



Extending GLASSO to non-Gaussian settings: sparse concentration estimation via EM algorithm

György Terdik¹ · Abdaljbbar B. A. Dawod²

Received: 28 January 2026 / Accepted: 1 June 2026
© The Author(s) 2026

Abstract

This paper develops graphical modeling techniques for multivariate non-Gaussian data, in which the sparsity pattern of the concentration matrix encodes conditional uncorrelations among variables. This structure enables efficient covariance selection, yielding concentration graph models that extend beyond the Gaussian case. Within elliptically symmetric distributions and indeed in an even broader family, distributions for which the best prediction is linear, zero partial correlations imply zero conditional correlations. That preserves the interpretability of edges in the graphical model. Focusing on elliptically symmetric families such as the generalized hyperbolic and power exponential distributions, we propose a modified graphical LASSO (GLASSO) framework for estimating sparse concentration matrices. The method is implemented using an EM-type algorithm adapted from Gaussian GLASSO to accommodate non-Gaussian likelihoods. Simulation studies and real-data applications demonstrate the effectiveness and robustness of the proposed estimators in recovering underlying graphical structures despite deviations from normality.

Keywords Graphical LASSO (GLASSO) · Sparse precision matrix estimation · EM algorithm · Power exponential distribution · Generalized symmetric hyperbolic distribution

✉ Abdaljbbar B. A. Dawod
dawod.abdaljbbar@inf.unideb.hu

György Terdik
terdik.gyorgy@inf.unideb.hu

¹ Department of Information Technology, Faculty of Informatics, University of Debrecen, Kassai, Debrecen 4028, Hungary

² Doctoral School of Informatics, University of Debrecen, Kassai, Debrecen 4028, Hungary

1 Introduction

In the multivariate normal distribution, the relationship between partial correlation and conditional independence is well established: a partial correlation of 0 indicates conditional independence between the variables. Partial correlations can be obtained by inverting the covariance matrix, whose inverse $\Theta = \Sigma^{-1}$, known as the concentration matrix,¹ plays a central role in probabilistic graphical modeling. In fact, in the multivariate Gaussian case, the partial correlation is zero if and only if the variables are conditionally independent.

Kelker (1970) and Bh (1998) examined the role of partial correlation in extending results from the Gaussian case to the broader class of elliptically symmetric distributions. They demonstrated that a zero partial correlation implies conditional independence if and only if the distribution is multivariate normal. Consequently, in non-Gaussian elliptical families, the interpretation of a zero partial correlation is limited to indicating zero conditional correlation. The elliptically symmetric variates $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, h)$ is defined by the location-scale extension of a spherically distributed variate $\mathbf{W}$ so that $\mathbf{X} = \boldsymbol{\mu} + \Sigma^{1/2}\mathbf{W}$, where $\boldsymbol{\mu} \in \mathbb{R}^d$, Σ is a full rank variance matrix, and the function h is called density generator; see Fang et al. (2017), Muirhead (2009). It is worth noticing that the variance matrix Σ is equal to the variance of $\mathbf{X}$ multiplied by a constant.

Elliptically symmetric variates provide a broad framework for modeling dependence structures in multivariate data. Within this class, the Multivariate Symmetric Generalized Hyperbolic (MSGH) and Multivariate Power Exponential (MPE) distributions offer greater flexibility in capturing heavy tails, peakedness, and non-Gaussian behavior—making them particularly valuable for realistic modeling and robust inference in high-dimensional settings. These distributions play a pivotal role in modern statistical modeling, particularly in areas where data exhibit heavy tails, or kurtosis beyond Gaussian assumptions (Mcneil et al. 2005; Weibel et al. 2024; Rossell and Zwiernik 2021). These flexible distributions extend the normal framework, enabling more accurate modeling of complex real-world phenomena in finance, signal processing, environmental science, and machine learning (Prause 1999). MSGH models provide a unifying structure that encompasses several important distributions, such as Student's t , Laplace, etc., and hyperbolic distributions, offering versatility in capturing tail behavior (Browne and McNicholas 2015). Similarly, the MPE family generalizes the Gaussian model by controlling kurtosis via shape parameters, enabling robust estimation under outlier contamination (Gómez et al. 1998). The methods developed by Ravikumar et al. (2011) and Janková and van de Geer (2015) provide consistent, asymptotically normal, and accurate sparse concentration matrix estimation in high-dimensional contexts. Recent studies further demonstrate their effectiveness in robust clustering and financial risk analysis, where they outperform Gaussian mixtures in capturing elliptical and heavy-tailed structures (Franczak et al. 2013). Their flexibility in modeling a wide range of tail behaviors makes these distributions superior to the Gaussian model for accurate density estimation and risk

¹ We prefer usage of "concentration matrix" rather than "precision matrix" since the latter is used in a more general sense.

measurement. Accordingly, understanding their association structures—via concentration matrix estimation—is essential for identifying valid, conditionally uncorrelated networks and extending statistical modeling beyond light-tailed assumptions.

Graphical models facilitate inference by mapping dependencies via undirected graphs, where edges correspond to non-zero entries in the concentration matrix Θ . In high-dimensional contexts, penalized likelihood methods like the Graphical least absolute shrinkage and selection operator (GLASSO) are employed to induce sparsity in Θ , thereby efficiently recovering the structure of Gaussian Graphical Models (GGMs) (Friedman et al. 2008). This regularization reveals conditional dependencies through the non-zero off-diagonal elements. At the same time, zero entries imply conditional independence—or conditionally uncorrelated associations in non-Gaussian models—providing a parsimonious representation of complex multivariate data (Yamada et al. 2025; Marzban Vaselabadi et al. 2025).

The concentration matrix is central to assessing conditional independence under Gaussian assumptions. Huang et al. (2016) introduced the Partial Correlation Screening method for row-wise estimation of concentration matrices. More recently, Erickson and Rydén (2025) developed a joint regression approach for estimating high-dimensional sparse concentration and partial correlation matrices. Estimating the concentration matrix Θ , particularly identifying its zero entries, is crucial for defining new parameters that facilitate analyses among Gaussian variables. The GLASSO provides an efficient computational framework for this estimation but relies on the assumption of normality. In high-dimensional settings of heavy-tailed elliptically symmetric distributions, various methods have been proposed to estimate the sparse concentration matrix. One major contribution is replacing the Gaussian likelihood with heavy-tailed ellipticals [e.g., Student- t 's likelihood (Finegold and Drton 2014)] to reduce the sensitivity of the heavy tail. The second mainstream adapts the penalty, either via Sorted- ℓ_1 Penalized Estimators (SLOPE), such as GSLOPE and (TSLOPE) for the Gaussian and Student- t variants (Riccobello et al. 2025), or via weights within GLASSO (WGLASSO). The third contribution uses robust M-estimators and shrinkage, such as Tyler's (Chen et al. 2011), Partial Correlation GLASSO (PC-GLASSO) (Zhao et al. 2014), and Maronna's M-estimators with regularization, to obtain a high-breakdown scatter estimate, and then plugs it into the procedure for sparse-concentration matrix estimation. At the same time, the fourth approach uses a divergence-based objective function (γ -LASSO) (Onizuka and Hashimoto 2025; Hirose et al. 2017) or a trimmed objective function (Yang and Lozano 2015) to downweigh the heavy-tail effect or the contamination. In addition, rank-based and nonparanormal transformations estimate correlation or copula structures without strong assumptions about tail behavior, such as constrained ℓ_1 minimization estimator (CLIME/T-CLIME) and their spatial-sign extensions (SGLASSO and SCLIME) (Lu and Feng 2025), which provide computationally attractive constrained- ℓ_1 solutions.

Despite their effectiveness, existing methods suffer from limited tail flexibility, sensitivity to misspecification, efficiency loss, and computational challenges in high dimensions. These limitations motivate the proposed emLASSO, which exploits the equivalence between zero partial and conditional correlations under linear prediction to recover sparse concentration matrices for heavy-tailed MSGH and MPE distribu-

tions. An EM-based penalized likelihood framework is developed for these families, providing flexible tail modeling and robust, positive-definite concentration estimates via closed-form E-steps and simple penalized M-steps, without ad hoc trimming or multi-stage procedures.

The remainder of the paper is organized as follows. Section 2 studies the relationship between partial and conditional uncorrelatedness under linear prediction, including elliptically symmetric variates. Section 3 introduces the proposed emLASSO for elliptically symmetric distributions and describes estimation and inference via stochastic and variational EM algorithms. Section 4 evaluates emLASSO against competing methods and presents an application to body-dimension network data. Section 5 concludes the paper.

2 Partially and conditionally uncorrelated elliptically symmetric variates

The importance of partial correlation lies in its direct connection to the variance matrix and straightforward inference about possible conditional independence between variates. Let $\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\Sigma}_{\mathbf{X}})$ be Gaussian multivariate and $\mathbf{X}_{\setminus(p,q)}$ denote the vector with dimension $d - 2$ when the p^{th} and q^{th} entries are missing from $\mathbf{X}$ and $\mathbf{X}_{p,q} = [X_p, X_q]^T$. The conditional independence, $X_p \perp\!\!\!\perp X_q | \mathbf{X}_{\setminus(p,q)}$, of two entries X_p and X_q , ($p \neq q$) with respect to $\mathbf{X}_{\setminus(p,q)}$ holds if and only if the p, q -th entry of the concentration matrix $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$ equals zero, see the Inverse Variance Lemma in Whittaker (2009). The off-diagonal entries of the scaled $\boldsymbol{\Theta}$ have a remarkable properties being the negative correlation coefficients ρ_{X_p, X_q} and in Gaussian case they are the partial correlation coefficients $\rho_{X_p, X_q | \bullet \mathbf{X}_{\setminus(p,q)}}$ as well. Since the best predictor $E[\mathbf{X}_{p,q} | \mathbf{X}_{\setminus(p,q)}]$ and the the linear least squares predictor $\hat{X}_{p,q | \bullet \mathbf{X}_{\setminus(p,q)}}$ (LLSP) are the same in Gaussian case. The conditional covariance and the partial covariance have an important role in these notions, hence we call attention to the difference between them, the conditional covariance $C_{X_p, X_q | \mathbf{X}_{\setminus(p,q)}}$ and partial covariance $C_{X_p, X_q | \bullet \mathbf{X}_{\setminus(p,q)}}$, namely

$$\begin{aligned} C_{X_p, X_q | \mathbf{X}_{\setminus(p,q)}} &= \text{Cov}(X_p, X_q | \mathbf{X}_{\setminus(p,q)}) \\ &= E[(X_p - E[X_p | \mathbf{X}_{\setminus(p,q)}])(X_q - E[X_q | \mathbf{X}_{\setminus(p,q)}]) | \mathbf{X}_{\setminus(p,q)}], \end{aligned}$$

and $C_{X_p, X_q | \bullet \mathbf{X}_{\setminus(p,q)}} = \text{Cov}(X_p - \hat{X}_{p | \bullet \mathbf{X}_{\setminus(p,q)}}, X_q - \hat{X}_{q | \bullet \mathbf{X}_{\setminus(p,q)}})$, where

$$\hat{X}_{p | \bullet \mathbf{X}_{\setminus(p,q)}} = \mu_{X_p} + \rho_{X_p, X_q} \sigma_{X_p} \boldsymbol{\Sigma}_{\mathbf{X}_{\setminus(p,q)}}^{-1} (\mathbf{X}_{\setminus(p,q)} - \boldsymbol{\mu}_{\mathbf{X}_{\setminus(p,q)}}),$$

say.

Beyond Gaussian variates, the main issue now is interpreting the partially uncorrelated property in the conditional inference. For notational simplicity, let $\mathbf{Y} = \mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}$.

Lemma 1 Consider the multivariate $\mathbf{Y} = [Y_p, Y_q, Y_{\setminus p,q}]$. Assume that the best prediction of Y_p with respect to $Y_{\setminus p}$ is linear i.e. $E(Y_p | Y_{\setminus p}) = \hat{Y}_{p|\bullet}[Y_{\setminus p}]$ then from $Y_p \perp Y_q | Y_{\setminus p,q}$ follows that $Y_p \perp Y_q | Y_{\setminus p,q}$, in other words, if Y_p and Y_q are partially uncorrelated, then they are conditionally uncorrelated as well.

See the proof in Appendix 1.

If $\mathbf{Y}$ is an elliptically symmetric random variable, $\mathbf{Y} \sim E_d(\mathbf{0}, \Sigma, h)$, then the conditional expectation is linear, see Fang et al. (2017), Rossell and Zwiernik (2021), therefore the message from the Lemma 1 is that finding zeros in the concentration matrix leads to conditional uncorrelation among elliptically symmetric random variables as well.

Remark 1 The Generalized Hyperbolic variate $\mathbf{X} \sim \mathcal{GH}(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$, is given by the stochastic equation

$$\mathbf{X} = \mu + V\gamma + \sqrt{V}\Sigma^{1/2}Z,$$

(see the next Section for details). The parameter γ refers to the asymmetry of $\mathbf{X}$, if $\gamma \neq 0$, then $\mathbf{X}$ is not symmetric. Let $[\mathbf{X}_1, \mathbf{X}_2]$ be a partition of $\mathbf{X}$, then one can show [see Terdik (2022) for conditional cumulants] that,

$$\begin{aligned} E[\mathbf{X}_2 | \mathbf{X}_1] &= \hat{\mathbf{X}}_{\mathbf{X}_2 | \mathbf{X}_1} + E(V | \mathbf{X}_1) \delta_2, \\ \Sigma_{\mathbf{X}_2 | \mathbf{X}_1} &= E(V | \mathbf{X}_1) \Sigma_{\mathbf{X}_2 | \mathbf{X}_1} + \sigma_{V | \mathbf{X}_1}^2 \delta_2 \delta_2^T, \end{aligned}$$

where $\delta_2 = \gamma_2 - \Sigma_{21} \Sigma_{11}^{-1} \gamma_1$, $\gamma = [\gamma_1, \gamma_2]$ accordingly, and conclude that among the hyperbolic distributions, the only symmetrical one fulfills the assumption of Lemma 1. Actually, it also concerns the family of mean-variance-mixed normal distributions.

3 emLASSO for some non-Gaussian elliptically symmetric variates

We now consider a modification of the LASSO algorithm for a subset of elliptically symmetric variates, namely, variance-mixtures of normal variates. Actually, we use the EM algorithm, which enables efficient model selection even when the data do not follow a Gaussian distribution. Finegold and Drton (2014) have used the Gaussian LASSO algorithm for the multivariate Student t distribution. The key idea is that the weight variate τ in the normal-variance-mixture representation of the multivariate Student t distribution can be treated as a latent parameter. Then, the sample-based prediction $\hat{\tau}$ of the weight-variate τ in the E-step enables the use of Gaussian LASSO in the M-step, yielding emLASSO. We maintain the robustness of LASSO during computation with emLASSO for heavy-tailed distributions as well.

3.1 emLASSO_H for multivariate symmetric generalized hyperbolic distributions

The MSGH distributions are constructed as a standard variance mixture with a Generalized Inverse Gaussian (GIG) weights, see Mcneil et al. (2005), Example 3.8. It turns out that MSGH is a subfamily of the elliptically symmetric variates, and as it is the zeros of the concentration matrix imply the conditional uncorrelation between random variables. In this section, we give a clear formula for the prediction $\hat{\tau}$ of the weight variate τ (see Eq. (1)) which inherits the LASSO algorithm. The MSGH is constructed as a normal variance mixture with a Generalized Inverse Gaussian (GIG) weights, Mcneil et al. (2005), Example 3.8. The mixing variates GIG include exceptional cases, such as the gamma (G), inverse gamma (IGam), exponential (Exp), positive hyperbolic (pHyp), Lévy, inverse Gaussian (IG), and Dirac distribution (as a limiting case), see Paoletta (2007), Sect. 9.4.

In this section, we present an explicit formula for predicting the weight variable τ , enabling the application of the LASSO algorithm. Let us start with the stochastic representation

$$\mathbf{X} = \boldsymbol{\mu} + \sqrt{V}\boldsymbol{\Sigma}^{1/2}Z$$

of an MSGH variates $\mathbf{X}$ with parameters $\mathcal{MSGH}(\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma})$. Here $Z \sim \mathcal{N}_d(\mathbf{0}, \mathbf{I}_d)$, $V \sim \mathcal{GIG}(\lambda, \chi, \psi)$ is a non-negative scalar-valued variate, which is independent on Z , where the associated parameters are listed as

- $\boldsymbol{\mu} \in \mathbb{R}^d$: location parameter.
- $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$: positive definite dispersion matrix governing scale and the shape of elliptical level sets.
- γ : skewness parameter; $\gamma = 0$ yields symmetry about $\boldsymbol{\mu}$, while $|\gamma|$ controls the magnitude of skewness.
- $\lambda \in \mathbb{R}$: shape parameter influencing tail behavior and kurtosis.
- $\chi > 0$: tail parameter; smaller values correspond to heavier tails, whereas larger values approach Gaussian behavior.
- $\psi > 0$: scale/tail parameter complementary to χ ; larger values imply lighter tails and reduced variability, while smaller values induce heavier tails and overdispersion.

Further properties are discussed in Paoletta (2007), and alternative parameterizations are provided in Weibel et al. (2024), see Terdik (2026) for the precision matrix. The density function of $\mathbf{X} \sim \mathcal{MSGH}(\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is

$$f_{\mathbf{X}}(\mathbf{X}) = c_{\lambda, \chi, \psi} |\boldsymbol{\Sigma}|^{-1/2} \frac{K_{\lambda-d/2} \left((\psi (\chi + q(\mathbf{X})))^{1/2} \right)}{(\psi (\chi + q(\mathbf{X})))^{(d/2-\lambda)/2}},$$

where $q(\mathbf{X}) = (\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Theta} (\mathbf{X} - \boldsymbol{\mu})$ and K_{λ} denotes a modified Bessel function of the third kind with index λ and the normalizing constant $c_{\lambda, \chi, \psi}$ is given by

$$c_{\lambda, \chi, \psi} = \frac{(\sqrt{\chi\psi})^{-\lambda} \psi^{d/2}}{(2\pi)^{d/2} K_{\lambda}(\sqrt{\chi\psi})}.$$

The corresponding mixing weight, $V \sim \mathcal{GIG}(\lambda, \chi, \psi)$ has the density

$$f_V(x) = c_V x^{\lambda-1} \exp\left(-\frac{1}{2}(\chi x^{-1} + \psi x)\right),$$

with the constant $c_V = \chi^{-\lambda} (\sqrt{\chi\psi})^{\lambda} / 2K_{\lambda}(\sqrt{\chi\psi})$.

Now, the desired stochastic representation

$$\mathbf{X} = \boldsymbol{\mu} + \frac{1}{\sqrt{\tau}} \boldsymbol{\Sigma}^{1/2} \mathbf{Z}, \quad (1)$$

can be reached using the transformation $V = 1/\tau$, with the Jacobian $dv = 1/\tau^2 d\tau$, hence we have

$$f_{\tau}(x) = \frac{c_V}{2} x^{(-\lambda+1)/2-2} \exp\left(-\frac{1}{2}(\chi x + \psi x^{-1})\right). \quad (2)$$

The Gaussianity of $\mathbf{X}$ shows up, by the stochastic representation(1), under the constrain when τ is given hence

$$f(\mathbf{X}|\tau) = \frac{\tau^{d/2}}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{\tau}{2} q(\mathbf{X})\right).$$

We obtain the conditional density when $\mathbf{X} = \mathbf{x}$ is given

$$\begin{aligned} f(\tau|\mathbf{X}) &= \frac{f(\mathbf{X}|\tau)f(\tau)}{f(\mathbf{X})} = c\tau^{d/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{\tau}{2} q(\mathbf{X})\right) \tau^{-\lambda-1} \exp\left(-\frac{1}{2}(\chi\tau + \psi\tau^{-1})\right) \\ &\quad \times |\boldsymbol{\Sigma}_d|^{1/2} \frac{(\psi(\chi + q(\mathbf{X})))^{(d/2-\lambda)/2}}{K_{\lambda-d/2}((\psi(\chi + q(\mathbf{X})))^{1/2})} \\ &= c \frac{(\psi(\chi + q(\mathbf{X})))^{(d/2-\lambda)/2}}{K_{\lambda-d/2}((\psi(\chi + q(\mathbf{X})))^{1/2})} \tau^{d/2-\lambda-1} \exp\left(-\frac{1}{2}((\chi + q(\mathbf{x}))\tau + \psi\tau^{-1})\right), \end{aligned}$$

with an appropriate constant c . We conclude that $\tau|\mathbf{X} \sim \mathcal{GIG}(\lambda_{\tau}, \chi_{\tau}, \psi_{\tau})$, with parameters $\lambda_{\tau} = d/2 - \lambda$, $\chi_{\tau} = \psi$, $\psi_{\tau} = \chi + q(\mathbf{X})$. The expected value of a $\mathcal{GIG}(\lambda, \chi, \psi)$ variate is known (Paolella 2007, 9.4, p. 306–309), to be

$$\mathbb{E}[V^r] = \eta^r \frac{K_{\lambda+r}(\tilde{\omega})}{K_{\lambda}(\tilde{\omega})},$$

where $\eta = \sqrt{\chi/\psi}$ is a new scale parameter and $\tilde{\omega} = \sqrt{\chi\psi}$. Therefore, putting the parameters of $\mathcal{GIG}(\lambda_\tau, \chi_\tau, \psi_\tau)$ we obtain

$$\begin{aligned} E[\tau|\mathbf{X}] &= \sqrt{\frac{\chi_\tau}{\psi_\tau} \frac{K_{\lambda_\tau+1}(\sqrt{\chi_\tau\psi_\tau})}{K_{\lambda_\tau}(\sqrt{\chi_\tau\psi_\tau})}} \\ &= \sqrt{\frac{\psi}{\chi + q(\mathbf{X})} \frac{K_{d/2-\lambda+1}(\sqrt{\psi(\chi + q(\mathbf{X}))})}{K_{d/2-\lambda}(\sqrt{\psi(\chi + q(\mathbf{X}))})}}. \end{aligned}$$

We arrived at the

Lemma 2 Let the variate $\mathbf{X} \sim \mathcal{MSGH}(\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ be given by the stochastic representation

$$\mathbf{X} = \boldsymbol{\mu} + \frac{1}{\sqrt{\tau}} \boldsymbol{\Sigma}^{1/2} \mathbf{Z},$$

with weights (2). Then the best prediction $\hat{\tau}$ for the weight-variate τ by $\mathbf{X}$ is given by

$$\hat{\tau}(\mathbf{X}) = \sqrt{\frac{\psi}{\chi + q(\mathbf{X})} \frac{K_{d/2-\lambda+1}(\sqrt{\psi(\chi + q(\mathbf{X}))})}{K_{d/2-\lambda}(\sqrt{\psi(\chi + q(\mathbf{X}))})}}. \quad (3)$$

Remark 2 The Gamma random variable $\tau \sim \Gamma(\nu/2, \nu/2)$ can be obtain evaluating the limit of $\mathcal{GIG}(\lambda = -\nu/2, \chi = \nu, \psi)$ as $\psi \rightarrow 0$, see Mcneil et al. (2005), p. 80. Let us apply the limit of the function $K_\nu(z)$ at zero

$$K_\nu(z) \sim \frac{1}{2} \Gamma(\nu) \left(\frac{1}{2}z\right)^{-\nu}, \quad z \rightarrow 0$$

see DLMF (10.30.2) (Olver et al. 2020), use (3) and obtain

$$\begin{aligned} &\lim_{\psi \rightarrow 0} \sqrt{\frac{\psi}{\nu + q(\mathbf{X})} \frac{K_{d/2+\nu/2+1}(\sqrt{\psi(\nu + q(\mathbf{X}))})}{K_{d/2+\nu/2}(\sqrt{\psi(\nu + q(\mathbf{X}))})}} \\ &= \lim_{\psi \rightarrow 0} \sqrt{\frac{\psi}{\nu + q(\mathbf{X})} \frac{d/2 + \nu/2}{\sqrt{\psi(\nu + q(\mathbf{X}))}/2}} \\ &= \sqrt{\frac{1}{\nu + q(\mathbf{X})} \frac{d + \nu}{\sqrt{\nu + q(\mathbf{X})}}} = \frac{d + \nu}{\nu + q(\mathbf{X})}. \end{aligned}$$

The result $\hat{\tau} = E[\tau|\mathbf{X}] = (\nu + d) / (\nu + q(\mathbf{X}))$ coincides with the prediction $\hat{\tau}$ in the case of t -distribution, see Finegold and Drton (2014).

Example 1 Consider the distribution $MSGH(\lambda, \chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\lambda = -1/2$, which is called multivariate normal inverse Gaussian (MNIG). Let $\mathbf{X} \sim \mathcal{MNIG}(\chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\psi > 0, \chi > 0$, then we obtain

$$E[\tau|\mathbf{X}] = \sqrt{\frac{\psi}{\chi + q(\mathbf{X})} \frac{K_{(d+3)/2}(\sqrt{\psi(\chi + q(\mathbf{X}))})}{K_{(d+1)/2}(\sqrt{\psi(\chi + q(\mathbf{X}))})}}$$

Put $d = 2$, then we have for $\mathcal{MNIG}(\chi, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma})$:

$$\begin{aligned} E[\tau|\mathbf{X}] &= \sqrt{\frac{\psi}{\chi + q(\mathbf{X})} \frac{K_{5/2}(\sqrt{\psi(\chi + q(\mathbf{X}))})}{K_{3/2}(\sqrt{\psi(\chi + q(\mathbf{X}))})}} \\ &= \frac{3 + 3\sqrt{\psi(\chi + q(\mathbf{X}))} + \psi(\chi + q(\mathbf{X}))}{(\chi + q(\mathbf{X}))(1 + \sqrt{\psi(\chi + q(\mathbf{X}))})}, \end{aligned}$$

One can obtain the limiting case $\psi \rightarrow 0$, from (3), of the MNIG distribution with parameters $\mathcal{MNIG}(0, \psi, \boldsymbol{\mu}, \boldsymbol{\Sigma})$:

$$\lim_{\psi \rightarrow 0} \frac{3 + 3\sqrt{\psi(\chi + q(\mathbf{X}))} + \psi(\chi + q(\mathbf{X}))}{(\chi + q(\mathbf{X}))(1 + \sqrt{\psi(\chi + q(\mathbf{X}))})} = \frac{3}{\chi + q(\mathbf{X})}.$$

We conclude from the previous section that GLASSO can be applied in two steps: using the EM algorithm.

3.2 emLASSO estimator

Let us consider the model (1), assume that the distribution and the corresponding parameters of $\mathbf{X}$ are known (estimated), and the formula for the conditional expectation $\hat{\tau} = E[\tau|\mathbf{X}]$ is available. Considering first $\mathbf{X}$ like to be a d -dimensional normal random vector $\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with mean vector $\boldsymbol{\mu}$, and variance-covariance matrix $\boldsymbol{\Sigma}$. The LASSO-regularized estimate $\hat{\boldsymbol{\Theta}}$ of the concentration matrix $\boldsymbol{\Theta}$ of $\mathbf{X}$ is obtained as follows:

$$\hat{\boldsymbol{\Theta}} = \arg \max_{\boldsymbol{\Theta}} \left(\frac{n}{2} \log |\boldsymbol{\Theta}| - \frac{n}{2} \text{tr} [\boldsymbol{\Theta} \hat{\mathbf{S}}_{\mathbf{X}}] - \rho \|\boldsymbol{\Theta}\|_1 \right), \quad (4)$$

where

$$\hat{\mathbf{S}}_{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n (\mathbf{X}_i - \hat{\boldsymbol{\mu}}) (\mathbf{X}_i - \hat{\boldsymbol{\mu}})^\top,$$

$\hat{\boldsymbol{\mu}} = 1/n \sum_{i=1}^n \mathbf{X}_i$, and $\text{tr}(\cdot)$ denotes the trace of a matrix.

Take into consideration that if the weight-variate τ for a normal variance mixture model (1), is known, then the LASSO estimator is based on the "Gaussian" log-likelihood

$$\ell(\boldsymbol{\mu}, \boldsymbol{\Theta} | \mathbf{X}, \tau) \simeq \frac{1}{2} \log |\boldsymbol{\Theta}| - \tau(\mathbf{X}) q(\mathbf{X}),$$

where

$$q(\mathbf{X}) = (\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Theta} (\mathbf{X} - \boldsymbol{\mu}) = \text{tr}[\boldsymbol{\Theta} (\mathbf{X} - \boldsymbol{\mu}) (\mathbf{X} - \boldsymbol{\mu})^\top],$$

and irrelevant additive constants are omitted. Now we have the estimates

$$\hat{\boldsymbol{\mu}} = \frac{1}{\sum_{i=1}^n \hat{\tau}_i} \sum_{i=1}^n \hat{\tau}_i \mathbf{X}_i,$$

with $\hat{\tau}_i = \hat{\tau}(\mathbf{X}_i)$ for short, and

$$\hat{\mathbf{S}}_{\mathbf{X}, \tau} = \frac{1}{n} \sum_{i=1}^n \hat{\tau}_i (\mathbf{X}_i - \hat{\boldsymbol{\mu}}) (\mathbf{X}_i - \hat{\boldsymbol{\mu}})^\top.$$

Hence, the LASSO estimate of the concentration matrix $\boldsymbol{\Theta}$ becomes

$$\hat{\boldsymbol{\Theta}} = \arg \max_{\boldsymbol{\Theta}} \left(\frac{n}{2} \log |\boldsymbol{\Theta}| - \frac{n}{2} \text{tr} \left[\boldsymbol{\Theta} \hat{\mathbf{S}}_{\mathbf{X}, \tau} \right] - \rho \|\boldsymbol{\Theta}\|_1 \right),$$

where the hidden variate τ is not observed, it is predicted by the conditional expectation $\hat{\tau} = \mathbb{E}[\tau | \mathbf{X}]$ from the sample. This result allows us to use the EM algorithm.

Estimate $\boldsymbol{\Theta}_0$ by (4), like in Gaussian case.

E-step. Use the current estimate $\hat{\boldsymbol{\Theta}}_k$ to calculate $\hat{\tau}_i, i = 1 : n$ by (3), put

$$\hat{\boldsymbol{\mu}}_k = \frac{1}{\sum_{i=1}^n \hat{\tau}_i} \sum_{i=1}^n \hat{\tau}_i \mathbf{X}_i,$$

and

$$\hat{\mathbf{S}}_{\mathbf{X}, \tau} = \frac{1}{n} \sum_{i=1}^n \hat{\tau}_i (\mathbf{X}_i - \hat{\boldsymbol{\mu}}_k) (\mathbf{X}_i - \hat{\boldsymbol{\mu}}_k)^\top,$$

respectively.

M-step. Obtain the next estimate $\hat{\Theta}_{i+1}$ of the concentration matrix by

$$\hat{\Theta}_{i+1} = \arg \max_{\Theta} \left(\frac{n}{2} \log |\Theta| - \frac{n}{2} \text{tr} \left[\Theta \hat{S}_{\mathbf{X}, \tau} \right] - \rho \|\Theta\|_1 \right).$$

The iteration of these E- and M-steps will be referred to as the emLASSO algorithm.

Now, one can follow the algorithm emLASSO for estimating the zeros of the concentration matrix. The pseudocode for the algorithm is given in Appendix 2.

3.3 emLASSO_P for Multivariate Power Exponential distribution

The MPE variate $\mathbf{X}$ of dimension d is defined by the density function

$$f(\mathbf{X}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \beta) = c_{d,\beta} |\boldsymbol{\Sigma}|^{-1/2} \exp \left(-\frac{1}{2} [(\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})]^\beta \right), \quad (5)$$

where

$$c_{d,\beta} = \frac{d\Gamma(d/2)}{\pi^{d/2}\Gamma(1 + d/2\beta) 2^{1+d/2\beta}}$$

and $\beta > 0$ is the shape parameter, denote it by $\text{MPE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \beta)$, if $\beta = 1$ the model reduces to the Gaussian, and to the Laplace distribution if $\beta = 1/2$, see Gómez et al. (1998).

The distribution $\text{MPE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \beta)$ belongs to the family of a scale mixture of normal distributions if and only if $0 < \beta < 1$, Gómez-Sánchez-Manzano et al. (2008). If $\beta < 1$, it is considered a heavy-tailed, leptokurtic distribution, illustrated in the contour plot in Fig. 1. That is again, a special case of the elliptically contoured distributions, hence the Lemma 1 applies.

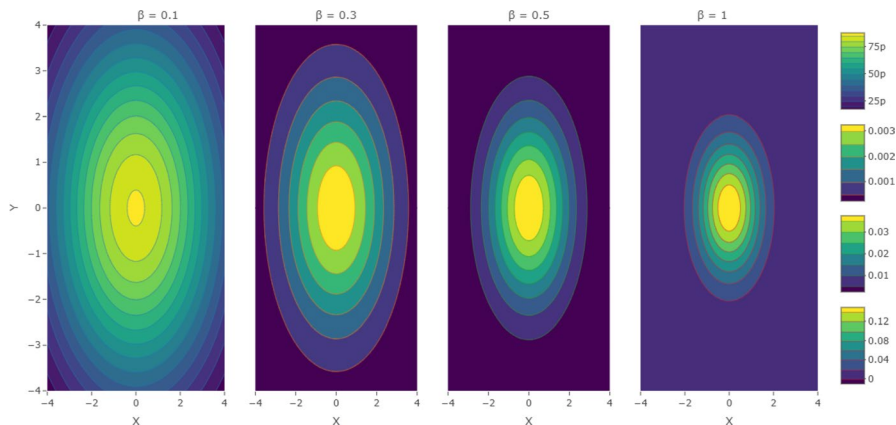


Fig. 1 Contour plot of bivariate MPE distribution with different shape parameters

First, we consider the Laplace distribution $\mathcal{MP}\mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, 1/2)$. We have the density function

$$f(\mathbf{X}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, 1/2) = c_{d,1/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2} \sqrt{q(\mathbf{X})}\right),$$

where $c_{d,1/2} = d\Gamma(d/2)/\pi^{d/2}2^{d+1}\Gamma(d+1)$. The expected value and the variance are $E[\mathbf{X}] = \boldsymbol{\mu}$ and $\text{Var}(\mathbf{X}) = 4(d+1)\boldsymbol{\Sigma}$, respectively. The stochastic representation of $\mathbf{X} \sim \mathcal{MP}\mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, 1/2)$ is

$$\mathbf{X} = \boldsymbol{\mu} + V\boldsymbol{\Sigma}^{1/2}\mathbf{Z},$$

where $\boldsymbol{\mu} \in \mathbb{R}^d$, $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{I}_d)$, $V > 0$, is independent of $\mathbf{Z}$. Here, the variate V is connected to the Lévy distribution

$$f_{V,1/2}(x) = c_{V,1/2} x^d e^{-x^2/8}, 0 < x < \infty,$$

where the constant is $c_{V,1/2} = 1/2^{(3d+1)/2}\Gamma((d+1)/2)$, see Gómez-Sánchez-Manzano et al. (2008). We apply the transformation $V = 1/\sqrt{\tau}$ with Jacobian $dx = 1/2\tau^{-3/2}d\tau$, and obtain

$$\mathbf{X} = \boldsymbol{\mu} + \frac{1}{\sqrt{\tau}}\boldsymbol{\Sigma}^{1/2}\mathbf{Z},$$

where the τ has the generalized Gamma density

$$f_{\tau,1/2}(\tau) = c_{\tau,1/2} \tau^{-d/2-3/2} e^{-1/8\tau},$$

with constant $c_{\tau,1/2} = 1/2^{3(d+1)/2}\Gamma((d+1)/2)$.

Now, the conditional distribution of $\mathbf{X}$ is $\mathbf{X}|\tau \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}/\tau)$ therefore

$$f_{\mathbf{X}|\tau}(\mathbf{X}|\tau) = \frac{\tau^{d/2}}{(2\pi)^{d/2}} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{\tau}{2} q(\mathbf{X})\right).$$

To find the conditional expectation $E[\tau|\mathbf{X}]$ we consider

$$\begin{aligned} f_{\tau|\mathbf{X}}(\tau|\mathbf{X}) &= \frac{f_{\mathbf{X}|\tau}(\mathbf{X}|\tau) f_{\tau,1/2}(\tau)}{f_{\mathbf{X}}(\mathbf{X})} \\ &= c_{\tau|\mathbf{X}} \tau^{d/2} \tau^{-(d+3)/2} e^{-1/8\tau} \exp\left(-\frac{\tau}{2} q(\mathbf{X}) + \frac{1}{2} \sqrt{q(\mathbf{X})}\right) \\ &= c_{\tau|\mathbf{X}} \tau^{-3/2} e^{\sqrt{q(\mathbf{X})}/2} \exp\left(-\frac{1}{2} \left(\tau q(\mathbf{X}) + \frac{1}{4\tau}\right)\right), \end{aligned}$$

where the normalizing constant is

$$c_{\tau|\mathbf{X}} = \frac{1}{2^{d+1/2}} \frac{\Gamma(d)}{\Gamma(d/2 + 1/2) \Gamma(d/2)}.$$

The conditional density $f_{\tau|\mathbf{X}}$ proves to be inverse Gaussian $\mathcal{IG}(\chi, \psi)$ with parameters $\chi = 1/4$, $\psi = q(\mathbf{X})$, see Paolella (2007), p. 313. Therefore, the conditional expectation follows from the expectation of the $\mathcal{IG}(\chi, \psi)$ variate Y , say, which is $E[Y] = \sqrt{\chi/\psi}$. Applying the actual parameters, we obtain

$$\hat{\tau} = E[\tau|\mathbf{X}] = \sqrt{\chi/\psi} = \frac{1}{2\sqrt{q(\mathbf{X})}}. \quad (6)$$

3.3.1 MPE distribution, general case

Now we consider the $\mathcal{MP}\mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \beta)$ distribution with $\beta \in (0, 1]$, (Chen et al. 2023). In this case, the variance mixture representation

$$\mathbf{X} = \boldsymbol{\mu} + V\boldsymbol{\Sigma}^{1/2}\mathbf{Z},$$

is given in general by Gómez-Sánchez-Manzano et al. (2008), Theorem 2.1, (2). The density function of V can not be expressed in terms of elementary functions, except when $\beta = 1/2$ in the previous section. Nevertheless, the scale mixture of normal distributions is valid if and only if $\beta \in (0, 1]$. In this way for $\beta \in (0, 1)$ we have the density of the weight-variate V by

$$f_{V,\beta}(x) = c_{V,\beta} x^{d-3} S_{\beta}(x^{-2}, 2^{1-1/\beta}), 0 < x < \infty,$$

where $S_{\beta}(x^{-2}, 2^{1-1/\beta})$ denotes the stable density of variate V and $c_{V,\beta} = 2^{(1+d/2-d/2\beta)} \Gamma(d/2 + 1) / \Gamma(d/2\beta + 1)$. The weight-variate V here is symmetric α -stable with density $S_{\beta}(u, \sigma)$, say, can be characterized by its Laplace transform $\mathcal{L}_{S_{\beta}}(u) = e^{-\sigma^{\beta} u^{\beta}}$.

Note that the parameter for the stable density belongs to the interval $(0, 2]$ and here we restrict it to the interval $(0, 1)$. We apply the transformation $V = 1/\sqrt{\tau}$, where the Jacobian is $dx = 1/2\tau^{-3/2}d\tau$, therefore the density of τ is

$$f_{\tau,\beta}(\tau) = c_{\tau,\beta} \tau^{-d/2} S_{\beta}(\tau, 2^{1-1/\beta}),$$

where $c_{\tau,\beta} = 2^{d/2(1-\beta)} \Gamma(d/2 + 1) / \Gamma(d/2\beta + 1)$. Then we use the density function $f(\mathbf{X}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \beta)$ of $\mathbf{X}$, see (5) and the conditional density of $\mathbf{X}$

$$f_{\mathbf{X}|\tau}(\mathbf{X}|\tau) = \frac{\tau^{d/2}}{(2\pi)^{d/2}} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{\tau}{2} q(\mathbf{X})\right),$$

when τ is given, to obtain the conditional density of τ when $\mathbf{X}$ is given, as usual. Indeed, we have

$$\begin{aligned} f_{\tau|\mathbf{X}}(\tau|\mathbf{X}) &= \frac{f_{\mathbf{X}|\tau}(\mathbf{X}|\tau) f_{\tau,\beta}(\tau)}{f_{\mathbf{X}}(\mathbf{X})} \\ &= c_{\tau|\mathbf{X}} \exp\left(-\frac{\tau}{2}q(\mathbf{X}) + \frac{1}{2}q(\mathbf{X})^\beta\right) S_\beta\left(\tau, 2^{1-1/\beta}\right) \\ &= c_{\tau|\mathbf{X}} e^{q(\mathbf{X})^\beta/2} \exp\left(-\frac{\tau}{2}q(\mathbf{X})\right) S_\beta\left(\tau, 2^{1-1/\beta}\right). \end{aligned}$$

Here, one may deduce that $c_{\tau|\mathbf{X}} = 1$.

Now the expected value $E[\tau|\mathbf{X}]$ follows by the integral

$$E[\tau|\mathbf{X}] = e^{q(\mathbf{X})^\beta/2} \int_0^\infty \exp\left(-\frac{\tau}{2}q(\mathbf{X})\right) \tau S_\beta\left(\tau, 2^{1-1/\beta}\right) d\tau.$$

Put $u = q(\mathbf{X})/2$ and notice that this integral here is connected to the Laplace Transform of the density $S_\beta(\tau, 2^{1-1/\beta})$, which is

$$\begin{aligned} \mathcal{L}_{S_\beta}(u) &= \int_0^\infty \exp(-\tau u) S_\beta\left(\tau, 2^{1-1/\beta}\right) d\tau \\ &= e^{-2^{\beta(1-1/\beta)}u^\beta} = e^{-2^{\beta-1}u^\beta}, u > 0, \end{aligned}$$

. The expected value according to the density $S_\beta(\tau, 2^{1-1/\beta})$ is the derivative

$$-\frac{d}{du} \mathcal{L}_{S_\beta}(u) = 2^{\beta-1} \beta u^{\beta-1} e^{-2^{\beta-1}u^\beta},$$

see Feller (1966), p. 435. Exactly, what we need to get $E[\tau|\mathbf{X}]$ is to put up $u = q(\mathbf{X})/2$ back into the derivative and obtaining

$$\begin{aligned} E[\tau|\mathbf{X}] &= e^{q(\mathbf{X})^\beta/2} 2^{\beta-1} \beta \left(\frac{2}{q(\mathbf{X})}\right)^{1-\beta} e^{-2^{\beta-1}(q(\mathbf{X})/2)^\beta} \\ &= \beta q(\mathbf{X})^{\beta-1}. \end{aligned}$$

This result provides the best prediction of the weight variate τ when the observation $\mathbf{X}$ is given.

Lemma 3 Let the variate $\mathbf{X} \sim \mathcal{MPE}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \beta)$ with $\beta \in (0, 1]$, then the predictor $\hat{\tau}$ is given by

$$\hat{\tau} = E[\tau|\mathbf{X}] = \beta q(\mathbf{X})^{\beta-1}.$$

Remark 3 If $\beta = 1/2$ then

$$\hat{\tau} = \frac{1}{2\sqrt{q(\mathbf{X})}}, \quad (7)$$

which equals to (6), see above.

Remark 4 The EM algorithm can be highly generalized. One can implement it directly not only for the MPE and MSGH models using the predictor $\hat{\tau}$ for τ (i.e., for hidden parameters), but also for other models, as it is outlined in the Algorithm 1 in Appendix 2.

While robust graphical models based on the multivariate t-distribution are well established, their heavy tails can induce high estimator variance, potentially affecting the stability of the recovered sparsity pattern. The proposed emLASSO framework addresses this limitation by employing the MSGH and MPE families, which allow more flexible control of tail behavior. Similar motivations for using hyperbolic-type distributions—a subclass of the MSGH family—have recently been explored in Bayesian regression (Kawakami and Hashimoto 2023; Hu and Zhang 2024). Given the computational efficiency of the proposed EM-type algorithms, a natural extension is to develop a fully Bayesian graphical modeling framework. Such an approach could incorporate hierarchical priors to provide uncertainty quantification while retaining robust edge-selection properties.

3.4 Estimation of nuisance parameters

In the context of the emLASSO framework, the estimation of the concentration matrix Θ implicitly depends on additional distribution-specific parameters, which we refer to as nuisance distributional parameters. These include the parameters (λ, χ, ψ) for the MSGH family and the shape parameter β for the MPE family. Since these parameters directly influence the conditional expectation of the latent weights w_i in (3) and (7), their treatment is an essential component of the methodology. In this study, we adopted a two-step estimation approach, in which nuisance parameters are estimated separately from the concentration matrix. Specifically:

- **Step 1 (Nuisance parameter estimation):** The distributional parameters η which represents the parameters (λ, χ, ψ) in the case of MSGH and λ for MPE, are estimated using marginal maximum likelihood based on the observed data, independently of the sparsity constraint on Θ . Denote the resulting estimator by

$$\hat{\eta} = \arg \max_{\eta} \sum_{i=1}^n \log f(\mathbf{X}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \eta)$$

- **Step 2 (Sparse concentration estimation):** Conditional on $\hat{\eta}$, the emLASSO algorithm is applied to estimate Θ , where the latent weights are computed as

$$\hat{\tau}_i^{(k+1)} = \mathbb{E} \left[\tau_i | \mathbf{X}_i; \hat{\eta}, \boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)} \right]$$

This separation ensures computational tractability and avoids the need for high-dimensional joint optimization over both η and Θ .

Therefore, the EM algorithm is applied only to the latent scale variables τ , while the nuisance parameters η are treated as fixed after their initial estimation. In practice, for the MPE family, β is estimated via maximum likelihood using the `normalp` package in R. Similarly, for the MSGH family, the `fit.ghypuv` function from the `ghyp` package is used to estimate the shape parameter λ .

In addition, to distinguish between idealized and data-driven performance, we consider two scenarios:

- **Oracle setting:** nuisance parameters are fixed at their true values.
- **Practical setting:** nuisance parameters are estimated from the data as described above.

3.5 Theoretical considerations for the estimator of Θ

While the proposed emLASSO extends the GLASSO framework to non-Gaussian settings, its theoretical properties require careful discussion due to the presence of the latent variable τ and nuisance distributional parameters η .

Connection to GLASSO:

A key observation is that, conditional on the latent weights $\{\tau_i\}_{i=1}^n$, the model reduces to a weighted Gaussian likelihood. In particular, the M-step of the EM algorithm solves a penalized likelihood problem of the form (4), which is equivalent to

the classic GLASSO problem with a reweighted empirical covariance matrix $\hat{\mathbf{S}}_{\mathbf{X}, \tau}$. Therefore, conditional on consistent estimation of the weights, the estimator inherits the structural properties of the classical GLASSO estimator, including sparsity and positive definiteness.

Consistency under known nuisance parameters:

Consider first the oracle setting in which the nuisance parameters η are known. In this case, the latent weights satisfy

$$\hat{\tau}_i = \mathbb{E} \left[\tau_i | \mathbf{X}_i; \eta, \boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)} \right],$$

and the estimator reduces to a weighted GLASSO with the convergence rate of the classical GLASSO as

$$\|\hat{\Theta} - \Theta\|_F = O_p \left(\sqrt{\frac{s \log d}{n}} \right)$$

where s denotes the number of nonzero entries in Θ . Similar guarantees can be established for element-wise consistency and support recovery. Check (Ravikumar et al. 2011; Mazumder and Hastie 2012) regarding the consistency of the ℓ_1 -penalized estimator.

Effect of nuisance parameter estimation:

When estimating the nuisance parameters η from the data, it yields the weights

$$\hat{\tau}_i = \mathbb{E} \left[\tau_i | \mathbf{X}_i; \hat{\eta}, \boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)} \right],$$

In this case, the estimation error can be decomposed as

$$\left\| \hat{\Theta} - \Theta \right\|_F \leq \underbrace{\left\| \hat{\Theta} - \Theta(\hat{\eta}) \right\|_F}_{\text{GLASSO error}} + \underbrace{\left\| \hat{\Theta}(\hat{\eta}) - \Theta(\eta) \right\|_F}_{\text{nuisance error}}$$

Therefore, if $\hat{\eta}$ is a consistent estimator of η , and the mapping from η to τ is sufficiently smooth, the second term is asymptotically negligible. Consequently, the estimator $\hat{\Theta}$ retains the same convergence rate as in the oracle case.

4 Simulated results and real example

4.1 Simulation results

A preconfigured simulation has been used to evaluate the performance of the proposed method against competing methods across different environments. We benchmarked the proposed emLASSO, GLASSO, γ -LASSO, and SGLASSO using the Receiver Operating Characteristic (ROC) curve to assess the accuracy of graph structure recovery. The proposed approach is computationally more demanding than GLASSO because GLASSO optimization is performed at each M-step of the EM algorithm. The simulation results were evaluated using ROC curves for all comparisons. For SGLASSO, the tuning parameter was set to $\lambda = 10^{-5}$ following (Lu and Feng 2025), while for γ -LASSO, we used $\gamma = 0.05$ as suggested by Onizuka and Hashimoto (2025). All methods were implemented using *glasso* or *glassoFast* R packages. A cutoff threshold of 10^{-3} was applied to define the true graph structure. To ensure stable results, the simulation was repeated 200 times, and the average ROC curve was computed. In addition, under the settings, when $d > n$, to address computational issues, the approximation was set to true in the *glasso* function of the R package.

Common methods for selecting the GLASSO penalty parameter ρ include K-fold cross-validation, Information criteria (AIC/BIC), extended BIC (EBIC), and the stability approach to regularization selection (StARS) (Yamada et al. 2025; Chen et al. 2023). We adopted StARS for our simulations as it selects the tuning parameters based on edge stability across subsamples. This approach yields robust, reproducible,

and interpretable networks and avoids the pitfalls of likelihood-based methods by capturing the genuine graph structure of the relationships.

4.1.1 Case 1: $\mathcal{MSGH}$, $d = 50$; sparsity = 30%; $n = 100$

True sparse concentration matrices of size 50×50 with density 30% were generated using the algorithm described in Algorithm 1. The algorithm enforces strict diagonal dominance to ensure positive definiteness. In R, a sparse adjacency matrix was generated using the *rsparsematrix* function from the **Matrix** package, and the corresponding precision matrix was obtained via the *SimulatePrecision* function from the **fake** package. Under the settings of the $\mathcal{MSGH}$ distribution, datasets were simulated with sample sizes $n = 50$ and $n = 100$ from $\mathcal{MSGH}_d(\lambda, \chi = 0.5, \psi = 0.5, \mu = \mathbf{0}, \Sigma = \Theta^{-1})$, where $\lambda \in (2, 5, 10, 15)$.

As a result, Fig. 2 highlights the performance of the GLASSO, emLASSO_H , γ -LASSO, and SGLASSO estimators. Based on the provided ROC curves, the performance of the estimators varies substantially with λ . For $\lambda = 2$, all five estimators—GLASSO, emLASSO_H , γ -LASSO, and SGLASSO—exhibit tightly clustered performance with comparable recovery of the true sparsity pattern. However, as λ increases to 5, 10, and 15, a distinct separation emerges where the emLASSO_H esti-

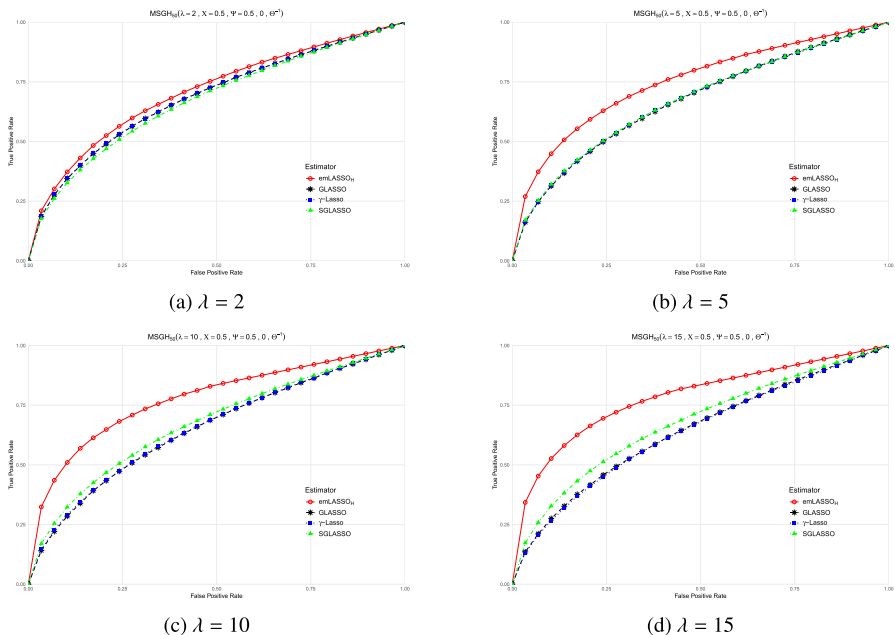


Fig. 2 ROC curves comparing the performance of the GLASSO, emLASSO_H , γ -LASSO, and SGLASSO estimators in recovering the true sparsity pattern of the concentration matrix. The true conditional uncorrelated graphs are generated from the $\mathcal{MSGH}_{50}(\lambda, \chi = 0.5, \psi = 0.5, \mu = \mathbf{0}, \Sigma = \Theta^{-1})$. Where $\lambda \in (2, 5, 10, 15)$

mator consistently demonstrates superior performance, achieving the highest True Positive Rate for a given False Positive Rate. The SGLASSO estimator generally ranks second, outperforming the remaining methods but falling below emLASSO_H in these higher λ settings. Meanwhile, the GLASSO and γ -LASSO estimators exhibit nearly identical lower-tier performance profiles at higher λ values, clustering together at the bottom of the ROC space.

4.1.2 Case 2: $\mathcal{MSGH}$, $d = 100$; sparsity = 30%; $n = 50$

In the high-dimensional settings $d > n$ under the $\mathcal{MSGH}$ distribution, the emLASSO_H estimator consistently exhibits superior performance in recovering the true sparsity pattern. While all estimators display comparable efficacy at the lowest λ parameter ($\lambda = 2$), a clear performance hierarchy emerges as λ increases to 5, 10, and 15. In these heavier-tailed scenarios, emLASSO_H dominates the ROC space, achieving the highest True Positive Rate for any given False Positive Rate. The SGLASSO estimator consistently ranks second, outperforming the remaining methods. Meanwhile, the GLASSO and γ -LASSO estimators cluster together with notably lower sensitivity, indicating that they are less robust to the heavy tails introduced by the higher λ values in this regime, as highlighted in Fig. 3.

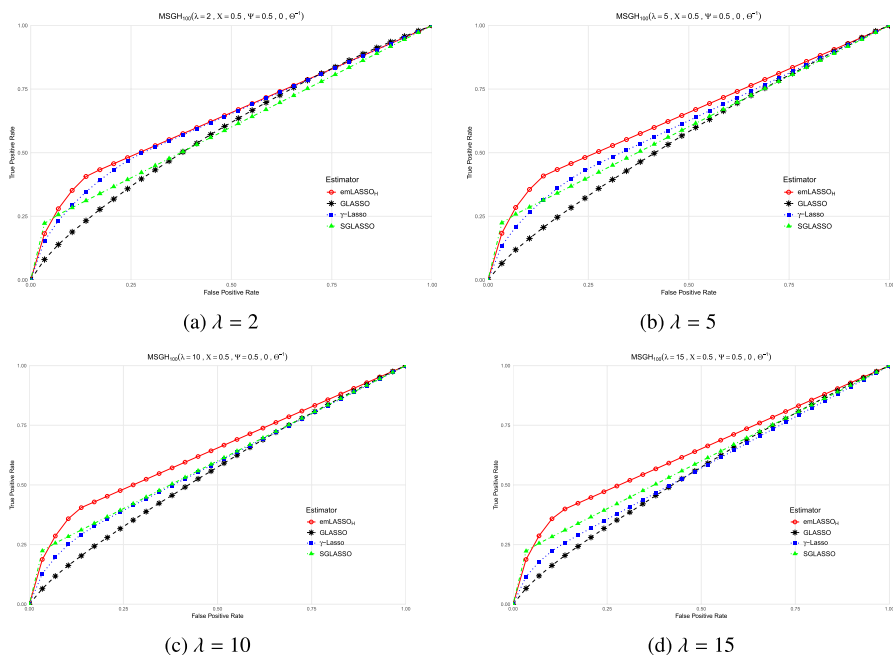


Fig. 3 ROC curves comparing the performance of the GLASSO, emLASSO_H , γ -LASSO, and SGLASSO estimators in recovering the true sparsity pattern of the concentration matrix. The true conditional uncorrelated graphs are generated from the $\mathcal{MSGH}_{100}(\lambda, \chi = 0.5, \psi = 0.5, \mu = \mathbf{0}, \Sigma = \Theta^{-1})$.

Where $\lambda \in (2, 5, 10, 15)$ and sample size $n = 50$

4.1.3 Case 3: $\mathcal{MP}\mathcal{E}$, $d = 50$; sparsity = 30%; $n = 100$

In this case, datasets were simulated from the MPE distribution $\mathcal{MP}\mathcal{E}_d(\mathbf{0}, \Theta^{-1}, \beta)$, across different sample sizes ($n = 50$ and 100). To evaluate the impact of the shape parameter on the performance of each method, we examined varying values of $\beta \in \{1/3, 1/2, 2/3, 3/4, 1\}$. After the simulation, Fig. 4 shows different behaviors of the GLASSO, emLASSO_P, γ -LASSO, and SGLASSO estimators, when the real concentration matrix is generated from the $\mathcal{MP}\mathcal{E}_{50}(\mathbf{0}, \Theta^{-1}, \beta)$ distribution. The proposed method outperforms the GLASSO approach when the shape parameter is small; however, its performance deteriorates as β approaches 1 (in the Gaussian case), at which point it performs comparably to GLASSO. For small to moderate values of β , emLASSO_P outperforms the competing methods. The results depicted by Fig. 4 reveal that the emLASSO_P estimator demonstrates a clear and consistent superiority in recovering the true sparsity pattern across all considered values of the shape parameter β . Regardless of whether the distribution is more heavy-tailed ($\beta = 1/3$) or approaches a lighter-tailed shape ($\beta = 3/4$), emLASSO_P consistently yields the highest True Positive Rate (TPR) for any given False Positive Rate (FPR), effectively dominating the ROC space. The SGLASSO estimator generally achieves

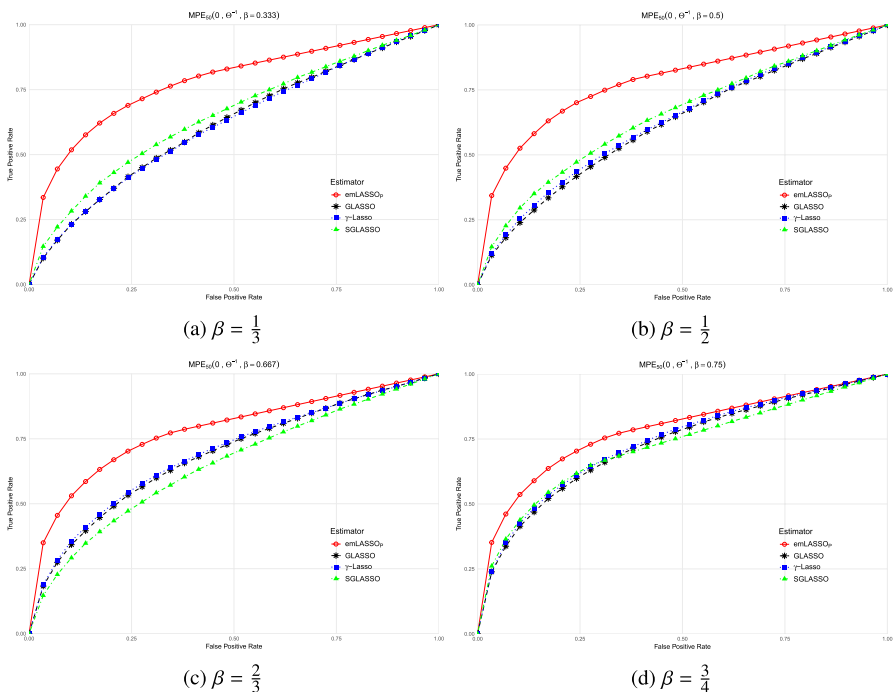


Fig. 4 ROC curves comparing the performance of the GLASSO, emLASSO_P, γ -LASSO, and SGLASSO estimators in recovering the true sparsity pattern of the concentration matrix. The true conditional uncorrelated graphs are generated from the $\mathcal{MP}\mathcal{E}_{50}(\mathbf{0}, \Theta^{-1}, \beta)$ distribution, and different values of the shape parameter (β)

second-best performance, showing a distinct advantage over standard methods but remaining below the emLASSO_P curve in all scenarios. In contrast, the GLASSO and γ -LASSO estimators exhibit the lowest performance, clustering together with significantly lower sensitivity. This hierarchy remains stable across all variations of β , underscoring the robustness of emLASSO_P to the specific tail behavior of the multivariate power exponential distribution.

4.1.4 Case 4: $\mathcal{MPE}$, $d = 100$; sparsity= 30%; $n = 50$

In the case of a high-dimensional context as $d > n$, Fig. 5 shows the ROC curves with a high dimension $d = 100$ and a limited sample size ($n = 50$). The emLASSO_P estimator demonstrates a commanding performance advantage over competing methods across all considered shape parameters ($\beta \in \{1/3, 1/2, 2/3, 3/4\}$). As illustrated in the ROC curves, emLASSO_P consistently achieves the highest True Positive Rate for low False Positive Rates, exhibiting significant robustness to the heavy-tailed nature of the MPE distribution. The SGLASSO estimator consistently ranks second, clearly separating itself from the remaining estimators. Conversely, the GLASSO and γ -LASSO estimators show substantially degraded performance in this "large d , small n " regime, clustering together with much lower sensitivity compared to the top-performing emLASSO_P .

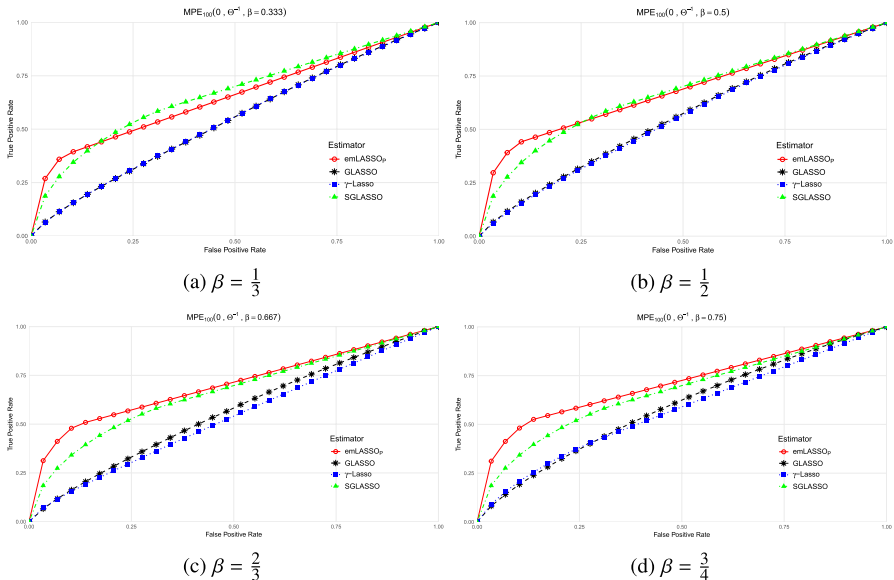


Fig. 5 ROC curves comparing the performance of the GLASSO, emLASSO_P , γ -LASSO, and SGLASSO estimators in recovering the true sparsity pattern of the concentration matrix. The true conditional uncorrelated graphs are generated from the $\mathcal{MPE}_{100}(\mathbf{0}, \Theta^{-1}, \beta)$ distribution, and different values of the shape parameter (β) and a small sample size $n = 50$

4.1.5 Case 5: Performance evaluation in contaminated 80/20 Gaussian and $MPE/MSGH$ scenarios

In the contaminated setting, the data comprise 80% Gaussian observations and 20% non-Gaussian observations (from the MPE or MSGH families), with a sparsity level of 30%. We assess the performance and robustness of the proposed $emLASSO_H$ and $emLASSO_P$ estimators.

Performance under MSGH contamination:

When Gaussian data are contaminated with MSGH samples, $emLASSO_H$ exhibits strong robustness. Figures 6 and 7 indicate that:

- **Recovery:** For both $d = 50$ and $d = 100$, $emLASSO_H$ consistently attains the highest true positive rate for any given false positive rate.
 - **Sensitivity:** As the tail-weight parameter λ increases from 2 to 10, $emLASSO_H$ maintains a clear advantage over SGLASSO and GLASSO.
 - **Stability:** The method remains effective in high-dimensional regimes ($d > n$), continuing to dominate the ROC space for $d = 100$.
- Performance under MPE contamination:*

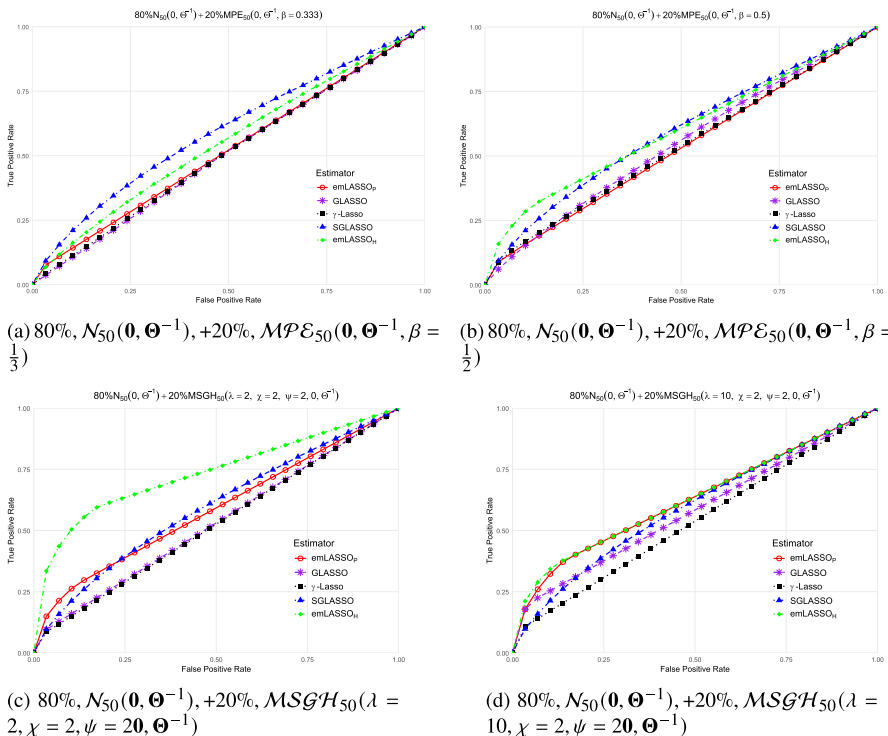


Fig. 6 ROC curves comparing the performance of the GLASSO, $emLASSO_H$, γ -LASSO, and SGLASSO estimators in recovering the true sparsity pattern of the concentration matrix. The true conditional uncorrelated graphs are generated from the contaminated Normal distribution as $80\%, N_{50}(0, \Theta^{-1})$, 20% MPE/MSGH, with $n = 100$ and $d = 50$

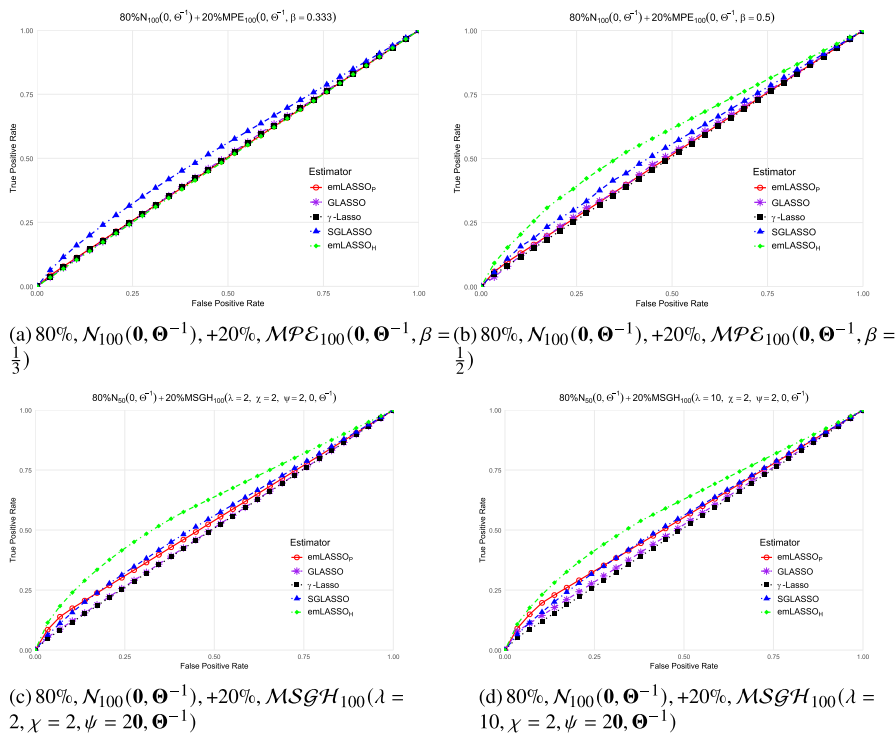


Fig. 7 ROC curves comparing the performance of the GLASSO, emLASSO_H, γ -LASSO, and SGLASSO estimators in recovering the true sparsity pattern of the concentration matrix. The true conditional uncorrelated graphs are generated from the contaminated Normal distribution as 80%, $N_{50}(\mathbf{0}, \Theta^{-1})$, 20% MPE and MSGH, with $n = 100$ and $d = 50$

When the contamination follows an MPE distribution, the relative performance of the estimators varies with the shape parameter β . Specifically:

- **Robustness:** For $\beta = 0.5$, emLASSO_H achieves the best performance, followed by SGLASSO.
- **Heavy-tailed regimes:** Under stronger contamination ($\beta = 0.333$), SGLASSO exhibits increased robustness, often matching or slightly outperforming emLASSO_H in recovering sparsity.
- **emLASSO_P:** In the 80/20 mixture setting, emLASSO_P performs comparably to GLASSO and γ -LASSO, with lower sensitivity to the underlying sparsity structure than emLASSO_H.

GLASSO and γ -LASSO show substantial performance degradation under heavy-tailed contamination, whereas emLASSO_H remains comparatively robust, particularly in MSGH-based mixtures. This behavior suggests that the EM-based weighting scheme in emLASSO_H effectively downweights contaminated observations, thereby improving the accuracy of concentration matrix estimation.

4.1.6 Can MSGH yield the same results as MPE?

Preliminary descriptive analysis and measures of multivariate kurtosis can guide the choice between the MPE and MSGH models. In particular, Mardia's multivariate kurtosis provides a useful diagnostic for assessing tail behavior relative to the Gaussian benchmark [See Terdik (2022)]. Here are the highlights of their preferences:

- **Tail Flexibility Comparison:** MSGH is exceptionally powerful for **heavy-tailed (leptokurtic)** data. It can model more complex "tail-decay" patterns than MPE. If your data is strictly heavy-tailed with outliers, emLASSO_H will often produce results that are equal to or better than emLASSO_P . MPE is unique because it can also model thin-tailed (**platykurtic**) data ($\beta > 1$). MSGH is generally restricted to the leptokurtic domain. Therefore, if your dataset contains "flat" distributions, MSGH cannot replicate the MPE results.
- **The "Heavy-Tailed" Convergence:** In cases of moderate heavy-tailedness, the two methods often converge on the same sparsity pattern. As seen in the ROC results for MSGH contamination, emLASSO_H and emLASSO_P both significantly outperform standard GLASSO, often showing overlapping performance curves when the contamination is not extreme.
- **Parameter Sensitivity:** MSGH requires tuning multiple nuisance parameters (λ, χ, ψ). If these are tuned perfectly, MSGH can "mimic" the robust behavior of an MPE model for heavy-tailed data. However, the MPE's single β parameter makes it easier to achieve these results reliably without additional uncertainty.

4.2 Estimating the conditional correlation graph for Body Dimension data

To demonstrate the practical relevance of the proposed method, we consider the Body Dimensions dataset introduced by Heinz et al. (2003). The dataset includes various anthropometric measurements—such as height, weight, limb lengths, and body circumferences—collected from adult men and women. The adopted dataset comprised 507 individuals (247 males and 260 females), with slight differences between the groups, as indicated by selected variables in Fig. 8. Because these measurements are inherently correlated, the dataset provides an appropriate setting for examining a conditional uncorrelated graph among body dimensions. Figure 9 illustrates the conditional independence of the nine skeletal breadth variables—Biacromial Diameter, Bicep Girth, Biiliac Diameter, Chest Depth, Chest Diameter, Elbow Diameter, Wrist Diameter, Knee Diameter, and Ankle Diameter—and stratified by gender (Male = M, Female = F).

To assess the joint distribution of the 24 attributes, we first computed descriptive statistics, as shown in Table 1. We then conducted marginal normality tests for each variable, and the results indicate that most variables deviate from normality. In addition, a principal component analysis (PCA) of the data reveals that the Q-Q plot of the first principal component scores deviates substantially from the Gaussian assumption, further suggesting non-normality in this high-dimensional setting. The results align with the preliminary analysis of Chen et al. (2023). In addition, to further assess the multivariate normality of the dataset, we applied Mardia's test of

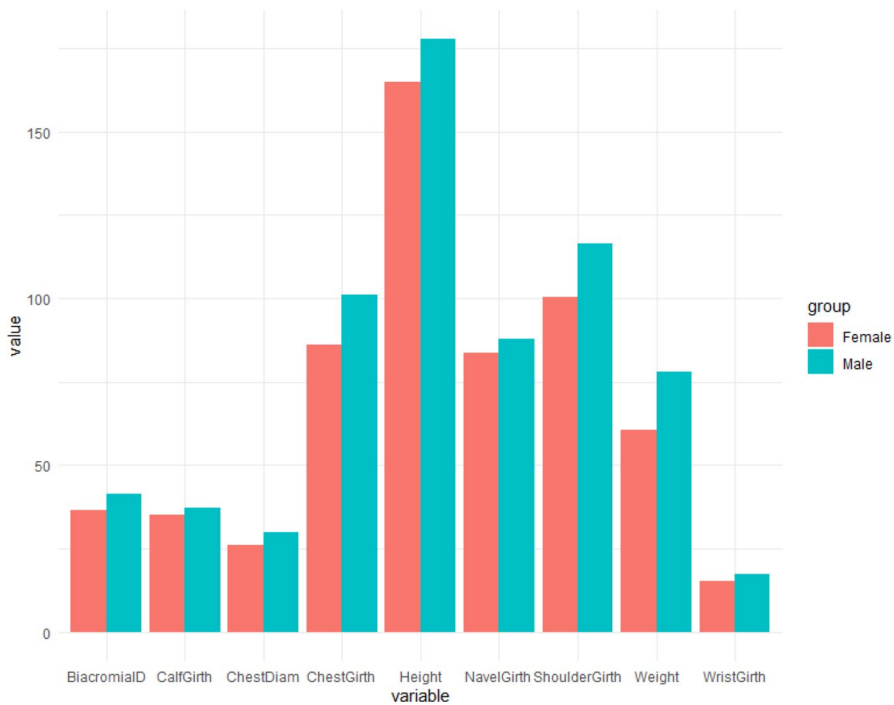


Fig. 8 Bar plot comparing the average levels of some measurements between men and women. The chart highlights the magnitude and direction of sex-based differences across all measured parameters

kurtosis. The results yielded a multivariate kurtosis statistic of 732.90 with a p-value of 1.97×10^{-70} . This extreme departure from normality ($p < 0.001$) confirms the presence of significant heavy-tailedness, necessitating the use of robust estimation frameworks, such as the proposed emLASSO methods. Hence, these findings motivate the consideration of more flexible distributional families, such as the MPE and MSGH models.

While both the MSGH and MPE distributions provide robust alternatives to the Gaussian framework, we prioritized the MPE distribution for the primary analysis of the dataset. This choice is primarily driven by the structural parsimony of the MPE family and its unique flexibility in modeling diverse tail behaviors. Specifically, the MPE shape parameter β allows the model to accommodate both leptokurtic (heavy-tailed) and platykurtic (thin-tailed) variables within a single framework. Moreover, the descriptive analysis of the current dataset in Table 1 revealed a significant mixture of positive and negative excess kurtosis across variables, as confirmed by Mardia's multivariate test. The MPE distribution offers a more computationally efficient and theoretically consistent fit. This parsimonious approach reduces the risk of overparameterization associated with the more complex MSGH family while maintaining high sensitivity in recovering the underlying sparse concentration structure.

Following the approach described in Sect. 3.4, the nuisance parameters can be estimated before fitting the graphical model. In this context, Chen et al. report estimated shape parameters of $\beta = 0.603$ for males and $\beta = 0.622$ for females, support-

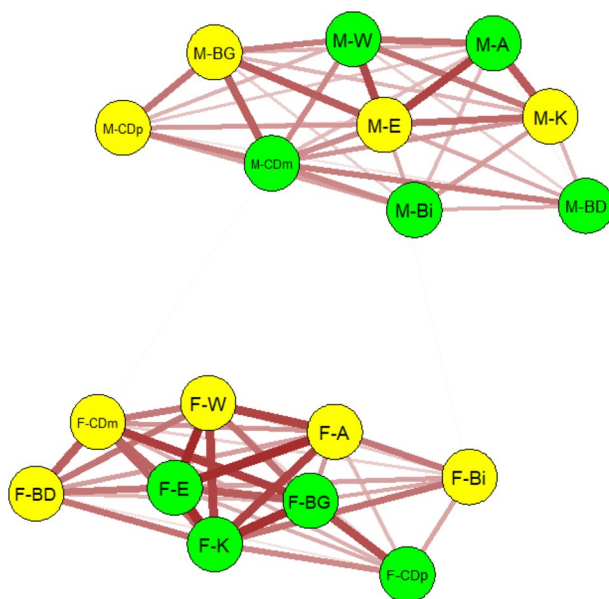


Fig. 9 Partial correlations of the nine Skeletal Breadth variables, presented separately for males (M) and females (F), indicating that they are conditionally uncorrelated between groups. The variables are: Biacromial Diameter, Bicep Girth, Biiliac Diameter, Chest Depth, Chest Diameter, Elbow Diameter, Wrist Diameter, Knee Diameter, and Ankle Diameter

ing the suitability of the MPE model for these data. Accordingly, the shape parameter β can be estimated via maximum likelihood.

In this case study, we estimate the sparse concentration matrix under the MPE assumption using the proposed *emLASSO* approach. The resulting estimate is used to construct a network graph that visualizes the conditional relationships among the variables. This analysis demonstrates the proposed method's ability to capture complex dependence structures in real-world multivariate data.

The conditional, uncorrelated graph estimated using *emLASSO*, as shown in Fig. 10, revealed clear interrelationships among male body dimensions. Strong links were observed among upper-body variables—shoulder, chest, and torso measures—indicating coordinated variation in thoracic size. The waist, hip, and thigh girths formed another cohesive cluster, representing lower-trunk proportionality. Weight served as a central node connecting several regions, particularly the chest and waist, whereas height showed inverse associations with girth measures, reflecting proportional differences in body build. Peripheral measures, such as wrist and knee diameters, showed limited connections, consistent with their anatomical uncorrelation. Overall, the model captured distinct conditional correlations, providing a coherent representation of the male body morphology.

Similarly, the *emLASSO*-estimated conditional uncorrelated graph for the female group (Fig. 11) clearly depicts conditional correlations among body measurements. Compared with Gaussian-based models, it produces a sparser, more interpretable structure that emphasizes biologically meaningful links while reducing the influence

Table 1 Descriptive statistics of the body dimension Measurements

	Mean	SD	Median	Trimmed	Min	Max	Skew	Kurtosis
BiacromialD	38.81	3.06	38.70	38.74	32.40	47.40	0.16	-0.84
BiiliacD	27.83	2.21	28.00	27.89	18.70	34.70	-0.42	1.07
BitrocD	31.98	2.03	32.00	32.00	24.70	38.00	-0.09	0.02
ChestDepth	19.23	2.52	19.00	19.09	14.30	27.50	0.49	-0.14
ChestDiam	27.97	2.74	27.80	27.89	22.20	35.60	0.26	-0.74
ElbowD	13.39	1.35	13.30	13.38	9.90	16.70	0.05	-0.74
WristD	10.54	0.94	10.50	10.54	8.10	13.30	0.05	-0.49
KneeD	18.81	1.35	18.70	18.77	15.70	24.30	0.34	0.07
AnkleD	13.86	1.25	13.80	13.85	9.90	17.20	0.07	-0.38
ShoulderGirth	108.20	10.37	108.20	107.99	85.90	134.80	0.14	-0.90
ChestGirth	93.33	10.03	91.60	93.04	72.60	118.70	0.24	-0.89
WaistGirth	76.98	11.01	75.80	76.27	57.90	113.20	0.54	-0.23
NavelGirth	85.65	9.42	84.60	85.21	64.00	121.10	0.45	0.13
HipGirth	96.68	6.68	96.00	96.41	78.80	128.30	0.50	0.76
ThighGirth	56.86	4.46	56.30	56.56	46.30	75.70	0.69	0.72
BicepGirth	31.17	4.25	31.00	31.06	22.40	42.40	0.22	-0.83
ForearmGirth	25.94	2.83	25.80	25.88	19.60	32.50	0.15	-0.98
KneeGirth	36.20	2.62	36.00	36.12	29.00	49.00	0.47	1.00
CalfGirth	36.08	2.85	36.00	36.02	28.40	47.70	0.28	0.29
AnkleGirth	22.16	1.86	22.00	22.08	16.40	29.30	0.40	0.36
WristGirth	16.10	1.38	16.10	16.07	13.00	19.60	0.15	-0.73
Age	30.18	9.61	27.00	28.99	18.00	67.00	1.13	0.94
Weight	69.15	13.35	68.20	68.58	42.00	116.40	0.40	-0.36
Height	171.14	9.41	170.30	170.93	147.20	198.10	0.15	-0.46
Gender	0.49	0.50	0.00	0.48	0.00	1.00	0.05	-2.00

of extreme values. Upper-body measures—shoulder, chest, and chest diameter—form a cohesive subnetwork reflecting coordinated variation in torso morphology, whereas abdominal and lower-body variables (waist, navel, hip, and thigh girths) cluster tightly, indicating anatomical consistency. Height and weight are inversely associated with several girths, suggesting that increases in stature and mass do not affect all body regions uniformly. Overall, the conditional-uncorrelated graph reveals modular organization across anatomical areas and demonstrates the effectiveness of MPE-based emLASSO in capturing stable, physiologically coherent relationships.

4.3 Computational costs

The computational complexity of the proposed emLASSO procedures is theoretically summarized in Table 2 and can be compared with that of the classical GLASSO estimator. The GLASSO algorithm involves a single penalized-likelihood optimization with a computational cost of $\mathcal{O}(d^3)$. In contrast, the proposed methods rely on an EM-type iterative scheme. For the Student- t and MPE models, the E-step is computationally inexpensive, and the dominant cost within each iteration arises from the computation of the weighted sample covariance matrix and the subsequent GLASSO update, yielding a per-iteration complexity of order $\mathcal{O}(nd^2 + d^3)$. The overall computational cost is therefore $\mathcal{O}(T(nd^2 + d^3))$, where T denotes the number of EM

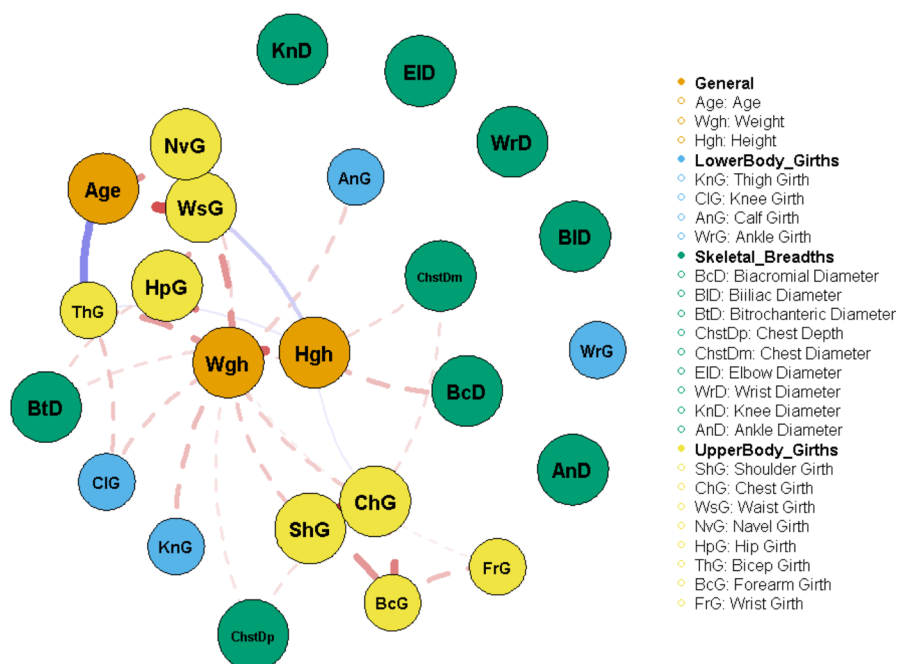
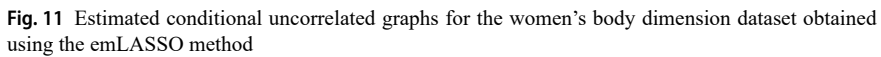


Fig. 10 Estimated conditional uncorrelated graphs for the male's body dimension dataset obtained using the emLASSO_P method

iterations. For the MSGH model, the evaluation of modified Bessel functions introduces an additional computational component, resulting in a per-iteration cost of order $\mathcal{O}(nd^2C_{\text{Bessel}} + d^3)$. As a consequence, this variant is computationally more demanding than the Student- t and MPE cases. Overall, the additional cost of the proposed framework is primarily driven by the EM iterations and the complexity of the underlying distributional model.

The empirical benchmarking results in Table 3 indicate that emLASSO_P strikes the most favorable balance between computational speed and memory overhead among the proposed variants. Although GLASSO remains the most efficient (0.61ms), emLASSO_P (4.8ms) remains feasible for high-dimensional applications. In contrast, emLASSO_H exhibits significant memory churn (24.41MB/itr), which is reflected in the high frequency of garbage collection events (gc/sec=8.26).

Therefore, these theoretical findings in Table 2 are consistent with the empirical computation times reported in Table 3, where the Student- t and MPE variants exhibit moderate overhead relative to GLASSO. At the same time, the MSGH-based approach is comparatively slower.



Method	Model	E-step Cost	Per-Iteration	Total Complexity	Speed
GLASSO	Gaussian	–	$\mathcal{O}(d^3)$	$\mathcal{O}(d^3)$	Fastest
emLASSO _T	Student t	$\mathcal{O}(1)$	$\mathcal{O}(nd^2 + d^3)$	$\mathcal{O}(T(nd^2 + d^3))$	Fast
emLASSO _P	MPE	$\mathcal{O}(1)$ (power)	$\mathcal{O}(nd^2 + d^3)$	$\mathcal{O}(T(nd^2 + d^3))$	Fast
emLASSO _H	MSGH	$\mathcal{O}(C_{\text{Bessel}})$	$\mathcal{O}(nd^2 C_{\text{Bessel}} + d^3)$	$\mathcal{O}(T(nd^2 C_{\text{Bessel}} + d^3))$	Slow

Table 3 Empirical Computational Performance Comparison of emLASSO variants and GLASSO with $d = 50, n = 100$ and iterations of 1000 times.

Method	Median time	Relative speed	Memory Alloc.	Garbage collection/s
GLASSO	0.61 ms	$1.0\times$	0.25 MB	6.43
emLASSO _P	4.80 ms	$7.8\times$	2.13 MB	8.28
emLASSO _t	12.90 ms	$21.1\times$	6.45 MB	10.10
emLASSO _H	94.60 ms	$154.4\times$	24.41 MB	8.26

5 Conclusion

In this study, we showed that zero partial correlation implies zero conditional correlation for general elliptically symmetric variates. We further construct conditional, uncorrelated graphs for the MSGH and MPE distributions and propose the EM-based LASSO (emLASSO) method to recover the true concentration matrix under these distributions. The proposed approach extends penalized likelihood estimation by incorporating the MPE shape parameter, thereby providing greater flexibility in modeling both light- and heavy-tailed data. The proposed methods have been assessed against standard benchmarks, including GLASSO, γ -LASSO, and SGLASSO.

Our results demonstrate that the proposed estimators, emLASSO_P and emLASSO_H, consistently outperform competing methods in recovering the true graph structure, particularly in heavy-tailed and high-dimensional environments ($d > n$). The emLASSO_P estimator exhibited remarkable robustness across varying values of the shape parameters, dominating the ROC curve even as the distribution has a severe heavy-tail. Similarly, emLASSO_H proved superior in the MSGH settings, establishing a clear performance hierarchy as the tail-weight parameter λ increased.

Compared with SGLASSO, a competitive second-best alternative, the standard GLASSO and its variants showed significant performance degradation under heavy-tailed distributions. Although the proposed approach incurs a higher computational cost due to the EM algorithm's iterative nature, this trade-off yields substantial gains in accuracy and stability. Ultimately, this study confirms that the emLASSO framework is a reliable, robust tool for graphical modeling of complex, non-Gaussian, high-dimensional data.

Appendix 1. Proof of Lemma 1

Proof Let us assume that

$$E(Y_p | Y_q, Y_{\setminus(p,q)}) = E(Y_p | Y_{\setminus(p,q)}). \quad (8)$$

then $Y_p \perp Y_q | Y_{\setminus(p,q)}$, i.e. Y_p and Y_q are conditional uncorrelated. Now the (8) implies that

$$\begin{aligned} E(Y_p | Y_{\setminus(p,q)}) E(Y_q | Y_{\setminus(p,q)}) &= E(Y_q E(Y_p | Y_{\setminus(p,q)}) | Y_{\setminus(p,q)}) \\ &= E(Y_q E(Y_p | Y_q, Y_{\setminus(p,q)}) | Y_{\setminus(p,q)}) \\ &= E(E(Y_p Y_q | Y_q, Y_{\setminus(p,q)}) | Y_{\setminus(p,q)}) \\ &= E(Y_p Y_q | Y_{\setminus(p,q)}). \end{aligned}$$

The conditional covariance is written as

$$C_{Y_p, Y_q | Y_{\setminus(p,q)}} = E(Y_p Y_q | Y_{\setminus(p,q)}) - E(Y_p | Y_{\setminus(p,q)}) E(Y_q | Y_{\setminus(p,q)}),$$

hence

$$\text{Cov}(Y_p, Y_q | Y_{\setminus(p,q)}) = 0,$$

follows from (8). Hence, we need to show that the condition (8), i.e.

$$\hat{Y}_{p|\bullet}[Y_q, Y_{\setminus(p,q)}] = \hat{Y}_{p|\bullet} Y_{\setminus(p,q)},$$

fulfills provided $C_{Y_p, Y_q | \bullet} Y_{\setminus(p,q)} = 0$. Now the LLSP $\hat{Y}_{p|\bullet}[Y_q, Y_{\setminus(p,q)}]$ can be resolved

$$\hat{Y}_{p|\bullet}[Y_q, Y_{\setminus(p,q)}] = \hat{Y}_{p|\bullet} Y_{\setminus(p,q)} + \hat{Y}_{p|\bullet}(Y_q - \hat{Y}_{q|\bullet} Y_{\setminus(p,q)}),$$

(see Whittaker 2009 p.137, Additivity of LLSP). Therefore, in one way

$$\hat{Y}_{p|\bullet}[Y_q, Y_{\setminus(p,q)}] = \hat{Y}_{p|\bullet}[Y_q, Y_{\setminus(p,q)}] - \hat{Y}_{p|\bullet} Y_{\setminus(p,q)}$$

and

$$\hat{Y}_{p|\bullet}(Y_q - \hat{Y}_{q|\bullet} Y_{\setminus(p,q)}) = C_{Y_p, Y_q | \bullet} Y_{\setminus(p,q)} \Sigma_{Y_q | \bullet}^{-1} Y_{\setminus(p,q)} (Y_q - \hat{Y}_{q|\bullet} Y_{\setminus(p,q)}),$$

in the other way. Now if $C_{Y_p, Y_q | \bullet} Y_{\setminus(p,q)} = 0$, then

$$0 = \hat{Y}_{p|\bullet}[Y_q, Y_{\setminus(p,q)}] - \hat{Y}_{p|\bullet} Y_{\setminus(p,q)} = E(Y_p | Y_q, Y_{\setminus(p,q)}) - E(Y_p | Y_{\setminus(p,q)}).$$

Therefore if $C_{Y_p, Y_q | \bullet} Y_{\setminus(p,q)} = 0$, i.e. the partial correlation is zero then the assumption (8) is fulfilled and $\hat{Y}_p \perp Y_q | Y_{\setminus(p,q)}$. $\square$

Appendix 2. EM algorithm for estimating the emLASSO

Input:

Data: $\mathbf{X}_{1:n} \in \mathbb{R}^{n \times d}$

Function of the weight variate $\widehat{\tau}(\mathbf{x}) = \mathbb{E}[\tau|\mathbf{X}]$ be given

Regularization parameter: $\rho > 0$

Convergence threshold: ϵ

Output: μ, Θ

Initialization:

Initialize $\mu^{(0)}, \Theta^{(0)}, t = 0$

repeat

Expectation step:

for $i = 1$ to n do

Compute Mahalanobis distance:

$$q^{(t)}(\mathbf{x}_i) = \left(\mathbf{x}_i - \mu^{(t)} \right)^\top \Theta^{(t)} \left(\mathbf{x}_i - \mu^{(t)} \right)$$

Compute Weight:

$$\widehat{\tau}^{(t+1)}(\mathbf{x}_i)$$

end for

Maximization step:

Update the weighted mean vector:

$$\mu^{(t+1)} = \frac{\sum_{i=1}^n \widehat{\tau}^{(t+1)}(\mathbf{x}_i) \mathbf{x}_i}{\sum_{i=1}^n \widehat{\tau}^{(t+1)}(\mathbf{x}_i)}$$

Update the weighted empirical covariance matrix:

$$\mathbf{S}^{(t+1)} = \frac{1}{n} \sum_{i=1}^n \tau^{(t+1)} \left(\mathbf{x}_i - \mu^{(t+1)} \right) \left(\mathbf{x}_i - \mu^{(t+1)} \right)^\top$$

Maximize the concentration matrix with ℓ_1 penalty:

$$\Theta^{(t+1)} = \arg \max_{\Theta > 0} \left(\log |\Theta| - \text{tr} \left(\mathbf{S}^{(t+1)} \Theta \right) - \rho \|\Theta\|_1 \right)$$

Convergence Check:

if $\|\Theta^{(t+1)} - \Theta^{(t)}\|_F < \epsilon$ then

Break

end if

Increment $t = t + 1$

until convergence

Return: $\mu = \mu^{(t+1)}, \Theta = \Theta^{(t+1)}$

Algorithm 1 Implementation of the EM Algorithm for estimating the sparse concentration matrix of MSGH and MPE Distributions.

Acknowledgements The authors thank the reviewers for their careful reading of the manuscript and for their constructive and insightful comments, which have substantially improved the quality and clarity of this work.

Author contributions All authors contributed to this work equally.

Funding Open access funding provided by University of Debrecen. Not applicable.

Data availability Publicly available data sets were analyzed in this study. These data can be found here: <https://www.tandfonline.com/doi/figure/10.1080/10691898.2003.11910711?scroll=top&needAccess=true>. The package is published in <https://github.com/AbdoDawod/emLASSO>.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Browne RP, McNicholas PD (2015) A mixture of generalized hyperbolic distributions. *Can J Stat* 43(2):176–198
- Bh M (1998) Propriétés asymptotiques d'un estimateur bi-spectral pour un processus p-adique strictement homogène. *C R Acad Sci Paris* 326(1):987–990
- Chen Y, Wiesel A, Hero AO (2011) Robust shrinkage estimation of high-dimensional covariance matrices. *IEEE Trans Signal Process* 59(9):4097–4107
- Chen X, Feng Z, Peng H (2023) Estimation and order selection for multivariate exponential power mixture models. *J Multivar Anal* 195:105140. <https://doi.org/10.1016/j.jmva.2022.105140>
- Erickson S, Rydén T (2025) Sparse estimation of inverse covariance and partial correlation matrices via joint partial regression. <https://doi.org/10.48550/arXiv.2502.08414>
- Fang KW, Kotz S, Ng KW (2017) Symmetric multivariate and related distributions. Chapman and Hall/CRC
- Feller W (1966) An introduction to probability theory and its application, vol II. John Wiley, New York
- Finegold M, Drton M (2014) Robust graphical modeling with t-distributions. <https://doi.org/10.48550/ARXIV.1408.2033>
- Franzack BC, Browne RP, McNicholas PD (2013) Mixtures of shifted asymmetric Laplace distributions. *IEEE Trans Pattern Anal Mach Intell* 36(6):1149–1157
- Friedman J, Hastie T, Tibshirani R (2008) Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9(3):432–441
- Gómez E, Gómez-Villegas M, Marín JM (1998) A multivariate generalization of the power exponential family of distributions. *Commun Stat-Theory Methods* 27(3):589–600
- Gómez-Sánchez-Manzano E, Gómez-Villegas M, Marín J (2008) Multivariate exponential power distributions as mixtures of normal distributions with Bayesian applications. *Commun Stat-Theory Methods* 37(6):972–985
- Heinz G, Peterson LJ, Johnson RW et al (2003) Exploring relationships in body dimensions. *J Stat Educ*. <https://doi.org/10.1080/10691898.2003.11910711>
- Hirose K, Fujisawa H, Sese J (2017) Robust sparse Gaussian graphical modeling. *J Multivar Anal* 161:172–190. <https://doi.org/10.1016/j.jmva.2017.07.012>
- Huang S, Jin J, Yao Z (2016) Partial correlation screening for estimating large precision matrices, with applications to classification. *Ann Stat*, 44 (5), pp 2018–2057

- Hu W, Zhang W (2024) Flexible Bayesian quantile regression based on the generalized asymmetric huberised-type distribution. *Stat Comput* 34(4):144. <https://doi.org/10.1007/s11222-024-10459-z>
- Jankov J, van de Geer S (2015) Confidence intervals for high-dimensional inverse covariance estimation. *Electron J Stat* 9(1):1205–1229. <https://doi.org/10.1214/15-EJS1031>
- Kawakami J, Hashimoto S (2023) Approximate Gibbs sampler for Bayesian huberized lasso. *J Stat Comput Simul* 93(1):128–162. <https://doi.org/10.1080/00949655.2022.2104332>
- Kelker D (1970) Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhy Indian J Stat Ser A*, 32(4), pp 419–430 <https://www.jstor.org/stable/25049690>
- Lu Z, Feng L (2025) Robust sparse precision matrix estimation and its application. [arXiv:2503.03575](https://arxiv.org/abs/2503.03575)
- Marzban Vasselabadi N, Tahmasebi S, Mohammadi R et al (2025) Robust Bayesian structure learning for graphical models with T-distributions using G-Wishart prior. *Comput Stat* 40(8):4277–4305. <https://doi.org/10.1007/s00180-025-01621-6>
- Mazumder R, Hastie T (2012) The graphical lasso: New insights and alternatives. *Electron J Stat* 6:2125–2149. <https://doi.org/10.1214/12-EJS740>
- McNeil AJ, Frey R, Embrechts P (2005) Quantitative risk management, Princeton Series in Finance. Princeton University Press, Princeton (**concepts, techniques and tools**)
- Muirhead RJ (2009) Aspects of multivariate statistical theory, vol 197. John Wiley & Sons
- Olver FWJ, Olde Daalhuis AB, Lozier DW, Schneider BI, Boisvert RF, Clark CW, Miller BR, Saunders BV, Cohl HS, McClain MA (eds) (2020) NIST Digital Library of Mathematical Functions. Release 1.0.26 of 2020-03-15. <http://dlmf.nist.gov/>
- Onizuka T, Hashimoto S (2025) Robust Bayesian graphical modeling using γ -divergence. *J Multivar Anal*, 209, p 105461
- Paoletta MS (2007) Intermediate probability: a computational approach. Wiley, <https://books.google.hu/books?id=9SHARfvyiR4C>
- Prause K (1999) The generalized hyperbolic model: estimation, financial derivatives, and risk measures. PhD thesis, PhD thesis, University of Freiburg
- Ravikumar P, Wainwright MJ, Raskutti G et al (2011) High-dimensional covariance estimation by minimizing L1-penalized log-determinant divergence. *Electron J Stat* 5(none):935–980. <https://doi.org/10.1214/11-EJS631>
- Riccobello R, Bonaccolto G, Kremer PJ et al (2025) Sparse graphical modelling for global minimum variance portfolio. *CMS* 22(2):8. <https://doi.org/10.1007/s10287-025-00535-4>
- Rossell D, Zwiernik P (2021) Dependence in elliptical partial correlation graphs. *Electron J Stat* 15(2):4236–4263. <https://doi.org/10.1214/21-EJS1891>
- Terdik GY (2022) Multivariate statistical methods; going beyond the linear. In: *Frontiers in probability and the statistical sciences*. Springer International Publishing, Cham. <https://doi.org/10.1007/978-3-030-81392-5>
- Terdik G (2026) On Fisher information for multivariate elliptically contoured distributions. *Stat Probab Lett* 234:110690. <https://doi.org/10.1016/j.spl.2026.110690>
- Weibel M, Luethi D, Breymann HE (2024) ghyp: Generalized Hyperbolic Distribution and Its Special Cases. <https://doi.org/10.32614/cran.package.ghyp>, r package version 1.6.5
- Whittaker J (2009) Graphical models in applied multivariate statistics. Wiley Publishing
- Yamada T, Sakurai T, Fujikoshi Y (2025) Variable selection method based on BIC with consistency for non-zero partial correlations under a large-dimensional setting. *Comput Stat* 40(7):3585–3611. <https://doi.org/10.1007/s00180-025-01628-z>
- Yang E, Lozano AC (2015) Robust gaussian graphical modeling with the trimmed graphical lasso. *Advances in neural information processing systems* 28
- Zhao T, Roeder K, Liu H (2014) Positive semidefinite rank-based correlation matrix estimation with application to semiparametric graph estimation. *J Comput Graph Stat* 23(4):895–922. <https://doi.org/10.1080/10618600.2013.858633>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.