Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2026, 2026:76
https://doi.org/10.28919/cmbn/9924
ISSN: 2052-2541

ROBUST BAYESIAN PRECISION MATRIX ESTIMATION WITH A REGULARIZED HORSESHOE PRIOR

SAUMU ATHMAN ABDALLAH[1,*], ANTHONY WANJOYA[1], MUTUA KILAI[2]

[1]Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, P.O. Box 62000-00200, Nairobi, Kenya

[2]Department of Pure and Applied Sciences, Kirinyaga University, P.O. Box 143-10300, Kerugoya, Kenya

Abstract. Estimating high-dimensional precision matrices in Gaussian graphical models is challenging due to sensitivity to outliers and limitations of uniform shrinkage methods. We propose a robust Bayesian approach based on a sparse Cholesky factorization, combining the γ-divergence for outlier resistance with a regularized horseshoe prior to enable adaptive shrinkage. Estimation is performed via MAP optimization using the Adam algorithm, with uncertainty quantified through a Laplace approximation. Theoretical properties including posterior propriety, existence of the MAP estimator, and robustness are established. Simulation results show improved performance in terms of Frobenius norm and root mean squared error under data contamination. An application to gene expression data illustrates the method's ability to identify potentially meaningful association patterns. The proposed framework provides a computationally tractable approach to robust precision matrix estimation in high-dimensional settings.

Keywords: Gaussian graphical models; precision matrix estimation; γ-divergence; regularized horseshoe prior; robust Bayesian inference; high-dimensional statistics; Laplace approximation.

2020 AMS Subject Classification: 62H12.

*Corresponding author

E-mail address: athman.saumu@students.jkuat.ac.ke

Received April 16, 2026

1. INTRODUCTION

Gaussian graphical models (GGMs) provide a powerful framework for representing conditional independence relationships among multivariate continuous variables. In these models, the precision matrix encodes the underlying graph structure, with zero entries corresponding to conditional independence between variable pairs given the remaining variables [1]. GGMs are widely used in applications such as genomics [2], finance [3], and neuroimaging [4].

In modern data-driven settings, it is increasingly common to encounter high-dimensional scenarios where the number of variables exceeds the number of observations ($p > n$). In such regimes, classical maximum likelihood estimation becomes ill-posed due to the non-invertibility of the sample covariance matrix, leading to unstable or unreliable estimates. To address this challenge, regularization techniques have been introduced to impose structural assumptions on the precision matrix, most notably sparsity. The sparsity assumption, which posits that many variables are conditionally independent, reduces estimation variance and enhances interpretability in high-dimensional settings.

A major breakthrough in sparse estimation was the introduction of the lasso by Tibshirani [5], which employs an ℓ_1 penalty to encourage sparsity in regression models. This idea was subsequently extended to Gaussian graphical models through penalized likelihood approaches [6, 7], culminating in the graphical lasso algorithm of Friedman et al. [8], which provides an efficient framework for estimating sparse precision matrices. Theoretical developments have further established that such ℓ_1-penalized estimators can recover the underlying graph structure under suitable sparsity assumptions.

While frequentist approaches such as the graphical lasso are computationally efficient, they do not naturally provide measures of uncertainty, which are essential in many scientific applications. This limitation has motivated the development of Bayesian approaches to graphical modeling. Early Bayesian methods employed discrete mixture priors, such as spike-and-slab formulations [9, 10], which explicitly model edge inclusion. Although theoretically appealing, these approaches suffer from computational challenges due to the combinatorial complexity of the graph space.

To improve scalability, continuous shrinkage priors have been proposed as an alternative. The Laplace (double-exponential) prior, which mirrors the ℓ_1 penalty, has been widely used in Bayesian graphical lasso formulations [11, 12]. However, the Laplace prior applies uniform shrinkage across coefficients, which may lead to over-shrinkage of large signals and biased estimation. This limitation has motivated the development of global–local shrinkage priors, such as the horseshoe prior [13], which allows for adaptive shrinkage by combining global and local scale parameters. The regularized horseshoe prior [14] further refines this approach by controlling tail behavior while preserving the ability to retain strong signals.

In the context of Gaussian graphical models, extending global–local shrinkage priors presents additional challenges due to the requirement that the precision matrix remain symmetric and positive definite. Approaches based on Cholesky parameterizations have been proposed to address these constraints [15, 16], though they may complicate direct interpretation of conditional independence relationships.

Another critical issue in GGM estimation is robustness. Standard Gaussian likelihood-based methods are sensitive to outliers and model misspecification. To mitigate this, robust alternatives have been proposed, including heavy-tailed distributions such as the multivariate t-distribution [17], as well as divergence-based approaches that downweight extreme observations. Among these, the γ-divergence introduced by Fujisawa and Eguchi [18] has gained attention due to its redescending influence function, which effectively limits the impact of outliers. However, its performance depends on the choice of the tuning parameter, which controls the trade-off between robustness and efficiency [19, 20].

Recent work by Onizuka and Hashimoto [21] integrates the γ-divergence into a Bayesian graphical modeling framework using Laplace priors to induce sparsity. While this approach improves robustness and enables uncertainty quantification, it inherits the limitations of uniform shrinkage associated with the Laplace prior. In particular, it may fail to adequately distinguish between noise and strong signals in high-dimensional settings.

Motivated by these limitations, this paper proposes a robust Bayesian approach to precision matrix estimation that combines the γ-divergence with a regularized horseshoe prior. By replacing the Laplace prior with an adaptive global–local shrinkage prior, the proposed method

aims to improve graph structure recovery and uncertainty quantification in high-dimensional settings, particularly in the presence of contamination. The performance of the proposed approach is evaluated through simulation studies and real data analysis, focusing on estimation accuracy, robustness, and posterior uncertainty.

2. METHODOLOGY

2.1. Multivariate Normal Model and Cholesky Parameterization. Let $\mathbf{Y} = (y_1, \ldots, y_n)^\top \in \mathbb{R}^{n\times p}$ denote the centered and standardized observation matrix, where each $y_i \in \mathbb{R}^p$ represents the ith observation. We assume the data follow a multivariate normal distribution with mean zero and covariance matrix Σ, so that the precision matrix is $\Omega = \Sigma^{-1} \succ 0$. The multivariate normal density is

$$f(y_i \mid \Omega) = (2\pi)^{-p/2}|\Omega|^{1/2}\exp\left(-\frac{1}{2}y_i^\top \Omega y_i\right). \tag{1}$$

Our primary objective is the estimation of Ω in high-dimensional settings where p may be comparable to or larger than n. To ensure positive definiteness and facilitate computational efficiency, we adopt the Cholesky decomposition

$$\Omega = \mathbf{L}\mathbf{L}^\top, \tag{2}$$

where $\mathbf{L}$ is a lower triangular matrix with strictly positive diagonal entries. This parameterization provides a bijection between the cone of positive definite matrices $\mathscr{M}^+$ and the set $\mathscr{L}_+ = \{\mathbf{L} \in \mathbb{R}^{p\times p} : \mathbf{L} \text{ lower triangular}, \mathbf{L}_{ii} > 0\}$.

Sparsity is imposed on the off-diagonal elements of $\mathbf{L}$ to regularize estimation and improve interpretability of the factorization. However, sparsity in $\mathbf{L}$ does not imply sparsity in Ω; rather, it serves as a structured regularization mechanism that facilitates stable estimation in high dimensions. The resulting estimate $\widehat{\Omega} = \widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top$ is subsequently used for inference.

2.2. Model Assumptions. We assume that the observations $\mathbf{y}_i \in \mathbb{R}^p$ are independently generated from a multivariate distribution with mean zero and precision matrix $\Omega \in \mathscr{M}^+$. The baseline model assumes Gaussianity, while robustness to deviations from this assumption, including contamination by outliers or heavy-tailed observations, is introduced through the γ-divergence. The precision matrix is parameterized via its Cholesky factor $\Omega = \mathbf{L}\mathbf{L}^\top$. Off-diagonal elements

of $\mathbf{L}$ are assigned a regularized horseshoe prior, which induces sparsity through adaptive shrinkage. Diagonal elements are constrained to be positive to ensure positive definiteness. We further assume that the underlying precision matrix is sparse and that the sample size is sufficient relative to the dimensionality to allow stable estimation under the imposed regularization.

2.3. Model formulation.

2.3.1. *Definition of the γ-divergence Objective.* The γ-divergence, introduced by Fujisawa and Eguchi [18], is a robust alternative to the Kullback-Leibler divergence for measuring the discrepancy between a true unknown data-generating distribution and a specified statistical model. It is particularly useful in the presence of outliers or contamination, as it naturally downweights the influence of observations that deviate substantially from the model. Let $g(y)$ be the true data-generating distribution (unknown), and let $f(y \mid \Omega)$ be the model. The γ-divergence between g and f_Ω is defined as:

$$D_\gamma(g, f_\Omega) = \frac{1}{\gamma(1+\gamma)} \log \int g(y)^\gamma dy - \frac{1}{\gamma} \log \int g(y) f_\Omega(y)^\gamma dy + \frac{1}{1+\gamma} \log \int f_\Omega(y)^{1+\gamma} dy$$

for $\gamma > 0$, and since $g(y)$ is unknown, we use the empirical version of the divergence. Minimization of the resulting empirical (γ)-divergence yields the negative (γ)-log-likelihood function

$$\ell_\gamma(\Omega) = -\frac{1}{\gamma} \log \left(\frac{1}{n} \sum_{i=1}^{n} f(y_i \mid \Omega)^\gamma \right) + \frac{1}{1+\gamma} \log \int f(y \mid \Omega)^{1+\gamma} dy. \tag{3}$$

Substituting (1) into (3) and ignoring constants that do not depend on Ω yields the negative γ-divergence loss

$$\ell_\gamma(\Omega) = -\frac{1}{2(1+\gamma)} \log |\Omega| - \frac{1}{\gamma} \log \left(\frac{1}{n} \sum_{i=1}^{n} \exp \left\{ -\frac{\gamma}{2} y_i^\top \Omega y_i \right\} \right). \tag{4}$$

The γ-divergence improves robustness by downweighting observations with large Mahalanobis distances. In the term $\exp\left(-\frac{\gamma}{2} y_i^\top \Omega y_i\right)$, observations that deviate substantially from the model contribute exponentially small weights, effectively reducing their influence on the estimated precision matrix. This redescending property ensures that outliers have vanishing impact as their magnitude increases, while observations consistent with the model retain full influence. The parameter γ controls the threshold for downweighting: smaller γ values produce behavior

closer to the standard Gaussian likelihood, while larger γ values provide stronger protection against outliers.

When $\gamma = 0$, the loss reduces smoothly to the standard Gaussian negative log-likelihood,

$$\ell_0(\Omega) = -\frac{1}{2}\log\det(\Omega) + \frac{1}{2}\operatorname{tr}(\mathbf{S}\Omega)\,. \tag{5}$$

where $\mathbf{S} = n^{-1}\sum_{i=1}^{n} y_i y_i^\top$ denotes the sample covariance matrix.

2.3.2. *Regularized Horseshoe Shrinkage and Diagonal Stabilization.* Sparsity in the precision matrix is induced through a regularized horseshoe prior on the off-diagonal elements of the Cholesky factor $\mathbf{L}$. Let β_j denote the jth off-diagonal element of $\mathbf{L}$ (in a suitable vectorized ordering). The prior hierarchy is specified as

$$\beta_j \mid \lambda_j, \tau, c^2 \sim \mathcal{N}\left(0,\, \tau^2\tilde{\lambda}_j^2\right), \tag{6}$$

where τ is a global shrinkage parameter controlling overall sparsity, λ_j is a local parameter allowing coefficient-specific adaptation, and c^2 is a slab parameter that bounds the prior variance and prevents excessively large coefficients.

To control tail behavior while preserving strong shrinkage near zero, we use the regularized horseshoe transformation

$$\tilde{\lambda}_j^2 = \frac{c^2\lambda_j^2}{c^2 + \tau^2\lambda_j^2}. \tag{7}$$

The slab parameter is treated as unknown with hyperprior

$$c^2 \sim \text{Inverse-Gamma}(2, 1), \tag{8}$$

allowing the slab scale to adapt to the data.

The shrinkage parameters follow half-Cauchy priors

$$\lambda_j \sim \mathrm{C}^+(0, \lambda_{\text{scale}}), \qquad \tau \sim \mathrm{C}^+(0, \tau_{\text{scale}}), \tag{9}$$

providing heavy-tailed shrinkage that strongly regularizes small coefficients while allowing larger signals to remain. Positivity constraints for λ_j, τ, and c^2 are enforced via softplus transformations of unconstrained parameters during optimization.

For the diagonal elements of the Cholesky factor, we impose an exponential prior

$$\mathbf{L}_{ii} \sim \mathrm{Exp}(\alpha), \tag{10}$$

which regularizes the scale of the precision matrix and ensures positive definiteness. In the objective function, this contributes the term

$$\alpha \sum_{i=1}^{p} \mathbf{L}_{ii},$$

together with the Jacobian correction induced by the exponential reparameterization. This prior stabilizes estimation and improves numerical conditioning during optimization.

2.3.3. *Objective Function for MAP Estimation.* Estimation is performed using the maximum a posteriori (MAP) principle. The objective corresponds to the negative log-posterior distribution.

Under the Cholesky parameterization

$$\Omega = \mathbf{L}\mathbf{L}^{\top},$$

the posterior is expressed in terms of the Cholesky factor $\mathbf{L}$. The corresponding negative log-posterior objective can be written as

$$\begin{aligned} \mathscr{L}\left(\mathbf{L}, \lambda, \tau, c^2\right) &= \ell_{\gamma}\left(\mathbf{L}\mathbf{L}^{\top}; \mathbf{Y}\right) \\ &\quad + \mathscr{P}_{\mathrm{RHS}}\left(\mathbf{L}, \lambda, \tau, c^2\right) + \alpha \sum_{i=1}^{p} L_{ii}. \end{aligned} \tag{11}$$

where $\ell_{\gamma}(\mathbf{L}\mathbf{L}^{\top}; \mathbf{Y})$ denotes the γ-divergence log-likelihood, $\mathscr{P}_{\mathrm{RHS}}(\cdot)$ represents the negative log-density of the regularized horseshoe prior applied to the off-diagonal elements together with the hyperpriors on (λ, τ, c^2), and the third term corresponds to the contribution of the exponential prior on the diagonal elements of the Cholesky factor.

The MAP estimator is therefore defined as

$$\left(\widehat{\mathbf{L}}, \widehat{\lambda}, \widehat{\tau}, \widehat{c^2}\right) = \underset{\mathbf{L} \in \mathscr{L}_+,\, \lambda > 0,\, \tau > 0,\, c^2 > 0}{\arg\min} \mathscr{L}\left(\mathbf{L}, \lambda, \tau, c^2\right). \tag{12}$$

where $\mathscr{L}_+$ denotes the set of lower triangular matrices with positive diagonal entries. The corresponding precision matrix estimate is obtained as $\widehat{\Omega} = \widehat{\mathbf{L}}\widehat{\mathbf{L}}^{\top}$.

For numerical stability and to guarantee positive definiteness, the optimization is carried out using unconstrained parameters, with the diagonal elements reparameterized as

$$\mathbf{L}_{ii} = \exp(\tilde{L}_{ii}),$$

ensuring $\mathbf{L}_{ii} > 0$ throughout the optimization.

2.4. Optimization Strategy: Adam Optimizer for MAP Estimation. Maximum a posteriori (MAP) estimation is performed using the Adam optimizer [22] for all values of the robustness parameter γ. Adam is a first-order gradient-based optimization algorithm that combines ideas from AdaGrad [23] and RMSProp [24], providing adaptive learning rates and momentum-based updates.

Algorithm formulation. Let $\mathscr{L}(\theta)$ denote the objective function with parameter vector θ. At iteration t, Adam maintains exponentially weighted moving averages of the first and second moments of the gradient,

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t, \tag{13}$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2, \tag{14}$$

where $g_t = \nabla_\theta \mathscr{L}(\theta_{t-1})$ and $\beta_1, \beta_2 \in [0,1)$ are decay parameters. Bias-corrected estimates are given by

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}, \qquad \hat{v}_t = \frac{v_t}{1-\beta_2^t}. \tag{15}$$

The parameter update is

$$\theta_t = \theta_{t-1} - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon}, \tag{16}$$

where α is the learning rate and ε is a small constant for numerical stability.

Choice of optimizer. The objective function in (11) is non-convex due to the γ-divergence loss and the hierarchical regularized horseshoe prior, resulting in multiple local minima and parameters operating on different scales. Although positivity of the diagonal elements of the Cholesky factor is enforced via the reparameterization $\mathbf{L}_{ii} = \exp(\tilde{L}_{ii})$, the optimization remains unconstrained but exhibits heterogeneous curvature across parameters.

Adam is adopted for its stable performance in such settings. Its adaptive per-parameter learning rates effectively handle the differing scales of the Cholesky coefficients and shrinkage hyperparameters, while momentum-based updates improve stability in regions of low curvature, particularly when coefficients are strongly shrunk toward zero. These properties make Adam well-suited for high-dimensional non-convex optimization problems [22].

Optimization is carried out over unconstrained parameters, ensuring that the resulting precision matrix $\Omega = \mathbf{L}\mathbf{L}^\top$ remains positive definite throughout. However, due to non-convexity, the resulting estimates correspond to approximate MAP solutions and may depend on initialization. Alternative methods such as L-BFGS [25] were considered but exhibited less stable behavior in this setting.

Algorithm 1 MAP Estimation for the Robust Precision Matrix Estimation using Adam

Require: Data matrix $\mathbf{Y} \in \mathbb{R}^{n \times p}$, configuration parameters

1: Initialize parameters $\theta = (\beta, \lambda, \tau, c^2)$
2: Compute weighted covariance matrix $\mathbf{S}$
3: Initialize precision matrix using Cholesky factorization
4: **for** $t = 1, 2, \ldots, T$ **do**
5: Construct precision matrix $\Omega(\theta)$
6: Compute negative log-posterior

$$\mathscr{L}(\theta) = \ell_\gamma(\Omega; \mathbf{Y}) + \text{RHS prior} + \text{diagonal prior}$$

7: Compute gradient $g_t = \nabla_\theta \mathscr{L}(\theta)$ using automatic differentiation
8: Clip gradient magnitude for numerical stability
9: Update parameters using Adam

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon}$$

10: Check convergence using early stopping
11: **end for**
12: Compute MAP estimate $\widehat{\Omega}$
13: Compute Laplace covariance $H(\hat{\theta})^{-1}$ for uncertainty quantification

2.5. MAP Estimation Algorithm.

2.6. Uncertainty Quantification via Laplace Approximation. Posterior uncertainty was approximated using a Laplace approximation around the maximum a posteriori (MAP) estimate,

$$\hat{\theta} = \arg\min_{\theta} \left[-\log p(\theta \mid \mathbf{Y}) \right]. \tag{17}$$

Let

$$H(\hat{\theta}) = \nabla^2 \left[-\log p(\theta \mid \mathbf{Y}) \right] \Big|_{\theta=\hat{\theta}} \tag{18}$$

be the Hessian at the MAP. The posterior is then approximated as

$$\theta \mid \mathbf{Y} \approx N\left(\hat{\theta}, H(\hat{\theta})^{-1}\right), \tag{19}$$

where $H(\hat{\theta})^{-1}$ estimates the posterior covariance. Samples from this approximation were used to compute RMSE, coverage probability, and credible interval length.

This approach is justified as the regularized horseshoe prior yields a smooth posterior near the MAP, allowing curvature-based approximation.

Diagnostics under a contaminated AR(2) setting ($p = 50$, $n = 200$, $\gamma = 0.01$) supported validity: the Hessian was positive definite (eigenvalues 2.87×10^{-4} to 5.25×10^{-1}, condition number 1.83×10^{3}), variances were stable across small and large edges, and a posterior predictive check gave $p = 0.32$, indicating no lack of fit. These diagnostics support the use of the Laplace approximation as a computationally tractable and numerically stable approach for posterior uncertainty quantification in this setting.

2.7. Implementation. All methods are implemented in Python using PyTorch for automatic differentiation. Numerical stability is ensured through Cholesky parameterization of the precision matrix, log-sum-exp evaluation for the γ-divergence likelihood, and explicit symmetrization of matrix operations to maintain positive definiteness.

MAP estimation is performed using the Adam optimizer as described in Section 2.4. Posterior uncertainty is computed using the Laplace approximation described in Section 2.6. Reproducible code and scripts are provided in the supplementary materials.

TABLE 1. Implementation and experimental settings used in the study.

Component	**Setting**
Simulation design	
Sample size	$n = 200$
Number of variables	$p = 12, 50, 100$
Robustness parameter	$\gamma = 0.01, 0.02$
Prior specification	
Off-diagonal prior	Regularized horseshoe
Global shrinkage prior	$\tau \sim \mathrm{C}^{+}(0,1)$
Slab parameter	$c^2 \sim$ Inverse-Gamma$(2,1)$
Diagonal prior	$\mathbf{L}_{ii} \sim \mathrm{Exp}(\alpha)$
Computation	
MAP optimizer	Adam
Maximum iterations	800
Posterior approximation	Laplace approximation
Posterior samples	500

The settings in Table 1 are chosen to reflect moderate to high-dimensional precision matrix estimation scenarios while ensuring stable estimation. The sample size $n = 200$ and dimensions $p = 12, 50, 100$ allow evaluation of the method across increasing model complexity. The robustness parameter ($\gamma = 0.01, 0.02$) introduces a mild robustness against contamination while maintaining efficiency under near-Gaussian data. The remaining settings follow standard choices in sparse Bayesian graphical modeling; in particular, a brief sensitivity analysis comparing diag_rate $= 0.5$ and diag_rate $= 0.1$ produced nearly identical estimates and graph recovery metrics, indicating that the results are not sensitive to moderate changes in this hyperparameter.

3. THEORETICAL PROPERTIES OF THE γ-POSTERIOR OBJECTIVE

We consider a robust Bayesian approach to precision matrix estimation based on the γ-divergence combined with a sparsity-inducing regularized horseshoe prior on the Cholesky

factor. This section establishes key theoretical properties of the resulting posterior distribution, including posterior propriety, existence of a maximum a posteriori (MAP) estimator, and posterior robustness. Theoretical properties of γ-divergence-based posteriors have been established in prior work, including Onizuka and Hashimoto [21], Hashimoto and Sugasawa [26], and the general Bayesian framework of Bissiri et al. [27]. The results presented here adapt these findings to the present model, which employs a Cholesky parameterization for computational stability and a regularized horseshoe prior for adaptive shrinkage.

Let $\Omega \in \mathscr{M}^+$ denote the precision matrix of a p-dimensional Gaussian graphical model, where $\mathscr{M}^+$ is the cone of symmetric positive definite matrices. Let $\mathbf{Y} = (y_1, \ldots, y_n)^\top \in \mathbb{R}^{n \times p}$ denote the observed data, with

$$y_i \sim \mathscr{N}_p(0, \Omega^{-1}).$$

Under the γ-divergence framework [18], the posterior distribution is defined as follows:

$$\pi_\gamma(\Omega \mid \mathbf{Y}) = \frac{|\Omega|^{\frac{1}{2(1+\gamma)}} \left(\sum_{i=1}^n \exp\left(-\frac{\gamma}{2} y_i^\top \Omega y_i\right)\right)^{1/\gamma} \pi(\Omega)}{\int_{\mathscr{M}^+} |\Omega|^{\frac{1}{2(1+\gamma)}} \left(\sum_{i=1}^n \exp\left(-\frac{\gamma}{2} y_i^\top \Omega y_i\right)\right)^{1/\gamma} \pi(\Omega)\, d\Omega}. \tag{20}$$

for $\gamma > 0$. The case $\gamma = 0$ corresponds to the limit that recovers the standard Gaussian log-likelihood.

3.1. Prior Specification. To ensure positive definiteness and facilitate computation, we parameterize the precision matrix via the Cholesky decomposition

$$\Omega = \mathbf{L}\mathbf{L}^\top, \tag{21}$$

where $\mathbf{L}$ is a lower triangular matrix with strictly positive diagonal entries. This parameterization provides a bijection between $\mathscr{M}^+$ and the set $\mathscr{L}_+ = \{\mathbf{L} \in \mathbb{R}^{p \times p} : \mathbf{L} \text{ lower triangular}, \mathbf{L}_{ii} > 0\}$.

The prior is specified directly on the Cholesky factor $\mathbf{L}$ as

$$\pi(\mathbf{L}) = \left(\prod_{i>j} \pi_{\text{RHS}}(\mathbf{L}_{ij})\right) \left(\prod_{i=1}^p \pi_{\text{Exp}}(\mathbf{L}_{ii})\right), \tag{22}$$

where:

- For $i > j$, the off-diagonal elements follow a regularized horseshoe (RHS) prior [14]. This hierarchical prior is proper and heavy-tailed, with density bounded above by a constant multiple of a Cauchy density. A full specification is given in Section 2.3.2.
- For $i = 1, \ldots, p$, the diagonal elements follow an exponential prior $\mathbf{L}_{ii} \sim \text{Exp}(\alpha)$ with rate $\alpha > 0$, so that $\pi(\mathbf{L}_{ii}) = \alpha e^{-\alpha \mathbf{L}_{ii}}$ for $\mathbf{L}_{ii} > 0$.

Since the prior on $\mathbf{L}$ is proper and supported on $\mathscr{L}_+$, it induces a proper prior on Ω supported on $\mathscr{M}^+$ via the transformation $\Omega = \mathbf{L}\mathbf{L}^\top$. Although the induced prior on Ω does not admit a closed-form expression, its support and tail behavior are inherited from the prior on $\mathbf{L}$ through the transformation.

When working with the Cholesky parameterization, we write $\pi_\gamma(\mathbf{L} \mid \mathbf{Y})$ to denote the posterior expressed in terms of the Cholesky factor, where $\Omega = \mathbf{L}\mathbf{L}^\top$.

3.2. Posterior Propriety. We first establish that the γ-posterior in (20) is well-defined under the proposed prior.

Theorem 3.1 (Posterior propriety; Onizuka and Hashimoto [21])**.** *Assume that the prior $\pi(\Omega)$ is supported on $\mathscr{M}^+$ and that for each $i = 1, \ldots, p$, there exists an integrable function g_i such that*

$$\omega_{ii}^{\frac{1}{2(1+\gamma)}} \pi(\omega_{ii}) \leq g_i(\omega_{ii})$$

almost everywhere, where $\pi(\omega_{ii})$ denotes the marginal prior density of the ith diagonal element of Ω. Then the γ-posterior $\pi_\gamma(\Omega \mid \mathbf{Y})$ is proper for all $\mathbf{Y} \in \mathbb{R}^{n \times p}$.

We now verify that the prior induced by our specification satisfies the conditions of Theorem 3.1.

Proposition 3.2. *Under the Cholesky parameterization $\Omega = \mathbf{L}\mathbf{L}^\top$ with $\mathbf{L}_{ii} \sim Exp(\alpha)$ ($\alpha > 0$) and proper priors on the off-diagonal elements, the induced prior $\pi(\Omega)$ satisfies the integrability condition of Theorem 3.1. Consequently, the γ-posterior is proper.*

Proof. The proof proceeds in three steps.

Step 1: Off-diagonal contribution. The regularized horseshoe prior is proper with a finite normalizing constant [14]. Since propriety is unaffected by the presence of proper priors on other parameters, it suffices to analyze the diagonal contribution.

Step 2: Transformation to Ω. For any fixed off-diagonal elements, the mapping from the diagonal elements $\{\mathbf{L}_{ii}\}$ to $\{\omega_{ii}\}$ is given by $\omega_{ii} = \sum_{k=1}^{i} \mathbf{L}_{ik}^2$. In particular, $\omega_{ii} \geq \mathbf{L}_{ii}^2$, so that $\mathbf{L}_{ii} \leq \sqrt{\omega_{ii}}$. The Jacobian of the transformation from $\mathbf{L}$ to Ω is

$$J(\mathbf{L} \to \Omega) = \prod_{i=1}^{p} \mathbf{L}_{ii}^{p-i+1},$$

which is a polynomial function of the diagonal elements.

Step 3: Marginal tail bound. The prior on $\mathbf{L}_{ii}$ is exponential: $\pi(\mathbf{L}_{ii}) = \alpha e^{-\alpha \mathbf{L}_{ii}}$. Combining the prior with the Jacobian and integrating over the off-diagonal elements (which contribute a bounded factor due to propriety), we obtain the following bound for the marginal prior density of ω_{ii}:

$$\pi(\omega_{ii}) \leq C \cdot \omega_{ii}^{m/2} e^{-\alpha\sqrt{\omega_{ii}}},$$

for some constants $C > 0$ and integer $m \geq 0$. The factor $\omega_{ii}^{m/2}$ arises from the polynomial growth of the Jacobian and the inequality $\mathbf{L}_{ii} \leq \sqrt{\omega_{ii}}$; the exponential factor arises from the exponential prior.

Multiplying by the term $\omega_{ii}^{\frac{1}{2(1+\gamma)}}$ required in Theorem 3.1 yields

$$\omega_{ii}^{\frac{1}{2(1+\gamma)}} \pi(\omega_{ii}) \leq C \cdot \omega_{ii}^{\frac{1}{2(1+\gamma)} + \frac{m}{2}} e^{-\alpha\sqrt{\omega_{ii}}}.$$

For any exponents $a, b > 0$, the function $\omega^a e^{-b\sqrt{\omega}}$ is integrable on $(0, \infty)$ because exponential decay dominates any polynomial growth. Hence the right-hand side is integrable, and the condition of Theorem 3.1 is satisfied. □

Remark 3.3. Proposition 3.2 holds for any fixed $\gamma > 0$ and for any $\alpha > 0$. The result is conditional on the hyperparameters; in practice, hyperparameters are either fixed or assigned proper hyperpriors, which preserves propriety.

3.3. Existence of a MAP Estimator. We now establish that the negative log-posterior attains a minimum, guaranteeing the existence of a MAP estimator. Let $\mathscr{L}(\mathbf{L}) = -\log \pi_\gamma(\mathbf{L} \mid \mathbf{Y})$ denote the objective function, where the posterior is expressed in terms of the Cholesky factor $\mathbf{L}$.

Proposition 3.4. *Under the prior specification in Section 2.3.2, the objective function $\mathscr{L}(\mathbf{L})$ is continuous on the domain $\mathscr{L}_+ = \{\mathbf{L} : \mathbf{L}$ lower triangular$, \mathbf{L}_{ii} > 0\}$ and is coercive in the sense that $\mathscr{L}(\mathbf{L}) \to \infty$ as either $\|\mathbf{L}\|_F \to \infty$ or any diagonal element $\mathbf{L}_{ii} \to 0$. Consequently, $\mathscr{L}(\mathbf{L})$ attains a minimum on $\mathscr{L}_+$, and a MAP estimator exists.*

Proof. The γ-divergence loss is continuous in Ω on $\mathscr{M}^+$ [18], and the mapping $\mathbf{L} \mapsto \Omega = \mathbf{L}\mathbf{L}^\top$ is continuous. The prior terms are continuous on $\mathscr{L}_+$. Hence $\mathscr{L}$ is continuous.

We establish coercivity. The negative log-posterior consists of three contributions: the γ-divergence loss, the exponential prior on the diagonals, and the regularized horseshoe prior on the off-diagonals.

First, consider the behavior as $\mathbf{L}_{ii} \to 0$. The γ-divergence loss contains a term $-\frac{1}{2(1+\gamma)} \log |\mathbf{L}\mathbf{L}^\top| = -\frac{1}{1+\gamma} \sum_i \log \mathbf{L}_{ii}$, which diverges to $+\infty$ as any $\mathbf{L}_{ii} \to 0$ because $-\log \mathbf{L}_{ii} \to \infty$. Thus $\mathscr{L}(\mathbf{L}) \to \infty$ as any diagonal element approaches zero.

Next, consider $\|\mathbf{L}\|_F \to \infty$. If a diagonal element $\mathbf{L}_{ii} \to \infty$, then the exponential prior contributes $\alpha \mathbf{L}_{ii} \to \infty$.

If only off-diagonal elements diverge, the regularized horseshoe prior remains bounded below due to its heavy-tailed but proper structure. In this case, coercivity is driven by the γ-divergence loss. For any fixed $y_i \neq 0$, note that

$$y_i^\top \Omega y_i = y_i^\top \mathbf{L}\mathbf{L}^\top y_i = \|\mathbf{L}^\top y_i\|^2.$$

As $\|\mathbf{L}\|_F \to \infty$ while diagonal elements remain bounded, at least one off-diagonal entry of $\mathbf{L}$ diverges. Since $\mathbf{L}$ is lower triangular, $\mathbf{L}^\top y_i$ is a linear combination of the rows of $\mathbf{L}^\top$ with coefficients given by the components of y_i. Because $y_i \neq 0$, at least one component of $\mathbf{L}^\top y_i$ diverges, implying $\|\mathbf{L}^\top y_i\|^2 \to \infty$. Consequently,

$$\exp\left(-\frac{\gamma}{2} y_i^\top \Omega y_i\right) = \exp\left(-\frac{\gamma}{2} \|\mathbf{L}^\top y_i\|^2\right) \to 0.$$

Hence the empirical average inside the logarithm in the γ-divergence loss converges to zero as $\|\mathbf{L}\|_F \to \infty$. It follows that the negative log-likelihood term diverges to $+\infty$.

Since the prior contribution is bounded below and does not offset this divergence, the overall objective function $\mathscr{L}(\mathbf{L})$ diverges as $\|\mathbf{L}\|_F \to \infty$, establishing coercivity. Hence $\mathscr{L}(\mathbf{L})$ is coercive.

Since $\mathscr{L}$ is continuous and coercive, its sublevel sets

$$\{\mathbf{L} : \mathscr{L}(\mathbf{L}) \leq M\}$$

are bounded. Being closed subsets of a finite-dimensional space, they are therefore compact. Hence, $\mathscr{L}$ attains its minimum on each sublevel set.

Moreover, since $\mathscr{L}(\mathbf{L}) \to \infty$ as any $\mathbf{L}_{ii} \to 0$, the minimum cannot occur on the boundary where some $\mathbf{L}_{ii} = 0$. Therefore, a minimizer exists in the interior $\mathscr{L}_+$. □

Remark 3.5. The objective $\mathscr{L}(\mathbf{L})$ is non-convex due to the γ-divergence loss and the hierarchical shrinkage prior. Consequently, the MAP estimator is not guaranteed to be unique, and numerical optimization may converge to a local rather than global minimum. In practice, we initialize from a sensible starting point and employ early stopping to improve stability.

3.4. Posterior Robustness. A key motivation for using the γ-divergence is its robustness to outliers. The following result formalizes that extreme observations have vanishing influence on the posterior.

Theorem 3.6 (Posterior robustness)**.** *Assume that the γ-posterior is proper. Let $D = \{y_1, \ldots, y_n\}$ be a dataset containing a fixed set of m observations $K = \{y_1, \ldots, y_m\}$ and $n-m$ clean observations $L = \{y_{m+1}, \ldots, y_n\}$. Let $D^* = \{y_{m+1}, \ldots, y_n\}$ denote the dataset with the m observations removed. Then as $\|y_j\| \to \infty$ for $j = 1, \ldots, m$,*

$$\pi_\gamma(\Omega \mid D) \to \pi_\gamma(\Omega \mid D^*)$$

in total variation distance.

Proof. The posterior depends on the data through the term

$$\left(\sum_{i=1}^{n} \exp\left(-\frac{\gamma}{2} y_i^\top \Omega y_i \right) \right)^{1/\gamma}.$$

Decompose the sum as

$$\sum_{i=1}^{n} \exp\left(-\frac{\gamma}{2} y_i^\top \Omega y_i \right) = \underbrace{\sum_{i \in K} \exp\left(-\frac{\gamma}{2} y_i^\top \Omega y_i \right)}_{\text{outliers}} + \underbrace{\sum_{i \in L} \exp\left(-\frac{\gamma}{2} y_i^\top \Omega y_i \right)}_{\text{clean observations}}.$$

For each fixed Ω and each $i \in K$, as $\|y_i\| \to \infty$, we have $y_i^\top \Omega y_i \geq \lambda_{\min}(\Omega)\|y_i\|^2$, where $\lambda_{\min}(\Omega) > 0$ since $\Omega \in \mathscr{M}^+$. Hence $y_i^\top \Omega y_i \to \infty$, so

$$\exp\left(-\frac{\gamma}{2} y_i^\top \Omega y_i\right) \to 0.$$

Thus the contribution of the outliers to the sum converges pointwise to zero. Let

$$S_n(\Omega) = \sum_{i=1}^{n} \exp\left(-\frac{\gamma}{2} y_i^\top \Omega y_i\right), \qquad S_{n-m}(\Omega) = \sum_{i \in L} \exp\left(-\frac{\gamma}{2} y_i^\top \Omega y_i\right).$$

Since $0 \leq \exp\left(-\frac{\gamma}{2} y_i^\top \Omega y_i\right) \leq 1$, we have

$$\sum_{i \in K} \exp\left(-\frac{\gamma}{2} y_i^\top \Omega y_i\right) \leq m,$$

and consequently $S_n(\Omega) \leq S_{n-m}(\Omega) + m$. Because the function $x \mapsto x^{1/\gamma}$ is continuous on $(0, \infty)$, it follows that $S_n(\Omega)^{1/\gamma} \to S_{n-m}(\Omega)^{1/\gamma}$ pointwise as $\|y_i\| \to \infty$ for $i \in K$.

The γ-posterior can be written as

$$\pi_\gamma(\Omega \mid D) = \frac{|\Omega|^{\frac{1}{2(1+\gamma)}} S_n(\Omega)^{1/\gamma} \pi(\Omega)}{\int |\Omega|^{\frac{1}{2(1+\gamma)}} S_n(\Omega)^{1/\gamma} \pi(\Omega)\, d\Omega}.$$

Since $0 \leq S_n(\Omega)^{1/\gamma} \leq (S_{n-m}(\Omega) + m)^{1/\gamma}$ and the function $|\Omega|^{\frac{1}{2(1+\gamma)}} (S_{n-m}(\Omega) + m)^{1/\gamma} \pi(\Omega)$ is integrable by Proposition 3.2, it provides a dominating integrable function. Hence, by the dominated convergence theorem, as $\|y_i\| \to \infty$ for $i \in K$,

$$|\Omega|^{\frac{1}{2(1+\gamma)}} S_n(\Omega)^{1/\gamma} \pi(\Omega) \to |\Omega|^{\frac{1}{2(1+\gamma)}} S_{n-m}(\Omega)^{1/\gamma} \pi(\Omega)$$

pointwise, and the normalizing constant converges to $\int |\Omega|^{\frac{1}{2(1+\gamma)}} S_{n-m}(\Omega)^{1/\gamma} \pi(\Omega)\, d\Omega$. Therefore,

$$\pi_\gamma(\Omega \mid D) \to \frac{|\Omega|^{\frac{1}{2(1+\gamma)}} S_{n-m}(\Omega)^{1/\gamma} \pi(\Omega)}{\int |\Omega|^{\frac{1}{2(1+\gamma)}} S_{n-m}(\Omega)^{1/\gamma} \pi(\Omega)\, d\Omega} = \pi_\gamma(\Omega \mid D^*)$$

pointwise almost everywhere. Convergence in total variation follows from Scheffé's lemma since the densities are non-negative and integrate to one. □

Remark 3.7. Theorem 3.6 shows that the posterior becomes arbitrarily close to the posterior based only on the clean observations as the magnitude of the outliers diverges. This property holds for any fixed number of outliers and relies critically on the redescending nature of the

γ-likelihood. Under the prior specification of Section 2.3.2, posterior propriety is guaranteed by Proposition 3.2, so the robustness property applies.

4. Simulation Study

4.1. Simulation Design. A simulation study was conducted to evaluate the performance of the proposed robust Bayesian precision matrix estimation method with the regularized horseshoe prior in comparison with the Laplace prior method of [21]. The objective of the simulation was to examine the ability of the two priors to recover strong associations and accurately estimate the precision matrix under both clean and contaminated data settings.

Data were generated from Gaussian graphical models with autoregressive structures, which serve as data generating mechanisms for the precision matrix.

- AR(1): $\Omega_{i,i-1} = \Omega_{i-1,i} = -0.5$ for $i = 2,\ldots,p$, with all other off-diagonal elements equal to zero.
- AR(2): $\Omega_{i,i-1} = \Omega_{i-1,i} = -0.5$ for $i = 2,\ldots,p$, and $\Omega_{i,i-2} = \Omega_{i-2,i} = -0.25$ for $i = 3,\ldots,p$, with all remaining off-diagonal elements equal to zero.

These autoregressive constructions are standard benchmarks in Gaussian graphical model simulations and are commonly used in the literature [6, 12]. The simulations were performed under three dimensional settings, namely $p = 12$, $p = 50$, and $p = 100$, with a fixed sample size of $n = 200$.

To assess robustness to misspecification and outliers, four scenarios were considered:

- Clean data (CLN): observations generated from $N(0, \Omega_0^{-1})$.
- Variance contamination (VAR): a proportion $\varepsilon = 0.1$ of observations had inflated covariance.
- Mean-shift contamination (MC): a proportion $\varepsilon = 0.1$ had shifted mean with magnitude $\eta = 10$.
- Heavy-tailed contamination (HT): a proportion $\varepsilon = 0.15$ was drawn from a t_3 distribution.

Similar contamination schemes have been used in robustness studies for Gaussian graphical models [28, 21].

4.1.1. *Model Estimation.* For each generated dataset, the precision matrix was estimated using the proposed γ-divergence-based method under two sparsity-inducing priors: the Laplace prior and the regularized horseshoe (RHS) prior. Two robustness levels were considered, $\gamma = 0.01$ and $\gamma = 0.02$.

Estimation was performed via maximum a posteriori (MAP) optimization using the Adam algorithm with early stopping for stable convergence. The resulting precision matrices were used to compute partial correlations for identifying strong associations and evaluating performance.

Performance metrics were computed over 100 Monte Carlo replications for each setting, except in the high-dimensional case ($p = 100$), where 10 replications were used due to computational constraints. For the Laplace prior, the global shrinkage parameter was set according to the theoretically motivated choice

$$\lambda = \sqrt{\frac{\log(p)}{n}},$$

which is commonly used in graphical lasso–type estimators and provides a dimension–dependent level of regularization.

Evaluation metrics: Model performance was assessed using the following measures:

- Normalized Frobenius error:

$$\frac{\|\widehat{\Omega} - \Omega_0\|_F}{\|\Omega_0\|_F}.$$

- Root Mean Squared Error (RMSE): letting

$$\widehat{\Omega} = \frac{1}{B}\sum_{b=1}^{B} \Omega^{(b)},$$

$$\mathrm{RMSE} = \sqrt{\frac{2}{p(p-1)}\sum_{i<j}\left(\widehat{\omega}_{ij} - \omega_{0,ij}\right)^2}.$$

- Average Length (AL) of 95% credible intervals:

$$\mathrm{AL} = \frac{2}{p(p-1)}\sum_{i<j}\left(\widehat{\omega}_{ij}^{(97.5\%)} - \widehat{\omega}_{ij}^{(2.5\%)}\right).$$

- Coverage Probability (CP):

$$\mathrm{CP} = \frac{2}{p(p-1)}\sum_{i<j}\mathbb{I}\left(\widehat{\omega}_{ij}^{(2.5\%)} \le \omega_{0,ij} \le \widehat{\omega}_{ij}^{(97.5\%)}\right).$$

Posterior Sampling. Posterior uncertainty was characterized differently for the two priors due to their structural properties.

For the regularized horseshoe (RHS) prior, the posterior is smooth near the MAP estimate $\hat{\theta}$, allowing a Laplace approximation. Let

$$H(\hat{\theta}) = \nabla^2 \left[-\log p(\theta \mid \mathbf{Y}) \right] \Big|_{\theta=\hat{\theta}}.$$

Then $\theta \mid \mathbf{Y} \approx \mathcal{N}(\hat{\theta}, H(\hat{\theta})^{-1})$, and samples from this approximation were used to compute RMSE, coverage probability, and interval length.

For the Laplace prior, the ℓ_1 penalty is non-differentiable at zero, making the Laplace approximation unsuitable. Uncertainty was instead estimated via non-parametric bootstrap: $B = 200$ resamples were drawn, and MAP estimates $\{\hat{\theta}^{(b)}\}_{b=1}^{B}$ were obtained.

These choices are dictated by the priors' properties. As a result, uncertainty metrics are not directly comparable across models; each is assessed by calibration to the nominal 95% level. In contrast, point estimation metrics (Frobenius error, RMSE) remain comparable since they depend only on MAP estimates.

4.2. Simulation Results.

4.2.1. *Frobenius Norm Error.* The normalized Frobenius errors for both AR(1) and AR(2) precision matrix structures are reported in Tables 2 and 3. Across all dimensions and contamination scenarios, the regularized horseshoe prior achieves lower Frobenius error across all approaches, particularly under data contamination. This improvement stems from its adaptive shrinkage mechanism of applying strong shrinkage to noise while allowing true signals to escape regularization. In contrast, the Laplace prior imposes uniform shrinkage, which either undershrinks noise in low dimensions or overshrinks signals in high dimensions. This advantage persists under both robustness levels, $\gamma = 0.01$ and $\gamma = 0.02$, and becomes more pronounced as dimensionality increases.

TABLE 2. Normalized Frobenius error for AR(1) structures.

	CLN		VAR		MC		HT	
p	Lap	RHS	Lap	RHS	Lap	RHS	Lap	RHS
$\gamma = 0.01$								
12	1.226	0.541	0.383	0.573	0.945	0.634	0.537	0.546
50	1.282	0.505	1.072	0.573	1.037	0.540	0.783	0.544
100	1.268	0.558	1.117	0.558	1.090	0.526	1.094	0.529
$\gamma = 0.02$								
12	1.227	0.550	0.679	0.540	0.977	0.632	0.624	0.552
50	1.282	0.517	1.066	0.580	1.115	0.549	1.065	0.563
100	1.266	0.570	1.116	0.567	1.198	0.539	1.253	0.560

TABLE 3. Normalized Frobenius error for AR(2) structures.

	CLN		VAR		MC		HT	
p	Lap	RHS	Lap	RHS	Lap	RHS	Lap	RHS
$\gamma = 0.01$								
12	1.403	0.584	0.517	0.606	1.059	0.648	0.654	0.583
50	1.925	0.763	1.592	0.650	1.648	0.681	1.720	0.640
100	1.867	0.788	1.606	0.652	1.629	0.681	1.696	0.634
$\gamma = 0.02$								
12	1.404	0.597	0.779	0.578	1.095	0.648	0.748	0.591
50	1.916	0.809	1.575	0.682	1.786	0.726	1.865	0.705
100	1.860	0.840	1.601	0.686	1.804	0.733	1.831	0.710

Shrinkage Parameters behavior: The hierarchical regularized horseshoe prior introduces two key parameters: the global shrinkage parameter τ and the slab parameter c^2. The parameter τ controls the overall level of shrinkage applied to the precision matrix entries, while c^2 determines the width of the slab component that allows large signals to escape shrinkage.

Table 4 reports the estimated global shrinkage parameter τ and slab variance c^2 for the regularized horseshoe prior. The values of τ remain small across all dimensions, indicating strong shrinkage toward sparse precision matrices. The slab parameter c^2 is estimated adaptively from the data, allowing the model to adjust shrinkage for larger signals. Under contamination scenarios, slightly smaller τ values are observed, reflecting stronger shrinkage in the presence of noisy observations.

TABLE 4. Adaptive shrinkage parameters of the regularized horseshoe prior (τ and c^2) across dimensions and contamination scenarios.

	CLN		VAR		MC		HT	
p	τ	c^2	τ	c^2	τ	c^2	τ	c^2
AR(1)								
12	0.080	0.514	0.034	0.451	0.068	0.479	0.040	0.449
50	0.056	0.843	0.048	0.404	0.051	0.736	0.044	0.456
100	0.057	0.814	0.053	0.547	0.054	0.721	0.050	0.569
AR(2)								
12	0.082	0.514	0.034	0.451	0.073	0.502	0.043	0.448
50	0.064	0.866	0.055	0.519	0.057	0.744	0.050	0.500
100	0.065	0.839	0.060	0.644	0.061	0.750	0.058	0.644

Note: Values represent Monte Carlo means. The standard deviations for τ range approximately from 0.000 to 0.006 across scenarios, while the standard deviations for c^2 range approximately from 0.0001 to 0.033.

4.3. Posterior Uncertainty Metrics. Tables 5, 6, and 7 present the posterior uncertainty metrics under both AR(1) and AR(2) structures. These metrics characterize each model's uncertainty properties using the method most appropriate to its prior structure: the Laplace approximation for the smooth regularized horseshoe posterior, and bootstrap resampling for the non-differentiable Laplace prior. Direct cross-model comparisons of coverage probability or interval length are not intended, as the approximation methods differ fundamentally. Instead, uncertainty calibration for each model is evaluated independently against the nominal 95% target. In

contrast, point estimation metrics, such as RMSE and Frobenius error, remain directly comparable across models since they depend only on the MAP point estimates and are unaffected by the choice of uncertainty approximation.

RMSE (Point Estimate Accuracy). As shown in Table 5, the regularized horseshoe prior yields low RMSE values across both AR(1) and AR(2) structures, with errors decreasing as the dimension increases from $p = 12$ to $p = 50$. This pattern is observed across both clean and contaminated scenarios, indicating improved estimation accuracy in moderate-dimensional settings. The consistent reduction in RMSE reflects the effect of adaptive shrinkage, which allows small coefficients to be strongly regularized while preserving larger signals.

The Laplace prior exhibits a different pattern across the same settings, with RMSE values that do not decrease as markedly with increasing dimension. This behavior is consistent with its uniform shrinkage mechanism, which applies the same level of regularization across coefficients. The patterns observed in Table 5 highlight the influence of the prior structure on estimation accuracy.

TABLE 5. Root Mean Squared Error (RMSE) under AR(1) and AR(2) structures with $\gamma = 0.01$. Standard errors are in parentheses.

Scenario	Method	AR(1)		AR(2)	
		$p = 12$	$p = 50$	$p = 12$	$p = 50$
Clean	Laplace	0.2733 (0.030)	0.3075 (0.018)	0.2905 (0.033)	0.4441 (0.036)
	Horseshoe	0.1739 (0.009)	0.0563 (0.003)	0.1987 (0.007)	0.0788 (0.003)
Variance	Laplace	0.1846 (0.013)	0.2219 (0.010)	0.2084 (0.012)	0.3358 (0.022)
	Horseshoe	0.2041 (0.0001)	0.0979 (0.001)	0.2261 (0.0001)	0.1094 (0.001)
Mean shift	Laplace	0.2071 (0.022)	0.2532 (0.013)	0.2374 (0.022)	0.3570 (0.024)
	Horseshoe	0.1920 (0.004)	0.0756 (0.002)	0.2145 (0.003)	0.0929 (0.002)
Heavy tail	Laplace	0.1178 (0.012)	0.1896 (0.013)	0.1472 (0.012)	0.2814 (0.021)
	Horseshoe	0.2037 (0.0001)	0.0932 (0.002)	0.2255 (0.001)	0.1053 (0.002)

Coverage Probability (Uncertainty Calibration). As shown in Table 6, coverage probabilities for both methods move closer to the nominal level of 0.95 as the dimension increases, reflecting improved uncertainty calibration in higher dimensions. At $p = 50$, the horseshoe prior achieves

coverage probabilities generally close to the nominal level across all contamination scenarios under the AR(1) structure, with similarly strong performance under AR(2). The Laplace prior also shows reasonable calibration at $p = 50$, though with greater variability across scenarios. These results indicate that both methods produce reasonably well-calibrated uncertainty estimates in moderate-dimensional settings, with the horseshoe prior showing slightly tighter alignment with the nominal level.

TABLE 6. Coverage Probability (CP) under AR(1) and AR(2) structures with $\gamma = 0.01$. Target coverage is 0.95. Standard errors in parentheses.

Scenario	Method	AR(1)		AR(2)	
		$p = 12$	$p = 50$	$p = 12$	$p = 50$
Clean	Laplace	0.802 (0.031)	0.927 (0.008)	0.806 (0.035)	0.918 (0.009)
	Horseshoe	0.881 (0.031)	0.990 (0.003)	0.734 (0.022)	0.951 (0.003)
Variance	Laplace	0.847 (0.050)	0.957 (0.009)	0.835 (0.041)	0.953 (0.009)
	Horseshoe	0.833 (0.0001)	0.960 (0.001)	0.682 (0.0001)	0.922 (0.001)
Mean shift	Laplace	0.828 (0.040)	0.927 (0.007)	0.808 (0.044)	0.930 (0.007)
	Horseshoe	0.844 (0.012)	0.974 (0.003)	0.696 (0.005)	0.935 (0.003)
Heavy tail	Laplace	0.922 (0.034)	0.949 (0.009)	0.906 (0.030)	0.951 (0.009)
	Horseshoe	0.833 (0.0001)	0.962 (0.001)	0.682 (0.003)	0.924 (0.002)

Average Length (Interval Efficiency). As shown in Table 7, the regularized horseshoe prior produces tightly concentrated credible intervals across both AR(1) and AR(2) structures, with interval lengths remaining small as the dimension increases from $p = 12$ to $p = 50$. This indicates more localized posterior uncertainty in moderate-dimensional settings, suggesting greater precision in the interval estimates. Across contamination scenarios, the interval lengths remain relatively stable, indicating that the posterior is not overly dispersed even in the presence of noise. This behavior is consistent with the shrinkage structure of the prior, which concentrates uncertainty around relevant signals while limiting the spread from weaker coefficients. The Laplace prior exhibits a different pattern across the same settings, with interval lengths that are more variable across scenarios, reflecting the combined effect of uniform shrinkage and bootstrap-based uncertainty quantification.

TABLE 7. Average Credible Interval Length (AL) under AR(1) and AR(2) structures with $\gamma = 0.01$. Standard errors in parentheses.

Scenario	Method	AR(1)		AR(2)	
		$p = 12$	$p = 50$	$p = 12$	$p = 50$
Clean	Laplace	0.4888 (0.026)	0.9649 (0.042)	0.6049 (0.037)	1.4450 (0.080)
	Horseshoe	0.1851 (0.016)	0.1344 (0.004)	0.1946 (0.018)	0.1594 (0.006)
Variance	Laplace	0.4850 (0.033)	0.9349 (0.024)	0.5687 (0.035)	1.3822 (0.070)
	Horseshoe	0.0476 (0.002)	0.1178 (0.002)	0.0478 (0.003)	0.1428 (0.004)
Mean shift	Laplace	0.4434 (0.021)	0.8694 (0.035)	0.5252 (0.026)	1.2589 (0.067)
	Horseshoe	0.1145 (0.010)	0.1139 (0.003)	0.1222 (0.008)	0.1317 (0.004)
Heavy tail	Laplace	0.4043 (0.028)	0.7566 (0.030)	0.4858 (0.038)	1.1259 (0.062)
	Horseshoe	0.0763 (0.009)	0.1016 (0.003)	0.0837 (0.009)	0.1200 (0.004)

The results shown correspond to $\gamma = 0.01$. Simulations with $\gamma = 0.02$ yielded nearly identical RMSE and coverage probability values, with only small changes in interval length.

5. APPLICATION

5.1. Golub Leukemia Dataset Description. The proposed method was applied to the Golub leukemia gene expression dataset [29], available via `Bioconductor` [30]. The dataset consists of 7,129 genes measured across 72 samples. The analysis was conducted on the 50 genes with the largest sample variances, after standardizing the data to zero mean and unit variance.

5.2. Data Analysis. The proposed method was applied to the Golub leukemia dataset ($n = 72, p = 50$) and compared with the Laplace prior approach. Uncertainty was quantified via bootstrap for the Laplace model and via Laplace approximation for the horseshoe model, consistent with the simulation setup.

Strong associations were identified by thresholding absolute partial correlations from $\hat{\Omega}$, with results reported at $|pc_{ij}| > 0.05$. This threshold balances sensitivity and specificity, and sensitivity analysis (Table 8) confirms that conclusions are robust to its choice.

As shown in Table 8, the horseshoe model consistently detects more associations across all thresholds, and the Jaccard indices between the two methods remain relatively stable, indicating that the differences in inferred association patterns are systematic rather than threshold-dependent.

TABLE 8. Threshold sensitivity analysis for association selection.

Threshold	Laplace Edges	Horseshoe Edges	Shared Edges	Jaccard
0.01	213	238	67	0.174
0.03	145	183	43	0.151
0.05	106	160	37	0.162
0.10	69	104	26	0.177

5.2.1. *Association Patterns (Threshold = 0.05).* Focusing on the threshold $|\rho_{ij}| > 0.05$, we compare the properties of the inferred association patterns derived from the estimated precision matrices. As shown in Table 9, the horseshoe prior yields a substantially denser set of associations (160 associations, density 13.1%) compared to the Laplace prior (106 associations, density 8.7%). Additionally, the horseshoe model identifies higher-degree hub variables (maximum degree 24 vs 11), indicating stronger connectivity patterns. We refer to these structures as association networks to emphasize that they represent thresholded partial correlations rather than conditional independence claims.

This difference reflects the effect of adaptive shrinkage, the horseshoe prior selectively shrinks small partial correlations while allowing larger signals to remain relatively unshrunk. In contrast, the Laplace prior enforces uniform shrinkage, which can overshrink strong associations and produce overly sparse patterns. These differences are further illustrated in Figure 1.

TABLE 9. Summary statistics for identified associations ($|\rho_{ij}| > 0.05$)

Method	Associations	Density	Max Degree	Mean $\lvert\rho_{ij}\rvert$	Median $\lvert\rho_{ij}\rvert$
Laplace	106	0.0865	11	0.0185	0.0003
Horseshoe	160	0.1306	24	0.0288	0.0003

5.2.2. *Uncertainty Quantification.* Table 10 presents uncertainty diagnostics for each model. For the regularized horseshoe model, the Laplace approximation yields an average credible interval length of 0.1876 and a posterior standard deviation of 0.3001. These reflect the adaptive shrinkage properties of the prior, which preserves uncertainty around strong signals by allowing them to escape excessive regularization. The model identifies 60 associations with credible intervals that exclude zero, and 95.1% of intervals contain zero, consistent with appropriate sparsity calibration.

For the Laplace prior model, bootstrap resampling produces an average credible interval length of 0.0771 and a posterior standard deviation of 0.0739. The uniform shrinkage imposed by the ℓ_1 penalty yields narrower intervals and identifies 43 associations whose credible intervals exclude zero, with 96.5% of intervals containing zero.

Together, these results demonstrate that both methods achieve high rates of intervals containing zero, reflecting their ability to induce sparsity in high-dimensional settings. The differences in interval width and number of significant associations reflect the distinct shrinkage mechanisms of each prior: the Laplace prior enforces uniform regularization, while the regularized horseshoe prior adaptively differentiates between noise and signal.

TABLE 10. Uncertainty comparison between Laplace and horseshoe models.

Method	AL	AL Std	Sig Associations	% Zero in CI	% Extreme	Posterior SD
Laplace	0.0771	0.1019	43	96.49	0.00	0.0739
Horseshoe	0.1876	0.1288	60	95.10	0.00	0.3001

5.2.3. *Agreement Between Methods.* We assess agreement between the two methods in terms of shared associations and association strength correlation. As shown in Table 11, the overlap between the two sets of identified associations is limited, with only 37 shared associations out of 1225 possible pairs and a Jaccard index of 0.162. This indicates substantial differences in the inferred association patterns. However, the correlation of association strengths (0.549) suggests moderate agreement in the relative magnitude of partial correlations.

TABLE 11. Agreement between Laplace and horseshoe models.

Shared Associations	Total Pairs	Jaccard Index	Overlap (%)	Correlation
37	1225	0.162	3.02	0.549

5.2.4. *Hub Variable Analysis.* The hub variable analysis (Table 12) identifies genes with high connectivity in the estimated association structure. The horseshoe prior detects highly connected hubs such as *ELANE*, a leukemia-related neutrophil gene, and ribosomal proteins (*RPL41*, *RPL18A*, *RPLP1*) associated with cellular proliferation.

In contrast, the Laplace prior identifies less connected hubs, including immune and regulatory genes such as *CXCL8*, *NFKBIA*, and *GPX1*. Overall, the horseshoe prior produces a more densely connected association pattern, consistent with adaptive shrinkage, whereas the Laplace prior yields a more conservative structure due to uniform shrinkage. Genes appearing across both methods, such as *RPL41* and *RPLP1*, suggest robust components of the association structure that warrant further biological investigation.

These findings should be interpreted as exploratory hypotheses rather than definitive conclusions about gene regulatory networks. Validation in independent datasets or through experimental follow-up is necessary to confirm these patterns.

TABLE 12. Top 10 hub genes identified by each method. Probe identifiers are reported with corresponding gene symbols in parentheses where available.

Rank	Horseshoe		Laplace	
	Gene	Degree	Gene	Degree
1	M27783_s_at (ELANE)	24	Y00787_s_at (CXCL8)	11
2	Z12962_at (RPL41)	16	AFFX-HSAC07/X00351_5_at	9
3	Z84721_cds2_at	16	Y00433_at (GPX1)	8
4	M96326_rna1_at	15	X80822_at (RPL18A)	8
5	X80822_at (RPL18A)	15	M27891_at	8
6	J04617_s_at	15	Z12962_at (RPL41)	8
7	AFFX-HSAC07/X00351_M_at	14	V00599_s_at (TUBB)	7
8	M17886_at (RPLP1)	14	M69043_at(NFKBIA)	6
9	M27891_at	12	M17886_at(RPLP1)	6
10	AFFX-HSAC07/X00351_3_at	11	AFFX-HUMRGE/M10098_5_at	6

Probe identifiers are reported alongside mapped gene symbols (in parentheses) using the `hgu95av2.db` annotation package. Probes without available mappings are retained using their original identifiers.

5.2.5. *Shrinkage Behaviour.* The regularized horseshoe prior employs adaptive shrinkage governed by the global shrinkage parameter τ and the slab parameter c^2. The global parameter τ controls the overall level of shrinkage applied to the off-diagonal elements of the Cholesky factor L, while c^2 controls the slab component that allows larger signals to escape excessive shrinkage. In the Golub leukemia data application, the estimated values were $\tau = 0.2201$ and $c^2 = 1.7867$. The small value of τ indicates strong global shrinkage, while the moderately large value of c^2 allows stronger associations in the resulting precision matrix $\hat{\Omega} = \hat{\mathbf{L}}\hat{\mathbf{L}}^\top$ to remain relatively unshrunk.

Compared with the Laplace prior, the regularized horseshoe prior produced a more flexible association structure. This reflects the difference between uniform shrinkage and adaptive

shrinkage: the Laplace prior tends to produce a more conservative structure, whereas the regularized horseshoe prior can retain stronger partial correlations while shrinking weaker signals. Therefore, the denser association patterns obtained from the regularized horseshoe prior should be interpreted as exploratory evidence of potentially relevant associations rather than definitive biological conclusions.

The network plot and heatmaps in Figures 1 and 2 visually support these results. The regularized horseshoe prior produces a denser association structure, while the Laplace prior yields a more conservative network.

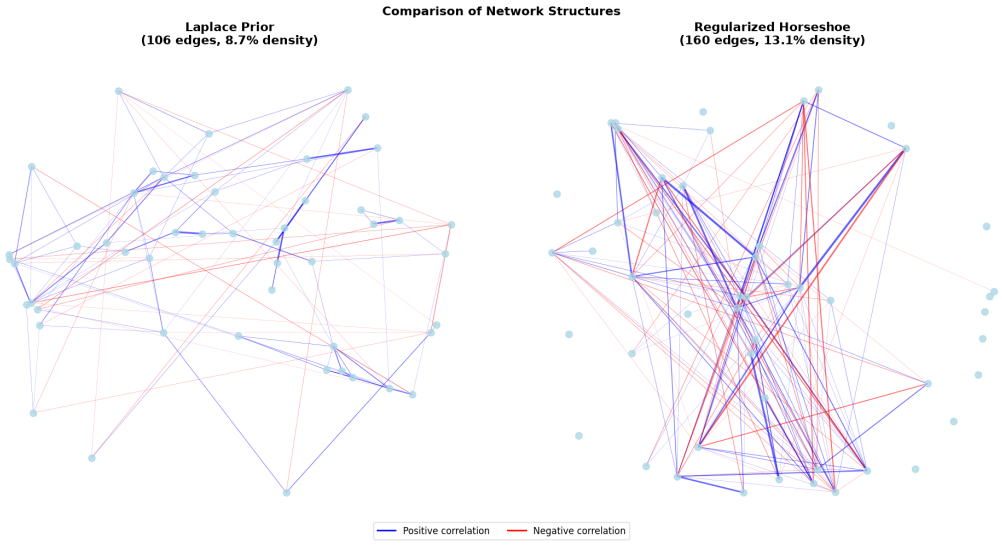

FIGURE 1. Comparison of association networks estimated using the Laplace and regularized horseshoe priors at threshold $|\rho_{ij}| > 0.05$.

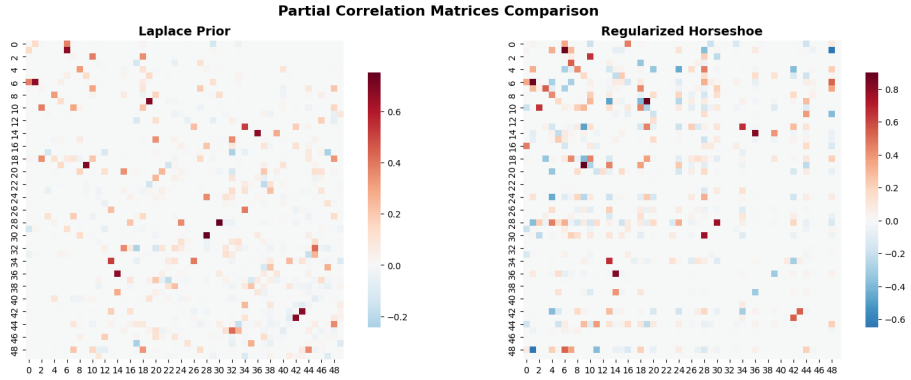

FIGURE 2. Heatmap of partial correlations estimated using the regularized horseshoe prior and the Laplace prior.

6. Discussion

The results show that combining the γ-divergence with the regularized horseshoe prior improves robust sparse precision matrix estimation when compared with the Laplace prior framework. Previous robust Bayesian graphical modelling work based on γ-divergence used Laplace-type shrinkage to induce sparsity [21]. Although this approach improves robustness to outliers, the Laplace prior applies more uniform shrinkage. In contrast, the regularized horseshoe prior provides adaptive shrinkage, allowing weak signals to be strongly regularized while preserving stronger associations [13, 14].

The simulation results suggest that this adaptive shrinkage improves estimation accuracy, especially in moderate-dimensional and larger model settings. The real-data application further shows that the proposed method identifies a richer association structure than the Laplace prior. However, the additional associations should be interpreted as exploratory, since the networks are based on thresholded partial correlations and require further validation.

7. Conclusion

This paper proposed a robust Bayesian framework for precision matrix estimation combining the γ-divergence with a regularized horseshoe prior under a Cholesky parameterization. Theoretical properties including posterior propriety, existence of a MAP estimator, and robustness to extreme observations were established. Simulation results show the proposed approach achieves lower estimation error than the Laplace prior baseline, with pronounced gains under contamination, while coverage probabilities approach nominal levels. Application to the Golub leukemia dataset demonstrates practical utility, identifying hub genes with known relevance to leukemia biology. A limitation is that inference relies on MAP estimation with Laplace approximation, which may not fully capture posterior uncertainty in highly non-convex settings. Future work can focus on developing more robust uncertainty quantification strategies, extending the framework to non-Gaussian and dynamic settings, conducting broader comparative evaluations, and applying the method to additional high-dimensional biological datasets. Overall, the proposed framework provides a principled and computationally efficient approach to robust precision matrix estimation.

ACKNOWLEDGEMENT

The authors gratefully acknowledge Pan African University Institute for Basic Sciences, Technology, and Innovation (PAUSTI) for its invaluable support in facilitating this study.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] S.L. Lauritzen, Graphical Models, Oxford University Press, Oxford, 1996. https://doi.org/10.1093/oso/9780198522195.001.0001.

[2] A. Dobra, C. Hans, B. Jones, J.R. Nevins, G. Yao, et al., Sparse Graphical Models for Exploring Gene Expression Data, J. Multivar. Anal. 90 (2004), 196–212. https://doi.org/10.1016/j.jmva.2004.02.009.

[3] T.-H. Lee, E. Seregina, Optimal Portfolio Using Factor Graphical Lasso, J. Financ. Econom. 22 (2024), 670–695. https://doi.org/10.1093/jjfinec/nbad011.

[4] S.M. Smith, K.L. Miller, G. Salimi-Khorshidi, M. Webster, C.F. Beckmann, et al., Network Modelling Methods for fMRI, NeuroImage 54 (2011), 875–891. https://doi.org/10.1016/j.neuroimage.2010.08.063.

[5] R. Tibshirani, Regression Shrinkage and Selection via the Lasso, J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1996), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

[6] M. Yuan, Y. Lin, Model Selection and Estimation in the Gaussian Graphical Model, Biometrika 94 (2007), 19–35. https://doi.org/10.1093/biomet/asm018.

[7] O. Banerjee, L. El Ghaoui, A. d'Aspremont, Model Selection through Sparse Maximum Likelihood Estimation for Multivariate Gaussian or Binary Data, J. Mach. Learn. Res. 9 (2008), 485–516. https://jmlr.org/papers/v9/banerjee08a.html.

[8] J. Friedman, T. Hastie, R. Tibshirani, Sparse Inverse Covariance Estimation with the Graphical Lasso, Biostatistics 9 (2008), 432–441. https://doi.org/10.1093/biostatistics/kxm045.

[9] B. Jones, C. Carvalho, A. Dobra, C. Hans, C. Carter, et al., Experiments in Stochastic Computation for High-Dimensional Graphical Models, Stat. Sci. 20 (2005), 388–400. https://doi.org/10.1214/088342305000000304.

[10] H. Wang, Scaling It Up: Stochastic Search Structure Learning in Graphical Models, Bayesian Anal. 10 (2015), 351–377. https://doi.org/10.1214/14-ba916.

[11] T. Park, G. Casella, The Bayesian Lasso, J. Am. Stat. Assoc. 103 (2008), 681–686. https://doi.org/10.1198/016214508000000337.

[12] H. Wang, Bayesian Graphical Lasso Models and Efficient Posterior Computation, Bayesian Anal. 7 (2012), 867–886. https://doi.org/10.1214/12-ba729.

[13] C.M. Carvalho, N.G. Polson, J.G. Scott, The Horseshoe Estimator for Sparse Signals, Biometrika 97 (2010), 465–480. https://doi.org/10.1093/biomet/asq017.

[14] J. Piironen, A. Vehtari, Sparsity Information and Regularization in the Horseshoe and Other Shrinkage Priors, Electron. J. Stat. 11 (2017), 5018–5051. https://doi.org/10.1214/17-ejs1337si.

[15] Y. Li, B.A. Craig, A. Bhadra, The Graphical Horseshoe Estimator for Inverse Covariance Matrices, J. Comput. Graph. Stat. 28 (2019), 747–757. https://doi.org/10.1080/10618600.2019.1575744.

[16] D.R. Williams, Bayesian Estimation for Gaussian Graphical Models: Structure Learning, Predictability, and Network Comparisons, Multivar. Behav. Res. 56 (2021), 336–352. https://doi.org/10.1080/00273171.2021.1894412.

[17] M. Finegold, M. Drton, Robust Graphical Modeling of Gene Networks Using Classical and Alternative t-Distributions, Ann. Appl. Stat. 5 (2011), 1057–1080. https://doi.org/10.1214/10-aoas410.

[18] H. Fujisawa, S. Eguchi, Robust Parameter Estimation with a Small Bias against Heavy Contamination, J. Multivar. Anal. 99 (2008), 2053–2081. https://doi.org/10.1016/j.jmva.2008.02.004.

[19] S. Yonekura, S. Sugasawa, Adaptation of the Tuning Parameter in General Bayesian Inference with Robust Divergence, Stat. Comput. 33 (2023), 39. https://doi.org/10.1007/s11222-023-10205-7.

[20] T. Kawashima, H. Fujisawa, Robust Regression against Heavy Heterogeneous Contamination, Metrika 86 (2023), 421–442. https://doi.org/10.1007/s00184-022-00874-1.

[21] T. Onizuka, S. Hashimoto, Robust Bayesian Graphical Modeling Using γ-Divergence, J. Multivar. Anal. 209 (2025), 105461. https://doi.org/10.1016/j.jmva.2025.105461.

[22] D.P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, arXiv preprint arXiv:1412.6980, (2014). https://doi.org/10.48550/arXiv.1412.6980.

[23] J. Duchi, E. Hazan, Y. Singer, Adaptive Subgradient Methods for Online Learning and Stochastic Optimization, J. Mach. Learn. Res. 12 (2011), 2121–2159. https://jmlr.org/papers/v12/duchi11a.html.

[24] T. Tieleman, G. Hinton, Lecture 6.5-RMSprop: Divide the Gradient by a Running Average of its Recent Magnitude, Coursera: Neural Netw. Mach. Learn. 4 (2012), 26–31.

[25] D.C. Liu, J. Nocedal, On the Limited Memory BFGS Method for Large Scale Optimization, Math. Program. 45 (1989), 503–528. https://doi.org/10.1007/bf01589116.

[26] S. Hashimoto, S. Sugasawa, Robust Bayesian Regression with Synthetic Posterior Distributions, Entropy 22 (2020), 661. https://doi.org/10.3390/e22060661.

[27] P.G. Bissiri, C.C. Holmes, S.G. Walker, A General Framework for Updating Belief Distributions, J. R. Stat. Soc. Ser. B Stat. Methodol. 78 (2016), 1103–1130. https://doi.org/10.1111/rssb.12158.

[28] T. Kawashima, H. Fujisawa, Robust and Sparse Regression via γ-Divergence, Entropy 19 (2017), 608. https://doi.org/10.3390/e19110608.

[29] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, et al., Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring, Science 286 (1999), 531–537. https://doi.org/10.1126/science.286.5439.531.

[30] R.C. Gentleman, V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, et al., Bioconductor: Open Software Development for Computational Biology and Bioinformatics, Genome Biol. 5 (2004), R80. https://doi.org/10.1186/gb-2004-5-10-r80.