Posterior Contraction rate for one group global-local shrinkage priors in sparse normal means problem

Sayantan Paul\textsuperscript{1} and Arijit Chakrabarti\textsuperscript{1}

\textsuperscript{1}Applied Statistics Unit, Indian Statistical Institute, Kolkata, India
email id:sayantanp-r@isical.ac.in, arc@isical.ac.in

Abstract

We consider a high-dimensional sparse normal means model where the goal is to estimate the mean vector assuming the proportion of non-zero means is unknown. We consider a Bayesian setting and model the mean vector by a one-group global-local shrinkage prior belonging to a broad class of such priors that includes the horseshoe prior. Such a class of priors was earlier considered in Ghosh and Chakrabarti (2017) \cite{ghosh2017}. We address some questions related to asymptotic properties of the resulting posterior distribution of the mean vector for the said class priors. As is well-known, for global-local shrinkage priors, the global shrinkage parameter plays a pivotal role in capturing the sparsity in the model. We consider two ways to model this parameter in this paper. Firstly, we consider this as an unknown fixed parameter and estimate it by an empirical Bayes estimate. In the second approach, we do a hierarchical Bayes treatment by assigning a suitable non-degenerate prior distribution on it. We first show that for the class of priors under study, the posterior distribution of the mean vector contracts around the true parameter at a near minimax rate when the empirical Bayes approach is used. Next we prove that in the hierarchical Bayes approach, the corresponding Bayes estimate attains a near minimax rate asymptotically under the squared error loss function. We also show that the posterior contracts around the true parameter at a near minimax rate. These results generalize those of van der Pas et al. (2014) \cite{pas2014}, (2017) \cite{pas2017}, proved for the horseshoe prior, for a broad class of priors containing the horseshoe prior. We have also studied in this work the asymptotic optimality of the horseshoe+ prior in this context. We prove that using the empirical Bayes estimate of the global parameter, the corresponding Bayes estimate attains a near
minimax rate asymptotically under the squared error loss function and also show that the posterior distribution contracts around the true parameter at a near minimax rate. While most of our proofs require invoking several new technical arguments and discovering interesting sharper bounds on quantities of interest, few of our results are obtained by building on ideas for proving existing results.

1 Introduction

Over the last few decades, analysis and inference on datasets with large number of variables have become important problems for the statisticians. Such datasets regularly arise in various scientific fields such as genomics, medicine, finance, astronomy and economics, to name a few. Naturally, this has become an area of very active research. We are interested in the high-dimensional asymptotic setting when the number of parameters grows at least at the same rate as the number of observations. In such problems, often it is the case that only few of the total number of parameters is really important. For example in high-dimensional regression problems, it is quite common that the proportion of relevant regressors is very small compared to the total number of covariates. This is the phenomenon of sparsity. We study in this paper a very important high-dimensional problem, namely the well-known normal means problem. Precisely we have an observation vector $X \in \mathbb{R}^n$, $X = (X_1, X_2, \ldots, X_n)$, where each observation $X_i$ is expressed as,

$$X_i = \theta_i + \epsilon_i, \ i = 1, 2, \cdots, n, \quad (1.1)$$

where the unknown mean parameters $\theta_1, \theta_2, \cdots, \theta_n$ denote the effects under investigation and $\epsilon_i$’s are independent $N(0, 1)$ random variables. In this scenario, our goal is to estimate the corresponding unknown mean vector $\theta$. The true mean vector $\theta_0 = (\theta_{01}, \theta_{02}, \cdots, \theta_{0n})$ is assumed to be sparse, in a nearly black sense in the terminology of Donoho et al. (1992) [9], that is, $\theta_0 \in l_0[p_n]$ where $l_0[p_n] = \{\theta \in \mathbb{R}^n : \#(1 \leq i \leq n : \theta_i \neq 0) \leq p_n\}$ with $p_n = o(n)$ as $n \to \infty$ which means the number of non-zero means grows at much slower rate than $n$. This model has been widely studied in the literature, see Johnstone and Silverman (2004) [15], Efron (2004) [10], Jiang and Zhang (2009) [14] in this context. As mentioned by Johnstone and Silverman (2004) [15], this model also bears potential application in several contexts containing astronomical and other signal and image processing, model selection and data mining.

In the Bayesian setting, a natural way to model $\theta$ satisfying (1.1) is to use a two-groups prior like a
spike and slab prior due to Mitchell and Beauchamp (1988) [16], which is a mixture of Dirac measure at 0 and a heavy-tailed absolutely continuous distribution $F$ over $\mathbb{R}$, given as

$$\theta_i \overset{iid}{\sim} (1 - \nu)\delta_0 + \nu F, i = 1, 2, \cdots, n,$$

where $\nu$ is probability that true value of $\theta_i$ would be non-zero. Although it is natural to use the two-groups prior in sparse settings, analyzing this model can be computationally prohibitive, especially in high-dimensional problems and complex parametric frameworks. To overcome this, an alternative Bayesian approach has become very popular in recent times. Here one uses instead a continuous one-group prior for modelling parameters satisfying sparsity constraints (e.g. (1.1)) and such priors entail much simpler computational effort for posterior analyses than their two-group counterparts. Some examples of one-groups priors to model sparsity are the $t$-prior due to Tipping (2001) [20], the Laplace prior of Park and Casella (2008) [17], the normal-exponential-gamma prior due to Griffin and Brown (2010) [4], the horseshoe prior of Carvalho et al. (2009 [6], 2010 [7]) etc. All of these above mentioned priors can be expressed as a “global-local” scale mixture of normals as

$$\theta_i | \lambda_i, \tau \overset{iid}{\sim} \mathcal{N}(0, \lambda_i^2 \tau^2), \lambda_i^2 \overset{iid}{\sim} \pi_1(\lambda_i^2), \tau \sim \pi_2(\tau).$$

(1.2)

where $\lambda_i$ is the local shrinkage parameter used for detecting the signals and $\tau$ is the global shrinkage parameter used to control the overall sparsity in the model. As recommended by Polson and Scott (2010) [18], in sparse one-group model, the priors on $\lambda^2_i$ and $\tau$ should fulfill the following two criteria:- (i) $\pi_1(\lambda^2_i)$ should have thick tails. (ii)$\pi_2(\tau)$ should have substantial mass near zero. The horseshoe prior is one example of such a prior. These criteria ensure that the prior on $\theta_i$ assigns significant probability near zero but also has a heavy tail so that it can accommodate large signals. This makes the noise observations getting shrunk towards zero while the large signals are left almost unshrunk. This property is known as tail robustness which was proved by Polson and Scott (2010) [18] in Theorem 1 of their paper.

van der Pas et al. (2014) [21] used the horseshoe prior for estimating $\theta$ by choosing the global shrinkage parameter $\tau$ in terms of the proportion of non-zero means assuming it to be known. They proved that under the squared error loss, choosing $\tau = \left(\frac{\kappa}{n}\right)^\alpha$, for $\alpha \geq 1$, the Bayes estimate corresponding to the horseshoe prior asymptotically attains the minimax risk up to a multiplicative constant and also showed that the corresponding posterior distribution contracts around the true mean vector at the minimax rate. Ghosh and Chakrabarti
(2017) \cite{11} proved a similar result for a class of priors given by

$$\pi_1(\lambda_i^2) = K(\lambda_i^2)^{-a-1}L(\lambda_i^2), \tag{1.3}$$

where $K \in (0, \infty)$ is the constant of proportionality, $a$ is a positive real number and $L : (0, \infty) \to (0, \infty)$ is measurable non-constant slowly varying function satisfying assumption $(A1)$ and $(A2)$ of section 2. Recall that $L(\cdot)$ is said to be slowly varying, if for any $\alpha > 0$, $\frac{L(\alpha x)}{L(x)} \to 1$ as $x \to \infty$. Since, the class of priors (mentioned in (1.3)) contains three parameter beta normal (which contains horseshoe), generalized double Pareto, half t priors and all of them satisfy the above mentioned assumptions on $L(\cdot)$, their results extend those of van der Pas et al. (2014) \cite{21}.

Using the asymptotic framework of Bogdan et al. (2011) \cite{3}, Datta and Ghosh (2013) \cite{8} proved that the Bayes risk for the horseshoe estimator attains the Bayes oracle risk if the global shrinkage parameter is of the same order as the proportion of non-zero means present in the model. It is also evident from the results of van der Pas et al. (2014) \cite{21} and Ghosh and Chakrabarti (2017) \cite{11} that the choice of global shrinkage parameter $\tau$ (based on the proportion of non-zero means) plays a very important role in proving asymptotic minimaxity. In practice, generally, the proportion of non-zero means may not be known. To address this issue, van der Pas et al. (2014) \cite{21} used horseshoe prior and proposed to use a ‘plug-in’ estimator of $\tau$ which is learnt from the data and use it in the formula for the posterior mean. They showed in case of horseshoe prior, the posterior mean corresponding to the empirical Bayes estimate of $\tau$ attains “near minimax” rate for squared error loss and also proved that the posterior distribution of the mean vector contracts around the truth at a near minimax rate. Later Ghosh and Chakrabarti (2017) \cite{11} extended the result of the posterior mean corresponding to the empirical Bayes estimate for the class of priors mentioned above but did not address the issue regarding the contraction rate of the posterior distribution. In this scenario, a natural question is whether the results obtained by van der Pas et al. (2014) \cite{21} regarding the posterior contraction rate based on the empirical Bayes estimate of the horseshoe prior can be extended for the broader class of priors mentioned by Ghosh and Chakrabarti (2017) \cite{11}. To the best of our knowledge, no such result still exists in the literature. Towards this, we observed that the upper bound corresponding to the mean square error of the empirical Bayes estimate of the horseshoe prior matches that of general class of global-local shrinkage priors mentioned in \cite{13}. It was our hunch that results similar to Ghosh and Chakrabarti (2017) \cite{11} for optimal posterior contraction rate (assuming the proportion of non-zero means as known) can be established using the empirical Bayes estimate.
of $\tau$. Theorem 2 of this paper answers our hunch in the affirmative. Hence, our result is indeed a generalization to that of van der Pas et al. (2014) [21]. It is important to mention that, Theorem 1 of our paper can not be obtained directly by using the empirical Bayes estimate of $\tau$ in place of $\tau$ when used as tuning parameter in the upper bound of the posterior variance of the mean vector given in Theorem 3 of Ghosh and Chakrabarti (2017) [11]. To obtain the upper bound, first we have to divide our calculations corresponding to different range of $\tau$ and then have to use the monotonicity of $\tau$ in the definition of the posterior mean of the shrinkage coefficient. Now, for $a \in [\frac{1}{2}, 1)$ after following the above steps, we can use some arguments similar to that of Ghosh and Chakrabarti (2017) [11] to obtain the desired result. But this is not enough for the case $a \geq 1$, as there were some soft spots in the algebra though the final result was true. We were required to obtain sharper upper bound on the posterior mean of the shrinkage coefficient for $a \geq 1$ (given in Lemma 1) using several new techniques. This has been one of the key statements used in proving all theorems regarding the posterior contraction rate of the broad class of priors in this paper. Not only this, instead of using an upper bound on the quantity $J(x, \tau)$ (defined in the proof of Theorem 1), we have dealt it differently by using several new arguments. A broad discussion can be found in Case-2 in the proof of Theorem 1.

An alternative way regarding the use of $\tau$ when it is assumed to be unknown is to use a non-degenerate prior on it. Now it becomes an interesting question whether the same optimal posterior contraction results hold for the general class of priors or at least for the horseshoe prior when a non-degenerate prior distribution is used on $\tau$. It also becomes necessary to answer to the question that what conditions are required to be used on the prior density of $\tau$ to obtain the desirable results. It was van der Pas et al. (2017) [23] who first came up with two conditions on the prior density of $\tau$ (later discussed in section 3.2), hereby referred as $\pi_n(\tau)$ for the horseshoe prior so that the posterior distribution of $\theta$ can contract near the true value at a near minimax rate. One of our goals in this paper was to extend the results of van der Pas et al. (2017) [23] for the general class of priors mentioned in (1.3). We have obtained the optimal posterior contraction result for the broad class of priors using full Bayes estimate of $\tau$, which is given in Theorems 6-8. In the case of full Bayes approach, the novelty of our work lies in using the corresponding posterior mean as a weighted average of the posterior mean of $\theta$ when $\tau$ is used as a tuning parameter with the weight as the posterior distribution of $\tau$ given data. For obtaining the optimal posterior contraction rate, we have to establish upper bounds on the posterior mean of the shrinkage coefficient and on the variance of the posterior distribution of the corresponding mean vector. As already mentioned, for obtaining the upper bound on the posterior mean of the shrinkage coefficient, we have used Lemma 1 after using above mentioned relationship. In full Bayes method, too, after using this relationship,
arguments similar to that of Ghosh and Chakrabarti (2017) [11] still will be useful for deriving the upper bound on the variance of the posterior distribution of the corresponding mean vector when \( a \in \left[ \frac{1}{2}, 1 \right] \). For \( a \geq 1 \), like the empirical Bayes procedure, in this case too, we have to obtain an upper bound of the posterior mean of the shrinkage coefficient derived in this paper (Lemma 1) which is sharper than that of Ghosh and Chakrabarti (2017) [11]. This has been one of the key ingredients used in proving the optimal posterior contraction rate for the hierarchical Bayes procedure, too. Here also we have to use some new arguments similar to that of the empirical Bayes version for obtaining the desired results. Our results are proved under the conditions similar to that of van der Pas et al. (2017) [23] which shows that some of the techniques mentioned there can be used for a broad class of priors.

Till now we have studied posterior contraction properties of one group global-local shrinkage priors of the form (1.2) where the local shrinkage parameter is modeled using equation (1.3). Next, we move on to another global-local shrinkage prior named horseshoe+ due to Bhadra et al. (2017) [1]. As mentioned by Bhadra et al. (2017) [1], the hierarchical model for horseshoe+ is given as,

\[
\begin{align*}
\theta_i | \lambda_i & \sim \mathcal{N}(0, \lambda_i^2), \\
\lambda_i | \eta_i, \tau & \sim \mathcal{C}^+(0, \tau \eta_i), \\
\eta_i & \sim \mathcal{C}^+(0, 1).
\end{align*}
\]

(1.4)

The result of Datta and Ghosh (2013) [8] regarding the Bayes risk of the horseshoe prior was the intuition of Bhadra et al. (2017) [1] to extend the concept of horseshoe prior in such a way that the resultant new sparse signal recovery prior not only attains the oracle risk up to a multiplicative constant, but can also improve upon the error rates in theory as well as in practice. Though horseshoe+ prior is a natural extension of horseshoe prior, due to the presence of the another level of local shrinkage parameter \( \eta_i \), after integrating out \( \eta_i \), the hierarchical form can not be expressed in the form of (1.2) and (1.3) without assuming \( \tau = 1 \). Even after assuming \( \tau = 1 \), (1.4) while expressed in the form of (1.3), the slowly varying function becomes unbounded in this case, which is contrary to the assumption of Ghosh and Chakrabarti (2017) [11] (see assumption (A1) and (A2)). Due to these two reasons, we have studied the posterior contraction property of the horseshoe+ prior in a separate section. Our result proves that when the proportion of non-zero means is unknown, using the empirical Bayes estimate of the global parameter, not only the upper bound of the mean square error of \( \theta \) for horseshoe+ prior attains a near minimax rate, but also the posterior distribution contracts at the truth at a
Briefly, our contributions in this paper are as follows. Firstly, in case of the empirical Bayes procedure, we have shown that the results of posterior contraction rate due to van der Pas et al. (2014) can be extended for the general class of priors mentioned in (1.3) assuming some conditions on the slowly varying function $L$ similar to that of Ghosh and Chakrabarti (2017). Secondly, using similar conditions on the prior distribution on $\tau$ as mentioned by van der Pas et al. (2017), the results on the posterior concentration based on the hierarchical Bayes estimate of the global shrinkage parameter derived in this paper is indeed an extension to that of van der Pas et al. (2017). For both empirical Bayes and full Bayes approach, techniques used by Ghosh and Chakrabarti (2017) are helpful for extending the results of van der Pas et al. (2014), (2017), when $a \in [\frac{1}{2},1)$, but for $a \geq 1$, we have to find sharper upper bound on the posterior mean of the shrinkage coefficient compared to that obtained by Ghosh and Chakrabarti (2017) using some new arguments (given in Lemma 1) for proving the statements of Theorems 1, 3 and 4. Thirdly, for horseshoe+ prior, using some results of Bhadra et al. (2017) we have obtained upper bounds on the posterior mean of the shrinkage coefficient and on the posterior variance of the mean vector, given in Lemma 2, which is the key to obtain the Theorems 6, 7 and 8.

The rest of the paper is organized as follows. In section 2 we briefly describe the existing results on the posterior concentration rates assuming the proportion of non-zero means is known. In section 3.1 when the level of sparsity is unknown, using the empirical Bayes estimate of $\tau$, we prove that the upper bound of the total posterior variance using the empirical Bayes estimate of $\tau$ corresponding to this general class of priors in (1.3) is at a near minimax rate, this result coined with Theorem 2 of Ghosh and Chakrabarti (2017) ensures the first contribution in this paper. In section 3.2 we have generalized the result of posterior contraction of the full Bayes posterior distribution of $\theta$ for the above mentioned class of priors using the conditions given by van der Pas et al. (2017). In section 4 using the empirical Bayes estimate of $\tau$, we show that the posterior distribution of horseshoe+ prior contracts at a near minimax rate, up to some multiplicative constant. Section 5 contains the overall discussion along with some possible extensions of our work. All the proofs of those above mentioned theorems are given in section 6.

1.1 Notation

For any two sequences $\{a_n\}$ and $\{b_n\}$ with $\{b_n\} \neq 0$ for all large $n$, we write $a_n \asymp b_n$ to denote $0 < \liminf_{n \to \infty} \frac{a_n}{b_n} \leq \limsup_{n \to \infty} \frac{a_n}{b_n} < \infty$ and $a_n \lesssim b_n$ to denote for sufficiently large $n$, there exists a constant
$c > 0$ such that $a_n \leq cb_n$. Finally, $a_n = o(b_n)$ denotes $\lim_{n \to \infty} \frac{a_n}{b_n} = 0$.

$\phi(\cdot)$ denotes the density of a standard normal distribution. In full Bayes procedure, we use the notation $\tau_n(p_n)$ which denotes $(p_n/n)\sqrt{\log(n/(p_n))}$.

2 Posterior Contraction rate of global-local shrinkage prior with known level of sparsity

Before describing the main contributions, let us first give a brief outline of the existing work on the posterior concentration property of the general class of priors in (1.3). For the theoretical development of the paper, Ghosh and Chakrabarti (2017) [11] assumed that the slowly varying function $L(\cdot)$ defined in (1.3) satisfies the following two assumptions:

(A1) $\lim_{t \to \infty} L(t) \in (0, \infty)$, that is, there exists some $c_0(>0)$ such that $L(t) \geq c_0 \forall t \geq t_0$, for some $t_0 > 0$, which depends on both $L$ and $c_0$.

(A2) There exists some $M \in (0, \infty)$ such that $\sup_{t \in (0, \infty)} L(t) \leq M$.

Ghosh et al. (2016) [12] established that many popular one-group shrinkage priors such as the three parameter beta normal mixtures, the generalized double Pareto priors, half-t prior and inverse gamma prior can be expressed in the above general form (1.3) which satisfies the two assumptions [A1] and [A2].

Using the normal means model of the form (1.1) along with the hierarchical form of prior mentioned in (1.2), for global-local scale mixtures of normal,

$$\theta_i | X_i, \kappa_i, \tau \sim \text{ind}\ N((1 - \kappa_i)X_i, (1 - \kappa_i))$$

where $\kappa_i = \frac{1}{1 + \lambda_i^2 \tau}$, $i = 1, 2, \cdots, n$ is the shrinkage coefficient due to the $i^{th}$ parameter $\theta_i$. Hence, $E(\theta_i | X_i, \tau) = E(1 - \kappa_i | X_i, \tau)X_i$, which is denoted as $T_\tau(X_i)$. Thus the resulting vector of posterior means $(E(\theta_1 | X_1, \tau), E(\theta_2 | X_2, \tau), \cdots, E(\theta_n | X_n, \tau))$ is denoted as $T_\tau(X)$.

Using the knowledge of known level of sparsity, van der Pas et al. (2014) [21] proved that with appropriate choice of $\tau$, the mean square error of horseshoe estimator attains the minimax $l_2$ risk, upto a multiplicative constant. Their results also reveal that the corresponding posterior distribution contracts around both the true mean vector and the Bayes estimate at least as fast as the minimax rate. Later Ghosh and Chakrabarti...
(2017) [11] considered a general class of global-local priors of the form [120] with the slowly varying function $L(\cdot)$ satisfying two assumptions [A1] and [A2] and extended the result of posterior concentration rate for the broad class of prior. For detailed discussion, please see Ghosh and Chakrabarti (2017) [11].

3 Posterior Contraction rate of general class of prior when the level of sparsity is unknown

This section deals with the posterior contraction rates of both the empirical Bayes and full Bayes posterior distributions of the mean vector, $\theta$, which are one of the main findings of our article. The empirical Bayes posterior distribution of $\theta$ is obtained by replacing the global shrinkage parameter $\tau$ with its empirical Bayes version, a data-dependent estimate of $\tau$. On the other hand, using a non-degenerate prior on $\tau$ along with the hierarchical form mentioned above and finally integrating out all nuisance parameters gives the full Bayes posterior distribution of $\theta$.

Since the minimax rate has proven to be a useful benchmark for the speed of contraction of posterior distributions and according to Theorem 2.5 of Ghosal et al. (2000) [13], posterior distributions cannot contract around the truth faster than the minimax risk, it has been always a hunch for the researchers to find out the optimal contraction rate of the posterior distribution of the parameter of interest. In case of adaptive procedures, where the maximum number of non-zero means $p_n$ is unknown, the results available in the literature have been proved in the terms of “near minimax rate” $p_n \log n$, for example for the spike-and-slab Lasso [19], the Lasso [2], and the horseshoe due to van der Pas et al. (2014) [21]. In this section, to distinguish between two approaches on the use of $\tau$ when the proportion of non-zero means is unknown, we have divided the whole discussion in two sub-sections, one for the empirical Bayes and the other for full Bayes version. In both of these procedures, results on the posterior contraction rate attain a near minimax rate, which justifies that the results of van der Pas et al. (2014) [21], (2017) [23] hold for a broad class of priors including the horseshoe prior as a special case.

3.1 Empirical Bayes Procedure

As mentioned by van der Pas et al. (2014) [21] and Ghosh and Chakrabarti (2017) [11], proportion of non-zero means may not be known in general. In this situation, van der Pas et al. (2014) [21] proposed to use a
data-dependent estimate of $\tau$, known as the empirical Bayes estimate of $\tau$, which is given by

$$\hat{\tau} = \max\left\{ \frac{1}{n}, \frac{1}{c_2 n} \sum_{i=1}^{n} 1\{|X_i| > \sqrt{c_1 \log n}\} \right\},$$

(3.1)

where $c_1 \geq 2$ and $c_2 \geq 1$ are two positive constants. Under the assumption that at least one and at most all the parameters are non zero, $\hat{\tau}$ as defined above will always lie between $\frac{1}{n}$ to 1. Let $T_{\hat{\tau}}(X)$ be the Bayes estimate of $T_{\tau}(X)$ evaluated at $\tau = \hat{\tau}$. Initially, van der Pas et al. (2014) shown that the empirical Bayes estimate $T_{\hat{\tau}}(X)$ of horseshoe prior (i.e. $a = 0.5$ and $L(t) = t/(t+1)$ in (1.3)) attains the near minimax $l_2$ risk up to some multiplicative constants, which was further generalized by Ghosh and Chakrabarti (2017) as they obtained the same contraction rate when $a \geq 0.5$ provided $p_n \propto n^\beta$, $0 < \beta < 1$. In this section, we study the optimal property of global-local shrinkage priors which can be expressed in the form of (1.2) and the local shrinkage parameter can be written using (1.3), horseshoe+ prior expressed using the hierarchical form (1.4) will be discussed later.

This theorem provides an upper bound to the total posterior variance corresponding to the general class of priors under study using the empirical Bayes estimate of $\tau$ when $a \geq \frac{1}{2}$. The proof of the theorem is given in section 6.

**Theorem 1.** Suppose $X \sim N_n(\theta_0, I_n)$ where $\theta_0 \in l_0[p_n]$ with $p_n \propto n^\beta$ where $0 < \beta < 1$. Consider the class of priors (1.3) with $a \geq \frac{1}{2}$ where $L(\cdot)$ satisfies Assumptions (A1) and (A2). Then the total posterior variance using the empirical Bayes estimate of $\tau$ corresponding to this general class of priors satisfies, as $n \to \infty$,

$$\sup_{\theta_0 \in l_0[p_n]} \mathbb{E}_{\theta_0} \sum_{i=1}^{n} \text{Var}(\theta_i | X_i, \hat{\tau}) \lesssim p_n \log n$$

where $\text{Var}(\theta_i | X_i, \hat{\tau})$ denotes $\text{Var}(\theta_i | X_i, \tau)$ evaluated at $\tau = \hat{\tau}$.

The next theorem provides upper bounds on the rates of posterior contraction for the chosen class of priors with $a \geq 0.5$, both around the true parameter $\theta_0$ and the corresponding Bayes estimate $T_{\hat{\tau}}(X)$.

**Theorem 2.** Suppose the assumptions of Theorem 1 hold, then empirical Bayes posterior contracts around both the true parameter $\theta_0$ and the corresponding Bayes estimate $T_{\hat{\tau}}(X)$ at the following rate,

$$\sup_{\theta_0 \in l_0[p_n]} \mathbb{E}_{\theta_0} \Pi_{\hat{\tau}} \left( \theta : ||\theta - \theta_0||^2 > M_n p_n \log n | X \right) \to 0,$$
and
\[
\sup_{\theta_0 \in \ell^0(p_n)} \mathbb{E}_{\theta_0} \Pi_{\hat{\tau}}(\theta : ||\theta - T_{\hat{\tau}}(X)||^2 > M_n p_n \log n | X) \to 0.
\]

for any \(M_n \to \infty\), \(p_n \to \infty\) with \(p_n = o(n)\) as \(n \to \infty\).

Proof. Using Markov’s inequality along with Theorem 1 and Theorem 2 of Ghosh and Chakrabarti (2017) [11] provides the first part, while use of Markov’s inequality together with Theorem 1 is sufficient for proving the 2nd part. ■

The first part of Theorem 2 states that the posterior distribution of \(\theta\) contracts around the true \(\theta_0\) at least as fast as the minimax \(l_2\) risk. On the other hand, the Theorem 2.5 of Ghosal et al. (2000) [13] reveals that the posterior distribution cannot contract around the truth faster than the minimax risk. As a combination of these two, the conclusion can be drawn is that the rate at which the posterior distribution of the parameter vector \(\theta\) contracts around \(\theta_0\) is the near minimax rate. However, it is not necessarily true for the corresponding Bayes estimates as the posterior spread may be of smaller order than the rate at which the full Bayes estimator approaches the underlying mean vector. A similar result was also proved by van der Pas et al. (2017) [23] for the horseshoe prior. Our results show that the existing result regarding the near minimax rate of the full Bayes posterior for the Horseshoe prior can be extended for the general class of priors under study.

van der Pas et al. (2014) [21] provided near-minimax rate for horseshoe prior when the empirical Bayes estimate of \(\tau\) is used. These two results can be looked upon as a generalization of the existing result as it considers a broad class of priors.

3.2 Full Bayes Approach

Motivated by van der Pas et al. (2017) [23], the following two conditions are required for contraction of the full Bayes posterior distribution at a near minimax rate for the class of priors of the form (1.2) and (1.3) with the two assumptions \([A1]\) and \([A2]\) on the slowly varying function.

(C1) The prior density \(\pi_n\) is supported within \([\frac{1}{n}, 1]\).

(C2) Let \(t_n = C_u \pi^2 \tau_n(p_n)\) where \(C_u\) is a finite positive constant defined in Lemma 3.7(i) in van der Pas et al. (2017) [23]. Then \(\pi_n\) satisfies
\[
(p_n/n)^{M_1} \int_{\frac{1}{n}}^{t_n} \pi_n(\tau)d\tau \geq e^{-c p_n}, \text{ for some } c \leq C_u/2, M_1 \geq 2.
\]
The condition (C1) is exactly same as that of van der Pas et al. (2017) [23], whereas the Condition (C2) is a slightly modified version of the condition given by van der Pas et al. (2017) [23], which gives the idea at which rate the tail of the posterior distribution of \( \tau \) should go to zero as the dimension increases.

The motivation behind the Condition (C1) i.e. the restriction of the support on \( \pi_n \) within the interval \([\frac{1}{n}, 1]\) is similar to that of the empirical Bayes estimate of \( \tau \) defined in (3.1). Additionally, this truncation also solves the computational issues when \( \tau \) is very small. This relation also helps to understand that upper bound of full Bayes estimate of \( \theta \) can attain the near minimax risk asymptotically.

Condition (C2) implies that there should have sufficient prior mass around the optimal values of \( \tau \) so that the full Bayes posterior distribution can contract at a near minimax rate. This condition is satisfied by many prior densities, except the very sparse case when \( p_n \gtrsim \log n \). As mentioned by van der Pas et al. (2017) [23], the half Cauchy distribution supported on \([\frac{1}{n}, 1]\) having density \( \pi_n(\tau) = (\arctan(1) - \arctan(1/n))^{-1}(1 + \tau^2)^{-\frac{1}{2}}1_{\tau \in [\frac{1}{n}, 1]} \) satisfies condition (C2) provided \( p_n \gtrsim \log n \). Similarly, uniform prior on \([\frac{1}{n}, 1]\) with density \( \pi_n(\tau) = n/(n^2 - 1)1_{\tau \in [\frac{1}{n}, 1]} \) satisfies condition (C2) under the same assumption. Though we have slightly modified the second condition of van der Pas et al. (2017) [23], but it is important to observe that all the above-mentioned priors which satisfies the condition of van der Pas et al. (2017) [23], also obeys our condition (C2). The main reason behind this is that both the conditions are asymptotically equivalent.

The next theorem gives an upper bound to the mean square error for the hierarchical Bayes posterior estimate of \( \theta, \hat{\theta} \). The proof of the theorem is given in section 6.

**Theorem 3.** Suppose \( X \sim N_n(\theta_0, I_n) \) where \( \theta_0 \in l_0[p_n] \) with \( p_n \propto n^\beta \) where \( 0 < \beta < 1 \). Consider the class of priors (1.3) with \( a \geq \frac{1}{2} \) where \( L(\cdot) \) satisfies Assumptions (A1) and (A2). If the prior on \( \tau \) satisfies Conditions (C1) and (C2), then the hierarchical Bayes posterior estimate of \( \theta \) satisfies the following

\[
\sup_{\theta_0 \in l_0[p_n]} E_{\theta_0}||\hat{\theta} - \theta_0||^2 \asymp 2pn \log n
\]

as \( p_n \to \infty \) and \( p_n = o(n) \) as \( n \to \infty \).

Assuming \( p_n \propto n^\beta \), \( 0 < \beta < 1 \), the corresponding minimax error rate under the squared \( l_2 \) norm is of the order of \( 2p_n \log n \), which signifies that like the data-dependent estimate \( T_\tau(X) \), the full Bayes estimate \( \hat{\theta} \) also attains a near minimax \( l_2 \) risk up to some multiplicative constants, under the condition \( p_n \propto n^\beta \), \( 0 < \beta < 1 \). This also proves that in order to obtain the asymptotic minimaxity of the full Bayes estimate, a sharp peak near zero like horseshoe prior is not always needed.
The next theorem provides an upper bound to the total posterior variance corresponding to the general class of priors under study using the full Bayes estimate of \( \tau \) when \( a \geq \frac{1}{2} \). The proof of the theorem is given in section 6.

**Theorem 4.** Suppose \( X \sim N_n(\theta_0, I_n) \) where \( \theta_0 \in l_0[p_n] \) with \( p_n \propto n^\beta \) where \( 0 < \beta < 1 \). Consider the class of priors (1.3) with \( a \geq \frac{1}{2} \) where \( L(\cdot) \) satisfies Assumptions \( (A1) \) and \( (A2) \). If the prior on \( \tau \) satisfies Conditions \( (C1) \) and \( (C2) \), then the total posterior variance using the full Bayes estimate of \( \tau \) corresponding to this general class of priors satisfies

\[
\sup_{\theta_0 \in l_0[p_n]} \mathbb{E}_{\theta_0} \sum_{i=1}^{n} \text{Var}(\theta_i|X) \lesssim p_n \log n
\]

as \( p_n \to \infty \) and \( p_n = o(n) \) as \( n \to \infty \).

The next theorem provides upper bounds on the rates of posterior contraction for the chosen class of priors with \( a \geq \frac{1}{2} \), both around the true parameter \( \theta_0 \) and the corresponding Bayes estimate \( \hat{\theta} \).

**Theorem 5.** Suppose the assumptions of Theorem 4 hold, then full Bayes posterior contracts around both the true parameter \( \theta_0 \) and the corresponding Bayes estimate \( \hat{\theta} \) at the following rate,

\[
\sup_{\theta_0 \in l_0[p_n]} \mathbb{E}_{\theta_0} \Pi \left( \theta : ||\theta - \theta_0||^2 > M_n p_n \log n | X \right) \to 0,
\]

and

\[
\sup_{\theta_0 \in l_0[p_n]} \mathbb{E}_{\theta_0} \Pi \left( \theta : ||\theta - \hat{\theta}||^2 > M_n p_n \log n | X \right) \to 0,
\]

for any \( M_n \to \infty \), \( p_n \to \infty \) as \( n \to \infty \).

**Proof.** Using Markov’s inequality along with Theorem 3 and 4 provides the first part, while use of Markov’s inequality together with Theorem 4 is sufficient for proving the 2nd part.

The first part of Theorem 4 states that the posterior distribution of \( \theta \) contracts around the true \( \theta_0 \) at least as fast as the minimax \( l_2 \) risk. On the other hand, the Theorem 2.5 of Ghosal et al. (2000) \[13\] reveals that the posterior distribution cannot contract around the truth faster than the minimax risk. As a combination of these two, the conclusion can be drawn is that the rate at which the posterior distribution of the parameter vector \( \theta \) contracts around \( \theta_0 \) is the near minimax rate. However, it is not necessarily true for the corresponding Bayes estimates as the posterior spread may be of smaller order than the rate at which the full Bayes estimator approaches the underlying mean vector. Similar result was also proved by van der Pas et al. (2017) \[23\] for the
horseshoe prior. Our results show that the existing result regarding the near minimax rate of the full Bayes posterior for the Horseshoe prior can be extended for the general class of prior under study.

4 Posterior Contraction rate of horseshoe+ prior with unknown level of sparsity

In this section, we confine our attention to horseshoe+ prior under the condition that induced level of sparsity is unknown. As mentioned by Bhadra et al. (2017) [1], they introduced a half-Cauchy mixing variable \( \eta_i \) in the hierarchical form (1.4) and as a result of this, the local shrinkage parameter \( \lambda_i \) are not marginally independent after conditioning the global parameter \( \tau \), the same can be visualized after integrating out \( \eta_i \). The prior distribution of \( \lambda_i \) given \( \tau \) is of the form,

\[
\pi(\lambda_i|\tau) = \frac{4}{\pi^2 \tau} \log(\lambda_i/\tau) (\lambda_i/\tau)^2 - 1
\]

Due to the presence of \( \tau \) in (4.1), it is obvious that it can not be expressed in the form (1.3) without the assumption \( \tau = 1 \). Even under the assumption of \( \tau = 1 \), when the prior on \( \lambda_i \) is expressed in the form of (1.3), the slowly varying function \( L(\cdot) \) does not satisfy the assumptions (A1) and (A2), hence we have studied the posterior contraction property of this prior separately. van der Pas et al. (2016) [22] proved that the horseshoe+ prior enjoys the same upper bound on the posterior contraction rate as the horseshoe using \( \tau \) as a tuning parameter, hence we are interested to find out whether results similar to Theorem 2 of our paper can be established for the horseshoe+ prior when \( \tau \) is estimated from the data. Results of Bhadra et al. (2017) [1] reveal that the role played by \( \tau \) in horseshoe+ is exactly same as that in the horseshoe prior, which motivates us to use the same empirical Bayes estimate of \( \tau \) that was used by van der Pas et al. (2014) [21] for horseshoe prior (defined in (3.1)).

The next theorem gives an upper bound to the mean square error of the empirical Bayes posterior estimate of \( \theta, T_\tau(X) \) corresponding to horseshoe+ prior. The proof of the theorem is given in section 6.

**Theorem 6.** Suppose \( X \sim \mathcal{N}_n(\theta_0, I_n) \) where \( \theta_0 \in l_0[p_n] \) with \( p_n \propto n^\beta \) where \( 0 < \beta < 1 \). Consider the hierarchical form (1.4). Then the empirical Bayes posterior estimate of \( \theta \) satisfies the following

\[
\sup_{\theta_0 \in l_0[p_n]} \mathbb{E}_{\theta_0} ||T_\tau(X) - \theta_0||^2 \precsim 2p_n \log n
\]
as \( p_n \to \infty \) and \( p_n = o(n) \) as \( n \to \infty \).

Under the assumption \( p_n \propto n^\beta \) where \( 0 < \beta < 1 \), the corresponding minimax error rate under the squared \( l_2 \) norm is of the order of \( 2p_n \log n \), which implies that like the general class of priors of the form \( \text{(1.3)} \), for horseshoe+ prior too, the data-dependent estimate \( \hat{T}_\tau(X) \) attains a near minimax \( l_2 \) risk upto some multiplicative constant even though the slowly varying function is not bounded above.

The next theorem provides upper bounds to the total posterior variance corresponding to horseshoe+ prior using the empirical Bayes estimate of \( \tau \). The proof of this theorem is in section \( \text{6.} \)

**Theorem 7.** Suppose \( X \sim N_n(\theta_0, I_n) \) where \( \theta_0 \in l_0[p_n] \) with \( p_n \propto n^\beta \) where \( 0 < \beta < 1 \). Consider the hierarchical form \( \text{(1.4)} \). Then the total posterior variance using the empirical Bayes estimate of \( \tau \) corresponding to the horseshoe+ prior satisfies, as \( n \to \infty \),

\[
\sup_{\theta_0 \in l_0[p_n]} \mathbb{E}_{\theta_0} \left( \sum_{i=1}^n \text{Var}(\theta_i|X_i, \hat{\tau}) \right) \lesssim p_n \log n
\]

where \( \text{Var}(\theta_i|X_i, \hat{\tau}) \) denotes \( \text{Var}(\theta_i|X_i, \tau) \) evaluated at \( \tau = \hat{\tau} \).

The next theorem provides upper bounds to the rates of posterior contraction for the horseshoe+ prior, both around the true parameter \( \theta_0 \) and the corresponding Bayes estimate \( \hat{T}_\tau(X) \).

**Theorem 8.** Suppose the assumptions of Theorem \( \text{6.} \) hold, then empirical Bayes posterior contracts around both the true parameter \( \theta_0 \) and the corresponding Bayes estimate \( \hat{T}_\tau(X) \) at the following rate,

\[
\sup_{\theta_0 \in l_0[p_n]} \mathbb{E}_{\theta_0} \Pi_{\hat{\tau}} \left( \theta : ||\theta - \theta_0||^2 > M_n p_n \log n | X \right) \to 0,
\]

and

\[
\sup_{\theta_0 \in l_0[p_n]} \mathbb{E}_{\theta_0} \Pi_{\hat{\tau}} \left( \theta : ||\theta - \hat{T}_\tau(X)||^2 > M_n p_n \log n | X \right) \to 0 .
\]

for any \( M_n \to \infty \), \( p_n \to \infty \) with \( p_n = o(n) \) as \( n \to \infty \).

**Proof.** Using Markov’s inequality along with Theorem \( \text{6.} \) and Theorem \( \text{7.} \) of this paper provides the first part, while use of Markov’s inequality together with Theorem \( \text{4.} \) is sufficient for proving the 2nd part.

The first statement of Theorem \( \text{8.} \) concludes that the rate of contraction of the posterior distribution of \( \theta \) around the true \( \theta_0 \) is at least as fast as the minimax \( l_2 \) risk. On the other hand, the Theorem 2.5 of Ghosal et
al. (2000) reveals that the posterior distribution cannot contract around the truth faster than the minimax risk. Combining these two results in stating that the rate at which the posterior distribution of the parameter vector \( \theta \) contracts around \( \theta_0 \) is a near minimax rate. However, the same can not be stated necessarily for the corresponding Bayes estimates of the mean vector as the posterior spread may be of smaller order than the rate at which the full Bayes estimator approaches the underlying mean vector. An important conclusion that can be drawn that in spite of having unbounded slowly varying function, still using an adaptive estimate of \( \tau \), the rate of contraction of the posterior distribution is exactly same as that of van der Pas et al. (2017) for horseshoe prior. To the best of our knowledge, this is the first result regarding the concentration property of the posterior distribution of the mean vector when it is modeled using horseshoe+ prior assuming the proportion of non-zero means being unknown.

5 Concluding remarks

In this paper, we study a high-dimensional sparse normal means model. Ghosh and Chakrabarti (2017) had shown that in case of a sparse normal means model when the proportion of non-zero means is known, Bayes estimate corresponding to the general class of prior of the form asymptotically attain the exact minimax risk with respect to the \( l_2 \) norm, possibly up to some multiplicative constants. They also proved that by appropriately choosing the global shrinkage parameter \( \tau \), the posterior contracts around the true value at the minimax rate. However no such result related to the rate of contraction of the posterior distribution was proved when either \( \tau \) is learnt from the data or a non-degenerate prior is used on \( \tau \). As mentioned earlier, van der Pas et al. (2014) proved that when the level of sparsity is unknown, the empirical Bayes posterior of horseshoe prior still contracts around the truth at a near minimax risk with respect to the \( l_2 \) norm, possibly up to some multiplicative constants. Later van der Pas et al. (2017) stated that under some conditions on the prior density on \( \tau \), as discussed in section 3.2, the same result holds for the full Bayes estimate, too. This gives raise of the question whether results similar to that of van der Pas et al. (2014), (2017) hold generally for the class of priors modeled through equations (1.2) and (1.3). Theorem 2 and 5 of this paper give an affirmative answer to these questions and generalize the above mentioned results of van der Pas for the general class of one group tail robust shrinkage priors mentioned in (1.3).

One interesting thing that we should highlight is that using some of the arguments of Ghosh and Chakrabarti (2017), the possible extension of the results of van der Pas et al. (2014, 2017) is not that difficult
when $a \in \left[\frac{1}{2}, 1\right)$, but if one is interested to know whether similar results hold for $a \geq 1$, this question can not be answered using the techniques of Ghosh and Chakrabarti (2017) [11] since the upper bound for the posterior variance claimed there for $a \geq 1$ had a soft up of. However employing some novel arguments and establishing some upper bounds related to the posterior variance and the posterior mean of the shrinkage coefficient, we have been able to prove the corresponding posterior contraction rate is of the order of the minimax rate when the global shrinkage parameter is either estimated from the data or a non-degenerate prior is used on it assuming the proportion of non-zero means is unknown. Lemma [1] proved in this context is the backbone of extending the results even if $a \geq 1$. As a consequence of this, the results of van der Pas et al. (2014 [21], 2017 [23]) can be generalized for the above mentioned class of one group tail robust shrinkage priors of the form (1.3) for $a \geq 0.5$.

Though our chosen class of priors in (1.3) is rich enough to include three parameter beta normal mixtures, the generalized double Pareto priors, and the horseshoe-type priors like the inverse– gamma priors, the half-t priors, but due to the assumption on the slowly varying function $L$, horseshoe+ prior is not included in this class. Since, all of our results hinge upon the assumption, none of the results mentioned in section 3.1 or section 3.2 hold for horseshoe+ prior. When the proportion of non-zero means is known, van der Pas et al. (2016) [22] modeled the mean vector by horseshoe+ prior. They proved that the posterior concentration result also holds for horseshoe+ prior using the appropriate choice of $\tau$, but whether same result holds for horseshoe+ prior when the proportion of non-zero means is unknown still remains unanswered. As a answer to this question, using the empirical Bayes estimate due to van der Pas et al. (2014) [21], our results in the section 4 reveals about the same when the underlying unknown sparsity is estimated by a data-dependent approach. To the best of our knowledge, our results related to the rate of the contraction of the posterior distribution of the mean vector is still first one to be available in the literature for horseshoe+ prior when the underlying sparsity is measured by an adaptive estimate of the global shrinkage parameter. Finally, we expect that the theoretical findings of our work provide some useful contributions to the literature of the posterior contraction rate of the mean vector when it is studied using the class of global-local shrinkage priors in the case of proportion of non-zero means being unknown.

6 Proofs

Before starting the proofs, let us first find out a relationship between $\hat{\theta}_i$ and $T_\tau(X_i)$, which will be one of the key steps in proving the theorems of this section. Using (1.1)-(1.3) and the condition (C1) for the full Bayes
approach, posterior mean of $\theta_i$ can be written as

$$\hat{\theta}_i = E(\theta_i | X) = \int_{1/n}^1 E(\theta_i | X, \tau) \pi_n(\tau | X) d\tau$$

$$= \int_{1/n}^1 E(1 - \kappa_i | X_i, \tau) X_i \cdot \pi_n(\tau | X) d\tau$$

$$= \int_{1/n}^1 T_\tau(X_i) \pi_n(\tau | X) d\tau$$

where the equality in the second line follows due to the fact that posterior distribution of $\theta_i$ given $(X, \tau)$ depends only on $(X_i, \tau)$ only.

Next comes a very important Lemma, which is one of the main pillars for proving the upcoming Theorems 1-5 in this article.

**Lemma 1.** Suppose $X \sim \mathcal{N}_n(\theta_0, I_n)$. Consider the class of priors with $a \geq 1$. Then for any $\tau \in (0, 1)$ and $x \in \mathbb{R}$,

$$E(1 - \kappa | x, \tau) \leq \left( \tau^2 e^{x^2} + K \int_1^\infty \frac{t\tau^2}{1 + t\tau^2} \cdot \frac{1}{\sqrt{1 + t\tau^2}} t^{-a-1} L(t) e^{t^2 \frac{x^2}{1+\tau^2}} dt \right) (1 + o(1)) \quad (L.1.1)$$

where $o(1)$ depends only on $\tau$ such that $\lim_{\tau \to 0} o(1) = 0$ and $\int_0^\infty t^{-a-1} L(t) dt = K^{-1}$.

**Proof.** Posterior distribution of $\kappa$ given $x$ and $\tau$ is,

$$\pi(\kappa | x, \tau) \propto \kappa^{-\frac{1}{2}}(1 - \kappa)^{-a-1} L \left( \frac{1}{\tau^2} (\frac{1}{\kappa} - 1) \right) \exp \left( \frac{(1 - \kappa)x^2}{2} \right), 0 < \kappa < 1.$$

Using the transformation $t = \frac{1}{\tau^2} (\frac{1}{\kappa} - 1)$, $E(1 - \kappa | x, \tau)$ becomes

$$E(1 - \kappa | x, \tau) = \frac{\tau^2 \int_0^\infty (1 + t\tau^2)^{-\frac{1}{2}} t^{-a} L(t) e^{t^2 \frac{x^2}{1+\tau^2}} dt}{\int_0^\infty (1 + t\tau^2)^{-\frac{1}{2}} t^{-a-1} L(t) e^{t^2 \frac{x^2}{1+\tau^2}} dt}.$$

Note that,

$$\int_0^\infty (1 + t\tau^2)^{-\frac{1}{2}} t^{-a-1} L(t) e^{t^2 \frac{x^2}{1+\tau^2}} dt \geq \int_0^\infty (1 + t\tau^2)^{-\frac{1}{2}} t^{-a-1} L(t) dt = K^{-1}(1 + o(1)).$$
The equality in the last line follows due to the Dominated Convergence Theorem. Hence,

\[ E(1 - \kappa | x, \tau) \leq K(A_1 + A_2)(1 + o(1)), \tag{L-1.2} \]

where

\[ A_1 = \int_0^1 \frac{t^2}{1 + t^2} \cdot \frac{1}{\sqrt{1 + t^2}} t^{-a-1} L(t) e^{\frac{a^2}{2} \frac{t^2}{1 + t^2}} dt. \]

and

\[ A_2 = \int_1^\infty \frac{t^2}{1 + t^2} \cdot \frac{1}{\sqrt{1 + t^2}} t^{-a-1} L(t) e^{\frac{a^2}{2} \frac{t^2}{1 + t^2}} dt. \]

Since, for any \( t \leq 1 \) and \( \tau \leq 1, \frac{t^2}{1 + t^2} \leq \frac{1}{2} \) and using the fact \( \int_0^\infty t^{-a-1} L(t) dt = K^{-1} \) (obtained from (1.3)),

\[ A_1 \leq K^{-1} \tau^2 e^{\frac{a^2}{4}}. \tag{L-1.3} \]

Using the (L-1.2) and (L-1.3) along with the form of \( A_2 \), for any \( \tau \in (0, 1) \) and \( x \in \mathbb{R} \), we get the result of the form (L-1.1).

**Remark 1.** This Lemma is a refinement of the Lemma 2 of Ghosh and Chakrabarti (2017) [11], which is obtained by a careful inspection and division of the range of the integral in \((0, 1)\) and \((1, \infty)\). Note that, the upper bound of \( E(1 - \kappa | x, \tau) \) within the interval \((0, 1)\) is sharper than that of Ghosh and Chakrabarti (2017) [11]. This order is important for obtaining the optimal posterior contraction rate for both the data-dependent estimate of \( \tau \) and prior used on \( \tau \) when \( a \geq 1 \) and hence has been used as a key ingredient for proving the Theorems 1, 3 and 4 when \( a \geq 1 \), which is one of the main contributions of our work in this paper. It is not difficult to provide an upper bound on the second term in \( E(1 - \kappa | x, \tau) \) such that calculations based on the upper bound can still attain the near minimax rate, which we will provide later when the range of integration is finite. In case of the integration having infinite range, we will use the bound of the form (L-1.1) and deal two situations \( a = 1 \) and \( a > 1 \) separately. For full calculations, go through Case-2 in the proof of Theorem 1 when \( a \geq 1 \). Another important point that we need to mention here is that the upper bound on the posterior shrinkage coefficient is independent of any assumption on \( L(\cdot) \), but assumptions (A1) and (A2) will definitely be used in proving the subsequent results.

Now we prove very important lemma for horseshoe+ prior, which leads to establish the posterior contraction rate, that are given in Theorems 6-8.
Lemma 2. Let us consider the hierarchical model of the form \( \text{(1.4)} \) and define \( \kappa = 1/(1 + \lambda^2 \tau^2) \). Then for any \( \tau \in (0, 1) \) and \( x \in \mathbb{R} \), we have the following bounds for horseshoe+ prior,

\[
(1) \quad E(1 - \kappa | x, \tau) \leq \tau^2 .
\]

(2) The absolute value of the difference between the posterior mean \( T_\tau(x) \) and the observation \( x \) can be bounded above by a non-negative real valued function \( g(\cdot, \tau) \) depending on \( c \) satisfies, for any \( \rho > c \),

\[
\lim_{\tau \downarrow 0} \sup_{|x| > \sqrt{\rho \log 1/\tau}} g(x, \tau) = 0 .
\]

(3) \( \text{Var}(\theta | x, \tau) \) can be bounded above by another non-negative real valued function \( \tilde{g}(\cdot, \tau) \) depending on \( C_1 \) satisfies, for any \( \rho > C_1 \),

\[
\lim_{\tau \downarrow 0} \sup_{|x| > \sqrt{\rho^2 \log 1/\tau}} \tilde{g}(x, \tau) = 1 .
\]

Proof. The posterior distribution of \( \kappa \) given \( x \) and \( \tau \) is of the form,

\[
\pi(\kappa | x, \tau) \propto (1 - \kappa)^{-\frac{1}{2}} \exp\left(-\frac{\kappa x^2}{2}\right) |\log\left(\frac{\kappa x^2}{(1-\kappa) \tau^2}\right)|^{\frac{1}{2}} \left(\kappa \tau^2 + 1\right), 0 < \kappa < 1 . \tag{L-2.1}
\]

Using the inequality \( 1 - \frac{1}{y} < \log(y) < y - 1 \) for \( y > 0 \), we have the following inequality

\[
\frac{1}{\kappa \tau^2} < \frac{|\log\left(\frac{\kappa x^2}{(1-\kappa) \tau^2}\right)|}{|\kappa (\tau^2 + 1) - 1|} < \frac{1}{1 - \kappa} . \tag{L-2.2}
\]

Now, for any \( x \in \mathbb{R} \) and \( \tau > 0 \),

\[
E(1 - \kappa | x, \tau) = \int_0^1 (1 - \kappa) \pi(\kappa | x, \tau) d\kappa
\]

\[
= \frac{\int_0^1 (1 - \kappa)(1 - \kappa)^{-\frac{1}{2}} \exp\left(-\frac{\kappa x^2}{2}\right) |\log\left(\frac{\kappa x^2}{(1-\kappa) \tau^2}\right)|^{\frac{1}{2}} \left(\kappa \tau^2 + 1\right) d\kappa}{\int_0^1 (1 - \kappa)^{-\frac{1}{2}} \exp\left(-\frac{\kappa x^2}{2}\right) |\log\left(\frac{\kappa x^2}{(1-\kappa) \tau^2}\right)|^{\frac{1}{2}} \left(\kappa \tau^2 + 1\right) d\kappa} . \quad [\text{Using } \text{(L-2.1)}]
\]
With the use of \((L-2.2)\) and using the range of \(\kappa\),

\[
E(1 - \kappa|x, \tau) \leq \frac{\int_0^1 (1 - \kappa)^{-\frac{1}{2}} \exp\left\{-\frac{\kappa x^2}{2}\right\} d\kappa}{\int_0^1 (1 - \kappa)^{-\frac{1}{2}} \exp\left\{-\frac{\kappa x^2}{2}\right\} \frac{1}{\kappa^{\frac{1}{2}}} d\kappa}
\leq \tau^2 .
\]

Therefore for any \(x \in \mathbb{R}\) and \(\tau > 0\), \(E(1 - \kappa|x, \tau) \leq \tau^2\). [1] is proved.

Now, for any \(x \in \mathbb{R}\) and \(\tau \in (0,1)\),

\[
|T_\tau(x) - x| = |xE(\kappa|x, \tau)|
\]

\[
= |x||\int_0^1 \kappa(1 - \kappa)^{-\frac{1}{2}} \exp\left\{-\frac{\kappa x^2}{2}\right\} \left\{\frac{\log\left(\frac{\kappa x^2}{\kappa(x + 1)}\right)}{\kappa(x + 1)}\right\} d\kappa
\]

\[
\leq \tau^2 .
\]

Fix any \(\eta \in (0,1)\) and define \(I_1(x, \tau) = |xE(\kappa1\{\kappa \leq \eta\}|x, \tau)\), \(I_2(x, \tau) = |xE(\kappa1\{\kappa > \eta\}|x, \tau)\). For any \(x \in \mathbb{R}\) and \(\tau \in (0,1)\)

\[
I_1(x, \tau) \leq |x||\int_0^\eta \kappa(1 - \kappa)^{-\frac{1}{2}} \exp\left\{-\frac{\kappa x^2}{2}\right\} \left\{\frac{\log\left(\frac{\kappa x^2}{\kappa(x + 1)}\right)}{\kappa(x + 1)}\right\} d\kappa
\]

First consider the transformation \(t = (\frac{1}{\kappa} - 1)\) and hence the upper bound of \(I_1(x, \tau)\) becomes,

\[
I_1(x, \tau) \leq |x||\int_0^\eta (1 + t)^{-\frac{1}{2}} t^{-\frac{1}{2}} \exp\left\{-\frac{x^2}{2(1+t)}\right\} dt
\]

Now consider the transformation \(u = \frac{x^2}{(1+t)}\), hence for the numerator of \((L-2.2)\),

\[
\int_0^\eta (1 + t)^{-\frac{1}{2}} t^{-\frac{1}{2}} \exp\left\{-\frac{x^2}{2(1+t)}\right\} dt = \int_0^{\eta x^2} \left(\frac{u}{x^2}\right)^{\frac{1}{2}} \left(\frac{x^2}{u} - 1\right)^{-\frac{1}{2}} e^{-\frac{x^2}{2u}} du
\]

\[
= \int_0^{\eta x^2} \left(\frac{u}{x^2}\right)^{\frac{3}{2}} (1 - \frac{u}{x^2})^{-\frac{3}{2}} e^{-\frac{x^2}{2u}} du
\]

\[
\leq \frac{1}{x^4} (1 - \eta)^{\frac{1}{2}} \int_0^{\eta x^2} u e^{-\frac{x^2}{2u}} du ,
\]

where the inequality in the last line follows due to the fact that for \(0 < u < \eta x^2\), \(1 - \eta < 1 - \frac{u}{x^2} < 1\). Similarly,
for the denominator in (L-2.3),
\[
\int_0^\infty (1 + t)^{-\frac{x}{2}} t^{-\frac{1}{2}} \exp\left\{-\frac{x^2}{2(1 + t)}\right\} dt = \int_0^{x^2} \left(\frac{u}{x^2}\right)^\frac{1}{2} \left(\frac{x^2}{u} - 1\right)^{-\frac{1}{2}} e^{-\frac{x^2}{u^2}} du \\
\geq \frac{1}{x^2} \int_0^{x^2} e^{-\frac{x^2}{u}} du.
\]

Using (L-2.3)-(L-2.5),
\[
I_1(x, \tau) \leq (1 - \eta)^{-\frac{x}{2}} \frac{1}{x} \left\{ \frac{\int_0^\infty u e^{-\frac{x^2}{u}} du}{\int_0^{x^2} e^{-\frac{x^2}{u}} du} \right\} = g_1(x, \tau), \text{ say,}
\]

where \( g_1(x, \tau) = C_* [x \cdot \int_0^{x^2} e^{-\frac{x^2}{u}} du]^{-1} \) where \( C_* \) is a global constant independent of both \( x \) and \( \tau \). Since, \( g_1(x, \tau) \) is independent of any \( \tau \) and is decreasing in \( |x| \), so
\[
\sup_{|x| > \sqrt{\rho \log \frac{1}{\tau}}} g_1(x, \tau) \leq g_1(\sqrt{\rho \log \frac{1}{\tau}}, \tau),
\]
which implies
\[
\lim_{\tau \downarrow 0} \sup_{|x| > \sqrt{\rho \log \frac{1}{\tau}}} g_1(x, \tau) = 0.
\]

Now, \( I_2(x, \tau) = |x| E(\kappa 1\{\kappa > \eta\}|x, \tau) \). Using Theorem 4 of Bhadra et al. (2017) [1],
\[
P(\kappa > \eta|x, \tau) \leq \exp\left\{-\frac{\eta(1 - \delta)x^2}{2}\right\} \frac{1}{\tau^2} C(\eta, \delta),
\]
where \( C(\eta, \delta) \) is a constant independent of \( x \). Hence,
\[
I_2(x, \tau) \leq \frac{1}{\tau^2} C(\eta, \delta) |x \exp\left\{-\frac{\eta(1 - \delta)x^2}{2}\right\}| = g_2(x, \tau), \text{ say.}
\]

Since, \( g_2(x, \tau) \) is decreasing in \( |x| \) for \( |x| > 1/\sqrt{\eta(1 - \delta)} \),
\[
\lim_{\tau \downarrow 0} \sup_{|x| > \sqrt{\rho \log \frac{1}{\tau}}} g_2(x, \tau) = \begin{cases} 0, \rho > c = \frac{2}{\eta(1 - \delta)}; \\ \infty, \text{ otherwise.} \end{cases}
\]
Hence, $|T_{\tau}(x) - x| \leq g(x, \tau) = g_1(x, \tau) + g_2(x, \tau)$ such that for any $\rho > c > 2$,

\[
\lim_{\tau \downarrow 0} \sup_{|x| > \sqrt{\rho \log \frac{1}{\tau}}} g(x, \tau) = 0.
\]

This proves \([2]\). For \([3]\), first we use the fact obtained from Lemma A.1. of Ghosh and Chakrabarti (2017) \([11]\), for any $x \in \mathbb{R}$ and $\tau \in (0, 1),$

\[
\text{Var}(\theta| x, \tau) = E[(1 - \kappa)|x, \tau] + x^2 E[\kappa^2|x, \tau] - x^2 E^2[\kappa|x, \tau] \leq 1 + x^2 E[\kappa^2|x, \tau]
\]

Again, we fix any $\eta \in (0, 1)$ and define $\hat{I}_1(x, \tau) = x^2 E(\kappa^2 1\{\kappa \leq \eta\}|x, \tau)$, $\hat{I}_2(x, \tau) = x^2 E(\kappa^2 1\{\kappa > \eta\}|x, \tau)$. Using exactly the same procedure as used in \([2]\) of Lemma \(2\) we obtain the following statements,

\[
\hat{I}_1(x, \tau) \leq \hat{g}_1(x, \tau) \quad \text{such that} \quad \lim_{\tau \downarrow 0} \sup_{|x| > \sqrt{\rho^2 \log \frac{1}{\tau}}} \hat{g}_1(x, \tau) = 0.
\]

and

\[
\hat{I}_2(x, \tau) \leq \hat{g}_2(x, \tau) \quad \text{such that for} \quad \rho > C_1 = \sqrt{\frac{2}{\eta(1 - \delta)}} \lim_{\tau \downarrow 0} \sup_{|x| > \sqrt{\rho^2 \log \frac{1}{\tau}}} \hat{g}_2(x, \tau) = 0.
\]

Now, define $\hat{g}(x, \tau) = 1 + \hat{g}_1(x, \tau) + \hat{g}_2(x, \tau)$ and using the above set of arguments, we obtain \([3]\) ·

Remark 2. The importance of these results is that in spite of having unbounded slowly varying function $L(\cdot)$, the upper bounds corresponding to the posterior shrinkage coefficient and that of the posterior variance is of the same order as that of the general class of priors expressed in the form of \([1.2]\) and \([1.3]\). In fact, the upper bound of the posterior shrinkage coefficient is smaller order than that of the above mentioned class of prior, which works as an intuition that the corresponding Bayes estimate of $\theta$ using the empirical Bayes estimate of $\tau$ in horseshoe+ prior can still contract around the truth at near minimax rate.

Proof of Theorem \([1]\).

Proof. Let us define $\hat{p}_n = \sum_{i=1}^{n} 1_{\{\theta_i \neq 0\}}$. Thus, $\hat{p}_n \leq p_n$. 

23
First we split $E_{\theta_0} \sum_{i=1}^{n} \text{Var} (\theta_i | X_i, \hat{\tau})$ as

$$E_{\theta_0} \sum_{i=1}^{n} \text{Var} (\theta_i | X_i, \hat{\tau}) = \sum_{i: \theta_{0i} \neq 0} E_{\theta_0} \text{Var} (\theta_i | X_i, \hat{\tau}) + \sum_{i: \theta_{0i} = 0} E_{\theta_0} \text{Var} (\theta_i | X_i, \hat{\tau}).$$  \hspace{1cm} (T-1.1)

We shall follow the following steps to prove the desired result:-

**Step-1** For any $i$ such that $\theta_{0i} \neq 0$, we will show that, for sufficiently large $n$, $E_{\theta_0} \text{Var} (\theta_i | X_i, \hat{\tau}) \lesssim \log n$, which results in proving, as $n \to \infty$,

$$\sum_{i: \theta_{0i} \neq 0} E_{\theta_0} \text{Var} (\theta_i | X_i, \hat{\tau}) \lesssim \bar{p}_n \log n .$$

**Step-2** For any $i$ such that $\theta_{0i} = 0$, we deal the cases $a \in \left[\frac{1}{2}, 1 \right)$ and $a \geq 1$ separately. For both of the above mentioned cases, we will prove that as $n \to \infty$

$$\sum_{i: \theta_{0i} = 0} E_{\theta_0} \text{Var} (\theta_i | X_i, \hat{\tau}) \lesssim p_n \log n .$$

Combining the two steps, we get the final result. Now let us derive **Step-1** and **Step-2**.

**Step-1** Fix any $i$ such that $\theta_{0i} \neq 0$. Now we split $E_{\theta_0} \text{Var} (\theta_i | X_i, \hat{\tau})$ as

$$E_{\theta_0} \text{Var} (\theta_i | X_i, \hat{\tau}) = E_{\theta_0} [\text{Var} (\theta_i | X_i, \hat{\tau})1_{\{|X_i| \leq \sqrt{4a\rho^2 \log n}\}}] + E_{\theta_0} [\text{Var} (\theta_i | X_i, \hat{\tau})1_{\{|X_i| > \sqrt{4a\rho^2 \log n}\}}].$$ \hspace{1cm} (T-1.2)

Since, for any fixed $x \in \mathbb{R}$ and $\tau > 0$, $\text{Var} (\theta | x, \tau) \leq 1 + x^2$, as $n \to \infty$,

$$E_{\theta_0} [\text{Var} (\theta_i | X_i, \hat{\tau})1_{\{|X_i| \leq \sqrt{4a\rho^2 \log n}\}}] \lesssim \log n .$$ \hspace{1cm} (T-1.3)

Note that, for any fixed $x \in \mathbb{R}$, $x^2 E(\kappa^2 | x, \tau) = x^2 E\left(\frac{1}{1+\lambda x \tau^2} | x, \tau\right)$ is non-increasing in $\tau$. Also using Lemma A.1 in Ghosh and Chakrabarti (2017) [11], $\text{Var} (\theta | x, \tau) \leq 1 + x^2 E(\kappa^2 | x, \tau)$. Using these two results,

$$\text{Var} (\theta | x, \hat{\tau}) \leq 1 + x^2 E(\kappa^2 | x, \hat{\tau})$$

$$\leq 1 + x^2 E(\kappa^2 | x, \frac{1}{n}) \quad \text{[Since } \hat{\tau} \geq \frac{1}{n}]$$

$$\leq \tilde{h}(x, \frac{1}{n}).$$

Using arguments similar to Lemma 3 of Ghosh and Chakrabarti (2017) [11], one can show that there exists a
non-negative and measurable real-valued function $\tilde{h}(x, \tau)$ satisfying $\tilde{h}(x, \tau) = 1 + \tilde{h}_1(x, \tau) + \tilde{h}_2(x, \tau)$ with

$$\tilde{h}_1(x, \tau) = C_\ast \left[ x^2 \int_0^{\frac{x^2 \log n}{1 + \rho_0}} \exp\left(-\frac{u}{2} \right) u^{a + \frac{1}{2} - 1} du \right]^{-1}$$

where $C_\ast$ is a global constant independent of both $x$ and $\tau$ and

$$\tilde{h}_2(x, \tau) = x^2 \frac{H(a, \eta, \delta)}{\Delta(\tau^2, \eta, \delta)} \tau^{-2a} e^{-\frac{2a(1-\eta)\tau^2}{2}}$$

where

$$H(a, \eta, \delta) = \frac{(a + \frac{1}{2})(1 - \eta \delta)^a}{K(\eta \delta)^{a + \frac{1}{2}}}$$

and

$$\Delta(\tau^2, \eta, \delta) = \int_0^{\infty} \left\{ \frac{t^{-\frac{1}{2}}}{\tau^2} \left( \frac{1}{\eta \delta} - 1 \right)^{-(a + \frac{1}{2})} \right\} L(t) dt$$

for any $\eta, \delta \in (0, 1)$. Since, $\tilde{h}_1(x, \tau)$ is strictly decreasing in $|x|$, hence for sufficiently large $n$,

$$\sup_{|x| > \sqrt{4a\rho^2 \log n}} \tilde{h}_1(x, \frac{1}{n}) \leq C_\ast \left[ \rho^2 \log n \int_0^{\frac{4a\rho^2 \log n}{1 + \rho_0}} \exp\left(-\frac{u}{2} \right) u^{a + \frac{1}{2} - 1} du \right]^{-1} \lesssim \frac{1}{\log n}$$

Also noting that $\tilde{h}_2(x, \tau)$ is strictly decreasing in $|x| > C_1 = \sqrt{\frac{2}{\eta(1-\delta)}}$, for any $\rho > C_1$,

$$\sup_{|x| > \sqrt{4a\rho^2 \log n}} \tilde{h}_2(x, \frac{1}{n}) \leq 4a\rho^2 \frac{H(a, \eta, \delta)}{\Delta(\tau^2, \eta, \delta)} \log n \cdot n^{2a} e^{-2a\rho^2 \eta(1-\delta) \log n} = 4a\rho^2 \frac{H(a, \eta, \delta)}{\Delta(\tau^2, \eta, \delta)} \log n \cdot n^{-2a(\frac{2a}{C_1^2} - 1)}$$

Since, $\lim_{n \to \infty} \Delta(\frac{1}{n^2}, \eta, \delta)$ is finite for every fixed $\eta$ and $\delta \in (0, 1)$,

$$\sup_{|x| > \sqrt{4a\rho^2 \log n}} \tilde{h}(x, \frac{1}{n}) \leq 1 + \sup_{|x| > \sqrt{4a\rho^2 \log n}} \tilde{h}_1(x, \frac{1}{n}) + \sup_{|x| > \sqrt{4a\rho^2 \log n}} \tilde{h}_2(x, \frac{1}{n}) \lesssim 1.$$ 

Using above arguments, as $n \to \infty$,

$$\mathbb{E}_{\theta_i} [\text{Var}(\hat{\theta}_i\mid X_i, \tilde{\tau}) 1_{\{|X_i| > \sqrt{4a\rho^2 \log n}\}}] \lesssim 1.$$ 

(T-1.4)
Combining (T-1.2), (T-1.3) and (T-1.4), we obtain for sufficiently large $n$,

$$
E_{\theta_{0i}} Var(\theta_i | X_i, \hat{\tau}) \lesssim \log n .
$$

(T-1.5)

Since all of the above arguments hold true for any $i$ such that $\theta_{0i} \neq 0$, as $n \to \infty$,

$$
\sum_{i: \theta_{0i} \neq 0} E_{\theta_{0i}} Var(\theta_i | X_i, \hat{\tau}) \lesssim \tilde{p}_n \log n .
$$

(T-1.6)

**Step-2** Fix any $i$ such that $\theta_{0i} = 0$.

**Case-1** First we consider the case when $a \in [\frac{1}{2}, 1)$. Now we split $E_{\theta_{0i}} Var(\theta_i | X_i, \hat{\tau})$ as

$$
E_{\theta_{0i}} Var(\theta_i | X_i, \hat{\tau}) = E_{\theta_{0i}}[Var(\theta_i | X_i, \hat{\tau}) 1_{\{|X_i| > \sqrt{4a\rho^2 \log n}\}}] + E_{\theta_{0i}}[Var(\theta_i | X_i, \hat{\tau}) 1_{\{|X_i| \leq \sqrt{4a\rho^2 \log n}\}}] .
$$

(T-1.7)

Using $Var(\theta|x, \hat{\tau}) \leq 1 + x^2$ and the identity $x^2 \phi(x) = \phi(x) - \frac{d}{dx}[x\phi(x)]$, we obtain, as $n \to \infty$,

$$
E_{\theta_{0i}}[Var(\theta_i | X_i, \hat{\tau}) 1_{\{|X_i| > \sqrt{4a\rho^2 \log n}\}}]\leq 2 \int_{\sqrt{4a\rho^2 \log n}}^{\infty} (1 + x^2)\phi(x) dx \lesssim \log n \cdot n^{-2a\rho^2} .
$$

(T-1.8)

Now let us choose some $\gamma > 1$ such that $c_2 \gamma - 1 > 1$. Next, we decompose the second term as follows:

$$
E_{\theta_{0i}}[Var(\theta_i | X_i, \hat{\tau}) 1_{\{|X_i| \leq \sqrt{4a\rho^2 \log n}\}}] = E_{\theta_{0i}}[Var(\theta_i | X_i, \hat{\tau}) 1_{\{|X_i| > \sqrt{4a\rho^2 \log n}\}} 1_{\{\hat{\tau} \geq \frac{c_2 \gamma - 1}{2}\}} 1_{\{|X_i| \leq \sqrt{4a\rho^2 \log n}\}}] +
E_{\theta_{0i}}[Var(\theta_i | X_i, \hat{\tau}) 1_{\{\hat{\tau} \leq \frac{c_2 \gamma - 1}{2}\}} 1_{\{|X_i| \leq \sqrt{4a\rho^2 \log n}\}}] .
$$

(T-1.9)

Note that

$$
E_{\theta_{0i}}[Var(\theta_i | X_i, \hat{\tau}) 1_{\{\hat{\tau} \geq \frac{c_2 \gamma - 1}{2}\}} 1_{\{|X_i| \leq \sqrt{4a\rho^2 \log n}\}}] \leq 4a\rho^2 \log n P_{\theta_{0i}}[\hat{\tau} > \frac{p_n}{n}, |X_i| \leq \sqrt{4a\rho^2 \log n}]
$$

$$
\leq 4a\rho^2 \log n P_{\theta_{0i}}[\hat{\tau} > \frac{p_n}{n}, |X_i| \leq \sqrt{4a\rho^2 \log n}]
$$

$$
\leq 4a\rho^2 \log n P_{\theta_{0i}}[\frac{1}{c_2 n} \sum_{j=1(\neq i)}^{n} 1_{\{|X_j| > \sqrt{4a\rho^2 \log n}\}} > \frac{p_n}{n}] .
$$

Note that, Since $a \geq \frac{1}{2}$ and $\rho > 1$, so $4a\rho^2$ plays the same role $c_1$ used by van der Pas et al. (2014) [21]. Hence, by employing similar arguments used for proving Lemma A.7 in van der Pas et al. (2014) [21], we can
show that,

\[ \mathbb{P}_{\theta_0} \left[ \frac{1}{c_2 n} \sum_{j=1}^{n} \mathbb{1}_{\{|X_j| > \sqrt{4a\rho^2 \log n}\}} > \gamma \frac{p_n}{n} \right] \lesssim \frac{p_n}{n}. \]

Employing these two results, as \( n \to \infty \),

\[ \mathbb{E}_{\theta_0} \left[ \text{Var}(\theta_i \mid X_i, \hat{\tau})1_{\{\hat{\tau} > \gamma \frac{p_n}{n}\}} \mathbb{1}_{\{|X_i| \leq \sqrt{4a\rho^2 \log n}\}} \right] \lesssim \frac{p_n}{n} \log n. \quad (T-1.10) \]

For the final part, \( \mathbb{E}_{\theta_0} \left[ \text{Var}(\theta_i \mid X_i, \hat{\tau})1_{\{\hat{\tau} \leq \gamma \frac{p_n}{n}\}} \mathbb{1}_{\{|X_i| \leq \sqrt{4a\rho^2 \log n}\}} \right] \], since for any fixed \( x \in \mathbb{R} \) and \( \tau > 0 \),

\[ \text{Var}(\theta \mid x, \tau) \leq E(1 - \kappa \mid x, \tau) + J(x, \tau) \]

where \( J(x, \tau) = x^2 E(1 - \kappa \mid x, \tau) \). Since \( E(1 - \kappa \mid x, \tau) \) is non-decreasing in \( \tau \), so, \( E(1 - \kappa \mid \tau, \hat{\tau}) \leq E(1 - \kappa \mid \tau, \gamma \frac{p_n}{n}) \) whenever \( \hat{\tau} \leq \gamma \frac{p_n}{n} \). Using lemma 2 of Ghosh and Chakrabarti (2017) \[11\],

\[ E(1 - \kappa \mid \tau, \gamma \frac{p_n}{n}) \lesssim (\frac{p_n}{n})^2 e^{\frac{x^2}{2}}. \]

Similarly, \( J(x, \gamma \frac{p_n}{n}) \lesssim (\frac{p_n}{n})^2 e^{\frac{x^2}{2}} \).

Using the above two arguments, for sufficiently large \( n \),

\[ \mathbb{E}_{\theta_0} \left[ \text{Var}(\theta_i \mid X_i, \hat{\tau})1_{\{\hat{\tau} \leq \gamma \frac{p_n}{n}\}} \mathbb{1}_{\{|X_i| \leq \sqrt{4a\rho^2 \log n}\}} \right] \lesssim (\frac{p_n}{n})^2 \int_0^{\sqrt{4a\rho^2 \log n}} e^{\frac{x^2}{2}} e^{-\frac{x^2}{2}} \, dx = (\frac{p_n}{n})^2 \sqrt{\log n}. \quad (T-1.11) \]

Note that all these preceding arguments hold uniformly in \( i \) such that \( \theta_{0i} = 0 \). Combining all these results, for \( a \in \left[ \frac{1}{2}, 1 \right) \) using (T-1.7)-(T-1.11), as \( n \to \infty \),

\[ \sum_{i: \theta_{0i} = 0} \mathbb{E}_{\theta_0} \left[ \text{Var}(\theta_i \mid X_i, \hat{\tau}) \right] \lesssim (n - \hat{p}_n)[\sqrt{\log n \cdot n^{-2a\rho^2}} + \frac{p_n}{n} \cdot \log n + (\frac{p_n}{n})^2 a \sqrt{\log n}] \lesssim p_n \log n. \quad (T-1.12) \]

The second inequality follows due to the fact that \( \hat{p}_n \leq p_n \) and \( p_n = o(n) \) as \( n \to \infty \).
Case-2 Now we assume $a \geq 1$ and split $E_{\theta_0}Var(\theta_i|X_i, \hat{\tau})$ as

$$E_{\theta_0}Var(\theta_i|X_i, \hat{\tau}) = E_{\theta_0}[Var(\theta_i|X_i, \hat{\tau})1_{\{|X_i|>\sqrt{2\rho^2 \log n}\}}] + E_{\theta_0}[Var(\theta_i|X_i, \hat{\tau})1_{\{|X_i|\leq \sqrt{2\rho^2 \log n}\}}]$$  \hspace{1cm} (T-1.13)

Using exactly the same arguments when $a \in [\frac{1}{2}, 1)$, we have the followings for $a \geq 1$ as $n \to \infty$

$$E_{\theta_0}[Var(\theta_i|X_i, \hat{\tau})1_{\{|X_i|>\sqrt{2\rho^2 \log n}\}}] \lesssim \sqrt{\log n} \cdot n^{-\rho^2}. \hspace{1cm} (T-1.14)$$

and

$$E_{\theta_0}[Var(\theta_i|X_i, \hat{\tau})1_{\{|X_i|\leq \sqrt{2\rho^2 \log n}\}}] \lesssim \frac{p_n}{n} \log n. \hspace{1cm} (T-1.15)$$

For the part $E_{\theta_0}[Var(\theta_i|X_i, \hat{\tau})1_{\{|X_i|>\sqrt{2\rho^2 \log n}\}}]$, note that for fixed $x \in \mathbb{R}$ and any $\tau > 0$,

$$Var(\theta|x, \tau) \leq E(1-\kappa|x, \tau) + x^2E[(1-\kappa)^2|x, \tau]$$

$$\leq E(1-\kappa|x, \tau) + x^2E(1-2\kappa|x, \tau)$$

$$\leq E(1-\kappa|x, \tau)1_{\{|x|\leq 1\}} + 2x^2E(1-\kappa|x, \tau). \hspace{1cm} (T-1.16)$$

Note that the above set of inequalities are true for any value of $a > 0$, but results of optimal posterior contraction rate can be proved without using these bounds for $a \in [\frac{1}{2}, 1)$, but these are indeed useful for $a \geq 1$. Since for fixed $x \in \mathbb{R}$ and any $\tau > 0$, $E(1-\kappa|x, \tau)$ is non-decreasing in $\tau$ and hence with the use of (T-1.16),

$$E_{\theta_0}[Var(\theta_i|X_i, \hat{\tau})1_{\{|X_i|>\sqrt{2\rho^2 \log n}\}}] \leq E_{\theta_0}[E(1-\kappa_i|X_i, \gamma \frac{p_n}{n})1_{\{|X_i|\leq 1\}}] + 2E_{\theta_0}[X_i^2E(1-\kappa_i|X_i, \gamma \frac{p_n}{n})1_{\{|X_i|\leq \sqrt{2\rho^2 \log n}\}}]. \hspace{1cm} (T-1.17)$$

For the first term in (T-1.17), we use Lemma 1 along with using for any $\tau \in (0, 1)$,

$$\frac{t^2}{1+t^2} \cdot \frac{1}{\sqrt{1+t^2}} t^{-a-1} \leq \tau t^{-(a+\frac{1}{2})}$$

and the boundedness of $L(t)$, the second term $A_2$ can be bounded as,

$$A_2 \leq \frac{2M}{(2a-1)} \tau e^{\frac{\rho^2}{2}}.$$
Hence, we have for sufficiently large \( n \)

\[
\mathbb{E}_{\Theta_n} [E(1 - \kappa_i | X_i, \gamma \frac{p_n}{n}) 1_{\{|X_i| \leq 1\}}] \lesssim \frac{p_n}{n} \int_0^1 e^{-\frac{t^2}{2}} \, dx + \frac{p_n}{n}.
\]

Hence,

\[
\mathbb{E}_{\Theta_n} [E(1 - \kappa_i | X_i, \gamma \frac{p_n}{n}) 1_{\{|X_i| \leq 1\}}] \lesssim \frac{p_n}{n}.
\]

For the second term in \((T-1.17)\) we shall use the upper bound of \( E(1 - \kappa|x, \tau) \) of the form \((L \rightarrow 1)\) and hence,

\[
\mathbb{E}_{\Theta_n} [X^2 E(1 - \kappa_i | X_i, \gamma \frac{p_n}{n}) 1_{\{|X_i| \leq \sqrt{2 \rho^2 \log n}\}}] \lesssim \frac{p_n}{n} \int_0^{\sqrt{2 \rho^2 \log n}} x^2 e^{\frac{x^2}{2}} \phi(x) \, dx
\]

\[
+ \int_0^{\sqrt{2 \rho^2 \log n}} \int_1^\infty \frac{t(\gamma \frac{p_n}{n})^2}{1 + t(\gamma \frac{p_n}{n})^2} \frac{1}{\sqrt{1 + t(\gamma \frac{p_n}{n})^2}} t^{-a-1} L(t) e^{\frac{x^2}{2}} \left( \frac{2 \rho^2 \log n}{1 + t(\gamma \frac{p_n}{n})^2} \right)^{\frac{1}{2}} x^2 \phi(x) \, dt \, dx.
\]

Note that the first integral is bounded by a constant and for the second integral using Fubini’s theorem and the transformation \( y = \frac{x}{\sqrt{1 + t(\gamma \frac{p_n}{n})^2}} \), the second term becomes

\[
\int_1^\infty t^{-a-1} L(t) t(\gamma \frac{p_n}{n})^2 \left( \int_0^{\sqrt{2 \rho^2 \log n}} y^2 e^{-\frac{y^2}{2}} \, dy \right) \, dt.
\]

We handle the above integral separately for \( a = 1 \) and \( a > 1 \). For \( a > 1 \), using the boundedness of \( L(t) \) it is easy to show that,

\[
\int_1^\infty t^{-a-1} L(t) t(\gamma \frac{p_n}{n})^2 \left( \int_0^{\sqrt{2 \rho^2 \log n}} y^2 e^{-\frac{y^2}{2}} \, dy \right) \, dt \lesssim \frac{p_n}{n}.
\]

For \( a = 1 \) note that,

\[
\int_1^\infty t^{-a-1} L(t) t(\gamma \frac{p_n}{n})^2 \left( \int_0^{\sqrt{2 \rho^2 \log n}} y^2 e^{-\frac{y^2}{2}} \, dy \right) \, dt \leq \int_1^\infty t^{-a-1} L(t) t(\gamma \frac{p_n}{n})^2 \left( \int_0^{\sqrt{2 \rho^2 \log n}} y^2 e^{-\frac{y^2}{2}} \, dy \right) \, dt
\]

\[
= (\gamma \frac{p_n}{n})^2 \frac{2 \pi}{\sqrt{2 \rho^2 \log n}} \frac{1}{t} L(t) \, dt + (\sqrt{2 \rho^2 \log n})^3 (\gamma \frac{p_n}{n})^2 \int_1^\infty t^{-2} L(t) \, dt.
\]

\[
(T-1.21)
\]
Here the division in the range of $t$ in (T-1.21) occurs due to the fact the integral $(\int_0^{\sqrt{\frac{2\rho^2 \log n}{t(\gamma pn)^2}} y^2 e^{-y^2} dy})$ can be bounded by $(\frac{2\rho^2 \log n}{t(\gamma pn)^2(2\pi)^{\frac{3}{2}}})$ when $t \leq \frac{2\rho^2 \log n}{(\gamma pn)^2(2\pi)^{\frac{3}{2}}}$ and by $\frac{\sqrt{2\pi}}{2}$ when $t > \frac{2\rho^2 \log n}{(\gamma pn)^2(2\pi)^{\frac{3}{2}}}$. For the first term in (T-1.21) with the boundedness of $L(t)$,

\[(\gamma \frac{p_n}{n})^2 \int_1^{\frac{2\rho^2 \log n}{t(\gamma pn)^2(2\pi)^{\frac{3}{2}}}} \frac{1}{t} L(t) dt \leq (\gamma \frac{p_n}{n})^2 M \log \left( \frac{2\rho^2 \log n}{(\gamma pn)^2(2\pi)^{\frac{3}{2}}} \right).\]

Hence for sufficiently large $n$ with $p_n = o(n)$,

\[(\gamma \frac{p_n}{n})^2 \int_1^{\frac{2\rho^2 \log n}{t(\gamma pn)^2(2\pi)^{\frac{3}{2}}}} \frac{1}{t} L(t) dt \lesssim \frac{p_n}{n} \sqrt{\log n}.\]  

(T-1.22)

Now for the second term in (T-1.21) again using the boundedness of $L(t)$,

\[(\sqrt{2\rho^2 \log n})^3 (\gamma \frac{p_n}{n})^2 \int_1^{\infty} \frac{1}{t} \cdot t^{-2} L(t) \cdot \frac{1}{(t(\gamma pn)^2(2\pi)^{\frac{3}{2}})} dt \lesssim \frac{p_n}{n}.\]  

(T-1.23)

So, using (T-1.21)-(T-1.23), for $a = 1$,

\[\int_1^{\infty} t^{-a-1} L(t) t(\gamma \frac{p_n}{n})^2 \left( \int_0^{\sqrt{\frac{2\rho^2 \log n}{1+t(\gamma pn)^2}} y^2 e^{-y^2} dy} \right) dt \lesssim \frac{p_n}{n} \sqrt{\log n}.\]  

(T-1.24)

With the help of (T-1.20) and (T-1.24), we have, for $a \geq 1$

\[\int_1^{\infty} t^{-a-1} L(t) t(\gamma \frac{p_n}{n})^2 \left( \int_0^{\sqrt{\frac{2\rho^2 \log n}{1+t(\gamma pn)^2}} y^2 e^{-y^2} dy} \right) dt \lesssim \frac{p_n}{n} \sqrt{\log n}.\]  

(T-1.25)

Using all these arguments, we finally have

\[\mathbb{E}_{\theta_{0i}} [X_i^2 E(1 - \kappa_{ij} | X_i, \gamma \frac{p_n}{n}) 1_{|X_i| \leq \sqrt{2\rho^2 \log n}}] \lesssim \frac{p_n}{n} \sqrt{\log n}.\]  

(T-1.26)

Note that all these preceding arguments hold uniformly in $i$ such that $\theta_{0i} = 0$. Combining all these results, for
\(a \geq 1\), using (T-1.13)-(T-1.26), as \(n \to \infty\),

\[
\sum_{i:\theta_0i=0} \mathbb{E}_{\theta_0i} \text{Var} (\theta_i | X_i, \hat{\tau}) \lesssim (n - \tilde{p}_n)[\sqrt{\log n \cdot n^{-a}} + \frac{pn}{n} \cdot \log n + \frac{pn}{n} \sqrt{\log n}]
\]

\[
\lesssim pn \log n .
\]  

(T-1.27)

With the use of (T-1.12) and (T-1.27), for \(a \geq \frac{1}{2}\), as \(n \to \infty\)

\[
\sum_{i:\theta_0i=0} \mathbb{E}_{\theta_0i} \text{Var} (\theta_i | X_i, \hat{\tau}) \lesssim pn \log n .
\]  

(T-1.28)

Now using (T-1.1), (T-1.6) and (T-1.28), for sufficiently large \(n\),

\[
\mathbb{E}_{\theta_0} \sum_{i=1}^{n} \text{Var} (\theta_i | X_i, \hat{\tau}) \lesssim pn \log n .
\]

Finally, taking supremum over all \(\theta_0 \in l_0[p_n]\), the result is obtained.

\[\blacksquare\]

**Proof of Theorem 3.**

Proof. Let us define \(\tilde{p}_n = \sum_{i=1}^{n} 1_{(\theta_0i \neq 0)}\). In this case also, we will follow the same steps mentioned in previous theorem. Now, we split the mean square error as

\[
\sum_{i=1}^{n} \mathbb{E}_{\theta_0i} (\hat{\theta}_i - \theta_{0i})^2 = \sum_{i:\theta_0i \neq 0} \mathbb{E}_{\theta_0i} (\hat{\theta}_i - \theta_{0i})^2 + \sum_{i:\theta_0i = 0} \mathbb{E}_{\theta_0i} (\hat{\theta}_i - \theta_{0i})^2 .
\]  

(T-3.1)

**Step-1** Fix any \(i\) such that \(\theta_{0i} \neq 0\). Applying Cauchy-Schwartz inequality and using \(\mathbb{E}_{\theta_0i} (X_i - \theta_{0i})^2 = 1\), we get,

\[
\mathbb{E}_{\theta_0i} (\hat{\theta}_i - \theta_{0i})^2 \leq \left[ \sqrt{\mathbb{E}_{\theta_0i} (\hat{\theta}_i - X_i)^2} + 1 \right]^2 .
\]  

(T-3.2)
Using Condition (C1) on $\tau$,

$$
E_{\theta_0 i} (\hat{\Theta}_i - X_i)^2 = E_{\theta_0 i} \left( \int_{1/\rho}^{1} (T_{\tau}(X_i) - X_i) \pi_n(\tau|\bm{X}) d\tau \right)^2
$$

$$
\leq E_{\theta_0 i} \left( \int_{1/\rho}^{1} (T_{\tau}(X_i) - X_i)^2 \pi_n(\tau|\bm{X}) d\tau \right)
$$

$$
= A_1 + A_2, \text{ Say},
$$

where $A_1 = E_{\theta_0 i} \left( \int_{1/\rho}^{1} (T_{\tau}(X_i) - X_i)^2 1_{\{X_i \leq \sqrt{4a\rho^2 \log n}\}} \pi_n(\tau|\bm{X}) d\tau \right)$. Since, $|T_{\tau}(x) - x| \leq \sqrt{4a\rho^2 \log n}$ whenever $|x| \leq \sqrt{4a\rho^2 \log n}$, which implies, $A_1 \leq 4a\rho^2 \log n$.

Now, $A_2 = E_{\theta_0 i} \left( \int_{1/\rho}^{1} (T_{\tau}(X_i) - X_i)^2 1_{\{X_i > \sqrt{4a\rho^2 \log n}\}} \pi_n(\tau|\bm{X}) d\tau \right)$, note that, for each fixed $x$, $|T_{\tau}(x) - x| = |x| E(\lambda^2 \tau^2 |x, \tau)$ is non-increasing in $\tau$. So, using the same arguments used for proving Theorem 2 in Ghosh and Chakrabarti (2017) [11], $\forall \tau \in [1/n, 1]$,

$$
\sup_{|x| > \sqrt{4a\rho^2 \log n}} (T_{\tau}(x) - x)^2 \lesssim \frac{1}{\log n}.
$$

Note that, all of the above arguments hold for any $i$ such that $\theta_{0i} \neq 0$. So, Using the above inequalities, as $n \to \infty$,

$$
\sum_{i: \theta_{0i} \neq 0} E_{\theta_0 i} (\hat{\Theta}_i - \theta_{0i})^2 \lesssim \tilde{p}_n \log n. \quad (T-3.3)
$$

**Step-2** Fix any $i$ such that $\theta_{0i} = 0$.

**Case-1** First, consider $\frac{1}{2} \leq a < 1$. Now,

$$
E_{\theta_0 i} (\hat{\Theta}_i - \theta_{0i})^2 = E_{\theta_0 i} \left( \int_{1/\rho}^{1} T_{\tau}(X_i) \pi_n(\tau|\bm{X}) d\tau \right)^2
$$

$$
\leq E_{\theta_0 i} \left( \int_{1/\rho}^{1} T_{\tau}^2(X_i) \pi_n(\tau|\bm{X}) d\tau \right)
$$

$$
= B_1 + B_2, \text{ Say},
$$

where $B_1 = E_{\theta_0 i} \left( \int_{1/\rho}^{1} T_{\tau}^2(X_i) \pi_n(\tau|\bm{X}) d\tau \right)$. Since, for any fixed $x \in \mathbb{R}, |T_{\tau}(x)| = |x| E(\lambda^2 \tau^2 |x, \tau)$ is
non-decreasing in $\tau$, so, $T^2_\tau (x) \leq T^2_{\frac{\tau}{n}} (x)$ whenever $\tau \in \left[ \frac{1}{n}, \frac{p_n}{n} \right]$.

\[ B_1 \leq \mathbb{E}_{\theta_0} \left( T^2_{\frac{\tau}{n}} (X_i) \int_{\frac{\tau}{n}}^{\frac{p_n}{n}} \pi_n (\tau | X) d\tau \right) \]
\[ \leq \mathbb{E}_{\theta_0} (T^2_{\frac{\tau}{n}} (X_i)) = B_{11} + B_{12}, \text{ Say}. \]

Now, $B_{11} = \mathbb{E}_{\theta_0} (T^2_{\frac{\tau}{n}} (X_i) 1_{\{|X_i| \leq \sqrt{\frac{4a \log(n)}{p_n}}\}})$. Note that, $T^2_{\frac{\tau}{n}} (x) = x E(1 - \kappa |x, \frac{\tau}{p_n})$. Also, for any fixed $x \in \mathbb{R}$ and $p_n = o(n)$, $E(1 - \kappa |x, \frac{\tau}{p_n}) \leq \frac{KM}{a(1 - \alpha)} (\frac{\tau}{p_n})^{2a} e^{\frac{x^2}{2}} (1 + o(1))$. Using the above arguments,

\[ B_{11} \lesssim (\frac{p_n}{n})^{4a} \int_0^{\sqrt{\frac{4a \log(n)}{p_n}}} x^2 \exp(x^2/2) dx \]

Using $x^2 \exp(\frac{x^2}{2}) \leq \frac{d}{dx} [x \exp(\frac{x^2}{2})]$ as $n \to \infty$,

\[ B_{11} \lesssim (\frac{p_n}{n})^{2a} \sqrt{\log(\frac{n}{p_n})}. \quad (T-3.4) \]

For $B_{12} = \mathbb{E}_{\theta_0} (T^2_{\frac{\tau}{n}} (X_i) 1_{\{|X_i| > \sqrt{\frac{4a \log(n)}{p_n}}\}})$, using the fact that for any fixed $\tau$, $|T_\tau (x)| \leq |x| \forall x \in \mathbb{R}$ and $x^2 \phi(x) = \phi(x) - \frac{d}{dx} [x \phi(x)]$,

\[ B_{12} \leq 2 \int_{\sqrt{\frac{4a \log(n)}{p_n}}}^{\infty} x^2 \phi(x) dx \]
\[ \leq 2 \left[ \sqrt{\frac{4a \log(n)}{p_n}} \phi(\sqrt{\frac{4a \log(n)}{p_n}}) + \phi(\sqrt{\frac{4a \log(n)}{p_n}}) \right] \]
\[ \lesssim (\frac{p_n}{n})^{2a} \sqrt{\log(\frac{n}{p_n})}. \quad (T-3.5) \]

Using (T-3.4) and (T-3.5), as $n \to \infty$,

\[ B_1 \lesssim (\frac{p_n}{n})^{2a} \sqrt{\log(\frac{n}{p_n})}. \quad (T-3.6) \]

Note that the term $B_2 = \mathbb{E}_{\theta_0} \left( \int_{\frac{1}{n}}^{\frac{1}{n}} T^2_\tau (X_i) \pi_n (\tau | X) d\tau \right)$ can be split into 3 parts namely $B_{21}, B_{22}$ and $B_{23}$ where $B_{21} = \mathbb{E}_{\theta_0} \left( \int_{\frac{1}{n}}^{\frac{1}{n}} T^2_\tau (X_i) 1_{\{|X_i| \leq \sqrt{\frac{4a \log(n)}{p_n}}\}} \pi_n (\tau | X) d\tau \right)$. Using the same arguments used in $B_{11}, B_{21}$ can be bounded above as,

\[ B_{21} \lesssim t_n^{2a} \sqrt{\log(\frac{1}{t_n})}. \]
Now using $\log(1 + x) \geq \frac{x}{1 + x}$, $x > -1$ with $1 + x = \frac{n}{p_n}$ and $p_n = o(n)$, $\sqrt{\log(t_n)} \leq \sqrt{\log(\frac{n}{p_n})(1 + o(1))}$ and $t_n^{2a} \leq t_n, \forall a \geq \frac{1}{2}$. Using these two facts,

$$B_{21} \lesssim \frac{p_n \log(n)}{p_n}.$$  \hfill (T-3.7)

Next, for $B_{22} = \mathbb{E}_{\theta_0} \left( \int_{5t_n}^1 T^2(\tau) 1_{\{|X| \leq \frac{4a \log(t_n)}{p_n}\}} \pi_n(\tau|X) d\tau \right)$, using the fact that for any fixed $\tau$, $|T_\tau(x)| \leq |x| \forall x \in \mathbb{R}$ along with Cauchy-Schwartz inequality,

$$B_{22} \leq \mathbb{E}_{\theta_0} \left( X_i^2 1_{\{|X_i| \leq \frac{4a \log(t_n)}{p_n}\}} \right) \int_{5t_n}^1 \pi_n(\tau|X) d\tau \leq \sqrt{\mathbb{E}_{\theta_0} \left( X_i^4 1_{\{|X_i| \leq \frac{4a \log(t_n)}{p_n}\}} \right) \mathbb{E}_{\theta_0} \pi_n(\tau \geq 5t_n|X)}$$

Applying Condition (C2) on $\tau$ and using similar arguments used in Lemma 3.6 in Van der Pas et al. (2017) \[23\], $\mathbb{E}_{\theta_0} \pi_n(\tau \geq 5t_n|X) \lesssim \frac{p_n}{n}$. Hence,

$$B_{22} \lesssim \frac{p_n}{n}. \hfill (T-3.8)$$

Finally, using the similar arguments used in $B_{12}$, it is easy to see that

$$B_{23} = \mathbb{E}_{\theta_0} \left( \int_{\frac{1}{p_n}}^1 T^2(\tau) 1_{\{|X| > \frac{4a \log(t_n)}{p_n}\}} \pi_n(\tau|X) d\tau \right)$$

can be bounded above as,

$$B_{23} \lesssim \frac{p_n \log(n)}{p_n}. \hfill (T-3.9)$$

From (T-3.7)-(T-3.9), for sufficiently large $n$,

$$B_2 \lesssim \frac{p_n \log(n)}{p_n}. \hfill (T-3.10)$$

Note that all these preceding arguments hold uniformly in $i$ such that $\theta_0i = 0$. Hence, for any $a \in [0.5, 1)$ as $n \rightarrow \infty$,

$$\sum_{i: \theta_0 i = 0} \mathbb{E}_{\theta_0} (\hat{\theta}_i - \theta_0i)^2 \lesssim (n - \tilde{p}_n) \left[ \left(\frac{p_n}{n}\right)^{2a} \sqrt{\log\left(\frac{n}{p_n}\right)} + \frac{p_n}{n} \log\left(\frac{n}{p_n}\right) \right]$$

\[ \lesssim p_n \log\left(\frac{n}{p_n}\right). \]  \hfill (T-3.11)

The second inequality follows due to the fact that $\tilde{p}_n \leq p_n$ and $p_n = o(n)$ as $n \rightarrow \infty$. 

34
Case-2 Next we consider $a \geq 1$. In this case also, applying exactly the same techniques, $E_{\theta_0}((\hat{\theta}_i - \theta_0)^2$ can be bounded by the sum of $B_1$ and $B_2$ and then $B_1$ can be further bounded by the sum of $B_{11}$ and $B_{12}$, where $B_{11} = E_{\theta_0}(T_{\theta_0}^2(X_i)1_{\{|X_i| \leq \sqrt{2\log(p_n)}\}})$. Again using, $T_{\theta_0}^2(x) = xE(1 - \kappa|x|, \frac{a^2}{b^2})$ and for any fixed $x \in \mathbb{R}$ and for $a \geq 1$, using Lemma 4 and noting that $T_{\theta_0}^2(x) \leq x^2E(1 - \kappa|x|, \frac{a^2}{b^2})$, as $n \rightarrow \infty$,

$$B_{11} \leq E_{\theta_0} \left[ X_i^2E(1 - \kappa|X_i|, \frac{p_n}{n})1_{\{|X_i| \leq \sqrt{2\log(p_n)}\}} \right] \leq \frac{p_n}{n} \int_0^{\sqrt{2\log(p_n)}} x^2 e^{\frac{x^2}{2}} \phi(x) dx$$

$$+ \int_0^{\sqrt{2\log(p_n)}} \frac{t(p_n)^2}{1 + t(p_n)^2} \frac{1}{\sqrt{1 + t(p_n)^2}} e^{-a-1}L(t) \int_0^{\infty} \frac{t^a \cdot \pi(a)^2}{\sqrt{1 + t(p_n)^2}} x^2 \phi(x) dt dx$$

Note that these integrals are of the same form of \((T-1.19)\). Hence using the same arguments as used in that case when $a \geq 1$,

$$B_{11} \approx \frac{p_n}{n} \sqrt{\log\left(\frac{n}{p_n}\right)} . \tag{T-3.12}$$

For $B_{12} = E_{\theta_0}(T_{\theta_0}^2(X_i)1_{\{|X_i| > \sqrt{2\log(p_n)}\}})$, using the fact that for any fixed $\tau$, $|T_{\tau}(x)| \leq |x| \forall x \in \mathbb{R}$ and $x^2\phi(x) = \phi(x) - \frac{d}{dx}[x\phi(x)]$,

$$B_{12} \approx \frac{p_n}{n} \sqrt{\log\left(\frac{n}{p_n}\right)} . \tag{T-3.13}$$

Using \(T-3.12\) and \(T-3.13\), as $n \rightarrow \infty$,

$$B_1 \leq \frac{p_n}{n} \sqrt{\log\left(\frac{n}{p_n}\right)} . \tag{T-3.14}$$

Following the previous steps used in $B_2$ when $a \in \left[\frac{1}{2}, 1\right)$, in this case also, let us split $B_2$ into 3 parts, namely $B_{21}$, $B_{22}$ and $B_{23}$, where $B_{21} = E_{\theta_0} \left( \int_{t_{\theta_0}}^{5t_{\theta_0}} T_{\tau}^2(X_i)1_{\{|X_i| \leq \sqrt{2\log(p_n)}\}} \pi_n(\tau|X) d\tau \right)$, $B_{22} = E_{\theta_0} \left( \int_{5t_{\theta_0}}^{1} T_{\tau}^2(X_i)1_{\{|X_i| \leq \sqrt{2\log(p_n)}\}} \pi_n(\tau|X) d\tau \right)$ and $B_{23} = E_{\theta_0} \left( \int_{1}^{5t_{\theta_0}} T_{\tau}^2(X_i)1_{\{|X_i| > \sqrt{2\log(p_n)}\}} \pi_n(\tau|X) d\tau \right)$. Using similar arguments used when $a \in [0.5, 1)$, we can show that for $a \geq 1$, as $n \rightarrow \infty$,

$$B_{21} \leq \frac{p_n}{n} \log\left(\frac{n}{p_n}\right) . \tag{T-3.15}$$

$$B_{22} \leq \frac{p_n}{n} . \tag{T-3.16}$$
and
\[ B_{23} \lesssim \frac{p_n}{n} \log \left( \frac{n}{p_n} \right). \] (T-3.17)

Combining (T-3.15)-(T-3.17), we can say, for sufficiently large \( n \),
\[ B_2 \lesssim \frac{p_n}{n} \log \left( \frac{n}{p_n} \right). \] (T-3.18)

Since all these preceding arguments hold uniformly in \( i \) such that \( \theta_{0i} = 0 \). Hence, using (T-3.14) and (T-3.18), for any \( a \geq 1 \) as \( n \to \infty \),
\[ \sum_{i: \theta_{0i} = 0} \mathbb{E}_{\theta_{0i}} (\hat{\theta}_i - \theta_{0i})^2 \lesssim (n - \hat{p}_n) \left[ \frac{p_n}{n} n \sqrt{\log \left( \frac{n}{p_n} \right)} + \frac{p_n}{n} \log \left( \frac{n}{p_n} \right) \right] \lesssim p_n \log \left( \frac{n}{p_n} \right). \] (T-3.19)

Hence, with the help of (T-3.11) and (T-3.19), for any \( a \geq \frac{1}{2} \), for sufficiently large \( n \),
\[ \sum_{i: \theta_{0i} = 0} \mathbb{E}_{\theta_{0i}} (\hat{\theta}_i - \theta_{0i})^2 \lesssim p_n \log \left( \frac{n}{p_n} \right). \] (T-3.20)

Hence, using (T-3.1), (T-3.3) and (T-3.20) and taking supremum over all \( \theta_0 \in \ell_{0[p_n]} \), as \( n \to \infty \),
\[ \sup_{\theta_0 \in \ell_{0[p_n]}} \sum_{i=1}^{n} \mathbb{E}_{\theta_{0i}} (\hat{\theta}_i - \theta_{0i})^2 \lesssim p_n \log n. \] (T-3.21)

To get the final form of the result note that the mean square error in (T-3.21) is always bounded below by the corresponding minimax \( l_2 \) risk which is of the order of \( 2p_n \log n \). \( \blacksquare \)

Remark 3. Note that in Theorem 3, in the case of zero means, the truncation based on the absolute values of \( X_i \) plays a pivotal role for establishing that the upper bound of the mean square error is of the order of the near minimax rate, upto some multiplicative constants. For example, when \( a \in \left[ \frac{1}{2}, 1 \right) \) for the term denoted as \( B_{21} \), let us redefine \( B_{21} \) as \( B_{21} = \mathbb{E}_{\theta_0} \left( J_{\pi_n}^{5/2} T^2(X_i) 1_{\{|X_i| \leq \sqrt{4a \log \left( \frac{n}{p_n} \right)} \}} \pi_n (\tau | X) d\tau \right) \), then following exactly same steps, it is easy to see that, \( B_{21} \lesssim \left( \frac{p_n}{n} \log \left( \frac{n}{p_n} \right) \right)^{2a} \sqrt{\log \left( \frac{n}{p_n} \right)} \), which definitely exceeds the minimax rate for \( a \in \left[ \frac{1}{2}, 1 \right) \). Same argument goes for \( a \geq 1 \) too. In other words, to obtain the near minimax rate, the truncation proposed by us based on the absolute values of \( X_i \) is optimal in some sense.
Proof of Theorem 4:

Proof. Note that,

$$\text{Var}(\theta_i|X) = E_{\tau|X}[\text{Var}(\theta_i|X, \tau)] + Var_{\tau|X}[E(\theta_i|X, \tau)]$$

Since, the posterior distribution of $\theta_i$ given $(x, \tau)$ depends on $(x_i, \tau)$ only, hence

$$\text{Var}(\theta_i|X) = E_{\tau|X}[\text{Var}(\theta_i|X_i, \tau)] + Var_{\tau|X}[E(\theta_i|X_i, \tau)]$$

Next we split $E_{\theta_0} \sum^n_{i=1} \text{Var}(\theta_i|X)$ as

$$E_{\theta_0} \sum^n_{i=1} \text{Var}(\theta_i|X) = \sum_{i: \theta_0_i \neq 0} E_{\theta_0_i} \text{Var}(\theta_i|X) + \sum_{i: \theta_0_i = 0} E_{\theta_0_i} \text{Var}(\theta_i|X) . \quad (T-4.1)$$

Using the above argument, these terms can be further split into

$$\sum_{i: \theta_0_i \neq 0} E_{\theta_0_i} \text{Var}(\theta_i|X) = \sum_{i: \theta_0_i \neq 0} E_{\theta_0_i} E_{\tau|X}[\text{Var}(\theta_i|X_i, \tau)] + \sum_{i: \theta_0_i \neq 0} E_{\theta_0_i} Var_{\tau|X}[E(\theta_i|X_i, \tau)] \quad (T-4.2)$$

and

$$\sum_{i: \theta_0_i = 0} E_{\theta_0_i} \text{Var}(\theta_i|X) = \sum_{i: \theta_0_i = 0} E_{\theta_0_i} E_{\tau|X}[\text{Var}(\theta_i|X_i, \tau)] + \sum_{i: \theta_0_i = 0} E_{\theta_0_i} Var_{\tau|X}[E(\theta_i|X_i, \tau)] . \quad (T-4.3)$$

Let us define $\hat{p}_n = \sum^n_{i=1} 1_{\{\theta_0_i \neq 0\}}$. Thus, $\hat{p}_n \leq p_n$.

**Step-1** Fix any $i$ such that $\theta_0_i \neq 0$. Now we split $E_{\theta_0} E_{\tau|X}[\text{Var}(\theta_i|X_i, \tau)]$ as

$$E_{\theta_0} E_{\tau|X}[\text{Var}(\theta_i|X_i, \tau)] = E_{\theta_0} E_{\tau|X}[\text{Var}(\theta_i|X_i, \tau)1_{|X_i| \leq \sqrt{4a\rho^2 \log n}}] + E_{\theta_0} E_{\tau|X}[\text{Var}(\theta_i|X_i, \tau)1_{|X_i| > \sqrt{4a\rho^2 \log n}}] . \quad (T-4.4)$$
Since, for any fixed \(x \in \mathbb{R}\) and \(\tau > 0\), \(\text{Var}(\theta|x, \tau) \leq 1 + x^2\), as \(n \to \infty\),

\[
\mathbb{E}_{\theta_0}E_x[\text{Var}(\theta_i|X_i, \tau)1_{\{|X_i| \leq \sqrt{4a\rho^2 \log n}\}}] = \mathbb{E}_{\theta_0}\left( \int_{1/\tau}^1 \text{Var}(\theta_i|X_i, \tau)1_{\{|X_i| \leq \sqrt{4a\rho^2 \log n}\}} \right) \\
\leq \mathbb{E}_{\theta_0}\left( (1 + X_i^2)1_{\{|X_i| \leq \sqrt{4a\rho^2 \log n}\}} \right) \\
\leq 4a\rho^2 \log n(1 + o(1)). \tag{4.5}
\]

Note that, for any fixed \(x \in \mathbb{R}\), \(x^2 \mathbb{E}(\kappa^2|x, \tau) = x^2 \mathbb{E}(\frac{1}{(1 + x^2 \tau^2)^2}|x, \tau)\) is non-increasing in \(\tau\). Also using Lemma A.1 in Ghosh and Chakrabarti (2017) \[11\], \(\text{Var}(\theta|x, \tau) \leq 1 + x^2 \mathbb{E}(\kappa^2|x, \tau)\). Using these two results along with the same argument used in Step-1 of Theorem \[\text{II}\] for any \(\tau \in \left[\frac{1}{n}, 1\right]\)

\[
\text{Var}(\theta|x, \tau) \leq 1 + x^2 \mathbb{E}(\kappa^2|x, \tau) \\
\leq 1 + x^2 \mathbb{E}(\kappa^2|x, \frac{1}{n}) \\
\leq \hat{h}(x, \frac{1}{n}).
\]

Hence, using the same arguments used in Step-1 of Theorem \[\text{II}\]

\[
\sup_{|x| > \sqrt{4a\rho^2 \log n}} \text{Var}(\theta|x, \tau) \leq \sup_{|x| > \sqrt{4a\rho^2 \log n}} \hat{h}(x, \frac{1}{n}) \leq 1 + \sup_{|x| > \sqrt{4a\rho^2 \log n}} \hat{h}_1(x, \frac{1}{n}) + \sup_{|x| > \sqrt{4a\rho^2 \log n}} \hat{h}_2(x, \frac{1}{n}) \lesssim 1.
\]

Using above arguments, as \(n \to \infty\),

\[
\mathbb{E}_{\theta_0}E_{\tau}[\text{Var}(\theta_i|X_i, \tau)1_{\{|X_i| > \sqrt{4a\rho^2 \log n}\}}] = \mathbb{E}_{\theta_0}\left( \int_{1/\tau}^1 \text{Var}(\theta_i|X_i, \tau)1_{\{|X_i| > \sqrt{4a\rho^2 \log n}\}} \pi_n(\tau|X)d\tau \right) \lesssim 1. \tag{4.6}
\]

Combining (4.4), (4.5) and (4.6), we obtain for sufficiently large \(n\),

\[
\mathbb{E}_{\theta_0}E_{\tau}[\text{Var}(\theta_i|X_i, \tau)] \lesssim \log n.
\]

38
Since all of the above arguments hold true for any \( i \) such that \( \theta_{0i} \neq 0 \), as \( n \to \infty \),

\[
\sum_{i: \theta_{0i} \neq 0} \mathbb{E}_{\theta_{0i}} \mathbb{E}_{\tau_i | X} [\text{Var}(\theta_i | X_i, \tau)] \lesssim \tilde{p}_n \log n . \tag{T-4.7}
\]

Now, for the second term in (T-4.2), note that, \( \text{Var}_{\tau_i | X} [\mathbb{E}(\theta_i | X_i, \tau)] = \mathbb{E}_{\tau_i | X} (T_{\tau_i}(X_i) - \hat{\theta}_i)^2 \) and using the definition of \( \hat{\theta}_i \),

\[
\mathbb{E}_{\tau_i | X}(T_{\tau_i}(X_i) - \hat{\theta}_i)^2 = \mathbb{E}_{\tau_i | X}[(T_{\tau_i}(X_i) - X_i) + (X_i - \hat{\theta}_i)]^2 \\
\leq 2 \mathbb{E}_{\tau_i | X}(T_{\tau_i}(X_i) - X_i)^2 + 2(\hat{\theta}_i - X_i)^2
\]

Using these facts,

\[
\mathbb{E}_{\theta_{0i}} \text{Var}_{\tau_i | X} [\mathbb{E}(\theta_i | X_i, \tau)] \leq 2 \mathbb{E}_{\theta_{0i}} \mathbb{E}_{\tau_i | X} (T_{\tau_i}(X_i) - X_i)^2 + 2 \mathbb{E}_{\theta_{0i}} (\hat{\theta}_i - X_i)^2 \tag{T-4.8}
\]

Using the same arguments used in Step-1 in Theorem 3 as \( n \to \infty \)

\[
\mathbb{E}_{\theta_{0i}} (\hat{\theta}_i - X_i)^2 \leq 4 a \rho^2 \log n \tag{T-4.9}
\]

For the first term,

\[
\mathbb{E}_{\theta_{0i}} \mathbb{E}_{\tau_i | X} (T_{\tau_i}(X_i) - X_i)^2 = \mathbb{E}_{\theta_{0i}} \left( \int_{\tau} (T_{\tau_i}(X_i) - X_i)^2 \pi_n(\tau | X) d\tau \right) \lesssim \log n, \tag{T-4.10}
\]

where the inequality follows from the same arguments mentioned above. On combining (T-4.8), (T-4.9) and (T-4.10), for sufficiently large \( n \)

\[
\mathbb{E}_{\theta_{0i}} \text{Var}_{\tau_i | X} [\mathbb{E}(\theta_i | X_i, \tau)] \lesssim \log n .
\]

Again noting that all of the above arguments hold true for any \( i \) such that \( \theta_{0i} \neq 0 \), as \( n \to \infty \),

\[
\sum_{i: \theta_{0i} \neq 0} \mathbb{E}_{\theta_{0i}} \text{Var}_{\tau_i | X} [\mathbb{E}(\theta_i | X_i, \tau)] \lesssim \tilde{p}_n \log n . \tag{T-4.11}
\]
Using (T-4.2), (T-4.7) and (T-4.11), as \( n \to \infty \)

\[
\sum_{i: \theta_{0i} \neq 0} \mathbb{E}_{\theta_{0i}} \text{Var}(\theta_i|X) \lesssim \tilde{p}_n \log n .
\]  

(T-4.12)

**Step-2** Fix any \( i \) such that \( \theta_{0i} = 0 \). Now we split \( \mathbb{E}_{\theta_{0i}} \mathbb{E}_{\tau|X}[\text{Var}(\theta_i|X_i, \tau)] \) as

\[
\mathbb{E}_{\theta_{0i}} \mathbb{E}_{\tau|X}[\text{Var}(\theta_i|X_i, \tau)] = \mathbb{E}_{\theta_{0i}} \left[ \int_{\frac{1}{\tau}}^{1/\tau} \text{Var}(\theta_i|X_i, \tau) \pi_n(\tau|X) d\tau \right]
\]

\[
= \mathbb{E}_{\theta_{0i}} \left[ \int_{\frac{1}{\tau}}^{1} \text{Var}(\theta_i|X_i, \tau) \pi_n(\tau|X) d\tau \right] + \mathbb{E}_{\theta_{0i}} \left[ \int_{1}^{1/\tau} \text{Var}(\theta_i|X_i, \tau) \pi_n(\tau|X) d\tau \right]
\]

\[
\leq \mathbb{E}_{\theta_{0i}} \left[ \int_{\frac{1}{\tau}}^{1} \text{Var}(\theta_i|X_i, \tau) \pi_n(\tau|X) d\tau \right] + \mathbb{E}_{\theta_{0i}} \left[ (1 + X_i^2) \int_{1/\tau}^{1} \pi_n(\tau|X) d\tau \right], \quad (T-4.13)
\]

where the inequality in the last term follows due to the fact that \( \text{Var}(\theta|x, \tau) \leq 1 + x^2 \) for any \( x \in \mathbb{R} \) and any \( \tau > 0 \).

**Case-1** First we consider the case when \( a \in [\frac{1}{2}, 1) \). We again decompose \( \mathbb{E}_{\theta_{0i}} \left[ \int_{\frac{1}{\tau}}^{1} \text{Var}(\theta_i|X_i, \tau) \pi_n(\tau|X) d\tau \right] \) as

\[
\mathbb{E}_{\theta_{0i}} \left[ \int_{\frac{1}{\tau}}^{1} \text{Var}(\theta_i|X_i, \tau) \pi_n(\tau|X) d\tau \right] = \mathbb{E}_{\theta_{0i}} \left[ \int_{\frac{1}{\tau}}^{1} \text{Var}(\theta_i|X_i, \tau) 1_{|X_i| \leq \sqrt{4a \log(\frac{1}{\tau m})}} \pi_n(\tau|X) d\tau \right] +
\]

\[
\mathbb{E}_{\theta_{0i}} \left[ \int_{\frac{1}{\tau}}^{1} \text{Var}(\theta_i|X_i, \tau) 1_{|X_i| > \sqrt{4a \log(\frac{1}{\tau m})}} \pi_n(\tau|X) d\tau \right] \quad (T-4.14)
\]

Since for any fixed \( x \in \mathbb{R} \) and \( \tau > 0 \), \( \text{Var}(\theta|x, \tau) \leq E(1 - \kappa|x, \tau) + J(x, \tau) \) and \( E(1 - \kappa|x, \tau) \) is non-decreasing in \( \tau \) and using Lemma 2 of Ghosh and Chakrabarti (2017) [11] and Lemma A.2 of Ghosh and Chakrabarti (2017) [11], for \( a \in [0.5, 1) \), for any fixed \( x \in \mathbb{R} \) and \( \tau > 0 \),

\[
\mathbb{E}_{\theta_{0i}} \left[ \int_{\frac{1}{\tau}}^{1} \text{Var}(\theta_i|X_i, \tau) 1_{|X_i| \leq \sqrt{4a \log(\frac{1}{\tau m})}} \pi_n(\tau|X) d\tau \right] \leq \tilde{p}_n 2^a \int_{0}^{\sqrt{4a \log(\frac{1}{\tau m})}} e^{-\frac{x^2}{2} \phi(x)} dx
\]

\[
\lesssim \frac{p_n \log(n)}{p_n}, \quad (T-4.15)
\]

where inequality in the last step follows using the same argument used for providing upper bound to the term \( B_{21} \) in **Case-1** of Theorem 3 when \( a \in [0.5, 1) \). Using \( \text{Var}(\theta|x, \tau) \leq 1 + x^2 \) and the identity

40
On combining (T-4.14)-(T-4.16), for sufficiently large $n$,

\[ E_{\theta_0}\left[ \int_{\frac{1}{5}}^{5t_n} Var(\theta_i|X_i, \tau) \pi_n(\tau|X)d\tau \right] \lesssim \frac{p_n}{n} \log \left( \frac{n}{p_n} \right). \quad (T-4.17) \]

Next we split $E_{\theta_0}\left[ (1 + X_i^2) \int_{\frac{1}{5}t_n}^{1} \pi_n(\tau|X)d\tau \right]$ as

\[ E_{\theta_0}\left[ (1 + X_i^2) \int_{\frac{1}{5}t_n}^{1} \pi_n(\tau|X)d\tau \right] = E_{\theta_0}\left[ (1 + X_i^2)^1 \{ |X_i| > \sqrt{4a \log(\frac{1}{t_n})} \} \int_{\frac{1}{5}t_n}^{1} \pi_n(\tau|X)d\tau \right] + \]

\[ E_{\theta_0}\left[ (1 + X_i^2)^1 \{ |X_i| \leq \sqrt{4a \log(\frac{1}{t_n})} \} \int_{\frac{1}{5}t_n}^{1} \pi_n(\tau|X)d\tau \right]. \quad (T-4.18) \]

Note that

\[ E_{\theta_0}\left[ (1 + X_i^2)^1 \{ |X_i| > \sqrt{4a \log(\frac{1}{t_n})} \} \int_{\frac{1}{5}t_n}^{1} \pi_n(\tau|X)d\tau \right] \leq 2 \int_{\sqrt{4a \log(\frac{1}{t_n})}}^{\infty} (1 + x^2) \phi(x) dx \]

\[ \lesssim t_n 2a \sqrt{\log \frac{1}{t_n}} \]

\[ \lesssim \frac{p_n}{n} \cdot \log \frac{n}{p_n}. \quad (T-4.19) \]

Applying Cauchy-Schwartz inequality,

\[ E_{\theta_0}\left[ (1 + X_i^2)^1 \{ |X_i| \leq \sqrt{4a \log(\frac{1}{t_n})} \} \int_{\frac{1}{5}t_n}^{1} \pi_n(\tau|X)d\tau \right] \leq \sqrt{E_{\theta_0}\left[ (1 + X_i^2)^2 \right]} \sqrt{E_{\theta_0}\pi_n(\tau \geq 5t_n|X)} \]

Applying Condition \[ (C2) \] on $\tau$ and using the same arguments used in $B_{22}$ in Step-2 of Theorem for $a \in \left[ \frac{1}{2}, 1 \right]$, $E_{\theta_0}\pi_n(\tau \geq 5t_n|X) \lesssim \frac{p_n}{n}$, which ensures that, as $n \to \infty$

\[ E_{\theta_0}\left[ (1 + X_i^2)^1 \{ |X_i| \leq \sqrt{4a \log(\frac{1}{t_n})} \} \int_{\frac{1}{5}t_n}^{1} \pi_n(\tau|X)d\tau \right] \lesssim \frac{p_n}{n}. \quad (T-4.20) \]
Using (T-4.18)-(T-4.20), for sufficiently large \( n \)
\[
E_{\theta_0} \left[ (1 + X_i^2) \int_{5t_n}^1 \pi_n(\tau|X) d\tau \right] \lesssim \frac{p_n}{n} \log \frac{n}{p_n} .
\] (T-4.21)

Combining (T-4.13), (T-4.17) and (T-4.21), as \( n \to \infty \)
\[
E_{\theta_0} E_{\tau|X} [Var(\theta_i|X_i, \tau)] \lesssim \frac{p_n}{n} \log \left( \frac{n}{p_n} \right) .
\]

Note that all these preceding arguments hold uniformly in \( i \) such that \( \theta_{0i} = 0 \). Hence, for any \( a \in [0.5, 1) \) as \( n \to \infty \),
\[
\sum_{i: \theta_{0i} = 0} E_{\theta_0} E_{\tau|X} [Var(\theta_i|X_i, \tau)] \lesssim (n - \hat{p}_n) \frac{p_n}{n} \log \left( \frac{n}{p_n} \right)
\]
\[
\lesssim p_n \log \left( \frac{n}{p_n} \right) .
\] (T-4.22)

The second inequality follows due to the fact that \( \hat{p}_n \leq p_n \) and \( p_n = o(n) \) as \( n \to \infty \).

**Case-2** Next we consider the case when \( a \geq 1 \). We now decompose \( E_{\theta_0} \left[ \int_{\frac{1}{5}}^{5t_n} Var(\theta_i|X_i, \tau) \pi_n(\tau|X) d\tau \right] \) as
\[
E_{\theta_0} \left[ \int_{\frac{1}{5}}^{5t_n} Var(\theta_i|X_i, \tau) \pi_n(\tau|X) d\tau \right] = E_{\theta_0} \left[ \int_{\frac{1}{5}}^{5t_n} Var(\theta_i|X_i, \tau) 1_{\{|X_i| \leq \sqrt{2 \log \left( \frac{n}{p_n} \right)} \}} \pi_n(\tau|X) d\tau \right] +
E_{\theta_0} \left[ \int_{\frac{1}{5}}^{5t_n} Var(\theta_i|X_i, \tau) 1_{\{|X_i| > \sqrt{2 \log \left( \frac{n}{p_n} \right)} \}} \pi_n(\tau|X) d\tau \right]
\] (T-4.23)

Since for any fixed \( x \in \mathbb{R} \) and \( \tau > 0 \), \( Var(\theta|x, \tau) \leq E(1 - \kappa|x, \tau)1_{\{|x| \leq 1\}} + 2x^2 E(1 - \kappa|x, \tau) \) (Obtained from (T-1.16)) and \( E(1 - \kappa|x, \tau) \) is non-decreasing in \( \tau \), so,
\[
E_{\theta_0} \left[ \int_{\frac{1}{5}}^{5t_n} Var(\theta_i|X_i, \tau) 1_{\{|X_i| \leq \sqrt{2 \log \left( \frac{n}{p_n} \right)} \}} \pi_n(\tau|X) d\tau \right] \leq E_{\theta_0} \left[ Var(\theta_i|X_i, 5t_n) 1_{\{|X_i| \leq \sqrt{2 \log \left( \frac{n}{p_n} \right)} \}} \right]
\]
\[
\leq E_{\theta_0} \left[ E(1 - \kappa|X_i, 5t_n) 1_{\{|X_i| \leq 1\}} \right] + 2E_{\theta_0} \left[ X_i^2 E(1 - \kappa|X_i, 5t_n) 1_{\{|X_i| \leq \sqrt{2 \log \left( \frac{n}{p_n} \right)} \}} \right]
\]

Note that the terms involved in the above equation is of the similar form of (T-1.17) and hence applying the
same arguments as used before, for sufficiently large $n$,

$$
\mathbb{E}_{\theta_0} \left[ \int_{\frac{1}{n}}^{5t_n} \text{Var}(\theta_i|X_i, \tau) 1_{\{|X_i| \leq \sqrt{2 \log \left( \frac{1}{tn} \right)}\}} \pi_n(\tau|X) d\tau \right] \lesssim t_n \sqrt{\log \left( \frac{1}{tn} \right)}. \tag{T-4.24}
$$

Using $\text{Var}(\theta|x, \tau) \leq 1 + x^2$ and the identity $x^2 \phi(x) = \phi(x) - \frac{d}{dx}[x \phi(x)]$, we obtain, as $n \to \infty$,

$$
\mathbb{E}_{\theta_0} \left[ \int_{\frac{1}{n}}^{5t_n} \text{Var}(\theta_i|X_i, \tau) 1_{\{|X_i| > \sqrt{2 \log \left( \frac{1}{tn} \right)}\}} \pi_n(\tau|X) d\tau \right] \lesssim t_n \sqrt{\log \left( \frac{1}{tn} \right)}. \tag{T-4.25}
$$

On combining (T-4.23)-(T-4.25), for sufficiently large $n$,

$$
\mathbb{E}_{\theta_0} \left[ \int_{\frac{1}{n}}^{5t_n} \text{Var}(\theta_i|X_i, \tau) \pi_n(\tau|X) d\tau \right] \lesssim t_n \sqrt{\log \left( \frac{1}{tn} \right)}. \tag{T-4.26}
$$

Now to get the upper bound on the second term in (T-4.13) for $a \geq 1$, applying the same arguments when $a \in \left[ \frac{1}{2}, 1 \right)$, we can show, as $n \to \infty$,

$$
\mathbb{E}_{\theta_0} \left[ (1 + X_i^2) 1_{\{|X_i| > \sqrt{2 \log \left( \frac{1}{tn} \right)}\}} \int_{\frac{1}{n}}^{1} \pi_n(\tau|X) d\tau \right] \leq 2 \int_{\sqrt{2 \log \left( \frac{1}{tn} \right)}}^{\infty} (1 + x^2) \phi(x) dx \\
\approx t_n \sqrt{\log \left( \frac{1}{tn} \right)} \\
\approx \frac{p_n}{n} \cdot \log \frac{n}{p_n}. \tag{T-4.27}
$$

and

$$
\mathbb{E}_{\theta_0} \left[ (1 + X_i^2) 1_{\{|X_i| \leq \sqrt{2 \log \left( \frac{1}{tn} \right)}\}} \int_{5t_n}^{1} \pi_n(\tau|X) d\tau \right] \approx \frac{p_n}{n}. \tag{T-4.28}
$$

As a result of (T-4.27) and (T-4.28), for sufficiently large $n$,

$$
\mathbb{E}_{\theta_0} \left[ (1 + X_i^2) \int_{5t_n}^{1} \pi_n(\tau|X) d\tau \right] \lesssim \frac{p_n}{n} \cdot \log \frac{n}{p_n}. \tag{T-4.29}
$$

With the help of (T-4.13), (T-4.26) and (T-4.29), for $a \geq 1$,

$$
\mathbb{E}_{\theta_0} E_{\tau|X} [\text{Var}(\theta_i|X_i, \tau)] \lesssim \frac{p_n}{n} \cdot \log \frac{n}{p_n}. \tag{T-4.30}
$$
Note that all these preceding arguments hold uniformly in i such that $\theta_{0i} = 0$. Hence, for any $a \geq 1$ as $n \to \infty$,

$$
\sum_{i: \theta_{0i} = 0} \mathbb{E}_{\theta_{0i}} E_{\tau}[\text{Var}(\theta_i|X_i, \tau)] \lesssim (n - \tilde{p}_n) \frac{p_n}{n} \log \left( \frac{n}{p_n} \right)
$$

$$
\lesssim p_n \log \left( \frac{n}{p_n} \right)
$$

(T-4.31)

With the use of (T-4.32) and (T-4.31), for $a \geq \frac{1}{2}$,

$$
\sum_{i: \theta_{0i} = 0} \mathbb{E}_{\theta_{0i}} E_{\tau}[\text{Var}(\theta_i|X_i, \tau)] \lesssim p_n \log \left( \frac{n}{p_n} \right)
$$

(T-4.32)

For the second term in (T-4.3), note that

$$
\mathbb{E}_{\theta_{0i}} \text{Var}_{\tau}[\mathbb{E}(\theta_i|X_i, \tau)] \leq \mathbb{E}_{\theta_{0i}} \text{Var}_{\tau}\mathbb{X} T^2_\tau(X_i) = \mathbb{E}_{\theta_{0i}} \left( \int_{1/n}^1 T^2_\tau(X_i) \pi_n(\tau|\mathbb{X}) d\tau \right)
$$

Now applying the same argument used in Step-2 in Theorem 3 when $a \geq 0.5$, for sufficiently large $n$,

$$
\mathbb{E}_{\theta_{0i}} \left( \int_{1/n}^1 T^2_\tau(X_i) \pi_n(\tau|\mathbb{X}) d\tau \right) \lesssim \frac{\tilde{p}_n}{n} \log \left( \frac{\tilde{p}_n}{n} \right)
$$

Again noting that this argument holds uniformly in i such that $\theta_{0i} = 0$. Hence, for any $a \geq 0.5$ as $n \to \infty$,

$$
\sum_{i: \theta_{0i} = 0} \mathbb{E}_{\theta_{0i}} \text{Var}_{\tau}[\mathbb{E}(\theta_i|X_i, \tau)] \lesssim (n - \tilde{p}_n) \frac{p_n}{n} \log \left( \frac{n}{p_n} \right)
$$

$$
\lesssim p_n \log \left( \frac{n}{p_n} \right)
$$

(T-4.33)

With the use of (T-4.3), (T-4.32) and (T-4.33), as $n \to \infty$,

$$
\sum_{i: \theta_{0i} = 0} \mathbb{E}_{\theta_{0i}} \text{Var}(\theta_i|\mathbb{X}) \lesssim p_n \log \left( \frac{n}{p_n} \right)
$$

(T-4.34)

Finally on combining (T-4.1), (T-4.12) and (T-4.34) and taking supremum over all $\theta_0 \in l_0[p_n]$, for sufficiently large $n$

$$
\mathbb{E}_{\theta_0} \sum_{i=1}^n \text{Var}(\theta_i|\mathbb{X}) \lesssim p_n \log(n)
$$

Proof of Theorem 6.
Proof. Let us define \( \tilde{p}_n = \sum_{i=1}^{n} 1_{\{\theta_{0i} \neq 0\}} \). In this case also, we will follow the same steps mentioned in previous theorem. Now, we split the mean square error as

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

(T-6.1)

Step-1 Fix any \( i \) such that \( \theta_{0i} \neq 0 \). Using \( \mathbb{E}_{\theta_{0i}} (X_i - \theta_{0i})^2 = 1 \), we get,

\[
\mathbb{E}_{\theta_{0i}} (T_{\hat{\tau}}(X_i) - \theta_{0i})^2 \leq 2 \left[ \mathbb{E}_{\theta_{0i}} (T_{\hat{\tau}}(X_i) - X_i)^2 + 1 \right] .
\]  

(T-6.2)

Since, for any \( \tau > 0 \) and \( x \in \mathbb{R} \), \( |T_{\hat{\tau}}(x) - x| \leq |x| \), hence, \( |T_{\hat{\tau}}(x) - x| \leq \sqrt{2\rho \log n} \) whenever \( |x| \leq \sqrt{2\rho \log n} \), which implies

\[
\mathbb{E}_{\theta_{0i}} \left[ (T_{\hat{\tau}}(X_i) - X_i)^2 1_{\{|X_i| \leq \sqrt{2\rho \log n}\}} \right] \leq 2\rho \log n .
\]  

(T-6.3)

Now, using (2) of Lemma 2, we have, for \( \hat{\tau} \geq \frac{1}{n} \),

\[
(T_{\hat{\tau}}(x) - x)^2 \leq g^2(x, \frac{1}{n})
\]

and using the definition along with the decreasing property of \( g(\cdot, \cdot) \) obtained from Lemma 2 we have the following for sufficiently large \( n \),

\[
\sup_{|x| > \sqrt{2\rho \log n}} g(x, \frac{1}{n}) \leq g(\sqrt{2\rho \log n}, \frac{1}{n}) \lesssim \frac{1}{\sqrt{\log n}} + \sqrt{\log n \cdot n^{-2(\rho - 1)}} .
\]

Since, \( \rho > c \), we get as \( n \to \infty \)

\[
\mathbb{E}_{\theta_{0i}} \left[ (T_{\hat{\tau}}(X_i) - X_i)^2 1_{\{|X_i| \leq \sqrt{2\rho \log n}\}} \right] \lesssim \frac{1}{\log n} .
\]  

(T-6.4)

Using (T-6.2)-(T-6.4), we obtain

\[
\mathbb{E}_{\theta_{0i}} (T_{\hat{\tau}}(X_i) - \theta_{0i})^2 \lesssim \log n .
\]  

(T-6.5)

Noting that all above arguments are independent of any \( \theta_{0i} \) such that \( \theta_{0i} \neq 0 \), we get as \( n \to \infty \)

\[
\sum_{i: \theta_{0i} \neq 0} \mathbb{E}_{\theta_{0i}} (T_{\hat{\tau}}(X_i) - \theta_{0i})^2 \lesssim \tilde{p}_n \log n .
\]  

(T-6.6)
Step-2 Fix any \( i \) such that \( \theta_{0i} = 0 \). Choose \( \gamma > 1 \) such that \( c_2 \gamma - 1 > 1 \). Note that

\[
E_{\theta_{0i}}[T_\tau^2(X_i)] = E_{\theta_{0i}}[T_\tau^2(X_i)1\{\hat{\tau} \leq \gamma \frac{pn}{n}\}] + E_{\theta_{0i}}[T_\tau^2(X_i)1\{\hat{\tau} > \gamma \frac{pn}{n}\}] .
\] (T-6.7)

Since, \( T_\tau(x) = xE(1 - \kappa|x, \tau) \) and using (2) of Lemma 2 we get the following

\[
E_{\theta_{0i}}[T_\tau^2(X_i)1\{\hat{\tau} \leq \gamma \frac{pn}{n}\}] \leq E_{\theta_{0i}}[T_\tau^2(X_i)] \leq 2\gamma^2 \frac{pn}{n} \int_0^\infty x^2 \phi(x)dx .
\]

Hence, for sufficiently large \( n \),

\[
E_{\theta_{0i}}[T_\tau^2(X_i)1\{\hat{\tau} \leq \gamma \frac{pn}{n}\}] \lesssim \frac{pn}{n} .
\] (T-6.8)

Now, using exactly the same arguments as used in the proof of Theorem 2 of Ghosh and Chakrabarti (2017) [11] for Zero means, we obtain

\[
E_{\theta_{0i}}[T_\tau^2(X_i)1\{\hat{\tau} > \gamma \frac{pn}{n}\}1\{|X_i| > \sqrt{c_1 \log(pn)}\}] \lesssim \left(\frac{pn}{n}\right)^{\frac{1}{4}} \sqrt{\log\left(\frac{n}{pn}\right)} .
\] (T-6.9)

and

\[
E_{\theta_{0i}}[T_\tau^2(X_i)1\{\hat{\tau} > \gamma \frac{pn}{n}\}1\{|X_i| \leq \sqrt{c_1 \log(pn)}\}] \lesssim \frac{pn}{n} \log\left(\frac{n}{pn}\right) .
\] (T-6.10)

Combining (T-6.7)-(T-6.10), for sufficiently large \( n \)

\[
E_{\theta_{0i}}[T_\tau^2(X_i)] \lesssim \frac{pn}{n} \log\left(\frac{n}{pn}\right) .
\] (T-6.11)

Observing that the above arguments go through for any \( i \) such that \( \theta_{0i} = 0 \), as \( n \to \infty \)

\[
\sum_{i:\theta_{0i}=0} E_{\theta_{0i}}(T_\tau(X_i) - \theta_{0i})^2 \lesssim p_n \log\left(\frac{n}{pn}\right) .
\] (T-6.12)

Using (T-6.1), (T-6.6) and (T-6.12) for sufficiently large \( n \)

\[
\sum_{i=1}^n E_{\theta_{0i}}(T_\tau(X_i) - \theta_{0i})^2 \lesssim p_n \log n .
\]

The final result is obtained first taking supremum over all \( \theta_0 \in l_0[p_n] \) and then using the same reasoning used in the end of Theorem 3. ■
Proof of Theorem 7.

Proof. Using exactly the same techniques used in Theorem 1, we have

\[ \mathbb{E}_{\theta_0} \sum_{i=1}^{n} \text{Var}(\theta_i|X_i, \hat{\tau}) = \sum_{i: \theta_0i \neq 0} \mathbb{E}_{\theta_0} \text{Var}(\theta_i|X_i, \hat{\tau}) + \sum_{i: \theta_0i = 0} \mathbb{E}_{\theta_0i} \text{Var}(\theta_i|X_i, \hat{\tau}) . \]  

(T-7.1)

Step-1 Fix any \( i \) such that \( \theta_0i \neq 0 \). Again using \( \text{Var}(\theta|x, \tau) \leq 1 + x^2 \) for any \( x \in \mathbb{R} \) and any \( \tau > 0 \), we have

\[ \mathbb{E}_{\theta_0i} [\text{Var}(\theta_i|X_i, \hat{\tau}) 1_{\{|X_i| \leq \sqrt{2\rho^2 \log n}\}}] \leq 2\rho^2 \log n (1 + o(1)) . \]  

(T-7.2)

Using (3) of Lemma 2 and by the definition of \( \hat{\tau} \) obtained from (3.1),

\[ \text{Var}(\theta|x, \hat{\tau}) \leq \tilde{g}(x, \frac{1}{n}) \]

and following exactly the same lines as Step-1 of Theorem 1 with (3) of Lemma 2 implies that for any \( \rho > C_1 \), for sufficiently large \( n \)

\[ \sup_{|x| > \sqrt{2\rho^2 \log n}} \tilde{g}(x, \frac{1}{n}) \lesssim 1 . \]

Using above arguments, as \( n \to \infty \),

\[ \mathbb{E}_{\theta_0i} [\text{Var}(\theta_i|X_i, \hat{\tau}) 1_{\{|X_i| \leq \sqrt{2\rho^2 \log n}\}}] \lesssim 1 . \]  

(T-7.3)

Using (T-7.2) and (T-7.3) and noting that the above arguments hold true for any \( i \) such that \( \theta_0i \neq 0 \), for sufficiently large \( n \)

\[ \sum_{i: \theta_0i \neq 0} \mathbb{E}_{\theta_0i} \text{Var}(\theta_i|X_i, \hat{\tau}) \lesssim \tilde{p}_n \log n . \]  

(T-7.4)

Step-2 Fix any \( i \) such that \( \theta_0i = 0 \). Choose \( \gamma > 1 \) such that \( c_2\gamma - 1 > 1 \). Now employing exactly the same reasoning as used in Case-1 of Step-2 of Theorem 1, we have the following for sufficiently large \( n \)

\[ \mathbb{E}_{\theta_0i} [\text{Var}(\theta_i|X_i, \hat{\tau}) 1_{\{|X_i| \geq \sqrt{c_1 \log n}\}}] \lesssim \sqrt{\log n} \cdot n^{-\frac{1}{\sqrt{\gamma}}} . \]  

(T-7.5)

and

\[ \mathbb{E}_{\theta_0i} [\text{Var}(\theta_i|X_i, \hat{\tau}) 1_{\{|X_i| \geq \sqrt{c_1 \log n}\}} 1_{\{|\hat{\tau} - \rho^2 n^{-\frac{1}{2}}\} \leq \frac{\rho^2 n}{n} \log n .} \]  

(T-7.6)
Also, using the fact, $\text{Var}(\theta|x, \tau) \leq (1 + x^2)E(1 - \kappa|x, \tau)$ and with the use of [1] of Lemma 2,

$$E_{\theta_0} [\text{Var}(\theta_i|X_i, \hat{\tau}) 1_{\{\hat{\tau} \leq \gamma \log n\}} 1_{\{|X_i| > c_1 \log n\}}] \leq 2\gamma^2 \frac{p_n}{n} .$$  \hspace{1cm} (T-7.7)

Using (T-7.5)-(T-7.7) and noting that the above arguments hold true for any $i$ such that $\theta_{0i} = 0$, for sufficiently large $n$

$$\sum_{i: \theta_{0i} = 0} E_{\theta_0} \text{Var}(\theta_i|X_i, \hat{\tau}) \lesssim p_n \log n .$$  \hspace{1cm} (T-7.8)

Combining (T-7.1), (T-7.4) and (T-7.8) and taking supremum over all $\theta_0 \in l_0[p_n]$, we get the desired result. ■

**References**

[1] Bhadra, Anindya and Datta, Jyotishka and Polson, Nicholas G and Willard, Brandon (2017). The horseshoe+ estimator of ultra-sparse signals. *Bayesian Analysis*, 12(4), 1105–1131.

[2] Bickel, Peter J and Ritov, Ya’acov and Tsybakov, Alexandre B (2009). Simultaneous analysis of Lasso and Dantzig selector. *The Annals of statistics*, 37(4), 1705–1732.

[3] Bogdan, Malgorzata and Chakrabarti, Arijit and Frommlet, Florian and Ghosh, Jayanta K (2011). Asymptotic Bayes-optimality under sparsity of some multiple testing procedures. *The Annals of Statistics*, 39(3), 1551–1579.

[4] Brown, Philip J and Griffin, Jim E (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, 5(1), 171–188.

[5] Caron, François and Doucet, Arnaud (2008). Sparse Bayesian nonparametric regression. *Proceedings of the 25th international conference on Machine learning*, 88–95.

[6] Carvalho, Carlos M and Polson, Nicholas G and Scott, James G (2009). Handling sparsity via the horseshoe. *Artificial Intelligence and Statistics*, 73–80.

[7] Carvalho, Carlos M and Polson, Nicholas G and Scott, James G (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2), 465–480.
[8] Datta, Jyotishka and Ghosh, Jayanta K (2013). Asymptotic properties of Bayes risk for the horseshoe prior. *Bayesian Analysis*, 8(1), 111–132.

[9] Donoho, D. L., Johnstone, I. M. and Hoch, J. C. and Stern, A. S. (1992). Maximum entropy and the nearly black object. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(1), 41–67.

[10] Efron, Bradley (2004). Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *Journal of the american statistical association*, 99(465), 96–104.

[11] Ghosh, Prasenjit and Chakrabarti, Arijit (2017). Asymptotic optimality of one-group shrinkage priors in sparse high-dimensional problems. *Bayesian Analysis*, 12(4), 1133–1161.

[12] Ghosh, Prasenjit and Tang, Xueying and Ghosh, Malay and Chakrabarti, Arijit (2016). Asymptotic properties of Bayes risk of a general class of shrinkage priors in multiple hypothesis testing under sparsity. *Bayesian Analysis*, 11(3), 753–796.

[13] Ghosal, Subhashis and Ghosh, Jayanta K and Van Der Vaart, Aad W (2000). Convergence rates of posterior distributions. *The Annals of statistics*, 500–531.

[14] Jiang, Wenhua and Zhang, Cun-Hui (2009). General maximum likelihood empirical Bayes estimation of normal means. *The Annals of Statistics*, 37(4), 1647–1684.

[15] Johnstone, Iain M and Silverman, Bernard W (2004). Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences. *The Annals of Statistics*, 32(4), 1594–1649.

[16] Mitchell, Toby J and Beauchamp, John J (1988). Bayesian variable selection in linear regression. *Journal of the american statistical association*, 83(404), 1023–1032.

[17] Park, Trevor and Casella, George (2008). The bayesian lasso. *Journal of the american statistical association*, 103(482), 681–686.

[18] Polson, Nicholas G and Scott, James G (2010). Shrink globally, act locally: Sparse Bayesian regularization and prediction. *Bayesian statistics*, 9, 501-538.

[19] Ročková, Veronika (2018). Bayesian estimation of sparse signals with a continuous spike-and-slab prior. *The Annals of Statistics*, 46(1), 401–437.
[20] Tipping, Michael E (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of machine learning research, 1*(Jun), 211–244.

[21] Van Der Pas, Stéphanie L and Kleijn, Bas JK and Van Der Vaart, Aad W (2014). The horseshoe estimator: Posterior concentration around nearly black vectors. *Electronic Journal of Statistics, 8*(2), 2585–2618.

[22] Van Der Pas, SL and Salomond, J-B and Schmidt-Hieber, Johannes (2016). Conditions for posterior contraction in the sparse normal means problem. *Electronic journal of statistics, 10*(1), 976–1000.

[23] Van der Pas, Stéphanie and Szabó, Botond and Van der Vaart, Aad (2017). Adaptive posterior contraction rates for the horseshoe. *Electronic Journal of Statistics, 11*(2), 3196–3225.