Posterior Concentration Properties of a General Class of Shrinkage Estimators around Nearly Black Vectors

Prasenjit Ghosh and Arijit Chakrabarti

Applied Statistics Unit, Indian Statistical Institute, Kolkata, India

January 6, 2015

Abstract

We consider the problem of estimating a high-dimensional multivariate normal mean vector when it is sparse in the sense of being nearly black. Optimality of Bayes estimates corresponding to a very general class of continuous shrinkage priors on the mean vector is studied in this work. The class of priors considered is rich enough to include a wide variety of heavy-tailed priors including the horseshoe which are in extensive use in sparse high-dimensional problems. In particular, the three parameter beta normal mixture priors, the generalized double Pareto priors, the inverse gamma priors and the normal-exponential-gamma priors fall inside this class. We work under the frequentist setting where the data is generated according to a multivariate normal distribution with a fixed unknown mean vector. Under the assumption that the number of non-zero components of the mean vector is known, we show that the Bayes estimators corresponding to this general class of priors attain the minimax risk (possibly up to a multiplicative constant) corresponding to the $l_2$ loss. Further an upper bound on the rate of contraction of the posterior distribution around the estimators under study is established. We also provide a lower bound to the posterior variance for an important subclass of this general class of shrinkage priors that include the generalized double Pareto priors with shape parameter $\alpha = \frac{1}{2}$, the three parameter beta normal mixtures with parameters $a = \frac{1}{2}$ and $b > 0$ (including the horseshoe in particular), the inverse gamma prior with shape parameter $\alpha = \frac{1}{2}$ and many other shrinkage priors. This work is inspired by the recent work of van der Pas et al (2014) on the posterior contraction properties of the horseshoe prior under the present set-up. We extend their results for this general class of priors and come up with novel unifying proofs using properties of slowly varying functions. This work shows that the general scheme of arguments in van der Pas et al (2014) can be used in greater generality.

1 Introduction

With rapid advancements in modern technology and computing facilities, high throughput data have become commonplace in real life problems across diverse scientific fields such as genomics, biology, medicine, cosmology, finance, economics and climate studies. As a result inferential problems involving a large number of unknown parameters are coming to the fore. Problems where the number of unknown parameters grows as least as fast as the number of observations are typically called high-dimensional. In such problems, often times it is also true that only a few of these parameters are of real importance. For example, in a high dimensional regression problem, it is often true that the proportion of non-zero regressors or regressors with large magnitude is quite small compared to the total number of candidate regressors. This is called the phenomenon of sparsity. A common Bayesian approach to model sparse high-dimensional data is to use a two-component point mass mixture prior for the parameters and they put a positive mass at zero (to

\textsuperscript{1}Email for correspondence: prasenjit@isical.ac.in Arijit Chakrabarti: arc@isical.ac.in
induce sparsity) and a heavy tailed continuous distribution (to identify the non-zero coefficients). These are also referred to as “spike and slab priors” or two-groups priors. This is a very natural way of modelling data of this kind from a Bayesian viewpoint. See Johnstone and Silverman (2004) and Efron (2004) in this context.

Use of the two-groups prior, although very natural, poses a very daunting task computationally. Note that the cardinality of the model space becomes $2^p$ where $p$ is the number of parameters involved, and even for moderately large $p$ like 50, it is practically impossible to study posterior probabilities of the different models. Sometimes it is also possible that most of the parameters are very close to zero, but not exactly equal to zero. So in such a case a continuous prior may be able to capture sparsity in a more flexible manner. Due to these reasons, significant efforts have gone into modeling sparse high-dimensional data in recent times through hierarchical one-group continuous priors, which are also called one-group shrinkage priors. Bayesian analysis is computationally much more tractable than the two-group prior in such cases and easily implementable through standard MCMC techniques. But more importantly, these priors are suitable to capture sparsity since they accord a significant chunk of probability around zero while they have tails which are heavy enough to ensure a priori large probabilities for large parameter values. In general, such priors are expressed as multivariate scale-mixtures of normals that mix over two levels of parameters appearing in the scale, referred to as a “global” shrinkage parameter and a “local” shrinkage parameter. While the global shrinkage parameter accounts for the overall sparsity in the data by shrinking the noise observations to the origin, the local shrinkage parameters are helpful in detecting the obvious signals by leaving the large observations mostly unshrunk.

A great variety of one-group shrinkage priors have appeared in the literature over the years. Notable early examples are the $t$-prior in Tipping (2001), the double-exponential prior in Park and Casella (2008) and Hans (2008), and the normal-exponential-gamma priors in Griffin and Brown (2005). Very recently Carvalho et al. (2009, 2010) introduced the horseshoe prior, which has very appealing properties. Subsequently, many other one-group priors have been proposed in the literature, e.g., in Polson and Scott (2011, 2012), Armagan et al. (2011), Armagan et al. (2012) and Griffin and Brown (2010, 2012, 2013). The class of “three parameter beta normal” mixture priors was introduced in Armagan et al. (2011) and “generalized double Pareto” class of priors was introduced by Armagan et al. (2012). The three parameter beta normal mixture family of priors encompasses among others the horseshoe, the Strawderman-Berger and the normal-exponential-gamma priors. Very recently, a different class of one-group priors named Dirichlet-Laplace (DL) priors have been introduced in Bhattacharya et al. (2014). They investigated its various theoretical properties and demonstrated its good performances through extensive simulations.

As commented in Castillo and van der Vaart (2012), the Bayesian approach to sparsity is not driven by the ultimate goal of producing estimators that attain the minimax rate or for that matter posterior distributions with rate of contraction same as the minimax rate. However, for theoretical investigations, minimax rate can be taken as a benchmark and this is a motivation to study this kind of optimality properties for the Bayesian approach to sparsity. In an important article, Johnstone and Silverman (2004) focused on the case where a two-groups prior is used to model the mean parameters. They showed that if the unknown proportion of non-zero means is estimated by marginal maximum likelihood and a coordinatewise posterior median estimate is used, the resulting estimator attains the minimax rate with respect to $l_q$ loss, $q \in (0, 2]$. In Castillo and van der Vaart (2012), the full Bayes approach was studied where they found conditions on the two-groups prior that ensure contraction of the posterior distribution at the minimax rate. In recent times, researchers have started to investigate the optimality properties of estimators and testing rules based on these, and posterior contraction rate where a one-group shrinkage prior has been used instead. Amongst various one-group shrinkage priors, the horseshoe prior has acquired a prominent place in the Bayesian literature and it has been used extensively in inferential problems involving sparsity. Carvalho et al. (2010) have theoretically showed good performance of the horseshoe estimator (the Bayes estimate corresponding to the horseshoe prior) in
terms of the Kullback-Leibler risk when the true mean is zero. Datta and Ghosh (2013) showed a near oracle optimality property of multiple testing rules based on the horseshoe estimator in the context of multiple testing. Ghosh et al (2014) extended their work by theoretically showing that the multiple testing rules based on a general class of tail robust shrinkage priors enjoy similar optimality properties as the horseshoe. This general class of shrinkage priors is rich enough to include among others, the three parameter beta normal priors, the generalized double Pareto priors, the inverse gamma priors, and the normal-exponential-gamma priors, the horseshoe prior and the Strawderman-Berger prior, in particular. In an important recent article, van der Pas et al (2014) showed that for the problem of estimation of a sparse normal mean vector, the horseshoe estimator asymptotically achieves the minimax risk with respect to the $l_2$ loss, possibly up to a multiplicative constant and the corresponding posterior distribution contracts at least as fast as the minimax rate around the posterior mean. This was shown assuming that the number of non-zero means is known and the global shrinkage parameter tends to zero at an appropriate rate as the dimension grows to infinity. They also provide conditions under which the horseshoe estimator combined with an empirical Bayes estimate of the global variance component still attains the minimax quadratic risk even when the number of non-zero means is unknown. In a beautiful recent article, Bhattacharya et al (2014) showed that for the estimation of a sparse multivariate normal mean vector, under the quadratic risk function, the posterior arising from the Dirichlet-Laplace prior attains a minimax optimal rate of posterior contraction, that is, the corresponding posterior distribution contracts at the minimax rate for an appropriate choice of the underlying Dirichlet concentration parameter. See also Bickel, Ritov and Tsybakov (2009) for the minimax risk properties of the Lasso estimator which is the least squares estimator of the regression coefficients with an $l_1$ constraint on the regression coefficients. It was later shown by Castillo et al (2014) that the corresponding entire posterior distribution contracts at a much slower rate, thus indicating an inadequate measure of uncertainty in the estimate.

A natural question to ask, and also posed in section 6 of van der Pas et al (2014), is what aspects of the shrinkage priors are essential towards obtaining optimal posterior concentration properties as obtained for the case of the horseshoe prior. As mentioned earlier, in the context of simultaneous testing of a large number of independent normal means, Ghosh et al (2014) considered a general class of heavy-tailed shrinkage priors and showed some optimality properties of the multiple testing rules induced by the corresponding Bayes estimates. Polson and Scott (2011) suggested that in sparse problems, one should choose the prior distribution corresponding to the local shrinkage parameter to be appropriately heavy-tailed so that large signals can escape the “gravitational pull” of the corresponding global variance component and are almost left unshrunk which is essential for the recovery of large signals when the data is sparse. It is to be mentioned in this context that priors with exponential or lighter tails, such as the Laplace or the double-exponential prior and the normal prior, fail to meet this condition. Motivated by this, we consider in this article, the problem of estimating a sparse multivariate normal mean vector based on a very general class of “tail-robust” shrinkage priors that is rich enough to include a wide variety of shrinkage priors, such as, the three parameter beta normal mixtures (which generalizes the horseshoe prior in particular), the generalized double Pareto prior, the inverse-gamma priors, the half-t priors and many more. It is shown that when the underlying multivariate normal mean which is sparse in the nearly-black sense, the Bayes estimates corresponding to this general class of priors asymptotically attain the minimax-quadratic risk possibly up to a multiplicative factor and the entire posterior distribution contracts at least as fast as the minimax rate around the posterior mean assuming the number of non-zero means is known and that we are free to choose the global shrinkage parameter which tends to zero at an appropriate rate as the dimension grows to infinity. An important contribution of our theoretical investigation is showing that shrinkage priors which are appropriately heavy-tailed (to be defined in Section 2) and have sufficient mass around the origin, are good enough to attain the minimax optimal rate of contraction, provided that the global tuning parameter is carefully chosen. We also provide a lower bound to the corresponding posterior variance for an important subclass of this general class of shrinkage priors that include the generalized double Pareto priors with shape parameter $\alpha = \frac{1}{2}$, the three parameter
beta normal mixtures with parameters \( a = \frac{1}{2} \) and \( b > 0 \) (including the horseshoe in particular), the inverse gamma prior with shape parameter \( (\alpha = \frac{1}{2}) \) and many other shrinkage priors. We provide a general unifying argument that works for this general class under consideration and thus extends the work of van der Pas et al. (2014).

We organize the paper as follows. In Section 2, we describe the problem and the general class of shrinkage priors under consideration. Section 3 contains the main theoretical results, that estimators arising out of this general class of shrinkage priors attain the minimax quadratic risk up to some multiplicative constant and that the corresponding posterior distribution results in a minimax optimal rate of posterior contraction. Proofs of the main theorems and other theoretical results essential for their derivation are given in Section 4 (Appendix) followed by a discussion in Section 5.

**Notations:** In this paper, we adopted the same convention of notation used in van der Pas et al. (2014). Let \( \{A_n\} \) and \( \{B_n\} \) be two sequences of positive real numbers indexed by \( n \). We write \( A_n \asymp B_n \) to denote \( 0 < \lim_{n \to \infty} \inf \frac{A_n}{B_n} \leq \lim_{n \to \infty} \sup \frac{A_n}{B_n} < \infty \) and \( A_n \lesssim B_n \) to denote that there exists some \( c > 0 \) independent of \( n \) such that \( A_n \leq cB_n \).

## 2 A General Class of Tail Robust Shrinkage Priors

Let us suppose that we observe an \( n \)-component random observation \( (X_1, \ldots, X_n) \in \mathbb{R}^n \), such that

\[
X_i = \theta_i + \epsilon_i \quad \text{for} \quad i = 1, \ldots, n, \tag{2.1}
\]

where the unknown parameters \( \theta_1, \ldots, \theta_n \) denote the effects under investigation and \( \epsilon = (\epsilon_1, \ldots, \epsilon_n) \sim N_n(0, I_n) \).

Let \( l_0[p_n] \) denote the subset of \( \mathbb{R}^n \) given by,

\[
l_0[p_n] = \{ \theta \in \mathbb{R}^n : \#(1 \leq j \leq n : \theta_j \neq 0) \leq p_n \}. \tag{2.2}
\]

Suppose we want to estimate the true mean vector \( \theta_0 = (\theta_{01}, \ldots, \theta_{0n}) \) when \( \theta_0 \) is known to be sparse in the “nearly black sense”, that is, \( \theta_0 \in l_0[p_n] \) with \( p_n = o(n) \) as \( n \to \infty \). The corresponding squared minimax rate for estimating \( \theta_0 \) is known to be \( 2p_n \log(n/p_n)(1 + o(1)) \) as \( n \to \infty \) (see Donoho et al. (1992)), that is,

\[
\inf_{\delta} \sup_{\theta \in l_0[p_n]} \mathbb{E}_{\theta_0}[|\hat{\theta} - \theta_0|^2] \asymp p_n \log\left(\frac{n}{p_n}\right). \tag{2.3}
\]

In (2.3) above and throughout this paper \( E_{\theta_0} \) denotes an expectation with respect to the \( N_n(\theta_0, I_n) \) distribution. Our goal is to obtain an estimate \( \hat{\theta}_0 \) from a Bayesian viewpoint with some good theoretical properties. As stated already in the introduction that a natural Bayesian approach to model (2.1) is to use a two-component point mass prior for the \( \theta_i \)’s, given by,

\[
\theta_i \overset{i.i.d.}{\sim} (1 - \pi) \delta_{0} + \pi \cdot f, \quad i = 1, \ldots, m. \tag{2.4}
\]

where \( \delta_{0} \) denotes the distribution having probability mass 1 at the point 0, \( \pi \) denotes an absolutely continuous distribution over \( \mathbb{R} \). See Mitchell and Beauchamp (1988) and Johnstone and Silverman (2004) in this context. It is usually recommended to choose a heavy tailed absolutely continuous distribution \( f \) over \( \mathbb{R} \) so that large observations can be recovered with higher degree of accuracy. Johnstone and Silverman (2004) used a \( t \) distribution in this context and used an empirical Bayes approach in order to estimate the unknown mixing proportion \( \pi \) via the method of marginal maximum likelihood and showed that if the co-ordinatewise posterior median estimate is used, the resulting estimator of \( \hat{\theta}_0 \) attains the minimax rate with respect to the \( l_0 \) loss, \( q \in (0, 2] \). Castillo and van der Vaart (2012) studied the full Bayes approach where they found conditions...
on the two-groups prior that ensure contraction of the posterior distribution at the minimax rate. A detailed list of other empirical Bayes approaches to the two-group model can be found in Castillo and van der Vaart (2012), Efron (2008), Jiang and Zhang (2005), Yuan and Lin (2005) and references therein.

As already mentioned in the introduction that although the two groups prior (2.4) is considered to be the most natural formulation for handling sparsity from a Bayesian viewpoint, it offers a daunting computational challenge in high dimensional problems because of the enormously large model space (in this case it is $2^n$). Due to this reason, the one-group formulation to model sparse data has received considerable attention from researchers over the years, mostly due to the ease of their computational tractibility. Polson and Scott (2011) showed that almost all such shrinkage priors can be expressed as multivariate scale-mixture of normals which makes the computation based on these one-group shrinkage priors much easier compared to the corresponding two-group formulation. Standard Markov-chain Monte Carlo techniques are available in the Bayesian literature for the computation of the corresponding Bayes estimates of the underlying model parameters. In this article, we consider Bayes estimators based on a general class of one-group shrinkage priors given through the following hierarchical one-group formulation:

\[
\begin{align*}
X_i | \theta_i & \sim N(\theta_i, 1), \text{ independently for } i = 1, \ldots, m \\
\theta_i | (\lambda_i^2, \tau^2) & \sim N(0, \lambda_i^2 \tau^2), \text{ independently for } i = 1, \ldots, m \\
\lambda_i^2 & \sim \pi(\lambda_i^2), \text{ independently for } i = 1, \ldots, m
\end{align*}
\]

with $\pi(\lambda_i^2)$ being given by,

\[
\pi(\lambda_i^2) = K(\lambda_i^2)^{-a-1}L(\lambda_i^2),
\]

where $K \in (0, \infty)$ is the constant of proportionality, $a$ is a positive real number and $L: (0, \infty) \rightarrow (0, \infty)$ is a measurable, non-constant, slowly varying function satisfying the following:

**Assumption 2.1.**

1. $\lim_{t \to \infty} L(t) \in (0, \infty)$, that is, there exists some positive real number $c_0$ such that $L(t) > c_0$ for all $t \geq t_0$, for some finite positive real number $t_0$ depending on $L$ and $c_0$. Choose $t_0 > 0$ to be the minimum of all such $t$’s such that $L(t) > c_0$.

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

Recall that a measurable function $L: (0, \infty) \rightarrow (0, \infty)$ is said to be slowly varying if for each fixed $\alpha > 0$, $\frac{L(tx)}{L(x)} \rightarrow 1$ as $x \rightarrow \infty$. A simple sufficient condition for a function $L: (0, \infty) \rightarrow (0, \infty)$ to be slowly varying is that $\lim_{t \rightarrow \infty} L(x) \in (0, \infty)$. Hence, every constant function is a slowly varying function. However, since we assume the prior given in (2.5) to be proper, the possibility of $L(\cdot)$ being a constant function is immediately ruled out.

Each $\lambda_i^2$ is referred to as a local shrinkage parameter and the parameter $\tau^2$ is called the global shrinkage parameter. For the theoretical treatment of this paper, we assume the global shrinkage parameter $\tau^2$ to be known. We would like to mention here that a very broad class of one-group shrinkage priors actually fall inside this above general class. For example, it can be easily seen that the celebrated horseshoe prior is a member of this general class under study by simply taking $a = 0.5$ and $L(t) = t/(1 + t)$ in (2.5) satisfying both the conditions of Assumption (2.1). Ghosh et al (2014) observed that the three parameter beta normal mixtures (which include the horseshoe and the normal-exponential-gamma priors as special cases) and the generalized double Pareto priors can be expressed in the above general form by showing that the corresponding prior distribution of the local shrinkage parameters can be written in the form given in (2.5) with the corresponding $L(\cdot)$ satisfying Assumption (2.1). It is easy to verify that some other well known shrinkage priors such as the families of inverse-gamma priors and the half-t priors are also covered by this general class of prior distributions under consideration. We would like to mention in this context that the above general class exclude priors such as the double-exponential or Laplace prior.
or the normal prior which have exponential or lighter tails.

From Theorem 1 of Polson and Scott (2011) it follows that the above general class of one-group priors will be “tail-robust” in the sense that for any given \( \tau > 0 \), \( E(\theta_i|X, \tau^2) \approx X_i \), for large \( X_i \)’s, which means for such priors large observations will be almost left unshrunk even when the global shrinkage parameter \( \tau^2 \) is too small. We shall elucidate this fact in some greater detail in the forthcoming sections using properties of slowly varying functions. It was suggested in Polson and Scott (2011) that the global shrinkage parameter \( \tau^2 \) should be very small so that small \( X_i \)’s or the noise observations can be shrunk towards the origin while the prior distribution of the local shrinkage parameters \( \lambda_i^2 \) should have heavy tails so that large signals can escape the effect of \( \tau^2 \) and almost remain unshrunk. Thus \( \lambda_i^2 \) should result in a prior distribution for the \( \theta_i \)’s which has a high concentration of mass near the origin but have thick tails at the extremes to accomodate large signals. Polson and Scott (2011) also showed that for priors having exponential or lighter tails, such as the Laplace or the double-exponential prior, even the large \( X_i \)’s will always be shrunk towards the origin by some non-diminishing amount for small values of \( \tau \), which is certainly not desirable for the recovery of large signals in sparse situations.

Now for a general global-local scale mixture of normals we have,

\[
\theta_i|(X_i, \lambda_i^2, \tau^2) \sim N((1 - \kappa_i)X_i, (1 - \kappa_i)), \quad \kappa_i = 1/(1 + \lambda_i^2 \tau^2),
\]

independently for \( i = 1, \cdots, m \), so that for each \( i \), the posterior mean of \( \theta_i \) is given by,

\[
E(\theta_i|X_i, \lambda_i^2, \tau^2) = (1 - \kappa_i)X_i.
\]

Next, using the iterated expectation formula it follows that,

\[
E(\theta_i|X_i, \tau^2) = (1 - E(\kappa_i|X_i, \tau^2))X_i.
\]

The corresponding posterior mean \( E(\theta|X, \tau) = (E(\theta_1|X_1, \tau^2), \cdots, E(\theta_m|X_m, \tau^2)) \) will be the estimator arising out of the general class of shrinkage priors (2.5) and will be denoted by \( T_\tau(X) \). It will be shown in the next section that when \( \theta_0 \) is sparse in the “nearly black sense” and \( a \in [\frac{1}{2}, 1] \), the estimator \( T_\tau(X) \) of \( \theta_0 \) will asymptotically attain the minimax rate (2.3) up to some multiplicative constant assuming \( \tau = \frac{\kappa}{\kappa^*} \) and that the posterior distribution contracts at least as fast as the minimax rate around the posterior mean.

3 Theoretical Results

In this section, we first state two optimality results of the general class of shrinkage estimators when \( \frac{1}{2} \leq a < 1 \), assuming that the number of non-zero parameters \( p_n \) is known. Theorem 3.1 states that the general class of heavy-tailed shrinkage priors attain the minimax risk under the \( l_2 \)-norm, when \( \frac{1}{2} \leq a < 1 \), possibly up to a multiplicative constant. Theorem 3.2 provides an upper bound to the variance corresponding to the posterior distribution based on the chosen class of heavy-tailed distribution. Theorem 3.3 provides an upper bound to the rate of posterior contraction which is equal to the corresponding squared error minimax risk up to a multiplicative constant. Theorem 3.4 provides a lower bound to the posterior variance for an important subclass of this general class of shrinkage priors that gives more insight about the spread of the posterior distribution around these estimators for various choices of \( \tau \). Our proofs are based on novel unifying arguments crucially exploiting properties of slowly varying functions. We however followed the broad architecture of the proofs of the main theorems of van der Pas et al (2014). Lemmas 4.3, 4.4 and 4.5, given in the appendix, on which Theorems 3.1 through 3.3 crucially hinge upon, are completely independent of the work of van der Pas et al (2014). However, proofs of Lemma 4.6 and Theorem 3.4 have been derived following some key arguments of van der Pas et al (2014). This shows that the general scheme of arguments in van der Pas et al (2014) can be used in greater generality which will be evident in the next section.
Theorem 3.1. Suppose $X \sim N_n(\theta_0, I_n)$. Then the estimator $T_\tau(x)$ based on the general class of shrinkage priors (2.5), with $\frac{1}{2} \leq a < 1$, satisfies

$$\sup_{\theta_0 \in \Theta} E_{\theta_0} ||T_\tau(X) - \theta_0||^2 \asymp p_n \log(\frac{n}{p_n})$$

if $\tau = (\frac{p_n}{n})^a$, $a \geq 1$, as $n, p_n \to \infty$ and $p_n = o(n)$.

Proof. See Section 4.

A remarkable implication of Theorem 3.1 is that, for $\frac{1}{2} \leq a < 1$, Bayes estimators based on the general class of heavy-tailed shrinkage priors under consideration attain the minimax quadratic risk possibly up to some multiplicative factor. Theorem 3.1 therefore shows that the three parameter beta normal mixtures, the generalized double Pareto, the inverse gamma priors and the half-t priors, and in particular the normal-exponential-gamma priors and the horseshoe prior, result in Bayes estimators that attain the minimax risk in $l_2$ norm, thus extending the asymptotic minimaxity property of the horseshoe estimator obtained by van der Pas et al (2014) and can be considered as an important theoretical justification for the use of such priors when $\frac{1}{2} \leq a < 1$.

The next theorem gives an upper bound on the posterior variance corresponding to our general class of heavy tailed shrinkage priors which indicates that the resulting posterior distribution contracts fast enough to produce an adequate measure of uncertainty around the corresponding Bayes estimate.

Theorem 3.2. Suppose $X \sim N_n(\theta_0, I_n)$. Then the variance of the posterior distribution corresponding to the general class of shrinkage priors (2.5), with $\frac{1}{2} \leq a < 1$, satisfies

$$\sup_{\theta_0 \in \Theta} E_{\theta_0} \sum_{i=1}^{n} \text{Var}(\theta_{0i} | X_i) \asymp p_n \log(\frac{n}{p_n})$$

if $\tau = (\frac{p_n}{n})^a$, $a \geq 1$, as $n, p_n \to \infty$ and $p_n = o(n)$.

Proof. See Section 4.

Theorems 3.1 and 3.2 allow us to find an upper bound on the rate of contraction of the full posterior distribution, both around the underlying mean vector and around the Bayes estimates arising out of this general class of shrinkage priors (2.5).

Theorem 3.3. Under the assumptions of Theorem 3.2,

$$\sup_{\theta_0 \in \Theta} E_{\theta_0} \Pi \left( \theta : ||\theta - \theta_0||^2 > M_n p_n \log(\frac{n}{p_n}) | X \right) \to 0,$$

and

$$\sup_{\theta_0 \in \Theta} E_{\theta_0} \Pi \left( \theta : ||\theta - T_\tau(X)||^2 > M_n p_n \log(\frac{n}{p_n}) | X \right) \to 0,$$

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

Proof. A straight forward application of Markov’s inequality coupled with the results of Theorem 3.1 and Theorem 3.2 leads to (3.3), while (3.4) is followed from the result of Theorem 3.2.

Ghosal et al (2000) showed that the posterior distribution cannot contract faster than the minimax rate around the true mean vector. Hence, the upper bound on the rate of contraction in (3.3) is sharp. However, the same cannot be inferred about the upper bound in (3.4). To get a better insight about the spread of the posterior distribution around these estimators, let us confine our attention to the case when $L(\cdot)$ given in (2.5) is non-decreasing over $(0, \infty)$ with $a = 0.5$. This subclass include the generalized double Pareto priors with shape parameter $\alpha = \frac{1}{2}$,
the three parameter beta normal mixtures with parameters $a = 1/2$ and $b > 0$ (including the horseshoe in particular), the inverse gamma prior with shape parameter $(\alpha = 1/2)$ and many more (see Ghosh et al. (2014) in this context). The next theorem gives a lower bound on the posterior variance corresponding to this restricted subclass that provide more insight into the effect of the choice of $\alpha$.

**Theorem 3.4.** Suppose $X \sim N_n(\theta_0, I_n)$ and $\theta_0 \in \mathbb{I}_0[p_n]$. Further assume that the function $L(.)$ given by (2.5) satisfies Assumption 2.1 and is non-decreasing over $(0, \infty)$. Then for $\alpha = 1/2$, the variance of the posterior distribution corresponding to the general class of shrinkage priors, satisfies

$$
\sum_{i=1}^{n} E_{\theta_i} Var(\theta_0 | X_i) \gtrsim \frac{\rho_n^2}{n^{\alpha-1}} \sqrt{\log \left( \frac{n}{p_n} \right)}
$$

(3.5)

if $\tau = (\frac{n}{n})^\alpha$, $\alpha > 0$, as $n, p_n \to \infty$ and $p_n = o(n)$.

**Proof.** See Section 4.

Following the arguments of van der Pas et al. (2014) we conclude that for $0 < \alpha < 1$ the posterior distribution corresponding to this restricted sub-class contracts at a sub-optimal rate, while for $\alpha > 1$ it contracts too quickly resulting in an inadequate measure of uncertainty about the corresponding Bayes estimates. On the other hand, the choice $\alpha = 1$ seems to be optimal in the following sense: the lower bound obtained in Theorem 3.4 is of the order of $p_n \sqrt{\log(n/p_n)}$ which misses the minimax rate by a factor of $\sqrt{\log(n/p_n)}$ and this suggests that the posterior distribution corresponding to this restricted subclass concentrates around the corresponding Bayes estimates at a rate close to the minimax rate (2.3).

## 4 Appendix

### 4.1 Appendix A: Proofs

**Lemma 4.1.** For the general class of shrinkage priors (2.5) satisfying Assumption (2.1) the following holds true for any $0 < a < 1$:

$$
E(1 - \kappa | x, \tau) \leq \frac{KM}{a(1-a)} e^{-\frac{a}{2a} (1 + o(1))}, \text{ each fixed } x \in \mathbb{R},
$$

where $\kappa = \frac{1}{1 + \lambda^2 \tau^2}$ denote the shrinkage coefficients and the $o(1)$ term depends only on $\tau^2$ such that $\lim_{\tau \to 0} o(1) = 0$.

**Proof.** See Theorem 3.1 and Theorem 3.2 and subsequent discussions of Ghosh et al. (2014).

**Lemma 4.2.** For every fixed $\tau > 0$, and each fixed $\eta, \delta \in (0, 1)$, the posterior distribution of the shrinkage coefficients $\kappa = 1/(1 + \lambda^2 \tau^2)$ based on the general class of shrinkage priors (2.5) satisfying Assumption (2.1), with $a > 0$, satisfies the following concentration inequality:

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

where $\Delta(\tau^2, \eta, \delta) = \xi(\tau^2, \eta, \delta) L \left( \frac{1}{\tau^2} \left( \frac{1}{\eta^2} - 1 \right) \right)$,

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

$$
H(a, \eta, \delta) = \frac{(a + \frac{1}{2})(1 - \eta \delta)^a}{K(\eta \delta)^{a+\frac{1}{2}}},
$$
where the term \(\Delta(\tau^2, \eta, \delta)\) is such that \(\lim_{\tau \to 0} \Delta(\tau^2, \eta, \delta)\) is a finite positive quantity for every fixed \(\eta \in (0, 1)\) and every fixed \(\delta \in (0, 1)\).

**Proof.** See Theorem 3.3 of Ghosh et al (2014).

**Corollary 4.1.** Under the assumptions of Lemma 4.1, for every fixed \(0 < a < 1\), and each fixed \(x \in \mathbb{R}\),

\[
E(\kappa | x, \tau) \to 1 \text{ as } \tau \to 0.
\]

**Proof.** The proof follows immediately from Lemma 4.1.

**Lemma 4.3.** For \(0 < \tau^2 < 1\), given any \(c > 2\), the absolute value of the difference between the Bayes estimators \(T_\tau(x)\) based on the general class of shrinkage priors \((2.5)\) satisfying Assumption (2.1) and an observation \(x\), can be bounded above by a real valued function \(h(x, \tau)\), depending on \(c\), and satisfying the following:

For any \(\rho > c\),

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

**Proof.** By definition,

\[
| T_\tau(x) - x | = | x E(\kappa | x, \tau^2) |
\]

\[
= \left| \int_0^1 \kappa \kappa^{a + \frac{1}{2} - 1}(1 - \kappa)^{-a - 1} L\left(\frac{1}{\kappa} x \tau \right) e^{-\kappa x^2 / 2} d\kappa \right|
\]

\[
= \int_0^1 \kappa^{a + \frac{1}{2} - 1}(1 - \kappa)^{-a - 1} L\left(\frac{1}{\kappa} x \tau \right) e^{-\kappa x^2 / 2} d\kappa
\]

\[
= I(x, \tau), \text{ say.}
\]

Fix \(\eta \in (0, 1)\) and \(\delta \in (0, 1)\).

Observe that

\[
I(x, \tau) \leq I_1(x, \tau) + I_2(x, \tau)
\]

where \(I_1(x, \tau) = | x E(\kappa \{ \kappa < \eta \} | x, \tau^2) | \) and \(I_2(x, \tau) = | x E(\kappa \{ \kappa > \eta \} | x, \tau^2) | \).

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

\[
I_1(x, \tau) = \left| x E(\kappa \{ \kappa < \eta \} | x, \tau^2) \right|
\]

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

\[
= \int_0^{\infty} \kappa^{a + \frac{1}{2} - 1}(1 - \kappa)^{-a - 1} L\left(\frac{1}{\kappa} \frac{1}{\tau^2} \frac{1}{\kappa} x \tau \right) e^{-\kappa x^2 / 2} d\kappa
\]

\[
\leq \left| \int_0^{\infty} \frac{1}{(1 + t^2)^{1/2}} \kappa^{a - 1} L(t) e^{-t x^2} dt \right|
\]

\[
= J_1(x, \tau) \text{ say,}
\]

Next observe that \(\frac{dx}{dt} > 0\) as \(\tau^2 < 1\). Hence \(L(t) \geq c_0\) for every \(t \geq \frac{dx}{dt}\). Also, the function \(L\) is bounded by the constant \(M > 0\). Utilizing these two observations and using the variable transformation \(u = x^2 / (1 + t^2)\) in both the numerator and the denominator of the ratios of two integrals.
on the right hand side of (4.2), and writing \( s = \frac{1}{1+\eta} \in (0, 1) \), the term \( J_1(x, \tau) \) can be bounded above as follows:

\[
J_1(x, \tau) \leq \frac{M}{c_0} | x | x^{1/2} \left[ \frac{\int_0^{sx^2} e^{-u/2} u^{a+3/2-1} du}{\int_0^{sx^2} e^{-u/2} u^{a+1/2-1} du} \right]^{-a-1}
\]

\[
= \frac{M}{c_0} | x | x^{1/2} \left[ \frac{\int_0^{sx^2} e^{-u/2} u^{a+3/2-1} du}{\int_0^{sx^2} e^{-u/2} u^{a+1/2-1} du} \right]^{-a-1}
\]

Note that \( 0 < u < \eta x^2 \Rightarrow 0 < \frac{u}{x^2} < \eta < 1 \Rightarrow 1 - \eta < 1 - \frac{u}{x^2} < 1 \). Similarly, \( 0 < u < sx^2 \Rightarrow 1 - s < 1 - \frac{u}{x^2} < 1 \). Therefore we have,

\[
J_1(x, \tau) \leq \frac{M}{c_0(1-\eta)^{1+a}} | x | x^{1/2} \left[ \frac{\int_0^{sx^2} e^{-u/2} u^{a+3/2-1} du}{\int_0^{sx^2} e^{-u/2} u^{a+1/2-1} du} \right]^{-a-1}
\]

\[
= h_1(x, \tau), \quad (4.3)
\]

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

Next we observe that,

\[
I_2(x, \tau) = | x E(\kappa \{ \kappa > \eta \} | x, \tau^2) | \leq | x \Pr(\kappa > \eta) | \leq | x \frac{H(a, \eta, \delta)}{\tau^{2a} \Delta(\tau^2, \eta, \delta)} | = h_2(x, \tau), \quad (4.4)
\]

Let \( h(x, \tau) = h_1(x, \tau) + h_2(x, \tau) \). Therefore combining (4.1), (4.2), (4.3) and (4.4) we finally obtain for every \( x \in \mathbb{R} \) and \( \tau > 0 \),

\[
| T_\tau(x) - x | \leq h(x, \tau), \quad (4.5)
\]

Now observe that the function \( h_1(x, \tau) \) is strictly decreasing in \( | x | \). Therefore, for any fixed \( \tau > 0 \) and every \( \rho > 0 \),

\[
\sup_{|x| > \sqrt{\rho \log(1/\tau^2)}} h_1(x, \tau) \leq C_\ast \left[ \frac{\rho \log(1/\tau^2)}{\tau^{2a}} \right]^{-a-1} \int_{0}^{\sqrt{\rho \log(1/\tau^2)}} e^{-u/2} u^{a+1/2-1} du
\]

implying that

\[
\lim_{\tau \downarrow 0} \sup_{|x| > \sqrt{\rho \log(1/\tau^2)}} h_1(x, \tau) = 0. \quad (4.6)
\]

Again the function \( h_2(x, \tau) \) is eventually decreasing in \( | x | \). Therefore, for all sufficiently small \( \tau > 0 \),

\[
\sup_{|x| > \sqrt{\rho \log(1/\tau^2)}} h_2(x, \tau) \leq h_2\left( \sqrt{\rho \log(1/\tau^2)}, \tau \right).
\]
Let \( \beta = \lim_{\tau \to 0} \Delta(\tau^2, \eta, \delta) \) for every fixed \( \eta, \delta \in (0, 1) \). Then \( 0 < \beta < \infty \) as followed from Lemma 4.2. Then,

\[
\lim_{\tau \to 0} h_2(\sqrt{\rho \log(\frac{1}{\tau^2})}, \tau) = \frac{1}{\beta} \lim_{\tau \to 0} |\tau^{-2a} \sqrt{\rho \log(\frac{1}{\tau^2})} e^{-\frac{a(1-\delta)}{2a} \rho \log(\frac{1}{\tau^2})}| = \sqrt{\frac{\alpha}{\beta}} \lim_{\tau \to 0} (\sqrt{2a} e^{-\frac{1}{2} a(1-\delta)} (\rho^{-\frac{1}{2}} - \frac{2}{\eta a(1-\delta)}) e^{-\frac{a(1-\delta)}{2a} \rho \log(\frac{1}{\tau^2})})
\]

whence it follows that

\[
\lim_{\tau \to 0} \sup_{|x| > \sqrt{\rho \log(\frac{1}{\tau^2})}} h_2(x, \tau) = \begin{cases} 0 & \text{if } \rho > \frac{2}{\eta (1-\delta)} \\ \infty & \text{otherwise,} \end{cases} \tag{4.7}
\]

Combining (4.4) and (4.7) together with the fact that

\[
\lim_{\tau \to 0} \sup_{|x| > \sqrt{\rho \log(\frac{1}{\tau^2})}} h(x, \tau) \leq \lim_{\tau \to 0} \sup_{|x| > \sqrt{\rho \log(\frac{1}{\tau^2})}} h_1(x, \tau) + \lim_{\tau \to 0} \sup_{|x| > \sqrt{\rho \log(\frac{1}{\tau^2})}} h_2(x, \tau)
\]

it immediately follows that

\[
\lim_{\tau \to 0} \sup_{|x| > \sqrt{\rho \log(\frac{1}{\tau^2})}} h(x, \tau) = \begin{cases} 0 & \text{if } \rho > \frac{2}{\eta (1-\delta)} \\ \infty & \text{otherwise,} \end{cases} \tag{4.8}
\]

Observe that by choosing \( \eta \) appropriately close to 1 and \( \delta \) sufficiently close to 0, any real number larger than 2 can be expressed in the form \( \frac{2}{\eta (1-\delta)} \). For example, choosing \( \eta = \frac{5}{\delta} \) and \( \delta = \frac{1}{\eta} \) we obtain \( \frac{2}{\eta (1-\delta)} = 3 \). Hence, given \( c > 2 \), let us choose \( 0 < \eta, \delta < 1 \) such that \( c = \frac{2}{\eta (1-\delta)} \). Clearly, the choice of the function \( h(x, \tau) \) depends on \( c > 2 \). Using the preceding arguments it therefore follows that given any \( c > 2 \), the absolute difference between the posterior mean \( \bar{T}_{\tau}(x) \) and an observation \( x \) can be bounded above by a function \( h(x, \tau) \) depending on \( c \) such that \( \lim_{\tau \to 0} \sup_{|x| > \sqrt{\rho \log(\frac{1}{\tau^2})}} h(x, \tau) = 0 \) for all \( \rho > c \). \( \Box \)

Remark 4.1 Observe that the function \( h \) defined in the proof of Lemma 4.3 also satisfies the following:

For each fixed \( \tau > 0 \),

\[
\lim_{|x| \to \infty} h(x, \tau) = 0.
\]

Proof of Theorem 3.1

Proof. Suppose that \( X \sim \mathcal{N}(\theta, I_n) \), \( \theta \in I_0[p_n] \) and \( \tilde{p}_n = \#\{i : \theta \neq 0\} \). Note that \( \tilde{p}_n \leq p_n \). Assume without any loss of generality that for \( i = 1, \ldots, p_n, \theta_i \neq 0 \), while for \( i = \tilde{p}_n + 1, \ldots, n, \theta_i = 0 \). We split up the expectation \( E_{\theta_0}[T_{\tau}(X) - \theta]^2 \) into the two corresponding parts:

\[
\sum_{i=1}^{p_n} E_{\theta_i}(T_{\tau}(X_i) - \theta_i)^2 = \sum_{i=1}^{\tilde{p}_n} E_{\theta_i}(T_{\tau}(X_i) - \theta_i)^2 + \sum_{i=\tilde{p}_n+1}^{n} E_{\theta_i}(T_{\tau}(X_i) - \theta_i)^2
\]

We will now show that these two terms can be bounded by \( \tilde{p}_n(1+2 \log(\frac{1}{\tilde{p}_n})) \) and \( (n-\tilde{p}_n)\tau^{2a} \sqrt{\log(\frac{1}{\tilde{p}_n})} \) respectively, up to multiplicative constants, for any choice of \( \tau \in (0, 1) \).
Non-zero $\theta_i$'s:

Let $\zeta_\tau = \sqrt{2\log(\frac{1}{\tau a})}$.

Then
\[
E_{\theta_i}(T_\tau(X_i) - \theta_i)^2 = E_{\theta_i}((T_\tau(X_i) - X_i) + (X_i - \theta_i))^2 \\
\leq 2E_{\theta_i}(T_\tau(X_i) - X_i)^2 + 2E_{\theta_i}(X_i - \theta_i)^2 \\
\leq 2 + 2\left(\sup_x |T_\tau(x) - x|\right)^2
\]

Using Lemma 4.3, given any $c > 1$, one can obtain a non-negative real-valued function $h(x, \tau)$, depending on $c$, which satisfies the following:
\[
\lim_{\tau \downarrow 0} \sup_{|x| > \rho \zeta_\tau} h(x, \tau) = 0 \text{ for all } \rho > c. \tag{4.9}
\]

Claim: As $\tau \to 0$ :
\[
\arg \max_x |T_\tau(x) - x| \lesssim \zeta_\tau. \tag{4.10}
\]

Proof of Claim: Let $x_0(\tau) = \arg \max_x |T_\tau(x) - x|$. Using the observation in Remark 4.1, it can be easily established that $|x_0(\tau)| < \infty$ for each fixed $\tau > 0$. On contrary, let us now assume the Claim to be false. Then, for all $c > 0$, $|x_0(\tau)| > c\zeta_\tau$ infinitely often. Let us fix any $x^* \in \mathbb{R} - \{0\}$, any $c > 1$ and any $\rho > c$. Then, by the definition we have, 
\[
|x^*||E(\kappa|x^*, \tau)| = |T_\tau(x^*) - x^*| \\
\leq |T_\tau(x_0(\tau)) - x_0(\tau)| \\
\leq \sup_{|x| > \rho \zeta_\tau} h(x, \tau)
\]

which would be a contradiction because as $\tau \to 0$, $\sup_{|x| > \rho \zeta_\tau} h(x, \tau) \to 0$ along a subsequence (using (4.9)), whereas $|x^*||E(\kappa|x^*, \tau)| \to |x^*|$ as $\tau \to 0$ which follows as an immediate consequence of Corollary 4.1.

Equation (4.10) together with the fact $|T_\tau(x)| \lesssim |x|$ immediately leads to the following:
\[
\left(\sup_x |T_\tau(x) - x|\right)^2 \lesssim \zeta_\tau^2 \tag{4.11}
\]

whence it follows that
\[
E_{\theta_i}(T_\tau(X_i) - \theta_i)^2 \lesssim 1 + \zeta_\tau^2. \tag{4.12}
\]

Parameters equal to zero: We split up the term for the zero means into two parts:
\[
E_0T_\tau(X)^2 = E_0T_\tau(X)^21\{|X| \leq \zeta_\tau\} + E_0T_\tau(X)^21\{|X| > \zeta_\tau\}, \tag{4.13}
\]

where $\zeta_\tau = \sqrt{2\log(\frac{1}{\tau a})}$.

Next using Lemma 4.1 we have, for any $0 < \tau < 1$ and for every $0 < a < 1$,
\[
E(1 - \kappa|X, \tau|^2) \leq \frac{KM}{a(1-a)}\zeta_\tau^2 \tau^2(1 + o(1)),
\]

where the $o(1)$ term depends only on $\tau^2$ with $\lim_{\tau \to 0} o(1) = 0$. 

12
Therefore, letting \( g_1(a) = \frac{KM}{a(1-a)} \) we have,

\[
E_0 T_\tau(X)^2 1\{|X| \leq \zeta_\tau\} \leq \frac{g_1(a)^2}{\sqrt{2\pi}} (\tau^{2a})^2 \int_{-\zeta_\tau}^{\zeta_\tau} x^2 e^{-\frac{x^2}{2}} dx (1 + o(1)) \text{ as } \tau \to 0
\]

\[
= \frac{2g_1(a)^2}{\sqrt{2\pi}} (\tau^{2a})^2 \int_{0}^{\zeta_\tau} x^2 e^{-\frac{x^2}{2}} dx (1 + o(1)) \text{ as } \tau \to 0
\]

\[
\leq \sqrt{\frac{2}{\pi}} g_1(a) (\tau^{2a})^2 \zeta_\tau \frac{1}{\tau^{2a}} (1 + o(1)) \text{ as } \tau \to 0
\]

\[
\lesssim \zeta_\tau \tau^{2a}
\]

(4.14)

For the second term we have:

\[
E_0 T_\tau(X)^2 1\{|X| > \zeta_\tau\} = 2 \int_{\zeta_\tau}^{\infty} x^2 \phi(x) dx
\]

\[
= 2[\zeta_\tau \phi(\zeta_\tau) + (1 - \Phi(\zeta_\tau))]
\]

\[
\leq 2\zeta_\tau \phi(\zeta_\tau) + 2\frac{\phi(\zeta_\tau)}{\zeta_\tau}
\]

\[
= \sqrt{\frac{2}{\pi}} \zeta_\tau \tau^{2a} (1 + o(1)) \text{ as } \tau \to 0
\]

\[
\lesssim \zeta_\tau \tau^{2a}
\]

(4.15)

Combining equations (4.13), (4.14) and (4.15), it follows that for all sufficiently small \( \tau \),

\[
\sum_{i=1}^{n} E_{\theta_i} (T_\tau(X_i) - \theta_i)^2 \lesssim \tilde{p}_n (1 + 2 \log\left(\frac{1}{\tau^{2a}}\right)) + (n - \tilde{p}_n) \tau^{2a} \sqrt{\log\left(\frac{1}{\tau^{2a}}\right)}
\]

(4.16)

Putting \( \tau = \left(\frac{a}{n}\right)^\alpha \) and taking supremum over all \( \theta \in l_0[p_n] \) on both sides of (4.16) we obtain:

\[
\sup_{\theta \in l_0[p_n]} \sum_{i=1}^{n} E_{\theta_i} (T_\tau(X_i) - \theta_i)^2 \lesssim \tilde{p}_n + \tilde{p}_n \log\left(\frac{n}{\tilde{p}_n}\right) + (n - \tilde{p}_n) \left(\frac{\tilde{p}_n}{n}\right)^{2a} \sqrt{\log\left(\frac{n}{\tilde{p}_n}\right)}
\]

which, for \( \alpha \geq 1 \) and \( \frac{1}{2} \leq \alpha < 1 \), will be at most of the order \( p_n \log\left(\frac{p_n}{\tilde{p}_n}\right) \) if \( p_n = o(n) \), because \( \tilde{p}_n \leq p_n \). Since the minimax quadratic risk (2.3) for this problem is always smaller than \( \sup_{\theta \in l_0[p_n]} \sum_{i=1}^{n} E_{\theta_i} (T_\tau(X_i) - \theta_i)^2 \), the stated result follows immediately. \( \blacksquare \)

**Lemma 4.4.** The posterior variance arising out of the general class of shrinkage priors (4.20) can be represented by the following identity:

\[
\text{Var}(\theta|x, \tau) = \frac{T_\tau(x)^2}{x} - (T_\tau(x) - x)^2 + x^2 \int_{0}^{\infty} \frac{1}{(1 + \tau^{2a})^{1/2}} \frac{t^{-a-1} L(t) e^{-2a^2 t^{2a+1}/2}}{t^{2a+1} L(t) e^{-2a^2 t^{2a+1}/2}} dt
\]

which can be bounded from above by

1. \( \text{Var}(\theta|x, \tau) \leq 1 + x^2 \).
2. \( \text{Var}(\theta|x, \tau) \leq (\frac{1}{2} + x) T_\tau(x) - T_\tau(x)^2 \).
Proof. By the law of iterated variance it follows that

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

which can equivalently be represented as by the following identity as well:

\[ \text{Var}(\theta|x, \tau) = \frac{T_\tau(x)}{x} - \left(\frac{T_\tau(x)}{x}\right)^2 + x^2 \int_0^\infty \int_0^\infty \frac{1}{(1 + t^2)^{1/2}} t^{-a-1} L(t) e^{-\frac{x^2 t^2}{2(1 + t^2)^{1/2}}} dt \]

Lemma 4.5. Suppose \( J(x, \tau) = x^2 \int_0^\infty \int_0^\infty \frac{t^{1/2} \theta^2}{(1 + t^2)^{3/2}} t^{-a-1} L(t) e^{-\frac{x^2 t^2}{2(1 + t^2)^{1/2}}} dt \). Then for any \( \rho > 0 \),

\[ \int_{-\infty}^{\infty} J(x, \tau) \phi(x) dx \lesssim \tau^4 + \tau^2 a + \tau^2 \zeta \tau, \]

where the function \( L \) in the above expression is already defined in (2.5) and satisfies Assumption (2.1), with \( \frac{1}{2} \leq a < 1 \) and \( \zeta = \sqrt{2 \log(\frac{1}{a})} \).

Proof. Since \( L \) is slowly varying, there exists some \( A_0 > 1 \) such that \( L \) is bounded over every compact subsets of \([A_0, \infty)\) and

\[ \lim_{x \to \infty} \int_0^{A_0} t^{-a} L(t) dt = \frac{1}{1 - a} \]

for \( 0 < a < 1 \).

Let us now split the term \( J(x, \tau) \) as follows:

\[ J(x, \tau) = J_1(x, \tau) + J_2(x, \tau) + J_3(x, \tau) \]  

where

\[ J_1(x, \tau) = x^2 \int_0^{A_0} \frac{(t^2)^{1/2} \theta^2}{(1 + t^2)^{3/2}} t^{-a-1} L(t) e^{-\frac{x^2 t^2}{2(1 + t^2)^{1/2}}} dt \]

\[ J_2(x, \tau) = x^2 \int_0^\infty \int_0^{A_0} \frac{t^{1/2} \theta^2}{(1 + t^2)^{3/2}} t^{-a-1} L(t) e^{-\frac{x^2 t^2}{2(1 + t^2)^{1/2}}} dt \]

\[ J_3(x, \tau) = x^2 \int_0^\infty \int_{A_0}^\infty \frac{t^{1/2} \theta^2}{(1 + t^2)^{3/2}} t^{-a-1} L(t) e^{-\frac{x^2 t^2}{2(1 + t^2)^{1/2}}} dt \]
Now we give upper bounds to the terms \( J_1(x, \tau) \), \( J_2(x, \tau) \) and \( J_3(x, \tau) \) as follows:

Note that

\[
J_1(x, \tau) \leq x^2 A_0^{-2} e^{\frac{A_0 x^2}{2(1 + t \tau^2)}} \int_0^\infty \frac{t^{-a} L(t) dt}{1 + t \tau^2} \leq A_0^{-2} x^2 e^{\frac{A_0 x^2}{2}} (1 + o(1))
\]

where we used the fact that \( \int_0^\infty (1 + t \tau^2)^{-1/2} t^{-a} L(t) dt = \int_0^\infty t^{-a-1} L(t) dt (1 + o(1)) \), the \( o(1) \) term being independent of \( x \) and is such that \( \lim_{\tau \to 0} o(1) = 0 \).

Therefore,

\[
\int_{\rho \tau}^{\rho_0 \tau} J_1(x, \tau) \phi(x) dx \leq A_0^{-2} \tau^4 x^2 \int_{\rho \tau}^{\rho_0 \tau} e^{\frac{A_0 x^2}{2}} (1 + o(1)) \leq A_0^{-2} \tau^4 (1 + A_0 \tau^2)^{3/2} (1 + o(1)) \lesssim \tau^4
\]  

(4.20)

Next,

\[
J_2(x, \tau) \leq \tau^4 x^2 \int_0^\infty A_0^{-2} x^2 e^{\frac{A_0 x^2}{2}} L(t) dt \leq \frac{K}{2-a} \tau^4 x^2 e^{\frac{A_0 x^2}{2}} L(\frac{A_0}{\tau^2})(1 + o(1)) as \tau \to 0
\]

\[
= \frac{K}{2-a} A_0^{-2} x^2 \int_0^\infty e^{\frac{A_0 x^2}{2}} L(t) dt (1 + o(1)) as \tau \to 0
\]

\[
\leq \frac{K M}{2-a} A_0^{-2} x^2 \int_0^\infty e^{\frac{A_0 x^2}{2}} L(t) dt (1 + o(1)) as \tau \to 0
\]

(4.21)

since \( \int_0^\infty t^{-a} L(t) dt = \left( \frac{A_0}{2-a} \right)^{2-a} (1 + o(1)) \), which follows from Karamata’s integral theorem as \( 0 < a < 1; \int_0^\infty \frac{1}{t^{a-1}} dt = K^{-1}(1 + o(1)) \) and \( L(\frac{A_0}{2-a}) = L(\frac{2-a}{2-a})(1 + o(1)) \) as \( \tau \to 0 \), the \( o(1) \) term being independent of \( x \) and is such that \( \lim_{\tau \to 0} o(1) = 0 \).

Hence

\[
\int_{\rho \tau}^{\rho_0 \tau} J_2(x, \tau) \phi(x) dx \lesssim \tau^{2a}
\]

Finally observe that

\[
J_3(x, \tau) \leq K M x^2 e^{\frac{x^2}{2}} \int_0^\infty \frac{e^{-\frac{t}{x}} t^{-a-1} dt}{(1 + t \tau^2)^{1/2}} (1 + o(1)) as \tau \to 0
\]

\[
= K M x^2 e^{\frac{x^2}{2}} \int_0^\infty e^{-\frac{t}{x}} t^{-a} t^{1/2} \left( \frac{1}{\tau^2} \left( \frac{x^2}{u} - 1 \right) \right)^{-a-1} du (1 + o(1)) as \tau \to 0
\]

\[
= K M x^2 e^{\frac{x^2}{2}} \int_0^\infty e^{-\frac{t}{x}} u^{a+1/2} \left( 1 - \frac{u}{x^2} \right)^{-a-1} du (1 + o(1)) as \tau \to 0
\]

\[
\leq K M x^2 e^{\frac{x^2}{2}} \int_0^\infty e^{-\frac{t}{x}} u^{a+1/2} \left( 1 - \frac{u}{x^2} \right)^{-a-1} du (1 + o(1)) as \tau \to 0 [since A_0 > 1]
\]
Note that \( 0 < u < \frac{x^2}{2} \Rightarrow 0 < \frac{u}{x} < \frac{1}{2} < 1 \Rightarrow \frac{1}{2} < 1 - \frac{u}{x} < 1 \). Also, as \( \frac{1}{2} \leq a < 1 \) we have \( u^{a-\frac{1}{2}} \leq (\frac{x}{2})^{a-\frac{1}{2}} \) for all \( u \leq \frac{x}{2} \). It therefore, follows that,

\[
J_3(x, \tau) \leq 2^{a+1} K M \tau^{2a} x^{1-2a} \left( \frac{x^2}{2} \right)^{a-\frac{1}{2}} \int_0^{\frac{\pi}{2}} e^{-\frac{x^2}{2}} du(1 + o(1)) \text{ as } \tau \to 0
\]

\[
\leq 2 \sqrt{2} K M \tau^{2a} e^{\frac{x^2}{2}} \int_0^\infty e^{-\frac{x^2}{2}} du(1 + o(1)) \text{ as } \tau \to 0
\]

\[
= 4 \sqrt{2} K M \tau^{2a} e^{\frac{x^2}{2}} (1 + o(1)) \text{ as } \tau \to 0
\]

so that

\[
\int_{-\rho \zeta}^{\rho \zeta} J_3(x, \tau) \phi(x) dx \lesssim \tau^{2a} \zeta
\]

(4.22)

Combining equations (4.19) - (4.22) the result follows immediately.

**Proof of Theorem 3.2**

**Proof**. Suppose that \( X \sim N_n(\theta, I_n) \), \( \theta \in l_0[p_n] \) and \( \tilde{p}_n = \#\{i : \theta_i \neq 0\} \). Note that \( \tilde{p}_n \leq p_n \). Assume without any loss of generality that for \( i = 1, \ldots, p_n, \theta_i \neq 0 \), while for \( i = \tilde{p}_n + 1, \ldots, n, \theta_i = 0 \). Let \( \zeta = \sqrt{2 \log(\frac{1}{\tau \pi})} \).

**Nonzero means:**

By applying the same reasoning as in Lemma 4.3 to the final term of \( \text{Var}(\theta|x) \) in (4.17), there exists a non-negative real-valued function \( h(x, \tau) \) such that \( \text{Var}(\theta|x) \leq h(x, \tau) \), where \( h(x, \tau) \to 1 \) as \( x \to \infty \) for any fixed \( \tau \in (0, 1) \). If \( \tau \to 0 \), the function \( h(x, \tau) \) satisfies the following for any \( c > 1 \):

\[
\lim_{\tau \downarrow 0} \sup_{|x| > \rho \zeta} h(x, \tau) = 1 \text{ for all } \rho > c.
\]

Hence \( \text{Var}(\theta|x) \lesssim 1 \), for any \( x > \zeta \) as \( \tau \to 0 \). Now suppose \( x \leq \zeta \). Then by the bound \( \text{Var}(\theta|x) \leq 1 + x^2 \) from Lemma 4.4, we find:

\[
\text{Var}(\theta|x) \leq 1 + \zeta^2.
\]

(4.23)

Therefore:

\[
\sum_{i=1}^{\tilde{p}_n} E_\theta \text{Var}(\theta_i|X_i) \lesssim \tilde{p}_n (1 + \zeta^2).
\]

(4.24)

**Zero means:**

By the bound \( \text{Var}(\theta|x) \leq 1 + x^2 \), we find for any \( \rho \geq 1 \):

\[
E_\theta \text{Var}(\theta|X) 1_{\{|X| > \rho \zeta\}} \leq 2 \int_{\rho \zeta}^\infty (1 + x^2) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx
\]

\[
\leq \frac{\tau^{2a}}{\zeta} + \zeta \tau^{2a}.
\]

(4.25)

When \( |x| < \rho \zeta \), we consider the upper bound \( \text{Var}(\theta|x) \leq \frac{T_n(x)}{x} + J(x, \tau) \) from Lemma 4.4, where the term \( J(x, \tau) \) denote simply the third term on the right hand side of (4.13). Again using the upper bound given in Lemma 4.1 it follows that:

\[
\int_{-\rho \zeta}^{\rho \zeta} \frac{T_n(x)}{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \lesssim \tau^{2a} \zeta.
\]

(4.26)
Therefore, using equations (4.25) and (4.26) and Lemma 4.5, it follows that:

\[ \sum_{i=\rho_n+1}^{n} E_0 \text{Var}(\theta_i|X_i) \lesssim (n - \rho_n)(\zeta_{\tau} + \tau^{4-2\alpha} + 1)\tau^{2\alpha}. \]  

(4.27)

From equations (4.24) and (4.27) it finally follows that:

\[ E_0 \sum_{i=1}^{n} \text{Var}(\theta_i|X_i) \lesssim \rho_n(1 + \zeta_{\tau}^2) + (n - \rho_n)(\zeta_{\tau} + \tau^{4-2\alpha} + 1)\tau^{2\alpha}. \]

Putting \( \tau = (\frac{\rho_n}{n})^\alpha \), we obtain:

\[ E_0 \sum_{i=1}^{n} \text{Var}(\theta_i|X_i) \lesssim \rho_n(1 + \log(\frac{n}{\rho_n})) + \left( \frac{\rho_n}{n} \right)^{2\alpha} (n - \rho_n) \left( \sqrt{\log(\frac{n}{\rho_n})} + \left( \frac{\rho_n}{n} \right)^{(4-2\alpha)\alpha} + 1 \right) \]

which will be at most of the order \( \rho_n \log(\frac{n}{\rho_n}) \) as \( n \to \infty \) for \( \frac{1}{2} \leq a < 1 \) and \( \alpha \geq 1 \), if \( \rho_n = o(n) \). ■

**Lemma 4.6.** Suppose the function \( L(\cdot) \) given by (2.3) satisfies Assumption 2.1 and is non-decreasing over \((0, \infty)\), with \( a = \frac{1}{2} \). Let us define for fixed \( y \geq 0 \) and for fixed \( k > 0 \),

\[ I_k = \int_0^\infty \frac{(t\tau^2)^{k-\frac{1}{2}}}{(1 + t\tau^2)^{\frac{3}{2}}} t^{-3/2} L(t)e^{\frac{t\tau^2}{1+t\tau^2}y} dt. \]

Then,

\[ I_\frac{1}{2} \geq L(1) \tau \left[ \frac{\tau}{y} (e^{\tau^2} - e^{-\tau^2}) + \frac{1}{\sqrt{2y}} (e^\tau - e^{-\tau}) \right], \text{ for } \tau < \frac{1}{\sqrt{2}} \]

\[ I_\frac{1}{2} \leq \left[ e^{\tau^2} + 2 Me^{-\frac{1}{\sqrt{2}}} e^{\frac{1}{\sqrt{2}}} (1 - \frac{1}{\sqrt{2}}) + 2M \sqrt{y} (e^\tau - e^{-\tau}) \right], \text{ for } \tau < \frac{1}{\sqrt{2}} \]

\[ I_\frac{1}{2} \leq M \tau \left[ e^{\tau^2} + 2 e^{\frac{1}{\sqrt{2}}} (1 - \frac{1}{\sqrt{2}}) + \sqrt{y} (e^\tau - e^{-\tau}) \right], \text{ for } \tau < \frac{1}{\sqrt{2}} \]

\[ I_\frac{1}{2} \geq L(1) \tau \left[ e^{\tau^2} (1 - \frac{1}{\sqrt{2}}) + \sqrt{y} (e^\tau - e^{-\tau}) + \frac{1}{2y} (e^\tau - e^{-\tau}) \right], \text{ for } \tau < \frac{1}{\sqrt{2}} \]

**Proof.** Note that since \( L \) is non-decreasing over \((0, \infty)\), \( L(t) \geq L(1) \) for all \( t \geq 1 \). Therefore,

\[ I_\frac{1}{2} = \int_0^\infty \frac{(t\tau^2)^2}{(1 + t\tau^2)^{5/2}} t^{-3/2} L(t)e^{\frac{t\tau^2}{1+t\tau^2}y} dt \]

\[ = \frac{1}{\tau} \int_0^\infty \frac{(t^2 + (t\tau^2))^2}{(1 + t^2)^{5/2}} t^{-3/2} L(t)e^{\frac{t^2}{1+t^2}y} dt \]

\[ \geq \frac{L(1)}{\tau} \int_1^\infty \frac{(t^2 + (t\tau^2))^2}{(1 + t^2)^{5/2}} t^{-2} e^{\frac{t^2}{1+t^2}y} dt \]  

(4.28)

Now putting \( u = \frac{t^2}{1+t^2} \) in (4.28) we obtain,

\[ I_\frac{1}{2} \geq L(1) \tau \int_{t^2}^{1} u^{1/2} e^{uy} du \]

\[ \geq L(1) \tau \int_{t^2}^{1} u^{1/2} e^{uy} du \text{[since } \tau^2 < 1] \]

\[ = L(1) \tau \left[ \frac{\tau}{y} (e^{\tau^2} - e^{-\tau^2}) + \frac{1}{\sqrt{2y}} (e^\tau - e^{-\tau}) \right], \text{ for } \tau < \frac{1}{\sqrt{2}} \]
where the last equality follows using the same set arguments in the proof of Lemma (A.1) of van der Pas et al. (2014).

Next observe that
\[ I_+ = \int_0^\infty \frac{1}{(1 + t\tau^2)^{1/2}} t^{-3/2} L(t)e^{t\tau^2} dt \]
\[ = \frac{1}{\tau} \int_0^\infty \tau^{-3/2} L(\frac{t}{\tau})e^{\frac{t\tau^2}{\tau^2}} dt \]
\[ = \tau \int_0^1 u^{-3/2} L(\frac{1}{\tau} 1 - \frac{u}{1 - u})e^{uy} du \]
\[ = \tau \left[ \int_0^1 u^{-3/2} L(\frac{1}{\tau} 1 - \frac{u}{1 - u})e^{uy} du + \int_1^\infty u^{-3/2} L(\frac{1}{\tau} 1 - \frac{u}{1 - u})e^{uy} du \right] \quad (4.29) \]

Now observe that \( e^{uy} \leq e^{t^2y} \) for all \( u \leq \tau^2 \). Using this fact and applying the change of variable \( t = \frac{1}{\tau} \frac{u}{1 - u} \) in the first integral on the right hand side of (4.29), we obtain,
\[ \int_0^{\tau^2} u^{-3/2} L(\frac{1}{\tau^2} 1 - \frac{u}{1 - u})e^{uy} du \leq e^{t^2y} \int_0^{\tau^2} u^{-3/2} L(\frac{1}{\tau^2} 1 - \frac{u}{1 - u}) du \]
\[ = e^{t^2y} \int_0^1 \frac{1}{\tau^2} L(t)dt \]
\[ \leq e^{t^2y} \int_0^\infty \frac{1}{\tau^2} L(t)dt \quad \text{[since \( \frac{1}{\sqrt{1 + t\tau^2}} \leq 1 \)]} \]
\[ = \frac{K^{-1}e^{t^2y}}{\tau} \quad (4.30) \]

For the second integral on the right hand side of (4.29), we observe that the function \( L \) is uniformly bounded by the constant \( M > 0 \). Using this observation and then apply the same arguments given in the proof of Lemma (A.1) of van der Pas et al. (2014), we obtain,
\[ \int_1^\infty u^{-3/2} L(\frac{1}{\tau^2} 1 - \frac{u}{1 - u})e^{uy} du \leq 2M e^{t^2y} \left( \frac{1}{\tau} - \frac{1}{\sqrt{\tau}} \right) + e^{\frac{t^2y}{2}} \left( \frac{1}{\sqrt{\tau}} - \sqrt{2} \right) + \frac{\sqrt{2}}{y} (e^y - e^{\frac{t^2y}{2}}) \quad \text{for} \quad \tau < \frac{1}{2} \]
\[ (4.29), (4.30) \quad \text{and} \quad (4.31) \]

together immediately give the stated upper bound on \( I_+ \).

Again note that \( \tau^2 < u < 1 \implies \frac{1}{\tau^2(1 - u)} > \frac{1}{1 - \tau} \) which is strictly greater than 1. Hence \( L(\frac{1}{\tau^2(1 - u)}) \geq L(\frac{1}{1 - \tau}) \geq L(1) \) as \( L \) is nondecreasing. Using this observation and (4.29) and then applying the same reasoning given in the proof of Lemma (A.1) of van der Pas et al. (2014), we obtain,
\[ I_+ \geq \tau \int_0^{\tau^2} u^{-3/2} L(\frac{1}{\tau^2} 1 - \frac{u}{1 - u})e^{uy} du \]
\[ \geq L(1)\tau \int_0^{\tau^2} u^{-3/2} e^{uy} du \]
\[ = L(1)\tau \left[ e^{t^2y} \left( \frac{1}{\tau} - \frac{1}{\sqrt{\tau}} \right) + \frac{\sqrt{2}}{y} (e^y - e^{\frac{t^2y}{2}}) \right] \quad \text{for} \quad \tau < \frac{1}{2}. \]

Stated upper bound for \( I_+ \) follows immediately by noting that \( L \) is uniformly bounded by the constant \( M > 0 \) and subsequently by change the variable \( u = \frac{t^2y}{\tau^2} \) followed by the same set of arguments given in the proof of Lemma (A.1) of van der Pas et al. (2014), thereby completing the proof of Lemma 4.6. ■
Proof of Theorem 3.4

Proof. From (4.18) we have,

\[\text{Var}(\theta|x, \tau) \geq x^2 E \left[(1 - \kappa)^2|x, \tau\right] - x^2 E^2 \left[(1 - \kappa)|x, \tau\right]\]

\[= x^2 \left[\int_0^\infty \frac{t^2}{(1+t^2)^{3/2}} L(t) e^{-\frac{x^2}{2(1+t^2)}} dt - \int_0^\infty \frac{1}{(1+t^2)^{3/2}} \left(t^{-3/2} L(t) e^{-\frac{x^2}{2(1+t^2)}} dt \right)^2\right]\]

\[= 2y \left[I_2 - \left(I_2 \right)^2\right]\]

by putting \(y = \frac{x^2}{2}\) and rest of the proof follows by applying Lemma 4.6 and the same set of arguments given in the proof of Theorem (3.4) of van der Pas et al (2014).

4.2 Appendix B: Some Properties of Slowly Varying Functions

Lemma 4.7. If \(L\) is any slowly varying function then there exists \(A_0 > 0\) such that \(L\) is locally bounded in \([A_0, \infty)\), that is, \(L\) is bounded in all compact subsets of \([A_0, \infty)\).

Proof. See Lemma 1.3.2 and subsequent discussion in Bingham et al (1987).

Lemma 4.8. If \(L\) is any slowly varying function, \(A_0\) is so large such that \(L\) is locally bounded in \([A_0, \infty)\) and \(\alpha > -1\), then

\[\int_{x}^{\infty} t^\alpha L(t) dt \rightarrow \frac{1}{1 + \alpha} \text{ as } x \rightarrow \infty.\]

Proof. See Proposition 1.5.8 of Bingham et al (1987).

Lemma 4.9. If \(L\) is any slowly varying function and \(\alpha < -1\), then

\[\int_{x}^{\infty} t^\alpha L(t) dt \rightarrow \frac{-1}{\alpha + 1} \text{ as } x \rightarrow \infty.\]

Proof. See Proposition 1.5.10 of Bingham et al (1987).

5 Discussion

We studied in this paper various theoretical properties of a general class of heavy-tailed shrinkage priors in terms of the quadratic minimax risk for estimating a multivariate normal mean vector which is known to be sparse in the sense of being nearly black. It is shown that Bayes estimators arising out of this general class asymptotically attain the minimax risk in the \(l_2\) norm possibly up to some multiplicative constants. Optimal rate of posterior contraction of these prior distributions in terms of the corresponding quadratic minimax rate has also been established. We provided a unifying theoretical treatment through exploiting properties of slowly varying functions that holds for a very broad class of shrinkage priors including some well-known prior distributions such as the horseshoe prior, the normal-exponential-gamma priors, the three parameter beta normal priors, the generalized double Pareto priors, the inverse gamma priors and many others. Another major contribution of this work is to show that shrinkage priors which are heavy-tailed and have sufficient mass around the origin, are good enough in order to attain the minimax optimal rate of contraction and one does not require a pole at the origin, provided that the global tuning parameter is carefully chosen, a question that was posed in van der Pas et al (2014). We observed that (though not reported in this paper) when the number of non-zero means is unknown, the Bayes estimators based on this general class of one-group priors, combined with the empirical
Bayes estimate of the global shrinkage parameter as suggested in van der Pas et al (2014), still attain the minimax risk upto a multiplicative constant in the $l_2$ norm. In this sense, our work can be considered as a full extension of the posterior concentration properties for the horseshoe prior obtained by van der Pas et al (2014) over a very large class of global-local scale mixture of normals. Moreover, the range $[\frac{1}{2}, 1)$ of the hyperparameter $a$ used in the definition of this general class is in concordance with that obtained in the context of multiple testing considered in Ghosh et al (2014). Therefore, the results obtained in this paper can be thought of as another theoretical justification for the use of such priors. However, a more interesting question would be to investigate whether the posterior contraction properties of such prior distributions still remain when a hyperprior is placed over the global shrinkage parameter $\tau$. We hope to address this problem elsewhere in future.

Acknowledgement The authors would like to thank Professor Jayanta Kumar Ghosh for letting them aware about the recent work of van der Pas et al (2014) on the posterior contraction properties of the horseshoe prior.

References

Armagan, A., Dunson, D. B., and Clyde, M. (2011). “Generalized Beta Mixtures of Gaussians.” In Shawe-Taylor, J., Zemel, R. S., Bartlett, P. L., Pereira, F. C. N, and Weinberger, K. Q. (eds.), Advances in Neural Information Processing Systems 24, 523-531.

Armagan, A., Dunson, D. B., and Lee, J. (2012). “Generalized double Pareto Shrinkage.” Statistica Sinica, 23(1): 119-143.

Armagan, A., Dunson, D. B., Lee, J., Bajwa, W. U. and Strawn, N.(2013). “Posterior consistency in linear models under shrinkage priors.” Biometrika, 100(4): 1011-1018.

Bhattacharya, A., Pati, D., Pillai, N., and Dunson, D. (2012). “Dirichlet-Laplace priors for optimal shrinkage” Arxiv preprint arXiv:1401.5398v1.

Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1987). Regular Variation. In Rota, G.-C. (ed.), Encyclopedia of mathematics and its applications, vol 27, University Press, Cambridge, Great Britain.

Carvalho, C., Polson, N., and Scott, J. (2009). “Handling sparsity via the horseshoe.” Journal of Machine Learning Research W&CP, 5: 73-80.

Carvalho, C., Polson, N., and Scott, J. (2010). “The horseshoe estimator for sparse signals.” Biometrika, 97(2): 465-480.

Castillo, I., Schmidt-Heiber, J., and van der Vaart, A.(2014). “Bayesian Linear Regression with Sparse Priors” arXiv:1403.0735.

Castillo, I. and van der Vaart, A.(2012). “Needles and straw in a haystack: Posterior Concentration For Possibly Sparse Sequences” The Annals of Statistics, 40(4): 2069-2101.

Datta, J. and Ghosh, J. K. (2013). “Asymptotic Properties of Bayes Risk for the horseshoe Prior.” Bayesian Analysis, 8(1): 111-132.

Donoho, D. L., Johnstone, I. M., Hoch, J. C. and Stern, A. S. (1992). “Maximum Entropy and the Nearly Black Object (with Discussion).” Journal of the Royal Statistical Society. Series B (Methodological), 54: 41–81.

Efron, B. (2004). “Large-scale simultaneous hypothesis testing: The choice of a null hypothesis.” Journal of the American Statistical Association, 99(465): 96-104.
Efron, B. (2008). “Microarrays, Empirical Bayes and the two-groups Model.” *Statistical Science*, 23(1): 1-22.

Gelman, A. (2006). “Prior distributions for variance parameters in hierarchical models.” *Bayesian Analysis*, 1(3): 515-533.

Ghosal, S., Ghosh, J. K., and van der Vaart, A. W. (2000). “Convergence Rates of Posterior Distribution.” *The Annals of Statistics*, 28(2): 500-531.

Ghosh, P., Tang, X., Ghosh, M. and Chakrabarti, A. (2014). “Asymptotic Properties of Bayes Risk for a General Class of Shrinkage Priors in Multiple Hypothesis Testing Under Sparsity.” *arXiv:1310.7462 (Submitted)*.

Griffin, J. E. and Brown, P. J. (2005). “Alternative prior distributions for variable selection with very many more variables than observations.” Technical report, University of Warwick.

Griffin, J. E. and Brown, P. J. (2010). “Inference with normal-gamma prior distributions in regression problems.” *Bayesian Analysis*, 5(1): 171-188.

Griffin, J. E. and Brown, P. J. (2012). “Structuring shrinkage: some correlated priors for regression.” *Biometrika*, 99(2): 481-487.

Griffin, J. E. and Brown, P. J. (2013). “Some priors for sparse regression modeling.” *Bayesian Analysis*, 8(3): 691-702.

Ghosal, S., Ghosh, J. K., and van der Vaart, A. W. (2000). “Convergence Rates of Posterior Distribution.” *The Annals of Statistics*, 28(2): 500-531.

Hans, C. (2009). “Bayesian lasso regression.” *Biometrika*, 96(4): 835-845.

Jiang, W., and Zhang, C. H. (2009). “General Maximum Likelihood Empirical Bayes Estimation of Normal Means.” *The Annals of Statistics*, 37: 1647–1684.

Johnstone, I. and Silverman, B. W. (2004). “Needles and straw in haystacks: Empirical-Bayes estimates of possibly sparse sequences.” *The Annals of Statistics*, 32(4): 1594-1649.

Mitchell, T. and Beauchamp, J. (1988). “Bayesian variable selection in linear regression (with discussion).” *Journal of the American Statistical Association*, 83(404): 1023-1036.

Park, T. and Casella, G. (2008). “The Bayesian lasso.” *Journal of the American Statistical Association*, 103(482): 681-686.

van der Pas, S. L., Kleijn, B. J. K. and van der Vaart, A. W. (2014). “The horseshoe estimator: Posterior concentration around nearly black vectors.” *arXiv:1404.0202 (Preprint)*.

Polson, N. G. and Scott., J. G. (2010). “Large Scale Simultaneous Testing with Hypergeometric Inverted Beta Priors.” Technical report.

Polson, N. G. and Scott., J. G. (2011). “Shrink Globally, Act Locally: Sparse Bayesian Regularization and Prediction.” In Bernardo, J. M., Bayarri, J. M., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M. (eds.), *Bayesian Statistics 9, Proceedings of the 9th Valencia International Meeting*, 501-538. Oxford University Press.

Polson, N. G. and Scott., J. G. (2012). “On the Half-Cauchy Prior for a Global Scale parameter.” *Bayesian Analysis*, 7(2): 1-16.

Scott, J. G. (2011). “Bayesian estimation of intensity surfaces on the sphere via needlet shrinkage and selection.” *Bayesian Analysis*, 6(2): 307-327.

Tipping, M. (2001). “Sparse Bayesian learning and the Relevance Vector Machine.” *Journal of Machine Learning Research*, 1: 211-244.

Yuan, M. and Lin,Y.(2005). “Efficient Empirical Bayes Variable Selection and Estimation in Linear Models.” *Journal of the American Statistical Association*, 100: 1215-1225.