

ORIGINAL ARTICLE OPEN ACCESS

# Concentration of a Sparse Bayesian Model With Horseshoe Prior in Estimating High-Dimensional Precision Matrix

The Tien Mai 

Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway

**Correspondence:** The Tien Mai ([the.t.mai@ntnu.no](mailto:the.t.mai@ntnu.no))**Received:** 4 August 2024 | **Revised:** 28 August 2024 | **Accepted:** 4 September 2024**Keywords:** high-dimensional data | horseshoe prior | misspecified model | posterior concentration rate | precision matrix | sparsity

## ABSTRACT

Precision matrices are crucial in many fields such as social networks, neuroscience and economics, representing the edge structure of Gaussian graphical models (GGMs), where a zero in an off-diagonal position of the precision matrix indicates conditional independence between nodes. In high-dimensional settings where the dimension of the precision matrix  $p$  exceeds the sample size  $n$  and the matrix is sparse, methods like graphical Lasso, graphical SCAD and CLIME are popular for estimating GGMs. While frequentist methods are well-studied, Bayesian approaches for (unstructured) sparse precision matrices are less explored. The graphical horseshoe estimate, applying the global-local horseshoe prior, shows superior empirical performance, but theoretical work for sparse precision matrix estimations using shrinkage priors is limited. This paper addresses these gaps by providing concentration results for the tempered posterior with the fully specified horseshoe prior in high-dimensional settings. Moreover, we also provide novel theoretical results for model misspecification, offering a general oracle inequality for the posterior. A concise set of simulations is performed to validate our theoretical findings.

## 1 | Introduction

Precision matrices play a pivotal role in numerous fields, including social networks, neuroscience and economics (Pourahmadi 2011; Ryali et al. 2012; Fan, Liao, and Liu 2016; Callot et al. 2021). The configuration of the precision matrix  $\Omega = \Sigma^{-1} \in \mathbb{R}^{p \times p}$  represents the edge structure of a Gaussian graphical model (GGM), where the nodes are normally distributed as  $\mathcal{N}(0, \Sigma)$  (Lauritzen 1996). GGMs continue to be favoured for network estimation due to the straightforward interpretation of the precision matrix: a zero in an off-diagonal position indicates conditional independence between the two corresponding nodes, given all other nodes. Due to these important implications, the estimation of precision matrices has received substantial attention over the past few decades.

In practical applications, users often encounter situations where the dimension  $p$  of the precision matrix is comparable to or even exceeds the sample size  $n$ , and the precision matrix

exhibits sparsity. In such contexts, leveraging the sparsity of the precision matrix is crucial to discern the edge structure of the graph and assess the conditional dependencies among the nodes. Among the most widely used frequentist methods for estimating GGMs are the graphical lasso (Friedman, Hastie, and Tibshirani 2008), the graphical SCAD (Fan, Feng, and Wu 2009), and the CLIME estimator (Cai, Liu, and Luo 2011). For a comprehensive overview, see Cai, Ren, and Zhou (2016) and Janková and van de Geer (2017). These methods provide estimates of high-dimensional inverse covariance matrices under various sparsity patterns.

The body of work on Bayesian methodologies for *unstructured precision matrices* is relatively limited. Wang (2012) introduced a Bayesian variant of the graphical lasso, utilising block Gibbs sampling. Banerjee and Ghosal (2015) employed a similar prior structure to the Bayesian graphical lasso but included a substantial point-mass at zero for the off-diagonal elements of the precision matrix. They derived posterior convergence

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Author(s). *Stat* published by John Wiley & Sons Ltd.

rates in the Frobenius norm under sparsity assumptions and proposed a Laplace approximation method for calculating marginal posterior probabilities of models. Additionally, Gan, Narisetty, and Liang (2019) proposed spike-and-slab variants with double exponential priors. Atchadé (2019) combined the spike-and-slab prior with a pseudo-likelihood approach, resulting in a quasi-posterior distribution with favourable contraction properties.

A common issue with the spike-and-slab approach is the presence of binary indicator variables, which can hinder posterior exploration. The graphical horseshoe method, introduced by Li, Craig, and Bhadra (2019), addresses these issues by applying the global-local horseshoe prior (Carvalho, Polson, and Scott 2010) to GGMs. Li, Craig, and Bhadra (2019) provided substantial empirical evidence of the graphical horseshoe estimate's superior performance compared to several competing Bayesian and frequentist methods in both low- and high-dimensional settings. However, there is limited research on the theoretical properties of shrinkage priors for sparse precision matrix estimation. Recent studies by Zhang, Yao, and Ghosh (2022) and Sagar et al. (2024) have adapted several versions of Horseshoe prior in different ways. Specifically, Zhang, Yao, and Ghosh (2022) consider a pseudo-likelihood approach similar to Atchadé (2019), while Sagar et al. (2024) explore a horseshoe-like prior. Moreover, both studies offer theoretical results only in an asymptotic sense and do not address high-dimensional settings. This work provides an attempt to address these gaps in the literature.

In this paper, we provide concentration results for the posterior in estimating precision matrices utilising the horseshoe prior as in Li, Craig, and Bhadra (2019). Following Banerjee and Ghosal (2015), Zhang, Yao, and Ghosh (2022) and Sagar et al. (2024), we assume that the underlying precision matrix is sparse. Importantly, our results are derived in a high-dimensional setting where  $p > n$ . Moreover, while Zhang, Yao, and Ghosh (2022) and Sagar et al. (2024) consider either a horseshoe-like prior or a horseshoe prior with a fixed global shrinkage parameter, our results are obtained for the fully specified horseshoe prior used in practice as in (Li, Craig, and Bhadra 2019). Another novel aspect of our work is that we also provide theoretical results for the case of model misspecification. Specifically, we prove a general oracle inequality for the posterior under an abstract setting that allows for model misspecification. A brief simulation study is conducted to validate our theoretical results.

The rest of the paper comprises the following sections Section 2 presents the problem of sparse precision matrix estimation, a Bayesian approach using tempered posterior and the Horseshoe prior. Section 3 contains our main theoretical results. Several simulations are conducted in Section 4 to further illustrate and validate our theoretical findings. We conclude in Section 5. All technical proofs are gathered in Appendix A.

## 2 | Problem and Method

Notations: Let  $P, R$  be two probability measures and  $\mu$  be any measure such that  $P \ll \mu$  and  $R \ll \mu$ . For  $\alpha \in (0, 1)$ , the

$\alpha$ -Rényi divergence between  $P$  and  $R$  is defined by  $D_\alpha(P, R) = \frac{1}{\alpha-1} \log \int \left(\frac{dP}{d\mu}\right)^\alpha \left(\frac{dR}{d\mu}\right)^{1-\alpha} d\mu$ , and the Kullback-Leibler divergence is defined by  $\mathcal{K}(P, R) = \int \log\left(\frac{dP}{dR}\right) dP$  if  $P \ll R$ , and  $+\infty$  otherwise. Let  $\|\cdot\|_q$  denote the  $\ell_q$ -norm,  $\|\cdot\|_\infty$  denote the max-norm of vectors and let  $\|\cdot\|_0$  denote the  $\ell_0$  (quasi)-norm (the number nonzero entries) of vectors.

### 2.1 | Model

We are interested in estimating the precision matrix  $\Omega \in \mathbb{R}^{p \times p}$  from the sample data  $Y \in \mathbb{R}^{n \times p}$ , where  $n$  denotes the sample size, and  $p$  is the dimension of precision matrix or equivalently the number of nodes in the corresponding graph. The rows of  $Y$  are assumed to be i.i.d.  $\mathcal{N}(0, \Omega^{-1})$  distributed. The likelihood function based on the data can be written as

$$L(Y|\Omega) = (2\pi)^{-\frac{np}{2}} \det(\Omega)^{\frac{n}{2}} \exp\left\{-\frac{1}{2} \text{tr}(Y\Omega Y^T)\right\}. \quad (1)$$

Similar to Atchadé (2019), Zhang, Yao, and Ghosh (2022) and Sagar et al. (2024), we assume that the diagonal elements  $\omega_{ii}$  of  $\Omega$  are known. Without loss of generality, we assume the diagonal  $\omega_{ii} = 1$  for all  $i = 1, \dots, p$ . In this work, we consider high-dimensional setting that  $p > n$ . We assume that the true precision matrix  $\Omega_0$  is sparse in the sense that its number of non-zero off-diagonal elements is small.

### 2.2 | Method

As in Li, Craig, and Bhadra (2019), the horseshoe prior (Carvalho, Polson, and Scott 2010) is independently specified on the elements of  $\Omega$  as

$$\omega_{ij} | \lambda_{ij}, \tau \sim N(0, \lambda_{ij}^2 \tau^2), \quad \lambda_{ij} \sim \text{Ca}_+(0, 1), \quad \tau \sim \text{Ca}_+(0, 1), \quad (2)$$

for  $1 \leq i < j \leq p$ , where  $\text{Ca}_+(0, 1)$  denotes the truncated standard half-Cauchy distribution with density proportional to  $(1+t^2)^{-1} \mathbb{1}_{(0, \infty)}(t)$ . We shall denote this prior induced by the hierarchy in (2) by  $\Pi_{HS}$ .

It is important to note that Zhang, Yao, and Ghosh (2022) present theoretical results exclusively for the case where  $\tau$  is a fixed value. However, in their numerical studies, they use a half-Cauchy prior, as in Li, Craig, and Bhadra (2019). Therefore, our work extends their theoretical framework.

We focus on the tempered posterior given as

$$\Pi_{n,\alpha}(\Omega) \propto L(Y|\Omega)^\alpha \Pi_{HS}(\Omega).$$

for  $\alpha \in (0, 1)$ . It is also called a fractional posterior (Bhattacharya, Pati, and Yang 2019; Alquier and Ridgway 2020). The regular posterior is obtained when  $\alpha = 1$ ; however, using a smaller value in general would be benefit in both computational (Friel and Pettitt 2008) or in model misspecified cases (Grünwald and Van Ommen 2017). It is worth mentioning that tempered/fractional posteriors has received significant attention in recent

years, as in Agapiou and Castillo (2024); (Bissiri, Holmes, and Walker 2016); Chakraborty, Bhattacharya and Mallick (2020), Liu et al. (2021), L'Huillier et al. (2023); (Medina et al. 2022), Syring and Martin (2019), Yonekura and Sugawara (2023) and Yang, Pati, and Bhattacharya (2020).

Put  $\hat{\Omega} = \int \Omega \Pi_{n,\alpha}(d\Omega)$  as our mean estimator.

### 3 | Results

#### 3.1 | Consistency of Tempered Posterior

Let  $p_\Omega$  and  $P_\Omega$  denote the density and the distribution of a Gaussian  $\mathcal{N}(0, \Omega^{-1})$  random variable. Initially, we enumerate our assumptions regarding the truth.

**Assumption 1. (Bounded spectrum).** The true precision matrix  $\Omega_0$  satisfies that there exists a universal constant  $\varepsilon_0 > 0$  that  $0 < \varepsilon_0^{-1} \leq \text{eig}_{\min}(\Omega) \leq \text{eig}_{\max}(\Omega) \leq \varepsilon_0 < \infty$ .

**Assumption 2.** The true precision matrix  $\Omega_0$  is sparse in the sense that  $\Omega_0 \in \mathcal{S}(s) := \left\{ \Omega: \sum_{1 \leq i < j \leq p} \mathbf{1}_{(\omega_{ij} \neq 0)} \leq s \right\}$ .

**Assumption 3. (Sparsity).** It is assumed that  $s < n$ .

Assumption 1 and 2 are standard in both frequentist and Bayesian literature for sparse precision matrix estimation (Cai, Ren, and Zhou 2016; Janková and van de Geer 2017; Banerjee and Ghosal 2015; Sagar et al. 2024; Zhang, Yao, and Ghosh 2022). Assumption 3 is a sparsity condition often assumed in high-dimensional settings (Bellec, Lecué, and Tsybakov 2018; Gao, van der Vaart, and Zhou 2020; Banerjee, Castillo, and Ghosal 2021). We now present first a consistency result of the tempered posterior in  $p > n$  setting.

**Theorem 1.** For any  $\alpha \in (0,1)$ , assume that Assumptions 1–3 are satisfied. Given any  $\Omega_0$  such that  $\|\Omega_0\|_\infty \leq C_1$ , then

$$\mathbb{E} \left[ \int D_\alpha(P_\Omega, P_{\Omega_0}) \Pi_{n,\alpha}(d\Omega) \right] \leq C \frac{1+\alpha}{1-\alpha} \varepsilon_n,$$

where  $\varepsilon_n = \frac{s \log(p/s)}{n}$  and  $C$  is a universal positive constant depending only on  $\varepsilon_0, C_1$ .

Our proof technique leverages the general concentration theory for tempered posteriors as described in Bhattacharya, Pati, and Yang (2019) and Alquier and Ridgway (2020), which ensures the desired convergence with respect to the  $\alpha$ -Rényi divergence. However, in the context of precision matrices, the Frobenius norm is more interpretable than the  $\alpha$ -Rényi divergence. Using results from Banerjee and Ghosal (2015) and Van Erven and Harremos (2014), the results for the Frobenius norm are stated in the following corollary.

Put  $c_\alpha := 1$  for  $\alpha \in [0.5, 1)$  and  $c_\alpha := (1 - \alpha)/\alpha$  for  $\alpha \in (0, 0.5)$ .

**Corollary 3.1.** Assuming that Theorem 1 holds, then we have that

$$\mathbb{E} \left[ \int \|\Omega - \Omega_0\|_2^2 \Pi_{n,\alpha}(d\Omega) \right] \leq C' c_\alpha \frac{1+\alpha}{1-\alpha} \varepsilon_n,$$

and an application of Jensen's inequality,

$$\mathbb{E} \left[ \|\hat{\Omega} - \Omega_0\|_2^2 \right] \leq C' c_\alpha \frac{1+\alpha}{1-\alpha} \varepsilon_n,$$

where  $C' > 0$  is a universal constant depending only on  $\varepsilon_0, C_1$ .

#### 3.2 | Concentration Rates

We now provide our main result concerning the concentration of the tempered posterior relative to the  $\alpha$ -Rényi divergence of the densities.

**Theorem 2.** Given any  $\alpha \in (0,1)$ , assume that Assumptions 1–3 are satisfied. For all  $\Omega_0$  such that  $\|\Omega_0\|_\infty \leq C_1$ , then

$$\mathbb{P} \left[ \int D_\alpha(P_\Omega, P_{\Omega_0}) \Pi_{n,\alpha}(d\Omega) \leq \frac{2(\alpha+1)}{1-\alpha} \varepsilon_n \right] \geq 1 - \frac{2}{n\varepsilon_n},$$

where  $\varepsilon_n = C \frac{s \log(p/s)}{n}$  and  $C$  is a universal positive constant depending only on  $\varepsilon_0, C_1$ .

*Remark 1.* Our results imply that the concentration rate  $\varepsilon_n$  is adaptive to the unknown sparsity level  $s$  of the true precision matrix. The concentration rate of order  $s \log(p/s)/n$  is known to be minimax-optimal in high-dimensional setting (Bellec, Lecué, and Tsybakov 2018; Gao, van der Vaart, and Zhou 2020). Readers may prefer to Banerjee, Castillo, and Ghosal (2021) for a recent review of Bayesian methods in different high-dimensional models. It is noted that our results hold with high probability, specifically depending on  $p$  rather than  $n$ . More precisely, the probability is greater than or equal to  $1 - 2(n\varepsilon_n)^{-1} = 1 - 2(Cs \log(p/s))^{-1}$ .

The same results apply when using the Frobenius norm that is stated in the following corollary.

**Corollary 3.2.** As a special case, Theorem 2 leads to

$$\mathbb{P} \left[ \int \|\Omega - \Omega_0\|_2^2 \Pi_{n,\alpha}(d\Omega) \leq 2c_\alpha \frac{1+\alpha}{1-\alpha} \varepsilon_n \right] \geq 1 - \frac{2}{n\varepsilon_n}, \text{ and}$$

$$\mathbb{P} \left[ \|\hat{\Omega} - \Omega_0\|_2^2 \leq 2c_\alpha \frac{1+\alpha}{1-\alpha} \varepsilon_n \right] \geq 1 - \frac{2}{n\varepsilon_n}.$$

Corollary 3.2 shows that the tempered posterior distribution of  $\Omega$  concentrates around its true value,  $\Omega_0$ , at a specified rate relative to the squared Frobenius norm. As follows with the previous works on tempered posterior (Chakraborty, Bhattacharya, and Mallick 2020; Alquier and Ridgway 2020; Bhattacharya, Pati, and Yang 2019), our results do not require that the true value can be tested against sufficiently separated other values in some suitable sieve as in Sagar et al. (2024) and Zhang, Yao, and Ghosh (2022).

One may note that our tempered posterior estimates are not necessarily positive definite. In some cases, a positive semidefinite (PSD) matrix may be more preferable. To obtain a PSD matrix  $\tilde{\Omega}$ , a postprocessing step from Zhang, Yao, and Ghosh (2022) can be used as follows:

$$\tilde{\Omega} = \arg \min_{\Omega \text{ is PSD}} \|\Omega - \bar{\Omega}\|_2,$$

where  $\bar{\Omega}$  represents random samples from the posterior distribution  $\Pi_{n,\alpha}(\Omega)$ . As shown in (Higham 1988), this optimisation problem has a closed-form solution given by  $\tilde{\Omega} = (B + H)/2$ , where  $H$  is the symmetric polar factor of  $B = (\bar{\Omega} + \bar{\Omega}^T)/2$ . This leads to the inequality  $\|\tilde{\Omega} - \Omega_0\|_2 \leq 2\|\Omega - \Omega_0\|_2$ , which immediately ensures the posterior concentration for the PSD matrix  $\tilde{\Omega}$  from Corollary 3.2, so the proof is omitted and summarised in the following corollary.

**Corollary 3.3.** Assume that Theorem 2 holds, then

$$\mathbb{P} \left[ \int \|\tilde{\Omega} - \Omega_0\|_2^2 \Pi_{n,\alpha}(d\Omega) \leq 2c_\alpha \frac{1+\alpha}{1-\alpha} \varepsilon_n \right] \geq 1 - \frac{2}{n\varepsilon_n}.$$

### 3.3 | Result in the Misspecified Case

In this section, we show that our previous results can be extended to the misspecified setting. Assume that the true data generating distribution is parametrised by  $\Omega_0 \notin S(s)$  and define  $P_{\Omega_0}$  as the true distribution. Put

$$\Omega_* := \arg \min_{\Omega \in S(s)} \mathcal{K}(P_{\Omega_0}, P_\Omega).$$

In order not to change all the notation, we define an extended parameter set  $\{\Omega_0\} \cup S(s)$ . Here, we clearly state that only  $\Omega_*$  is in  $S(s)$ . Example for this circumstance is that  $\Omega_0$  may no longer be sparse.

**Assumption 4.** The minimal precision matrix  $\Omega_*$  satisfies that

$$\Omega_* \in S(s, \varepsilon_0) := \left\{ \Omega: \sum_{1 \leq i < j \leq p} \mathbf{1}_{(\omega_{ij} \neq 0)} \leq s, 0 < \varepsilon_0^{-1} \leq \text{eig}_{\min}(\Omega) \leq \text{eig}_{\max}(\Omega) \leq \varepsilon_0 < \infty \right\}.$$

**Theorem 3.** For any  $\alpha \in (0,1)$ , let assume that Assumptions 4 and 3 hold. For all  $\Omega_0$  such that  $\|\Omega_*\|_\infty \leq C_1$ , then

$$\mathbb{E} \left[ \int D_\alpha(P_{\Omega_0}, P_\Omega) \Pi_{n,\alpha}(d\Omega) \right] \leq \frac{\alpha}{1-\alpha} \min_{\Omega \in S(s, \varepsilon_0)} \mathcal{K}(P_{\Omega_0}, P_\Omega) + \frac{1+\alpha}{1-\alpha} \varepsilon_n,$$

where  $\varepsilon_n = C \frac{s \log(p/s)}{n}$  and  $C$  is a universal positive constant depending only on  $\varepsilon_0, C_1$ .

In Theorem 3, it is important to note that assumptions are made solely regarding  $\Omega_*$  and not  $\Omega_0$ . Consequently, if  $\Omega_0 \in S(s, \varepsilon_0)$ , meaning the model is well-defined and  $\mathcal{K}(P_{\Omega_0}, P_\Omega) = 0$  within  $S(s, \varepsilon_0)$ , then Theorem 1 can be viewed as a specific instance of Theorem 3. Otherwise, the result establishes a general oracle inequality.

Although it is not a sharp oracle inequality due to the differing risk measures on both sides, this observation remains valuable, particularly when  $\mathcal{K}(P_{\Omega_0}, P_{\Omega_*})$  is minimal. Nonetheless, under additional assumptions, we can further derive an oracle

inequality result with  $\ell_2$  distance on both sides. The result is as follows.

**Corollary 3.4.** Assume that Theorem 3 holds. Then, there exist a universal constant  $C' > 0$  depending only on  $\varepsilon_0, C_1$  such that

$$\mathbb{E} \left[ \int \|\Omega - \Omega_0\|_2^2 \Pi_{n,\alpha}(d\Omega) \right] \leq C' \left( \frac{\alpha}{1-\alpha} \min_{\Omega \in S(s, \varepsilon_0)} \|\Omega - \Omega_0\|_2^2 + \frac{1+\alpha}{1-\alpha} \varepsilon_n \right),$$

$$\mathbb{E} \left[ \|\hat{\Omega} - \Omega_0\|_2^2 \right] \leq C' \left( \frac{\alpha}{1-\alpha} \min_{\Omega \in S(s, \varepsilon_0)} \|\Omega - \Omega_0\|_2^2 + \frac{1+\alpha}{1-\alpha} \varepsilon_n \right).$$

To the best of our knowledge, results in Corollary 3.4 are completely novel for estimating sparse precision matrix from a Bayesian perspective.

## 4 | Numerical Examples

In this section, we conduct several simulations to provide a thorough demonstration of our theoretical findings. It is worth noting that more extensive numerical studies involving the graphical horseshoe method, particularly in comparison with other approaches, have been previously presented in the works of Li, Craig, and Bhadra (2019) and Zhang, Yao, and Ghosh (2022). For the purpose of our analysis, we utilise the Gibbs sampling algorithm, as detailed in Li, Craig, and Bhadra (2019), to implement our framework. Additionally, the R code for the graphical horseshoe method can be accessed through the R package ‘GHS’ (Srivastava and Bhadra 2018). Our R code for simulations is available online ([https://github.com/tienmt/f\\_graphical\\_horseshoe](https://github.com/tienmt/f_graphical_horseshoe)).

We consider two distinct settings for the true precision matrix in our analysis:

- Sparse setting: In this scenario, the dimension of the precision matrix is set at  $p = 100$ , with a sample size of  $n = 30$ . The true precision matrix  $\Omega_0$  is characterised by sparsity, having  $s = 86$  nonzero off-diagonal elements.
- Dense setting: Here, the dimension of the precision matrix is  $p = 50$ , with a sample size of  $n = 30$ . The true precision matrix  $\Omega_0$  is defined as the inverse of the covariance matrix  $\Sigma$ , where the diagonal entries of  $\Sigma$  are all equal to 1, and each off-diagonal entry is 0.2, reflecting a denser structure.

In addition to considering the GGM, where the data  $Y_i \sim \mathcal{N}(0, \Omega_0^{-1})$ , we also explore a scenario in which the data are generated from a multivariate Student- $t$  distribution with 3 degrees of freedom, denoted as  $Y_i \sim \text{mvT}_3(0, \Omega_0^{-1})$ . For all settings, the Gibbs sampler is executed for a total of 1100 iterations, with the initial 100 iterations discarded as burn-in steps. Each

**TABLE 1** | Estimation errors are measured using the mean squared error, calculated across 50 simulated repetitions.

Method	Dense, $p = 50, n = 30$		Sparse, $p = 100, n = 30$	
	$Y_i \sim \mathcal{N}(\mathbf{0}, \Omega^{-1})$	$Y_i \sim \text{mvT}_3(\mathbf{0}, \Omega^{-1})$	$Y_i \sim \mathcal{N}(\mathbf{0}, \Omega^{-1})$	$Y_i \sim \text{mvT}_3(\mathbf{0}, \Omega^{-1})$
GHS	0.013 (0.003)	0.009 (0.004)	0.0028 (0.0005)	0.0041 (0.0019)
fGHS $_{\alpha=0.9}$	0.011 (0.003)	0.007 (0.003)	0.0024 (0.0004)	0.0032 (0.0011)
fGHS $_{\alpha=0.5}$	0.003 (0.001)	0.005 (0.001)	0.0025 (0.0003)	0.0031 (0.0006)
fGHS $_{\alpha=0.1}$	0.015 (0.005)	0.006 (0.002)	0.0088 (0.0009)	0.0042 (0.0006)

simulation setting is independently repeated 50 times to ensure robust results.

We consider three distinct values for the tempered/fractional parameter  $\alpha$  in our analysis: 0.9, 0.5 and 0.1. We refer to our proposed method corresponding to these values as ‘fGHS $_{\alpha=0.9}$ ’, ‘fGHS $_{\alpha=0.5}$ ’ and ‘fGHS $_{\alpha=0.1}$ ’, respectively. For comparison, we denote the method from Li, Craig, and Bhadra (2019), which corresponds to  $\alpha = 1$ , as ‘GHS’. The average squared error for estimating the precision matrix, along with the corresponding standard deviations, is reported in Table 1.

The results provided in Table 1 offer a comprehensive evaluation of our proposed method, which leverages the tempered posterior to handle model misspecification effectively. These findings suggest that our method remains robust even when the underlying model is not perfectly specified. Specifically, in situations where the data are generated from the true Gaussian model, our approach exhibits behaviour that closely mirrors that of the conventional Bayesian method, particularly when the tempered/fractional parameter  $\alpha$  is near 1. This similarity underscores the consistency of our method with traditional approaches under ideal conditions.

Moreover, our simulations highlight that certain intermediate values of  $\alpha$ , such as  $\alpha = 0.9$  or  $\alpha = 0.5$ , can potentially enhance performance, yielding better results than when  $\alpha = 1$  is used. This observation points to the potential benefits of fine-tuning the parameter  $\alpha$  to optimise outcomes in practical applications, see (Wu and Martin 2023) for more discussions. However, it is also evident from our studies that selecting a small value of  $\alpha$ , e.g., 0.1, can lead to a noticeable increase in error, indicating that there is a delicate balance to be maintained when choosing this parameter.

Overall, these simulation studies provide clear and convincing evidence in support of our theoretical framework. They not only validate the robustness of our method but also demonstrate its flexibility in adapting to different scenarios, thereby offering valuable insights into its practical utility.

## 5 | Conclusion

This study has investigated the concentration properties of the tempered posterior when estimating sparse precision matrices using the horseshoe prior. Unlike prior studies that employed either a horseshoe-like prior or a simplified horseshoe prior with a fixed shrinkage parameter, our analysis focuses on the fully specified horseshoe prior. We extend the previous theoretical

works by considering a high-dimensional setting ( $p > n$ ). We also present a novel aspect by deriving theoretical results for model misspecification, establishing a general oracle inequality for the posterior in an abstract setting.

We note that our findings are primarily derived using the Frobenius norm, so extending these results to other norms represents a valuable avenue for future research. Another potential gap in our work is that establishing a sharp oracle inequality for the model misspecified case remains an open question for future research.

### Acknowledgements

The author wishes to acknowledge the support provided by the Centre for Geophysical Forecasting, funded by the Norwegian Research Council under grant no. 309960, at NTNU. Special thanks are extended to the associate editor and the anonymous reviewer for their insightful feedback, which greatly improved the manuscript.

### Conflicts of Interest

The author declares no conflicts of interest.

### Data Availability Statement

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

### References

- Agapiou, S., and I. Castillo. 2024. “Heavy-Tailed Bayesian Nonparametric Adaptation.” *Annals of Statistics* 1–28. (to appear) arXiv:2308.04916.
- Alquier, P., and J. Ridgway. 2020. “Concentration of Tempered Posteriors and of Their Variational Approximations.” *Annals of Statistics* 48, no. 3: 1475–1497.
- Atchadé, Y. F. 2019. “Quasi-Bayesian Estimation of Large Gaussian Graphical Models.” *Journal of Multivariate Analysis* 173: 656–671.
- Banerjee, S., I. Castillo, and S. Ghosal. 2021. “Bayesian Inference in High-Dimensional Models.” arXiv preprint arXiv:2101.04491.
- Banerjee, S., and S. Ghosal. 2015. “Bayesian Structure Learning in Graphical Models.” *Journal of Multivariate Analysis* 136: 147–162.
- Bellec, P. C., G. Lecué, and A. B. Tsybakov. 2018. “Slope Meets Lasso: Improved Oracle Bounds and Optimality.” *Annals of Statistics* 46, no. 6B: 3603–3642.
- Bhattacharya, A., D. Pati, and Y. U. N. Yang. 2019. “Bayesian Fractional Posteriors.” *Annals of Statistics* 47, no. 1: 39–66.
- Bissiri, P. G., C. C. Holmes, and S. G. Walker. 2016. “A General Framework for Updating Belief Distributions.” *Journal of the Royal Statistical Society Series B: Statistical Methodology* 78, no. 5: 1103–1130.

- Cai, T., W. Liu, and X. Luo. 2011. "A Constrained l-1 Minimization Approach to Sparse Precision Matrix Estimation." *Journal of the American Statistical Association* 106, no. 494: 594–607.
- Cai, T. T., Z. Ren, and H. H. Zhou. 2016. "Estimating Structured High-Dimensional Covariance and Precision Matrices: Optimal Rates and Adaptive Estimation." *Electronic Journal of Statistics* 10: 1–59.
- Callot, L., M. Caner, A. O. Önder, and E. Ulaşan. 2021. "A Nodewise Regression Approach to Estimating Large Portfolios." *Journal of Business & Economic Statistics* 39, no. 2: 520–531.
- Carvalho, C. M., N. G. Polson, and J. G. Scott. 2010. "The Horseshoe Estimator for Sparse Signals." *Biometrika* 97, no. 2: 465–480.
- Chakraborty, A., A. Bhattacharya, and B. K. Mallick. 2020. "Bayesian Sparse Multiple Regression for Simultaneous Rank Reduction and Variable Selection." *Biometrika* 107, no. 1: 205–221.
- Fan, J., Y. Feng, and Y. Wu. 2009. "Network Exploration Via the Adaptive Lasso and Scad Penalties." *Annals of Applied Statistics* 3, no. 2: 521–541.
- Fan, J., Y. Liao, and H. Liu. 2016. "An Overview of the Estimation of Large Covariance and Precision Matrices." *Econometrics Journal* 19, no. 1: C1–C32.
- Friedman, J., T. Hastie, and R. Tibshirani. 2008. "Sparse Inverse Covariance Estimation With the Graphical Lasso." *Biostatistics* 9, no. 3: 432–441.
- Friel, N., and A. N. Pettitt. 2008. "Marginal Likelihood Estimation via Power Posteriors." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 70, no. 3: 589–607.
- Gan, L., N. N. Narisetty, and F. Liang. 2019. "Bayesian Regularization for Graphical Models With Unequal Shrinkage." *Journal of the American Statistical Association* 114, no. 527: 1218–1231.
- Gao, C., A. W. van der Vaart, and H. H. Zhou. 2020. "A General Framework for Bayes Structured Linear Models." *Annals of Statistics* 48, no. 5: 2848–2878.
- Grünwald, P., and T. Van Ommen. 2017. "Inconsistency of Bayesian Inference for Misspecified Linear Models, and a Proposal for Repairing It." *Bayesian Analysis* 12, no. 4: 1069–1103.
- Higham, N. J. 1988. "Computing a Nearest Symmetric Positive Semidefinite Matrix." *Linear Algebra and Its Applications* 103: 103–118.
- Janková, J., and S. van de Geer. 2017. "Honest Confidence Regions and Optimality in High-Dimensional Precision Matrix Estimation." *Test* 26: 143–162.
- L'Huillier, A., L. Travis, I. Castillo, and K. Ray. 2023. "Semiparametric Inference Using Fractional Posteriors." *Journal of Machine Learning Research* 24, no. 389: 1–61.
- Lauritzen, S. L. 1996. *Graphical Models*. Oxford University Press.
- Li, Y., B. A. Craig, and A. Bhadra. 2019. "The Graphical Horseshoe Estimator for Inverse Covariance Matrices." *Journal of Computational and Graphical Statistics* 28, no. 3: 747–757.
- Liu, C., Y. Yang, H. Bondell, and R. Martin. 2021. "Bayesian Inference in High-Dimensional Linear Models Using an Empirical Correlation-Adaptive Prior." *Statistica Sinica* 31, no. 4: 2051–2072.
- Medina, M. A., J. L. M. Olea, C. Rush, and A. Velez. 2022. "On the Robustness to Misspecification of  $\alpha$ -Posteriors and Their Variational Approximations." *Journal of Machine Learning Research* 23, no. 147: 1–51.
- Pourahmadi, M. 2011. "Covariance Estimation: The GLM and Regularization Perspectives." *Statistical Science* 26, no. 3: 369–387.
- Ryali, S., T. Chen, K. Supekar, and V. Menon. 2012. "Estimation of Functional Connectivity in fMRI Data Using Stability Selection-Based Sparse Partial Correlation With Elastic Net Penalty." *NeuroImage* 59, no. 4: 3852–3861.
- Sagar, K., S. Banerjee, J. Datta, and A. Bhadra. 2024. "Precision Matrix Estimation Under the Horseshoe-Like Prior–Penalty Dual." *Electronic Journal of Statistics* 18, no. 1: 1–46.
- Srivastava, A., and A. Bhadra. 2018. "Ghs: Graphical Horseshoe MCMC Sampler Using Data Augmented Block Gibbs Sampler." <https://CRAN.R-project.org/package=GHS>, R package version 0.1.
- Syring, N., and R. Martin. 2019. "Calibrating General Posterior Credible Regions." *Biometrika* 106, no. 2: 479–486.
- Van Erven, T., and P. Harremoës. 2014. "Rényi Divergence and Kullback-Leibler Divergence." *IEEE Transactions on Information Theory* 60, no. 7: 3797–3820.
- Wang, H. 2012. "Bayesian Graphical Lasso Models and Efficient Posterior Computation." *Bayesian Analysis* 7, no. 4: 867–886.
- Wu, P.-S., and R. Martin. 2023. "A Comparison of Learning Rate Selection Methods in Generalized Bayesian Inference." *Bayesian Analysis* 18, no. 1: 105–132.
- Yang, Y., D. Pati, and A. Bhattacharya. 2020. " $\alpha$ -Variational Inference With Statistical Guarantees." *Annals of Statistics* 48, no. 2: 886–905.
- Yonekura, S., and S. Sugawara. 2023. "Adaptation of the Tuning Parameter in General Bayesian Inference With Robust Divergence." *Statistics and Computing* 33, no. 2: 39.
- Zhang, R., Y. Yao, and M. Ghosh. 2022. "Contraction of a Quasi-Bayesian Model With Shrinkage Priors Imprecision Matrix Estimation." *Journal of Statistical Planning and Inference* 221: 154–171.

## Appendix A: Proofs

### A.1 | Proof of Theorem 1

*Proof.* We can check the hypotheses on the KL between the likelihood terms as required in Theorem 2.6 in Alquier and Ridgway (2020). For

$$\mathcal{K}(p_1, p_2) = \int p_1 \log(p_1/p_2),$$

and let  $d_1, \dots, d_p$  denote the eigenvalues of  $\Omega_0^{-1/2} \Omega \Omega_0^{-1/2}$ . Then, using Lemma 1, we have

$$\mathcal{K}(p_{\Omega_0}, p_{\Omega}) = -\frac{1}{2} \sum_{i=1}^p \log d_i - \frac{1}{2} \sum_{i=1}^n (1 - d_i).$$

As  $\sum_{i=1}^n (1 - d_i)^2 = \|I_p - \Omega_0^{-1/2} \Omega \Omega_0^{-1/2}\|_2^2$ , when  $\|I_p - \Omega_0^{-1/2} \Omega \Omega_0^{-1/2}\|_2^2$  is small, we have  $\max_{1 \leq i \leq p} |1 - d_i| < 1$ ; see Banerjee and Ghosal (2015). This leads to that

$$\begin{aligned} \mathcal{K}(p_{\Omega_0}, p_{\Omega}) &\lesssim \sum_{i=1}^n (1 - d_i)^2 = \|I_p - \Omega_0^{-1/2} \Omega \Omega_0^{-1/2}\|_2^2 \\ &= \|\Omega_0^{-1/2} (\Omega - \Omega_0) \Omega_0^{-1/2}\|_2^2 \\ &\leq \|\Omega_0^{-1}\|_2^2 \|\Omega - \Omega_0\|_2^2 \leq \varepsilon_0^{-2} \|\Omega - \Omega_0\|_2^2. \end{aligned}$$

Integrating with respect to  $\rho_n \propto \mathbf{1}_{\|\Omega - \Omega_0\|_2 < \delta} \Pi_{HS}$  where  $\delta = [s \log(p/s)/n]^{1/2}$ , we have that

$$\int \mathcal{K}(p_{\Omega_0}, p_{\Omega}) \rho_n(d\Omega) \leq 2\varepsilon_0^{-2} \delta^2 \leq 2\varepsilon_0^{-2} \frac{s \log(p/s)}{n}. \quad (\text{A1})$$

Applying Lemma 3 with  $d = p(p-1)/2$ , we have that

$$\frac{1}{n} \mathcal{K}(\rho_n, \pi) \leq \frac{1}{n} \log \frac{1}{\Pi_{HS}(\|\Omega - \Omega_0\|_2 < \delta)} \leq \frac{1}{n} K_S \log(p/s).$$

Consequently, we can now apply Theorem 2.6 from Alquier and Ridgway (2020) to obtain the result, with  $\epsilon_n = 2\epsilon_0^{-2} \frac{s \log(p/s)}{n}$ . The proof is completed.  $\square$

**Proof of Corollary 3.1.** From Van Erven and Harremos (2014), we have that

$$h^2(P_\Omega, P_{\Omega_0}) \leq D_{1/2}(P_\Omega, P_{\Omega_0}) \leq D_\alpha(P_\Omega, P_{\Omega_0}),$$

for  $\alpha \in [0, 0.5]$ . In addition, from Van Erven and Harremos (2014), all  $\alpha$ -Rényi divergences are all equivalent for  $0 < \alpha < 1$ , through the formula  $\frac{\alpha}{\beta} \frac{1-\beta}{1-\alpha} D_\beta \leq D_\alpha \leq D_\beta$  for  $\alpha < \beta$ ; thus, we have that

$$D_{1/2}(P_\theta, P_{\theta_0}) \leq \frac{(1-\alpha)1/2}{\alpha(1-1/2)} D_\alpha(P_\theta, P_{\theta_0}) = \frac{(1-\alpha)}{\alpha} D_\alpha(P_\theta, P_{\theta_0}),$$

for  $\alpha \in (0, 0.5)$ . Thus, with  $c_\alpha = 1$  for  $\alpha \in [0, 0.5]$  and  $c_\alpha = (1-\alpha)/\alpha$  for  $\alpha \in (0, 0.5)$ , we obtain from Theorem 1 that

$$\mathbb{E} \left[ \int \frac{1}{c_\alpha} h^2(P_\Omega, P_{\Omega_0}) \pi_{n,\alpha}(d\Omega) \right] \leq \frac{1+\alpha}{1-\alpha} \epsilon_n.$$

From Lemma 2, one gets that

$$\mathbb{E} \left[ \int \frac{c_0}{c_\alpha} \|\Omega_1 - \Omega_2\|_2^2 \pi_{n,\alpha}(d\Omega) \right] \leq \frac{1+\alpha}{1-\alpha} \epsilon_n.$$

The proof is completed.

### A.1.1 | Proof of Theorem 2

*Proof.* We utilise the general framework of posterior concentration rates by confirming the criteria outlined in Theorem 2.4 of Alquier and Ridgway (2020). This involves assessing the prior concentration rate within Kullback–Leibler neighbourhoods.

Put

$$\mathcal{K}(p_1, p_2) = \int p_1 \log(p_1/p_2), \text{ and } V(p_1, p_2) = \int p_1 \log^2(p_1/p_2).$$

and let  $d_1, \dots, d_p$  denote the eigenvalues of  $\Omega_0^{-1/2} \Omega \Omega_0^{-1/2}$ . Controlling the term  $\mathcal{K}(p_1, p_2) = \int p_1 \log(p_1/p_2)$  is presented in the proof of Theorem 1, given in page 27. Now, using Lemma 1, we have

$$V(p_{\Omega_0}, p_\Omega) = \frac{1}{2} \sum_{i=1}^n (1-d_i)^2 + \mathcal{K}(p_{\Omega_0}, p_\Omega)^2.$$

As  $\sum_{i=1}^n (1-d_i)^2 = \|I_p - \Omega_0^{-1/2} \Omega \Omega_0^{-1/2}\|_2^2$ , when  $\|I_p - \Omega_0^{-1/2} \Omega \Omega_0^{-1/2}\|_2^2$  is small, we have  $\max_{1 \leq i \leq p} |1-d_i| < 1$ . This gives

$$\begin{aligned} V(p_{\Omega_0}, p_\Omega) &\lesssim \sum_{i=1}^n (1-d_i)^2 = \|I_p - \Omega_0^{-1/2} \Omega \Omega_0^{-1/2}\|_2^2 \\ &= \|\Omega_0^{-1/2} (\Omega - \Omega_0) \Omega_0^{-1/2}\|_2^2 \\ &\leq \|\Omega_0^{-1}\|_2^2 \|\Omega - \Omega_0\|_2^2 \leq \epsilon_0^{-2} \|\Omega - \Omega_0\|_2^2. \end{aligned}$$

Integrating with respect to  $\rho_n \propto \mathbf{1}_{\|\Omega - \Omega_0\|_2 < \delta} \Pi_{HS}$  where  $\delta = [s \log(p/s)/n]^{1/2}$ , we have that

$$\int V(p_{\Omega_0}, p_\Omega) \rho_n(d\Omega) \leq 2\epsilon_0^{-2} \delta^2 \leq 2\epsilon_0^{-2} \frac{s \log(p/s)}{n},$$

and from (A1),  $\int \mathcal{K}(p_{\Omega_0}, p_\Omega) \rho_n(d\Omega) \leq 2\epsilon_0^{-2} \delta^2 \leq 2\epsilon_0^{-2} \frac{s \log(p/s)}{n}$ . From Lemma 3, we have that

$$\frac{1}{n} \mathcal{K}(\rho_n, \pi) \leq \frac{1}{n} \log \frac{1}{\Pi_{HS}(\|\Omega - \Omega_0\|_2 < \delta)} \leq \frac{1}{n} K_S \log(p/s).$$

Consequently, we can now apply Theorem 2.4 from Alquier and Ridgway (2020) to obtain the result, with  $\epsilon_n = 2\epsilon_0^{-2} \frac{s \log(p/s)}{n}$ . The proof is completed.  $\square$

**Proof of Corollary 3.2.** The proof is similar to the proof of Corollary 3.1.

**Lemma 1.** Lemma B1 in (Sagar et al. 2024) Let  $p_k$  denote the density of a  $\mathcal{N}_d(0, \Sigma_k)$  random variable,  $k = 1, 2$ . Denote the corresponding precision matrices by  $\Omega_k = \Sigma_k^{-1}$ ,  $k = 1, 2$ . Then,

$$\begin{aligned} \mathbb{E}_{p_1} \left\{ \log \frac{p_1}{p_2}(X) \right\} &= \frac{1}{2} \{ \log \det \Omega_1 - \log \det \Omega_2 + \text{trace}(\Omega_1^{-1} \Omega_2 - I_d) \}, \\ \text{Var}_{p_1} \left\{ \log \frac{p_1}{p_2}(X) \right\} &= \frac{1}{2} \text{trace} \{ (\Omega_1^{-1/2} \Omega_2 \Omega_1^{-1/2} - I_d)^2 \}. \end{aligned}$$

**Lemma 2.** Lemma A.1 in Banerjee and Ghosal (2015) Let  $p_k$  denote the density of a  $\mathcal{N}_d(0, \Sigma_k)$  random variable,  $k = 1, 2$ . Denote the corresponding precision matrices by  $\Omega_k = \Sigma_k^{-1}$ ,  $k = 1, 2$ . Then, there exists a positive constant  $c_0$  such that  $\|\Omega_1 - \Omega_2\|_2^2 \leq c_0 h^2(p_1, p_2)$ .

The subsequent lemma introduces a novel prior concentration result for the horseshoe prior, offering a lower bound on the probability assigned to a Euclidean neighbourhood of a sparse vector. This lemma represents a vital improvement over Lemma 1 in Chakraborty et al. (2020), crucial for achieving the optimal rate  $s \log(p/s)/n$  instead of the suboptimal rate  $s \log(p)/n$  in high-dimensional setting; see Bellec et al. (2018) and Gao et al. (2020).

**Lemma 3.** Suppose  $\theta_0 \in \mathbb{R}^d$  such that  $\|\theta_0\|_0 = s$  and that  $s < n < d$  and  $\|\theta_0\|_\infty \leq C_1$ . Suppose  $\theta \sim \Pi_{HS}$ . Define  $\delta = [s \log(d/s)/n]^{1/2}$ . Then we have for some constant  $K > 0$  that  $\Pi_{HS}(\|\theta - \theta_0\|_2 < \delta) \geq e^{-Ks \log(d/s)}$ .

**Proof of Lemma 3.** Using the scale-mixture formulation of  $\Pi_{HS}$ ,

$$\begin{aligned} \Pi_{HS}(\|\theta - \theta_0\|_2 < \delta) &= \int_{\tau} pr(\|\theta - \theta_0\|_2 < \delta | \tau) f(\tau) d\tau \\ &\geq \int_{I_{\tau^*}} pr(\|\theta - \theta_0\|_2 < \delta | \tau) f(\tau) d\tau, \end{aligned}$$

where  $I_{\tau_*} = [\tau_*/2, \tau_*]$  with  $\tau_* = s(s/p)^{3/2} \{\log(d/s)/n\}^{1/2}$ . Let  $S = \{1 \leq j \leq d: \theta_{0j} \neq 0\}$ . For  $\tau \in I_{\tau_*}$ , one can lower bound the conditional probability as

$$\begin{aligned} \text{pr}(\|\theta - \theta_0\|_2 < \delta | \tau) &\geq \text{pr}(\|\theta_S - \theta_{0S}\|_2 < \delta/2 | \tau) \text{pr}(\|\theta_{S^c}\|_2 < \delta/2 | \tau) \\ &\geq \prod_{j \in S} \text{pr}\left(|\theta_j - \theta_{0j}| < \frac{\delta}{2s} | \tau\right) \prod_{j \in S^c} \text{pr}\left(|\theta_j| < \frac{\delta}{2\sqrt{d}} | \tau\right). \end{aligned} \tag{A2}$$

For a fixed  $\tau \in I_{\tau_*}$ , we will lower bound each of the terms in the right-hand side of (A2).

Initially, we examine  $\text{pr}\{|\theta_j| < \delta/(2d^{1/2}) | \tau\}$  with  $\tau$  in  $I_{\tau_*}$ . Given  $\tau$  and  $\lambda_j$ ,  $\theta_j$  follows a normal distribution  $N(0, \lambda_j^2 \tau^2)$ . Leveraging the Chernoff bound for a Gaussian random variable, we obtain the following:

$$\text{pr}\{|\theta_j| > \delta/(2d^{1/2}) | \lambda_j, \tau\} \leq 2e^{-\delta^2/(8p\lambda_j^2\tau^2)} \leq 2e^{-\delta^2/(8p\lambda_j^2\tau_*^2)}.$$

Thus,

$$\begin{aligned} \text{pr}\{|\theta_j| < \delta/(2d^{1/2}) | \tau\} &= \int_{\lambda_j} \text{pr}\{|\theta_j| < \delta/(2d^{1/2}) | \lambda_j, \tau\} f(\lambda_j) d\lambda_j \\ &\geq \int_{\lambda_j} \left\{1 - 2 \exp\left(-\delta^2/(8d\lambda_j^2\tau_*^2)\right)\right\} f(\lambda_j) d\lambda_j \\ &= 1 - \frac{4}{\pi} \int_{\lambda_j} \exp\left(-\delta^2/(8d\lambda_j^2\tau_*^2)\right) (1 + \lambda_j^2)^{-1} d\lambda_j = 1 - \frac{4}{\pi} I. \end{aligned}$$

We now upper-bound the integrand as follows:

$$\begin{aligned} I &= \int_{\lambda_j} \exp\left(-\frac{\delta^2}{8d\tau_*^2\lambda_j^2}\right) (1 + \lambda_j^2)^{-1} d\lambda_j \\ &\leq \int_{\lambda_j} \exp\left(-\frac{\delta^2}{8d\tau_*^2\lambda_j^2}\right) \lambda_j^{-2} d\lambda_j \\ &= \frac{1}{2} \int_0^\infty z^{-1/2} \exp\left(-\frac{\delta^2 z}{8d\tau_*^2}\right) dz \\ &= \frac{\Gamma(1/2)}{\{2\delta^2/(8d\tau_*^2)\}^{1/2}} = \frac{2s^2\sqrt{\pi}}{d}, \end{aligned}$$

where a substitution  $z = 1/\lambda^2$  at the third step was made. Thus, for  $\tau \in I_{\tau_*}$ ,  $\text{pr}\{|\theta_j| < \delta/2d^{1/2} | \tau\} \geq 1 - \frac{8s^2}{d\sqrt{\pi}}$ .

Now, we lower bound  $\text{pr}\{|\theta_j - \theta_{0j}| < \delta_0 | \tau\}$  with  $\tau \in I_{\tau_*}$ . Letting  $\delta_0 = \delta/(2\sqrt{s})$ , then

$$\begin{aligned} \text{pr}\{|\theta_j - \theta_{0j}| < \delta_0 | \tau\} &= \left(\frac{2}{\pi^3}\right)^{1/2} \int_{\lambda_j} \int_{|\theta_j - \theta_{0j}| < \delta_0} \exp\{-\theta_j^2/(2\lambda_j^2\tau^2)\} \frac{1}{\lambda_j\tau(1 + \lambda_j^2)} d\theta_j d\lambda_j \\ &\geq \left(\frac{2}{\pi^3}\right)^{1/2} \int_{|\theta_j - \theta_{0j}| < \delta_0} \int_{1/\tau}^{2/\tau} \exp\{-\theta_j^2/(2\lambda_j^2\tau^2)\} \frac{1}{\lambda_j\tau(1 + \lambda_j^2)} d\lambda_j d\theta_j \\ &\geq \left(\frac{2}{\pi^3}\right)^{1/2} \int_{|\theta_j - \theta_{0j}| < \delta_0} \exp(-\theta_j^2/2) \left(\int_{1/\tau}^{2/\tau} \frac{1}{1 + \lambda_j^2} d\lambda_j\right) d\theta_j, \end{aligned}$$

since for  $\lambda_j \in [1/\tau, 2/\tau]$ ,  $1/(\lambda_j\tau) \geq 1/2$  and  $\exp\{-\theta_j^2/(2\lambda_j^2\tau^2)\} \geq \exp(-\theta_j^2/2)$ . Continuing,

$$\begin{aligned} \text{pr}\{|\theta_j - \theta_{0j}| < \delta_0 | \tau\} &\geq (2/\pi^3)^{1/2} \frac{\tau}{4 + \tau^2} \int_{|\theta_j - \theta_{0j}| < \delta_0} \exp(-\theta_j^2/2) d\theta_j \\ &\geq (2/\pi^3)^{1/2} \frac{\tau}{4 + \tau^2} \exp\{-(C_1 + 1)^2/2\} \delta_0 \\ &\geq K \tau \delta_0 \geq K \frac{\tau_* \delta}{4\sqrt{s}} \geq Ks \left(\frac{s}{p}\right)^{3/2} \frac{\log(p/s)}{n} \\ &\geq K \left(\frac{s}{p}\right)^{5/2} = Ke^{-(5/2)\log(p/s)}, \end{aligned}$$

where in the third step, we used  $4 + \tau^2 < 5$  and  $n < p$  in the final step.

By substituting these bounds into (A2), we obtain for  $\tau \in I_{\tau_*}$ ,

$$\text{pr}(\|\theta - \theta_0\|_2 < \delta | \tau) \geq (Ke^{-(5/2)\log(p/s)})^s \left(1 - \frac{8s^2}{p\sqrt{\pi}}\right)^{p-s} \geq e^{-Ks \log(p/s)},$$

where  $K$  is a positive constant. The proof concludes by noting that the probability of  $\tau$  being within the interval  $I_{\tau_*}$  is greater than or equal to  $\tau_*/(2\pi)$ . Thus, with a minor deviation in notation, we obtain

$$\Pi_{HS}(\|\theta - \theta_0\|_2 < \delta) \geq e^{-Ks \log(p/s)},$$

for some positive constant  $K$ . The proof is completed.

### A.1.2 | Proof of Theorem 3

*Proof.* We can check the assumptions as in Theorem 2.7 in Alquier and Ridgway (2020).

Let us define  $A = \Omega_1^{-1/2} \Omega_2 \Omega_1^{-1/2}$ . Note that for a random variable  $Z \sim \mathcal{N}_d(0, \Sigma)$ , we have  $\mathbb{E}(Z^T A Z) = \text{trace}(A\Sigma)$ . Then, for  $X \sim \mathcal{N}_d(0, \Sigma_0)$ ,

$$\begin{aligned} \mathbb{E}_{p_0} \left\{ \log \frac{P_1}{P_2}(X) \right\} &= \frac{1}{2} \left\{ \log \det \Omega_1 - \log \det \Omega_2 + \mathbb{E}_{p_0}(X^T (\Omega_2 - \Omega_1) X) \right\} \\ &= \frac{1}{2} \left\{ \log \det \Omega_1 - \log \det \Omega_2 + \text{trace}[(\Omega_2 - \Omega_1)\Sigma_0] \right\}. \end{aligned}$$

Thus, we have that

$$\begin{aligned} \mathbb{E}_{\Omega_0} \left[ \log \frac{P_{\Omega_*}}{P_{\Omega}}(X) \right] &\leq \frac{1}{2} \left\{ \log \left( \frac{\det(\Omega_*)}{\det(\Omega)} \right) + \text{trace}[(\Omega - \Omega_*)\Omega_0^{-1}] \right\} \\ &\leq \frac{1}{2} \left\{ \log \left( \frac{\det(\Omega_*)}{\det(\Omega)} \right) + \|(\Omega - \Omega_*)\Omega_*^{-1}\Omega_*\|_2 \|\Omega_0^{-1}\|_2 \right\} \\ &\leq \frac{1}{2} \left\{ \log \left( \frac{\det(\Omega_*)}{\det(\Omega)} \right) + \|\Omega\Omega_*^{-1} - I_p\|_2 \|\Omega_*\|_2 \epsilon_0^{-1} \right\}. \end{aligned}$$

Let  $d_1, \dots, d_p$  denote the eigenvalues of  $\Omega_0^{-1}$ . As  $\sum_{i=1}^n (1 - d_i)^2 = \|I_p - \Omega_0^{-1}\|_2^2$ , when  $\|I_p - \Omega_0^{-1}\|_2^2$  is small, we have,  $\max_{1 \leq i \leq p} |1 - d_i| < 1$ . This leads to that



$$\begin{aligned} \mathbb{E}_{\Omega_0} \left[ \log \frac{P_{\Omega_*}}{P_{\Omega}}(X) \right] &\leq \frac{1}{2} \left\{ \sum_{i=1}^p \log d_i + \|\Omega_*\|_2 \varepsilon_0^{-1} \sum_{i=1}^p (1-d_i) \right\} \\ &\leq K \|\Omega_*\|_2 \varepsilon_0^{-1} \sum_{i=1}^p (1-d_i)^2 \leq K_{\varepsilon_0} \|\Omega - \Omega_*\|_2^2, \end{aligned}$$

for some absolute constant  $K_{\varepsilon_0} > 0$  depending only on  $\varepsilon_0$ . Integrating with respect to  $\rho_n \propto \mathbf{1}_{\|\Omega - \Omega_*\|_2 < \delta} \Pi_{HS}$  where  $\delta = \lceil \log(p/s)/n \rceil^{1/2}$ , we have that

$$\int \mathbb{E}_{\Omega_0} \left[ \log \frac{P_{\Omega_*}}{P_{\Omega}}(X) \right] \rho_n(d\Omega) \leq K_{\varepsilon_0} \delta^2 \leq K_{\varepsilon_0} \frac{s \log(p/s)}{n}.$$

To apply Theorem 2.7 in Alquier and Ridgway (2020), it remains to compute the KL between the approximation  $\rho_n$  and the prior. From Lemma 3, we have that

$$\frac{1}{n} \mathcal{K}(\rho_n, \pi) \leq \frac{1}{n} \log \frac{1}{\Pi_{HS}(\|\Omega - \Omega_0\|_2 < \delta)} \leq \frac{1}{n} Ks \log(p/s).$$

To obtain an estimate of the rate  $\varepsilon_n$  as in Theorem 2.7 in Alquier and Ridgway (2020), we put together those bounds and choosing  $\varepsilon_n = K_{\varepsilon_0} \frac{s \log(p/s)}{n}$ . The results are followed, and the proof is completed.  $\square$

**Proof of Corollary 3.4.** A similar argument as in the proof of Corollary 3.1, we have that  $h^2(p_{\Omega}, p_{\Omega_0}) \leq c_{\alpha} D_{\alpha}(P_{\Omega}, P_{\Omega_0})$ . Then, using Lemma 2, one has that  $\|\Omega_1 - \Omega_2\|_2^2 \lesssim h^2(p_1, p_2)$ . Additionally, from the proof of Theorem 2, page 27, one has that  $\mathcal{K}(p_{\Omega_0}, p_{\Omega}) \lesssim \varepsilon_0^{-2} \|\Omega - \Omega_0\|_2^2$ . Putting all together into Theorem 3, we get that

$$\mathbb{E} \left[ \int \|\Omega - \Omega_0\|_2^2 \Pi_{n,\alpha}(d\Omega) \right] \leq C \left( \frac{\alpha}{1-\alpha} \min_{\Omega \in \mathcal{S}(s, \varepsilon_0)} \|\Omega - \Omega_0\|_2^2 + \frac{1+\alpha}{1-\alpha} \varepsilon_n \right),$$

for some universal constant  $C > 0$  that do not depend on  $n, s, p$ . Subsequently, an application of Jensen's inequality yields the result for the mean estimator. The proof is completed.