

Asymptotic Optimality of One-Group Shrinkage Priors in Sparse High-dimensional Problems

Prasenjit Ghosh* and Arijit Chakrabarti†

Abstract. We study asymptotic optimality of inference in a high-dimensional sparse normal means model using a broad class of one-group shrinkage priors. Assuming that the proportion of non-zero means is known, we show that the corresponding Bayes estimates asymptotically attain the minimax risk (up to a multiplicative constant) for estimation with squared error loss. The constant is shown to be 1 for the important sub-class of “horseshoe-type” priors proving exact asymptotic minimaxity property for these priors, a result hitherto unknown in the literature. An empirical Bayes version of the estimator is shown to achieve the minimax rate in case the level of sparsity is unknown. We prove that the resulting posterior distributions contract around the true mean vector at the minimax optimal rate and provide important insight about the possible rate of posterior contraction around the corresponding Bayes estimator. Our work shows that for rate optimality, a heavy tailed prior with sufficient mass around zero is enough, a pole at zero like the horseshoe prior is not necessary. This part of the work is inspired by van der Pas et al. (2014). We come up with novel unifying arguments to extend their results over the general class of priors. Next we focus on simultaneous hypothesis testing for the means under the additive 0 – 1 loss where the means are modeled through a two-groups mixture distribution. We study asymptotic risk properties of certain multiple testing procedures induced by the class of one-group priors under study, when applied in this set-up. Our key results show that the tests based on the “horseshoe-type” priors asymptotically achieve the risk of the optimal solution in this two-groups framework up to the correct constant and are thus asymptotically Bayes optimal under sparsity (ABOS). This is the first result showing that in a sparse problem a class of one-group priors can exactly mimic the performance of an optimal two-groups solution asymptotically. Our work shows an intrinsic technical connection between the theories of minimax estimation and simultaneous hypothesis testing for such one-group priors.

Keywords: asymptotic minimaxity, posterior contraction, ABOS, sparsity, one-group shrinkage priors, horseshoe prior.

1 Introduction

In this paper we consider inference in a high-dimensional normal means model characterized by sparsity. Suppose we observe a random vector $\mathbf{X} = (X_1, \dots, X_n) \in \mathbb{R}^n$, such that

$$X_i = \theta_i + \epsilon_i, \quad \text{for } i = 1, \dots, n, \quad (1)$$

*Address: Applied Statistics Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Kolkata, Pincode – 700108, India, prasenjit.r@isical.ac.in

†Address: Applied Statistics Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Kolkata, Pincode – 700108, India, arc@isical.ac.in

where the unknown mean parameters $\theta_1, \dots, \theta_n$ denote the effects under investigation and ϵ_i 's are independent $N(0, 1)$ random variables. We assume that the fraction of non-zero θ_i 's is small and tends to zero as the dimension n grows to infinity. The normal means model (1) appears in many applications like signal and image processing, model selection, microarray experiments and nonparametric density estimation. A natural Bayesian approach to model (1) is to use a two-component mixture (with mixing proportion p) of a distribution $\delta_{\{0\}}$ degenerate at 0 and an absolutely continuous distribution F over \mathbb{R} , given by

$$\theta_i \stackrel{i.i.d.}{\sim} (1-p)\delta_{\{0\}} + p \cdot F, \quad i = 1, \dots, n. \quad (2)$$

For sparse modeling, F is typically chosen to be a heavy tailed distribution over \mathbb{R} . Carvalho et al. (2009) commented that a carefully chosen “two-groups” model can be considered a “gold standard” for sparse problems. Use of such two-groups priors, although very natural, often poses daunting computational challenges, especially in high-dimensional problems and complex parametric frameworks. Sometimes it is also possible that most of the parameters are very close to zero, but not exactly equal to zero. So in such a case a continuous prior may be able to capture sparsity in a more flexible manner. Due to these reasons, significant efforts have gone into modeling sparse high-dimensional data in recent times through hierarchical one-group continuous shrinkage priors which are computationally much simpler compared to the two-groups priors. An amazing variety of such one-group shrinkage priors is available in the literature. Some notable examples include the t -prior (Tipping 2001), the Laplace prior (Park and Casella 2008, Hans 2009), the normal–exponential–gamma priors (Griffin and Brown 2005), the horseshoe prior (Carvalho et al. 2009, Carvalho et al. 2010), the three parameter beta normal mixture priors (Armagan et al. 2011), the generalized double Pareto priors (Armagan et al. 2012), the Dirichlet–Laplace priors (Bhattacharya et al. 2012, Bhattacharya et al. 2014) etc. Most of these priors can be expressed as “global-local” scale-mixtures of normals as

$$\theta_i | (\lambda_i^2, \tau^2) \sim N(0, \lambda_i^2 \tau^2), \quad \lambda_i^2 \sim \pi_1(\lambda_i^2), \quad \tau^2 \sim \pi_2(\tau^2). \quad (3)$$

Here τ is the “global shrinkage parameter” accounting for the overall sparsity and λ_i^2 's are the “local shrinkage parameters” helpful in detecting the obvious signals. In some prior specifications, τ is not assigned a non-degenerate prior distribution but is assigned a fixed value or it is estimated from the data. Polson and Scott (2011) argued that, in sparse problems, one should choose the prior π_1 to be appropriately heavy tailed and π_2 should have a large probability near zero. Consequently, the resulting prior distribution for θ_i 's assigns a significant amount of probability near zero while their tails are heavy enough to accommodate large signals. As a result, noise observations are shrunk towards zero while large signals are left almost unshrunk, that is, $E(\theta_i | \mathbf{X}) \approx X_i$ for large X_i 's. The latter property is referred to as the “tail robustness” property and is enjoyed by a broad class of one-group priors. Polson and Scott (2011) advocated the use of such tail robust priors for recovery of sparse signals in high dimensional problems.

Several studies of asymptotic optimality of inference with one-group priors under sparsity have been reported in the literature. In the context of estimation of the mean

vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ in (1), van der Pas et al. (2014) showed that under the usual squared error loss, the horseshoe estimator (Bayes estimator corresponding to the horseshoe prior) asymptotically achieves the minimax risk *up to a multiplicative constant*. They also showed that the corresponding posterior distribution contracts around the true mean vector at the minimax rate and around the horseshoe estimator at least as fast as the minimax rate. Similar results on posterior contraction was shown by Bhattacharya et al. (2012, 2014) for the Dirichlet–Laplace prior. Castillo and van der Vaart (2012) commented that priors are not constructed with the goal of producing posteriors contracting at the minimax rate. However, for theoretical investigations, the minimax rate can be taken as a benchmark and therein lies the importance of such studies. Moreover, such results are also useful for quantification of uncertainty in estimation. See Bhattacharya et al. (2014) and van der Pas et al. (2014) in this context. Datta and Ghosh (2013) followed by Ghosh et al. (2015) considered the problem of testing $H_{0i}: \theta_i = 0$ against $H_{Ai}: \theta_i \neq 0$ simultaneously for $i = 1, \dots, n$, within a decision theoretic framework when each θ_i is *truly* generated from a two-groups model. They studied asymptotic risk properties of certain testing rules derived from the horseshoe and a general class of one-group tail robust priors, respectively, in this two-groups set up. Under an additive 0–1 loss function, they showed that the asymptotic risks of these rules match *up to a multiplicative constant* that of the optimal rule in the two-groups formulation, called the Bayes Oracle, the constant being close to 1. This implies that the optimal two-groups performance may be reasonably approximated by appropriate use of one-group priors.

The above results leave open some natural questions which motivate the present article. van der Pas et al. (2014) asked what aspects of the horseshoe prior are essential towards attaining optimal posterior concentration properties and whether its pole at zero is necessary for that purpose. They also wondered whether a heavy tailed prior distribution with sufficient mass around zero would work as well. Bhattacharya et al. (2012) conjectured earlier that heavy tailed prior distributions, such as the horseshoe, the normal–exponential–gamma and the generalized double Pareto, should possess minimax optimal concentration properties. In the context of multiple testing, a very natural question is whether the one-group priors can *exactly* achieve the performance of the optimal two-groups solution asymptotically.

This paper is an attempt to answer some of the questions mentioned above. Towards that, we investigate optimality properties of a very general class of one-group tail robust shrinkage priors described below. We do not assign any non-degenerate prior distribution to τ and treat it either as a tuning parameter that we are free to choose or estimate it from the data. Specifically, our chosen class of priors is given by

$$\theta_i | (\lambda_i^2, \tau^2) \sim N(0, \lambda_i^2 \tau^2), \quad \lambda_i^2 \sim \pi_1(\lambda_i^2) = K(\lambda_i^2)^{-a-1} L(\lambda_i^2), \quad (4)$$

independently for $i = 1, \dots, n$. Here $K \in (0, \infty)$ is the constant of proportionality, a is a positive real number and $L: (0, \infty) \rightarrow (0, \infty)$ is a measurable non-constant slowly varying function, that is, $\lim_{x \rightarrow \infty} L(\alpha x)/L(x) = 1$, for every fixed $\alpha > 0$. The aforesaid class is rich enough to include a great variety of one-group priors such as the three parameter beta normal mixtures, the generalized double Pareto priors, the inverse–gamma priors, the half–t priors, and many more. In particular, the horseshoe prior,

the Strawderman–Berger prior and the normal–exponential–gamma priors fall inside this class. Our work is based on several versions of a technical report by the authors available in *arXiv* since late 2014 (see Ghosh and Chakrabarti 2015). To the best of our knowledge, it is the first attempt to study posterior concentration properties of one-group shrinkage priors under a general unified framework. Very recently, we have come to know of a related article by van der Pas et al. (2016) which also addresses and answers some of the questions raised in the previous paragraph regarding the estimation problem. A pointed discussion of the work of van der Pas et al. (2016) and its comparison with our work is given in Section 3.1 of this paper.

We first describe our results on estimation of the normal mean vector θ under the squared error loss. Upon appropriate choice of τ based on the proportion of non-zero means, we prove that Bayes estimates based on our chosen class of priors are asymptotically minimax up to some multiplicative constant. The constant becomes 1 for the “horseshoe-type” priors (taking $a = 0.5$ in (4)) when τ is taken to be the proportion of non-zero means or up to a logarithmic factor of it. This is the first result in the literature proving exact asymptotic minimaxity using such priors. When the proportion of non-zero means is unknown, we show that an empirical Bayes version of the estimate, originally proposed by van der Pas et al. (2014) for the horseshoe estimator, attains the minimax risk up to a constant. We further study the rates of contraction of these posteriors assuming the knowledge of the level of sparsity. We show that the posterior distributions contract around the true mean vector at the minimax rate and around the corresponding Bayes estimates at least as fast as this rate. We also derive a lower bound to the total posterior spread for an important sub-family of the *horseshoe-type* priors. This provides important insights regarding the rates of convergence of these posterior distributions towards their respective means and gives important pointers about the optimal choice of τ for achieving optimal posterior concentration rates. Our work establishes the fact that shrinkage priors which are appropriately heavy tailed and put sufficient mass near zero through proper choice of τ are able to attain the minimax optimal rate of contraction. It also shows that one does not need a sharp peak at the origin for this to happen. This settles to a large extent the questions raised in van der Pas et al. (2014) discussed before. We provide some novel unifying arguments exploiting properties of slowly varying functions and extend the work of van der Pas et al. (2014) over for the class of priors under study. As an immediate consequence of our general theoretical results, we settle the conjecture of Bhattacharya et al. (2012) mentioned already.

We now describe our work in the context of multiple testing. As discussed earlier, an important question that remains unanswered in Datta and Ghosh (2013) and Ghosh et al. (2015) is whether testing rules based on one-group priors can asymptotically attain the optimal Bayes risk *exactly*. A key inequality established for our proof of asymptotic minimaxity turns out to be handy in this context. Using some novel arguments based on this inequality we show that the answer to our question is indeed in the affirmative for the “horseshoe-type” priors when τ is proportional to the theoretical proportion of true alternatives. Similar result holds for an empirical Bayes version of the same rule based on the empirical Bayes estimate of τ due to van der Pas et al.

(2014). Such rules are called asymptotically Bayes optimal under sparsity (ABOS), after Bogdan et al. (2011). As far as we know, this is the first such result in the literature which establishes the major fact that when applied to a two-groups formulation, the use of certain one-group priors can exactly mimic the performance of the optimal two-groups answer asymptotically. This reinforces and formally validates the basic wisdom behind proposing the one-group priors as an alternative to the two-groups priors in sparse problems. Our work also demonstrates an interesting technical connection between the theories of minimax estimation and simultaneous hypothesis testing under sparsity.

The paper is organized as follows. Section 2 describes the general class of priors under study. Section 3 presents results on the estimation problem while Section 4 contains the results on the multiple testing problem. Section 5 contains some concluding remarks. Because of space constraints, only proofs of *exact* asymptotic minimaxity and asymptotic Bayes optimality properties of horseshoe-type priors plus some important supporting results are presented in Section 6. Proofs of rest of the results are given in the supplementary material (Ghosh and Chakrabarti, 2016).

2 A General Class of One-Group Tail Robust Priors

In this paper, we consider a general class of one-group shrinkage priors of the kind (4). Then given τ , the posterior distribution of θ_i depends on the data \mathbf{X} only through X_i . From Theorem 1 of Polson and Scott (2011) it follows that the above general class of one-group priors will be “tail robust” in the sense that given $\tau > 0$, $E(\theta_i|X_i, \tau) \approx X_i$ for large X_i ’s, that is, $\lim_{X_i \rightarrow \infty} |E(\theta_i|X_i, \tau) - X_i| = 0$ for any fixed $\tau > 0$. For the theoretical development of this paper, we assume that the slowly varying component $L(\cdot)$ in (4) satisfies the following:

Assumption 1.

1. $\lim_{t \rightarrow \infty} L(t) \in (0, \infty)$, that is, there some exists $c_0(> 0)$ such that $L(t) \geq c_0$ for all $t \geq t_0$, for some $t_0 > 0$, which depends on both L and c_0 .
2. There exists some $0 < M < \infty$ such that $\sup_{t \in (0, \infty)} L(t) \leq M$.

Note that condition (1) of Assumption 1 ensures that $L(\cdot)$ is indeed slowly varying. Ghosh et al. (2015) established that many popular one-group shrinkage priors such as the three parameter beta normal mixtures and the generalized double Pareto priors can be expressed in the above general form. They showed that the corresponding prior distribution of the local shrinkage parameters can be written as in (4). Specifically $\pi_1(\lambda_i^2)$ for the three parameter beta normal mixtures can be written as

$$\pi_1(\lambda_i^2) = K(\lambda_i^2)^{-\beta-1}L(\lambda_i^2) \tag{5}$$

where $L(\lambda_i^2) = (1 + \lambda_i^{-2})^{-(\alpha+\beta)}$, $K = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$ and $\alpha > 0, \beta > 0$. Here $\Gamma(\alpha)$ denotes the gamma function evaluated at $\alpha > 0$. That this $L(\cdot)$ is slowly varying and satisfies Assumption 1 are easily verifiable. The three parameter beta normal mixtures

is rich enough to generalize some well known one-group shrinkage priors, such as the horseshoe ($\alpha = 0.5$, $\beta = 0.5$), the Strawderman-Berger ($\alpha = 1$, $\beta = 0.5$) and the normal-exponential-gamma ($\alpha = 1$, $\beta > 0$) priors; see Armagan et al. (2011). On the other hand the prior $\pi_1(\lambda_i^2)$ for the generalized double Pareto priors can be expressed as

$$\pi_1(\lambda_i^2) = K(\lambda_i^2)^{-\frac{\alpha}{2}-1}L(\lambda_i^2), \quad (6)$$

where $L(\lambda_i^2) = 2^{\frac{\alpha}{2}-1} \int_0^\infty e^{-\beta\sqrt{2u/\lambda_i^2}} e^{-u} u^{(\frac{\alpha}{2}+1)-1} du$, $K = \beta^\alpha/\Gamma(\alpha)$, with $\alpha > 0$, $\beta > 0$. Using the dominated convergence theorem Ghosh et al. (2015) showed that this satisfies condition 1 of Assumption 1, and hence it is slowly varying. Moreover, an easy application of the monotone convergence theorem shows that this $L(\cdot)$ is bounded above by the constant $2^{\frac{\alpha}{2}-1}\Gamma(\frac{\alpha}{2} + 1)$ thereby satisfying condition 2 of Assumption 1. See Section 2 of Ghosh et al. (2015) in this context.

It should be noted that some other well known shrinkage priors such as the inverse-gamma priors (with $\pi_1(\lambda_i^2) \propto (\lambda_i^2)^{-\alpha-1}e^{-\beta/\lambda_i^2}$, $\alpha > 0$, $\beta > 0$) and the half-t priors (with $\pi_1(\lambda_i^2) \propto (1 + \lambda_i^2/\nu)^{-(\nu+1)/2}$, $\nu > 0$) are also covered by our chosen class of priors. We refer the class of priors with $a = 0.5$ in (4) as the *horseshoe-type* priors. The class of horseshoe-type priors contains the three parameter beta normal mixtures with $\alpha = 0.5$, $\beta > 0$ (e.g. horseshoe, Strawderman-Berger), the generalized double Pareto priors with $\alpha = 1$ (e.g. standard double Pareto), and the inverse-gamma priors with $\alpha = 0.5$, etc.

Note that not all members of the class of horseshoe-type priors have a spike at zero like the horseshoe. But appropriately chosen small values of τ ensure that all such priors assign sufficient mass near the origin which is necessary for handling sparsity. Moreover, such a prior may have a tail which is even heavier than that of the horseshoe prior. For example, the inverse-gamma prior with $\alpha = 0.5$ and $\beta = 0.5$ results in a Cauchy distribution for the θ_i 's which does not have a sharp peak at the origin and has a tail heavier than that of the horseshoe prior. The hyperparameter a plays a crucial role in determining the shrinkage profile of such one-group priors. This is why the aforesaid differences do not affect the overall performances of these one-group priors asymptotically as will be seen later in this paper. More specifically, with $a = 0.5$ and upon proper choices of τ , such priors asymptotically attain the same optimal decision theoretic benchmarks in both the problems of simultaneous estimation and multiple hypothesis testing. This will be made more precise in Section 3 and Section 4 of this paper.

Before describing our theoretical results in the forthcoming sections, we introduce here some notations used therein. Let $\{A_n\}$ and $\{B_n\}$ be sequences of non-negative reals such that $B_n \neq 0$ for all large n . We write $A_n \asymp B_n$ to denote $0 < \liminf_n \frac{A_n}{B_n} \leq \limsup_n \frac{A_n}{B_n} < \infty$, and $A_n \lesssim B_n$ to denote that $A_n \leq cB_n$ for all sufficiently large n and some global constant $c > 0$. We write $A_n \sim B_n$ if $\lim_{n \rightarrow \infty} \frac{A_n}{B_n} = 1$, and $A_n = o(B_n)$ if $\lim_{n \rightarrow \infty} \frac{A_n}{B_n} = 0$. If $A_n = cB_n$ for all $n \geq 1$ and some global constant $c > 0$, we denote it as $A_n \propto B_n$.

3 Asymptotic Minimavity and Posterior Contraction Rates

Consider the normal means model in (1), where the mean vector θ is modeled through the general class of one-group priors (4) and is estimated by the corresponding posterior means. For our work we take $a \geq 0.5$. We present in this section the theoretical results involving the mean square error for these Bayes estimates and the spread of the corresponding posterior distributions. Let $\theta_0 = (\theta_{01}, \dots, \theta_{0n})$ denote the true mean vector which is assumed to be sparse in the nearly black sense, that is, $\theta_0 \in \ell_0[q_n]$, where $\ell_0[q_n] = \{\theta \in \mathbb{R}^n : \#\{1 \leq j \leq n : \theta_j \neq 0\} \leq q_n\}$ and $q_n = o(n)$. It is assumed that the maximum number of non-zero components q_n is known. The corresponding minimax error rate under the usual squared ℓ_2 norm $\|\hat{\theta} - \theta\|^2 = \sum_{i=1}^n (\hat{\theta}_i - \theta_i)^2$ is given by (see Donoho et al. 1992),

$$\inf_{\hat{\theta}} \sup_{\theta_0 \in \ell_0[q_n]} \mathbb{E}_{\theta_0} \|\hat{\theta} - \theta_0\|^2 = 2q_n \log\left(\frac{n}{q_n}\right)(1 + o(1)), \text{ as } n \rightarrow \infty. \tag{7}$$

For global-local scale mixtures of normals, $\theta_i | (X_i, \kappa_i, \tau) \stackrel{ind.}{\sim} N((1 - \kappa_i)X_i, (1 - \kappa_i))$, where $\kappa_i = 1/(1 + \lambda_i^2 \tau^2)$ denotes the i -th shrinkage coefficient. Thus, $E(\theta_i | X_i, \tau) = (1 - E(\kappa_i | X_i, \tau))X_i$, for $i = 1, \dots, n$. The resulting vector of posterior means $(E(\theta_1 | X_1, \tau), \dots, E(\theta_n | X_n, \tau))$ will be denoted by $T_\tau(\mathbf{X})$. For notational convenience, let us denote $E(\theta_i | X_i, \tau)$ by $T_\tau(X_i)$. Note that given τ , $T_\tau(\mathbf{X})$ is the Bayes estimate of θ under the squared error loss.

Theorem 1 below gives an upper bound to the mean square error for $T_\tau(\mathbf{X})$. Using this result, we show that for a broad range of choices of τ , depending on the proportion $\frac{q_n}{n}$, $T_\tau(\mathbf{X})$ attains the minimax risk (7) up to a multiplicative constant. In particular, for the class of horseshoe-type priors, $T_\tau(\mathbf{X})$ is shown to be exactly asymptotically minimax. This is presented in Corollary 1. Proofs of both these results are presented in Section 6.

Theorem 1. *Suppose $\mathbf{X} \sim \mathcal{N}_n(\theta_0, \mathbf{I}_n)$, where $\theta_0 \in \ell_0[q_n]$. Consider the class of priors (4) with $a > 0$ where $L(\cdot)$ satisfies Assumption 1. Then*

$$\sup_{\theta_0 \in \ell_0[q_n]} \mathbb{E}_{\theta_0} \|T_\tau(\mathbf{X}) - \theta_0\|^2 \lesssim \begin{cases} q_n \log\left(\frac{1}{\tau^{2a}}\right) + (n - q_n)\tau^{2a} \sqrt{\log\left(\frac{1}{\tau^{2a}}\right)} & \text{if } a \in (0, 1), \\ q_n \log\left(\frac{1}{\tau^{2a}}\right) + (n - q_n)\tau \sqrt{\log\left(\frac{1}{\tau}\right)} & \text{if } a \geq 1, \end{cases}$$

provided $\tau \rightarrow 0$, $q_n \rightarrow \infty$ and $q_n = o(n)$ as $n \rightarrow \infty$.

Corollary 1. *Under the assumptions of Theorem 1, take $\tau = (\frac{q_n}{n})^\alpha$, $\alpha \geq 1$, or $\tau = \frac{q_n}{n} \sqrt{\log(\frac{n}{q_n})}$. Then for $a \geq 0.5$ we have*

$$\sup_{\theta_0 \in \ell_0[q_n]} \mathbb{E}_{\theta_0} \|T_\tau(\mathbf{X}) - \theta_0\|^2 \asymp 2q_n \log\left(\frac{n}{q_n}\right). \tag{8}$$

In particular, for the horseshoe-type priors (with $a = 0.5$) if $\tau = \frac{q_n}{n}$, or, $\tau = \frac{q_n}{n} \sqrt{\log(\frac{n}{q_n})}$

$$\sup_{\theta_0 \in \ell_0[q_n]} \mathbb{E}_{\theta_0} \|T_\tau(\mathbf{X}) - \theta_0\|^2 \sim 2q_n \log\left(\frac{n}{q_n}\right). \tag{9}$$

Observe that (9) is a refinement of the corresponding result on asymptotic minimaxity (up to a multiplicative constant) of the horseshoe estimator obtained by van der Pas et al. (2014) and cannot be improved further. This is achieved by producing a sharper upper bound to the mean square error term due to the non-zero means. Note that the aforesaid asymptotic minimaxity property depends on the knowledge of the proportion $\frac{q_n}{n}$. However, $\frac{q_n}{n}$ may not always be known in practice. In such cases, van der Pas et al. (2014) proposed the following empirical Bayes estimate of τ , given by

$$\hat{\tau} = \max \left\{ \frac{1}{n}, \frac{1}{c_2 n} \sum_{j=1}^n 1\{|X_j| > \sqrt{c_1 \log n}\} \right\} \quad (10)$$

with $c_1 \geq 2$ and $c_2 \geq 1$ being some predetermined finite real numbers. Let $T_{\hat{\tau}}(\mathbf{X})$ be the Bayes estimate $T_{\tau}(\mathbf{X})$ evaluated at $\tau = \hat{\tau}$. Theorem 2 below shows that when the proportion $\frac{q_n}{n}$ is unknown, the empirical Bayes estimates $T_{\hat{\tau}}(\mathbf{X})$ still attain the minimax ℓ_2 risk up to some multiplicative constants, provided $q_n \propto n^{\beta}$ for some $0 < \beta < 1$. Proof of this theorem can be found in the supplementary file.

Theorem 2. *Suppose $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\theta}_0, \mathbf{I}_n)$, where $\boldsymbol{\theta}_0 \in \ell_0[q_n]$ with $q_n \propto n^{\beta}$ for some $0 < \beta < 1$. Consider the class of priors (4) with $a \geq 0.5$ where $L(\cdot)$ satisfies Assumption 1. Then*

$$\sup_{\boldsymbol{\theta}_0 \in \ell_0[q_n]} \mathbb{E}_{\boldsymbol{\theta}_0} \|T_{\hat{\tau}}(\mathbf{X}) - \boldsymbol{\theta}_0\|^2 \asymp 2q_n \log n (1 + o(1)) \text{ as } n \rightarrow \infty. \quad (11)$$

Note that when $q_n \propto n^{\beta}$ for $0 < \beta < 1$, the corresponding minimax error rate under the squared ℓ_2 norm is of the order of $2q_n \log n$. The significance of the above theorem therefore lies in the fact that it provides asymptotic minimaxity property (up to a multiplicative constant) of a completely data adaptive estimate $T_{\hat{\tau}}(\mathbf{X})$ under a very general set up. van der Pas et al. (2016) expressed hope that such a result might be true for a general class of priors. van der Pas et al. (2014) remarked that for asymptotic minimaxity property of the empirical Bayes estimates to hold, a sharp peak near zero like the horseshoe prior is probably not needed. Thus Theorem 2 proves both of these statements to be correct.

The next theorem gives an upper bound to the total posterior variance corresponding to the general class of priors under study when $a \geq 0.5$. This theorem is crucial for proving Theorem 4 about the corresponding posterior concentration rates presented below. Proof of Theorem 3 is presented in the supplementary file.

Theorem 3. *Suppose $\mathbf{X} \sim \mathcal{N}_n(\boldsymbol{\theta}_0, \mathbf{I}_n)$, where $\boldsymbol{\theta}_0 \in \ell_0[q_n]$. Consider the class of priors (4) with $a \geq 0.5$ where $L(\cdot)$ satisfies Assumption 1. Then the total posterior variance corresponding to this general class of priors satisfies*

$$\sup_{\boldsymbol{\theta}_0 \in \ell_0[q_n]} \mathbb{E}_{\boldsymbol{\theta}_0} \sum_{i=1}^n \text{Var}(\theta_i | X_i) \lesssim \begin{cases} q_n \log\left(\frac{1}{\tau^{2a}}\right) + (n - q_n) \tau^{2a} \sqrt{\log\left(\frac{1}{\tau^{2a}}\right)} & \text{if } a \in [0.5, 1), \\ q_n \log\left(\frac{1}{\tau^{2a}}\right) + (n - q_n) \tau \log\left(\frac{1}{\tau}\right) & \text{if } a \geq 1, \end{cases}$$

provided $\tau \rightarrow 0$, $q_n \rightarrow \infty$ and $q_n = o(n)$ as $n \rightarrow \infty$.

Theorem 4 below provides upper bounds on the rates of posterior contraction for our chosen class of priors with $a \geq 0.5$, both around the true θ_0 and the Bayes estimates $T_\tau(\mathbf{X})$.

Theorem 4. *Under the assumptions of Theorem 3, if $\tau = (\frac{q_n}{n})^\alpha$ with $\alpha \geq 1$, or $\tau = \frac{q_n}{n} \sqrt{\log(\frac{n}{q_n})}$, then*

$$\sup_{\theta_0 \in \ell_0[q_n]} \mathbb{E}_{\theta_0} \Pi \left(\left\{ \theta \in \mathbb{R}^n : \|\theta - \theta_0\|^2 > M_n q_n \log \left(\frac{n}{q_n} \right) \right\} | \mathbf{X} \right) \rightarrow 0, \quad (12)$$

and

$$\sup_{\theta_0 \in \ell_0[q_n]} \mathbb{E}_{\theta_0} \Pi \left(\left\{ \theta \in \mathbb{R}^n : \|\theta - T_\tau(\mathbf{X})\|^2 > M_n q_n \log \left(\frac{n}{q_n} \right) \right\} | \mathbf{X} \right) \rightarrow 0, \quad (13)$$

for every $M_n \rightarrow \infty$ as $n \rightarrow \infty$.

Proof. A simple application of Markov’s inequality coupled with Theorem 1 and Theorem 3, when $\tau = (\frac{q_n}{n})^\alpha$ with $\alpha \geq 1$, lead to (12), while (13) follows from Theorem 3 together with Markov’s inequality. \square

Using (12) and (13), it follows that the posterior distributions considered in Theorem 3 contract around both the true θ_0 and the corresponding Bayes estimates $T_\tau(\mathbf{X})$ at least as fast as the minimax ℓ_2 risk. On the other hand, according to Theorem 2.5 of Ghosal et al. (2000), these posterior distributions cannot contract around the true θ_0 faster than the minimax risk. Hence, the rate of contraction of these posterior distributions around θ_0 must be the minimax optimal rate in (7). However, the same may not be true for contraction around the corresponding Bayes estimates meaning that credible sets around these Bayes estimates cannot guarantee the coverage of the true θ_0 . It seems that a deeper investigation by invoking substantially newer techniques and arguments is needed to deal with this issue satisfactorily. One natural step towards that would be to find a lower bound to the total posterior spread. This is a highly non-trivial and mathematically sophisticated result to achieve, as also remarked in van der Pas et al. (2016).

In Theorem 5 below, we establish a lower bound of this kind, albeit for a smaller subclass of priors in (4). We confine our attention to the class of horseshoe-type priors, when $L(\cdot)$ is non-decreasing over $(0, \infty)$. This sub-family of priors covers the three parameter beta normal mixtures with $\alpha = 0.5$, $\beta > 0$ (e.g. horseshoe, Strawderman–Berger), the generalized double Pareto priors with $\alpha = 1$ and the inverse-gamma priors with $\alpha = 0.5$ etc. This result gives important pointers towards the appropriate choice of τ depending on $\frac{q_n}{n}$ for optimal posterior concentration of the horseshoe-type priors. This will be clear form the discussion that follows. Proof of this theorem is presented in the supplementary file.

Theorem 5. *Suppose $\mathbf{X} \sim \mathcal{N}_n(\theta_0, \mathbf{I}_n)$, where $\theta_0 \in \ell_0[q_n]$. Consider the class of priors (4) with $a = 0.5$, where $L(\cdot)$ satisfies Assumption 1 and is non-decreasing over $(0, \infty)$.*

Then, the total posterior variance corresponding to this class of priors, satisfies

$$\inf_{\theta_0 \in \ell_0[q_n]} \mathbb{E}_{\theta_0} \sum_{i=1}^n \text{Var}(\theta_i | X_i) \gtrsim (n - q_n) \tau \sqrt{\log\left(\frac{1}{\tau}\right)}, \quad (14)$$

provided $\tau \rightarrow 0$, $q_n \rightarrow \infty$ and $q_n = o(n)$ as $n \rightarrow \infty$.

If $\tau = (\frac{q_n}{n})^\alpha$ with $0 < \alpha < 1$, the lower bound in (14) exceeds the minimax rate (7). Thus the corresponding posterior distribution may have a sub-optimal spread and a credible set centered around its mean $T_\tau(\mathbf{X})$ may be too large to draw any meaningful inference. If $\tau = (\frac{q_n}{n})^\alpha$ with $\alpha \geq 1$, the lower bound in Theorem 5 is of a smaller order compared to the minimax rate (7). Therefore, for such choices of τ , credible sets centered around $T_\tau(\mathbf{X})$ may be too narrow to guarantee the coverage of the true θ_0 . However, if we choose $\tau = \frac{q_n}{n} \sqrt{\log(n/q_n)}$, then Theorem 3 and Theorem 5 together imply that the total posterior spread is asymptotically of the order of the minimax ℓ_2 rate. This provides a plausibility argument that, for $\tau = \frac{q_n}{n} \sqrt{\log(n/q_n)}$, the posterior distributions contract around their respective means at the minimax ℓ_2 rate. Again, for this choice of τ all the desired upper bounds in Theorem 1 and Theorem 3 are asymptotically of the order of the minimax rate (7). This suggests that for optimal recovery of a sparse normal mean vector as well as for attaining optimal posterior contraction rates, $\tau = \frac{q_n}{n} \sqrt{\log(n/q_n)}$ may be regarded as the desirable choice of τ for the horseshoe-type priors. Similar observations were also made by van der Pas et al. (2014) for the horseshoe prior.

Remark 1. Although the theoretical results presented in this section are built on certain ideas of the proofs of the main theorems of van der Pas et al. (2014), we have to employ novel unifying arguments using properties of slowly varying functions that work for the one-group priors under study. In particular, Lemma 3 presented in Section 6 and Lemmas A.1 – A.2 of the supplementary file, which are at the core of the arguments used for proving most of the theoretical results of this section, are completely independent of the work of van der Pas et al. (2014). However, proofs of Theorem 2 and Theorem 5 follow using some key arguments of van der Pas et al. (2014). Our work shows that some of the technical arguments used in van der Pas et al. (2014) can be used in greater generality.

3.1 A Comparison with the Work of van der Pas et al. (2016)

van der Pas et al. (2016) considered the following scale mixtures of normals as a prior for the θ_i 's in (1):

$$\theta_i | \sigma_i^2 \stackrel{ind.}{\sim} N(0, \sigma_i^2), \quad \sigma_i^2 \stackrel{ind.}{\sim} \pi(\sigma_i^2), \quad i = 1, \dots, n. \quad (15)$$

Their aim was to find answers to the question that which features of one-group shrinkage priors lead to minimax optimal posterior concentration rates under the ℓ_2 norm. Towards that they came up with interesting general conditions on the prior π in (15) which ensure such minimax rate optimality under sparsity. Technically speaking,

their work introduces a new concept of uniformly regularly varying functions and using properties of such functions they obtained their conditions. These conditions require that the tails of π be as heavy as the Laplace and a large amount of mass is assigned around zero relative to the tails, specially when the sparsity is very pronounced. Their conditions were general enough to be satisfied by the class of priors (4) satisfying Assumption 1 of this paper and also the horseshoe+ prior (Bhadra et al. 2015), the normal-inverse Gaussian prior, the normal-gamma prior and the spike-and-slab Lasso prior.

We now compare and explain the relevance and importance of our work in relation to the work of van der Pas et al. (2016). By way of comparison, first it should be noted that we are trying to answer the same general question as in van der Pas et al. (2016). But both these works have been developed independently of each other. Our work is based on an earlier version of a technical report that is available in *arXiv* since late 2014. To the best of our knowledge, this is the first such attempt to answer the aforesaid question within a general unified framework. As compared to them, our work is based on carefully exploiting standard facts on slowly varying functions and the conditions of Assumption 1 turns out to be essential in this approach. When τ is taken as a tuning parameter, our conditions show that an appropriately heavy tailed π_1 coupled with proper choice of τ depending on sparsity (which ensures enough mass near the origin) implies minimax optimal rate of convergence for the corresponding posterior distributions. In this sense, there is a fundamental similarity between the conclusions obtained in these two papers. It should be noted that the class of priors considered in our paper is actually motivated by the observation of Polson and Scott (2011) on one-group tail robust priors. They showed certain undesirable properties of priors having exponential tails such as the lasso or the double exponential, the normal-gamma, the normal-inverse Gaussian etc which do not satisfy the tail robustness criterion. Specifically they showed that in sparse problems such priors are capable of handling sparsity, but this comes at the expense of overshrinking the large observations by non-diminishing amounts. However, van der Pas et al. (2016) showed that for optimal recovery as well as for attaining minimax optimal rates of contraction, their conditions on the prior π in (15) are general enough to be satisfied even by priors having exponential tails like the normal-gamma prior and the normal-inverse Gaussian prior.

Thus the results of van der Pas et al. (2016) prove minimax *rate optimality* for a wider class of priors. However, our work establishes the crucial fact that for appropriately chosen values of the global tuning parameter τ , Bayes estimates obtained from the horseshoe-type priors are exactly asymptotically minimax in terms of the squared ℓ_2 norm. This is achieved through a careful exploitation of the inequalities established in Lemma 2 and Lemma 3 followed by some novel arguments. Based on these arguments we obtain sharper bounds (including the multiplicative factors) to the mean square error terms due to the zero and non-zero θ_i 's. This leads to the stronger result on the minimax optimal recovery of θ by the horseshoe-type priors as presented in Corollary 1. It should further be noted that both our works show that the posterior distributions contract around their respective means at least as fast as the minimax rate (7). But we also prove an interesting lower bound to the total posterior spread for a large collection of horseshoe-type priors. As remarked in van der Pas et al. (2016), this is the only such result of its kind along with Theorem 3.4 of van der Pas et al. (2014). This result

gives important clues about the plausible rates of posterior concentration around their respective means and is the first step towards guaranteeing on coverage of the truth if a credible set centered around the posterior mean is used. This also gives, as commented before, the appropriate choice of τ in this problem. We also prove asymptotic minimaxity results for an empirical Bayes version of the Bayes estimator which is an important contribution of this work in the sense that the proportion of non-zero means may not always be known. van der Pas et al. (2016) expressed hope that such a result might be true for a general class of priors. On the other hand van der Pas et al. (2014) remarked that for asymptotic minimaxity property of the empirical Bayes estimates to hold, a sharp peak near zero like the horseshoe prior is probably not needed. Our work proves both of these statements to be correct.

Last but not the least, our technique of proof shows that the compound decision problems of estimation and testing involving the unknown mean vector $\boldsymbol{\theta}$ in (1), are intimately related. In the following section we prove optimality results for the testing problem using a technique developed for the minimaxity proof. This in fact addresses a question raised in van der Pas et al. (2016) about thresholding rules in the context of simultaneous testing.

4 Asymptotic Bayes Optimality Under Sparsity

Consider the problem of simultaneous testing of hypotheses $H_{0i} : \theta_i = 0$ against $H_{Ai} : \theta_i \neq 0$, for $i = 1, \dots, n$, in the normal means model (1). Our interest is in sparse situations when the dimension n is large and most of the null hypotheses are true. We assume that the unknown θ_i 's are truly generated according to a two-groups model. We establish in this section asymptotic optimality properties of certain multiple testing rules (defined in (23) and (25)) based on one-group continuous shrinkage priors when these rules are applied in this two-groups set up. Interestingly, it turns out that an analogous form of the key inequality derived in Lemma 3 for proving the asymptotic minimaxity and posterior concentration results presented in Section 3, is crucial for the proofs of the results presented in this section. In this way the theoretical developments for the simultaneous estimation and multiple hypothesis testing problems based on our chosen class of priors are intimately connected.

We start by describing the two-groups formulation and the related asymptotic and decision theoretic framework followed by the multiple testing rules under investigation. To describe the two-groups prior, let us introduce a set of latent indicator binary random variables ν_1, \dots, ν_n . Here ν_1, \dots, ν_n are *i.i.d.* Bernoulli(p) random variables, with $\nu_i = 0$ denotes the event that H_{0i} is true while $\nu_i = 1$ corresponds to the event H_{0i} is false. Given $\nu_i = 0$, $\theta_i \sim \delta_{\{0\}}$, the distribution having probability mass 1 at the point 0, while given $\nu_i = 1$, θ_i is assumed to follow a $N(0, \psi^2)$ distribution with $\psi^2 > 0$. Marginalizing over the ν_i 's, given (p, ψ^2) , θ_i 's are *i.i.d.* with the following marginal distribution:

$$\theta_i \stackrel{i.i.d.}{\sim} (1-p)\delta_{\{0\}} + pN(0, \psi^2), \quad i = 1, \dots, n. \quad (16)$$

The marginal distribution of X_i 's is given by the following *two-groups* model:

$$X_i \stackrel{i.i.d.}{\sim} (1 - p)N(0, 1) + pN(0, 1 + \psi^2), \quad i = 1, \dots, n. \tag{17}$$

Under this set up, the given testing problem now boils down to testing simultaneously

$$H_{0i} : \nu_i = 0 \text{ versus } H_{Ai} : \nu_i = 1 \text{ for } i = 1, \dots, n. \tag{18}$$

We define the overall loss of a multiple testing procedure for the above testing problem as the number of misclassifications made by that test. Then the Bayes risk R of a multiple testing procedures is given by

$$R = \sum_{i=1}^n [(1 - p)t_{1i} + pt_{2i}]. \tag{19}$$

Here t_{1i} and t_{2i} denote the probabilities of a type I error and a type II error respectively for the i -th testing problem. Under this set up, Bogdan et al. (2011) derived the Bayes rule which minimizes the Bayes risk (19). It is the test which, for each $i = 1, \dots, n$, declares the i -th null hypothesis H_{0i} to be significant if

$$\pi(\nu_i = 1|X_i) > 0.5, \text{ or equivalently, } X_i^2 > c^2, \tag{20}$$

where $\pi(\nu_i = 1|X_i)$ denotes the posterior probability of the i -th alternative hypothesis H_{Ai} to be true and $c^2 \equiv c_{\psi, f, \delta}^2 = \frac{1+\psi^2}{\psi^2}(\log(1 + \psi^2) + 2 \log(\frac{1-p}{p}))$. The above testing rule is also referred to as the Bayes Oracle since it involves the unknown model parameters p and ψ^2 . By introducing two new parameters $u \equiv u_n = \psi_n^2$ and $v \equiv v_n = \psi_n^2(\frac{1-p_n}{p_n})^2$, Bogdan et al. (2011) considered the following asymptotic scheme given by,

Assumption 2. $p_n \rightarrow 0, u_n = \psi_n^2 \rightarrow \infty$ and $\frac{\log v_n}{u_n} \rightarrow C \in (0, \infty)$ as $n \rightarrow \infty$.

Bogdan et al. (2011) showed that under Assumption 2, the Bayes risk of the Bayes Oracle (20), denoted R_{Opt}^{BO} , has the following asymptotic expression

$$R_{Opt}^{BO} = np(2\Phi(\sqrt{C}) - 1)(1 + o(1)). \tag{21}$$

Within this set up, a multiple testing procedure with Bayes risk R is said to be asymptotically Bayes optimal under sparsity (ABOS), if

$$\frac{R}{R_{Opt}^{BO}} \rightarrow 1 \text{ as } n \rightarrow \infty. \tag{22}$$

Carvalho et al. (2010) modeled the unknown θ_i 's in the normal means model by the horseshoe prior and made the following interesting observation. They observed through simulations that the posterior probability of H_{Ai} being true under the discrete mixture model (2) with an appropriately heavy tailed F can be well approximated by $1 - \hat{\kappa}_i$,

where $\hat{\kappa}_i$ denotes the i -th posterior shrinkage coefficient based on the horseshoe prior. They proposed a natural classification rule under a symmetric 0 – 1 loss function based on the horseshoe prior, given by,

$$\text{reject } H_{0i} \text{ if } 1 - \hat{\kappa}_i > 0.5, \quad i = 1, \dots, n.$$

Carvalho et al. (2010) empirically observed that the estimated misclassification risk for the above thresholding rule is close to that of the optimal Bayes rule within their chosen two-groups framework. Datta and Ghosh (2013) made a formal asymptotic study of the aforesaid phenomena. They investigated the asymptotic risk properties of the thresholding rule above when applied to data generated from the two-groups model (17). Formally, taking τ as a tuning parameter, the thresholding rule considered in Datta and Ghosh (2013) is given by

$$\text{reject } H_{0i} \text{ if } 1 - E(\kappa_i | X_i, \tau) > 0.5, \quad i = 1, \dots, n. \quad (23)$$

They showed that under Assumption 2, the induced decision (23) attains the optimal Bayes risk (21) up to a multiplicative constant if $\tau \sim p$. Ghosh et al. (2015) proved that similar results hold for testing rule (23) using posterior shrinkage coefficients based on a general class of one-group shrinkage priors that includes the horseshoe. Specifically, they considered the class of priors of the form (4) with either (I) $0.5 < a < 1$ or, (II) $a = 0.5$ and $L(t)/\sqrt{\log(t)} \rightarrow 0$ as $t \rightarrow \infty$. Assuming that $\lim_{n \rightarrow \infty} \tau/p \in (0, \infty)$, they showed that the Bayes risk of the multiple testing rules (23), denoted R_{OG} , satisfies as $n \rightarrow \infty$

$$\frac{2\Phi(\sqrt{2aC}) - 1}{2\Phi(\sqrt{C}) - 1} (1 + o(1)) \leq \frac{R_{OG}}{R_{Opt}^{EB}} \leq \frac{2\Phi(\sqrt{\frac{2aC}{\eta(1-\delta)}}) - 1}{2\Phi(\sqrt{C}) - 1} (1 + o(1)), \quad (24)$$

for every fixed $\eta \in (0, \frac{1}{2})$ and $\delta \in (0, 1)$. This was a generalization and improvement over Datta and Ghosh (2013). In case p is unknown, Ghosh et al. (2015) considered a data adaptive version of (23), given by

$$\text{reject } H_{0i} \text{ if } 1 - E(\kappa_i | X_i, \hat{\tau}) > 0.5, \quad i = 1, \dots, n, \quad (25)$$

where $\hat{\tau}$ has already been defined in (10). They showed that, if $p \propto n^{-\beta}$ for $\beta \in (0, 1)$, then R_{OG}^{EB}/R_{Opt}^{BO} has the same upper bound as in (24), where R_{OG}^{EB} denotes the Bayes risk of the empirical Bayes procedure (25).

Thus, the induced decisions (23) and (25) based on a general family of heavy tailed one-group prior distributions, asymptotically attain the Oracle risk (21) up to a multiplicative factor. However, an interesting and natural follow up question is whether this multiplicative factor can be 1. This is equivalent to ask whether such induced multiple testing rules can be ABOS under Assumption 2. This question is investigated in Theorem 6 and Theorem 7 presented below. It is shown that the answer to the aforesaid question of asymptotic Bayes optimality is indeed in the affirmative for the horseshoe prior distributions. Proofs of both these theorems are given in Section 6.

Theorem 6. Let X_1, \dots, X_n be i.i.d. having distribution (17) where the sequence of vectors (ψ^2, p) satisfies Assumption 2. Suppose we wish to test (18) using the classification rule (23) induced by the class of priors (4) with $a \in [0.5, 1)$, where $L(\cdot)$ satisfies Assumption 1. Suppose $p \rightarrow 0, \tau \rightarrow 0$ as $n \rightarrow \infty$ such that $\lim_{n \rightarrow \infty} \tau/p^\alpha \in (0, \infty)$, for $\alpha \geq 1$. Then

$$\lim_{n \rightarrow \infty} \frac{R_{OG}}{R_{Opt}^{BO}} = \frac{2\Phi(\sqrt{2a\alpha}\sqrt{C}) - 1}{2\Phi(\sqrt{C}) - 1}. \tag{26}$$

In particular, for $a = 0.5$ and $\alpha = 1$ we have,

$$\lim_{n \rightarrow \infty} \frac{R_{OG}}{R_{Opt}^{BO}} = 1.$$

Theorem 6 therefore shows that, if τ is asymptotically of the order of p , such induced decisions based on the horseshoe-type priors asymptotically attain the risk of the Bayes Oracle up to the right constant, and hence, are ABOS. This is the first result of its kind as far as we know, and sharpens all known theoretical results about asymptotic risk properties of this kind of induced decisions in a two-groups framework. It shows that a broad class of one-group shrinkage priors can asymptotically mimic the overall performance of a more natural two-groups model. This resolves a long awaited technical question in the Bayesian literature that whether the optimal “two-groups answer” can exactly be matched by the “one-group solution”. This demonstrates in a very strong sense that appropriately chosen one-group shrinkage priors can be used with confidence in an inference problem under sparsity, where a two-groups modeling is more natural in a strict mathematical sense.

Note that the above limiting expression in (26) gives exact asymptotic expressions for the ratio of Bayes risk of the induced testing rule (23) to that of the Bayes Oracle, for any $a \in [0.5, 1)$ and any $\alpha \geq 1$. It also shows that the limiting value of the ratio of Bayes risks is an increasing function of both $\alpha \geq 1$ and $a \in [0.5, 1)$. This suggests that the hyperparameter a in the definition of $\pi_1(\lambda_i^2)$ in (4) may be set at $a = 0.5$ as a default choice for the present multiple testing problem.

The above theorem also establishes a necessary and sufficient condition on the choice of τ for the horseshoe-type priors for achieving optimal or near optimal performance for the present multiple testing problem. It will be shown in Section 6 that the type I and type II error probabilities of the i -th decision in (23), do not depend on i . Thus, the Bayes risk of the induced decision (23) is given by $R_{OG} = np(\frac{1-p}{p}t_1 + t_2)$ (using (19)). Suppose that $\lim_{n \rightarrow \infty} \frac{\tau}{p^\alpha} \in (0, \infty)$ with $\alpha > 0$. If $\alpha \in (0, 1)$, then using (53) of Section 6, one can show easily that $\frac{t_1}{p}$ diverges to ∞ as $n \rightarrow \infty$. Thus, $R_{OG}/R_{Opt}^{BO} \rightarrow \infty$ as $n \rightarrow \infty$ implying that the aforesaid Bayesian optimality property fails to hold in such cases. On the other hand, the limiting value of R_{OG}/R_{Opt}^{BO} as given by (26), is non-decreasing in $\alpha \geq 1$, thereby attaining its minimum value 1 at $\alpha = 1$. Therefore, for the horseshoe-type priors, the optimal choice of τ is asymptotically of the order of p , when the latter is assumed to be known.

The next theorem shows that, within the asymptotic framework of Bogdan et al. (2011), when $p_n \propto n^{-\beta}$ for $0 < \beta < 1$, the empirical Bayes induced testing procedures

(25) based on the horseshoe-type priors asymptotically attain the optimal Bayes risk in (21) up to the correct constant, and hence, are ABOS. Proof of this result is given in Section 6.

Theorem 7. *Let X_1, \dots, X_n be i.i.d. having distribution (17) and suppose Assumption 2 is satisfied by the sequence of vectors (ψ^2, p) , where $p \propto n^{-\beta}$ for some $0 < \beta < 1$. Suppose we wish to test (18) using the classification rule (25) induced by the class of priors (4) with $a = 0.5$, where $L(\cdot)$ satisfies Assumption 1. Then the Bayes risk of the empirical Bayes testing procedure (25), denoted R_{OG}^{EB} , satisfies*

$$\lim_{n \rightarrow \infty} \frac{R_{OG}^{EB}}{R_{Opt}^{BO}} = 1,$$

that is, the corresponding empirical Bayes decisions (25) will be ABOS.

Ghosh et al. (2015) showed that under the assumptions of Theorem 7, $\frac{\hat{\tau}}{p}$ converges to $2\beta_0$ in probability as n tends to ∞ , for some $\beta_0 > 0$. Thus the asymptotic Bayes optimality property of the empirical procedure (25) based on the horseshoe-type priors is in concordance with the conclusions of Theorem 6. The importance of this result lies in the fact that it is the first such result in the literature showing a completely data adaptive procedure based on one-group shrinkage priors to exactly achieve the optimum decision theoretic benchmark in a high dimensional problem under sparsity.

5 Discussion

In this paper, we study a high-dimensional sparse normal means model. Our focus is on various theoretical properties of inference using a general class of one-group tail robust shrinkage priors. We first show, assuming that the proportion of non-zero means is known, that the Bayes estimates arising out of this general class asymptotically attain the minimax risk with respect to the ℓ_2 norm, possibly up to some multiplicative constants. In particular, for the sub-class of ‘‘horseshoe-type’’ priors exact asymptotic minimaxity is established. In case the level of sparsity is unknown, an empirical Bayes version of the estimator is shown to achieve the minimax rate. Optimal rates of posterior contraction of these prior distributions around the truth in terms of the minimax error rate has also been established for proper choice of τ depending on the knowledge of sparsity. We provide a novel unifying theoretical treatment that holds for a very broad class of one-group shrinkage priors. Another major contribution of this work is to show that shrinkage priors which are appropriately heavy tailed are good enough in order to attain the minimax optimal rate of contraction and that one does not need a pole at the origin for this to happen, provided that τ is carefully chosen. This provides a partial answer to the question raised in van der Pas et al. (2014) already discussed in the introduction. We believe that one possible reason for such good performance of the kind of one-group shrinkage priors studied in this paper, is their ability to shrink the noise observations back to the origin, while leaving the large signals mostly unshrunk.

In the latter half of this paper, we also study, within a decision theoretic framework, the asymptotic risk properties of induced decisions based on our chosen class of priors

in the context of multiple testing. A major theoretical contribution of this work is to show that such induced decisions based on the horseshoe-type priors are asymptotically Bayes optimal under sparsity. To the best of our knowledge, this is the first such result in the literature where the optimal two-groups answer can be exactly achieved asymptotically by an one-group formulation under the assumption of sparsity. Our work also demonstrates an interesting technical connection between the theories of minimax estimation and simultaneous hypothesis testing under sparsity. An important consequence of this work is to theoretically establish the fact that for the class of horseshoe-type priors with $a = 0.5$, the optimal choice of τ in the multiple testing problem should be asymptotically of the order of the proportion of true alternatives p , when p is assumed to be known. Moreover, the present work also provides strong theoretical support in favor of using $a = 0.5$ as a default choice in our one-group prior specification.

Over the past few years, one-group shrinkage priors have been gaining increasing popularity in the Bayesian literature for modeling sparse high-dimensional data instead of the more natural two-groups models. We hope that the present work is a useful contribution towards understanding various theoretical properties of such one-group priors. It provides important theoretical justifications in favor of the use of such kind of one-group priors together with some useful guidelines regarding the choice of the underlying hyperparameters. We are hopeful that the techniques employed in the present article would prove to be important ingredients for optimality studies of such one-group shrinkage priors. However, an interesting problem that remains open till date is to show asymptotic optimality properties of a full Bayes approach by assigning a hyperprior to the global shrinkage parameter τ . This applies in equal measure in both the problems of simultaneous estimation and testing. We hope to address this problem elsewhere in future.

6 Proofs

We start this section by first noting that under (4), the shrinkage coefficients κ_i 's are independently distributed given (\mathbf{x}, τ) , with the posterior distribution of κ_i only depending on (x_i, τ) and is given by

$$\pi(\kappa_i|x_i, \tau) \propto \kappa_i^{a+\frac{1}{2}-1}(1-\kappa_i)^{-a-1}L\left(\frac{1}{\tau^2}\left(\frac{1}{\kappa_i}-1\right)\right)e^{-\frac{\kappa_i x_i^2}{2}}, \quad \kappa_i \in (0, 1).$$

We next present three very important lemmas which are crucial for deriving most of the results of this paper. Lemma 1 gives an upper bound to the tail probability $\Pr(\kappa_i > \eta|x_i, \tau)$, while Lemma 2 presents an upper bound to $E(1-\kappa_i|x_i, \tau)$. Lemma 3 provides a non-trivial upper bound to the difference of the coordinate-wise posterior means $T_\tau(x_i)$ and the observation x_i . This upper bound is one of the core facts used for proving most of the results described in this paper. For notational convenience, we suppress the dependence of κ_i and x_i on i while presenting these lemmas. Lemma 1 follows as an immediate consequence of Theorem 5 of Ghosh et al. (2015) and hence the proof is skipped.

Lemma 1. Consider the class of priors (4) with $a > 0$ where $L(\cdot)$ satisfies Assumption 1. For every fixed $\tau > 0$ and each fixed $\eta \in (0, 1)$, $\delta \in (0, 1)$, the posterior distribution of the shrinkage coefficients $\kappa = 1/(1 + \lambda^2\tau^2)$ satisfies the following:

$$\Pr(\kappa > \eta|x, \tau) \leq \frac{H(a, \eta, \delta)e^{-\frac{\eta(1-\delta)x^2}{2}}}{\tau^{2a}\Delta(\tau^2, \eta, \delta)}, \text{ uniformly in } x \in \mathbb{R},$$

where

$$\begin{aligned} \Delta(\tau^2, \eta, \delta) &= \Psi(\tau^2, \eta, \delta)L\left(\frac{1}{\tau^2}\left(\frac{1}{\eta\delta} - 1\right)\right), \\ \Psi(\tau^2, \eta, \delta) &= \frac{\int_{\frac{1}{\tau^2}\left(\frac{1}{\eta\delta} - 1\right)}^{\infty} t^{-(a+\frac{1}{2}+1)}L(t)dt}{\left(a + \frac{1}{2}\right)^{-1}\left(\frac{1}{\tau^2}\left(\frac{1}{\eta\delta} - 1\right)\right)^{-(a+\frac{1}{2})}L\left(\frac{1}{\tau^2}\left(\frac{1}{\eta\delta} - 1\right)\right)}, \text{ and} \\ H(a, \eta, \delta) &= \frac{\left(a + \frac{1}{2}\right)(1 - \eta\delta)^a}{K(\eta\delta)^{\left(a+\frac{1}{2}\right)}}. \end{aligned}$$

Furthermore, $\lim_{\tau \rightarrow 0} \Delta(\tau^2, \eta, \delta)$ is a finite positive quantity for every fixed $\eta \in (0, 1)$ and $\delta \in (0, 1)$.

Lemma 2. Consider the class of priors (4) with $a > 0$ where $L(\cdot)$ satisfies Assumption 1. Then, for each fixed $x \in \mathbb{R}$ and every fixed $\tau > 0$, we have

$$E(1 - \kappa|x, \tau) \leq \begin{cases} KMa^{-1}(1 - a)^{-1}e^{\frac{x^2}{2}}\tau^{2a}(1 + o(1)) & \text{if } a \in (0, 1), \\ 2M(2a - 1)^{-1}e^{\frac{x^2}{2}}\tau(1 + o(1)) & \text{if } a \geq 1, \end{cases}$$

where the $o(1)$ term depends only on τ such that $\lim_{\tau \rightarrow 0} o(1) = 0$.

Proof. The case for $a \in (0, 1)$ follows directly from Theorem 3 of Ghosh et al. (2015). So, let us now consider the case $a \geq 1$.

First we observe that

$$\begin{aligned} E(1 - \kappa|x, \tau) &= \frac{\int_0^1 \kappa^{a+\frac{1}{2}-1}(1 - \kappa)^{-a}L\left(\frac{1}{\tau^2}\left(\frac{1}{\kappa} - 1\right)\right)e^{-\frac{\kappa x^2}{2}}d\kappa}{\int_0^1 \kappa^{a+\frac{1}{2}-1}(1 - \kappa)^{-a-1}L\left(\frac{1}{\tau^2}\left(\frac{1}{\kappa} - 1\right)\right)e^{-\frac{\kappa x^2}{2}}d\kappa} \\ &\leq \frac{e^{\frac{x^2}{2}} \int_0^1 \kappa^{a+\frac{1}{2}-1}(1 - \kappa)^{-a}L\left(\frac{1}{\tau^2}\left(\frac{1}{\kappa} - 1\right)\right)d\kappa}{\int_0^1 \kappa^{a+\frac{1}{2}-1}(1 - \kappa)^{-a-1}L\left(\frac{1}{\tau^2}\left(\frac{1}{\kappa} - 1\right)\right)d\kappa}. \end{aligned} \tag{27}$$

Then, using the change of variable $t = \frac{1}{\tau^2}\left(\frac{1}{\kappa} - 1\right)$ in the numerator of (27) and applying the fact $\int_0^1 \kappa^{a+\frac{1}{2}-1}(1 - \kappa)^{-a-1}L\left(\frac{1}{\tau^2}\left(\frac{1}{\kappa} - 1\right)\right)d\kappa = K^{-1}\tau^{-2a}(1 + o(1))$ (see Lemma A.5 of Ghosh et al. 2015) to its denominator, we obtain

$$E(1 - \kappa|x, \tau) \leq Ke^{\frac{x^2}{2}} \int_0^{\infty} \frac{t\tau^2}{1 + t\tau^2} \frac{t^{-a-1}L(t)}{\sqrt{1 + t\tau^2}} dt(1 + o(1))$$

$$\begin{aligned}
 &= Ke^{\frac{x^2}{2}} \left(\int_0^1 + \int_1^\infty \right) \frac{t\tau^2}{1+t\tau^2} \frac{t^{-a-1}L(t)}{\sqrt{1+t\tau^2}} dt(1+o(1)) \\
 &= Ke^{\frac{x^2}{2}} (J_{1\tau} + J_{2\tau}), \text{ say.}
 \end{aligned} \tag{28}$$

Now note that the term $J_{1\tau}$ can be bounded above as

$$J_{1\tau} \leq \tau^2 \int_0^1 t^{-a-1}L(t)dt(1+o(1)) \leq K^{-1}\tau^2(1+o(1)). \tag{29}$$

Also, since $\frac{t\tau^2}{1+t\tau^2} \frac{t^{-a-1}L(t)}{\sqrt{1+t\tau^2}} \leq M\tau t^{-(a-\frac{1}{2})}$, we have

$$J_{2\tau} \leq M\tau \int_1^\infty t^{-(a-\frac{1}{2})}dt(1+o(1)) = 2M(2a-1)^{-1}\tau(1+o(1)). \tag{30}$$

Combining (27)–(30), the desired upper bound for $a \geq 1$ follows. That the corresponding $o(1)$ term depends only on τ and tends to 0 as $\tau \rightarrow 0$ is clear from the proof. \square

Lemma 3. Consider the class of priors (4) satisfying Assumption 1, with $a > 0$. Then, for each fixed $\tau \in (0, 1)$ and given any $c > 2$, the absolute difference between the coordinate-wise posterior mean $T_\tau(x)$ and an observation x , can be bounded above by a non-negative real valued function $h(\cdot, \tau)$, depending on c satisfying the following:

for any $\rho > c$,

$$\lim_{\tau \downarrow 0} \sup_{|x| > \sqrt{\rho \log\left(\frac{1}{\tau^{2a}}\right)}} h(x, \tau) = 0. \tag{31}$$

Proof. By definition,

$$\begin{aligned}
 |T_\tau(x) - x| &= |xE(\kappa|x, \tau)| \\
 &= \left| \frac{\int_0^1 \kappa \cdot \kappa^{a+\frac{1}{2}-1}(1-\kappa)^{-a-1}L\left(\frac{1}{\tau^2}\left(\frac{1}{\kappa}-1\right)\right)e^{-\kappa x^2/2}d\kappa}{\int_0^1 \kappa^{a+\frac{1}{2}-1}(1-\kappa)^{-a-1}L\left(\frac{1}{\tau^2}\left(\frac{1}{\kappa}-1\right)\right)e^{-\kappa x^2/2}d\kappa} \right|.
 \end{aligned}$$

Now observe that, for each fixed $\tau \in (0, 1)$, the function $|T_\tau(x) - x|$ is symmetric in x and it takes the value 0 when $x = 0$. Therefore, it would be enough to find any non-negative function $h(x, \tau)$ that is symmetric in x and satisfies the stated conditions. Hence, without any loss of generality, let us assume that $|x| > 0$.

Let us fix any $\eta, \delta \in (0, 1)$. Observe that, for any $x \in \mathbb{R}$ and for each fixed $\tau \in (0, 1)$,

$$|T_\tau(x) - x| \leq I_1(x, \tau) + I_2(x, \tau) \tag{32}$$

where $I_1(x, \tau) = |xE(\kappa 1\{\kappa \leq \eta\}|x, \tau)|$ and $I_2(x, \tau) = |xE(\kappa 1\{\kappa > \eta\}|x, \tau)|$.

Now using the variable transformation $t = \frac{1}{\tau^2}\left(\frac{1}{\kappa}-1\right)$ in $I_1(x, \tau)$ we have the following:

$$I_1(x, \tau) = \left| \frac{x \int_0^\eta \kappa \cdot \kappa^{a+\frac{1}{2}-1}(1-\kappa)^{-a-1}L\left(\frac{1}{\tau^2}\left(\frac{1}{\kappa}-1\right)\right)e^{-\kappa x^2/2}d\kappa}{\int_0^1 \kappa^{a+\frac{1}{2}-1}(1-\kappa)^{-a-1}L\left(\frac{1}{\tau^2}\left(\frac{1}{\kappa}-1\right)\right)e^{-\kappa x^2/2}d\kappa} \right|$$

$$\begin{aligned} &\leq \left| \frac{x \int_{\frac{1}{\tau^2}(\frac{1}{\eta}-1)}^{\infty} \frac{1}{(1+t\tau^2)^{3/2}} t^{-a-1} L(t) e^{-\frac{x^2}{2(1+t\tau^2)}} dt}{\int_{\frac{t_0}{\tau^2}}^{\infty} \frac{1}{(1+t\tau^2)^{1/2}} t^{-a-1} L(t) e^{-\frac{x^2}{2(1+t\tau^2)}} dt} \right| \\ &= J_1(x, \tau), \text{ say.} \end{aligned} \tag{33}$$

Observe that $\frac{t_0}{\tau^2} > t_0$ as $\tau^2 < 1$. Hence $L(t) \geq c_0$ for all $t \geq \frac{t_0}{\tau^2}$. Also, note that L is bounded above by the constant $M > 0$. Therefore, using the transformation $u = \frac{x^2}{1+t\tau^2}$ in both the numerator and the denominator of $J_1(x, \tau)$ in (33), and writing $s = \frac{1}{1+t_0} \in (0, 1)$, we obtain

$$\begin{aligned} J_1(x, \tau) &\leq \frac{M}{c_0} \left| x \frac{\int_0^{\eta x^2} e^{-u/2} \left(\frac{u}{x^2}\right)^{3/2} \left(\frac{1}{\tau^2} \left(\frac{x^2}{u} - 1\right)\right)^{-a-1} \frac{x^2}{\tau^2 u^2} du}{\int_0^{sx^2} e^{-u/2} \left(\frac{u}{x^2}\right)^{1/2} \left(\frac{1}{\tau^2} \left(\frac{x^2}{u} - 1\right)\right)^{-a-1} \frac{x^2}{\tau^2 u^2} du} \right| \\ &\leq \frac{M}{c_0} \left| \frac{1}{x} \cdot \frac{\int_0^{\eta x^2} e^{-u/2} u^{a+3/2-1} \left(1 - \frac{u}{x^2}\right)^{-a-1} du}{\int_0^{sx^2} e^{-u/2} u^{a+1/2-1} \left(1 - \frac{u}{x^2}\right)^{-a-1} du} \right|. \end{aligned}$$

Note that when $0 < u < \eta x^2$ we have $0 < \frac{u}{x^2} < \eta < 1$, that is, $1 - \eta < 1 - \frac{u}{x^2} < 1$. Similarly, we have $1 - s < 1 - \frac{u}{x^2} < 1$ when $0 < u < sx^2$. Using these observations we get

$$\begin{aligned} J_1(x, \tau) &\leq \frac{M}{c_0(1-\eta)^{1+a}} \left| \frac{1}{x} \cdot \frac{\int_0^{\eta x^2} e^{-u/2} u^{a+3/2-1} du}{\int_0^{sx^2} e^{-u/2} u^{a+1/2-1} du} \right| \\ &\leq \frac{M}{c_0(1-\eta)^{1+a}} \left| \frac{1}{x} \cdot \frac{\int_0^{\infty} e^{-u/2} u^{a+3/2-1} du}{\int_0^{sx^2} e^{-u/2} u^{a+1/2-1} du} \right| \\ &= h_1(x, \tau), \text{ say,} \end{aligned} \tag{34}$$

where $h_1(x, \tau) = C_* \left[\int_0^{sx^2} e^{-u/2} u^{a+1/2-1} du \right]^{-1}$ for some global constant $C_* \equiv C_*(a, \eta, L) > 0$ which is independent of both x and τ .

Next, from Lemma 1 it follows that

$$I_2(x, \tau) \leq |x \Pr(\kappa > \eta | x, \tau)| \leq \left| x \frac{H(a, \eta, \delta) e^{-\frac{\eta(1-\delta)x^2}{2}}}{\tau^{2a} \Delta(\tau^2, \eta, \delta)} \right| = h_2(x, \tau), \text{ say.} \tag{35}$$

Let $h(x, \tau) = h_1(x, \tau) + h_2(x, \tau)$. Therefore combining (32), (33), (34) and (35), we finally obtain for every $x \in \mathbb{R}$ and $\tau \in (0, 1)$,

$$|T_\tau(x) - x| \leq h(x, \tau). \tag{36}$$

Note that $h(x, \tau)$ defined above is symmetric in x about the origin. Now observe that the function $h_1(x, \tau)$ is strictly decreasing in $|x|$ for $|x| > 0$. Also, the function $x \mapsto |x| e^{-\eta(1-\delta)x^2/2}$ is strictly decreasing in x when $|x| > 1/\sqrt{\eta(1-\delta)}$. Using these facts

and the definition of the function h , it follows after some straightforward calculation that

$$\lim_{\tau \rightarrow 0} \sup_{|x| > \sqrt{\rho \log\left(\frac{1}{\tau^{2a}}\right)}} h(x, \tau) = \begin{cases} 0 & \text{if } \rho > \frac{2}{\eta(1-\delta)} \\ \infty & \text{otherwise.} \end{cases} \tag{37}$$

Note that by choosing η appropriately close to 1 and δ sufficiently close to 0, any real number larger than 2 can be expressed in the form $\frac{2}{\eta(1-\delta)}$. Hence, given any $c > 2$, let us choose $\eta, \delta \in (0, 1)$ such that $c = \frac{2}{\eta(1-\delta)}$. Clearly, the choice of $h(\cdot, \tau)$ depends on c . This, coupled with (36) and (37), completes the proof of Lemma 3. \square

Proof of Theorem 1. Define $\tilde{q}_n = \#\{i : \theta_{0i} \neq 0\}$. Let us split the mean square error $\mathbb{E}_{\theta_0} \|T_\tau(\mathbf{X}) - \theta_0\|^2 = \sum_{i=1}^n \mathbb{E}_{\theta_{0i}} (T_\tau(X_i) - \theta_{0i})^2$ as

$$\sum_{i=1}^n \mathbb{E}_{\theta_{0i}} (T_\tau(X_i) - \theta_{0i})^2 = \sum_{i:\theta_{0i} \neq 0} \mathbb{E}_{\theta_{0i}} (T_\tau(X_i) - \theta_{0i})^2 + \sum_{i:\theta_{0i} = 0} \mathbb{E}_{\theta_{0i}} (T_\tau(X_i) - \theta_{0i})^2. \tag{38}$$

Non-zero means: For $\theta_{0i} \neq 0$, using the Cauchy-Schwartz inequality and the fact that $\mathbb{E}_{\theta_{0i}} (X_i - \theta_{0i})^2 = 1$, we get

$$\mathbb{E}_{\theta_{0i}} (T_\tau(X_i) - \theta_{0i})^2 \leq [\sqrt{\mathbb{E}_{\theta_{0i}} (T_\tau(X_i) - X_i)^2} + 1]^2. \tag{39}$$

Let $\zeta_\tau = \sqrt{2 \log\left(\frac{1}{\tau^{2a}}\right)}$. Fix any $c > 1$ and choose any $\rho > c$. Now using Lemma 3, there exists a non-negative real-valued function h , depending on c , such that $|T_\tau(x) - x| \leq h(x, \tau)$ for all $x \in \mathbb{R}$, where the function h satisfies (31). Also note that $|T_\tau(x) - x| \leq x$ for all $x \in \mathbb{R}$. Using these facts and further noting that $\zeta_\tau \rightarrow \infty$ when $\tau \rightarrow 0$, we obtain as $\tau \rightarrow 0$

$$\begin{aligned} \mathbb{E}_{\theta_{0i}} (T_\tau(X_i) - X_i)^2 &= \mathbb{E}_{\theta_{0i}} [(T_\tau(X_i) - X_i)^2 1\{|X_i| \leq \rho\zeta_\tau\}] \\ &\quad + \mathbb{E}_{\theta_{0i}} [(T_\tau(X_i) - X_i)^2 1\{|X_i| > \rho\zeta_\tau\}] \\ &\leq \rho^2 \zeta_\tau^2 (1 + o(1)). \end{aligned} \tag{40}$$

Note that (40) holds uniformly for any i such that $\theta_i \neq 0$. Therefore, using (39) and (40), we obtain as $\tau \rightarrow 0$

$$\sum_{i:\theta_{0i} \neq 0} \mathbb{E}_{\theta_{0i}} (T_\tau(X_i) - \theta_{0i})^2 \lesssim \tilde{q}_n \zeta_\tau^2. \tag{41}$$

Zero means: We treat the cases $0 < a < 1$ and $a \geq 1$ separately.

Case (I) When $0 < a < 1$: For $\theta_{0i} = 0$, we split the corresponding mean square error as

$$\mathbb{E}_0 [T_\tau^2(X_i)] = \mathbb{E}_0 [T_\tau^2(X_i) 1\{|X_i| \leq \zeta_\tau\}] + \mathbb{E}_0 [T_\tau^2(X_i) 1\{|X_i| > \zeta_\tau\}], \tag{42}$$

where $\zeta_\tau = \sqrt{2 \log \left(\frac{1}{\tau^{2a}} \right)}$. Now using Lemma 2, we obtain as $\tau \rightarrow 0$

$$\mathbb{E}_0 T_\tau^2(X_i) 1\{|X_i| \leq \zeta_\tau\} \lesssim (\tau^{2a})^2 \int_0^{\zeta_\tau} x^2 e^{\frac{x^2}{2}} dx (1 + o(1)) \lesssim \zeta_\tau \tau^{2a}, \tag{43}$$

where we use the identity $\frac{d}{dx} [x e^{\frac{x^2}{2}}] = x^2 e^{\frac{x^2}{2}} + e^{\frac{x^2}{2}}$ to bound the term $\int_0^{\zeta_\tau} x^2 e^{\frac{x^2}{2}} dx$. For the second term, we first observe that $|T_\tau(x)| \leq |x|$ for any $x \in \mathbb{R}$. Using this, together with the fact that $x^2 \phi(x) = \phi(x) - \frac{d}{dx} [x \phi(x)]$, we obtain as $\tau \rightarrow 0$

$$\mathbb{E}_0 T_\tau^2(X_i) 1\{|X_i| > \zeta_\tau\} \leq 2 \int_{\zeta_\tau}^\infty x^2 \phi(x) dx \leq 2 \zeta_\tau \phi(\zeta_\tau) + 2 \frac{\phi(\zeta_\tau)}{\zeta_\tau} \lesssim \zeta_\tau \tau^{2a}. \tag{44}$$

Combining equations (42), (43) and (44), it follows that, for $0 < a < 1$, we have as $\tau \rightarrow 0$

$$\sum_{i:\theta_{0i}=0} \mathbb{E}_{\theta_{0i}} (T_\tau(X_i) - \theta_{0i})^2 \lesssim (n - \tilde{q}_n) \tau^{2a} \sqrt{\log \left(\frac{1}{\tau^{2a}} \right)}. \tag{45}$$

Case (II) When $a \geq 1$: Letting $\tilde{\zeta}_\tau = \sqrt{2 \log \left(\frac{1}{\tau} \right)}$, we split the corresponding mean square error as

$$\mathbb{E}_0 [T_\tau^2(X_i)] = \mathbb{E}_0 [T_\tau^2(X_i) 1\{|X_i| \leq \tilde{\zeta}_\tau\}] + \mathbb{E}_0 [T_\tau^2(X_i) 1\{|X_i| > \tilde{\zeta}_\tau\}].$$

Then applying the same reasoning as in *Case (I)* together with Lemma 2, it follows, for $a \geq 1$, we have as $\tau \rightarrow 0$

$$\sum_{i:\theta_{0i}=0} \mathbb{E}_{\theta_{0i}} (T_\tau(X_i) - \theta_{0i})^2 \lesssim (n - \tilde{q}_n) \tau \sqrt{\log \left(\frac{1}{\tau} \right)}. \tag{46}$$

Finally, on combining (38), (41), (45) and (46), we get, as $\tau \rightarrow 0$, the following:

$$\mathbb{E}_{\theta_0} \|T_\tau(\mathbf{X}) - \theta_0\|^2 \lesssim \begin{cases} \tilde{q}_n \log \left(\frac{1}{\tau^{2a}} \right) + (n - \tilde{q}_n) \tau^{2a} \sqrt{\log \left(\frac{1}{\tau^{2a}} \right)} & \text{if } a \in (0, 1) \\ \tilde{q}_n \log \left(\frac{1}{\tau^{2a}} \right) + (n - \tilde{q}_n) \tau \sqrt{\log \left(\frac{1}{\tau} \right)} & \text{if } a \geq 1. \end{cases}$$

The required result now follows immediately by observing that $\tilde{q}_n \leq q_n$ and $q_n = o(n)$ and then taking supremum over all $\theta_0 \in \ell_0[q_n]$. This completes the proof of Theorem 1. □

Proof of Corollary 1. Fix any constant $c > 1$ and choose any $\rho > c$ in the proof of Theorem 1. Then the corresponding multiplicative factor in the upper bound to the worst case ℓ_2 risk of $T_\tau(\mathbf{X})$ over $\ell_0[q_n]$ as obtained in Theorem 1 can at most be $4a\rho^2$. Now taking $\tau = \left(\frac{q_n}{n}\right)^\alpha$ with $\alpha \geq 1$, or $\tau = \frac{q_n}{n} \sqrt{\log(n/q_n)}$ in Theorem 1, it follows that the corresponding mean square error can at most be of the order of $2q_n \log \left(\frac{n}{q_n} \right)$

up to the multiplicative factor $2a\rho^2 \max\{1, \alpha\}$. It is also always bounded below by the minimax ℓ_2 risk which is of the order of $2q_n \log\left(\frac{n}{q_n}\right)$. This immediately leads to (8).

Note that for the horseshoe-type priors we have $a = 0.5$. Therefore, when $\tau = \frac{q_n}{n}$ or $\tau = \frac{q_n}{n} \sqrt{\log\left(\frac{n}{q_n}\right)}$, by using the proof of Theorem 1, it follows that

$$\sup_{\theta_0 \in \ell_0[q_n]} \mathbb{E}_{\theta_0} \|T_\tau(\mathbf{X}) - \theta_0\|^2 \leq 2\rho^2 q_n \log\left(\frac{n}{q_n}\right) (1 + o(1)) \text{ as } n \rightarrow \infty, \tag{47}$$

where the $o(1)$ term depends on $\rho > c > 1$. Now, using (47), and the fact that the minimax ℓ_2 risk in (7) is the greatest lower bound to the mean square error term in (47), we obtain

$$1 \leq \frac{\sup_{\theta_0 \in \ell_0[q_n]} \mathbb{E}_{\theta_0} \|T_\tau(\mathbf{X}) - \theta_0\|^2}{\inf_{\hat{\theta}} \sup_{\theta_0 \in \ell_0[q_n]} \mathbb{E}_{\theta_0} \|\hat{\theta} - \theta_0\|^2} \leq \rho^2 (1 + o(1)) \text{ as } n \rightarrow \infty. \tag{48}$$

Hence taking limit inferior and limit superior in (48) as $n \rightarrow \infty$, we have,

$$1 \leq \liminf_{n \rightarrow \infty} \frac{\sup_{\theta_0 \in \ell_0[q_n]} \mathbb{E}_{\theta_0} \|T_\tau(\mathbf{X}) - \theta_0\|^2}{\inf_{\hat{\theta}} \sup_{\theta_0 \in \ell_0[q_n]} \mathbb{E}_{\theta_0} \|\hat{\theta} - \theta_0\|^2} \leq \limsup_{n \rightarrow \infty} \frac{\sup_{\theta_0 \in \ell_0[q_n]} \mathbb{E}_{\theta_0} \|T_\tau(\mathbf{X}) - \theta_0\|^2}{\inf_{\hat{\theta}} \sup_{\theta_0 \in \ell_0[q_n]} \mathbb{E}_{\theta_0} \|\hat{\theta} - \theta_0\|^2} \leq \rho^2. \tag{49}$$

Note that the Bayes estimator $T_\tau(\mathbf{X})$ does not depend on the choice of $\rho > c > 1$. Hence the ratio in (48) is independent of the choices of $\rho > c > 1$. Consequently, the limit inferior and limit superior terms in (49) are also independent of the choices of $\rho > c > 1$ which are arbitrary. Therefore, taking infimum over all possible choices of $\rho > c > 1$ in (49), we get

$$\sup_{\theta_0 \in \ell_0[q_n]} \mathbb{E}_{\theta_0} \|T_\tau(\mathbf{X}) - \theta_0\|^2 \sim \inf_{\hat{\theta}} \sup_{\theta_0 \in \ell_0[q_n]} \mathbb{E}_{\theta_0} \|\hat{\theta} - \theta_0\|^2. \tag{50}$$

(50) together with (7) completes the proof of (9). □

Proof of Theorem 6. Let us first note that for each i , $X_i \sim N(0, 1)$ under H_{0i} . Therefore, the type I error probability of the i -th induced decision in (23) does not depend on i . Let t_1 denote the common value of t_{1i} 's, $i = 1, \dots, n$. Then, by definition

$$t_1 = \mathbb{P}_{H_{01}}(E(1 - \kappa_1 | X_1, \tau) > 0.5), \tag{51}$$

where $\mathbb{P}_{H_{01}}(B)$ denotes the probability of an event B under the distribution of X_1 under H_{01} . Again, for each i , $X_i \sim N(0, 1 + \psi^2)$ under H_{Ai} . Hence the corresponding type II error probability t_{2i} does not depend on i . Let t_2 denote the common value of t_{2i} 's, $i = 1, \dots, n$. Therefore, by definition

$$t_2 = \mathbb{P}_{H_{A1}}(E(1 - \kappa_1 | X_1, \tau) \leq 0.5), \tag{52}$$

where $\mathbb{P}_{H_{A1}}(B)$ denotes the probability of an event B under the distribution of X_1 under H_{A1} . Thus, using (51) and (52), we get $R_{OG} = n((1-p)t_1 + pt_2)$. Let us now fix any $0 < \eta, \delta < 1$ and choose any $\xi > \frac{2}{\eta(1-\delta)}$. It will be shown that t_1 can be bounded as

$$\frac{G_1(a, \eta, \delta)(\tau^{2a})^{\frac{\xi}{2}}}{\sqrt{\log(\frac{1}{\tau^2})}}(1 + o(1)) \leq t_1 \leq \frac{H_1(a, \eta, \delta)\tau^{2a}}{\sqrt{\log(\frac{1}{\tau^2})}}(1 + o(1)) \text{ as } n \rightarrow \infty, \tag{53}$$

where $G_1(a, \eta, \delta)$ and $H_1(a, \eta, \delta)$ are some finite positive constants each being independent of n , but depend on $a \in (0, 1)$, $\eta \in (0, 1)$ and $\delta \in (0, 1)$. On the other hand, one can show that t_2 can be bounded as

$$[2\Phi(\sqrt{2a\alpha}\sqrt{C}) - 1](1 + o(1)) \leq t_2 \leq [2\Phi(\sqrt{\xi a\alpha}\sqrt{C}) - 1](1 + o(1)) \text{ as } n \rightarrow \infty. \tag{54}$$

For the time being, let us assume that (53) and (54) are true. It will now be shown how they lead to the proof of Theorem 6.

First of all, observe that one can rewrite R_{OG} as $R_{OG} = np(t_2 + \frac{t_1}{p})$. Then using the bounds given in (53) and (54), and the fact that $\lim_{n \rightarrow \infty} \tau/p^\alpha = 0$ for $\alpha \geq 1$, it follows easily that the Bayes risk R_{OG} satisfies

$$\frac{2\Phi(\sqrt{2a\alpha}\sqrt{C}) - 1}{2\Phi(\sqrt{C}) - 1}(1 + o(1)) \leq \frac{R_{OG}}{R_{Opt}^{BO}} \leq \frac{2\Phi(\sqrt{\xi a\alpha}\sqrt{C}) - 1}{2\Phi(\sqrt{C}) - 1}(1 + o(1)) \text{ as } n \rightarrow \infty. \tag{55}$$

Here the $o(1)$ terms depend on η, δ and ξ . Now taking limit inferior and limit superior in (55) as $n \rightarrow \infty$, it follows

$$\frac{2\Phi(\sqrt{2a\alpha}\sqrt{C}) - 1}{2\Phi(\sqrt{C}) - 1} \leq \liminf_{n \rightarrow \infty} \frac{R_{OG}}{R_{Opt}^{BO}} \leq \limsup_{n \rightarrow \infty} \frac{R_{OG}}{R_{Opt}^{BO}} \leq \frac{2\Phi(\sqrt{\xi a\alpha}\sqrt{C}) - 1}{2\Phi(\sqrt{C}) - 1}. \tag{56}$$

Now observe that the multiple testing rules under study do not depend on how $\eta \in (0, 1)$, $\delta \in (0, 1)$ and $\xi > 2/(\eta(1-\delta))$ are chosen. Hence the ratio $\frac{R_{OG}}{R_{Opt}^{BO}}$ is free of any $\eta, \delta \in (0, 1)$ and any $\xi > 2/(\eta(1-\delta))$, for all $n \geq 1$. Thus, the limit inferior and the limit superior terms in (56) are also independent of the choices of η, δ and ξ . But $\xi > 2/(\eta(1-\delta))$ in (56) is arbitrary. Therefore, taking infimum over all such ξ 's in (56) and subsequently over all possible choices of $(\eta, \delta) \in (0, 1) \times (0, 1)$ and repeatedly using the continuity of $\Phi(\cdot)$, we get

$$\lim_{n \rightarrow \infty} \frac{R_{OG}}{R_{Opt}^{BO}} = \frac{2\Phi(\sqrt{2a\alpha}\sqrt{C}) - 1}{2\Phi(\sqrt{C}) - 1}.$$

That $\lim_{n \rightarrow \infty} \frac{R_{OG}}{R_{Opt}^{BO}} = 1$ if $a = 0.5$ and $\alpha = 1$ is now obvious. This completes the proof of the theorem, modulo the proof of (53) and (54) which are now presented below. \square

Proof of (53). First note that, following the same arguments as in the proof of Lemma 3, one can show that for each fixed $x \in \mathbb{R}$ and every fixed $0 < \tau < 1$, there exists a

measurable non-negative real valued function $g(x, \tau)$, depending on η and δ , given by

$$g(x, \tau) = \begin{cases} C_* \left[|x|^2 \int_0^{|x|^2} e^{-u/2} u^{a+1/2-1} du \right]^{-1} + \frac{H(a, \eta, \delta) e^{-\frac{\eta(1-\delta)x^2}{2}}}{\tau^{2a} \Delta(\tau^2, \eta, \delta)} & \text{if } |x| > 0 \\ 1 & \text{if } x = 0. \end{cases} \quad (57)$$

Here $C_* \equiv C_*(a, \eta, L)$ is some finite positive constant that is independent of both x and τ . The function $g(x, \tau)$ defined in (57) is such that $E(\kappa|x, \tau) \leq g(x, \tau)$, for each $x \in \mathbb{R}$ and each $\tau \in (0, 1)$, and satisfies the following:

given any $\xi > \frac{2}{\eta(1-\delta)}$,

$$\lim_{\tau \downarrow 0} \sup_{|x| > \sqrt{\xi \log(\frac{1}{\tau^{2a}})}} g(x, \tau) = 0. \quad (58)$$

Hence, for each fixed $\tau < 1$ we have

$$\{E(1 - \kappa_1|X_1, \tau) > 0.5\} \supseteq \{g(X_1, \tau) < 0.5\} \equiv B_n^c, \text{ say,} \quad (59)$$

where $B_n \equiv \{g(X_1, \tau) \geq 0.5\}$. Let $\zeta_\tau = \sqrt{\xi \log(\frac{1}{\tau^{2a}})}$ and define $C_n \equiv \{|X_1| > \zeta_\tau\}$.

Then using (51) and (59), we have,

$$\begin{aligned} t_1 &\geq \mathbb{P}_{H_{01}}(g(X_1, \tau) < 0.5) \\ &= \mathbb{P}_{H_{01}}(B_n^c) \\ &\geq \mathbb{P}_{H_{01}}(B_n^c \cap C_n) \\ &= [1 - \mathbb{P}_{H_{01}}(B_n|C_n)]\mathbb{P}_{H_{01}}(C_n). \end{aligned} \quad (60)$$

Note that the function $g(x, \tau)$ is measurable, non-negative and continuously decreasing in $|x|$ for $|x| \neq 0$. Hence $\mathbb{E}_{H_{01}}(g(X_1, \tau)||X_1| > \zeta_\tau)$ is well defined and is bounded for all sufficiently small $\tau \in (0, 1)$. Using these facts and applying Markov's inequality, coupled with (58) and the condition that $\tau \rightarrow 0$ as $n \rightarrow \infty$, it follows that,

$$\lim_{n \rightarrow \infty} \mathbb{P}_{H_{01}}(B_n|C_n) = 0. \quad (61)$$

Now, under H_{01} , $X_1 \stackrel{d}{=} Z$, where Z denotes a $N(0, 1)$ random variable having probability density function $\phi(\cdot)$ and cumulative distribution function $\Phi(\cdot)$. Further note that $\zeta_\tau \rightarrow \infty$ as $\tau \rightarrow 0$ and $\lim_{n \rightarrow \infty} \tau = 0$. Using these observations and Mill's ratio, we obtain as $n \rightarrow \infty$

$$\mathbb{P}_{H_{01}}(C_n) = \mathbb{P}(|Z| > \zeta_\tau) \geq 2 \frac{\phi(\zeta_\tau)}{\zeta_\tau} \left(1 - \frac{1}{\zeta_\tau^2}\right) = G_1(a, \eta, \delta) \frac{(\tau^{2a})^{\frac{\xi}{2}}}{\sqrt{\log(\frac{1}{\tau^2})}} (1 + o(1)), \quad (62)$$

for some appropriately chosen finite positive constant $G_1(a, \eta, \delta)$ which is independent of n , but depends on $a \in (0, 1)$, $\eta \in (0, 1)$ and $\delta \in (0, 1)$. Combining (60) – (62), the stated lower bound in (53) follows.

On the other hand, the corresponding upper bound in (53) follows as an immediate consequence of Theorem 6 of Ghosh et al. (2015) with some appropriately chosen finite positive constant $H_1(a, \eta, \delta)$ which does not depend on n , but depends on $a \in (0, 1)$, $\eta \in (0, 1)$ and $\delta \in (0, 1)$. This completes the proof of (53). \square

Proof of (54). Using the fact that $E(\kappa|x, \tau) \leq g(x, \tau)$ for each fixed $x \in \mathbb{R}$ and each fixed $\tau \in (0, 1)$, we obtain

$$\begin{aligned} t_2 &\leq \mathbb{P}_{H_{A1}}(g(X_1, \tau) \geq 0.5) \\ &= \mathbb{P}_{H_{A1}}(B_n) \\ &= \mathbb{P}_{H_{A1}}(B_n|C_n)\mathbb{P}_{H_{A1}}(C_n) + \mathbb{P}_{H_{A1}}(B_n|C_n^c)\mathbb{P}_{H_{A1}}(C_n^c) \\ &\leq \mathbb{P}_{H_{A1}}(B_n|C_n) + \mathbb{P}_{H_{A1}}(C_n^c), \end{aligned} \tag{63}$$

where the events B_n and C_n have already been defined before. Now applying the same argument for proving (61), it can be shown that

$$\lim_{n \rightarrow \infty} \mathbb{P}_{H_{A1}}(B_n|C_n) = 0. \tag{64}$$

Again, as $\lim_{n \rightarrow \infty} \frac{\tau}{p^\alpha} \in (0, \infty)$ for $\alpha > 0$, using Assumption 2, it follows that $\lim_{n \rightarrow \infty} \frac{\log(\frac{1}{\tau^2})}{1+\psi^2} = \alpha C$. Therefore, we have $n \rightarrow \infty$

$$\begin{aligned} \mathbb{P}_{H_{A1}}(C_n^c) &= \mathbb{P}_{H_{A1}}(|X_1| \leq \zeta_\tau) \\ &= \mathbb{P}\left(|Z| \leq \sqrt{\xi a} \sqrt{\frac{\log(\frac{1}{\tau^2})}{1+\psi^2}}\right) \\ &= \mathbb{P}(|Z| \leq \sqrt{\xi a \alpha} \sqrt{C})(1 + o(1)) \\ &= [2\Phi(\sqrt{\xi a \alpha} \sqrt{C}) - 1](1 + o(1)). \end{aligned} \tag{65}$$

Combining (63) – (65), the stated lower bound in (54) follows. On the other hand, the corresponding lower bound in (54) follows immediately following the same line of arguments for proving Theorem 8 of Ghosh et al. (2015). This completes the proof of (54). \square

Proof of Theorem 7. Let us fix any $\eta \in (0, 1)$, $\delta \in (0, 1)$ and any $\xi > \frac{2}{\eta(1-\delta)}$. Now we want to find upper bounds on the probability \tilde{t}_{1i} of type I error and the probability \tilde{t}_{2i} of type II error of the i -th empirical Bayes decision in (25). Using the same techniques employed in proving Theorem 10 and Theorem 11 of Ghosh et al. (2015), together with (53) and (54) used in the proof of Theorem 6, it follows that

$$\tilde{t}_{1i} \leq \frac{B_1^* \alpha_n}{\sqrt{\log(\frac{1}{\alpha_n^2})}}(1 + o(1)) + \frac{1/\sqrt{\pi}}{n^{c_1/2} \sqrt{\log n}} + e^{-2(2 \log 2 - 1)\beta_0 n p(1+o(1))} \text{ as } n \rightarrow \infty, \tag{66}$$

and

$$\tilde{t}_{2i} \leq \left[2\Phi\left(\sqrt{\frac{\xi C}{2}}\right) - 1\right](1 + o(1)) \text{ as } n \rightarrow \infty. \tag{67}$$

In (66) above, B_1^* and β_0 are some finite positive constants, each being independent of both i and n , while $\alpha_n = \mathbb{P}(|X_1| > \sqrt{c_1 \log n}) \sim 2\beta_0 p$ depends on n only. It should further be noted that the $o(1)$ terms as displayed in both (66) and (67) are independent of the index i . Then, using the upper bounds given in (66) and (67), and the arguments used for proving Theorem 2 of Ghosh et al. (2015), one can establish it easily that the Bayes risk R_{OG}^{EB} of the empirical Bayes decision (25), satisfies

$$R_{OG}^{EB} \leq np \left[2\Phi \left(\sqrt{\frac{\xi C}{2}} \right) - 1 \right] (1 + o(1)) \text{ as } n \rightarrow \infty, \quad (68)$$

for every fixed $\xi > \frac{2}{\eta(1-\delta)}$. Rest of the proof follows analogously by first taking the ratio R_{OG}^{EB}/R_{Opt}^{BO} and then employing the same arguments used in the proof of Theorem 6. \square

Supplementary Material

Supplementary Materials to the article “Asymptotic Optimality of One-Group Shrinkage Priors in Sparse High-dimensional Problems” (DOI: [10.1214/16-BA1029SUPP](https://doi.org/10.1214/16-BA1029SUPP); .pdf).

References

- Armagan, A., Dunson, D. B., and Clyde, M. (2011). “Generalized Beta Mixtures of Gaussians.” In Shawe-Taylor, J., Zemel, R. S., Bartlett, P. L., Pereira, F. C. N., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems*, volume 24, 523–531. [1134](#), [1138](#)
- Armagan, A., Dunson, D. B., and Lee, J. (2012). “Generalized Double Pareto Shrinkage.” *Statistica Sinica*, 23(1): 119–143. [MR3076161](#). [1134](#)
- Bhadra, A., Datta, J., Polson, N., and Willard, B. (2015). “The Horseshoe+ Estimator of Ultra-Sparse Signals.” *arXiv:1502.00560v2*. [1143](#)
- Bhattacharya, A., Pati, D., Pillai, N., and Dunson, D. B. (2012). “Bayesian Shrinkage.” *arXiv:1212.6088v1*. [1134](#), [1135](#), [1136](#)
- Bhattacharya, A., Pati, D., Pillai, N., and Dunson, D. B. (2014). “Dirichlet-Laplace priors for optimal shrinkage.” *arXiv:1401.5398v1*. [1134](#), [1135](#)
- Bogdan, M., Chakrabarti, A., Frommlet, F., and Ghosh, J. K. (2011). “Asymptotic Bayes-Optimality under Sparsity of Some Multiple Testing Procedures.” *The Annals of Statistics*, 39(3): 1551–1579. [MR2850212](#). doi: <http://dx.doi.org/10.1214/10-AOS869>. [1137](#), [1145](#), [1147](#)
- Carvalho, C., Polson, N., and Scott, J. (2009). “Handling Sparsity via the Horseshoe.” *Journal of Machine Learning Research W&CP*, 5: 73–80. [1134](#)
- Carvalho, C., Polson, N., and Scott, J. (2010). “The Horseshoe Estimator for Sparse Signals.” *Biometrika*, 97(2): 465–480. [MR2650751](#). doi: <http://dx.doi.org/10.1093/biomet/asq017>. [1134](#), [1145](#), [1146](#)

- Castillo, I. and van der Vaart, A. W. (2012). “Needles and Straw in a Haystack: Posterior Concentration For Possibly Sparse Sequences.” *The Annals of Statistics*, 40(4): 2069–2101. MR3059077. doi: <http://dx.doi.org/10.1214/12-AOS1029>. 1135
- Datta, J. and Ghosh, J. K. (2013). “Asymptotic Properties of Bayes Risk for the Horseshoe Prior.” *Bayesian Analysis*, 8(1): 111–132. MR3036256. doi: <http://dx.doi.org/10.1214/13-BA805>. 1135, 1136, 1146
- Donoho, D. L., Johnstone, I. M., Hoch, J. C., and Stern, A. S. (1992). “Maximum Entropy and the Nearly Black Object (with Discussion).” *Journal of the Royal Statistical Society. Series B (Methodological)*, 54: 41–81. MR1157714. 1139
- Ghosal, S., Ghosh, J. K., and van der Vaart, A. W. (2000). “Convergence Rates of Posterior Distribution.” *The Annals of Statistics*, 28(2): 500–531. MR1790007. doi: <http://dx.doi.org/10.1214/aos/1016218228>. 1141
- Ghosh, P. and Chakrabarti, A. (2015). “Posterior Concentration Properties of a General Class of Shrinkage Estimators around Nearly Black Vectors.” *arXiv:1412.8161v2*. 1136
- Ghosh, P. and Chakrabarti, A. (2016). “Supplementary Materials to the Article “Asymptotic Optimality of One-Group Shrinkage Priors in Sparse High-dimensional Problems”.” *Bayesian Analysis*. doi: <http://dx.doi.org/10.1214/16-BA1029SUPP>. 1137
- Ghosh, P., Tang, X., Ghosh, M., and Chakrabarti, A. (2015). “Asymptotic Properties of Bayes Risk of a General Class of Shrinkage Priors in Multiple Hypothesis Testing Under Sparsity.” *Bayesian Analysis (Advance Publications, DOI:10.1214/15-BA973)*. MR3498045. doi: <http://dx.doi.org/10.1214/15-BA973>. 1135, 1136, 1137, 1138, 1146, 1148, 1149, 1150, 1158, 1159
- Griffin, J. E. and Brown, P. J. (2005). “Alternative prior distributions for variable selection with very many more variables than observations.” Technical report, University of Warwick. 1134
- Hans, C. (2009). “Bayesian Lasso Regression.” *Biometrika*, 96(4): 835–845. MR2564494. doi: <http://dx.doi.org/10.1093/biomet/asp047>. 1134
- Park, T. and Casella, G. (2008). “The Bayesian Lasso.” *Journal of the American Statistical Association*, 103(482): 681–686. MR2524001. doi: <http://dx.doi.org/10.1198/016214508000000337>. 1134
- Polson, N. G. and Scott, J. G. (2011). “Shrink Globally, Act Locally: Sparse Bayesian Regularization and Prediction.” In *Bayesian Statistics 9, Proceedings of the 9th Valencia International Meeting*, 501–538. Oxford University Press. MR3204017. doi: <http://dx.doi.org/10.1093/acprof:oso/9780199694587.003.0017>. 1134, 1137, 1143
- Tipping, M. (2001). “Sparse Bayesian Learning and the Relevance Vector Machine.” *Journal of Machine Learning Research*, 1: 211–244. MR1875838. doi: <http://dx.doi.org/10.1162/15324430152748236>. 1134

- van der Pas, S. L., Kleijn, B. J. K., and van der Vaart, A. W. (2014). “The Horseshoe Estimator: Posterior Concentration Around Nearly Black Vectors.” *Electronic Journal of Statistics*, 8: 2585–2618. MR3285877. doi: <http://dx.doi.org/10.1214/14-EJS962>. 1133, 1135, 1136, 1140, 1142, 1143, 1144, 1148, 1161
- van der Pas, S. L., Salomond, J. B., and Schmidt-Hieber, J. (2016). “Conditions for Posterior Contraction in the Sparse Normal Means Problem.” *Electronic Journal of Statistics*, 10: 976–1000. MR3486423. doi: <http://dx.doi.org/10.1214/16-EJS1130>. 1136, 1140, 1141, 1142, 1143, 1144

Acknowledgments

We would like to thank the Editorial Board of Bayesian Analysis for their thoughtful comments on an earlier version of this work which have led to substantial improvement of the paper in terms of overall presentation and exposition. We would also like to thank Professor Jayanta Kumar Ghosh, Professor of Statistics, Purdue University, and Emeritus Professor, Indian Statistical Institute, for letting us aware about the work of van der Pas et al. (2014) on the posterior contraction properties of the horseshoe prior.