Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Bayesian Graphical Models under Positivity Constraints: A Scalable Generalized Likelihood Approach

Swarnali Raha
University of Florida
Partha Sarkar
ps24v@fsu.edu
Florida State University
Sirani M. Perera
Embry–Riddle Aeronautical University
Kshitij Khare
University of Florida

Research Article

Keywords:

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11616131/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Bayesian Graphical Models under Positivity Constraints: A Scalable Generalized Likelihood Approach

Swarnali Raha[1], Partha Sarkar[2*], Sirani M. Perera[3], Kshitij Khare[1]

[1]Department of Statistics, University of Florida, Gainesville, Florida, USA.
[2]Department of Statistics, Florida State University, Tallahassee, Florida, USA.
[3]Department of Mathematics, Embry-Riddle Aeronautical University, Daytona Beach, Florida, USA.

*Corresponding author(s). E-mail(s): ps24v@fsu.edu;
Contributing authors: swarnali.raha@ufl.edu; Sirani.Perera@erau.edu; kdkhare@ufl.edu;

Abstract

We develop a computationally scalable Bayesian framework for precision matrix estimation in Gaussian graphical models under total positivity constraints, which arise naturally in a wide range of applications involving positively associated variables. To overcome the high computational cost associated with using the Gaussian likelihood, we adopt a generalized Bayesian approach based on the D-trace loss, which eliminates the log-determinant term and enables efficient computation. Sparsity is induced via spike-and-slab priors, and the resulting generalized posterior is shown to be proper under conditions that hold almost universally in practice. Our primary contribution is a suite of efficient posterior sampling algorithms tailored to high-dimensional settings. Starting from a component-wise Gibbs sampler, we introduce a novel data augmentation scheme that induces conditional independence among precision matrix entries, enabling joint updates. By exploiting the Gram structure of the sample covariance matrix, we further develop a fast matrix-normal sampler that significantly reduces per-iteration complexity in high-dimensional settings. An interweaving strategy combines augmented and direct updates to improve mixing without sacrificing scalability. Experiments on synthetic and financial data demonstrate substantial computational gains over existing methods, while maintaining competitive estimation accuracy and improved recovery of structured dependencies.

Keywords: Gaussian graphical models, total positivity, generalized Bayesian inference, data augmentation, Gibbs sampling

1 Introduction

Graphical models represent dependence among random variables through a graph, where vertices correspond to variables and edges encode conditional associations. Let $\boldsymbol{Y} = (Y_1, \ldots, Y_p)^\top$ be a p-variate random vector with covariance matrix Σ and precision matrix $\Omega = \Sigma^{-1} = [\omega_{ij}]_{i,j=1}^p$. Under a Gaussian model, $\omega_{ij} = 0$ if and only if Y_i and Y_j are conditionally independent given all remaining variables. More generally, zeros in Ω correspond to vanishing partial correlations. Thus, graph estimation reduces to identifying the zero and nonzero off-diagonal entries of Ω.

The high-dimensional setting, where p is comparable to or exceeds the sample size n, has motivated

extensive work on sparse precision matrix estimation under the assumption of a sparse underlying graph (Friedman et al. 2008; Ravikumar et al. 2011; Lam and Fan 2009; Friedman et al. 2010; Cai et al. 2011; Ni et al. 2022; Shiratori and Takano 2024; Bhadra et al. 2024).

A particularly important class of graphical models imposes total positivity constraints (Ying et al. 2023; Lauritzen et al. 2019, 2021; Slawski and Hein 2015; Wang et al. 2020). In the Gaussian setting, this is equivalent to restricting Ω to be an M-matrix, that is, a positive definite matrix with nonpositive off-diagonal entries. Such models imply nonnegative partial correlations and constitute the Gaussian specialization of the broader MTP_2 framework, which encodes strong positive dependence (Lauritzen et al. 2019; Lake and Tenenbaum 2010; Agrawal et al. 2022; Egilmez et al. 2016). These dependence structures arise naturally in applications including finance, psychometrics, cognition, and signal processing.

Despite substantial progress on sparse graphical models, most work under total positivity constraints is frequentist. Bayesian methods, while offering uncertainty quantification, remain relatively underdeveloped. Existing Gaussian-likelihood-based samplers become difficult to scale in the presence of sign constraints, as enforcing the M-matrix restriction requires sampling from high-dimensional truncated distributions. This often leads to slow mixing and substantial computational overhead. Moreover, even in unconstrained settings, Bayesian samplers typically have per-iteration complexity $\mathcal{O}(p^4)$, with reductions to $\mathcal{O}(p^3)$ only under restricted regimes (Gao et al. 2026) (such as $n = o(\sqrt{p})$), making the constrained setting even more computationally challenging.

To address these challenges, we develop a generalized Bayesian framework based on the D-trace loss, replacing the Gaussian likelihood with a computationally tractable surrogate. Such loss-based formulations avoid the log-determinant term and have been widely used in graphical modeling (Meinshausen and Bühlmann 2006; Peng et al. 2009; Khare et al. 2015; Cai et al. 2011; ZHANG and ZOU 2014). Combined with spike-and-slab priors (George and McCulloch 1993), the resulting generalized posterior is shown to be proper and supports both sparsity learning and uncertainty quantification.

We first develop a direct element-wise Gibbs sampler for posterior computation. This direct sampler is straightforward and has per-iteration complexity $\mathcal{O}(p^3)$ for arbitrary n and p, which is lower than that of standard Gaussian-likelihood-based samplers. Nevertheless, this approach can still be computationally challenging in the high-dimensional regime $n \ll p$. Moreover, since the elements of the precision matrix are updated sequentially and are typically strongly dependent, the resulting MCMC updates may exhibit suboptimal results (see Section 4).

To alleviate these computational bottlenecks, we introduce a data-augmentation scheme motivated by the anti-correlated Gaussian approach of Zheng and Duan (2025). This framework was originally developed for multivariate regression, where augmentation reduces posterior dependence by inducing conditional independence. Generalizing this idea to the matrix-valued setting is nontrivial and requires a substantial reformulation; see Section 4.2. The resulting augmentation yields conditional independence among the entries of the precision matrix given the augmented variables, leading to the augmented Gibbs sampler in Algorithm 2.

Our augmented Gibbs sampler enjoys several benefits. First, it improves computational scalability relative to the direct sampler, particularly in high-dimensional regimes. While Gaussian-likelihood-based (unconstrained) samplers typically scale as $\mathcal{O}(p^4)$, with recent improvements (Gao et al. 2026) achieving $\mathcal{O}(p^3)$ under $n = o(\sqrt{p})$, our method has per-draw complexity $\mathcal{O}(min\{p^2n, p^3\})$, which becomes $o(p^{2.5})$ in the same regime. Second, the D-trace loss, together with the relaxation of the positive-definiteness constraint and the induced conditional independence, allows the M-matrix constraint to be incorporated naturally.

Finally, motivated by Yu and Meng (2011), we propose an interweaving Gibbs sampler that combines the efficiency of the augmented sampler with occasional updates from the direct sampler under the original parameterization. The augmented sampler enables efficient global exploration, while the direct sampler provides complementary local

updates. Empirically, this yields improved stability and mixing with minimal additional cost (see Algorithm 3 and Section 4.3).

Simulation studies demonstrate that the proposed methods achieve competitive, and often superior, performance in graph recovery and precision matrix estimation while maintaining scalability in high-dimensional settings. In an analysis of S&P 500 returns, the estimated networks recover sectoral structure more effectively than competing approaches, yielding higher modularity scores and clearer within-sector connectivity.

The remainder of the paper is organized as follows. Section 2 introduces the notation, background on generalized likelihood approaches, and the D-trace loss. Section 3 presents the proposed generalized Bayesian framework based on the D-trace loss and spike-and-slab priors, and establishes the propriety of the resulting (generalized) posterior. Section 4 develops the posterior sampling algorithms, including the direct Gibbs sampler, the augmented Gibbs sampler, and the interweaving Gibbs sampler. Section 5 studies the empirical performance of the proposed methods through simulation experiments and a real financial time-series data analysis. We conclude with a brief discussion.

2 Preliminaries

Notation. We begin by introducing the notations used in the paper. We define δ_c as the degenerate distribution at $c \in \mathbb{R}$. $N(\mu, \sigma^2; a, b)$ denotes the truncated normal distribution with mean μ and variance σ^2, restricted to the interval (a, b), $\Phi(.)$ is the cumulative distribution function of standard normal variable and $\text{IG}(a, b)$ denote the Inverse-Gamma distribution with shape parameter a and rate b. $I_A(\cdot)$ denotes the indicator function of a set A, which takes the value 1 on A and 0 elsewhere. Also, let $\mathcal{S}(p)$, $\mathcal{S}^*(p)$ and $\mathcal{S}^{**}(p)$ denote the collection of all $p \times p$ symmetric positive definite matrices with non-positive off-diagonal entries, the collection of all $p \times p$ symmetric matrices with positive diagonal and non-positive off-diagonal entries and the collection of all $p \times p$ symmetric strictly diagonally dominant matrices with non-positive off-diagonal entries, respectively. Finally, we write $\|A\|_F$ to denote the Frobenius norm of a matrix A. We use the standard asymptotic notations $O(\cdot)$ and $o(\cdot)$. Specifically, $f(n) = O(g(n))$ if there exist constants $(C > 0)$ and n_0 such that $|f(n)| \leq C|g(n)|$ for all $n \geq n_0$, whereas $f(n) = o(g(n))$ if $\lim_{n\to\infty} f(n)/g(n) = 0$.

2.1 Generalized likelihood and loss-based Bayesian inference

In many statistical problems, the full likelihood corresponding to a probabilistic data-generating model may be difficult to evaluate or computationally burdensome to work with, particularly in high-dimensional settings. This has motivated the use of likelihood-type objectives based on loss functions, which serve as tractable surrogates for the full likelihood. The use of such loss-based approaches has a long history in the statistical literature; see, for example, (Besag 1975; Lindsay 1988; Varin et al. 2011). These methods are often designed to target the parameter or functional of interest directly, while avoiding the additional complexity associated with specifying a complete probabilistic model for the data.

A key advantage of loss-based formulations is that they can lead to computationally efficient procedures by replacing components of the likelihood that are difficult to evaluate. In the context of Gaussian graphical models, a prominent example is the log-determinant term in the Gaussian likelihood, which incurs $\mathcal{O}(p^3)$ computational cost and requires positive definiteness of the precision matrix at each evaluation. To mitigate this burden, several alternative objectives have been proposed that avoid the log-determinant term while retaining desirable statistical properties. These include methods based on conditional likelihoods and pseudo-likelihoods, such as neighborhood selection (Meinshausen and Bühlmann 2006), SPACE (Peng et al. 2009), and SPLICE (Rocha et al. 2008), as well as formulations based on surrogate loss functions, including the symmetric lasso (Friedman et al. 2010), CONCORD (Khare et al. 2015), quasi-GHS (Zhang et al. 2022), CLIME (Cai et al. 2011), and the D-trace loss (ZHANG and ZOU 2014). These approaches typically reduce computational complexity and can be less sensitive to model misspecification, making them well suited for high-dimensional problems.

In the Bayesian setting, Bissiri et al. (2016) developed a principled framework for incorporating such loss-based objectives. In this approach, the likelihood contribution is replaced by an exponentiated negative loss function, yielding a *generalized posterior* of the form

$$\pi(\theta \mid \mathbf{Y}) \propto \exp\{-\lambda L(\theta; \mathbf{Y})\}\, \pi(\theta),$$

where $L(\theta; \mathbf{Y})$ is a chosen loss function, $\pi(\theta)$ is a prior distribution, and $\lambda > 0$ is a scaling parameter controlling the relative influence of the data. This construction provides a coherent updating rule for prior beliefs even when a fully specified likelihood is unavailable. The resulting generalized posterior retains many of the formal properties of Bayesian inference while allowing greater flexibility in modeling and computation.

In this paper, we adopt this perspective and construct a generalized Bayesian model based on the D-trace loss, which yields a computationally tractable posterior formulation while naturally accommodating structural constraints on the precision matrix. The specific form of the generalized likelihood used in our work is described in the sequel.

2.2 The D-trace loss

We conclude the preliminaries by introducing the D-trace loss and its key properties. The D-trace loss function for estimating Ω using a $p \times p$ positive definite matrix $\bar{\Sigma}$ is defined as

$$L_D(\Omega, \bar{\Sigma}) = \frac{1}{2} \operatorname{tr}(\Omega^2 \bar{\Sigma}) - \operatorname{tr}(\Omega). \tag{1}$$

The loss function is a difference of two trace terms, which motivates the name *D-trace*. The key properties of L_D include the following. First, $L_D(\Omega, \cdot)$ is a smooth convex function in Ω, leading to computational advantages. Second, the unique minimizer of $L_D(\Omega, \bar{\Sigma})$ is $(\bar{\Sigma})^{-1}$, enabling recovery of the precision matrix through minimization of the loss. Motivated by these properties, we employ the generalized likelihood induced by the D-trace loss for estimation of Ω. Finally, the loss is formulated in terms of the covariance matrix (second argument) rather than the precision matrix, allowing direct application in settings where only the covariance matrix is available.

Although the D-trace generalized likelihood shares certain properties with the standard Gaussian likelihood such as convexity and minimization at the inverse of the sample covariance matrix it also offers important advantages. In particular, the $\operatorname{tr}(\Omega)$ term replaces the $\log\det(\Omega)$ term in the Gaussian likelihood, reducing the associated $\mathcal{O}(p^3)$ computational cost to $\mathcal{O}(p)$. Moreover, evaluation of $\log\det(\Omega)$ requires Ω to be positive definite, which introduces additional computational overhead in MCMC implementations, as positive definiteness must be enforced at each iteration. This issue is naturally avoided in the D-trace framework. Indeed, in the subsequent section, we relax the positive definiteness constraint on Ω and allow Ω to vary in the space of symmetric matrices with positive diagonal entries. Further discussion of this relaxation is provided in Section 3.

3 Model Description

In this section, we provide a detailed description of the model used in the subsequent sections. Let $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ be a random sample from a p-variate mean-zero Gaussian distribution with covariance matrix $\Sigma = \Omega^{-1}$, where Ω denotes the corresponding precision matrix. We write $\mathbf{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n)^T$ for the resulting $n \times p$ observation matrix. As described in Section 1, our primary objective is to estimate the precision matrix $\Omega = ((\omega_{ij}))_{i,j=1}^{p}$ under the totally positive graphical model framework, where the off-diagonal entries of Σ are positive. In this setting, the problem reduces to estimating Ω subject to the M-matrix constraint (Ying et al. 2023), namely, $\omega_{ij} \leq 0$ for all $i \neq j$, $i, j \in \{1, \ldots, p\}$. As discussed before, rather than relying on a loss function derived from the strict Gaussian likelihood, we adopt a generalized likelihood based on the D-trace loss introduced in (1).

In particular, for a data matrix $\mathbf{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n)^T$ of dimension $n \times p$, the generalized likelihood of Ω based on the D-trace loss can be written as the following.

$$L_D(\Omega|\mathbf{Y}) = \exp\left\{ n \operatorname{tr}(\Omega) - \frac{n}{2} \operatorname{tr}(\Omega^2 S) \right\} \tag{2}$$

where, $S = n^{-1}\mathbf{Y}^T\mathbf{Y}$ with (i,j)-th element s_{ij}. Following Bissiri et al. (2016), we use the exponentiated negative D-trace loss function as our generalized likelihood.

Since our goal is to estimate the precision matrix in Bayesian setting, we will now introduce a prior distribution on Ω. We impose independent spike-and-slab priors on the off-diagonal entries of Ω to induce sparsity, and for the diagonal entries, we introduce iid flat priors. The prior assumptions on the entries of Ω are as formulated below. For $1 \leq i < j \leq p$, we assume

$$\begin{aligned} \omega_{ij} &\overset{iid}{\sim} (1-q)\delta_0 + q\,\mathcal{N}(0,\tau^2;-\infty,0) \\ \omega_{ii} &\propto I_{(0,\infty)}(\omega_{ii}) \end{aligned} \tag{3}$$

where $0 < q < 1$ and $\tau^2 > 0$ are the hyperparameters denoting the prior probability and the prior variance of the slab respectively.

Note that, the prior is supported on $\mathcal{S}^*(p)$, the collection of all $p \times p$ matrices with positive diagonal and non-positive off-diagonal entries, rather than $\mathcal{S}(p)$, the collection of all symmetric positive definite matrices with non-positive off-diagonal entries. This relaxation is feasible due to the structure of the D-trace generalized likelihood, which replaces the log-determinant term with a trace function. It does not create any issues for the task of sparsity selection, but needs to be addressed if positive definite estimates are needed for downstream applications. If a positive definite (PD) estimate of Ω is needed for a downstream application, then the PD relaxation may require an additional post-processing step. In particular, for a posterior draw or point estimate Ω that is not positive definite, one may project it to the positive definite cone by shifting its spectrum. Specifically, following the projection idea used in Samanta et al. (2022), one can replace Ω by

$$B(\Omega) = \begin{cases} \Omega, & \text{if } \Omega \text{ is PD}, \\ \Omega - \lambda_{\min}(\Omega) I_p + \epsilon I_p, & \text{if } \lambda_{\min}(\Omega) \leq 0 \end{cases}$$

where $\lambda_{\min}(\Omega)$ is the minimum eigenvalue of Ω and $\epsilon > 0$ is a small user-specified constant. This transformation leaves positive definite matrices unchanged and shifts any non-positive-definite matrix into the positive definite cone by increasing its diagonal entries. However, our empirical results under various settings show that a substantial proportion of posterior draws of Ω from our proposed samplers in Section 4 are already positive definite (see Section S.3.1 in the online resource document). If we further restrict $\mathcal{S}^*(p)$ to another class, say $\mathcal{S}^{**}(p)$, consisting of all $p \times p$ symmetric matrices with positive diagonal entries, non-positive off-diagonal entries, and strict diagonal dominance, that is,

$$\mathcal{S}^{**}(p) = \left\{ \Omega \in \mathcal{S}^*(p) : \omega_{ii} > \sum_{j \neq i} |\omega_{ij}|,\ i = 1, \ldots, p \right\}$$

then, every matrix in $\mathcal{S}^{**}(p)$ is positive definite. Therefore, restricting the parameter space from $\mathcal{S}^*(p)$ to $\mathcal{S}^{**}(p)$ provides an alternative way to obtain positive definite posterior draws without requiring any additional post-processing or projection onto the positive definite cone. Such formulations are common in the literature; see, for example, Ying et al. (2023); Egilmez et al. (2016). Due to space constraints, a detailed analysis of this restricted formulation, including the corresponding sampling algorithm and additional simulation results, is provided in the online resource document.

3.1 The (Generalized) Posterior Distribution

With the generalized likelihood and the specified prior in place, we now derive the joint generalized posterior distribution of Ω. A direct calculation yields the following lemma.

Lemma 1 (Joint posterior kernel) Let $S = n^{-1}\mathbf{Y}^T\mathbf{Y}$. Under the D-trace generalized likelihood in (2) and the spike-and-slab prior in (3), the joint generalized posterior density of Ω, given the data matrix $\mathbf{Y}$, is determined, up to a normalizing constant, by the

kernel

$$\pi_D(\Omega \mid \mathbf{Y}) \propto \exp\left\{ n \sum_{i=1}^{p} \omega_{ii} - \frac{n}{2} \operatorname{tr}(\Omega^2 S) \right\} \times \prod_{\substack{i,j=1 \\ i<j}}^{p} \left[(1-q) I_{\{0\}}(\omega_{ij}) + q \frac{2}{\tau\sqrt{2\pi}} \exp\left\{ -\frac{\omega_{ij}^2}{2\tau^2} \right\} \right] I_{\mathcal{S}^+(p)}(\Omega). \tag{4}$$

It is important to note that the generalized posterior density in (4) is not induced by an ordinary likelihood arising from a fully specified probability model. Rather, the generalized likelihood is obtained by exponentiating a loss function. In addition, the priors imposed on the model parameters are not all proper. Consequently, the propriety of the resulting generalized joint posterior is not automatic. Equivalently, the finiteness of the normalizing constant associated with the kernel in Lemma 1 must be verified. Examples of improper priors leading to improper posteriors are discussed in Dixit and Roy (2021). The implications of posterior impropriety for Gibbs samplers have been studied in Gelman et al. (2013), Hobert and Casella (1996), and Hobert and Casella (1998). In particular, it has been shown that an MCMC output may fail to exhibit any obvious numerical instability or unreasonable behavior even when the corresponding posterior is improper Hobert and Casella (1996); moreover, standard MCMC-based estimates may converge to zero with probability one under posterior impropriety. These considerations make it essential to rigorously establish the propriety of the generalized posterior density $\pi_D(\cdot \mid \mathbf{Y})$.

The following lemma addresses this issue. It shows that, although the construction does not start from an exact likelihood and the prior specification includes improper components, the posterior distribution stated in Lemma 1 is a proper probability distribution. This result provides the foundation for the coherence of the subsequent posterior analysis.

Lemma 2 (Posterior propriety) Suppose that $s_{ii} > 0,\ i = 1, \ldots, p$. Then, for any sample size n and ambient dimension p, the joint generalized posterior density of Ω, given $\mathbf{Y}$, is proper. Equivalently, the normalizing constant associated with the posterior kernel in Lemma 1 is finite.

The condition $s_{ii} > 0,\ i = 1, \ldots, p$, in Lemma 2 is mild. Since $S = n^{-1}\mathbf{Y}^T\mathbf{Y}$, $s_{ii} = 0$ if and only if the ith coordinate is zero for all observations. Thus, it only excludes coordinates with zero empirical second moment, an event of probability zero for continuous variables with no atom at zero. The proof is given in Section S.1.1 of the online resource document. Consequently, the posterior is proper and posterior inference is well-defined in the usual probabilistic sense.

4 Posterior Computation and Sampling

Note that the form of the joint (generalized) posterior density $\pi_D(\Omega|\mathbf{Y})$ in (4) is not tractable for closed-form analytical evaluation or direct sampling. Therefore, in this section, we introduce three MCMC methods to sample from the posterior.

4.1 Componentwise (Direct) Gibbs Sampler

We consider a componentwise Gibbs sampler to sample from $\pi_D(\Omega|\mathbf{Y})$. To this end, we compute the conditional posterior densities of the entries of Ω, given the remaining entries. Straightforward calculations yield the following conditional posteriors.

Lemma 3 The conditional posterior distributions of the off-diagonal entries of Ω, given the remaining entries, are obtained as follows. For $1 \le i < j \le p$,

$$\omega_{ij} \mid \mathbf{Y}, \Omega_{-(ij)} \sim (1 - q_{ij})\delta_0 + q_{ij}\, \mathcal{N}(A_1, A_2; -\infty, 0) \tag{5}$$

where

$$q_{ij} = \frac{c_{ij}}{1 + c_{ij}},\ A_1 = -\frac{b_{ij}}{a_{ij}},\ A_2 = \frac{1}{n a_{ij}},$$

$$a_{ij} = s_{ii} + s_{jj} + \frac{1}{n\tau^2},\ b_{ij} = \sum_{j' \neq j} \omega_{ij'} s_{jj'} + \sum_{i' \neq i} \omega_{i'j} s_{ii'},$$

and

$$c_{ij} = \frac{2q}{(1-q)\sqrt{n a_{ij} \tau^2}} \exp\left(\frac{n b_{ij}^2}{2 a_{ij}} \right) \Phi\left(b_{ij} \sqrt{\frac{n}{a_{ij}}} \right).$$

Here, s_{ij} denotes the (i,j)-th entry of $S = \frac{1}{n}\mathbf{Y}^\top\mathbf{Y}$ and for any matrix A, $A_{-(ij)}$ denotes the vector of all the elements in the upper triangle except the (i,j)-th entry.

The conditional posterior distributions of the diagonal entries of Ω, given the remaining entries, are given by

$$\omega_{ii} \mid \mathbf{Y}, \Omega_{-(ii)} \sim \mathcal{N}\left(\frac{1-d_i}{s_{ii}}, \frac{1}{ns_{ii}}; 0, \infty\right), \qquad (6)$$

where $d_i = \sum_{i' \neq i} \omega_{ii'} s_{ii'}$, for $i = 1, 2, \ldots, p$.

Algorithm 1 Direct Gibbs Sampler

procedure D-GIBBS($\Omega^{init}, \mathbf{Y}, T, q, \tau^2$)
$\Omega^{(0)} = \Omega^{init}$
  $S = n^{-1}\mathbf{Y}^T\mathbf{Y}$
  for $iter = 1, 2, \ldots, T$ **do**
    $\Omega^* = \Omega^{(iter-1)}$
    for $i = 1, 2, \ldots, p-1$ **do** ▷ **Update off-diagonals**
      for $j = i+1, \ldots, p$ **do**
        $a \leftarrow s_{ii} + s_{jj} + 1/n\tau^2$
        $b \leftarrow (\Omega^*_{\cdot i})^T S_{\cdot j} + (\Omega^*_{\cdot j})^T S_{\cdot i} - (s_{ii} + s_{jj})\omega^*_{ij}$
        $C^* \leftarrow \exp\left[\frac{nb^2}{2a}\right]\Phi\left(b\sqrt{\frac{n}{a}}\right)$
        $P(0) \leftarrow 1, P(1) \leftarrow \sqrt{\frac{1}{na\tau^2}}\frac{2q}{1-q}C^*$
        if $P(1) = \infty$ **then**
          $\omega^*_{ij} \leftarrow \mathcal{N}\left(-\frac{b}{a}, \frac{1}{na}; -\infty, 0\right)$
          $\omega^*_{ji} \leftarrow \omega^*_{ij}$
        else
          $P \leftarrow P/\mathrm{sum}(P)$
          $\omega^*_{ij} \sim P(0)\delta_0 + P(1)\mathcal{N}\left(-\frac{b}{a}, \frac{1}{na}; -\infty, 0\right)$
          $\omega^*_{ji} \leftarrow \omega^*_{ij}$
      $d \leftarrow 1 - (\Omega^*_{\cdot i})^T S_{\cdot i} + \omega^*_{ii} s_{ii}$ ▷ **Update diagonal**
      $\omega^*_{ii} \leftarrow \mathcal{N}(\frac{d}{s_{ii}}, \frac{1}{ns_{ii}}; 0, \infty)$
    Repeat Steps 18 - 19 for $i = p$
    $\Omega^{(iter)} \leftarrow \Omega^*$
  return $\Omega^{(1)}, \Omega^{(2)}, \ldots, \Omega^{(T)}$

The proof of Lemma 3 is provided in Section S.1.2 of the online resource document. Posterior samples can be generated using a Gibbs sampler based on these conditional distributions. A schematic of this sampler, which we refer to as the *Direct Gibbs Sampler*, or in short, *D-Gibbs* is outlined in Algorithm 1 below. Our next lemma focuses on the computational complexity of this Direct Gibbs Sampler.

Lemma 4 The computational complexity of each iteration of the direct Gibbs sampler in Algorithm 1 is $\mathcal{O}(p^3)$.

The proof of Lemma 4 is provided in Section S.1.3 of the online resource document. Note that the $\mathcal{O}(p^3)$ computational complexity in Lemma 4 is comparable to that of inverting a $p \times p$ matrix. This results in substantial computational overhead, posing a significant limitation for implementing this sampler in high dimensions, particularly when $p \gg n$. Moreover, the dependence induced by the direct Gibbs updates may lead to inefficient MCMC estimates. In the following section, we address these issues—especially the computational bottleneck—by developing an alternative Gibbs sampler based on an augmented framework.

4.2 Augmented Gibbs Sampler

Note that the non-zero correlation between the entries of Ω restricts joint sampling of all elements, thus enforcing sequential one-by-one sampling, which contributes to the computational overhead. To address this issue, we develop an Augmented Gibbs Sampler framework by generalizing the *Anti-correlation Gaussian* approach proposed in Zheng and Duan (2025) to matrix-variate settings. The goal is to introduce latent variables such that the posterior distributions of the entries of Ω are conditionally independent given these variables, thereby enabling efficient joint sampling. The resulting augmented posterior must, by construction, remain consistent with the original posterior distribution. A computationally efficient generalization of the vector-based framework in Zheng and Duan (2025) to the more involved matrix-variate setting necessitates several nontrivial and conceptually novel ideas, which we present in detail below.

To motivate the augmented formulation, we introduce a matrix-variate latent variable $R \in \mathbb{R}^{p\times p}$ such that

$$R \mid \mathbf{Y}, \Omega \sim \mathcal{MN}_{p\times p}\big(\Omega(kI_p - nS),\, I_p,\, kI_p - nS\big), \qquad (7)$$

where k is chosen sufficiently large so that $kI_p - nS$ is positive definite. Equivalently, conditional on $\mathbf{Y}$

and Ω, the rows of R are independent and

$$R_{i\cdot} \mid \mathbf{Y}, \Omega \sim \mathcal{N}_p\big(\{\Omega(kI_p - nS)\}_{i\cdot},\, kI_p - nS\big)$$

for $i = 1, \ldots, p$..The proposed augmentation in (7) is valid since R is introduced through a *proper* matrix-variate normal distribution given Ω and $\mathbf{Y}$. Indeed, by Lemma 2, the original generalized posterior for Ω is *proper*. Moreover, for each fixed Ω and $\mathbf{Y}$, the conditional distribution of R has total mass one. Consequently, the augmented joint posterior of (R, Ω) given $\mathbf{Y}$ is also proper. In particular, integrating the augmented posterior with respect to R and then Ω yields one, while integrating with respect to R alone recovers the original generalized posterior for Ω. Next we'll show this augmentation yields a conditionally independent posterior structure for the elements of Ω, stated in the following lemma.

Lemma 5 (Augmented conditional posterior distribution of Ω) Under the augmented formulation in (7), the conditional posterior densities of the entries of Ω given R and $\mathbf{Y}$ are as follows. For $1 \le i < j \le p$,

$$\begin{aligned}\omega_{ij} \mid R, \mathbf{Y} \sim (1 - q^*_{ij})\delta_0 + \\ q^*_{ij}\mathcal{N}\left(\frac{R_{ij} + R_{ji}}{k^*}, \frac{1}{k^*}; -\infty, 0\right), \\ \omega_{ii} \mid R, \mathbf{Y} \sim \mathcal{N}\left(\frac{R_{ii} + n}{k}, \frac{1}{k}; 0, \infty\right) \qquad (8)\end{aligned}$$

where

$$k^* = 2k + \frac{1}{\tau^2},\; q^*_{ij} = \frac{c_{ij}}{1 + c_{ij}},\; c_{ij} = \frac{2q}{1-q}\frac{1}{\tau\sqrt{k^*}}C^*,$$

$$C^* = \exp\left\{\frac{(R_{ij} + R_{ji})^2}{2k^*}\right\}\Phi\left(-\frac{R_{ij} + R_{ji}}{\sqrt{k^*}}\right),$$

Moreover, conditional on R and $\mathbf{Y}$, all entries in the upper triangle of Ω are mutually independent under the augmented posterior.

The proof of this Lemma is provided in Section S.1.4 of the online resource document. It follows immediately from Lemma 5 that, conditional on the augmented variable R, sampling from the posterior distribution of Ω can be carried out in $\mathcal{O}(p^2)$ operations. However, at each iteration of the Gibbs sampler, one must also sample R conditional on the current value of Ω from the matrix-normal distribution in (7). Direct sampling from this distribution requires computing a matrix square root, or equivalently a Cholesky factor, of the $p \times p$ matrix $kI_p - nS$, which entails $\mathcal{O}(p^3)$ computational cost. Thus, despite the conditional independence in the augmented posterior, the direct implementation has the same cubic-order bottleneck as the naive elementwise Gibbs sampler described in Section 4.1. As mentioned before this becomes particularly restrictive in high-dimensional settings where $n \ll p$.

However, we leverage specific structure of the matrix $kI_p - nS$ to develop a significantly more efficient alternative. In particular, if $\mathbf{Y}$ denotes the $n \times p$ data matrix, then $S = n^{-1}\mathbf{Y}^T\mathbf{Y}$, and hence $kI_p - nS = kI_p - \mathbf{Y}^T\mathbf{Y}$. Exploiting a low-rank structure to reduce the computational burden is common in the literature, especially in high-dimensional settings where $p \gg n$. For example, in the regression context, Bhattacharya et al. (2016) consider sampling from a Gaussian distribution with covariance matrix $\left(X^T\Omega X + D\right)^{-1}$, where D is a $p \times p$ positive definite matrix and X is an $n \times p$ design matrix. They exploit the Woodbury matrix identity, which reduces the computation to the inversion of an $n \times n$ matrix instead of a $p \times p$ matrix when $p \gg n$. In a related setting, Nishimura and Suchard (2023) use a preconditioned conjugate gradient method to iteratively solve the corresponding linear system.

Our present setting differs from the aforementioned regression examples in two important respects. First, the covariance structure involves the matrix difference $kI_p - nS$, rather than a sum of a positive definite matrix and a low-rank term. Second, the augmented object to be sampled is matrix-valued rather than vector-valued. A direct, but computationally inefficient, approach would be to apply the sampling scheme of Zheng and Duan (2025, Theorem 2.1) separately to each row of R, thereby repeating the same procedure p times. As discussed in Zheng and Duan (2025), generating a sample for each row of R has computational complexity $\mathcal{O}(p^2 + n^2)$. Therefore, repeating this procedure p times leads to an overall complexity of $\mathcal{O}(p^3 + n^2 p)$, which remains of order $\mathcal{O}(p^3)$ when $n \ll p$. This inefficiency arises because the procedure treats the rows separately and does not exploit the matrix-normal structure of R. To overcome this obstacle, we propose a

novel and efficient sampling scheme for matrix-normal distributions of the form (7). The following lemma formalizes this scheme.

Lemma 6 (Fast matrix-normal sampling under Gram structure) Let $U \in \mathbb{R}^{p\times n}$ satisfy $nS = kUU^T$, where k is chosen such that $I_n - U^TU$, equivalently $kI_p - nS$, is positive definite. Let $A \in \mathbb{R}^{n\times n}$ denote the Cholesky factor of $I_n - U^TU$, so that $AA^T = I_n - U^TU$. Let $V_1 \in \mathbb{R}^{p\times p}$ and $V_2 \in \mathbb{R}^{p\times n}$ be independent random matrices with i.i.d. standard normal entries. Define

$$X = \sqrt{k}\left\{V_1(I_p - UU^T) + V_2A^TU^T\right\}.$$

Then $X \sim \mathcal{MN}_{p\times p}\big(0, I_p, kI_p - nS\big)$. Moreover, after precomputing A, one draw of X can be generated in $\mathcal{O}(p^2n + pn^2)$ operations, which reduces to $\mathcal{O}(p^2n)$ when $n \ll p$.

The proof of Lemma 6 is provided in Section S.1.5 of the online resource document. In particular, Lemma 6 can be used to sample from (7) by taking $U = k^{-1/2}\mathbf{Y}^T$, so that $kUU^T = \mathbf{Y}^T\mathbf{Y} = nS$. The key advantage of Lemma 6 is that the Cholesky factor A can be computed once before running the MCMC chain, since S is fixed after observing the data. This preprocessing step requires $\mathcal{O}(pn^2+n^3)$ operations. Conditional on this preprocessing, one draw from the distribution can be generated at each MCMC iteration in $\mathcal{O}(p^2n+pn^2)$ operations using only matrix multiplications. In particular, when $n \ll p$, the per-draw complexity reduces to $\mathcal{O}(p^2n)$. For $n > p$, we can sample from the matrix normal directly with $O(p^3)$ complexity.

In the general Bayesian covariance matrix estimation literature using shrinkage priors (even without sign constraints), under the exact Gaussian likelihood, the time complexity of the sampling algorithms is typically $\mathcal{O}(p^4)$ when $n > p$. Recently, Gao et al. (2026) reduce this cost to $\mathcal{O}(p^3)$ in the high-dimensional regime $n = o(\sqrt{p})$. In contrast, our per-draw complexity is $\mathcal{O}(p^2n)$, which becomes $\mathcal{O}(p^{2.5})$ when $n = o(\sqrt{p})$, and is therefore lower in this regime. Thus, even without sign constraints, our approach of employing the D-Trace generalized likelihood and relaxing the positive-definiteness constraint can substantially improve computational scalability compared to existing methods. The quadratic nature of the D-trace loss, relaxation of PD constraint, and the conditional independence entries of Ω given R also enable seamless integration of any sign constraint on Ω in our framework, which is not the case with the Gaussian likelihood-based method. Hence, the computation scalability issue for this method becomes even more acute under the sign constraints. Moreover, the empirical results in Section 5 demonstrate that this computational advantage is not obtained at the cost of any systematic loss of estimation accuracy.

A remaining practical issue is the choice of k in the proposed algorithm. Since the augmentation requires $kI_p - nS$ to be positive definite, it is enough to choose $k > n\lambda_{\max}(S)$. We set $k = \{n\epsilon\lambda_{\max}(S)\}$, where $\epsilon = 1.001$. Since S is fixed after observing the data, this quantity can be computed once before starting the MCMC chain. A direct computation requires $\mathcal{O}(p^3)$ operations, which is a one-time preprocessing cost and does not affect the per-iteration complexity of the sampler. Combining Lemmas 5 and 6, we now present the entrywise Gibbs sampler in Algorithm 2.

Note that, in Algorithm 2, the entries of Ω are conditionally independent given the augmented variable R. Hence each entry of Ω (given R) can now be independently sampled with $O(1)$ computations requiring a total of $O(p^2)$ computations for sampling Ω (this can be further improved through parallel sampling, if resources are available). The discussion on the computational complexity of the augmented A-Gibbs sampler in the high-dimensional regime is formalized in the following lemma.

Lemma 7 The computational complexity of a single iteration of the proposed augmented Gibbs sampler in Algorithm 2 is $\mathcal{O}(\min\{p^2n, p^3\})$. When $p \gg n$, this complexity reduces to $O(np^2)$, which is an order of magnitude smaller than the complexity of the direct Gibbs sampler.

4.3 The Interweaving Algorithm

As described in Section 4.1, the componentwise direct Gibbs sampler suffers from substantial computational overhead, with per-iteration complexity of $\mathcal{O}(p^3)$, and induces strong dependence among updates. Together, these factors can lead to slow mixing and inefficient MCMC estimation, particularly in high-dimensional settings. To

Algorithm 2 Augmented Gibbs Sampler

procedure A-GIBBS($\Omega^{init}, \mathbf{Y}, T, \varepsilon, q, \tau^2$)
$\Omega^{(0)} = \Omega^{init}$
  $S = n^{-1}\mathbf{Y}^T\mathbf{Y}$, $\lambda =$ Maximum eigenvalue of nS
  $k \leftarrow \varepsilon\lambda$, $k^* \leftarrow 2k + \tau^{-2}$
  for iter $= 1, 2, \ldots, T$ **do** $M \leftarrow \Omega^{(iter-1)}(kI_p - nS)$
    if $p > n$ **then** ▷ **Sample** R
      $R \leftarrow \mathcal{MN}_{p\times p}\big(M,\, I_p,\, kI_p - nS\big)$ using Lemma 6
    else
      $R \leftarrow \mathcal{MN}_{p\times p}\big(M,\, I_p,\, kI_p - nS\big)$ directly
    $\Omega^* = \Omega^{(iter-1)}$
    for $i = 1, 2, \ldots, p-1$ **do** ▷ **Update off-diagonals**
      for $j = i+1, \ldots, p$ **do**
        $R^* = \frac{R_{ij}+R_{ji}}{k^*}$
        $C^* = \exp\left\{\frac{k^* R^{*2}}{2}\right\} \Phi\left(-\sqrt{k^*}R^*\right)$
        $P(0) \leftarrow 1$
        $P(1) \leftarrow \frac{2q}{1-q}\frac{1}{\tau\sqrt{k^*}}C^*$
        if $P(1) = \infty$ **then**
          $\omega^*_{ij} \leftarrow \mathcal{N}\left(R^*, \frac{1}{k^*}; -\infty, 0\right)$
          $\omega^*_{ji} \leftarrow \omega^*_{ij}$
        else
          $P \leftarrow P/\text{sum}(P)$
          $\omega^*_{ij} \sim P(0)\delta_0 + P(1)\mathcal{N}\left(R^*, \frac{1}{k^*}; -\infty, 0\right)$
          $\omega^*_{ji} \leftarrow \omega^*_{ij}$
      $\omega^*_{ii} \leftarrow \mathcal{N}(\frac{R_{ii}+n}{k}, \frac{1}{k}; 0, \infty)$ ▷ **Update diagonal**
    Repeat Step 21 for $i = p$
    $\Omega^{(iter)} \leftarrow \Omega^*$
  return $\Omega^{(1)}, \Omega^{(2)}, \ldots, \Omega^{(T)}$

address these issues, we introduced a data augmentation framework in Section 4.2 and proposed an augmented Gibbs sampler.

Although the augmented sampler achieves significant computational gains, exhibits substantially improved mixing, and consistently delivers superior sparsity recovery in our empirical studies (see Sections 5 and 5.2), it operates in an expanded parameter space through the introduction of latent variables. In contrast, the direct Gibbs sampler updates the entries of Ω under the original posterior parameterization, thereby preserving the intrinsic dependence structure of the model.

These two samplers therefore correspond to different representations of the same posterior distribution, each inducing a distinct Markov chain geometry. The augmented sampler leverages conditional independence to enable efficient global exploration, which translates into improved sparsity selection, as observed in both our simulation studies and real data analysis. On the other hand, the direct sampler performs local updates in the original parameter space and can provide complementary refinements of the chain, although by itself it exhibits slower mixing and comparatively poorer sparsity recovery in practice.

Motivated by these observations, we propose an interweaving algorithm that combines the strengths of both sampling schemes. The proposed sampler is designed to improve exploration of the parameter space both across different sparsity patterns and within a fixed sparsity pattern. The high-level idea behind this interweaving strategy is inspired by the ancillary-sufficiency interweaving strategy of Yu and Meng (2011). In particular, we alternate between the augmented and direct representations of the posterior, thereby combining the rapid global exploration afforded by the augmented sampler with the corrective local updates of the direct sampler. The algorithm proceeds by iteratively using the augmented sampler and invoking the direct sampler at every N-th iteration, for some pre-specified positive integer N, to encourage exploration across distinct sparsity patterns. Our empirical choice of N is specified in Section 5. The resulting interweaving strategy is described in Algorithm 3. Empirically, we find that

Algorithm 3 Interweaving Gibbs Sampler

procedure I-GIBBS($\Omega^{init}, \mathbf{Y}, T, \varepsilon, q, \tau^2, N$)
$\Omega^{(0)} = \Omega^{init}$
  $S = n^{-1}\mathbf{Y}^T\mathbf{Y}$, $\lambda =$ Maximum eigenvalue of nS
  for *iter* $= 1, 2, \ldots, T$ **do**
    $\Omega^{init} = \Omega^{(iter-1)}$
    if iter $\equiv 0 \pmod N$ **then**
      Perform Steps 5-21 of D-GIBBS (Algorithm 1)
    else
      Perform Steps 6-25 of A-GIBBS (Algorithm 2)
    $\Omega^{(iter)} \leftarrow$ sampled matrix in Steps 5-9
  return $\Omega^{(1)}, \Omega^{(2)}, \ldots, \Omega^{(T)}$

while the augmented sampler already performs strongly, incorporating occasional direct updates further enhances stability and overall mixing without compromising computational efficiency.

4.4 Choice of Hyperparameters

The selection of hyperparameters is an important aspect of Bayesian modeling. Although cross-validation is widely used for hyperparameter tuning, it can be computationally prohibitive in high-dimensional settings with large p. We therefore adopt standard choices from the literature for q and τ^2. Specifically, motivated by Narisetty and He (2014); Samanta et al. (2022), q is selected through a grid search over $\{1/p, 1/p^{1.5}, 1/p^2\}$.

For the hyperparameter τ^2, one may either fix it at a value close to 1 or adopt a fully Bayesian specification by assigning an objective Inverse-Gamma prior with shape parameter 10^{-4} and rate parameter 10^{-8}; see Wang (2012); Samanta et al. (2022). If a fully Bayesian approach is employed, some modifications are necessary to Algorithms 1, 2 and 3. First, posterior propriety is not guaranteed in this setting if $n < p$ and improper priors are used for the diagonal entries of Ω. Hence, when $n < p$, we employ independent diffuse Gaussian priors with mean 0 and precision parameter 0.01 for each diagonal entry of Ω. As a result, Step 18 of Algorithm 1 will change to

$$\omega_{ii} \leftarrow \mathcal{N}\left(\frac{nd}{ns_{ii}+0.01}, \frac{1}{ns_{ii}+0.01}; 0, \infty\right),$$

and Step 22 of Algorithm 2 will change to

$$\omega_{ii} \leftarrow \mathcal{N}\left(\frac{R_{ii}+n}{k+0.01}, \frac{1}{k+0.01}; 0, \infty\right).$$

Finally, the following Inverse-Gamma update for τ^2 is needed in each iteration of the various Gibbs samplers.

$$\tau^2 \sim \text{IG}(10^{-4}+\frac{1}{2}\sum_{i<j} I_{\mathbb{R}\setminus\{0\}}(\omega_{ij}), 10^{-8}+\frac{1}{2}\sum_{i<j}\omega_{ij}^2)$$

4.5 Sparsity Selection and Estimation based on Gibbs Output

Suppose that we have the output sequence from implementing one of the three algorithms proposed above. After discarding the first B samples as burn-in, where $B < T$ is a suitable pre-defined number, we can use the remaining $T^* = T - B$ samples to estimate the graph structure and the corresponding precision matrix. Let us write the post burn-in samples as $\{\bar{\Omega}^{(t)}\}_{t=1}^{T^*}$. To estimate the sparsity pattern of the precision matrix, we adapt the majority voting approach (Barbieri and Berger 2004). For this, we compute the proportion of non-zero outputs out of the T^* final samples for each of the off-diagonal entries as given below,

$$\hat{q}_{ij} = \frac{1}{T^*}\sum_{t=1}^{T^*} I_{\mathbb{R}\setminus\{0\}}(\bar{\omega}_{ij}^{(t)})$$

where $\bar{\omega}_{ij}^{(t)}$ is the (i,j)-th entry of $\bar{\Omega}^{(t)}$. According to the majority voting approach, we include an edge in the graph structure if the corresponding $\hat{q}_{ij}$ is more than 1/2. Alternatively, we estimate the sparsity pattern $\eta_{ij} = I(\omega_{ij} \neq 0)$ of the precision matrix Ω as follows.

$$\hat{\eta}_{ij} = \begin{cases} 1, & \text{if } \hat{q}_{ij} \geq \frac{1}{2} \\ 0, & \text{otherwise.} \end{cases}$$

Once the sparsity pattern is determined, the non-zero entries of Ω are estimated by simply taking the average of the corresponding entry over the samples where that particular entry is non-zero. More specifically, we compute the entries of the final estimate $\widehat{\Omega}$

$$\widehat{\omega}_{ij} = \begin{cases} \dfrac{\sum_{t=1}^{T^*} \bar{\omega}_{ij}^{(t)} I(\bar{\omega}_{ij}^{(t)} \neq 0)}{\sum_{t=1}^{T^*} I(\bar{\omega}_{ij}^{(t)} \neq 0)}, & \text{if } \hat{\eta}_{ij} = 1 \\ 0, & \text{otherwise.} \end{cases}$$

Alternatively, one can simply consider the average of all the T^* outputs as an estimate of the non-zero entries. Entrywise credible intervals for the non-zero entries of Ω can also be constructed using the corresponding Gibbs outputs as an observed sample.

5 Numerical Experiments

We implement the proposed algorithms in various settings to analyze both synthetic and real-world data and compare the sparsity selection and estimation performances of the proposed methods

with the existing benchmark methods. Relevant approaches include GLASSO (Friedman et al. 2008), CLIME (Cai et al. 2011), GGL (Egilmez et al. 2016) and SLTP (Wang et al. 2020). Among these, the first two methods estimate graphs in general settings, whereas GGL and SLTP are designed specifically for the M-matrix constraint. GGL estimates a sparse precision matrix, but SLTP targets only the graph structure. We omit SLTP from the reported tables because it does not provide a well-defined final precision matrix estimate and, more critically, scales poorly with sample size and dimension, with some larger runs failing to terminate within a reasonable time. As specified in Algorithms 1, 2 and 3, we will refer to the three proposed samplers as D-Gibbs, A-Gibbs, and I-Gibbs, respectively, where D-Gibbs denotes the Componentwise Direct Gibbs Sampler, A-Gibbs denotes the Augmented Gibbs Sampler, and I-Gibbs denotes the Interweaving Direct-Augmented Gibbs Sampler. For the proposed methods, hyperparameters are chosen based on recommendations in Section 4.4. For GLASSO, GGL and CLIME, the respective hyperparameters are chosen based on cross-validation over author-recommended ranges. For I-Gibbs, we take $N = 100$, i.e., at every 100-th iteration, we invoke one iteration of the D-Gibbs algorithm.

5.1 Synthetic Data Analysis

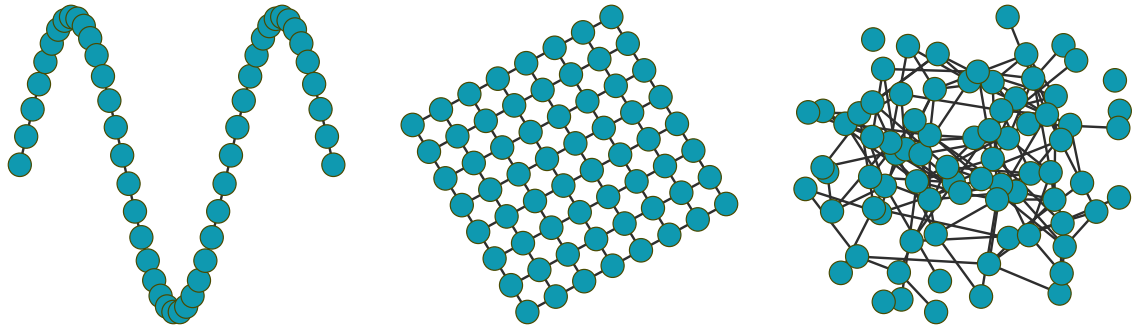
Experimental Setting. We evaluate the performance of the proposed algorithms with respect to the aforementioned state-of-the-art algorithms using synthetically generated datasets under various settings. We consider the *Line, Grid and Erdős-Rényi (with 5% edge density)* structures as data generating models. A *line graph* is a network where all vertices are arranged in a single continuous sequence. Each node has exactly two neighbors, except for the two endpoints which connect to only one neighbor each. A *grid graph* is a structured lattice, where vertices are organized into rows and columns, and each node, except those on the boundaries, connects to its four immediate neighbors (up, down, left, right). Finally, an *Erdős-Rényi graph* is constructed with a fixed number of nodes, and each possible pair of nodes has an independent, identical probability (5% in our experiments) of being connected by an edge. Figure 1 provides visual representations of these graph structures. For each structure, we consider six different simulation settings corresponding to the following combinations of (n, p): $(100, 20), (100, 50), (100, 100), (100, 200), (500, 100)$ and $(500, 250)$.

For each fixed graph structure and (n, p) combination, we generate 50 independent datasets with $6n/5$ observations following the data-generation framework in Slawski and Hein (2015). First we construct the graph according to the given structure, and then assign weights to the edges by sampling randomly from $U(2, 5)$ distribution to obtain the weighted adjacency matrix A. Defining $\delta = 1.05\lambda_{max}(A)$, where $\lambda_{max}(A)$ is the maximum eigenvalue of A, we set

$$\Omega^* = \delta I - A,\ \Omega_0 = E\Omega^* E,$$

where E is a diagonal matrix ensuring that the diagonal elements of $(\Omega_0)^{-1}$ are 1. We consider Ω_0 to be our true precision matrix, and draw n independent samples $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ from $N(\mathbf{0}, \Omega_0^{-1})$. The additional $n/5$ observations are used as test data to select slab inclusion probability q via the grid search described in Section 4.4.. We estimate the precision matrix using different algorithms as mentioned above, and evaluate the performance of the algorithms. Note again that SLTP only provides an estimate of the graph structure.

Evaluation Metrics. We first consider metrics for sparsity selection accuracy. Let TP, FP, TN and FN denote the true positives, false positives, true negatives and false negatives respectively. *Matthews Correlation Coefficient (MCC)* (Matthews 1975) and *F-score* (van Rijsbergen 1979) of an estimated graph measure how well the true graph structure has been recovered. The F-score of a graph ranges between 0 and 1, with $F = 1$ indicating a perfect structure recovery scenario, whereas MCC ranges between -1 and 1, with $-1, 0$ and 1 denoting a perfectly inverse prediction, a random prediction and a perfect prediction respectively. We further report the *Area Under the Receiver Operating Characteristic curve (AUROC)* as a measure of graph selection performance of the algorithms. To assess the quality of estimation, we compute the *Relative Estimation Error*, obtained as the Frobenius norm



(a) Line (b) Grid (c) Erdős–Rényi

Fig 1 Illustrations of graph structures used in the synthetic data study for comparison

of the difference of the estimated and true precision matrix, and scaled by the Frobenius norm of the true precision matrix.

Results. The MCC, F-score and the estimation errors corresponding to all the settings considered in our experiment are reported below in Table 1, 2 and 3 respectively. Results on AUROC have been included in Section S.3.3 of the online resource document.

Tables 1 and 2 report the MCC and F-scores across all 18 experimental settings. Across nearly all simulation settings, the proposed Gibbs samplers consistently achieve substantially higher MCC and F-scores than the competing methods. The sole exception, in which GGL outperforms the proposed algorithms, occurs under the Erdős–Rényi model with $n = 100$ and $p = 200$ (and 245 non-zero off-diagonal entries in the corresponding Ω_0), a severely data-deprived regime in which all six algorithms attain notably low MCC and F-score values. In all remaining settings, the proposed algorithms demonstrate substantial improvements over the general GLASSO and CLIME methods, which do not incorporate the sign constraint; the magnitude of improvement ranges from approximately 100% in most settings to as much as 200% in several others. While GGL, which does account for the sign constraints, improves upon GLASSO and CLIME, the proposed samplers exhibit considerably superior performance relative to GGL across nearly all settings.

Among the three proposed algorithms, all of which deliver significantly improved performance as described above, the interweaving I-Gibbs sampler achieves the highest overall performance. Specifically, I-Gibbs consistently attains the highest MCC and F-score values across all settings, reflecting a marked improvement in graph selection performance over existing state-of-the-art algorithms. The augmented A-Gibbs sampler yields results closely comparable to I-Gibbs, with differences typically appearing only at the third decimal place. In contrast, the graph selection performance of D-Gibbs is somewhat inferior to that of the other two Gibbs samplers, a finding that is consistent with its characterization as a local explorer, as discussed in Section 4.3.

Table 3 presents the precision matrix estimation errors across the 18 settings. Algorithms that incorporate the M-matrix constraint uniformly outperform GLASSO and CLIME, which do not account for this constraint. Compared to GGL, the proposed methods achieve lower estimation error in certain settings and remain broadly competitive across the remaining configurations. In summary, the three proposed Gibbs samplers yield substantial improvements in graph structure estimation under the M-matrix constraint. With respect to graph selection, I-Gibbs and A-Gibbs demonstrate the strongest performance, whereas D-Gibbs proves most effective for precision matrix estimation.

Table 1: *Matthews Correlation Coefficient (MCC)* for the benchmark and proposed algorithms across three data-generating models, namely, Line Graph, Grid Graph, and Erdős–Rényi Graph (5% edge density), and six (n, p) configurations.

Model	(n,p)	GLASSO	CLIME	GGL	D-Gibbs	A-Gibbs	I-Gibbs
Line Graph	$(100, 20)$	0.339362	0.162940	0.592964	0.839393	0.900792	**0.910718**
	$(100, 50)$	0.344382	0.366485	0.492189	0.803132	0.866734	**0.883758**
	$(100, 100)$	0.338153	0.469339	0.427420	0.779140	0.835501	**0.856129**
	$(100, 200)$	0.340614	0.545414	0.362629	0.750760	0.763749	**0.806506**
	$(500, 100)$	0.324436	0.442767	0.391187	0.889521	0.929744	**0.948483**
	$(500, 250)$	0.328980	0.472171	0.310010	0.918754	0.950831	**0.962887**
Grid Graph	$(100, 20)$	0.418464	NaN	0.589024	0.762005	0.782519	**0.792389**
	$(100, 50)$	0.422413	0.330798	0.544509	0.702119	0.706322	**0.718569**
	$(100, 100)$	0.408616	0.418420	0.489635	0.635538	0.645766	**0.648025**
	$(100, 200)$	0.384598	0.499235	0.462793	0.624958	0.576971	**0.638008**
	$(500, 100)$	0.408196	0.431054	0.523006	0.950699	0.940638	**0.963538**
	$(500, 250)$	0.396653	0.587283	0.469590	0.937282	0.915480	**0.949102**
Erdős-Rényi	$(100, 20)$	0.332911	0.242688	0.358103	0.767684	0.866433	**0.878072**
	$(100, 50)$	0.405567	0.326772	0.433457	0.600928	0.606377	**0.614399**
	$(100, 100)$	0.353774	0.300434	0.366720	0.358896	0.404450	**0.407736**
	$(100, 200)$	0.174604	0.133601	**0.194814**	0.103091	0.067407	0.100008
	$(500, 100)$	0.443320	0.432508	0.484373	0.805784	0.833517	**0.852969**
	$(500, 250)$	0.344685	0.003510	0.351726	0.365776	0.352156	**0.369464**

Table 2: *F-scores* for the benchmark and proposed algorithms across three data-generating models, namely, Line Graph, Grid Graph, and Erdős–Rényi Graph (5% edge density), and six (n, p) configurations.

Model	(n,p)	GLASSO	CLIME	GGL	D-Gibbs	A-Gibbs	I-Gibbs
Line Graph	$(100, 20)$	0.346516	0.223613	0.588181	0.849258	0.909195	**0.918645**
	$(100, 50)$	0.269845	0.290592	0.428643	0.803568	0.870196	**0.887556**
	$(100, 100)$	0.235832	0.384320	0.332391	0.777240	0.837927	**0.858570**
	$(100, 200)$	0.227745	0.482143	0.247804	0.752525	0.757187	**0.807213**
	$(500, 100)$	0.222446	0.350902	0.290577	0.885898	0.928729	**0.948192**
	$(500, 250)$	0.208240	0.372885	0.187323	0.916166	0.950052	**0.962504**
Grid Graph	$(100, 20)$	0.480015	0.299448	0.626990	0.799067	0.814697	**0.823955**
	$(100, 50)$	0.391348	0.302863	0.519985	0.714203	0.714059	**0.724332**
	$(100, 100)$	0.346849	0.370785	0.435944	0.628399	0.650666	**0.632516**
	$(100, 200)$	0.300669	0.470017	0.388500	0.611755	0.564937	**0.614718**
	$(500, 100)$	0.332863	0.355340	0.461504	0.952385	0.942336	**0.964802**
	$(500, 250)$	0.291539	0.525477	0.376971	0.938167	0.916677	**0.949584**
Erdős-Rényi	$(100, 20)$	0.280379	0.190686	0.290221	0.756816	0.864444	**0.877599**
	$(100, 50)$	0.376305	0.291812	0.394005	0.601390	0.601570	**0.605846**
	$(100, 100)$	0.351698	0.309478	0.358247	0.380263	**0.420035**	0.408998
	$(100, 200)$	0.207444	0.172611	**0.232161**	0.054220	0.035888	0.050440
	$(500, 100)$	0.389738	0.379392	0.433820	0.815546	0.841990	**0.859585**
	$(500, 250)$	0.346394	0.026335	0.351061	0.343584	0.342464	**0.359817**

5.2 Financial Time-Series Data

In this section, we analyze financial time-series data and compare the performance of the proposed algorithms with existing benchmarks. The $M-$matrix constraint is natural in this context, since latent market factors induce positive dependence among stock prices within related market sectors (Hennessy and Lapan 2002; Müller and Scarsini 2005). This provides a statistical justification for graphical model selection under M-matrix constraints; see, for example, Wang et al. (2020) and Ying et al. (2023).

We consider five industrial sectors classified according to the Global Industry Classification

Table 3: *Relative Estimation Errors* (in Frobenius norm) for the benchmark and proposed algorithms across three data-generating models, namely, Line Graph, Grid Graph, and Erdős–Rényi Graph (5% edge density), and six (n, p) configurations.

Model	(n,p)	GLASSO	CLIME	GGL	D-Gibbs	A-Gibbs	I-Gibbs
Line Graph	$(100, 20)$	0.402249	0.580012	0.193142	**0.170640**	0.172726	0.171740
	$(100, 50)$	0.537159	0.435991	0.229762	**0.191810**	0.195054	0.192894
	$(100, 100)$	0.577636	0.416646	0.265771	0.221955	0.223213	**0.219658**
	$(100, 200)$	0.564507	0.395472	0.317985	0.260906	0.261461	**0.257024**
	$(500, 100)$	0.389555	0.130876	0.104114	0.076846	0.078304	**0.076456**
	$(500, 250)$	0.388881	0.160640	0.122187	0.082941	0.084125	**0.082209**
Grid Graph	$(100, 20)$	0.368722	0.538054	**0.246963**	0.260354	0.265003	0.260866
	$(100, 50)$	0.413582	0.381189	**0.279902**	0.309609	0.315558	0.312877
	$(100, 100)$	0.438997	0.372740	**0.303163**	0.346133	0.354197	0.348000
	$(100, 200)$	0.464642	0.411157	**0.306747**	0.364031	0.386358	0.364370
	$(500, 100)$	0.259640	0.152300	0.126216	**0.103256**	0.108265	0.103862
	$(500, 250)$	0.281666	0.200332	0.130323	**0.112434**	0.122991	0.112953
Erdős-Rényi	$(100, 20)$	0.617947	0.647322	0.236652	0.178985	0.178695	0.178519
	$(100, 50)$	0.478910	0.399931	0.363733	0.334067	0.339354	0.335502
	$(100, 100)$	0.412339	0.372308	**0.346040**	0.471846	0.486807	0.474448
	$(100, 200)$	0.381610	0.358277	**0.337663**	0.422962	0.385606	0.417079
	$(500, 100)$	0.242779	0.179009	0.177906	0.161810	0.166856	0.163581
	$(500, 250)$	0.255073	0.562434	**0.235388**	0.301159	0.308389	0.303388

Standard (GICS): Information Technology, Financials, Health Care, Consumer Discretionary, and Consumer Staples. The list of S&P 500 constituent companies, including ticker symbols and GICS sector classifications, was obtained from Wikipedia's 'List of S&P 500 companies' page. The data consists of $p = 286$ stocks from these sectors included in the S&P 500 index. We collect daily closing prices from January 2022 to December 2023, yielding a total of $n = 501$ observations per stock. We compute daily log-returns and rescale them by multiplying with 100. Specifically, if $P_{i,j}$ denotes the closing price of the j-th stock on the i-th trading day, then the (i, j)-th entry of the log-return matrix $X \in \mathbb{R}^{501\times 286}$ is

$$X_{i,j} = 100 \log\left(\frac{P_{i+1,j}}{P_{i,j}}\right),$$

for $i = 1, \ldots, 501$, $j = 1, \ldots, 286$. We treat the rows of X as independent and identically distributed realizations of a random vector representing the scaled daily log-returns of the 286 stocks.

Estimating the covariance matrix of these log-returns is of fundamental importance in financial econometrics and quantitative portfolio management (Fan et al. 2016). The covariance matrix captures the co-movement structure of asset returns and serves as a key input for tasks such as mean–variance portfolio optimization, risk forecasting, and asset allocation (Markowitz 1952).

Moreover, two stocks are conditionally uncorrelated given the others if and only if the corresponding precision matrix entry is zero (Lauritzen 1996; Friedman et al. 2008). Thus, estimating a sparse precision matrix can be viewed as recovering a financial network that captures correlations between stocks after accounting for common market factors. Such networks provide insight into sectoral interactions, clustering behavior, and potential pathways of systemic risk and contagion (Torri and Giacometti 2017; Millington and Niranjan 2020). The M-matrix constraint further enhances interpretability and stability by enforcing positive dependence, consistent with economic intuition, and often leads to improved performance in high-dimensional settings.

Since the true underlying precision matrix is unknown, direct evaluation of the estimated graph through the true positive rate, false positive rate, or estimation error is not possible. However, stocks belonging to the same industrial sector are expected to exhibit stronger dependence. This motivates the use of *modularity* (Newman 2006) as an external measure for assessing the quality of the estimated graph, with the GICS sector

labels serving as the reference community structure. Modularity is a widely used measure to evaluate community detection in networks. For a graph $\mathcal{G}$ with vertex set V, edge set E and the adjacency matrix A, the modularity of the graph is defined as

$$Q := \frac{1}{2|E|} \sum_{i,j=1}^{|V|} \left(A_{ij} - \frac{d_i d_j}{2|E|} \right) I(c_i = c_j)$$

where, d_i denotes the degree of the i-th node, c_i is the type/community of the i-th node, and as defined in Section 2, $I(a = b) = 1$, when $a = b$, and otherwise it is 0. The value of Q for any graph or network ranges between -0.5 and 1. In our context, a high value of modularity indicates more connections between the stocks belonging to the same sector and lower connection between stocks belonging to different sectors. Hence, we expect the modularity to be high for a well-estimated precision matrix. We compare the performance of the proposed D-Gibbs, A-Gibbs and I-Gibbs sampler with GLASSO (Friedman et al. 2008), CLIME (Cai et al. 2011) and GGL (Egilmez et al. 2016). For the proposed samplers, the slab inclusion probability q is selected via the grid-search procedure described in Section 4.4. For GLASSO, CLIME and GGL, the respective hyper-parameters have been determined using 5-fold cross-validation. Here also, sparsity is induced by thresholding the off-diagonal entries at 10^{-3} in CLIME algorithm, following Cai et al. (2011). The values of modularity for various algorithms are provided in Table 4. Natural uncertainty quantification through the posterior is a key advantage of Bayesian methods. Following Barbieri and Berger (2004), we use marginal inclusion probabilities to assess the uncertainty associated with the graphs selected by the proposed algorithms. Further details and corresponding results are provided in Section S.4 of the online resource document.

As reported in Table 4, both A-Gibbs and I-Gibbs samplers yield significantly higher modularity values, indicating more effective community detection than GLASSO, CLIME and GGL. The low modularity achieved by D-Gibbs can be attributed to its recovery of a considerably larger number of *between-sector* edges, as shown in Table S6 in the online resource.

Figure 2 depicts the graph structures estimated by the various algorithms under comparison. The figure highlights clear differences in the recovered network topologies across methods. In particular, the graphs obtained from the augmented and interweaving Gibbs samplers exhibit more coherent community structure, with stronger within-sector connectivity and fewer spurious cross-sector edges. In contrast, the graphs estimated by competing methods appear either overly dense or fail to capture the underlying modular organization effectively. This visual evidence is consistent with the quantitative results reported in Table 4, where the augmented and interweaving methods achieve substantially higher modularity scores, indicating improved recovery of meaningful structural patterns in the data.

6 Discussion and Conclusion

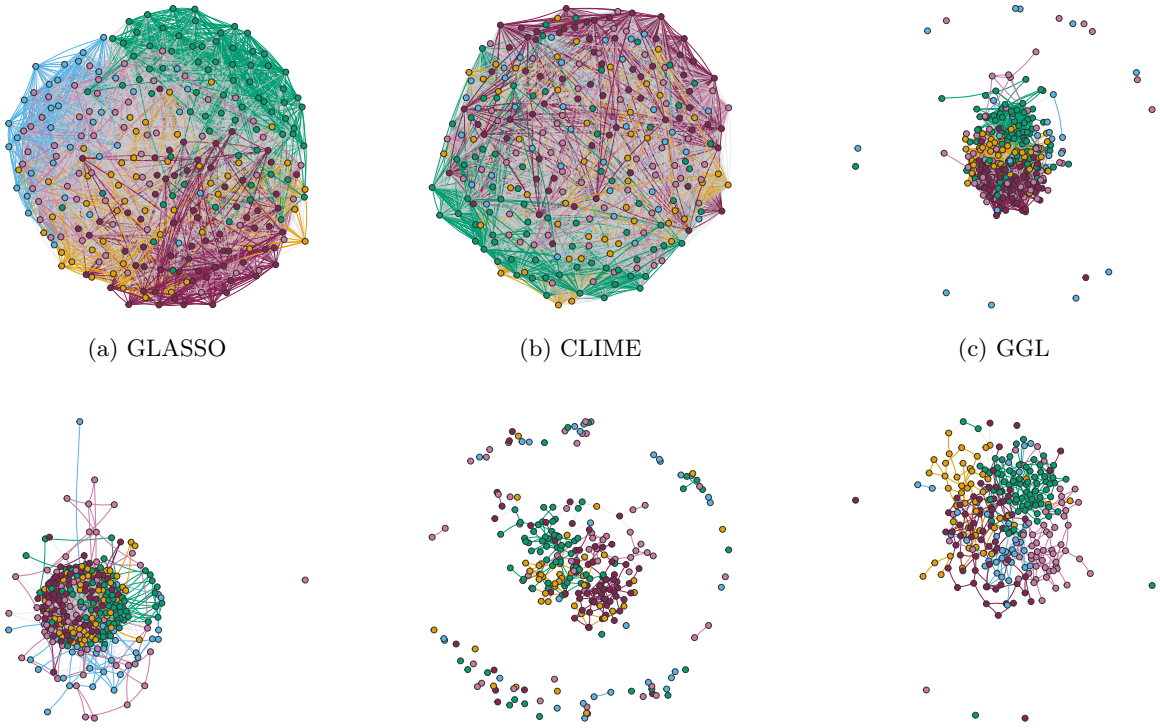
We proposed a scalable Bayesian framework for sparse precision matrix estimation under M-matrix constraints using a generalized posterior based on the convex D-trace loss. By avoiding the Gaussian likelihood and its log-determinant term, the method remains computationally tractable in high dimensions while supporting sparsity learning and uncertainty quantification through spike-and-slab priors.

Our main methodological contribution is the development of efficient posterior samplers. A novel matrix-variate augmentation induces conditional independence among precision matrix entries, enabling joint updates and fast matrix-normal sampling that substantially reduces computational cost. An interweaving sampler further improves mixing by combining augmented and direct parameterizations. Simulations and real-data analyses show improved graph recovery and competitive precision matrix estimation relative to existing methods while maintaining scalability. In financial data, the estimated networks recover meaningful sectoral structure, highlighting the practical utility of the approach.

Future work includes studying posterior contraction and graph selection consistency under the generalized framework, as well as extending the methodology to dynamic graphical models. More broadly, the combination of generalized Bayesian inference and augmentation-based computation

Table 4: Comparison of modularity scores for precision matrix estimation algorithms applied to S&P 500 financial time-series data from five GICS sectors. The graph induced by each estimated precision matrix is evaluated using modularity Q, with sector labels serving as the reference community structure.

Algorithm	**GLASSO**	**CLIME**	**GGL**	**D-Gibbs**	**A-Gibbs**	**I-Gibbs**
Modularity (Q)	0.136365	0.032529	0.283264	0.151645	0.454814	**0.570447**



(a) GLASSO (b) CLIME (c) GGL

(d) D-Gibbs (e) A-Gibbs (f) I-Gibbs

Fig 2 Network visualization of the estimated precision matrices for daily log-returns of 286 stocks from January 2022 to December 2023, using (a) GLASSO, (b) CLIME, (c) GGL, (d) D-Gibbs, (e) A-Gibbs and (f) I-Gibbs. The colored edges indicate edges that connect stocks from same sector, and the gray edges represent the cross-sector edges.

may be useful for large-scale structured covariance estimation problems.

Supplementary information. An online supplementary document (PDF), entitled "Supplementary Document for Bayesian Graphical Models under Positivity Constraints: A Scalable Generalized Likelihood Approach," accompanies this manuscript. It contains proofs of supporting lemmas and additional details of the numerical experiments.

Acknowledgements. The authors acknowledge the University of Florida Research Computing for providing access to the high-performance computing infrastructure of HiPerGator, which was used to perform the numerical experiments presented in the study. Artificial intelligence (AI)-assisted tools were used solely for spelling and grammatical checks during manuscript preparation. The authors assume full responsibility for the scientific content, accuracy, and conclusions of the manuscript.

Declarations

Funding

The work of Sirani M. Perera was supported by the Division of Mathematical Sciences (DMS) of the U.S. National Science Foundation (NSF) under Grant No. DMS-2410676; the work of Kshitij Khare was supported under Grant No. DMS-2410677; and the work of Partha Sarkar was supported under Grant No. DMS-2506060.

Ethics Approval

Not applicable.

Competing interests

The authors have no competing interests to declare that are relevant to the content of this article.

Consent for publication

All authors have read and approved the manuscript and consent to its publication.

Data Availability

The financial data used in this study consist of daily adjusted closing prices for S&P 500 constituent stocks from January 2022 to December 2023, obtained from Yahoo Finance via the `quantmod` R package (version 0.4.29). Sector classifications were obtained from the Wikipedia page https://en.wikipedia.org/wiki/List_of_S%26P_500_companies.

Code Availability

An anonymous version of the source code is available at https://github.com/srnlrh/DAI-Gibbs.

References

Agrawal, R., Roy, U., Uhler, C.: Covariance Matrix Estimation under Total Positivity for Portfolio Selection. Journal of Financial Econometrics **20**(2), 367–389 (2022) https://doi.org/10.1093/jjfinec/nbaa018

Barbieri, M.M., Berger, J.O.: Optimal predictive model selection. The Annals of Statistics **32**(3) (2004) https://doi.org/10.1214/009053604000000238

Bhattacharya, A., Chakraborty, A., Mallick, B.K.: Fast sampling with Gaussian scale mixture priors in high-dimensional regression. Biometrika **103**(4), 985–991 (2016) https://doi.org/10.1093/biomet/asw042

Besag, J.: Statistical Analysis of Non-Lattice Data. Journal of the Royal Statistical Society Series D: The Statistician **24**(3), 179–195 (1975) https://doi.org/10.2307/2987782

Bissiri, P.G., Holmes, C.C., Walker, S.G.: A general framework for updating belief distributions. Journal of the Royal Statistical Society Series B **78**(5), 1103–1130 (2016) https://doi.org/10.1111/rssb.12158

Bhadra, A., Sagar, K., Rowe, D., Banerjee, S., Datta, J.: Evidence Estimation in Gaussian Graphical Models Using a Telescoping Block Decomposition of the Precision Matrix. Journal of Machine Learning Research **25**(295), 1–43 (2024) https://doi.org/10.48550/arXiv.2205.01016

Cai, T., Liu, W., Luo, X.: A Constrained l_1 Minimization Approach to Sparse Precision Matrix Estimation. Journal of the American Statistical Association **106**(494), 594–607 (2011) https://doi.org/10.1198/jasa.2011.tm10155

Dixit, A., Roy, V.: Posterior impropriety of some sparse bayesian learning models. Statistics & Probability Letters **171**, 109039 (2021) https://doi.org/10.1016/j.spl.2021.109039

Egilmez, H., Pavez, E., Ortega, A.: Graph Learning from Data under Structural and Laplacian Constraints. IEEE Journal of Selected Topics in

Signal Processing **PP** (2016) https://doi.org/10.1109/JSTSP.2017.2726975

Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso. Biostatistics **9**(3), 432–441 (2008) https://doi.org/10.1093/biostatistics/kxm045

Friedman, J., Hastie, T., Tibshirani, R.: Applications of the lasso and grouped lasso to the estimation of sparse graphical models. Technical report, Stanford University (2010)

Fan, J., Liao, Y., Liu, H.: Approaches to High-Dimensional Covariance and Precision Matrix Estimations, pp. 100–134. Wiley-IEEE Press, Hoboken, NJ, USA (2016). https://doi.org/10.1002/9781118745540.ch6

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., Rubin, D.: Bayesian Data Analysis, 3rd ed edn. CreateSpace, United States (2013). https://doi.org/10.1201/9780429258411

George, E.I., McCulloch, R.E.: Variable selection via Gibbs sampling. Journal of the American Statistical Association **88**(423), 881–889 (1993) https://doi.org/10.1080/01621459.1993.10476353

Gao, Z., Sagar, K., Bhadra, A.: An Order of Magnitude Time Complexity Reduction for Gaussian Graphical Model Posterior Sampling Using a Reverse Telescoping Block Decomposition. arXiv preprint arXiv 2509.26385 (2026) https://doi.org/10.48550/arXiv.2509.26385 [stat.ME]

Hobert, J.P., Casella, G.: The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models. Journal of the American Statistical Association **91**(436), 1461–1473 (1996) https://doi.org/10.1080/01621459.1996.10476714

Hobert, J.P., Casella, G.: Functional Compatibility, Markov Chains, and Gibbs Sampling with Improper Posteriors. Journal of Computational and Graphical Statistics **7**(1), 42–60 (1998) https://doi.org/10.2307/1390768

Hennessy, D.A., Lapan, H.E.: The Use of Archimedean Copulas to Model Portfolio Allocations. Mathematical Finance **12** (2002) https://doi.org/10.1111/1467-9965.00136

Khare, K., Oh, S.-Y., Rajaratnam, B.: A convex pseudolikelihood framework for high dimensional partial correlation estimation with convergence guarantees. Journal of the Royal Statistical Society Series B: Statistical Methodology **77**(4), 803–825 (2015) https://doi.org/10.1111/rssb.12088

Lauritzen, S.L.: Graphical Models. Oxford University Press, Oxford (1996). https://doi.org/10.1093/9780191775284.001.0001

Lam, C., Fan, J.: Sparsistency and rates of convergence in large covariance matrix estimation. The Annals of Statistics **37**(6B), 4254–4278 (2009) https://doi.org/10.1214/09-AOS720

Lindsay, B.G.: Composite Likelihood Methods. In: Prabhu, N.U. (ed.) Statistical Inference from Stochastic Processes. Contemporary Mathematics, vol. 80, pp. 221–239. American Mathematical Society, Providence, RI (1988). https://doi.org/10.1090/conm/080/999014

Lake, B.M., Tenenbaum, J.B.: Discovering Structure by Learning Sparse Graphs (2010)

Lauritzen, S., Uhler, C., Zwiernik, P.: Maximum likelihood estimation in Gaussian models under total positivity. The Annals of Statistics **47**(4), 1835–1863 (2019) https://doi.org/10.1214/17-AOS1668

Lauritzen, S., Uhler, C., Zwiernik, P.: Total positivity in exponential families with application to binary variables. The Annals of Statistics **49**(3), 1436–1459 (2021) https://doi.org/10.1214/20-AOS2007

Markowitz, H.: Portfolio Selection. The Journal of Finance **7**(1), 77–91 (1952) https://doi.org/10.2307/2975974

Matthews, B.W.: Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et Biophysica Acta - Protein Structure **405**(2), 442–451 (1975) https://doi.org/10.1016/0005-2795(75)90109-9

Meinshausen, N., Bühlmann, P.: High-dimensional graphs and variable selection with the Lasso. The Annals of Statistics **34**(3), 1436–1462 (2006) https://doi.org/10.1214/009053606000000281

Millington, T., Niranjan, M.: Partial correlation financial networks. Applied Network Science **5** (2020) https://doi.org/10.1007/s41109-020-0251-z

Müller, A., Scarsini, M.: Archimedean copulae and positive dependence. Journal of Multivariate Analysis **93**(2), 434–445 (2005) https://doi.org/10.1016/j.jmva.2004.04.003

Newman, M.E.J.: Modularity and community structure in networks. Proceedings of the National Academy of Sciences **103**(23), 8577–8582 (2006) https://doi.org/10.1073/pnas.0601602103

Narisetty, N.N., He, X.: Bayesian variable selection with shrinking and diffusing priors. The Annals of Statistics **42**(2), 789–817 (2014) https://doi.org/10.1214/14-AOS1207

Nishimura, A., Suchard, M.A.: Prior-Preconditioned Conjugate Gradient Method for Accelerated Gibbs Sampling in "Large n, Large p" Bayesian Sparse Regression. Journal of the American Statistical Association **118**(544), 2468–2481 (2023) https://doi.org/10.1198/jasa.2009.0126

Ni, Y., Stingo, F.C., Baladandayuthapani, V.: Bayesian Covariate-Dependent Gaussian Graphical Models with Varying Structure. Journal of Machine Learning Research **23**, 1–48 (2022)

Peng, J., Wang, P., Zhou, N., Zhu, J.: Partial Correlation Estimation by Joint Sparse Regression Models. Journal of the American Statistical Association **104**(486), 735–746 (2009) https://doi.org/10.1198/jasa.2009.0126

Ravikumar, P., Wainwright, M.J., Raskutti, G., Yu, B.: High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. Electronic Journal of Statistics **5**, 935–980 (2011) https://doi.org/10.1214/11-EJS631

Rocha, G.V., Zhao, P., Yu, B.: A path following algorithm for Sparse Pseudo-Likelihood Inverse Covariance Estimation (SPLICE). arXiv 0807.3734 (2008) https://doi.org/10.48550/arXiv.0807.3734 [stat.ME]

Slawski, M., Hein, M.: Estimation of positive definite M-matrices and structure learning for attractive Gaussian Markov random fields. Linear Algebra and its Applications **473**, 145–179 (2015) https://doi.org/10.1016/j.laa.2014.04.020

Samanta, S., Khare, K., Michailidis, G.: A generalized likelihood-based Bayesian approach for scalable joint regression and covariance selection in high dimensions. Statistics and computing **32**(3), 47 (2022) https://doi.org/10.1007/s11222-022-10102-5

Shiratori, T., Takano, Y.: DC Algorithm for Estimation of Sparse Gaussian Graphical Models. PLoS ONE **19**(12), 0315740 (2024) https://doi.org/10.1371/journal.pone.0315740

Torri, G., Giacometti, R.: Sparse Precision Matrices for Minimum Variance Portfolios. SSRN Electronic Journal (2017) https://doi.org/10.1007/s10287-019-00344-6

Rijsbergen, C.J.: Information Retrieval, 2nd edn. Butterworths, London (1979)

Varin, C., Reid, N., Firth, D.: An overview of composite likelihood methods. Statistica Sinica, 5–42 (2011)

Wang, H.: Bayesian Graphical Lasso Models and Efficient Posterior Computation. Bayesian Analysis **7**(4), 867–886 (2012) https://doi.org/10.1214/12-BA729

Wang, Y., Roy, U., Uhler, C.: Learning High-dimensional Gaussian Graphical Models under Total Positivity without Adjustment of Tuning Parameters. In: Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics. Proceedings of

Machine Learning Research, vol. 108, pp. 2698–2708. PMLR, online (2020). https://doi.org/10.48550/arXiv.1906.05159

Ying, J., De Miranda Cardoso, J.V., Palomar, D.P.: Adaptive Estimation of Graphical Models under Total Positivity. In: Proceedings of the 40th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 202, pp. 40054–40074. PMLR, Honolulu, Hawaii, USA (2023). https://doi.org/10.48550/arXiv.2210.15471

Yu, Y., Meng, X.-L.: To Center or Not to Center: That Is Not the Question—An Ancillarity–Sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Efficiency. Journal of Computational and Graphical Statistics **20**(3), 531–570 (2011) https://doi.org/10.1198/jcgs.2011.203main

Zheng, Y., Duan, L.L.: Gibbs sampling using anti-correlation gaussian data augmentation, with applications to l1-ball-type models. Journal of Computational and Graphical Statistics **34**(4), 1601–1611 (2025) https://doi.org/10.1080/10618600.2025.2473932

Zhang, R., Yao, Y., Ghosh, M.: Contraction of a quasi-bayesian model with shrinkage priors in precision matrix estimation. Journal of Statistical Planning and Inference **221**, 154–171 (2022) https://doi.org/10.1016/j.jspi.2022.03.003

ZHANG, T., ZOU, H.: Sparse precision matrix estimation via lasso penalized D-trace loss. Biometrika **101**(1), 103–120 (2014) https://doi.org/10.1093/biomet/ash.059

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.pdf