy enough, as seen in the next subsection.

2.3. Optimal posterior convergence rate for the EB spike and Cauchy slab

The next result considers Cauchy tails, although other examples can be covered, as discussed below. In the sequel, we abbreviate by SAS prior a spike and slab prior with Cauchy slab.

Theorem 3. *Let Π_α be the SAS prior distribution (5) with slab distribution G equal to the standard Cauchy distribution. Let $\Pi_{\hat\alpha}[\cdot \,|\, X]$ be the corresponding plug-in posterior distribution given by (6), with $\hat\alpha$ chosen by the empirical Bayes procedure (10). There exist $C > 0$, $N_0 > 0$, and $c_0, c_1 > 0$ such that, for any $n \geq N_0$, for any s_n such that (11) is satisfied for such c_0, c_1,*

$$\sup_{\theta_0 \in \ell_0[s_n]} E_{\theta_0} \int \|\theta - \theta_0\|^2 d\Pi_{\hat\alpha}(\theta \,|\, X) \leq C r_n.$$

If one only assumes $s_n \leq c_0 n$ in (11), then the last statement holds with the bound $C r_n$ replaced by $C r_n + C \log^3 n$.

Theorem 3 confirms that the empirical Bayes plug-in posterior, with $\hat\alpha$ chosen by marginal maximum likelihood, converges at optimal rate with precise logarithmic factor, at least under the mild condition (11), if tails of the slab distribution are heavy enough. Inspection of the proof of Theorem 3 reveals that any slab density γ with tails of the order $x^{-1-\delta}$ with $\delta \in (0, 2)$ gives the same result. Sensibility to the tails, in particular in view of posterior convergence in terms of d_q-distances, will be further investigated in [5].

We note that the horseshoe prior on θ considered in [18]–[19] also has Cauchy-like tails, which seems to confirm that for empirical Bayes–calibrated (product–type) sparse priors, heavy tails are important to ensure optimal or near-optimal behaviour, see also the discussion [4].

The lower bound in condition (11) is specific to the estimate $\hat{\alpha}$. Note that in the very sparse regime where $s_n \leq c_1 \log^2 n$, the rate is no more than $C \log^3 n$, thus missing the minimax rate by at most a logarithmic factor. This lower bound on s_n can be removed and the minimax rate obtained over the whole range of sparsities s_n if one modifies slightly $\hat{\alpha}$, where the estimator is changed if $\hat{\alpha}$ is too close to the lower boundary of the maximisation interval, see Section 2.6.

2.4. Posterior convergence for the EB spike and slab LASSO

Now consider the following prior on θ with fixed parameter $\alpha \in [0, 1]$

$$\Pi_\alpha \sim \otimes_{i=1}^n (1 - \alpha)G_0(\cdot) + \alpha G_1(\cdot), \tag{13}$$

where for $k = 0, 1$, G_k is given by

$$G_0 = \mathrm{Lap}(\lambda_0), \quad G_1 = \left\{ \begin{array}{l} \mathrm{Lap}(\lambda_1) \\ \mathrm{or} \\ \mathrm{Cauchy}(1/\lambda_1), \end{array} \right.$$

which leads to the spike and slab LASSO prior of [17] in the case of a Laplace G_1, and to a heavy-tailed variant of the spike and slab LASSO if G_1 is $\mathrm{Cauchy}(1/\lambda_1)$, that is if its density is $\gamma_1(x) = (\lambda_1/\pi)(1 + \lambda_1^2 x^2)^{-1}$. In this setting γ_0, γ_1 denote the densities of G_0, G_1 with respect to Lebesgue measure. We call *SSL prior* a spike and slab LASSO prior with Cauchy slab.

By Bayes' formula the posterior distribution under (1) and (13) with fixed $\alpha \in [0, 1]$ is

$$\Pi_\alpha[\cdot \,|\, X] \sim \otimes_{i=1}^n (1 - a(X_i))G_{0,X_i}(\cdot) + a(X_i)G_{1,X_i}(\cdot), \tag{14}$$

where $g_k(x) = \phi * G_k(x) = \int \phi(x - u)dG_k(u)$ is the convolution of ϕ and G_k at point $x \in \mathbb{R}$ for $k = 0, 1$, the posterior weight $a(X_i)$ is defined through the function $a(\cdot)$ given by

$$a(x) = a_\alpha(x) = \frac{\alpha g_1(x)}{(1 - \alpha)g_0(x) + \alpha g_1(x)},$$

and if G_k has density γ_k with respect to Lebesgue measure, the distribution G_{k,X_i} has density

$$\gamma_{k,X_i}(\cdot) := \frac{\phi(X_i - \cdot)\gamma_k(\cdot)}{g_k(X_i)}.$$

In slight abuse of notation, we keep the same notation in the case of the SSL prior for quantities such as $a(x)$ or $\hat{\alpha}$ below, as it will always be clear from the context which prior we work with.

We consider the following specific choices for the constants λ_0, λ_1

$$\begin{cases} \lambda_0 &= L_0 n, & L_0 = 5\sqrt{2\pi}, \\ \lambda_1 &= L_1, & L_1 = 0.05. \end{cases} \tag{15}$$

The choice of the constants L_0, L_1 is mostly for technical convenience, and is similar to that of, e.g. Corollary 5.2 in [16]. Any other constant L_0 (resp. L_1) larger (resp. smaller) than the above value also works for the following result. The above numerical values may not be optimal.

Let $\hat{\alpha}$ be defined as the maximiser of the log-marginal likelihood,

$$\hat{\alpha} = \operatorname*{argmax}_{\alpha \in [\mathcal{C} \log n/n, 1]} \ell_n(\alpha; X), \tag{16}$$

for $\mathcal{C} = \mathcal{C}_0(\gamma_0, \gamma_1)$ a large enough constant to be chosen below (this ensures that $\hat{\alpha}$ belongs to an interval on which we can verify that β is increasing, see (40)). This time we do not have access to the threshold t, since for the SSL prior the posterior median is not a threshold estimator, so here $\mathcal{C} \log n/n$ plays the role of an approximated version of α_n in (10).

Theorem 4. *Let Π_α be the SSL prior distribution (13) with Cauchy slab and parameters (λ_0, λ_1) given by (15). Let $\Pi_{\hat{\alpha}}[\cdot \mid X]$ be the corresponding plug-in posterior distribution given by (14), with $\hat{\alpha}$ chosen by the empirical Bayes procedure (16). There exist $C > 0$, $N_0 > 0$, and $c_0, c_1 > 0$ such that, for any $n \geq N_0$, for any s_n such that (11) is satisfied for such c_0, c_1, then*

$$\sup_{\theta_0 \in \ell_0[s_n]} E_{\theta_0} \int \|\theta - \theta_0\|^2 d\Pi_{\hat{\alpha}}(\theta \mid X) \leq C s_n \log n.$$

If one only assumes $s_n \leq c_0 n$ in (11), then the last bound holds with $C s_n \log n$ replaced by $C(s_n \log n + \log^3 n)$.

This result is an SSL version of Theorem 3. It shows that a spike and slab LASSO prior with heavy-tailed slab distribution and empirical Bayes choice of the weight parameter leads to a nearly optimal contraction rate for the entire posterior distribution. Hence it provides a theoretical guarantee of a fully data-driven procedure of calibration of the smoothing parameter in SSL priors.

2.5. A brief numerical study

Theorems 1–2 imply that the posterior distribution for the spike and slab prior and Laplace(1) slab does not converge at optimal rate and the discrepancy between the actual rate and the minimax rate for some 'bad' θ_0s is at least of order

$$R_n = \frac{\exp\left(\sqrt{2\log(n/s_n)}\right)}{\log(n/s_n)},$$

up to a multiplicative constant factor, as both lower and upper bounds are up to a constant. Note that R_n grows more slowly than a polynomial in n/s_n, so

the sub-optimality effect will typically be only visible for quite large values of n/s_n. For instance, if $n = 10^4$ and $s_n = 10$, one has $R_n \approx 6$, which is quite small given that an extra multiplicative constant is also involved.

For the present simulation study we took $n = 10^7$, $s_n = 10$, for which $R_n \approx 13.9$, and the non-zero values of θ_0 equal to $\{2\log(n/s_n)\}^{1/2}$, as the lower bound proof of Theorems 1–2 suggests. We computed $\hat{\alpha}$ using the package EBayesThresh of Johnstone and Silverman [13] and computed $\int \|\theta - \theta_0\|_2^2 d\Pi_{\hat{\alpha}}(X)$ using its explicit expression, which can be obtained in closed form for a Laplace slab, with similar computations as in [13], Section 6.3. We then took the empirical average over 100 repetitions to estimate the target expectation $R_2 := E_{\theta_0} \int \|\theta - \theta_0\|_2^2 d\Pi_{\hat{\alpha}}(X)$. We first took $\gamma = \text{Lap}(1)$ a standard Laplace slab and obtained $\hat{R}_2 \approx 1110$. For comparison, we computed the empirical quadratic risk $\hat{R}_{mean}$ for the posterior mean (approximating $E_{\theta_0}\|\hat{\theta}^{mean} - \theta_0\|^2$) and $\hat{R}_{median}$ the posterior median of the same posterior, obtaining $\hat{R}_{mean} \approx 158$ and $\hat{R}_{median} \approx 167$. So, in this case $\hat{R}_2$ is already 6 to 7 times larger than the risk of either mean or median.

To further illustrate the 'blow-up' in the rate for the posterior second moment R_2, we took a Laplace slab $\text{Lap}(a)$ with inverse-scale parameter a, for which the numerator in the definition of R_n becomes $\exp\{a\sqrt{2\log(n/s_n)}\}$ (let us also note that the multiplicative constant we refer to above also depends on a). The same simulation experiment as above was conducted, with the standard Laplace slab replaced by a $\text{Lap}(a)$ slab, for different values of a. The numerical results are presented in Table 1, which feature a noticeable increase in the second moment $\hat{R}_2$, while the risks of posterior mean and median stay around the same value, as expected.

TABLE 1
Empirical risks $\hat{R}_2, \hat{R}_{med}, \hat{R}_{mean}$ for Laplace slabs $Lap(a)$ and $a \in [0.5, 3.5]$

a	0.5	1	1.5	2	2.5	3	3.5
Second moment	394	1110	2847	5716	8093	16530	34791
Median	173	167	169	174	185	209	219
Mean	157	158	166	172	182	224	336

We also performed the same experiments for the quasi-Cauchy slab prior introduced in [12]-[13] (it is very close to the standard Cauchy slab – in particular it has the same Cauchy tails – but more convenient from the numerical perspective, see [13], Section 6.4). We found $\hat{R}^{median} \approx 192$, $\hat{R}^{mean} \approx 191$ for the posterior mean and $\hat{R}_2 \approx 287$ for the posterior second moment. This time, as expected, the posterior second moment is not far from the two other risks.

2.6. Modified empirical Bayes estimator

For $n \geq 3$ and $A \geq 0$, let us set $t_n^2 = 2\log n - 5\log\log n$ and $t_A = \sqrt{2(1 + A)\log n}$. For Π_α the SAS prior with a Cauchy slab, let as before $t(\alpha)$ be the posterior median threshold for fixed α. It is not hard to check that $t(\cdot)$ is continuous and strictly decreasing so has an inverse (see [12], Section 5.3). In a similar fashion

as in [12], Section 4, let us introduce a modified empirical Bayes estimator as, for $A \geq 0$ and $\hat{t} := t(\hat{\alpha})$, $\alpha_A := t^{-1}(t_A)$,

$$\hat{\alpha}_A = \begin{cases} \hat{\alpha}, & \text{if } \hat{t} \leq t_n, \\ \alpha_A, & \text{if } \hat{t} > t_n. \end{cases} \tag{17}$$

Theorem 5. *Let Π_α be the SAS prior distribution with slab distribution G equal to the standard Cauchy distribution. For a fixed $A > 0$, let $\Pi_{\hat{\alpha}_A}[\cdot \,|\, X]$ be the corresponding plug-in posterior distribution, with $\hat{\alpha}_A$ the* modified *estimator* (17). *There exist $C, c_0 > 0$, $N_0 > 0$, such that, for any $n \geq N_0$, for any s_n such that $s_n \leq c_0 n$,*

$$\sup_{\theta_0 \in \ell_0[s_n]} E_{\theta_0} \int \|\theta - \theta_0\|^2 d\Pi_{\hat{\alpha}}(\theta \,|\, X) \leq C r_n.$$

Theorem 5 shows that the plug-in SAS posterior distribution using the modified estimator (17), $A > 0$, and a Cauchy slab attains the minimax rate of convergence r_n even in the very sparse regime $s_n \lesssim \log^2 n$, for which the unmodified estimate of Theorem 3 may lose a logarithmic factor.

2.7. Discussion

In this paper, we have developped a theory of empirical Bayes choice of the hyperparameter of spike and slab prior distributions. It extends the work of Johnstone and Silverman [12] in that here the complete EB posterior distribution is considered. One important message is that such a generalisation preserves optimal convergence rates at the condition of taking slab distributions with heavy enough tails. If the tails of the slab are only moderate (e.g. Laplace), then the complete EB posterior rate may be suboptimal. This is in contrast with the hierarchical case considered in [6], where a Laplace slab combined with a Beta distributed prior on α was shown to lead to an optimal posterior rate. On the one hand, the empirical Bayes method often leads to simpler or/and more easily tractable practical algorithms; on the other hand, we have illustrated here that the complete EB posterior may in some cases need slightly stronger conditions to conserve optimal theoretical guarantees. This phenomenon had not been pointed out so far in the literature, to the best of our knowledge.

We also note that Theorem 3 (or Theorem 5 if one allows for very sparse signals) enables one to recover the optimal form of the logarithmic factor $\log(n/s_n)$ in the minimax rate. This entails significant work, as one needs to control the empirical Bayes weight estimate $\hat{\alpha}$ both from above *and below*. This could work too in the SSL setting of Theorem 4, although this seems to need substantial extra technical work.

Looking at Theorems 1 and 2, it is natural to wonder why the Empirical Bayes approach fails for the Laplace slab where the full Bayes approach succeeds as seen in [6] Theorem 2.2. The reason why the hierarchical Bayes version works also for γ Laplace is the extra penalty in model size induced by the hierarchical

prior on dimension. Indeed, in the full Bayes approach, the posterior distribution of α given X has density

$$f_{\alpha \mid X}(\alpha) \propto p(X \mid \alpha)\pi(\alpha),$$

where $p(X \mid \alpha)$ is the marginal density one maximises when considering the MMLE $\hat\alpha$. Hence adding a term $\log \pi(\alpha)$ for well-chosen π – for instance that arising from a Beta$(1, n+1)$ prior on α as considered in [6] – to the log-marginal likelihood one maximises forces $\hat\alpha$ to concentrate on smaller values. For instance, in the present setting, one could consider a penalised log-marginal maximum likelihood, which would force the estimate $\hat\alpha$ to concentrate on slightly smaller values, which would allow one to avoid the extra $e^{\sqrt{\log n/s_n}}$ term arising in Theorems 1–2.

The present work can also serve as a basis for constructing confidence regions using spike-and-slab posterior distributions. This question is considered in the forthcoming paper [5].

3. Proofs for the spike and slab prior

Let us briefly outline the ingredients of the proofs to follow. For Theorems 1 and 3, our goal is to bound the expected posterior risk $R_n(\theta_0) = E_{\theta_0} \int \|\theta - \theta_0\|^2 d\Pi_{\hat\alpha}(\theta \mid X)$. There are three main tools. First, after introducing notation and basic bounds in Section 3.1, bounds on the posterior risk for fixed α are given in Section 3.2, as well as corresponding bounds for random α. Let us note that the corresponding upper bounds are different from those obtained on the quadratic risk for the posterior median in [12] (and in fact, must be, in view of the negative result in Theorem 1). Second, inequalities on moments of the score function are stated in Section 3.3. As a third tool, we obtain deviation inequalities on the location of $\hat\alpha$ in Section 3.4. One of the bounds sharpens the corresponding bound from [12] in case the signal belongs to the nearly-black class $\ell_0[s_n]$ which we assume here.

Proofs of Theorems 1 and 3 are given in Sections 3.5 and 3.6. For Theorem 3, we also needed to slightly complete the proof of one of the inequalities on thresholds stated in [12], see Lemma 11. The proof of Theorem 2, which uses ideas from both previous proofs, is given in Section 3.7. Proofs of technical lemmas for the SAS prior are given in Section 4.

3.1. Notation and tools for the SAS prior

Expected posterior L^2–squared risk. For a fixed weight α, the posterior distribution of θ is given by (6). On each coordinate, the mixing weight $a(X_i)$ is given by (7) and the density of the non-zero component γ_{X_i} by (8). In the sequel we will obtain bounds on the following quantity, already for a given $\alpha \in [0, 1]$,

$$\int \|\theta - \theta_0\|^2 d\Pi_\alpha(\theta \mid X) = \sum_{i=1}^n \int (\theta_i - \theta_{0,i})^2 d\Pi_\alpha(\theta_i \mid X_i).$$

To do so, we study $r_2(\alpha, \mu, x) := \int (u - \mu)^2 d\pi_\alpha(u \,|\, x)$, where $\pi_\alpha(\cdot \,|\, x) \sim (1 - a(x))\delta_0 + a(x)\gamma_x(\cdot)$. By definition

$$r_2(\alpha, \mu, x) = (1 - a(x))\mu^2 + a(x)\int (u - \mu)^2 \gamma_x(u)du.$$

This quantity is controlled by $a(x)$ and the term involving γ_x. From the definition of $a(x)$, bounding the denominator from below by one of its two components, and using $a(x) \leq 1$ yields, for any real x and $\alpha \in [0, 1]$,

$$\alpha \frac{g}{g \vee \phi}(x) \leq a(x) \leq 1 \wedge \frac{\alpha}{1 - \alpha}\frac{g}{\phi}(x). \tag{18}$$

The marginal likelihood in α. By definition, the empirical Bayes estimate $\hat{\alpha}$ in (10) maximises the logarithm of the marginal likelihood in α in (9). In case the maximum is not taken at the boundary, $\hat{\alpha}$ is a zero of the derivative (score) of the previous likelihood. Its expression is $S(\alpha) = \sum_{i=1}^n \beta(X_i, \alpha)$, where following [12] we set, for $0 \leq \alpha \leq 1$ and any real x,

$$\beta(x, \alpha) = \frac{\beta(x)}{1 + \alpha\beta(x)}, \qquad \beta(x) = \frac{g}{\phi}(x) - 1.$$

The study of $\hat{\alpha}$ below uses in a crucial way the first two moments of $\beta(X_i, \alpha)$, so we introduce the corresponding notation next. Let E_τ, for $\tau \in \mathbb{R}^n$, denote the expectation under $\theta_0 = \tau$. Define

$$\tilde{m}(\alpha) = -E_0\beta(X, \alpha) \tag{19}$$

and further denote

$$m_1(\tau, \alpha) = E_\tau[\beta(X, \alpha)] = \int_{-\infty}^{\infty} \beta(t, \alpha)\phi(t - \tau)dt.$$
$$m_2(\tau, \alpha) = E_\tau[\beta(X, \alpha)^2].$$

The thresholds $\zeta(\alpha)$, $\tilde{\tau}(\alpha)$ and $t(\alpha)$. Following [12], we introduce several useful thresholds. From Lemma 1 in [12], we know that g/ϕ, and therefore $\beta = g/\phi - 1$, is a strictly increasing function on $\mathbb{R}^+$. It is also continuous, so given α, a pseudo-threshold $\zeta = \zeta(\alpha)$ can be defined by

$$\beta(\zeta) = \frac{1}{\alpha}. \tag{20}$$

Further one can also define $\tau(\alpha)$ as the solution in x of

$$\Omega(x, \alpha) := \frac{a(x)}{1 - a(x)} = \frac{\alpha}{1 - \alpha}\frac{g}{\phi}(x) = 1.$$

Equivalently, $a(\tau(\alpha)) = 1/2$. Also, $\beta(\tau(\alpha)) = \alpha^{-1} - 2$ so $\tau(\alpha) \leq \zeta(\alpha)$. Define α_0 as $\tau(\alpha_0) = 1$ and set

$$\tilde{\tau}(\alpha) = \tau(\alpha \wedge \alpha_0). \tag{21}$$

Recall from Section 2 that $t(\alpha)$ is the threshold associated to the posterior median for given α. It is shown in [12], Lemma 3, that $t(\alpha) \leq \zeta(\alpha)$. Finally, the following bound in terms of $\tau(\alpha)$, see [12] p. 1623, is also useful for large x,

$$1 - a(x) \leq \mathbb{1}_{\{|x| \leq \tilde{\tau}(\alpha)\}} + e^{-\frac{1}{2}(|x| - \tilde{\tau}(\alpha))^2} \mathbb{1}_{\{|x| > \tilde{\tau}(\alpha)\}}. \tag{22}$$

3.2. Posterior risk bounds

Recall the notation $r_2(\alpha, \mu, x) = \int (u - \mu)^2 d\Pi_\alpha(u)$.

Lemma 1. *Let γ be the Cauchy or Laplace density. For any x and $\alpha \in [0, 1/2]$,*

$$r_2(\alpha, 0, x) \leq C\Big[1 \wedge \frac{\alpha}{1 - \alpha} \frac{g}{\phi}(x)\Big](1 + x^2)$$
$$r_2(\alpha, \mu, x) \leq (1 - a(x))\mu^2 + Ca(x)((x - \mu)^2 + 1).$$

Let γ be the Cauchy density. For any real x and $\alpha \in [0, 1/2]$,

$$E_0 r_2(\alpha, 0, x) \leq C\tau(\alpha)\alpha$$
$$E_\mu r_2(\alpha, \mu, x) \leq C(1 + \tilde{\tau}(\alpha)^2).$$

The following lower bound is used in the proof of Theorem 1.

Lemma 2. *Let γ be the Laplace density. There exists $C_0 > 0$ such that, for $x \in \mathbb{R}$ and $\alpha \in [0, 1]$*
$$r_2(\alpha, 0, x) \geq C_0 \alpha.$$

We now turn to bounding $r_2(\hat{\alpha}, \mu, x)$. This is the quantity $r_2(\alpha, \mu, x)$, where α (which comes in via $a(x) = a_\alpha(x)$) is replaced by $\hat{\alpha}$. This is done with the help of the threshold $\tilde{\tau}(\alpha)$.

Lemma 3 (no signal or small signal)**.** *Let γ be the Cauchy density. Let α be a fixed non-random element of $(0, 1)$. Let $\hat{\alpha}$ be a random element of $[0, 1]$ that may depend on $x \sim \mathcal{N}(0, 1)$ and on other data. Then there exists $C_1 > 0$ such that*
$$Er_2(\hat{\alpha}, 0, x) \leq C_1 \Big[\alpha \tilde{\tau}(\alpha) + P(\hat{\alpha} > \alpha)^{1/2}\Big].$$

There exists $C_2 > 0$ such that for any real μ, if $x \sim \mathcal{N}(\mu, 1)$,

$$Er_2(\hat{\alpha}, \mu, x) \leq \mu^2 + C_2.$$

Lemma 4 (signal)**.** *Let γ be the Cauchy density. Let α be a fixed non-random element of $(0, 1)$. Let $\hat{\alpha}$ be a random element of $[0, 1]$ that may depend on $x \sim \mathcal{N}(\mu, 1)$ and on other data and such that $\tilde{\tau}(\hat{\alpha})^2 \leq d \log(n)$ with probability 1 for some $d > 0$. Then there exists $C_2 > 0$ such that for all real μ,*

$$Er_2(\hat{\alpha}, \mu, x) \leq C_2 \Big[1 + \tilde{\tau}(\alpha)^2 + (1 + d \log n)P(\hat{\alpha} < \alpha)^{1/2}\Big].$$

3.3. Moments of the score function

The next three lemmas are borrowed from [12] and apply to any density γ such that $\log \gamma$ is Lipschitz on $\mathbb{R}$ and satisfies

$$\gamma(y)^{-1} \int_y^\infty \gamma(u)du \approx y^{\kappa-1}, \quad \text{as } y \to \infty. \tag{23}$$

Both Cauchy and Laplace densities satisfy (23), with $\kappa = 2$ and $\kappa = 1$ respectively, and their logarithm is Lipschitz.

Lemma 5. *For $\kappa \in [1,2]$ as in (23), as $\alpha \to 0$,*

$$\tilde{m}(\alpha) \asymp \zeta^{\kappa-1}g(\zeta).$$

Also, the function $\alpha \to \tilde{m}(\alpha)$ is nonnegative and increasing in α.

Lemma 6. *The function $\alpha \to m_1(\mu,\alpha)$ is decreasing in α. Also, $m_1(\zeta,\alpha) \sim 1/(2\alpha)$ as $\alpha \to 0$. For small enough α,*

$$m_2(\mu,\alpha) \leq C\alpha^{-1}m_1(\mu,\alpha), \quad \mu \geq 1.$$

Lemma 7. *There exist a constant c_1 such that for any x and α,*

$$|\beta(x,\alpha)| \leq \frac{1}{\alpha \wedge c_1},$$

and constants c_2, c_3, c_4 such that for any α, and κ as in (23),

$$\begin{aligned}
m_1(\mu,\alpha) &\leq -\tilde{m}(\alpha) + c_2\zeta(\alpha)\mu^2, & |\mu| &\leq 1/\zeta(\alpha) \\
m_1(\mu,\alpha) &\leq (\alpha \wedge c_3)^{-1} & &\text{for all } \mu
\end{aligned}$$

and

$$\begin{aligned}
m_2(\mu,\alpha) &\leq c_4 \frac{\tilde{m}(\alpha)}{\zeta(\alpha)^\kappa \alpha} & |\mu| &\leq 1/\zeta = 1/\zeta(\alpha) \\
m_2(\mu,\alpha) &\leq (\alpha \wedge c_3)^{-2} & &\text{for all } \mu.
\end{aligned}$$

3.4. In-probability bounds for $\hat{\alpha}$

Lemma 9 below implies that, for any possible θ_0, the estimate $\hat{\alpha}$ is smaller than a certain α_1 with high probability. One can interpret this as saying that $\hat{\alpha}$ does not lead to too much undersmoothing (i.e. too many nonzero coefficients). On the other hand, if there is enough signal in a certain sense, $\hat{\alpha}$ does not lead to too much oversmoothing (i.e. too many zero coefficients), see Lemma 10.

Although we generally follow the approach of [12], there is one significant difference. One needs a fairly sharp bound on α_1 below. Using the definition from [12] would lead to a loss in terms of logarithmic factors for the posterior L^2–squared moment. So we work with a somewhat different α_1, and shall thus

provide a detailed proof of the corresponding Lemma 9. For the oversmoothing case, one can borrow the corresponding Lemma of [12] as is.

Let $\alpha_1 = \alpha_1(d)$ be defined as the solution of the equation, with $\eta_n = s_n/n$,

$$d\alpha_1 \tilde{m}(\alpha_1) = \eta_n, \tag{24}$$

where d is a constant to be chosen (small enough for Lemma 9 to hold). A solution of (24) exists, as using Lemma 5, $\alpha \to \alpha\tilde{m}(\alpha)$ is increasing in α, and equals 0 at 0. Also, provided η_n is small enough, α_1 can be made smaller than any given arbitrary constant. The corresponding threshold ζ_1 is defined by $\beta(\zeta_1) = \alpha_1^{-1}$. From Lemma 5, we have $\tilde{m}(\alpha_1) \asymp \zeta g(\zeta_1)$ if γ is Cauchy and $\tilde{m}(\alpha_1) \asymp g(\zeta_1)$ if γ is Laplace.

Lemma 8. *Let κ be the constant in* (23)*. Let α_1 be defined by* (24) *for d a given constant and let ζ_1 be given by $\beta(\zeta_1) = \alpha_1^{-1}$. Then there exist real constants c_1, c_2 such that for large enough n,*

$$\log(n/s_n) + c_1 \leq \frac{\zeta_1^2}{2} \leq \log(n/s_n) + \frac{\kappa-1}{2} \log\log n + c_2,$$

with κ as in (23)*. Also, $\zeta_1^2 \sim 2\log(n/s_n)$ as n/s_n goes to ∞.*

Lemma 9. *Let α_1 be defined by* (24) *for d a given small enough constant and let ζ_1 be given by $\beta(\zeta_1) = \alpha_1^{-1}$. Suppose* (11) *holds. Then for some constant $C > 0$,*

$$\sup_{\theta \in \ell_0[s_n]} P_\theta[\hat{\zeta} < \zeta_1] \leq \exp(-Cs_n).$$

For the oversmoothing case, one denotes the proportion of signals above a level τ by

$$\tilde{\pi}(\tau; \mu) = \frac{1}{n}\#\{i : |\mu_i| \geq \tau\}. \tag{25}$$

We also set, recalling that α_0 is defined via $\tau(\alpha_0) = 1$,

$$\alpha(\tau, \pi) = \sup\{\alpha \leq \alpha_0 : \pi m_1(\tau, \alpha) \geq 2\tilde{m}(\alpha)\}. \tag{26}$$

One defines $\zeta_{\tau,\pi}$ as the corresponding pseudo-threshold $\beta^{-1}(\alpha(\tau, \pi)^{-1})$.

Lemma 10 ([12], Lemma 11)**.** *There exists C and π_0 such that if $\pi < \pi_0$, then for all $\tau \geq 1$,*

$$\sup_{\theta:\ \tilde{\pi}(\tau;\theta)\geq\pi} P_\theta[\hat{\zeta} > \zeta_{\tau,\pi}] \leq \exp\{-Cn\phi(\zeta_{\tau,\pi})\}.$$

3.5. Proof of Theorem 1

Proof. Let α^* be defined as the solution in α of the equation,

$$\alpha\tilde{m}(\alpha) = \eta_n/4, \tag{27}$$

where $\eta_n = s_n/n$ (that is $\alpha^* = \alpha_1(d)$ with $d = 4$ in (24)). Let ζ^* be defined via $\beta(\zeta^*) = \alpha^*$.

Let θ_0 be the specific signal defined by, for α^*, ζ^* as in (27),

$$\theta_{0,i} = \begin{cases} \zeta^*, & 1 \leq i \leq s_n \\ 0, & s_n < i \leq n \end{cases}.$$

Using Lemma 5, one gets $\tilde{m}(\alpha^*) \asymp g(\zeta^*) \asymp \gamma(\zeta^*)$ as $\zeta^* \to \infty$. Lemma 8 implies $\zeta^{*2} \geq 2 \log(1/\eta_n) + C$, for C a possibly negative constant. Combining this with the definition $\gamma(\zeta^*) = e^{-\zeta^*}/2$ leads to

$$\alpha^* \gtrsim \eta_n e^{\sqrt{\log(1/\eta_n)}}, \tag{28}$$

for c_0 in (11) small enough to have $2 \log(1/\eta_n) + C \geq \log(1/\eta_n)$. We next prove that, for $\hat{\alpha}$ given by (10), for small enough $c > 0$,

$$P_{\theta_0} [\hat{\alpha} < \alpha^*] \leq e^{-cs_n}. \tag{29}$$

If $\alpha^* \leq \alpha_n$ the probability at stake is 0, as $\hat{\alpha}$ belongs to $[\alpha_n, 1]$ by definition. For $\alpha^* > \alpha_n$, we have $\{\hat{\alpha} < \alpha^*\} = \{S(\alpha^*) < 0\}$. With $A = \sum_{i=1}^{n} m_1(\mu_i, \alpha^*)$,

$$P_{\theta_0} [\hat{\alpha} < \alpha^*] = P_{\theta_0} [S(\alpha^*) < 0] = P_{\theta_0} \left[\sum_{i=1}^{n} \beta(\theta_{0,i} + Z_i, \alpha^*) - m_1(\theta_{0,i}, \alpha^*) < -A \right]$$

Setting $W_i = m_1(\theta_{0,i}, \alpha^*) - \beta(\theta_{0,i} + Z_i, \alpha^*)$, we have $|W_i| \leq 2C/\alpha^* =: M$ and W_i are independent. So by Bernstein's inequality,

$$P_{\theta_0} \left[\sum_{i=1}^{n} W_i > A \right] \leq \exp \left[-\frac{1}{2} \frac{A^2}{V + \frac{1}{3}MA} \right],$$

where V is an upper-bound for $\sum_{i=1}^{n} \mathrm{Var}(W_i)$. The term A equals

$$A = (n - s_n)(-\tilde{m}(\alpha^*)) + s_n m_1(\zeta^*, \alpha^*).$$

The function $\alpha \to \alpha \tilde{m}(\alpha)$ is increasing, as $\tilde{m}(\cdot)$ is (Lemma 5), so by its definition (27), α^* can be made smaller than any given positive constant, provided c_0 in (11) is small enough, ensuring $\eta_n = s_n/n$ is small enough. Using Lemma 6, $m_1(\zeta, \alpha) \sim 1/(2\alpha)$ as $\alpha \to 0$. So, using (27), one obtains, for small enough c_0,

$$A \geq \frac{s_n}{3\alpha^*} - \frac{s_n}{4\alpha^*} = \frac{s_n}{12\alpha^*}.$$

On the other hand, the last part of Lemma 7 implies

$$V \leq \sum_{i \notin S_0} m_2(0, \alpha^*) + \sum_{i \in S_0} m_2(\zeta^*, \alpha^*)$$

$$\leq C(n - s_n) \frac{\tilde{m}(\alpha^*)}{\zeta^* \alpha^*} + C \frac{s_n}{\alpha^{*2}}.$$

Using the definition of α^*, one deduces $V \lesssim s_n/\alpha^{*2}$ and from this

$$\frac{V}{A^2} + \frac{MA}{3A^2} \lesssim \frac{1}{s_n},$$

which in turn implies (29), as then $P_{\theta_0}[\sum_{i=1}^n W_i > A] \leq \exp[-cs_n]$. Next one writes

$$\int \|\theta - \theta_0\|^2 d\Pi_{\hat{\alpha}}[\theta \mid X] \geq \int \|\theta - \theta_0\|^2 d\Pi_{\hat{\alpha}}[\theta \mid X] \mathbb{1}_{\hat{\alpha} \geq \alpha^*}$$

$$\geq \sum_{i \notin S_0} \int \theta_i^2 d\Pi_{\hat{\alpha}}(\theta \mid X) \mathbb{1}_{\hat{\alpha} \geq \alpha^*}$$

Lemma 2 implies, for any possibly data-dependent weight α, that

$$\int \theta_i^2 d\Pi_\alpha(\theta \mid X) \gtrsim \alpha,$$

so

$$\int \|\theta - \theta_0\|^2 d\Pi_{\hat{\alpha}}[\theta \mid X] \geq (n - s_n)\hat{\alpha}\mathbb{1}_{\hat{\alpha} \geq \alpha^*} \geq (n - s_n)\alpha^*\mathbb{1}_{\hat{\alpha} \geq \alpha^*}.$$

As $(n - s_n)\alpha^* P_{\theta_0}[\hat{\alpha} \geq \alpha^*] \gtrsim Cn\alpha^*(1 - e^{-cs_n})$, an application of (28) concludes the proof. $\qquad\square$

3.6. Proof of Theorem 3

Let us decompose the risk $R_n(\theta_0) = E_{\theta_0} \int \|\theta - \theta_0\|^2 d\Pi_{\hat{\alpha}}(\theta \mid X)$ according to whether coordinates of θ correspond to a 'small' or 'large' signal, the threshold being $\zeta_1 = \beta^{-1}(\alpha_1^{-1})$, with α_1 defined in (24). One can write

$$R_n(\theta_0) = \Big[\sum_{i:\ \theta_{0,i}=0} + \sum_{i:\ 0<|\theta_{0,i}|\leq\zeta_1} + \sum_{i:\ |\theta_{0,i}|>\zeta_1} \Big] E_{\theta_0} \int (\theta_i - \theta_{0,i})^2 d\Pi_{\hat{\alpha}}(\theta_i \mid X).$$

We next use the first part of Lemma 3 with $\alpha = \alpha_1$ and the second part of the Lemma to obtain, for any θ_0 in $\ell_0[s_n]$,

$$\Big[\sum_{i:\ \theta_{0,i}=0} + \sum_{i:\ 0<|\theta_{0,i}|\leq\zeta_1} \Big] E_{\theta_0} \int (\theta_i - \theta_{0,i})^2 d\Pi_{\hat{\alpha}}(\theta_i \mid X)$$

$$\leq C_1 \sum_{i:\ \theta_{0,i}=0} [\alpha_1 \tau(\alpha_1) + P_{\theta_0}(\hat{\alpha} > \alpha_1)] + \sum_{i:\ 0<|\theta_{0,i}|\leq\zeta_1} (\theta_{0,i}^2 + C)$$

$$\leq C_1 \Big[(n - s_n)\alpha_1 \tau(\alpha_1) + (n - s_n)e^{-c_1 \log^2 n} \Big] + (\zeta_1^2 + C)s_n,$$

where for the last inequality we use Lemma 9 and (11). From (24) one gets, with $\eta_n = s_n/n$,

$$n\alpha_1 \lesssim n\eta_n\zeta_1^{-1}g(\zeta_1)^{-1} \lesssim s_n\zeta_1.$$

Now using Lemma 8 and the fact that $\tau(\alpha_1) \leq \zeta_1$, one obtains that the contribution to the risk of the indices i with $|\theta_{0,i}| \leq \zeta_1$ is bounded by a constant times $s_n \log(n/s_n)$.

It remains to bound the part of the risk for indexes i with $|\theta_{0,i}| > \zeta_1$. To do so, one uses Lemma 4 with α chosen as $\alpha = \alpha_2 := \alpha(\zeta_1, \pi_1)$ and $\pi_1 = \tilde{\pi}(\zeta_1; \theta_0)$, following the definitions (25)–(26). One denotes by ζ_2 the pseudo-threshold associated to α_2. The following estimates are useful below

$$\zeta_1^2 < \zeta_2^2 \tag{30}$$

$$\pi_1 \zeta_2^2 \leq C \eta_n \log(1/\eta_n). \tag{31}$$

These are established in a similar way as in [12], but with the updated definition of α_1, ζ_1 from (24), so we include the proof below for completeness. One can now apply Lemma 4 with $\alpha = \alpha_2$,

$$\sum_{i: \, |\theta_{0,i}| > \zeta_1} E_{\theta_0} \int (\theta_i - \theta_{0,i})^2 d\Pi_{\hat{\alpha}}(\theta_i \mid X)$$

$$\leq C_2 n \pi_1 \left[1 + \zeta_2^2 + (1 + d \log n) P_{\theta_0}(\hat{\alpha} < \alpha_2)^{1/2} \right]$$

$$\leq C_2 n \pi_1 \left[1 + \zeta_2^2 + (1 + d \log n) P_{\theta_0}(\hat{\zeta} > \zeta_2)^{1/2} \right].$$

Let us verify that the term in brackets in the last display is bounded above by $C(1 + \zeta_2^2)$. If $\zeta_2 > \log n$, this is immediate by bounding $P_{\theta_0}(\hat{\zeta} > \zeta_2)$ by 1. If $\zeta_2 \leq \log n$, Lemma 10 implies $P_{\theta_0}(\hat{\zeta} > \zeta_2) \leq \exp(-Cn\phi(\zeta_2)) \leq \exp(-C\sqrt{n})$, so this is also the case. Conclude that the last display is bounded above by $Cn\pi_1(1 + \zeta_2^2) \leq C'n\pi_1\zeta_2^2$. Using (31), this term is itself bounded by $Cs_n \log(n/s_n)$, which concludes the proof of the Theorem, given (30)–(31).

We now check that (30)–(31) hold. We first compare α_1 and α_2. For small enough α, the bound on m_1 from Lemma 7 becomes $1/\alpha$, so that, using the definition (24) of α_1,

$$\frac{m_1(\zeta_1, \alpha_1)}{\tilde{m}(\alpha_1)} \leq \frac{1}{\alpha_1} \left(\frac{\eta_n}{d\alpha_1} \right)^{-1} \leq \frac{d}{\eta_n} \leq \frac{d}{\pi_1},$$

using the rough bound $\pi_1 \leq \eta_n$. Note that both functions $\tilde{m}(\cdot)^{-1}$ and $m_1(\zeta_1, \cdot)$ are decreasing via Lemmas 5–6, and so is their product on the interval where both functions are positive. As $d < 2$, by definition of α_2 this means $\alpha_2 < \alpha_1$ that is $\zeta_1 < \zeta_2$.

To prove (31), one compares ζ_2 first to a certain $\zeta_3 = \zeta(\alpha_3)$ defined by α_3 (largest) solution of

$$\bar{\Phi}(\zeta(\alpha_3) - \zeta_1) = \frac{8}{\pi_1} \alpha_3 \tilde{m}(\alpha_3),$$

with $\bar{\Phi}(x) = P[\mathcal{N}(0,1) > x]$. Using Lemma 11, which also gives the existence of ζ_3, one gets

$$\frac{m_1(\zeta_1, \alpha_3)}{\tilde{m}(\alpha_3)} \geq \frac{\frac{1}{4}\beta(\zeta_3)\bar{\Phi}(\zeta_3 - \zeta_1)}{\tilde{m}(\alpha_3)} = \frac{1}{4\alpha_3} \frac{8\alpha_3 \tilde{m}(\alpha_3)}{\pi_1 \tilde{m}(\alpha_3)} = \frac{2}{\pi_1}.$$

I. Castillo and R. Mismer

This shows, reasoning as above, that $\alpha_3 \leq \alpha_2$, that is $\zeta_2 \leq \zeta_3$. Following [12], one distinguishes two cases to further bound ζ_3.

If $\zeta_3 > \zeta_1 + 1$, using $\zeta_2^2 \leq \zeta_3^2$ and $\tilde{m}(\alpha_3) \lesssim \zeta_3 g(\zeta_3)$,

$$
\begin{aligned}
\pi_1 \zeta_2^2 &\leq \zeta_3^2 \frac{8\alpha_3 \tilde{m}(\alpha_3)}{\bar{\Phi}(\zeta_3 - \zeta_1)} \lesssim \zeta_3^3 \frac{g(\zeta_3)}{\beta(\zeta_3)} \frac{\zeta_3 - \zeta_1}{\phi(\zeta_3 - \zeta_1)} \\
&\leq C\zeta_3^4 \frac{\phi(\zeta_3)}{\phi(\zeta_3 - \zeta_1)} = C\zeta_3^4 \phi(\zeta_1) e^{-(\zeta_3 - \zeta_1)\zeta_1} \\
&\leq C(\zeta_1 + 1)^4 e^{-\zeta_1} \phi(\zeta_1),
\end{aligned}
$$

where for the last inequality we have used that $x \to x^4 e^{-(x-\zeta_1)\zeta_1}$ is decreasing for $x \geq \zeta_1 + 1$. Lemma 8 now implies that $\phi(\zeta_1) \lesssim \eta_n$. As ζ_1 goes to ∞ with n/s_n, one gets $\pi_1 \zeta_2^2 \lesssim \eta_n$.

If $\zeta_1 \leq \zeta_3 \leq \zeta_1 + 1$, let $\zeta_4 = \zeta(\alpha_4)$ with α_4 solution in α of

$$
\bar{\Phi}(1) = 8\alpha\tilde{m}(\alpha)\pi_1^{-1}.
$$

By the definition of ζ_3, since $\bar{\Phi}(1) \leq \bar{\Phi}(\zeta_3 - \zeta_1)$, we have $8\alpha_4\tilde{m}(\alpha_4) \leq 8\alpha_3\tilde{m}(\alpha_3)$ so that $\alpha_4 \leq \alpha_3$. Using Lemma 5 as before,

$$
\bar{\Phi}(1) \lesssim \frac{g(\zeta_4)}{\beta(\zeta_4)} \pi_1^{-1} \lesssim \phi(\zeta_4)\pi_1^{-1}.
$$

Taking logarithms this leads to

$$
\zeta_4^2 \leq C + 2\log(\pi_1^{-1}).
$$

In particular, $\zeta_2^2 \leq 2\log(\pi_1^{-1}) + C$. As $x \to x\log(1/x)$ is increasing, one gets, using $\pi_1 \leq \eta_n$,

$$
\pi_1 \zeta_2^2 \leq 2\eta_n \log(1/\eta_n) + C\eta_n,
$$

which concludes the verification of (30)–(31) and the proof of Theorem 3.

In checking (31), one needs a lower bound on m_1. In [12], the authors mention that it follows from their lower bound (82), Lemma 8. But this bound cannot hold uniformly for any smoothing parameter α (denoted by w in [12]), as $m_1(\mu, 0) = -\tilde{m}(w) < 0$ if $w \neq 0$. So, although the claimed inequality is correct, it does not seem to follow from (82). We state the inequality we use now, and prove it in Section 4.3.

Lemma 11. *Let* $\bar{\Phi}(t) = \int_t^\infty \phi(u)du$. *For* π_1, ζ_1 *as above, a solution* $0 < \alpha \leq \alpha_1$ *to the equation*

$$
\bar{\Phi}(\zeta(\alpha) - \zeta_1) = 8\pi_1^{-1}\alpha\tilde{m}(\alpha). \tag{32}
$$

exists. Let α_3 *be the largest such solution. Then for* c_0 *in* (11) *small enough,*

$$
m_1(\zeta_1, \alpha_3) \geq \frac{1}{4}\beta(\zeta_3)\bar{\Phi}(\alpha_3 - \zeta_1). \tag{33}
$$

3.7. Proof of Theorem 2

Let $\theta_0, \alpha^*, \zeta^*$ be defined as in the proof of Theorem 1. Below we show that the event $\mathcal{A} = \{\hat{\alpha} \in [\alpha^*, c\alpha^*]\}$, for c a large enough constant, has probability going to 1, faster than a polynomial in $1/n$. Recall from the proof of Theorem 1 that, if $\hat{\alpha} \geq \alpha^*$, so in particular on $\mathcal{A}$, we have $V_X \geq (n-s_n)\alpha^* \geq n\alpha^*/2 \geq C_1 s_n g(\zeta^*)^{-1}$. Denote

$$v_n = m s_n g(\zeta^*)^{-1}$$
$$V_X = \int \|\theta - \theta_0\|^2 d\Pi_{\hat{\alpha}}(\theta \mid X),$$

where m is chosen small enough so that $v_n \leq V_X/2$ on $\mathcal{A}$. Then,

$$\Pi_{\hat{\alpha}} \left[\|\theta - \theta_0\|^2 < v_n \mid X\right] \mathbb{1}_{\mathcal{A}} = \Pi_{\hat{\alpha}} \left[\|\theta - \theta_0\|^2 - V_X < v_n - V_X \mid X\right] \mathbb{1}_{\mathcal{A}}$$
$$\leq \Pi_{\hat{\alpha}} \left[\|\theta - \theta_0\|^2 - V_X < -V_X/2 \mid X\right] \leq 4V_X^{-2} \int \{\|\theta - \theta_0\|^2 - V_X\}^2 d\Pi_{\hat{\alpha}}(\theta \mid X),$$

where the second line follows from Markov's inequality. One now writes the L^2–norm in the previous display as sum over coordinates and one expands the square, while noting that given X the posterior $\Pi_{\hat{\alpha}}[\cdot \mid X]$ makes the coordinates of θ independent

$$\int \{\|\theta - \theta_0\|^2 - V_X\}^2 d\Pi_{\hat{\alpha}}(\theta \mid X)$$
$$= \int \sum_{i,j} \left[(\theta_i - \theta_{0,i})^2 - \int (\theta_i - \theta_{0,i})^2 d\Pi_{\hat{\alpha}}(\theta \mid X)\right]$$
$$\times \left[(\theta_j - \theta_{0,j})^2 - \int (\theta_j - \theta_{0,j})^2 d\Pi_{\hat{\alpha}}(\theta \mid X)\right] d\Pi_{\hat{\alpha}}(\theta \mid X)$$
$$= \sum_{i=1}^{n} \int \left[(\theta_i - \theta_{0,i})^2 - \int (\theta_i - \theta_{0,i})^2 d\Pi_{\hat{\alpha}}(\theta \mid X)\right]^2 d\Pi_{\hat{\alpha}}(\theta \mid X)$$
$$\leq \sum_{i=1}^{n} \int (\theta_i - \theta_{0,i})^4 d\Pi_{\hat{\alpha}}(\theta \mid X).$$

The last bound is the same as in the proof of the upper bound Theorem 3, except the fourth moment replaces the second moment. Denote $r_4(\alpha, \mu, x) = \int (u - \mu)^4 d\pi_{\hat{\alpha}}(u \mid x)$, then

$$r_4(\alpha, \mu, x) = (1 - a(x))\mu^4 + a(x) \int (u - \mu)^4 \gamma_x(u) du.$$

In a similar way as in the proof of Lemma 1, one obtains $\int (u - \mu)^4 \gamma_x(u) du \leq C(1 + (x - \mu)^4)$. Next, noting that since now γ is Laplace so g has Laplace tails, $x \to (1 + x^4)g(x)$ is integrable, proceeding as in the proof of Lemma 1, one gets

$E_0 r_4(\alpha, 0, x) \lesssim \alpha$ as well as $E_\mu r_4(\alpha, \mu, x) \lesssim 1 + \tilde{\tau}(\alpha)^4$, for any fixed α. Similarly as in Lemmas 3–4, one then derives the following random α bounds

$$Er_4(\hat{\alpha}, 0, x) \lesssim c\alpha^* + P(\hat{\alpha} > c\alpha^*)^{1/2}$$

and, for any μ,

$$Er_4(\hat{\alpha}, \mu, x) \lesssim 1 + \tau(\alpha^*)^4 + (1 + \log^2 n)P(\hat{\alpha} < \alpha^*)^{1/2}.$$

By using that the probabilities in the last displays go to 0 faster than $1/n$, which we show below, and gathering the bounds for all i,

$$E_{\theta_0} \sum_{i=1}^n \int (\theta_i - \theta_{0,i})^4 d\Pi_{\hat{\alpha}}(\theta \mid X) \lesssim s_n(1 + \tau(\alpha^*)^4) + n\alpha^*.$$

From this deduce that

$$E_{\theta_0} \Pi_{\hat{\alpha}} \left[\|\theta - \theta_0\|^2 < v_n \mid X \right] \lesssim P[\mathcal{A}^c] + [s_n(1 + \tau(\alpha^*)^4) + n\alpha^*]/(s_n g(\zeta^*)^{-1})^2$$
$$\lesssim P[\mathcal{A}^c] + s_n^{-1}(1 + \tau(\alpha^*)^4)g(\zeta^*) + s_n^{-1}g(\zeta^*).$$

The last bound goes to 0, as $\tau(\alpha^*) \leq \zeta_{\alpha^*} = \zeta^*$ and g has Laplace tails. To conclude the proof, we show that $P_{\theta_0}(\hat{\alpha} \in [\alpha^*, c\alpha^*])$ is small. From the proof of Theorem 1, one already has $P_{\theta_0}[\hat{\alpha} < \alpha^*] \leq \exp(-cs_n)$, which is a $o(1/n)$ using $s_n \gtrsim \log^2 n$. To obtain a bound on $P_{\theta_0}[\hat{\alpha} > c\alpha^*]$, one can now revert the inequalities in the reasoning leading to the Bernstein bound in the proof of Theorem 1. With $A = \sum_{i=1}^n m_1(\mu_i, \alpha)$, we have

$$P_{\theta_0}[\hat{\alpha} > c\alpha^*] = P_{\theta_0}[S(c\alpha^*) > 0]$$
$$= P_{\theta_0}\left[\sum_{i=1}^n \beta(\theta_{0,i} + Z_i, c\alpha^*) - m_1(\theta_{0,i}, c\alpha^*) > -A\right].$$

But here, $-A = (n - s_n)\tilde{m}(c\alpha^*) - s_n m_1(\zeta^*, c\alpha^*)$. As $\alpha \to \tilde{m}(\alpha)$ is increasing, $\tilde{m}(c\alpha^*) \geq \tilde{m}(\alpha^*)$. Now by Lemma 7,

$$m_1(\zeta^*, c\alpha^*) \leq (c\alpha^* \wedge c_3)^{-1} \leq \frac{1}{c\alpha^*},$$

provided $\alpha^* \leq c_3/c = c_3/16$, which is the case for η_n small enough. Since by definition $n\tilde{m}(\alpha^*) = s_n/(4\alpha^*)$, we have $-A \geq s_n/(8\alpha^*)$. From there one can carry over the same scheme of proof as for the previous Bernstein inequality, with now $\tilde{A} = -A$ and $\tilde{V}$ the variance proxy which is bounded by

$$\tilde{V} \leq (n - s_n)m_2(0, c\alpha^*) + s_n m_2(\zeta^*, c\alpha^*) \lesssim n\frac{\tilde{m}(c\alpha^*)}{\zeta_{c\alpha^*} c\alpha^*} + \frac{s_n}{(c\alpha^*)^2}.$$

Now $\tilde{m}(c\alpha^*) \lesssim Cg(\zeta_{c\alpha^*})$. Using bounds similar to those of Lemma 8, one can check that $C_1 + \zeta_{\alpha^*}^2 \leq \zeta_{c\alpha^*}^2 \leq C_2 + \zeta_{\alpha^*}^2$, which implies that $\tilde{m}(c\alpha^*)/\zeta_{c\alpha^*} \lesssim \tilde{m}(\alpha^*)/\zeta^* \lesssim \tilde{m}(\alpha^*)$. From this one deduces, with $\tilde{M} \leq C/s_n$,

$$\frac{\tilde{V}}{\tilde{A}^2} + \frac{\tilde{M}\tilde{A}}{3\tilde{A}^2} \lesssim \frac{C'}{s_n},$$

which by Bernstein's inequality implies $P_{\theta_0}[\hat{\alpha} > c\alpha^*] \leq \exp[-Cs_n]$, which completes the proof of Theorem 2.

4. Technical lemmas for the SAS prior

4.1. Proofs of posterior risk bounds: fixed α

Proof of Lemma 1. First one proves the first two bounds. To do so, we derive moment bounds on γ_x. Since $\gamma_x(\cdot)$ is a density function, we have for any x, $\int \gamma_x(u)du = 1$. This implies $(\log g)'(x) = \int (u - x)\gamma_x(u)du = \int u\gamma_x(u)du - x$. In [12], the authors check, see p. 1623, that $\int u\gamma_x(u)du =: \tilde{m}_1(x)$ is a shrinkage rule, that is $0 \leq \tilde{m}_1(x) \leq x$ for $x \geq 0$, so by symmetry, for any real x,

$$| \int u\gamma_x(u)du| \leq |x|.$$

Decomposing $u^2 = (u-x)^2 + 2x(u-x) + x^2$ and noting that $\int (u-x)^2\gamma_x(u)du = g''(x)/g(x) + 1$,

$$\int u^2\gamma_x(u)du = \frac{g''}{g}(x) + 1 + 2x\frac{g'}{g}(x) + x^2.$$

Note that for γ Laplace or Cauchy, we have $|\gamma'| \leq c_1\gamma$ and $|\gamma''| \leq c_2\gamma$. This leads to

$$|g'(x)| = | \int \gamma'(x - u)\phi(u)du| \leq c_1 \int \gamma(x - u)\phi(u)du = c_1 g(x)$$

and similarly $|g''| \leq c_2 g$, so that $\int u^2\gamma_x(u)du \leq C(1 + x^2)$ which gives the first bound using (18). We note, *en passant*, that the one but last display also implies for any real x that

$$\int u^2\gamma_x(u)du \geq 1 - c_2 - 2c_1|x| + x^2, \tag{34}$$

which implies that $\int u^2\gamma_x(u)du$ goes to ∞ with x. Also, for any real μ,

$$\int (u - \mu)^2\gamma_x(u)du = (x - \mu)^2 + \frac{g''}{g}(x) + 1 + 2(x - \mu)\frac{g'}{g}(x).$$

Now using again $g'/g \leq c_1$ and $g''/g \leq c_2$ leads to

$$\int (u - \mu)^2\gamma_x(u)du \leq C(1 + (x - \mu)^2).$$

By using the expression of $r_2(\alpha, \mu, x)$, this yields the second bound of the lemma.

We now turn to the bounds in expectation. For a zero signal $\mu = 0$, one notes that $x = \tau(\alpha)$ is the value at which both terms in the minimum in the first inequality of the lemma are equal. So

$$E_0 r_2(\alpha, 0, x) \lesssim \int \mathbb{1}_{|x| \leq \tau(\alpha)} \frac{\alpha}{1 - \alpha} \frac{g}{\phi}(x)\phi(x)(1 + x^2)dx$$

$$+ \int \mathbb{1}_{|x| > \tau(\alpha)}(1 + x^2)\phi(x)dx.$$

I. Castillo and R. Mismer

For γ Cauchy, g has Cauchy tails and $x \to (1+x^2)g(x)$ is bounded, so one gets, with $\alpha \le 1/2$,

$$E_0 r_2(\alpha, 0, x) \lesssim \alpha \int \mathbb{1}_{|x| \le \tau(\alpha)} dx + \tau(\alpha)\phi(\tau(\alpha)) + \phi(\tau(\alpha))/\tau(\alpha)$$
$$\lesssim \tau(\alpha)\alpha + \tau(\alpha)\phi(\tau(\alpha)) \lesssim \tau(\alpha)\alpha + \tau(\alpha)\alpha g(\tau(\alpha)) \lesssim \tau(\alpha)\alpha.$$

Turning to the last bound of the lemma, we distinguish two cases. Set for the remaining of the proof $T := \tilde{\tau}(\alpha)$ for simplicity of notation. The first case is $|\mu| \le 4T$, for which

$$E_\mu r_2(\alpha, \mu, x) \le \mu^2 + C \le C_1(1 + T^2).$$

The second case is $|\mu| > 4T$. We bound the expectation of each term in the second bound of the lemma (that for $r_2(\alpha, \mu, x)$) separately. First, $E[a(x)(1 + (x-\mu)^2)] \le C$. It thus suffices to bound $\mu^2 E_\mu[1 - a(x)]$. To do so, one uses the bound (22) and starts by noting that, if $Z \sim \mathcal{N}(0,1)$,

$$E[\mathbb{1}_{|Z+\mu| \le T}] \le P[|Z| \ge |\mu| - T] \le P[|Z| \ge |\mu|/2].$$

This implies, with $\bar{\Phi}(u) = \int_u^\infty \phi(t)dt \le \phi(u)/u$ for $u > 0$,

$$E_\mu[\mu^2 \mathbb{1}_{|x| \le T}] \le C_2 |\mu| \phi(|\mu|) \le C_3.$$

If $A = \{x, \ |x-\mu| \le |\mu|/2\}$ and A^c denotes its complement,

$$\sqrt{2\pi} E_\mu[e^{-\frac{1}{2}(|x|-T)^2}] \le \int_{A^c} e^{-\frac{1}{2}(x-\mu)^2} dx + \int_A e^{-\frac{1}{2}(|x|-T)^2} dx.$$

The first term in the last sum is bounded above by $2\bar{\Phi}(|\mu|/2)$. The second term, as $A \subset \{x, \ |x| \ge |\mu|/2\}$, is bounded above by $2\bar{\Phi}(|\mu|/4)$. This implies, in the case $|\mu| > 4T$, that

$$E_\mu r_2(\alpha, \mu, x) \le C_4 + 4\mu^2 \bar{\Phi}(|\mu|/4) + 5 \le C.$$

The last bound of the lemma follows by combining the previous bounds in the two cases. □

Proof of Lemma 2. From the expression of $r_2(\alpha, 0, x)$ it follows

$$r_2(\alpha, 0, x) \ge a(x) \inf_{x \in \mathbb{R}} \int u^2 \gamma_x(u) du \ge \alpha \frac{g}{\phi \vee g}(x) \inf_{x \in \mathbb{R}} \int u^2 \gamma_x(u) du$$
$$\ge \alpha \inf_{x \in \mathbb{R}} \frac{g}{\phi \vee g}(x) \inf_{x \in \mathbb{R}} \int u^2 \gamma_x(u) du \ge C_0 \alpha,$$

where $c_0 > 0$. Indeed, both functions whose infimum is taken in the last display are continuous in x, are strictly positive for any real x, and have respective limits 1 and $+\infty$ as $|x| \to \infty$, using (34), so these functions are bounded below on $\mathbb{R}$ by positive constants. □

4.2. Proofs of posterior risk bounds: random α

Proof of Lemma 3. Using the bound on $r_2(\alpha, 0, x)$ from Lemma 1,

$$
\begin{aligned}
r_2(\hat{\alpha}, 0, x) &= r_2(\hat{\alpha}, 0, x)\mathbb{1}_{\hat{\alpha} \leq \alpha} + r_2(\hat{\alpha}, 0, x)\mathbb{1}_{\hat{\alpha} > \alpha} \\
&\leq \left[\frac{\hat{\alpha}}{1 - \hat{\alpha}} \frac{g}{\phi}(x) \wedge 1 \right] (1 + x^2)\mathbb{1}_{\hat{\alpha} \leq \alpha} + C(1 + x^2)\mathbb{1}_{\hat{\alpha} > \alpha} \\
&\leq \left[\frac{\alpha}{1 - \alpha} \frac{g}{\phi}(x) \wedge 1 \right] (1 + x^2)\mathbb{1}_{\hat{\alpha} \leq \alpha} + C(1 + x^2)\mathbb{1}_{\hat{\alpha} > \alpha}.
\end{aligned}
$$

For the first term in the last display, one bounds the indicator from above by 1 and proceeds as in the proof of Lemma 1 to bound its expectation by $C\alpha\tilde{\tau}(\alpha)$. The first part of the lemma follows by noting that $E[(1 + x^2)\mathbb{1}_{\hat{\alpha} > \alpha}]$ is bounded from above by $(2 + 2E_0[x^4])^{1/2}P(\hat{\alpha} > \alpha)^{1/2} \leq C_1 P(\hat{\alpha} > \alpha)^{1/2}$ by Cauchy-Schwarz inequality. The second part of the lemma follows from the fact that using Lemma 1, $r_2(\alpha, \mu, x) \leq (1 - a(x))\mu^2 + Ca(x)((x - \mu)^2 + 1) \leq \mu^2 + C(x - \mu)^2 + C$ for any α. □

Proof of Lemma 4. Combining (22) and the third bound of Lemma 1,

$$
r_2(\hat{\alpha}, \mu, x) \leq \mu^2 \left[\mathbb{1}_{|x| \leq \tilde{\tau}(\hat{\alpha})} + e^{-\frac{1}{2}(|x| - \tilde{\tau}(\hat{\alpha}))^2}\mathbb{1}_{|x| > \tilde{\tau}(\hat{\alpha})} \right] + C((x - \mu)^2 + 1).
$$

Note that it is enough to bound the first term on the right hand side in the last display, as the last one is bounded by a constant under E_μ. Let us distinguish the two cases $\hat{\alpha} \geq \alpha$ and $\hat{\alpha} < \alpha$.

In the case $\hat{\alpha} \geq \alpha$, as $\tilde{\tau}(\alpha)$ is a decreasing function of α,

$$
\begin{aligned}
& \left[\mathbb{1}_{|x| \leq \tilde{\tau}(\hat{\alpha})} + e^{-\frac{1}{2}(|x| - \tilde{\tau}(\hat{\alpha}))^2}\mathbb{1}_{|x| > \tilde{\tau}(\hat{\alpha})} \right] \mathbb{1}_{\hat{\alpha} \geq \alpha} \\
&\leq \left[\mathbb{1}_{|x| \leq \tilde{\tau}(\hat{\alpha})} + \mathbb{1}_{\tilde{\tau}(\hat{\alpha}) < |x| \leq \tilde{\tau}(\alpha)} + e^{-\frac{1}{2}(|x| - \tilde{\tau}(\hat{\alpha}))^2}\mathbb{1}_{|x| > \tilde{\tau}(\alpha)} \right] \mathbb{1}_{\hat{\alpha} \geq \alpha} \\
&\leq \mathbb{1}_{|x| \leq \tilde{\tau}(\alpha)} + e^{-\frac{1}{2}(|x| - \tilde{\tau}(\alpha))^2}\mathbb{1}_{|x| > \tilde{\tau}(\alpha)},
\end{aligned}
$$

where we have used $e^{-\frac{1}{2}v^2} \leq 1$ for any v and that $e^{-\frac{1}{2}(u-c)^2} \leq e^{-\frac{1}{2}(u-d)^2}$ if $u > d \geq c$. As a consequence, one can borrow the fixed α bound obtained previously so that

$$
E\left[r_2(\hat{\alpha}, \mu, x)\mathbb{1}_{\hat{\alpha} \geq \alpha} \right] \leq 2E_\mu r_2(\alpha, \mu, x) \leq C\left[1 + \tilde{\tau}(\alpha)^2 \right].
$$

In the case $\hat{\alpha} < \alpha$, setting $b_n = \sqrt{d\log n}$ and noting that $\tilde{\tau}(\hat{\alpha}) \leq b_n$ with probability 1 by assumption, proceeding as above, with b_n now replacing $\tilde{\tau}(\alpha)$, one can bound

$$
\begin{aligned}
& \mathbb{1}_{|x| \leq \tilde{\tau}(\hat{\alpha})} + e^{-\frac{1}{2}(|x| - \tilde{\tau}(\hat{\alpha}))^2}\mathbb{1}_{|x| > \tilde{\tau}(\hat{\alpha})} \\
&\leq \mathbb{1}_{|x| \leq b_n} + e^{-\frac{1}{2}(|x| - b_n)^2}\mathbb{1}_{|x| > b_n}.
\end{aligned}
$$

From this one deduces that

$$E\left(\mu^2\left[\mathbb{1}_{|x|\leq\tilde{\tau}(\hat{\alpha})} + e^{-\frac{1}{2}(|x|-\tilde{\tau}(\hat{\alpha}))^2}\mathbb{1}_{|x|>\tilde{\tau}(\hat{\alpha})}\right]\mathbb{1}_{\hat{\alpha}<\alpha}\right)$$

$$\leq C\left(E_\mu\left[\mu^4\mathbb{1}_{|x|\leq b_n} + \mu^4 e^{-(|x|-b_n)^2}\right]\right)^{1/2}P(\hat{\alpha}<\alpha)^{1/2}.$$

Using similar bounds as in the fixed α case, one obtains

$$E_\mu\left[\mu^4\mathbb{1}_{|x|\leq b_n} + \mu^4 e^{-(|x|-b_n)^2}\right] \leq C(1+b_n^4).$$

Taking the square root and gathering the different bounds obtained concludes the proof. □

4.3. Proofs on pseudo-thresholds

Proof of Lemma 8. For small α, or equivalently large ζ, we have $(g/\phi)(\zeta) = \beta(\zeta) + 1 \asymp \beta(\zeta)$. Deduce that for large n, using $\eta_n = d\alpha_1\tilde{m}(\alpha_1)$ and Lemma 5 on $\tilde{m}$,

$$\eta_n \asymp \alpha_1\zeta_1^{\kappa-1}\frac{g(\zeta_1)}{\beta(\zeta_1)}\beta(\zeta_1) \asymp \zeta_1^{\kappa-1}\phi(\zeta_1) \asymp \zeta_1^{\kappa-1}e^{-\zeta_1^2/2}.$$

From this deduce that

$$|\log c + (\kappa-1)\log\zeta_1 - \frac{\zeta_1^2}{2} + \log(1/\eta_n)| \leq C.$$

In particular, using $\log\zeta \leq a + \zeta^2/4$ for some constant $a > 0$ large enough, one gets $\zeta_1^2 \leq 4(C + \log(1/\eta_n)) \leq 4(C + \log n)$. Inserting this back into the previous inequality leads to

$$\zeta_1^2/2 \leq \log(1/\eta_n) + C + (1/2)(\kappa-1)\log\log n.$$

The lower bound is obtained by bounding $(\kappa-1)\log(\zeta_1) \geq 0$, for small enough α_1. □

Proof of Lemma 9. Using (11), $\log(1/\eta_n) \leq \log(n) - 2\log\log n$, and the bound on ζ from Lemma 8 gives $\zeta_1^2 \leq 2\log n - \frac{3}{2}\log\log n$, so that $t(\alpha_1) \leq \zeta(\alpha_1) = \zeta_1 \leq \sqrt{2\log n} = t(\alpha_n)$. It follows that α_1 belongs to the interval $[\alpha_n, 1]$ over which the likelihood is maximised.

Then one notices that $\{\hat{\zeta} < \zeta_1\} = \{\hat{\alpha} > \alpha_1\} = \{S(\alpha_1) > 0\}$, regardless of the fact that the maximiser $\hat{\alpha}$ is attained in the interior or at the boundary of $[\alpha_n, 1]$. So

$$P_\theta[\hat{\zeta} < \zeta_1] = P_\theta[S(\alpha_1) > 0].$$

The score function equals $S(\alpha) = \sum_{i=1}^n \beta(X_i, \alpha)$, a sum of independent variables. By Bernstein's inequality, if W_i are centered independent variables with $|W_i| \leq M$ and $\sum_{i=1}^n \text{Var}(W_i) \leq V$, then for any $A > 0$,

$$P\left[\sum_{i=1}^n W_i > A\right] \leq \exp\{-\frac{1}{2}A^2/(V + \frac{1}{3}MA)\}.$$

Set $W_i = \beta(X_i, \alpha_1) - m_1(\theta_{0,i}, \alpha_1)$ and $A = -\sum_{i=1}^{n} m_1(\theta_{0,i}, \alpha_1)$. Then one can take $M = c_3/\alpha_1$, using Lemma 7. One can bound $-A$ from above as follows, using the definition of α_1,

$$-A \leq -\sum_{i \notin S_0} \tilde{m}(\alpha_1) + \sum_{i \in S_0} \frac{c}{\alpha_1} \leq -(n - s_n)\tilde{m}(\alpha_1) + cs_n/\alpha_1$$

$$\leq -n\tilde{m}(\alpha_1)/2 + cdn\tilde{m}(\alpha_1) \leq -n\tilde{m}(\alpha_1)/4,$$

provided d is chosen small enough and, using again the definition of α_1,

$$V \leq \sum_{i \notin S_0} m_2(0, \alpha_1) + \sum_{i \in S_0} m_2(\theta_{0,i}, \alpha_1) \leq \frac{C}{\alpha_1} \left[(n - s_n)\tilde{m}(\alpha_1)\zeta_1^{-\kappa} + cs_n/\alpha_1 \right]$$

$$\leq C\alpha_1^{-1} \left[n\tilde{m}(\alpha_1)\zeta_1^{-\kappa}/2 + cdn\tilde{m}(\alpha_1) \right] \leq C'dn\tilde{m}(\alpha_1)/\alpha_1,$$

where one uses that ζ_1^{-1} is bounded. This leads to

$$\frac{V + \frac{1}{3}MA}{A^2} \leq \frac{C'd}{n\alpha_1\tilde{m}(\alpha_1)} + \frac{4c_3}{3n\alpha_1\tilde{m}(\alpha_1)} \leq \frac{c_5^{-1}}{n\alpha_1\tilde{m}(\alpha_1)}.$$

One concludes that $P[\hat{\alpha} > \alpha_1] \leq \exp\{-c_5 n\alpha_1\tilde{m}(\alpha_1)\} = \exp\{-Cs_n\}$ using (24). $\qquad \square$

Proof of Lemma 11. First we check the existence of a solution. Set $\zeta_\alpha = \zeta(\alpha)$ and $R_\alpha := \bar{\Phi}(\zeta_\alpha - \zeta_1)/(\alpha\tilde{m}(\alpha))$. For $\alpha \to 0$ we have $\zeta_\alpha - \zeta_1 \to \infty$ so by using $\bar{\Phi}(u) \asymp \phi(u)/u$ as $u \to \infty$ one gets, treating terms depending on ζ_1 as constants and using $\phi(\zeta_\alpha) \asymp \alpha g(\zeta_\alpha)$,

$$\bar{\Phi}(\zeta_\alpha - \zeta_1) \asymp \frac{\phi(\zeta_\alpha - \zeta_1)}{\zeta_\alpha - \zeta_1} \asymp \alpha g(\zeta_\alpha)e^{\zeta_\alpha\zeta_1}.$$

As $\tilde{m}(\alpha) \asymp \zeta_\alpha g(\zeta_\alpha)$, one gets $R_\alpha \asymp e^{\zeta_\alpha\zeta_1}/\zeta_\alpha \to \infty$ as $\alpha \to 0$. On the other hand, with $\pi_1 \leq s_n/n$ and $\alpha_1\tilde{m}(\alpha_1) = ds_n/n$,

$$R_{\alpha_1} = \frac{1}{2\alpha_1\tilde{m}(\alpha_1)} = \frac{dn}{2s_n} \leq \frac{8}{\pi_1}\frac{d}{16},$$

so that $R_{\alpha_1} < 8/\pi_1$ as $d < 2$. This shows that the equation at stake has at least one solution for α in the interval $(0, \alpha_1)$.

By definition of $m_1(\mu, \alpha)$, for any μ and α, and $\zeta = \zeta(\alpha)$,

$$m_1(\mu, \alpha) = \int_{-\zeta}^{\zeta} \frac{\beta(x)}{1 + \alpha\beta(x)}\phi(x - \mu)dx \; + \int_{|x|>\zeta} \frac{\beta(x)}{1 + \alpha\beta(x)}\phi(x - \mu)dx$$

$$= \qquad\qquad (A) \qquad\qquad + \qquad\qquad (B).$$

By definition of ζ, the denominator in (B) is bounded from above by $2\alpha\beta(x)$ so

$$(B) \geq \frac{1}{2\alpha}\int_{|x|>\zeta} \phi(x - \mu)dx \geq \frac{1}{2}\beta(\zeta)\bar{\Phi}(\zeta - \mu).$$

One splits the integral (A) in two parts corresponding to $\beta(x) \geq 0$ and $\beta(x) < 0$. Let c be the real number such that $g/\phi(c) = 1$. By construction the part of the integral (A) with $c \leq |x| \leq \zeta$ is nonnegative, so, for $\alpha \leq |\beta(0)|^{-1}/2$,

$$
\begin{aligned}
(A) &\geq \int_{-c}^{c} \frac{\beta(x)}{1 + \alpha\beta(x)} \phi(x - \mu) dx \\
&\geq -\int_{-c}^{c} \frac{|\beta(0)|}{1 - \alpha|\beta(0)|} \phi(x - \mu) dx \\
&\geq -2|\beta(0)| \int_{-c}^{c} \phi(x - \mu) dx,
\end{aligned}
$$

where one uses the monotonicity of $y \to y/(1 + \alpha y)$. For $\mu \geq c$, the integral $\int_{-c}^{c} \phi(x - \mu) dx$ is bounded above by $2\int_{0}^{c} \phi(x - \mu) dx \leq 2c\phi(\mu - c)$. To establish (33), it thus suffices to show that

$$
(i) := 4|\beta(0)|c\phi(\zeta_1 - c) \leq \frac{1}{4}\beta(\zeta_3)\bar{\Phi}(\zeta_3 - \zeta_1) =: (ii).
$$

The right hand-side equals $2\tilde{m}(\alpha_3)/\pi_1$ by definition of ζ_3. Since γ is Cauchy, Lemma 5 gives $\tilde{m}(\alpha_3) \asymp \zeta_3 g(\zeta_3) \asymp \zeta_3^{-1}$. It is enough to show that $(\pi_1\zeta_3)^{-1}$ is larger than $C\phi(\zeta_1 - c)$, for suitably large $C > 0$.

Let us distinguish two cases. In the case $\zeta_3 \leq 2\zeta_1$, the previous claim is obtained, since ζ_1 goes to infinity with n/s_n by Lemma 8 and $\phi(\zeta_1 - c) = o(\zeta_1^{-1})$. In the case $\zeta_3 > 2\zeta_1$, we obtain an upper bound on ζ_3 by rewriting the equation defining it. For $t \geq 1$, one has $\bar{\Phi}(t) \geq C\phi(t)/t$. Since $\zeta_3 - \zeta_1 > \zeta_1$ in the present case, it follows from the equation defining ζ_3 that

$$
C\frac{\phi(\zeta_3 - \zeta_1)}{\zeta_3 - \zeta_1} \leq 8\alpha_3\tilde{m}(\alpha_3)/\pi_1.
$$

This can be rewritten using $\phi(\zeta_3 - \zeta_1) = \sqrt{2\pi}\phi(\zeta_3)\phi(\zeta_1)e^{\zeta_1\zeta_3}$, as well as $\phi(\zeta_3) = g(\zeta_3)\alpha_3/(1 + \alpha_3) \gtrsim \alpha_3 g(\zeta_3)$ and $\tilde{m}(\alpha_3) \asymp \zeta_3 g(\zeta_3)$. This leads to

$$
\frac{e^{\zeta_1\zeta_3}}{\zeta_3^2} \leq \frac{C}{\pi_1}e^{\zeta_1^2/2}.
$$

By using $e^x/x^2 \geq Ce^{x/2}$ for $x \geq 1$ one obtains $\zeta_1^2 e^{\zeta_1\zeta_3/2} \leq e^{\zeta_1^2/2}C/\pi_1$, that is, using $\zeta_1^2 \geq 1$,

$$
\pi_1\zeta_3 \leq \pi_1\zeta_1 + \frac{\pi_1\log(C/\pi_1)}{\zeta_1} \leq \pi_1\zeta_1 + C \leq C'\zeta_1,
$$

using that $u \to u\log(1/u)$ is bounded on $(0,1)$. So the previous claim is also obtained in this case, as $\phi(\zeta_1 - c)$ is small compared to $(C'\zeta_1)^{-1}$ for large ζ_1. $\square$

4.4. Proof of the convergence rate for the modified estimator

Proof of Theorem 5. The proof is overall in the same spirit as that of Theorem 2 in [12] and goes by distinguishing the two cases $s_n \geq \log^2 n$ and $s_n < \log^2 n$.

The main difference is that here we work with the full posterior distribution, and the risk bounds require Lemmas 1–4, that bound the posterior risk in various settings, as well as a result, Lemma 13 below, in the same vein.

Also, we need to work with a modified version of ζ_1, to make sure that the probability in Lemma 9 goes to 0 fast enough. We note that this version of ζ_1 is the one used in [12] for both their Theorems 1 and 2 (in our Theorem 3, such a modification is not needed and we worked with the simpler version there). To do so, one replaces $\eta_n = s_n/n$ in the definition (24) of α_1 by

$$\tilde{\eta}_n = \max\left(\eta_n, \frac{\log^2 n}{n}\right).$$

To keep notation simple, we still denote the corresponding threshold by ζ_1. In the first part of the proof below, $\eta_n \geq \log^2(n)/n$, so this is the same version as in definition (24). In the second part of the proof, we have $\tilde{\eta}_n = \log^2 n/n$ and we now indicate the relevant properties of the corresponding modified threshold ζ_1. First, the statement of Lemma 8 becomes, with $\kappa = 2$ (as γ is Cauchy),

$$\log(1/\tilde{\eta}_n) + c_1 \leq \frac{\zeta_1^2}{2} \leq \log(1/\tilde{\eta}_n) + \frac{1}{2}\log\log n + c_2. \tag{35}$$

Second, we need below a bound on $P[\hat{\zeta} < \zeta_1]$ with the modified version of ζ_1 as above. It is not hard to check from the proof of Lemma 9 that this proof goes through with the new version of ζ_1 and η_n replaced by $\tilde{\eta}_n$. The only difference is with the term cs_n/α_1 which is bounded by $cn\tilde{\eta}_n/\alpha_1 = n\tilde{m}(\alpha_1)$, so that Bernstein's inequality gives

$$P[\hat{\zeta} < \zeta_1] \leq \exp\{-C'n\alpha_1\tilde{m}(\alpha_1)\} \leq \exp\{-Cn\tilde{\eta}_n\} \leq e^{-C\log^2 n}. \tag{36}$$

We are now ready for the proof of Theorem 5. First consider the case $s_n \geq \log^2 n$ and let us show that the risk of the empirical Bayes posterior $\Pi_{\hat{\alpha}_A}[\cdot \mid X]$ is not larger than that of the non-modified one. One decomposes

$$E_{\theta_0} \int \|\theta - \theta_0\|^2 d\Pi_{\hat{\alpha}_A}(\theta \mid X)$$

$$= E_{\theta_0} \int \|\theta - \theta_0\|^2 d\Pi_{\hat{\alpha}}(\theta \mid X) 1_{\hat{t} \leq t_n} + E_{\theta_0} \int \|\theta - \theta_0\|^2 d\Pi_{\hat{\alpha}_A}(\theta \mid X) 1_{\hat{t} > t_n}$$

$$\leq E_{\theta_0} \int \|\theta - \theta_0\|^2 d\Pi_{\hat{\alpha}}(\theta \mid X) + E_{\theta_0} \int \|\theta - \theta_0\|^2 d\Pi_{\alpha_A}(\theta \mid X) 1_{\hat{t} > t_n} = (I) + (II).$$

The term (I) corresponds to the risk of the unmodified estimator, so is bounded as in Theorem 3. For (II), one splits it according to small and large signals $\theta_{0,i}$: $(II) = S + \tilde{S}$, with

$$S = \sum_{i:\ |\theta_{0,i}| \leq \zeta_1} E_{\theta_0} \int (\theta_i - \theta_{0,i})^2 d\Pi_{\alpha_A}(\theta_i \mid X) 1_{\hat{t} > t_n},$$

I. Castillo and R. Mismer

and $\tilde{S} = (II) - S$. From Lemma 1, one knows that $r_2(\alpha_A, \mu, x) \leq \mu^2 + C(1 + (x-\mu)^2)$, while for $\mu = 0$, one can use the bound in expectation $E_0 r_2(\alpha, 0, x) \leq C\alpha\tau(\alpha)$, so that

$$S \leq \{ \sum_{i:\ |\theta_{0,i}|=0} + \sum_{i:\ 0<|\theta_{0,i}|\leq\zeta_1} \} E_{\theta_0} \int (\theta_i - \theta_{0,i})^2 d\Pi_{\alpha_A}(\theta_i \mid X)$$
$$\leq Cn\alpha_A\tau(\alpha_A) + Cs_n\zeta_1^2.$$

We now use the definition of α_A to bound α_A and $\tau(\alpha_A)$. To bound $\tau(\alpha_A)$, note that for any $\alpha \in (0,1)$, by definition $a(\tau(\alpha)) = 1/2$, so for a signal of amplitude $\tau(\alpha)$, the posterior puts $1/2$ of its mass at zero, which means the posterior median is 0, implying $\tau(\alpha) \leq t(\alpha)$, so that $\tau(\alpha_A) \leq t_A$. Combining with the bound for α_A of Lemma 12,

$$n\alpha_A\tau(\alpha_A) \leq Cn^{-A}t_A^3.$$

For any fixed $A > 0$, this goes to 0 with n so it is a $o(s_n\zeta_1^2)$, while $s_n\zeta_1^2$ is bounded by $Cs_n \log(n/s_n)$ as follows from Lemma 8. Now to bound $\tilde{S}$, one adapts the last bound of Lemma 1 to accommodate for the indicator $1_{\hat{t}>t_n}$. This is done in Lemma 13 whose bound (38) implies $\tilde{S} \leq Cs_n t_A^2 P(\hat{t} > t_n)^{1/2}$. This bound coincides up to a universal constant with the corresponding bound (128) in [12] (taken for $p = 0$, $\tilde{p} = 1$ and $q = 2$, which corresponds to our setting, i.e. working with ℓ_0 classes and quadratic risk). So the remaining bounds of [12] for the case $s_n > c \log^2 n$ can be used directly (the distinction of the three cases as in [12] p. 1646-1647 can be reproduced word by word, and is omitted for brevity), leading to $\tilde{S} \leq Cs_n \log(n/s_n)$.

Second, consider the case where $s_n \leq \log^2 n$. We note that for this regime of s_n, the inequalities (35) become, using that by definition $\tilde{\eta}_n = \log^2 n/n$,

$$\log n - 2\log\log n + c_1 \leq \frac{\zeta_1^2}{2} \leq \log n - \frac{3}{2}\log\log n + c_2. \tag{37}$$

Let us show that the risk of the plug-in posterior using the modified estimator is at most of the order of the minimax risk. For ζ_1 as above,

$$E_{\theta_0} \int \|\theta - \theta_0\|^2 d\Pi_{\hat{\alpha}_A}(\theta \mid X)$$
$$= \Big[\sum_{i:\ \theta_{0,i}=0} + \sum_{i:\ 0<|\theta_{0,i}|\leq\zeta_1} + \sum_{i:\ |\theta_{0,i}|>\zeta_1} \Big] E_{\theta_0} \int (\theta_i - \theta_{0,i})^2 d\Pi_{\hat{\alpha}_A}(\theta_i \mid X)$$
$$=: (i) + (ii) + (iii).$$

For the terms (i) and (ii), apply respectively each bound of Lemma 3 with $\alpha = \alpha_A$ to get $(ii) \leq Cs_n \left[\zeta_1^2 + 1\right] \leq C's_n\zeta_1^2 \lesssim s_n \log n$ using (35), which is bounded from above by $Cs_n \log(n/s_n)$ in the regime $s_n \leq \log^2 n$. Also,

$$(i) \leq Cn \left[\alpha_A\tilde{\tau}(\alpha_A) + P[\hat{\alpha}_A > \alpha_A]^{1/2} \right].$$

For large enough n, we have $\tilde{\tau}(\alpha_A) = \tau(\alpha_A)$ which is less than $t(\alpha_A) = t_A$ as noted above. Now α_A is bounded using Lemma 12, so that $n\alpha_A\tilde{\tau}(\alpha_A) \lesssim t_A(1+A)(\log n)n^{-A} = o(1)$ for $A > 0$.

We now bound the probability $P[\hat{\alpha}_A > \alpha_A]^{1/2}$. Recall the inequality $t(\alpha)^2 \geq \zeta(\alpha)^2 - C$ (see e.g. (53) in [12]). Using (37), we have $\zeta_1^2 \geq 2\log n - 4\log\log n + 2c_1$ so, writing in slight abuse of notation $t(\zeta_1) = t(\alpha_1)$ seeing $t(\cdot)$ as a function of ζ_1 instead of α_1,

$$t(\zeta_1)^2 \geq t_n^2 + \log\log n - C + 2c_1$$

so that $t(\zeta_1) \geq t_n$ for n large enough. Deduce $\{\hat{\alpha}_A > \alpha_A\} = \{\hat{t} < t_n\} \subset \{\hat{t} < t(\zeta_1)\} = \{\hat{\zeta} < \zeta_1\}$. Using (36), we have $P[\hat{\zeta} < \zeta_1] \leq e^{-C\log^2 n}$, so that (i) goes to 0, and so is a $o(s_n\log(n/s_n))$.

Finally, for the term (iii) one uses Lemma 4 with $\alpha = \alpha_A$. Note $\{\hat{\alpha}_A < \alpha_A\} = \{t(\hat{\alpha}_A) > t_A\}$. But by definition note that $t(\hat{\alpha}_A)$ equals either t_A if $\hat{t} > t_n$ or $t(\hat{\alpha})$ if $\hat{t} = t(\hat{\alpha}) \leq t_n$, so that $t(\hat{\alpha}) \leq t_n$. As $t_n^2 < 2\log n < t_A^2$ for $A > 0$, conclude that in all cases $t(\hat{\alpha}_A) \leq t_A$ with probability one, so that $P[\hat{\alpha}_A < \alpha_A] = 0$. Thus

$$(iii) \leq \sum_{i:\ |\theta_{0,i}|>\zeta_1} E_{\theta_{0,i}} r_2(\hat{\alpha}_A, \theta_{0,i}, X_i) \leq C \sum_{i:\ |\theta_{0,i}|>\zeta_1} (1 + \tilde{\tau}(\alpha_A)^2 + 0) \leq Cs_n\tilde{\tau}(\alpha_A)^2,$$

which is no more than $2Cs_n(1+A)\log n \leq C's_n\log n$. As $s_n \leq c\log^2 n$, we have $\log n \lesssim \log(n/s_n)$ so $(iii) \leq Cs_n\log(n/s_n)$. Putting the previous bounds together, one gets $(i) + (ii) + (iii) \leq Cs_n\log(n/s_n)$, which concludes the proof. $\square$

Lemma 12. *For $A \geq 0$, with $t_A^2 = 2(1+A)\log n$ and $\alpha_A = t^{-1}(t_A)$, there exist $N_0 > 0$ and $C > 0$ both independent of A such that for $n \geq N_0$,*

$$\alpha_A \leq C(1+A)(\log n)n^{-1-A}.$$

Proof. First recall the bound $t(\alpha) < \zeta(\alpha)$. Setting $\alpha = t^{-1}(u)$ in this inequality leads, using $\zeta(u) = \beta^{-1}(1/u)$, to $u < \beta^{-1}(1/t^{-1}(u))$. As β is increasing on $\mathbb{R}^+$, one has $t^{-1}(u) < 1/\beta(u)$, so

$$\alpha_A < \frac{1}{\beta(t_A)} = \frac{g}{\phi-g}(t_A)\frac{\phi}{g}(t_A) \leq 2\frac{\phi}{g}(t_A) \leq Ct_A^2 e^{-t_A^2},$$

where we use that g has Cauchy tails. The result follows by using the expression of t_A. $\square$

Lemma 13. *For any real μ, for $B := \{\hat{t} > t_n\}$, and α_A, t_A as above,*

$$E_\mu[r_2(\alpha_A, \mu, x)\mathbb{1}_B] \leq C(t_A^2 + 1)P(B)^{1/2}. \tag{38}$$

Proof. Similar to the proof of Lemma 1, one sets $T := \tau(\alpha_A)$ and distinguishes two cases: if $|\mu| \leq 4T$, Lemma 1 implies $r_2(\alpha_A, \mu, x) \leq \mu^2 + (1 + (x - \mu)^2)$, so using Cauchy-Schwarz inequality,

$$E_\mu[r_2(\alpha_A, \mu, x)\mathbb{1}_B] \leq CT^2 P(B) + P(B) + E_\mu[(x-\mu)^4]^{1/2}P(B)^{1/2}$$
$$\leq C(1 + T^2)P(B)^{1/2}.$$

If $|\mu| > 4T$, one uses the bound on r_2 from Lemma 1 again keeping the dependence in $a(x)$. First,

$$E[a(x)\{1 + (x - \mu)^2\}\mathbb{1}_B] \leq E[\{1 + (x - \mu)^2\}^2]^{1/2}P(B)^{1/2} \leq CP(B)^{1/2}.$$

Let us now focus on $E_\mu[(1 - a(x))\mu^2\mathbb{1}_B] \leq E_\mu[\{1_{|x|\leq T} + e^{-(|x|-T)^2/2}1_{|x|>T}\}\mathbb{1}_B]$. The first term, using $P_\mu[|x| < T] \leq \bar{\Phi}(|\mu|/2)$, is bounded by

$$\mu^2\Phi(|\mu|/2)^{1/2}P(B)^{1/2} \leq CP(B)^{1/2}.$$

The second term is bounded by

$$\mu^2\{E_\mu[e^{(-|x|-T)^2}]\}^{1/2}P(B)^{1/2}.$$

In the proof of Lemma 1, we showed that $E_\mu[e^{(-|x|-T)^2/2}]^{1/2}$ is bounded by a universal constant times $\bar{\Phi}(|\mu|/4)$. As $e^{-y^2} \leq e^{-y^2/2}$, the term at stake is bounded from above by $\mu^2\bar{\Phi}(|\mu|/4)P(B)^{1/2} \leq CP(B)^{1/2}$, which implies (38). □

5. Proof of Theorem 4: the SSL prior

Recall that we use the notation of the SAS case, keeping in mind that every instance of g is replaced by g_1 and (some of the) ϕs by g_0. Similarly, $\beta(x, \alpha)$, $\tilde{m}$, m_1 and m_2 are defined as in Section 3.1, but with $\beta(x) = g_1/g_0 - 1$.

The main steps of the proof generally follow those of Theorem 3, although technically there are quite a few differences. In the SSL case, we do not know whether the function $\beta = g_1/g_0 - 1$ is nondecreasing over the whole $\mathbb{R}^+$. Yet, we managed to show that β, which is an even function, is nondecreasing on the interval

$$J_n = [2\lambda_1, \sqrt{2\log n}],$$

see Proposition 1 below. This allows us to define its inverse $\beta^{-1} = \beta_{|J_n}^{-1}$ on this interval. Further, we prove in Lemma 20 that β crosses the horizontal axis on the previous interval, is strictly negative on $[0, 2\lambda_1]$ and tends to ∞ when $x \to \infty$. As β is continuous, the graph of the function crosses any given horizontal line $y = c$, for any $c > 0$.

The threshold ζ in the SSL case. For every $\alpha \in (0, 1)$, one sets

$$\zeta = \zeta(\alpha) = \min\{s > 0, \ \beta(s) = 1/\alpha\}. \tag{39}$$

This is well defined by the property noted in the previous paragraph. Now one notes that $g_0 \leq 2\phi$ for $x \leq \lambda_0/2$, see Lemma 19, and that the function g_1/ϕ takes a value at $\sqrt{2\log n}$ not smaller than $Cn/\log n$, since $g_1 \lesssim \gamma_1$ has Cauchy tails. This implies the existence of a constant $\mathcal{C} > 1$ such that

$$\beta(\sqrt{2\log n}) \geq n/(\mathcal{C}\log n). \tag{40}$$

Now we claim that for any $\alpha \in (\mathcal{C} \log n/n, 1]$, we have the identity $\zeta(\alpha) = \beta^{-1}(\alpha^{-1})$. To see this, first note that for any $\alpha \in (\mathcal{C} \log n/n, 1]$, by (40) and $\beta(2\lambda_1) < 0$, we have $\alpha^{-1} \in \beta(J_n)$. This shows that $t = \beta^{-1}(\alpha^{-1})$ solves $\beta(t) = \alpha^{-1}$. Also, it is the smallest possible solution $t > 0$, as β takes negative values on $[0, 2\lambda_1]$, which establishes the identity.

The threshold ζ_1 in the SSL case. In the SSL case, the function $\alpha \to \tilde{m}(\alpha) = -E_0[\beta(X, \alpha)]$ is still nondecreasing, since for any real z, the map $\mathcal{M}_z : \alpha \to z/(1 + \alpha z)$ is nonincreasing and $\beta(X, \alpha) = \mathcal{M}_{\beta(X)}(\alpha)$. By Proposition 2, we also have that $\tilde{m}$ is positive for $\alpha \geq \mathcal{C} \log n/n$ and is of the order of a constant for $\alpha = 1$. So, the map $\alpha \to \alpha\tilde{m}(\alpha)$ is nondecreasing on $[\mathcal{C} \log n/n, 1]$, its value at $\mathcal{C} \log n/n$ is less than $C' \log n/n$, and its value at one is of the order of a constant. This shows, using $s_n \geq c_1 \log^2 n$ by (11), that the following equation has a unique solution $\alpha_1 \in (\mathcal{C} \log n/n, 1)$

$$\alpha_1 \tilde{m}(\alpha_1) = ds_n/n, \tag{41}$$

with d a small enough constant to be chosen later (see the proof of Lemma 21). Thus we can set

$$\zeta_1 = \beta^{-1}(\alpha_1^{-1}),$$

and by the above arguments we have $\zeta_1 \in J_n$. So Proposition 2 gives $\alpha_1^{-1} \asymp \frac{n}{s_n}\zeta_1 g_1(\zeta_1) \asymp \frac{n}{s_n\zeta_1}$. Now we can follow the same proof as in Lemma 8, replacing up to constants instances of $g_0(\zeta_1)$ by $\phi(\zeta_1)$ thanks to Lemma 17 and (46) (as $\zeta_1 \leq \sqrt{2 \log n} < \lambda_0/2$), to obtain

$$\zeta_1^2 \lesssim C \log(n/s_n).$$

Defining $\tau(\alpha)$ and $\tilde{\tau}(\alpha)$. In the SSL case, we set

$$\Omega(x, \alpha) = \frac{\alpha}{1 - \alpha} \frac{2g_1}{\phi}(x).$$

This definition is as in the SAS case except that g is replaced by $2g_1$. We still use the same notation for simplicity. As g_1 satisfies the same properties as g, one defines $\tau(\alpha)$ and $\tilde{\tau}(\alpha)$ similarly to the SAS case. More precisely, $\tau(\alpha)$ is the unique solution to the equation $\Omega(\tau(\alpha), \alpha) = 1$, whenever $\alpha \leq \alpha^*$, where $\Omega(0, \alpha^*) = 1$. One sets $\tau(\alpha) = 0$ for $\alpha \geq \alpha^*$ and $\tilde{\tau}(\alpha) = \tau(\alpha \wedge \alpha_0)$ with $\tau(\alpha_0) = \lambda_1$ (this slightly differs from the SAS case).

As in the proof of Theorem 3, one can now decompose the risk $R_n(\theta_0) = E_{\theta_0} \int \|\theta - \theta_0\|^2 d\Pi_{\hat{\alpha}}(\theta \,|\, X)$ according to whether coordinates of θ correspond to a 'small' or 'large' signal, the threshold being ζ_1 that we define next. One can write

$$R_n(\theta_0) = \Big[\sum_{i:\ \theta_{0,i}=0} + \sum_{i:\ 0<|\theta_{0,i}|\leq\zeta_1} + \sum_{i:\ |\theta_{0,i}|>\zeta_1} \Big] E_{\theta_0} \int (\theta_i - \theta_{0,i})^2 d\Pi_{\hat{\alpha}}(\theta_i \,|\, X).$$

We next use the first part of Lemma 16 with $\alpha = \alpha_1$ and the second part of the Lemma to obtain, for any θ_0 in $\ell_0[s_n]$,

$$\Big[\sum_{i:\ \theta_{0,i}=0} + \sum_{i:\ 0<|\theta_{0,i}|\leq\zeta_1} \Big] E_{\theta_0} \int (\theta_i - \theta_{0,i})^2 d\Pi_{\hat{\alpha}}(\theta_i \,|\, X)$$

$$\leq C \sum_{i: \ \theta_{0,i}=0} \left[\alpha_1 \tilde{\tau}(\alpha_1) + P_{\theta_0}(\hat{\alpha} > \alpha_1) + \lambda_0^{-2}\right] + \sum_{i: \ 0<|\theta_{0,i}|\leq \zeta_1} (\theta_{0,i}^2 + C)$$

$$\leq C(n-s_n)\left[\alpha_1\tilde{\tau}(\alpha_1) + e^{-C\log^2 n} + \lambda_0^{-2}\right] + (\zeta_1^2 + C)s_n,$$

where for the last inequality we use Lemma 21. From (41) one gets

$$n\alpha_1 \lesssim s_n \zeta_1^{-1} g(\zeta_1)^{-1} \lesssim s_n \zeta_1.$$

Let us now check that $\tilde{\tau}(\alpha_1) \leq \zeta_1$. First, $\beta(\zeta_1) = \alpha_1^{-1} > \alpha_1^{-1} - 1$. By definition of $\tau(\alpha_1)$, using $\phi \leq 2g_0$ by Lemma 17,

$$\alpha_1^{-1} - 1 = 2(g_1/\phi)(\tau(\alpha_1)) \geq \beta(\tau(\alpha_1)) + 1.$$

This gives us that $\beta(\zeta_1) \geq \beta(\tau(\alpha_1)) + 1$ which implies the result as β is increasing here. Now with the previous bound on ζ_1 one obtains that the contribution to the risk of the indices i with $|\theta_{0,i}| \leq \zeta_1$ is bounded by a constant times $s_n \log(n/s_n)$.

It remains to bound the part of the risk for indexes i with $|\theta_{0,i}| > \zeta_1$. To do so, one uses the second part of Lemma 16 with α chosen as $\alpha_2' = \mathcal{C}(\log n/n)$, with $\mathcal{C}$ as in (40). By definition of $\hat{\alpha}$ in (16), the probability that $\hat{\alpha}$ is smaller than α_2' equals zero. Also, one has $\tilde{\tau}(\alpha_2')^2 \leq C \log n$. Indeed, setting $\zeta_2' = \beta^{-1}(\alpha_2'^{-1})$, we have as before $\tau(\alpha_2') \leq \zeta_2' \leq \sqrt{2\log n}$. This implies

$$\sum_{i: \ |\theta_{0,i}|>\zeta_1} E_{\theta_0} \int (\theta_i - \theta_{0,i})^2 d\Pi_{\hat{\alpha}}(\theta_i \,|\, X) \leq C s_n \log n,$$

which concludes the proof of Theorem 4.

6. Technical lemmas for the SSL prior

6.1. Fixed α bounds

As in the SAS case, we use the notation $r_2(\alpha, \mu, x) = \int (u-\mu)^2 d\pi_\alpha(u\,|\,x)$, where now $\pi_\alpha(\cdot\,|\,x)$ is the posterior on one coordinate (X_1, say) for fixed α in the SSL case, given $X_1 = x$.

Lemma 14. *For a zero signal $\mu = 0$, we have for any x and $\alpha \in [0, 1/2]$,*

$$r_2(\alpha, 0, x) \leq C\left[1 \wedge \frac{\alpha}{1-\alpha}\frac{g_1}{\phi}(x)\right](1+x^2) + \int u^2 \gamma_{0,x}(u)du$$

$$E_0 r_2(\alpha, 0, x) \leq C\tau(\alpha)\alpha + 4/\lambda_0^2.$$

For an arbitrary signal $\mu \in \mathbb{R}$, we have that for any real x and $\alpha \in [0, 1/2]$,

$$r_2(\alpha, \mu, x) \leq (1 - a(x)) \int (u-\mu)^2 \gamma_{0,x}(u)du + Ca(x)((x-\mu)^2 + 1)$$

$$E_\mu r_2(\alpha, \mu, x) \leq C(1 + \tilde{\tau}(\alpha)^2).$$

Proof. By definition, in the SSL case, $r_2(\alpha, 0, x) = (1 - a(x)) \int u^2 \gamma_{0,x}(u) du + a(x) \int u^2 \gamma_{1,x}(u) du$. Similar to Lemma 1, we have $a(x) \int u^2 \gamma_{1,x}(u) du \leq C \big[1 \wedge \frac{\alpha}{1-\alpha} \frac{g_1}{g_0}(x) \big] (1 + x^2)$. The first bound now follows from the inequality $g_0 \geq \phi/2$ obtained in Lemma 17. For the bound in expectation,

$$
E_0 \left[\int u^2 \gamma_{0,x}(u) du \right] = \int \left(\int u^2 \frac{\phi(x-u)\gamma_0(u)}{g_0(x)} du \right) \phi(x) dx
$$
$$
\leq 2 \int u^2 \int \phi(x-u) dx \gamma_0(u) du = 2 \int u^2 \gamma_0(u) du = 4/\lambda_0^2,
$$

and one then proceeds as in Lemma 1 to obtain the bound for zero signal.

Now for a general signal μ, the bound for $r_2(\alpha, \mu, x)$ follows from the definition and the previous bound. For the bound in expectation, by symmetry one can assume $\mu \geq 0$. Also note that the term with the $a(x)$ factor is bounded in expectation by a constant, by using $a(x) \leq 1$. To handle the term with $1 - a(x)$, we distinguish two cases. First, one assumes that $\mu \leq \lambda_0/2$. We have, using $(a+b)^2 \leq 2a^2 + 2b^2$,

$$
(1 - a(x)) \int (u-\mu)^2 \gamma_{0,x}(u) du \lesssim (1 - a(x))\mu^2 + (1 - a(x)) \int u^2 \phi(x-u) \frac{\gamma_0(u)}{g_0(x)} du.
$$

For the first term we proceed as in Lemma 1, for the second using $g_0 \geq \phi/2$ from Lemma 17,

$$
E_\mu \left[(1 - a(x)) \int u^2 \phi(x-u) \frac{\gamma_0(u)}{g_0(x)} du \right] \leq 2 \int u^2 \gamma_0(u) \int \frac{\phi(x-u)\phi(x-\mu)}{\phi(x)} dx du
$$
$$
\lesssim \int u^2 \gamma_0(u) \int e^{-(x-(u+\mu))^2/2 + u\mu} dx du \lesssim \lambda_0 \int u^2 e^{-\lambda_0|u| + u\mu} du.
$$

As $\mu \leq \lambda_0/2$, this is in turn bounded by a constant times $(\lambda_0)^{-2}$. Now in the case that $\mu > \lambda_0/2$, recall from the proof of Lemma 1 that for any real x,

$$
\int (u-\mu)^2 \gamma_{0,x}(u) du = (x-\mu)^2 + 1 + \frac{g_0''}{g_0}(x) + 2(x-\mu)\frac{g_0'}{g_0}(x). \qquad (42)
$$

The first two terms are, in expectation, bounded by a constant. Next one writes

$$
E_\mu \left[(1 - a(x)) \frac{g_0''}{g_0}(x) \right] = \int (1 - a(x)) \frac{g_0''}{g_0}(x) \phi(x-\mu) dx
$$

By Lemma 17, we have $|g_0''| = \lambda_0^2 |g_0 - \phi| \leq 1$. One splits the integral on the last display in two parts. For $|x| \leq \mu/4$, one uses that g_0'' is bounded together with the bound $g_0 \geq \phi/2$. For $|x| > \mu/4$, one uses $g_0''/g_0 = \lambda_0^2(g_0 - \phi)/g_0 \leq \lambda_0^2$ together with $1 - a(x) \leq (g_0/g_1)(x)/\alpha$, which follows from the expression of $a(x)$. This leads to

$$
E_\mu \left[(1 - a(x)) \frac{g_0''}{g_0}(x) \right] \leq \int_{|x| \leq \mu/4} e^{x\mu - \frac{\mu^2}{2}} dx + \frac{\lambda_0^2}{\alpha} \int_{|x| > \mu/4} \frac{g_0}{g_1}(x) \phi(x-\mu) dx.
$$

The first term in the last expression is bounded. The second one is bounded by a constant given our choice of λ_0 by combining the following: $\alpha^{-1} \leq n$ from (16), $g_0 \lesssim \gamma_0$ for $\mu > \lambda_0/8$ from (47) and $g_1 \gtrsim \gamma_1$.

To conclude the proof, for the last term in (42), using (45), the bound on $1 - a(x)$ from Lemma 15 below, and the fact that $x \mapsto x\phi(x)$ is bounded, $E_\mu \left[2(1 - a(x))(x - \mu)\frac{g_0'}{g_0}(x) \right]$ is bounded by

$$
2 \int (1 - a(x))|\frac{g_0'}{g_0}(x)||(x - \mu)\phi(x - \mu)|dx \lesssim \int (1 - a(x))|x|dx
$$

$$
\lesssim \int_{|x| \leq \tilde{\tau}(\alpha)} |x|dx + \int_{\tilde{\tau}(\alpha) \leq |x| \leq \frac{\lambda_0}{2}} |x|e^{-\frac{(|x| - \tilde{\tau}(\alpha))^2}{2}}dx + \int_{|x| \geq \frac{\lambda_0}{2}} |x|(1 - a(x))dx
$$

$$
\lesssim \tilde{\tau}(\alpha)^2 + 2(1 - e^{-\frac{(\frac{\lambda_0}{2} - \tilde{\tau}(\alpha))^2}{2}}) + \tilde{\tau}(\alpha) + \int_{|x| \geq \frac{\lambda_0}{2}} n^3 |x|\frac{\gamma_0}{\gamma_1}(x)dx \lesssim 1 + \tilde{\tau}(\alpha)^2.
$$

$\square$

Lemma 15. *For any $x \in [0, \lambda_0/2]$ and $\alpha \in [0, 1]$,*

$$
1 - a(x) \leq \mathbb{1}_{|x| \leq \tilde{\tau}(\alpha)} + 4e^{-\frac{1}{2}(|x| - \tilde{\tau}(\alpha))^2} \mathbb{1}_{|x| > \tilde{\tau}(\alpha)}.
$$

Proof. One first notes that $1 - a(x) \leq 4\Omega(x, \alpha)^{-1}$ for $x \leq \lambda_0/2$, using the fact that for such x, $g_0(x) \leq 2\phi(x)$ as found in Lemma 19. The following inequalities hold for $\tilde{\tau}(\alpha) \leq x \leq \lambda_0/2$, using $\tilde{\tau}(\alpha) \geq \lambda_1$ by definition and that $|(\log g_1)'| \leq \lambda_1$ as seen in (44),

$$
\Omega(x, \alpha) = \Omega(\tilde{\tau}(\alpha), \alpha) \exp\left(\int_{\tilde{\tau}(\alpha)}^{x} ((\log g_1)'(u) - (\log \phi)'(u))du \right)
$$

$$
\geq \exp\left(\int_{\tilde{\tau}(\alpha)}^{x} (u - \lambda_1)du \right) \geq \exp\left(\int_{\tilde{\tau}(\alpha)}^{x} (u - \tilde{\tau}(\alpha))du \right) = e^{\frac{(x - \tilde{\tau}(\alpha))^2}{2}}.
$$

$\square$

6.2. *Random α bounds*

Lemma 16. *Let α be a fixed non-random element of $(0, 1)$. Let $\hat{\alpha}$ be a random element of $[0, 1]$ that may depend on $x \sim \mathcal{N}(0, 1)$ and on other data. Then there exists $C_1 > 0$ such that*

$$
Er_2(\hat{\alpha}, 0, x) \leq C_1 \left[\alpha\tilde{\tau}(\alpha) + P(\hat{\alpha} > \alpha)^{1/2} \right] + \frac{4}{\lambda_0^2}.
$$

There exists $C_2 > 0$ such that for any real μ, if $x \sim \mathcal{N}(\mu, 1)$,

$$
Er_2(\hat{\alpha}, \mu, x) \leq \mu^2 + C_2.
$$

Suppose now that $\tilde{\tau}(\hat{\alpha})^2 \leq d\log(n)$ with probability 1 for some $d > 0$, and that $x \sim \mathcal{N}(\mu, 1)$. Then there exists $C_2 > 0$ such that for all real μ,

$$Er_2(\hat{\alpha}, \mu, x) \leq C_2 \left[1 + \tilde{\tau}(\alpha)^2 + (1 + d\log n)P(\hat{\alpha} < \alpha)^{1/2}\right].$$

Proof of Lemma 16. For the first two inequalities, the proof is the same as in the SAS case in Lemma 3, the only difference being the presence of the term $4/\lambda_0^2$ coming from Lemma 14 for the first inequality. For the third inequality, it follows from Lemma 14 that

$$r_2(\hat{\alpha}, \mu, x) \leq (1 - a_{\hat{\alpha}}(x)) \int (u - \mu)^2 \gamma_{0,x}(u)du + C[(x - \mu)^2 + 1].$$

In expectation the last term is constant. For the first term, with Lemma 15,

$$1 - a_{\hat{\alpha}}(x) \leq \mathbb{1}_{|x| \leq \tilde{\tau}(\hat{\alpha})} + 4e^{-\frac{1}{2}(|x| - \tilde{\tau}(\hat{\alpha}))^2}\mathbb{1}_{\frac{\lambda_0}{2} \geq |x| > \tilde{\tau}(\hat{\alpha})} + \mathbb{1}_{|x| \geq \frac{\lambda_0}{2}}n\frac{g_0}{g_1}(x),$$

where the last estimate uses the bound $\alpha \geq 1/n$. As in Lemma 4, let us distinguish the two cases $\hat{\alpha} \geq \alpha$ and $\hat{\alpha} < \alpha$. In the case $\hat{\alpha} \geq \alpha$, as $\tilde{\tau}(\alpha)$ is a decreasing function of α,

$$\left[\mathbb{1}_{|x| \leq \tilde{\tau}(\hat{\alpha})} + 4e^{-\frac{1}{2}(|x| - \tilde{\tau}(\hat{\alpha}))^2}\mathbb{1}_{\frac{\lambda_0}{2} \geq |x| > \tilde{\tau}(\hat{\alpha})}\right]\mathbb{1}_{\hat{\alpha} \geq \alpha}$$

$$\lesssim \left[\mathbb{1}_{|x| \leq \tilde{\tau}(\hat{\alpha})} + \mathbb{1}_{\tilde{\tau}(\hat{\alpha}) < |x| \leq \tilde{\tau}(\alpha)} + e^{-\frac{1}{2}(|x| - \tilde{\tau}(\hat{\alpha}))^2}\mathbb{1}_{\frac{\lambda_0}{2} \geq |x| > \tilde{\tau}(\alpha)}\right]\mathbb{1}_{\hat{\alpha} \geq \alpha}$$

$$\lesssim \mathbb{1}_{|x| \leq \tilde{\tau}(\alpha)} + e^{-\frac{1}{2}(|x| - \tilde{\tau}(\alpha))^2}\mathbb{1}_{\frac{\lambda_0}{2} \geq |x| > \tilde{\tau}(\alpha)},$$

where we have used $e^{-\frac{1}{2}v^2} \leq 1$ for any v and that $e^{-\frac{1}{2}(u-c)^2} \leq e^{-\frac{1}{2}(u-d)^2}$ if $u > d \geq c$.

For the third term, we have to control

$$E_\mu\left[\mathbb{1}_{|x| \geq \frac{\lambda_0}{2}}n\frac{g_0}{g_1}(x)\int (u - \mu)^2\gamma_{0,x}(u)du\right].$$

To do so, one uses (42). In expectation, the term in factor of $(x - \mu)^2 + 1$ is bounded by a constant. Using (47) and the fact that $g_0''/g_0 \leq \lambda_0^2$, the term in factor g_0''/g_0 is bounded by

$$\lambda_0^2 n\int_{|x| \geq \frac{\lambda_0}{2}}\frac{g_0}{g_1}(x)\phi(x - \mu)dx \lesssim n^3\int_{|x| \geq \frac{\lambda_0}{2}}\frac{\gamma_0}{\gamma_1}(x)dx$$

$$\lesssim n^4\int_{|x| \geq \frac{\lambda_0}{2}}x^2 e^{-\lambda_0|x|}dx \lesssim n^4 e^{-Cn^2}.$$

Finally, using (45) and the fact that $x \mapsto x\phi(x)$ is bounded, one obtains

$$E_\mu\left[\mathbb{1}_{|x| \geq \frac{\lambda_0}{2}}n\frac{g_0}{g_1}(x)(x - \mu)\frac{g_0'}{g_0}(x)\right] \leq \int_{|x| \geq \frac{\lambda_0}{2}}n\frac{g_0}{g_1}(x)|x||(x - \mu)\phi(x - \mu)|dx$$

$$\lesssim \int_{|x| \geq \frac{\lambda_0}{2}}n\frac{g_0}{g_1}(x)|x|dx.$$

As a consequence, one can borrow the fixed α bound obtained previously so that

$$E\left[r_2(\hat{\alpha}, \mu, x) 1_{\hat{\alpha} \geq \alpha}\right] \lesssim E_\mu r_2(\alpha, \mu, x) \lesssim \left[1 + \tilde{\tau}(\alpha)^2\right].$$

In the case $\hat{\alpha} < \alpha$, setting $b_n = \sqrt{d \log n}$ and noting that $\tilde{\tau}(\hat{\alpha}) \leq b_n$ with probability 1 by assumption, proceeding as above, with b_n now replacing $\tilde{\tau}(\alpha)$, one can bound

$$\mathbb{1}_{|x| \leq \tilde{\tau}(\hat{\alpha})} + 4e^{-\frac{1}{2}(|x| - \tilde{\tau}(\hat{\alpha}))^2} \mathbb{1}_{\frac{\lambda_0}{2} \geq |x| > \tilde{\tau}(\hat{\alpha})} + \mathbb{1}_{|x| \geq \frac{\lambda_0}{2}} n \frac{g_0}{g_1}(x)$$

$$\lesssim \mathbb{1}_{|x| \leq b_n} + e^{-\frac{1}{2}(|x| - b_n)^2} \mathbb{1}_{\frac{\lambda_0}{2} \geq |x| > b_n} + \mathbb{1}_{|x| \geq \frac{\lambda_0}{2}} n \frac{g_0}{g_1}(x).$$

From this one deduces that $E\left[(1 - a_{\hat{\alpha}}(x)) \int (u - \mu)^2 \gamma_{0,x}(u) du\right]$ is bounded from above by a constant times

$$\left(E_\mu\left[\left(\int (u - \mu)^2 \gamma_{0,x}(u) du\right)^2 \left[\mathbb{1}_{|x| \leq b_n} + e^{-(|x| - b_n)^2}\right]\right]\right)^{1/2} P(\hat{\alpha} < \alpha)^{1/2}.$$

Using the same bounds but squared as in the fixed α case, one obtains that the expectation in the last display is bounded from above by $C(1 + b_n^4)$. Taking the square root and gathering the different obtained bounds concludes the proof. □

6.3. Properties of the functions g_0 and β for the SSL prior

Recall the notation ϕ, γ_0, g_0 from Section 2. For any real x, we also write $\psi(x) = \int_x^\infty e^{-u^2/2} du$. Our key result on β is the following.

Proposition 1. $\beta = \frac{g_1}{g_0} - 1$ *is strictly increasing on* $[2\lambda_1; \sqrt{2 \log n}]$.

We next state and prove some Lemmas used in the proof of Proposition 1 below.

Lemma 17. *The convolution* $g_0 = \phi * \gamma_0$ *satisfies* $g_0'' = \lambda_0^2(g_0 - \phi)$ *as well as*

$$\frac{1}{g_0} \leq \frac{2}{\phi} \qquad and \qquad |g_0 - \phi| \leq \frac{1}{\lambda_0^2}.$$

Proof. The first identity follows by differentiation. One computes $g_0(x)$ by separating the integral in a positive and negative part to get, for any real x,

$$g_0(x) = \frac{\lambda_0 e^{\frac{\lambda_0^2}{2}}}{2\sqrt{2\pi}} \left[e^{\lambda_0 x} \psi(\lambda_0 + x) + e^{-\lambda_0 x} \psi(\lambda_0 - x)\right]. \tag{43}$$

Now combining the standard inequality $(1 - x^{-2})e^{-x^2/2} \leq x\psi(x) \leq e^{-x^2/2}$, for $x > 0$, with the expression of $g_0(0)$ obtained from (43), we get $\frac{1}{2} \leq \frac{g_0}{\phi}(0) \leq 1$ for large enough n. By [12], Lemma 1, the function g_0/ϕ is increasing, which implies the first inequality of the lemma.

The approximation property of ϕ by g_0 is obtained by a Taylor expansion. For any $x, u \in \mathbb{R}$, there exists c between x and $x - u$ such that $\phi(x - u) - \phi(x) = ux\phi(x) + u^2(c^2 - 1)\phi(c)/2$, so that

$$2(g_0(x) - \phi(x)) = \int (2ux\phi(x) + u^2(c^2 - 1)\phi(c))\gamma_0(u)du = \int u^2(c^2 - 1)\phi(c)\gamma_0(u)du,$$

whose absolute value is bounded by $\int u^2|c^2 - 1|\phi(c)\gamma_0(u)du$. This is less than $\int u^2\gamma_0(u)du = \lambda_0^{-2}$. $\square$

Lemma 18. *Let $L_0 = 5\sqrt{2\pi}$. Then for all $x \in [0; \sqrt{2\log(\lambda_0/L_0)}]$,*

$$(\log g_0)'(x) \leq -x/2.$$

Proof. Let $g_{o+}(x) = \int_0^\infty \phi(v + x)\gamma_0(v)dv$ and $g_{o-}(x) = \int_{-\infty}^0 \phi(v + x)\gamma_0(v)dv$. First we check that for any x in the prescribed interval, we have

$$\lambda_0(g_{o+} - g_{o-})(x) \leq -x(\phi(x) - 2/\lambda_0) \leq 0.$$

For any real x, using the inequality $e^v \geq 1 + v$,

$$g_{o-}(x) = \int_0^\infty \phi(x - u)\gamma_0(u)du = \int_0^\infty \phi(x + u)e^{2xu}\gamma_0(u)du$$

$$\geq \int_0^\infty \phi(x + u)(1 + 2xu)\gamma_0(u)du$$

$$\geq g_{o+}(x) + \lambda_0 x \int_0^\infty u\phi(x + u)e^{-\lambda_0 u}du.$$

Setting $\Delta(x) = \int_0^\infty u\phi(x + u)e^{-\lambda_0 u}du$, one can write

$$\Delta(x) = \int_0^\infty u(\phi(x + u) - \phi(x))e^{-\lambda_0 u}du + \phi(x)\int_0^\infty ue^{-\lambda_0 u}du$$

$$= \int_0^\infty u(\phi(x + u) - \phi(x))e^{-\lambda_0 u}du + \phi(x)/\lambda_0^2.$$

As ϕ is 1–Lipshitz, one can bound from below $\phi(x + u) - \phi(x) \geq -u$, which leads to, for any $x \geq 0$,

$$\Delta(x) \geq -\int_0^\infty u^2 e^{-\lambda_0 u}du + \phi(x)/\lambda_0^2 \geq -2/\lambda_0^3 + \phi(x)/\lambda_0^2.$$

This leads to inequality on $g_{o+} - g_{o-}$ above, using that x belongs to the prescribed interval to get the nonpositivity. From this one deduces

$$g_0'(x) = \lambda_0(g_{o+} - g_{o-})(x) \leq -x(\phi(x) - 2/\lambda_0).$$

This now implies

$$\frac{g_0'}{g_0}(x) \leq -x\frac{\phi(x) - 2\lambda_0^{-1}}{\phi(x) + \lambda_0^{-2}}$$

On the prescribed interval $\phi(x) \geq 5/\lambda_0$, so using that $t \to (t-a)/(t+b)$ is increasing,

$$\frac{g_0'}{g_0}(x) \leq -x\frac{5\lambda_0^{-1} - 2\lambda_0^{-1}}{5\lambda_0^{-1} + \lambda_0^{-2}} = -\frac{3x}{5 + \lambda_0^{-1}} \leq -\frac{x}{2},$$

for large enough n, which concludes the proof. $\qquad\square$

Proof of Proposition 1. We will firstly note that if G_1 has a Cauchy$(1/\lambda_1)$ law,

$$|(\log g_1)'(x)| \leq \lambda_1. \qquad (44)$$

Indeed, for any real x, recalling that $\gamma_1(x) = (\lambda_1/\pi)(1 + \lambda_1^2 x^2)^{-1}$, one sees that $\gamma_1'(x)/\gamma_1(x) = (-2\lambda_1^2 x)/(1 + 2\lambda_1^2 x^2)$ and $|\gamma_1'(x)/\gamma_1(x)| \leq 2\sqrt{2}\lambda_1/3$. This implies (44), as

$$|(\log g_1)'(x)| = |\int \phi(x-u)\gamma_1'(u)du|/g_1(x)$$

$$\leq \frac{2\sqrt{2}}{3}\lambda_1 \int \phi(x-u)\gamma_1(u)du/g_1(x) \leq \frac{2\sqrt{2}}{3}\lambda_1 \leq \lambda_1.$$

Let $(x, y) \in [2\lambda_1; \lambda_0/4]^2$ with $x \leq y$. Using Lemma 18 one can find $c \in [x; y]$ with $\log(g_0(x)/g_0(y)) = (x-y)(\log g_0)'(c) \geq (x-y)(-c/2) \geq (y-x)x/2$. On the other hand, by (44) one deduces that for some $c \in [x; y]$, we have $\log(g_1(x)/g_1(y)) = (x-y)(\log g_1)'(c) \leq (y-x)\lambda_1$. Thus for any x, y as before,

$$\frac{g_1(x)}{g_1(y)} \leq e^{(y-x)\lambda_1} \qquad \text{and} \qquad e^{(y-x)\frac{x}{2}} \leq \frac{g_0(x)}{g_0(y)}.$$

As $x \geq 2\lambda_1$ by assumption, this leads to the announced inequality. $\qquad\square$

Lemma 19. *For n large enough, recalling that λ_0 depends on n, we have*

$$(\log g_0)'(x) \geq -x \qquad \text{for any } x > 0, \qquad (45)$$

$$g_0(x) \leq 2\phi(x) \qquad \text{for any } 0 \leq x \leq \lambda_0/2, \qquad (46)$$

$$g_0(x) \lesssim \gamma_0(x) \qquad \text{for any } x \geq \lambda_0/8. \qquad (47)$$

Proof. For any real x, we set $\mu_{0,1}(x) = \int u\frac{\phi(x-u)\gamma_0(u)}{g_0(x)}du$, the expectation of $\gamma_{0,x}$. A direct computation shows, for $x > 0$ that $(\log g_0)'(x) = -x + \mu_{0,1}(x)$. But

$$\mu_{0,1}(x) = \int_0^\infty u\frac{\lambda_0\phi(x-u)e^{-\lambda_0 u}}{2g_0(x)}du + \int_{-\infty}^0 u\frac{\lambda_0\phi(x-u)e^{\lambda_0 u}}{2g_0(x)}du$$

$$= \int_0^\infty u\frac{\lambda_0 e^{-\lambda_0 u}}{2g_0(x)}(\phi(x-u) - \phi(x+u))du$$

$$= \int_0^\infty u\frac{\lambda_0 e^{-\lambda_0 u}}{2g_0(x)}\phi(x+u)(e^{2xu} - 1)du \geq 0,$$

which leads to (45). For the second point, we first prove the identity, for $x > 0$,

$$g_0(x) = \frac{e^{\lambda_0^2/2}}{\sqrt{2\pi}} \psi(\lambda_0)\gamma_0(x) + \phi(x)\frac{\lambda_0}{2}\Big(e^{(\lambda_0-x)^2/2}(\psi(\lambda_0 - x) - \psi(\lambda_0)) \\ + e^{(\lambda_0+x)^2/2}\psi(\lambda_0 + x)\Big).$$

Indeed, $g_0(x) = \int_0^\infty \phi(u)(\gamma_0(x+u) + \gamma_0(x-u))du = \gamma_0(x)\int_0^\infty \phi(u)e^{-\lambda_0 u}du + \int_0^\infty \phi(u)\gamma_0(x-u)du$, for $x > 0$. The first term equals $e^{\lambda_0^2/2}\psi(\lambda_0)\gamma_0(x)/\sqrt{2\pi}$. The second one equals

$$\int_{-x}^\infty \phi(x+v)\gamma_0(v)dv = \phi(x)\int_{-x}^\infty e^{-\frac{v^2}{2}-vx}\gamma_0(v)dv$$

$$= \phi(x)\frac{\lambda_0}{2}\left(\int_{-x}^0 e^{-\frac{v^2}{2}-vx+\lambda_0 v}dv + \int_0^\infty e^{-\frac{v^2}{2}-vx-\lambda_0 v}dv\right)$$

$$= \phi(x)\frac{\lambda_0}{2}\left(\int_0^x e^{-\frac{v^2}{2}+vx-\lambda_0 v}dv + e^{\frac{(x+\lambda_0)^2}{2}}\int_0^\infty e^{-\frac{(v+x+\lambda_0)^2}{2}}dv\right)$$

$$= \phi(x)\frac{\lambda_0}{2}\left(e^{\frac{(\lambda_0-x)^2}{2}}\int_{\lambda_0-x}^{\lambda_0} e^{-\frac{u^2}{2}}du + e^{\frac{(x+\lambda_0)^2}{2}}\psi(x+\lambda_0)\right)$$

which gives the announced identity. If $x \le \lambda_0/2$, using the inequality $y\psi(y) \le e^{-y^2/2}$ for $y > 0$, we have

$$g_0(x) \le \lambda_0^{-1}\gamma_0(x)/\sqrt{2\pi} + \phi(x)(\lambda_0/2)\left[(\lambda_0 - x)^{-1} + (\lambda_0 + x)^{-1}\right].$$

This leads, using $\gamma_0(x)/\lambda_0 \le e^{-\lambda_0^2/2}$ for $x \le \lambda_0/2$, to $g_0(x) \le \phi(x)(1/2 + 1 + 1/2) = 2\phi(x)$.

For the third point, if $x \ge \lambda_0/8$, the first term is bounded as follows:

$$\lambda_0 e^{\lambda_0^2/2}e^{\lambda_0 x}\psi(\lambda_0 + x) \le \lambda_0 e^{\lambda_0^2/2}e^{\lambda_0 x}e^{-\lambda_0^2/2-x^2/2-\lambda_0 x}(\lambda_0 + x)^{-1}$$

$$\le \lambda_0(\lambda_0 + x)^{-1}e^{-x^2/2} \le \lambda_0(9\lambda_0/8)^{-1}e^{-x^2/2}.$$

Now $\psi(\lambda_0 - x) \le e^{-\lambda_0^2/2-x^2/2+\lambda_0 x}(\lambda_0 - x)^{-1} \le 4e^{-\lambda_0^2/2-x^2/2+\lambda_0 x}\lambda_0^{-1}$ if $\lambda_0/8 \le x \le 3\lambda_0/4$, which leads to $g_0(x) \lesssim \phi(x)$. If $x \ge 3\lambda_0/4$ one bounds the second term by $\lambda_0 e^{\lambda_0^2/2-\lambda_0 x} \le \lambda_0 e^{2\lambda_0 x/3-\lambda_0 x} \le \lambda_0 e^{-\lambda_0 x/3}$, so that, for $x \ge \lambda_0/8$,

$$g_0(x) \lesssim \gamma_0(x). \qquad \square$$

The next lemma is useful to control β outside $[2\lambda_1, \sqrt{2\log n}]$.

Lemma 20. *Set $\lambda_1 = 0.05$. For n large enough, for some $C > 0$, we have*

$$(g_1/g_0)(2\lambda_1) < 0.25,$$

$$\beta(x) < 0 \qquad\qquad\qquad \text{for all } x \in [0, 2\lambda_1],$$

$$\beta(x) \gtrsim n/\log n, \qquad\qquad \text{for all } \sqrt{2\log n} \le x \le \lambda_0/2,$$

$$\beta(x) \gtrsim e^{Cn^2}\gamma_1(n)/n \qquad\qquad \text{for all } x \ge \lambda_0/8.$$

Proof. 1) We have $\frac{g_1}{g_0}(2\lambda_1) \leq \frac{\lambda_1\sqrt{2\pi}}{\lambda_0 \int e^{-(u-2\lambda_1)^2/2}e^{-\lambda_0|u|}du}$. For the denominator, we have

$$\int e^{-(u-2\lambda_1)^2/2}e^{-\lambda_0|u|}du \geq \int_0^\infty e^{-(u-2\lambda_1)^2/2-\lambda_0 u}du$$

$$\geq e^{\lambda_0^2/2-2\lambda_1\lambda_0}\int_0^\infty e^{-(u-(2\lambda_1-\lambda_0))^2/2}du$$

$$\geq e^{-2\lambda_1^2}\psi(\lambda_0-2\lambda_1)/(\lambda_0-2\lambda_1)$$

$$\geq e^{-2\lambda_1^2}(\lambda_0-2\lambda_1)^{-1}(1-(\lambda_0-2\lambda_1)^{-2})$$

$$\geq 0.99e^{-2\lambda_1^2}(\lambda_0-2\lambda_1)^{-1}\text{for } n \text{ large enough.}$$

This implies $(g_1/g_0)(2\lambda_1) < 0.25$ for $\lambda_1 = 0.05$.

2) Let $x \in [0, 2\lambda_1]$, using Lemma 17, we have $\beta \leq 2g_1/\phi - 1$. As the last function is increasing as we know from the SAS case, we have $\beta(x) \leq 2(g_1/\phi)(2\lambda_1) - 1$. With (46) we end up with $\beta(x) \leq 4(g_1/\phi)(2\lambda_1) - 1$, which is strictly negative by the first point.

3) Let $x \in [\sqrt{2\log n}, \lambda_0/2]$. With (46), we have $\beta(x) \geq (g_1/2\phi)(x) - 1 \geq (g_1/2\phi)(\sqrt{2\log n}) - 1$, and as $g_1 \gtrsim \gamma_1$, we end up with $\beta(x) \gtrsim n/\log n$.

4) For $x \geq \lambda_0/8$, via (47) we have $\beta(x) + 1 \geq (\gamma_1/\gamma_0)(x) \geq (\gamma_1/\gamma_0)(\lambda_0/8)$ which gives the result. □

6.4. Bounds on moments of the score function

Recall that, for all $k \geq 1$, $\mu \in \mathbb{R}$ and $\alpha \in [0,1]$, $m_k(\mu,\alpha) = E[\beta(Z+\mu)^k]$ where $Z \sim \mathcal{N}(0,1)$, and $\tilde{m}(\alpha) = -m_1(0,\alpha) = -2\int_0^\infty \beta(z,\alpha)\phi(z)dz$.

Proposition 2. *With κ as in* (23), *there exist constants D_1 and D_2 such that for $\alpha \in (\mathcal{C}\log n/n, 1]$, $D_1\zeta^{\kappa-1}g_1(\zeta) \leq \tilde{m}(\alpha) \leq D_2\zeta^{\kappa-1}g_1(\zeta)$. Also, $c \leq \tilde{m}(1) \leq C$ with c, C independent of n.*

Proof. Recall that for $\alpha \in (\mathcal{C}\log n/n, 1]$, we have $\zeta = \beta^{-1}(\alpha^{-1})$ and $\zeta \leq \sqrt{2\log n}$.

$$\tilde{m}(\alpha) = -2\int_0^\infty \frac{\beta(z)}{1+\alpha\beta(z)}\phi(z)dz$$

$$= -2\int_0^\infty \beta(z)\phi(z)dz + 2\int_0^\infty \frac{\alpha\beta^2(z)}{1+\alpha\beta(z)}\phi(z)dz$$

$$= -2\int_0^\infty \beta(z)\phi(z)dz + 2\int_0^\zeta \frac{\alpha\beta^2(z)}{1+\alpha\beta(z)}\phi(z)dz$$

$$+ 2\int_\zeta^\infty \frac{\alpha\beta^2(z)}{1+\alpha\beta(z)}\phi(z)dz$$

$$:= A + B + C$$

- For the first term, with K a positive constant one can write

$$A = 2\int_0^\infty (\phi - \frac{g_1}{g_0}\phi) = 2\int_0^\infty (\phi - \frac{g_1}{g_0}(\phi - g_0 + g_0))$$

$$= 2\int_0^\infty (\phi - g_1) + 2\int_0^\infty \frac{g_1(g_0 - \phi)}{g_0}$$

$$= 0 + 2\int_0^{K\zeta} \frac{g_1(g_0 - \phi)}{g_0} + 2\int_{K\zeta}^\infty \frac{g_1(g_0 - \phi)}{g_0}$$

$$:= (i) + (ii).$$

Using the fact that g_1/ϕ is increasing, we have

$$|(i)| \le 2\lambda_0^{-2}\int_0^{K\zeta} g_1/g_0 \le 4\lambda_0^{-2}\int_0^{K\zeta} g_1/\phi$$

$$\le 4K\zeta g_1(K\zeta)\lambda_0^{-2}/\phi(K\zeta) \lesssim Kn^{K^2-2}\zeta g_1(K\zeta)$$

Taking $K = 6/5$, we end up with $|(i)| \lesssim \zeta n^{-2/5}g_1(6\zeta/5)$ and this term is strictly dominated by $\zeta^{\kappa-1}g_1(\zeta)$. By Lemma 17, and the fact that $g_1 \asymp \gamma_1$, we have:

$$|(ii)| \le 2\int_{K\zeta}^\infty g_1(1 + \phi/g_0) \le 6\int_{K\zeta}^\infty g_1$$

$$\lesssim (6\zeta/5)^{\kappa-1}g_1(6\zeta/5) \text{ using } (23)$$

This term too is dominated by $\zeta^{\kappa-1}g_1(\zeta)$.
- For the second term, we use the fact that on $(0,\zeta)$, $\alpha|\beta| < 1$, so $1 + b_0 \le 1 + \alpha\beta \le 2$, where $b_0 = g_1(2\lambda_1)/2\phi(0) - 1$ does not depend on n, so that

$$B \asymp \int_0^\zeta \alpha\beta^2(z)\phi(z)dz$$

We will now use the fact that, with $h := g_1^2/\phi$, $\int_0^\zeta h(z)dz \le 16h(\zeta)/\zeta$. This is a direct corollary of lemma 4 in [12]. We have, also using (46):

$$\int_0^\zeta \beta^2(z)\phi(z)dz \lesssim \int_0^\zeta (g_1^2/g_0^2)\phi \lesssim \int_0^\zeta g_1^2/\phi$$

$$\lesssim g_1^2(\zeta)/(\zeta\phi(\zeta)) \lesssim \beta(\zeta)g_1(\zeta)/\zeta \lesssim g_1(\zeta)(\alpha\zeta)^{-1}$$

hence $B \lesssim g_1(\zeta)\zeta^{-1}$, dominated by $\zeta^{\kappa-1}g_1(\zeta)$.
- For the last term, we first use the fact that $\alpha\beta(z) < 1 + \alpha\beta(z)$, so that $C \lesssim \int_\zeta^\infty \beta(z)\phi(z)dz$.

$$C \lesssim \int_\zeta^\infty g_1\phi/g_0 \lesssim \int_\zeta^\infty g_1(z)dz \text{ using Lemma 17}$$

$$\asymp \zeta^{\kappa-1}g_1(\zeta) \text{ using } (23)$$

For an upper bound we write

$$C = 2\int_\zeta^{\lambda_0/2} \frac{\alpha\beta^2(z)}{1+\alpha\beta(z)}\phi(z)dz + 2\int_{\lambda_0/2}^\infty \frac{\alpha\beta^2(z)}{1+\alpha\beta(z)}\phi(z)dz =: (i) + (ii).$$

For the first term, using (46), we have for every $z \in [\zeta, \lambda_0/2]$, $\beta(z) \geq \frac{g_1}{2\phi}(z) - 1 \geq \frac{g_1}{4\phi}(z)$ and $\alpha\frac{g_1}{4\phi}(z) \gtrsim \alpha\frac{n}{\log n} \gtrsim 1$, so that

$$(i) \geq 2\int_\zeta^{\lambda_0/2} \frac{\alpha(g_1^2/16\phi^2)(z)}{1+\alpha(g_1/4\phi)(z)}\phi(z)dz$$

$$\gtrsim \int_\zeta^{\lambda_0/2} g_1(z)dz \gtrsim \zeta^{\kappa-1}g_1(\zeta)$$

For the second term, we have

$$(ii) \lesssim \int_{\lambda_0/2}^\infty \beta(z)\phi(z)dz$$

$$\lesssim \int_{\lambda_0/2}^\infty g_1(z)dz \lesssim \lambda_0^{\kappa-1}g_1(\lambda_0) \lesssim \lambda_0^{-1}.$$

Putting the bounds together finally leads to $\tilde{m}(\alpha) \asymp g_1(\zeta)\zeta^{\kappa-1}$.
To prove $\tilde{m}(1) \lesssim \phi(0)/g_1(2\lambda_1)$, write $\tilde{m}(1) = -2\int_0^\infty \phi + 2\int_0^\infty \phi/(1+\beta)$.
Now $\int_0^\infty \phi/(1+\beta) = \int_0^{2\lambda_1} \phi/(1+\beta) + \int_{2\lambda_1}^{\lambda_0/2} \phi/(1+\beta) + \int_{\lambda_0/2}^{+\infty} \phi/(1+\beta)$.
Using that on $[0, 2\lambda_1]$, $1+\beta \geq 1+b_0 = g_1(2\lambda_1)/2\phi(0)$ and (46) and (47), we have

$$\int_0^\infty \phi/(1+\beta) \leq \int_0^{2\lambda_1} \phi/(1+b_0) + \int_{2\lambda_1}^{\lambda_0/2} \phi^2/g_1 + \int_{\lambda_0/2}^\infty \gamma_0\phi/g_1$$

$$\leq \int_0^{2\lambda_1} \phi/(1+b_0) + \int_{2\lambda_1}^\infty \phi^2/g_1 + \int_0^\infty \phi/g_1 \leq C.$$

For the lower bound, recall that $\tilde{m}(1) = -2\int_0^\infty \phi + 2\int_0^\infty \phi/(1+\beta)$ and use Lemma 17 to write $2\int_0^\infty \phi/(1+\beta) \geq \int_0^\infty \phi^2/g_1$ which does not depend on n. $\qquad\square$

Proposition 3. *Let $\alpha \in [\mathcal{C}\log n/n, 1]$.*
   1) For small enough α, we have $m_2(0, \alpha) \lesssim \tilde{m}(\alpha)(\alpha\zeta^\kappa)^{-1}$
   2) For $k = 1$ or 2, for all μ and all α small enough, $m_k(\mu, \alpha) \leq (\alpha \wedge |B_0|/(1+B_0))^{-k}$ with $B_0 = g_1(0)/2\phi(0) - 1$.

Proof. 1) Let $\alpha \in [0; 1]$, we have

$$m_2(0, \alpha) = 2\int_0^\infty \frac{\beta^2(z)}{(1+\alpha\beta(z))^2}\phi(z)dz$$

$$= 2\int_0^\zeta \frac{\beta^2(z)}{(1+\alpha\beta(z))^2}\phi(z)dz + 2\int_\zeta^\infty \frac{\beta^2(z)}{(1+\alpha\beta(z))^2}\phi(z)dz$$

For the first term, as in Proposition 2, and using Proposition 17, we have

$$\int_0^\zeta \frac{\beta^2(z)}{(1+\alpha\beta(z))^2}\phi(z)dz \lesssim \int_0^\zeta \beta^2(z)\phi(z)dz \lesssim g_1(\zeta)(\alpha\zeta)^{-1}$$

For the last term, by the fact that β is increasing on $[\zeta, \sqrt{2\log n}]$, (46) and (47) we have that $\beta > 0$ on $[\zeta, \infty]$ so that

$$\int_\zeta^\infty \frac{\beta^2(z)}{(1+\alpha\beta(z))^2}\phi(z)dz \lesssim 1/\alpha^2 \int_\zeta^\infty \phi(z)dz \lesssim \beta^2(\zeta)\phi(\zeta)/\zeta \lesssim \beta(\zeta)g_1(\zeta)/\zeta$$

hence $m_2(0,\alpha) \lesssim \frac{g_1(\zeta)}{\alpha\zeta}$. Yet $\tilde{m}(\alpha) \asymp \zeta^{\kappa-1}g_1(\zeta)$ when $\alpha \to 0$, which yields the first point.

2) Recall the definition $m_k(\mu,\alpha) = \int \left(\frac{\beta(t)}{1+\alpha\beta(t)}\right)^k \phi(t-\mu)dt$. If $\beta(t) \geq 0$, $\left|\frac{\beta(t)}{1+\alpha\beta(t)}\right| \leq 1/\alpha$. Otherwise we have $|t| < \lambda_0/2$ so using (46) for the numerator leads to $\beta(t) \geq g_1(0)/2\phi(0) - 1 = B_0$ and for the denominator $|1 + \alpha\beta(t)| = 1 + \alpha\beta(t) \geq 1 + \beta(t) \geq 1 + B_0$. $\qquad\square$

6.5. In-probability bounds

Lemma 21. *We take $\alpha = \alpha_1$ and $\zeta = \zeta_1$ as defined by* (41). *There exists $C > 0$ such that*

$$\sup_{\theta\in\ell_0(s_n)} P_\theta(\hat{\zeta} < \zeta) \leq \exp(-C(\log n)^2).$$

Proof. First note that, almost surely, $\hat{\alpha}^{-1} \geq 1 > \beta(2\lambda_1)$ with the help of the first point of Lemma 20, so $\hat{\zeta} = \beta^{-1}(\hat{\alpha}^{-1}) > 2\lambda_1$. Since β is increasing on $(2\lambda_1, \sqrt{2\log n})$ and $\zeta \leq \sqrt{2\log n}$, we have $\{\hat{\zeta} < \zeta\} = \{\hat{\alpha} > \alpha\}$, so $P(\hat{\zeta} < \zeta) = P(\hat{\alpha} > \alpha) = P(\hat{\alpha} > \alpha \cap S(\alpha) > 0) + P(\hat{\alpha} > \alpha \cap S(\alpha) \leq 0)$.

Let us now focus on the event $\{\hat{\alpha} > \alpha\} \cap \{S(\alpha) \leq 0\}$. If $S(\alpha) \leq 0$, since S is decreasing, $S < 0$ on $]\alpha, \hat{\alpha}[$. So the likelihood l is decreasing on $]\alpha, \hat{\alpha}[$. It implies that there exists $\alpha' \in]\alpha, \hat{\alpha}[$ such that $l(\alpha') > l(\hat{\alpha})$. But this contradicts the maximality of $\hat{\alpha}$. Therefore $\{\hat{\alpha} > \alpha\} \cap \{S(\alpha) \leq 0\} = \emptyset$. Hence $P(\hat{\zeta} < \zeta) = P(\hat{\alpha} > \alpha \cap S(\alpha) > 0) \leq P(S(\alpha) > 0)$.

The score function $S(\alpha) = \sum_{i=1}^n \beta(\theta_i + Z_i, \alpha)$ is a sum of independent random variables, each bounded by α^{-1}. We have $P(S(\alpha) > 0) = P(\sum_{i=1}^n W_i > A)$, with $A = -\sum_{i=1}^n m_1(\theta_i, \alpha)$ and $W_i = \beta(\theta_i + Z_i, \alpha) - m_1(\theta_i, \alpha)$ centered variables, bounded by $M = (1 + c)/\alpha$ using the second point of Proposition 3. Setting $V = \sum_{i=1}^n var(W_i)$, Bernstein's inequality gives

$$P(S(\alpha) > 0) \leq \exp(\frac{-A^2}{2(V + \frac{MA}{3})}).$$

Moreover, proceeding as in Lemma 9 in the SAS case, we have $-A \lesssim -n\tilde{m}(\alpha)$ and $V \lesssim n\frac{\tilde{m}(\alpha)}{\alpha}$, so $\left(\frac{A^2}{2(V+\frac{MA}{3})}\right)^{-1} = \frac{V}{A^2} + \frac{M}{3A} \leq \frac{C}{\alpha n\tilde{m}(\alpha)} + \frac{C'}{\alpha n\tilde{m}(\alpha)} \lesssim (\alpha n\tilde{m}(\alpha))^{-1}$

therefore $\frac{A^2}{2(V + \frac{MA}{3})} \gtrsim \alpha n \tilde{m}(\alpha) \gtrsim s_n \gtrsim (\log n)^2$ and finally

$$P(S(\alpha) > 0) \le \exp(-C(\log n)^2). \qquad \square$$

References

[1] A. Babenko and E. Belitser. Oracle convergence rate of posterior under projection prior and Bayesian model selection. *Math. Methods Statist.*, 19(3):219–245, 2010. MR2742927

[2] E. Belitser and N. Nurushev. Needles and straw in a haystack: robust empirical Bayes confidence for possibly sparse sequences. *ArXiv e-print 1511.01803*, Nov. 2015.

[3] I. Castillo. Lower bounds for posterior rates with Gaussian process priors. *Electronic Journal of Statistics*, 2:1281–1299, 2008. MR2471287

[4] I. Castillo. Discussion of: Uncertainty quantification with the horseshoe. *Bayesian Analysis*, 12(4):1250–1253, 2017.

[5] I. Castillo and B. T. Szabó. Spike and Slab Empirical Bayes sparse credible sets. *ArXiv e-print 1808.07721*, aug 2018.

[6] I. Castillo and A. W. van der Vaart. Needles and straw in a haystack: posterior concentration for possibly sparse sequences. *Ann. Statist.*, 40(4):2069–2101, 2012. MR3059077

[7] D. L. Donoho, I. M. Johnstone, J. C. Hoch, and A. S. Stern. Maximum entropy and the nearly black object. *J. Roy. Statist. Soc. Ser. B*, 54(1):41–81, 1992. With discussion and a reply by the authors. MR1157714

[8] E. I. George. The variable selection problem. *J. Amer. Statist. Assoc.*, 95(452):1304–1308, 2000. MR1825282

[9] E. I. George and D. P. Foster. Calibration and empirical Bayes variable selection. *Biometrika*, 87(4):731–747, 2000. MR1813972

[10] S. Ghosal, J. K. Ghosh, and A. W. van der Vaart. Convergence rates of posterior distributions. *Ann. Statist.*, 28(2):500–531, 2000. MR1790007

[11] W. Jiang and C.-H. Zhang. General maximum likelihood empirical Bayes estimation of normal means. *Ann. Statist.*, 37(4):1647–1684, 2009. MR2533467

[12] I. M. Johnstone and B. W. Silverman. Needles and straw in haystacks: empirical Bayes estimates of possibly sparse sequences. *Ann. Statist.*, 32(4):1594–1649, 2004. MR2089135

[13] I. M. Johnstone and B. W. Silverman. EbayesThresh: R Programs for Empirical Bayes Thresholding. *Journal of Statistical Software*, 12(8), 2005. MR2364426

[14] R. Martin and S. G. Walker. Asymptotically minimax empirical bayes estimation of a sparse normal mean vector. *Electron. J. Statist.*, 8(2):2188–2206, 2014. MR3273623

[15] T. J. Mitchell and J. J. Beauchamp. Bayesian variable selection in linear regression. *J. Amer. Statist. Assoc.*, 83(404):1023–1036, 1988. With comments by James Berger and C. L. Mallows and with a reply by the authors. MR0997578

[16] V. Ročková. Bayesian estimation of sparse signals with a continuous spike-and-slab prior. *Ann. Statist.*, 46(1):401–437, 2018. MR3766957

[17] V. Ročková and E. I. George. The spike-and-slab LASSO. *J. Amer. Statist. Assoc.*, 113(521):431–444, 2018. MR3803476

[18] S. van der Pas, B. Szabó, and A. van der Vaart. Adaptive posterior contraction rates for the horseshoe. *Electron. J. Stat.*, 11(2):3196–3225, 2017. MR3705450

[19] S. van der Pas, B. Szabó, and A. van der Vaart. Uncertainty quantification for the horseshoe (with discussion). *Bayesian Anal.*, 12(4):1221–1274, 2017. With a rejoinder by the authors. MR3724985

[20] S. L. van der Pas, J.-B. Salomond, and J. Schmidt-Hieber. Conditions for posterior contraction in the sparse normal means problem. *Electron. J. Stat.*, 10(1):976–1000, 2016. MR3486423

[21] M. Yuan and Y. Lin. Efficient empirical Bayes variable selection and estimation in linear models. *J. Amer. Statist. Assoc.*, 100(472):1215–1225, 2005. MR2236436