Joint learning of Gaussian graphical models in heterogeneous dependencies of high-dimensional transcriptomic data

Dung Ngoc Nguyen NGOCDUNG.NGUYEN@CSIRO.AU and **Zitong Li** ZITONG.LI@CSIRO.AU
CSIRO Agriculture and Food, Canberra, ACT, Australia.

Editors: Vu Nguyen and Hsuan-Tien Lin

Abstract

In biology, constructing gene co-expression networks presents a significant research challenge, largely due to the high dimensionality of the data and the heterogeneity of the samples. Furthermore, observations from two or more groups sharing the same biological variables require the comparison of gene co-expression patterns with some commonalities between the groups. In this context, we propose a mixture of Gaussian graphical models for paired data to estimate heterogeneous dependencies and uncover sub-population networks within complex biological datasets, incorporating sparsity and symmetry constraints between two groups of dependent variables. We develop an efficient generalized expectation-maximization (EM) algorithm for penalized maximum likelihood estimation with the fusion of a graphical lasso penalty. As a result, our simulation studies highlight the numerical performance of the proposed method, demonstrating its superior model fitting compared to the classical graphical lasso approach. We further demonstrate the practical application of our approach by estimating gene networks on a high-dimensional ecological transcriptomics data set of the nine-spined stickleback. Our new approach identified similarities and differences between groups of genes from the brain and liver tissues of samples collected from two habitats. These results show the efficiency of our approach to the identification of complicated interactions from high-dimensional and heterogeneous gene expression data.

Keywords: Mixture Gaussian graphical models; Paired data; Penalized maximum likelihood; EM algorithm; Unsupervised machine learning; Bioinformatics.

1. Introduction

With recent advances in the development of cost-effective high-throughput RNA sequencing technology (Stark et al. (2019)), gene co-expression network analysis (D'haeseleer et al. (2000); López-Kleine et al. (2013); Rao and Dixon (2019)) has become increasingly popular for studying the complex interactions between genes, proteins, and regulatory elements, identifying functionally related groups of genes and how they contribute to the expression of desirable traits, and understanding the biological factors underlying phenotypic diversity. The Gaussian graphical model (GGM), introduced by Dempster (1972), is one of the most commonly and widely applied tools for analyzing biological networks. This is a family of multivariate Gaussian models with restriction of the conditional independence of selected pairs of variables, given the others, in terms of an undirected graph. Each vertex in the graph represents a variable, and the absence of an edge implies that the corresponding entry in the concentration matrix, i.e. the inverse of the covariance matrix, is equal to zero, see Lauritzen (1996). Inference from these models, when applied to large-scale molecular biology experiments, enables us to account for the correlation of marker effects in predicting

gene functions, phenotypes, and molecular regulation patterns. See Ursem et al. (2008), Valcarcel et al. (2011), Ma et al. (2007), Chang and McGeachie (2011), Kurtz et al. (2015), Zheng et al. (2020) and references therein.

In biological applications, it is common for observed data to originate from various sources and exhibit heterogeneous dependencies across the entire population. Additionally, gene expression data are often collected from different treatments or across different tissues, cells, or phenotypes, generating interest in comparing gene co-expression patterns under distinct experimental conditions or between groups. In this paper, we focus on the joint learning of a mixture of Gaussian graphical models for paired data in heterogeneous populations, with an emphasis on high-dimensional scenarios. Different sub-populations or classes are modeled by distinct networks. For every sub-population, we assume the presence of exactly two dependent groups of homologous variables, i.e. representing two experimental conditions, with the association structure of each group captured by a corresponding sub-network. The two sub-networks within each sub-population are interconnected, with edges linking vertices both within and between the sub-networks. Similarities between groups are represented using graph coloring (Højsgaard and Lauritzen (2008)), where vertex coloring represents equality constraints on the diagonal entries of the concentration matrix and edge coloring represents equality constraints on the off-diagonal entries. We refer to this class of models as *a mixture of graphical models for paired data with restrictions on concentrations* or *mixture of pdRCON models* for short.

2. Related works and Contribution

Heterogeneity is a common characteristic in biological studies, where samples are often measured at different locations or originate from distinct populations or families. In the relevant literature, the mixture of graphical models has proven useful for uncovering genomic variations in high-throughput sequencing data that cannot be adequately captured with a single distribution (Liang and Jia (2023)). Recent studies employing this type of graphical model for analyzing omics data in quantitative genomics include Blein-Nicolas et al. (2024), Danaher et al. (2014). The former study extended the novel block-diagonal covariance for locally linear Gaussian mapping and applied this model to predict drought-related traits from protein abundance in maize. Whereas the latter introduced the joint graphical lasso to simultaneously construct multiple graphical models for distinct but related conditions, analyzing lung cancer microarray data. For further application, see Lotsi and Wit (2016), Lee and Xue (2018), Lartigue et al. (2021), and the references therein.

Moreover, comparing the distribution of a set of variables between two experimental conditions or groups is a key focus in numerous applications. When the association structure represented by a GGM is of interest, analyzing paired data can be framed as the joint learning of a graph structure of each group, with particular attention to the cross-sectional association structure between the two graphs. Joint learning of dependent GGMs is commonly applied in genomics to compare co-expression patterns between healthy and cancer tissues, as reflected in the transcriptional networks, see Hardcastle and Kelly (2013); Danaher et al. (2014); Aran et al. (2017). Another example involves the comparison of brain networks derived from fMRI data, which often exhibit naturally symmetrical structures between the two hemispheres. In this context, Roverato and Nguyen (2022, 2024) explored

the search space of pdRCON models using an efficient backward elimination procedure to better understand the structure of this model class. While this approach offered a fast model selection method, model identification remains challenging due to the high dimensionality and complexity of the model space. Alternatively, penalized maximum likelihood methods have been proposed to address high dimensionality by avoiding explicit exploration of the model space, potentially yielding solutions closer to the global optimum compared to the backward selection. Recent works of Ranciati et al. (2021) and Ranciati and Roverato (2023) introduced the graphical lasso for paired data (pdglasso) which extends the symmetric graphical lasso method for the class of GGMs to analyze paired data. These studies also developed an alternating directions method of multiplier (ADMM) algorithm to solve the pdglasso optimization tasks.

We develop a novel penalized expectation-maximization (EM) algorithm that simultaneously clusters individuals and infers the graph structure by adapting the original pdglasso methods, initially introduced by Ranciati et al. (2021); Ranciati and Roverato (2023), to each sub-population. Specifically, we use the graphical lasso to induce sparsity within groups and the fused graphical lasso to enforce graph symmetries between groups of variables. Regularization parameters are selected based on the asymptotic consistency of the extended Bayesian information criterion (eBIC) (Foygel and Drton (2010)), adapted for GGMs in scenarios where both the sample size and the number of variables are comparable. The efficiency of our approach is demonstrated by constructing gene networks using both synthetic data sets and real-world ecological transcriptome datasets from nine-spined sticklebacks, including a comparison of our method with the classical graphical lasso on synthetic data. Our method provides a robust tool for reconstructing gene co-expression networks, exploring similar expression patterns within and between condition groups, and clustering individuals based on shared biological characteristics. The potential applications of this approach extend beyond the gene co-expression network, including its use in constructing other high-dimensional biological networks such as microbial interaction networks (Faust (2021)), brain connectivity networks (Bullmore and Bassett (2011)), as well as in disease-gene association prediction (Miller and Bishop (2021)).

The rest of the paper is organized as follows. Section 3 provides an overview of the mixture of GGMs for addressing heterogeneity, along with penalized maximum likelihood estimations using graphical lasso penalty. In the same section, we introduce the family of mixtures of RCON models for paired data, incorporating a fused lasso penalty that enables simultaneously learning of both the network structure and similarity between two groups of variables. Section 4 describes a penalized EM algorithm for model estimation, with its application to both synthetic and real-world data presented in Section 5. Finally, Section 6 offers a brief discussion and concluding remarks. Technical details for EM algorithm and additional results from the numerical experiments are provided in the Supplementary.

3. The models

This section focuses on classes of Gaussian mixture graphical models, particularly those designed to address the paired data problems. It also covers the fusion of the graphical lasso applied to each sub-population in heterogeneous and high-dimensional datasets.

3.1. Mixture of Gaussian graphical models

Let $\mathbf{Y} = (Y_1, \ldots, Y_P)$ be a vector of continuous random variables indexed by $V = \{1, \ldots, P\}$ with the observation $\mathbf{y} \in \mathbb{R}^P$. In heterogeneous populations, observations are assumed to originate from one of K different network models. We define $\mathbf{Z} = (Z_1, \ldots, Z_K)$ as the vector of binary latent variables, where $Z_k = 1$ indicates that the observation $\mathbf{y}$ belongs to the k-th class. We model each class separately by assuming a Gaussian graphical model (GGM) for $\mathbf{Y}$, where $(\mathbf{Y} \mid Z_k = 1) \sim \mathcal{N}(0, \Theta_k^{-1})$, with the concentration matrix Θ_k corresponding to the undirected graph G_k. Specifically, each missing edge in the graph implies that the corresponding entry of the concentration matrix is zero. In GGMs, the zero pattern of the concentration matrix reflects the conditional independence between two corresponding variables in the joint distribution. Our interest lies in the structure of Θ. Therefore, without loss of generality, we assume throughout the paper that the random variables $\mathbf{Y}$ have zero means. The GGM of $\mathbf{Y} \mid \mathbf{Z}$ with respect to G_k can be expressed as

$$p(\mathbf{y} \mid Z_k = 1, \Theta_k) = (2\pi)^{-P/2} \det(\Theta_k)^{1/2} \exp\Big(- \frac{\mathbf{y}^T \Theta_k \mathbf{y}}{2} \Big), \tag{1}$$

where Θ_k is a positive-definite concentration matrix restricted on graph G_k. By marginalizing equations (1) according to the latent variable $\mathbf{Z}$, the density of $\mathbf{Y}$ is then specified as the weighted multivariate Gaussian graphical models, which is

$$p(\mathbf{y} \mid \mathbf{w}, \boldsymbol{\Theta}) = \sum_{k=1}^{K} w_k p(\mathbf{y} \mid Z_k = 1, \Theta_k) \tag{2}$$

with the parameter vectors $\mathbf{w} = (w_1, \ldots, w_K)$ and $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_K)$, where for $k = 1, \ldots, K$, the probability $\mathbb{P}(Z_k = 1) = w_k$ represents the mixture proportion, subject to $\sum_{k=1}^{K} w_k = 1$, and $p(\mathbf{y} \mid Z_k = 1, \Theta_k)$ refers to a GGM defined in (1) with respect to G_k.

For a sample of independent and identically distributed observations $\mathbf{y}_1, \ldots, \mathbf{y}_N$ and the allocation values $\mathbf{Z}_n = (Z_{n1}, \ldots, Z_{nK})$ associated with the observation $\mathbf{y}_n$, the maximum likelihood estimations (MLE) of $(\mathbf{w}, \boldsymbol{\Theta})$ are the values that maximize the log-likelihood function

$$l(\mathbf{w}, \boldsymbol{\Theta}) = \sum_{n=1}^{N} \log \Bigg\{ \sum_{k=1}^{K} w_k p(\mathbf{y}_n \mid Z_{nk} = 1, \Theta_k) \Bigg\}. \tag{3}$$

To handle high-dimensional settings, which are particularly commom in genomics, graphical lasso (glasso) (Yuan and Lin (2007), Friedman et al. (2008)) has been widely used to estimate precision matrices by incorporating a lasso penalty term into the likelihood function, producing sparse solutions. Specifically, in Gaussian mixture graphical models, sparse estimators of $\boldsymbol{\Theta}$ can be obtained by minimising the penalized log-likelihood function

$$l_{\lambda_1, \ldots, \lambda_K}(\mathbf{w}, \boldsymbol{\Theta}) = -l(\mathbf{w}, \boldsymbol{\Theta}) + \sum_{k=1}^{K} \lambda_k \|\Theta_k\|_1, \tag{4}$$

where $l(\cdot)$ is the log-likelihood function defined in (3), and $\|\cdot\|_1$ denotes the l_1-norm, which is the sum of the absolute values of the matrix entries. Here, the regularization parameters

$\lambda_1, \ldots, \lambda_K$ are non-negative and control the level of penalization for each sub-population. For every class k, as λ_k increase, the off-diagonal entries of the concentration matrix are shrunk towards zero. This allows the graphical lasso to perform the estimation and model selection simultaneously within the GGM framework. For various applications of the glasso and its variant in heterogeneous data, see Zhou et al. (2009); Lotsi and Wit (2016); Lartigue et al. (2021).

3.2. Mixture of RCON models for paired data

Paired data. In a paired data problem, the variables on every statistical unit are measured twice from two different conditions, e.g. across different tissues or treatments. Therefore, the random vector $\mathbf{Y}$ is partitioned into two sets of homologous variables $\mathbf{Y} = (\mathbf{Y}_L, \mathbf{Y}_R)$ so that every variable $Y_i \in \mathbf{Y}_L$ corresponds to a homologous variable $Y_j \in \mathbf{Y}_R$. Accordingly, the concentration matrix Θ is naturally divided into blocks such that

$$\Theta = \begin{pmatrix} \Theta^{LL} & \Theta^{LR} \\ \Theta^{RL} & \Theta^{RR} \end{pmatrix}.$$

The interest is in explicitly studying symmetries between and across the two sub-networks in the form of identities of concentrations in Θ^{LL} with the corresponding concentration in Θ^{RR} and identities of concentrations in Θ^{LR} with the corresponding concentrations in Θ^{RL}. Symmetries can be presented by graph colorings.

RCON models for paired data (pdRCON). Roverato and Nguyen (2022) approached the paired data problems by introducing the class of colored graphs for paired data (pdCGs) denoted by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Each vertex of the graph presents a random variable and the associated graph can be split into two sub-networks corresponding to the vertex sets, called $L = \{1, \ldots, Q\}$ and $R = \{1', \ldots, Q'\}$, with $Q = P/2$ and $i' = i + Q$ for $i \in L$, so that $V = L \cup R$ and $L \cap R = \emptyset$. The vertex coloring $\mathcal{V} = \{V_1, \ldots, V_v\}$ is a partition of V with specific types of color classes that is either twin-pairing $\{i, i'\}$ or atomic $\{i\}$, and the edge coloring $\mathcal{E} = \{E_1, \ldots, E_e\}$ is a partition of the edge set E into edge color classes that is twin-pairing $\{(i, j), (i', j')\}$ between groups or $\{(i, j'), (i', j)\}$ across groups, or atomic class with single edge element. In the graphical representation, if two homologous vertices or edges belong to a twin-pairing class, they are depicted in the same color. For vertices and edges of the atomic classes, they are all depicted in black.

RCON models for paired data (pdRCONs) are GGMs with additional equality constraints of the concentrations restricted by a pdCG $\mathcal{G}$. In particular, the vertex class $\{i, i'\}$ implies the equality of diagonal entries $\theta_{ii} = \theta_{i'i'}$ and the edge classes $\{(i, j), (i', j')\}$ and $\{(i, j'), (i', j)\}$ imply the equality of off-diagonal entries $\theta_{ij} = \theta_{i'j'}$ and $\theta_{ij'} = \theta_{i'j}$, respectively. For vertices and edges belonging to the atomic color classes, there are no equality restrictions of the associated parameters in the model. In this way, equality constraints reveal symmetries concerning both the structure of the network and the values of the parameters associated with vertices and edges and also have the practical advantage of reducing the number of parameters, see Roverato and Nguyen (2022, 2024) for more information.

Mixture of pdRCON models. With K classes in the heterogeneous data, we denote $\mathbf{\Theta}_{\mathcal{G}} = (\Theta_{\mathcal{G}_1}, \ldots, \Theta_{\mathcal{G}_K})$ the concentration matrices restricted on pdCGs $\mathcal{G} = (\mathcal{G}_1, \ldots, \mathcal{G}_K)$.

The density of $\mathbf{Y}$ is then specified as a mixture of weighted pdRCON models, which is

$$p(\mathbf{y} \mid \mathbf{w}, \boldsymbol{\Theta}_{\mathcal{G}}) = \sum_{k=1}^{K} w_k p(\mathbf{y} \mid Z_k = 1, \Theta_{\mathcal{G}_k}). \tag{5}$$

To learn both sparsity in the graph structures and similarities between two groups in the heterogeneous data, we apply the *fused lasso* to every sub-population model. The estimators of $(\mathbf{w}, \boldsymbol{\Theta}_{\mathcal{G}})$ are then obtained by

$$(\widehat{\mathbf{w}}, \widehat{\boldsymbol{\Theta}}_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2}) = \underset{\mathbf{w},\boldsymbol{\Theta}_{\mathcal{G}}}{\operatorname{argmin}} -\frac{1}{N}\sum_{n=1}^{N} \log\Big\{\sum_{k=1}^{K} w_k p(\mathbf{y}_n \mid Z_{nk} = 1, \Theta_{\mathcal{G}_k})\Big\} + \operatorname{pen}_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2}(\boldsymbol{\Theta}_{\mathcal{G}}) \tag{6}$$

where $p(\mathbf{y}_n \mid Z_{nk} = 1, \Theta_{\mathcal{G}_k})$ specified in (1) is a GGM with respect to pdCG $\mathcal{G}_k$ and the penalty function

$$\operatorname{pen}_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2}(\boldsymbol{\Theta}_{\mathcal{G}}) = \sum_{k=1}^{K} \lambda_k^{[1]} \|\Theta_{\mathcal{G}_k}\|_1 + \sum_{k=1}^{K} \lambda_k^{[2]} \|\Theta_{\mathcal{G}_k}^{LL} - \Theta_{\mathcal{G}_k}^{RR}\|_1 + \sum_{k=1}^{K} \lambda_k^{[2]} \|\Theta_{\mathcal{G}_k}^{LR} - \Theta_{\mathcal{G}_k}^{RL}\|_1, \tag{7}$$

with $\boldsymbol{\lambda}_1 = \big(\lambda_1^{[1]}, \ldots, \lambda_K^{[1]}\big)$, $\boldsymbol{\lambda}_2 = \big(\lambda_1^{[2]}, \ldots, \lambda_K^{[2]}\big)$ denoting the non-negative regularization parameter vectors. The first term of (7) encourages sparsity in the graph structure to each class k controlled by $\lambda_k^{[1]}$, and the last two terms encourage the identities of $\widehat{\Theta}_k^{LL}$ and $\widehat{\Theta}_k^{RR}$ between groups and the identities of $\widehat{\Theta}_k^{LR}$ and $\widehat{\Theta}_k^{RL}$ across groups controlled by $\lambda_k^{[2]}$. Here, we do not introduce any penalty to $\mathbf{w}$ given the fact that its dimension is unlikely to be high in most of the biological applications. Because (6) is a non-convex problem and it is difficult to obtain MLE in a direct way, we develop a penalized expectation-maximization (EM) algorithm to find maximum likelihood estimates for models with latent variables.

4. Penalized EM algorithm for the mixture of pdRCON models

In fact, if we know the variable $\mathbf{Z}$ we can simply derive the estimations through the samples of $\mathbf{Y}$ such that $(\mathbf{Y} \mid Z_k = 1) \sim \mathcal{N}(0, \Theta_{\mathcal{G}_k}^{-1})$. Generally, $\mathbf{Z}$ is unobserved, we thus use the posterior probability $p(\mathbf{Z} \mid \mathbf{Y})$ to approximate $\mathbf{Z}$. In this section, we describe a more abstract view of the penalized EM algorithm for the mixture of pdRCON models via complete data. A more detailed of the computational algorithm is given in Section S1 of the Supplementary.

4.1. Penalized complete log-likelihood function

For the complete data $\mathcal{D}_c = \{(\mathbf{y}_1, \mathbf{z}_1), \ldots, (\mathbf{y}_N, \mathbf{z}_N)\}$, the complete log-likelihood of $(\mathbf{w}, \boldsymbol{\Theta}_{\mathcal{G}})$ can be computed as

$$l_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2}(\mathcal{D}_c \mid \mathbf{w}, \boldsymbol{\Theta}_{\mathcal{G}}) = \sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}\Big(\log w_k + \log p(\mathbf{y_n} \mid \Theta_{\mathcal{G}_k})\Big) - \operatorname{pen}_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2}(\boldsymbol{\Theta}_{\mathcal{G}}),$$

where $\operatorname{pen}_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2}(\cdot)$ is the fused lasso penalty function defined in (7). In practice, we cannot derive the value of the (penalized) complete log-likelihood function due to unobserved

variables $\mathbf{Z}_n$, we consider the expectation of the (penalized) complete log-likelihood with respect to the posterior of the latent variables, which is

$$\mathbb{E}_{\mathbf{Z}|\mathbf{Y}}\Big(l_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2}(\mathcal{D}_c \mid \mathbf{w}, \boldsymbol{\Theta}_{\mathcal{G}})\Big) = \sum_{n=1}^{N}\sum_{k=1}^{K}\tau_{nk}\Big(\log w_k + \log p(\mathbf{y_n} \mid \Theta_{\mathcal{G}_k})\Big) - \text{pen}_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2}(\boldsymbol{\Theta}_{\mathcal{G}}) \tag{8}$$

where τ_{nk} is denoted the conditional expectation of Z_{nk} given observations $\mathbf{y}_n$, which can be specified by using Bayes' theorem, for every $n \in \{1,\ldots,N\},\ k \in \{1,\ldots,K\}$, as

$$\begin{aligned}\tau_{nk} &= \mathbb{E}_{\mathbf{Z}|\mathbf{Y}}(Z_{nk}) = \mathbb{P}(Z_{nk}=1 \mid \mathbf{y}_n, \mathbf{w}, \boldsymbol{\Theta}_{\mathcal{G}}) \\ &= \frac{p(\mathbf{y}_n, \mid Z_{nk}=1, \boldsymbol{\Theta}_{\mathcal{G}}, \mathbf{w}) \times \mathbb{P}(Z_{nk}=1 \mid \mathbf{w})}{p(\mathbf{y}_n \mid \boldsymbol{\Theta}_{\mathcal{G}}, \mathbf{w})} \\ &= \frac{w_k p(\mathbf{y_n} \mid \Theta_{\mathcal{G}_k})}{\sum_{l=1}^{K} w_l p(\mathbf{y_n} \mid \Theta_{\mathcal{G}_l})}.\end{aligned} \tag{9}$$

The quantity τ_{nk} is known as the posterior distribution of Z_{nk} given the observations and is used to find the MLE of the model parameters in the EM algorithm, which is described in the following section.

4.2. The algorithm

EM algorithm alternates between the expectation step (E-step), which computes the conditional expectation of the penalized complete log-likelihood with current values of parameters, and the maximization step (M-step), which updates the parameters based on maximizing the conditional expectation computed in E-step, until convergence, e.g., when there is no longer significant change in the variation of the parameter estimation. In particular,

(E-step) given the observed data $\mathbf{y}_1,\ldots,\mathbf{y}_N$ with current values of parameters $(\mathbf{w}^{(t)}, \boldsymbol{\Theta}_{\mathcal{G}}^{(t)})$ at t-th iteration of the algorithm, the posterior distribution of the latent variables is given by $\tau_{nk}^{(t)} = p(Z_{nk} \mid \mathbf{y}_n, \mathbf{w}^{(t)}, \boldsymbol{\Theta}_{\mathcal{G}}^{(t)})$ specified by (9).

(M-step) We use $\tau_{nk}^{(t)}$ to evaluate the conditional expectation of the penalized complete log-likelihood, which is defined by

$$O_{\text{pen}}\Big((\mathbf{w}, \boldsymbol{\Theta}_{\mathcal{G}}), (\mathbf{w}^{(t)}, \boldsymbol{\Theta}_{\mathcal{G}}^{(t)})\Big) = \sum_{n=1}^{N}\sum_{k=1}^{K}\tau_{nk}^{(t)}\Big(\log w_k + \log p(\mathbf{y_n} \mid \Theta_{\mathcal{G}_k})\Big) - \text{pen}_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2}(\boldsymbol{\Theta}_{\mathcal{G}}). \tag{10}$$

We observe that (10) can be decomposed into independent expressions as

$$O_{\text{pen}}\Big((\mathbf{w}, \boldsymbol{\Theta}_{\mathcal{G}}), (\mathbf{w}^{(t)}, \boldsymbol{\Theta}_{\mathcal{G}}^{(t)})\Big) = O(\mathbf{w}, \mathbf{w}^{(t)}) + O_{\text{pen}}(\boldsymbol{\Theta}_{\mathcal{G}}, \boldsymbol{\Theta}_{\mathcal{G}}^{(t)}),$$

where

$$\begin{aligned} O(\mathbf{w}, \mathbf{w}^{(t)}) &= \sum_{n=1}^{N}\sum_{k=1}^{K}\tau_{nk}^{(t)}\log w_k, \quad \text{and} \\ O_{\text{pen}}(\boldsymbol{\Theta}_{\mathcal{G}}, \boldsymbol{\Theta}_{\mathcal{G}}^{(t)}) &= \sum_{n=1}^{N}\sum_{k=1}^{K}\tau_{nk}^{(t)}\log p(\mathbf{y_n} \mid \Theta_{\mathcal{G}_k}) - \text{pen}_{\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2}(\boldsymbol{\Theta}_{\mathcal{G}}). \end{aligned}$$

We update new parameters $\left(\mathbf{w}^{(t+1)}, \mathbf{\Theta}_{\mathcal{G}}^{(t+1)}\right)$ by separately maximizing the two independent components of (10) as follows:

1. **Update mixture proportion $\mathbf{w}^{(t+1)}$.** By applying the Lagrange multiplier method to constraint $\sum_{k=1}^{K} w_k = 1$,we obtain the new update of w_k as

$$\widehat{w}_k^{(t+1)} = N_k^{(t)}/N \quad \text{with } N_k^{(t)} = \sum_{n=1}^{N} \tau_{nk}^{(t)} \tag{11}$$

   where $N_k^{(t)}$ is denoted as the effective number of observations assigned to class k.

2. **Update models' parameters $\mathbf{\Theta}_{\mathcal{G}}$.** The second term of (10) can be written as

$$\begin{aligned}
&O_{\text{pen}}(\mathbf{\Theta}_{\mathcal{G}}, \mathbf{\Theta}_{\mathcal{G}}^{(t)}) \\
&= \sum_{n=1}^{N}\sum_{k=1}^{K} \tau_{nk}^{(t)} \log p(\mathbf{y_n} \mid \Theta_{\mathcal{G}_k}) - \sum_{k=1}^{K} \lambda_k^{[1]} \|\Theta_{\mathcal{G}_k}\|_1 - \sum_{k=1}^{K} \lambda_k^{[2]} \Big(\|\Theta_{\mathcal{G}_k}^{LL} - \Theta_{\mathcal{G}_k}^{RR}\|_1 + \|\Theta_{\mathcal{G}_k}^{LR} - \Theta_{\mathcal{G}_k}^{RL}\|_1 \Big) \\
&= \frac{1}{2} \sum_{k=1}^{K} N_k^{(t)} \Big[\log\det(\Theta_{\mathcal{G}_k}) - tr\big(S_k^{(t)} \Theta_{\mathcal{G}_k}\big) \Big] \\
&- \sum_{k=1}^{K} \lambda_k^{[1]} \|\Theta_{\mathcal{G}_k}\|_1 - \sum_{k=1}^{K} \lambda_k^{[2]} \Big(\|\Theta_{\mathcal{G}_k}^{LL} - \Theta_{\mathcal{G}_k}^{RR}\|_1 + \|\Theta_{\mathcal{G}_k}^{LR} - \Theta_{\mathcal{G}_k}^{RL}\|_1 \Big),
\end{aligned} \tag{12}$$

   where, for $k \in \{1, \ldots, K\}$, $S_k^{(t)} = \sum_{n=1}^{N} \tau_{nk}^{(t)} \mathbf{y}_n^T \mathbf{y}_n / N_k^{(t)}$ is denoted as a weighted sample covariance matrix, and $tr(\cdot)$ is denoted the trace of a square matrix, i.e. the sum of elements on the main diagonal entries. As shown in (12), performing the update for the Gaussian networks' parameters corresponds to solving K separated fused lasso problems using the alternating direction method of multiplier (ADMM) algorithm proposed by Boyd et al. (2011). In particular, for every $k \in \{1, \ldots, K\}$,

$$\begin{aligned}
\widehat{\Theta}_{\mathcal{G}_k}^{(t+1)} = \text{argmin} \Big\{ &- N_k^{(t)} \Big[\log\det(\Theta_{\mathcal{G}_k}) - tr\big(S_k^{(t)} \Theta_{\mathcal{G}_k}\big) \Big] \\
&+ \lambda_k^{[1]} \|\Theta_{\mathcal{G}_k}\|_1 + \lambda_k^{[2]} \Big(\|\Theta_{\mathcal{G}_k}^{LL} - \Theta_{\mathcal{G}_k}^{RR}\|_1 + \|\Theta_{\mathcal{G}_k}^{LR} - \Theta_{\mathcal{G}_k}^{RL}\|_1 \Big) \Big\}.
\end{aligned} \tag{13}$$

   We refer the readers to Ranciati et al. (2021), Ranciati and Roverato (2023) for the application of ADMM to the graphical lasso for paired data. A more detailed technical computation for updating new parameter $\Theta_{\mathcal{G}_k}$ using ADMM method is provided in Section S1 of Supplementary.

In summary, the pseudocode of the penalized EM algorithm for a mixture of pdRCON models is given in Algorithm 1.

5. Application

The fused penalized EM algorithm for a mixture of pdRCON models is implemented by the programming language R on synthetic and real data. Parameter initialization, model selection and stopping rule of the EM algorithm will be considered in a specific case. Numerical

Algorithm 1: Penalzied EM for a mixture pdRCON model

Data: samples $(\mathbf{y}_1, \ldots, \mathbf{y}_N)$, regularizations $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$, maximum iteration number $T_{\max}$ initialization $\mathbf{w}^{(current)}$, $\boldsymbol{\Theta}_{\mathcal{G}}^{(current)}$, and $t \leftarrow 0$;

while *(convergence = false) and* $(t < T_{\max})$ **do**

(E-step) evaluate $\tau_{nk}^{(current)}$ using $\mathbf{w}^{(current)}$, $\boldsymbol{\Theta}_{\mathcal{G}}^{(current)}$ by equation (9);

(M-step) update $\mathbf{w}^{(new)}$ using $\tau_{nk}^{(current)}$ by equation (11);

update $\boldsymbol{\Theta}_{\mathcal{G}}^{(new)}$ using $\tau_{nk}^{(current)}$ by ADMM method solving (13);

check for convergence;

if *(convergence = true)* **then**

break and return $(\mathbf{w}^{(new)}, \boldsymbol{\Theta}_{\mathcal{G}}^{(new)})$;

else

$t \leftarrow t + 1$;

$\mathbf{w}^{(current)} \leftarrow \mathbf{w}^{(new)}$ and $\boldsymbol{\Theta}_{\mathcal{G}}^{(current)} \leftarrow \boldsymbol{\Theta}_{\mathcal{G}}^{(new)}$;

end

end

performance is presented for different sparsity and symmetry of parameters, including a comparison with the glasso method for graphical Gaussian mixture models.

5.1. Initialization, model selection and stopping rule

In the application, we implement the EM algorithm with initial values of $\mathbf{w}$ by the fractions of data points assigned to each class obtained by the k-means method, and the initial values of the concentration matrices are therefore considered as diagonal matrices whose diagonal entries are equal to the inverse of the sample variance of the data points within the sub-population, i.e. $1/(\tilde{S}_k)_{pp}$ for $p \in \{1, \ldots, P\}$. This is a reasonable choice, as the variables are generally at different scales in many real-life applications.

We apply the fused lasso for 5 different logarithmically spaced values of $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$, in particular, for every $k = 1, \ldots, K$, $\Lambda_k^{[1]}/5 \leq \lambda_k^{[1]} \leq \Lambda_k^{[1]}$ and $\Lambda_k^{[2]}/5 \leq \lambda_k^{[2]} \leq \Lambda_k^{[2]}$ where $\Lambda_k^{[1]} = \max|(\tilde{S}_k)_{ij}|$ and $\Lambda_k^{[2]} = \max\{|(\tilde{S}_k^{LL})_{ij} - (\tilde{S}_k^{RR})_{ij}|, |(\tilde{S}_k^{LR})_{ij} - (\tilde{S}_k^{RL})_{ij}|\}$, respectively. This setting is suitable for large-scale biological datasets which encourages more sparsity and symmetry constraints on parameters. Furthermore, implementing the EM algorithm on an exhaustive search for $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$ over K components leads to a very costly computation, hence, we will first fix $\boldsymbol{\lambda}_2$ to a low value, which could be zero, and perform the dense grid search for $\boldsymbol{\lambda}_1$ over K classes. After selecting the best value of $\boldsymbol{\lambda}_1$, a grid search for $\boldsymbol{\lambda}_2$ can be performed to select the final pair of optimal values of $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$. As the criteria for choosing the optimal value of regularization parameters, we apply an approximation of the extended BIC (eBIC) criterion (Foygel and Drton (2010)), which is computed as

$$\text{eBIC}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = -\frac{2}{N}\sum_{n=1}^{N}\log\Big\{\sum_{k=1}^{K} w_k p(\mathbf{y}_n \mid \widehat{\Theta}_{(\lambda_k^{[1]}, \lambda_k^{[2]})})\Big\} + d\log(N) + 4d\gamma\log(P), \qquad (14)$$

where $\widehat{\Theta}_{(\lambda_k^{[1]},\lambda_k^{[2]})}$ is the penalized maximum likelihood estimation, d is the number of parameters of the associated model and $\gamma \in [0,1]$. According to Foygel and Drton (2010), we set $\gamma = 0.5$. The optimal choice of $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$ is determined by a two-step procedure: (i) we first select the optimal values of $\boldsymbol{\lambda}_1$ that are $\boldsymbol{\lambda}_1^* = \operatorname{argmin} \text{eBIC}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 = 0)$, and (ii) given $\boldsymbol{\lambda}_1^*$, the optimal values of $\boldsymbol{\lambda}_1$ are obtained by $\boldsymbol{\lambda}_2^* = \operatorname{argmin} \text{eBIC}(\boldsymbol{\lambda}_1^*, \boldsymbol{\lambda}_2)$.

Regarding the stopping condition of the EM algorithm, we check the convergence by the change of the current estimate of solutions using a convenient matrix norm, i.e. if the Frobenius norm $\|\boldsymbol{\Theta}_{\mathcal{G}_k}^{(new)} - \boldsymbol{\Theta}_{\mathcal{G}_k}^{(current)}\|_F^2$ is less than a chosen tolerance threshold for all classes, the algorithm is stopped and has converged.

5.2. Simulation study

In this section, we conduct a simulation study of the mixture of pdRCON models in $K = 2$ sub-populations. We consider three scenarios, called A, B, and C, that differ in the graphs' density degree, i.e. the ratio between the edges present in a graph and the maximum number of edges. In particular, the density degrees of the two mixture components are approximately equal to $(0.6, 0.6)$ for scenario A, $(0.15, 0.15)$ for scenario B, and $(0.15, 0.6)$ for scenario C. For each scenario, we generate three pdCGs with $P = 50$ vertices on different symmetry densities between two components based on the edges present. For every pdCG $\mathcal{G}$, a concentration matrix $\boldsymbol{\Theta}_{\mathcal{G}}$ was randomly generated such that the Gaussian distribution $\mathcal{N}(0, \boldsymbol{\Theta}_{\mathcal{G}_k}^{-1})$ for each class k restricted on $\mathcal{G}_k$. Then 100 samples $\mathbf{y}_1, \ldots, \mathbf{y}_{100}$ were simulated from the two-component multivariate normal mixture model with two different settings of the (true) mixture proportions: $\mathbf{w}^{(1)} = (0.3, 0.7)$ and $\mathbf{w}^{(2)} = (0.5, 0.5)$. To each simulated dataset, we apply the penalized EM algorithm for a mixture of pdRCON models with fused lasso proposed in (6)-(7) compared to graphical lasso introduced in (4).

Moreover, averaged Kullback-Leibler (KL) loss was used as a measure of model performance which is computed as

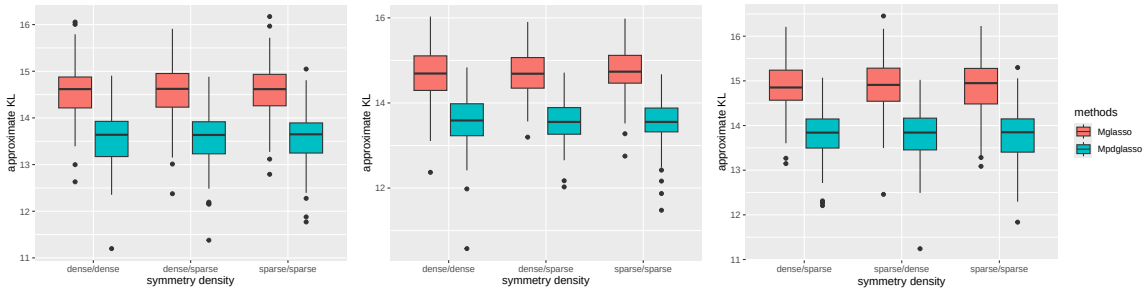
$$\frac{1}{N}\sum_{n=1}^{N} \log \frac{p(\mathbf{y}_n \mid \mathbf{w}^{true}, \boldsymbol{\Theta}_{\mathcal{G}}^{true})}{p(\mathbf{y}_n \mid \mathbf{w}^{estimate}, \boldsymbol{\Theta}_{\mathcal{G}}^{estimate})},$$

where $p(\cdot \mid \cdot)$ is the density function given by (5). Another measure we used here is the Frobenuis norm of the difference between the true and estimated concentration matrices for each sub-population, i.e. $\|\boldsymbol{\Theta}_{\mathcal{G}_k}^{true} - \boldsymbol{\Theta}_{\mathcal{G}_k}^{estimate}\|_F^2$ for $k \in \{1, \ldots, K\}$.

The quantile values of these measurements over 100 replicates are presented in Figures 1-3 for the first case of mixture proportion $\mathbf{w}^{(1)} = (0.3, 0.7)$ and in Section S2 of Supplementary material for the second case of $\mathbf{w}^{(2)} = (0.5, 0.5)$. The recorded results of using both KL (Figure 1) and Frobenuis measure (Figure 2 and 3) reveal that the fused graphical lasso for paired data we have proposed performs significantly better than the graphical lasso approach, observing that the median values for both measures from the graphical lasso are greater than the interquartile range of the fused lasso method.

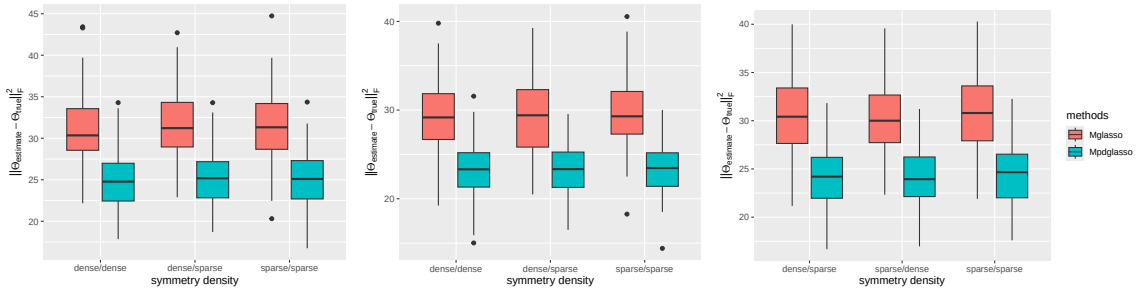
5.3. Application to transcriptomic data

We consider a transcriptomic dataset originally published in Wang et al. (2020). Briefly, the study focused on the marine-freshwater divergence in nine-spined sticklebacks, where



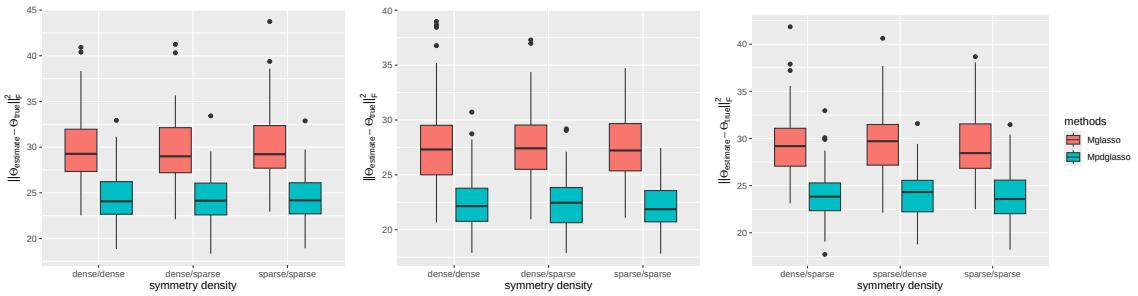
(a) (b) (c)

Figure 1: The quantile values of averaged Kullback-Leibler losses obtained from 100 replications of the graphical lasso method and fused lasso for the two-components pdRCON models with the mixture proportion $\mathbf{w} = (0.3, 0.7)$. Subfigures (a), (b), and (c) show the results recorded for scenario A, scenario B, and scenario C, respectively, of the generated concentrations.



(a) (b) (c)

Figure 2: The quantile values of Frobenius norm values of the difference between the true and estimated concentration matrices for sub-population $k = 1$. Subfigures (a), (b), and (c) show the results recorded for scenario A, scenario B, and scenario C, respectively, of the generated concentrations of two-component pdRCON models with the mixture proportion $\mathbf{w} = (0.3, 0.7)$.

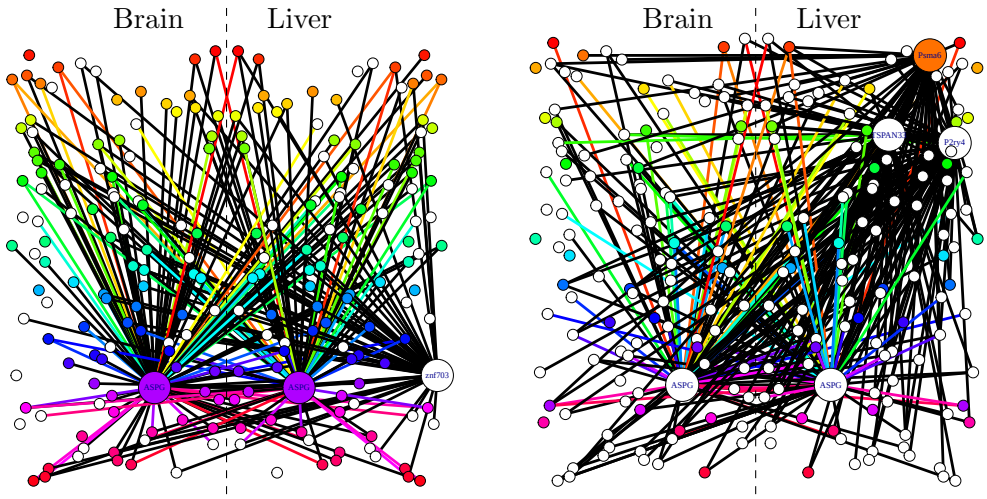


(a) (b) (c)

Figure 3: The quantile values of Frobenius norm values of the difference between the true and estimated concentration matrices for sub-population $k = 2$. Subfigures (a), (b), and (c) show the results recorded for scenario A, scenario B, and scenario C, respectively, of the generated concentrations of two-component pdRCON models with the mixture proportion $\mathbf{w} = (0.3, 0.7)$.

RNA-seq data were collected from 24 fish representing two marine and four freshwater populations in Finland and Sweden. The two groups of variables correspond to brain tissue and liver tissue, with each gene in brain tissue paired with its homologous gene in liver tissue. After filtering out genes with low variance and outliers, we selected expression data of 214 genes, comprising 107 genes from the brain and their 107 homologous genes from the liver, which were identified as top differential expressed genes in Wang et al. (2020), as a basis to estimate the gene network. We choose $K = 2$ representing the two ecological populations, i.e. marine and freshwater. In this application, we apply a mixture of pdRCON models with the following aims: (1) to evaluate whether the mixture of pdRCON models can accurately classify data points into marine and freshwater groups, and (2) to learn graphical networks that reveal distinct topological structures between the subpopulations, (3) to explore the similarity between variables in the two groups corresponding to RNA-seq data collected from brain and liver tissues, and (4) to identify a set of genes that may play a key role within the gene network.

The selected model classifies 6 sticklebacks into class 1 and 18 into class 2. Notably, two individuals are misclassified based on their known habitat origin. However, a principal component (PC) analysis on the same RNA-seq data, conducted by Wang et al. (2020) revealed that some marine fish collected from the Helsinki Baltic Sea are genetically closer to the freshwater population than to the marine population, according to the first two PCs. Thus, our method performs well and aligns with existing approaches in the clustering task. Figure 4 illustrates the gene co-expression networks of brain and liver tissues derived from GGMs using a fused graphical lasso. This representation effectively highlights key features of the model, such as network structures and symmetries that differ between classes and tissues. Interestingly, in both habitats, the ASPG gene emerges as a hub, connected to numerous other genes. The ASPG gene has previously been reported to be expressed in response to salinity and is implicated in salt-sensitive hypertension in three-spined sticklebacks (Wang et al. (2020); Gibbons et al. (2017)), underscoring its important role in explaining fresh-



(a) Class 1 (b) Class 2

Figure 4: Colored graphical representations of gene co-expression networks highlight genes that are highly correlated with other genes in brain and liver tissues of the nine-spined stickleback collected in two habitats with (a) presenting for habitat class 1 and (b) presenting for habitat class 2.

water/marine divergence in sticklebacks. In addition, the network for class 2 reveals more parsimonious gene connections in the brain, while more hub genes are identified in the liver, providing an interesting direction for further biological investigation.

6. Concluding remarks

We consider high-dimensional and heterogeneous gene expression data, where observations from each sub-population originate from two dependent groups of variables across tissues, cells, or observable physical properties of an organism. We address this problem within the framework of a mixture of Gaussian graphical models, represented by colored graphs for paired data. We propose a fused graphical lasso method for maximum likelihood estimation in a mixture of GGMs for paired data, aimed at uncovering relationships between genes with expression measured under different conditions and comparing group-specific gene networks. Our simulation studies demonstrate that the fused graphical lasso to the mixture GGMs for paired data outperforms the standard graphical lasso method in model estimation. Additionally, we applied our method to a high-dimensional transcriptomic dataset of nine-spined sticklebacks, collected from marine and freshwater environments across brain and liver tissues, where the number of genes greatly exceeds the number of individuals (e.g. $214 > 24$). The results align with other studies in terms of estimating gene networks, identifying hub genes, classifying individuals according to common biological characteristics, and providing new insights into the differentiation of gene networks across habitats. It is also noteworthy that the mixture of pdRCON models with the fused graphical lasso can be

effectively applied in clustering scenarios with an unknown number of mixture components, which necessitates model selection based on specific criteria to determine the appropriate number of classes K. Furthermore, improving the estimation process involves the selection of an appropriate set of starting values for the parameters, as well as the development of theoretical theorems and practical techniques concerning the consistency and convergence rate of the fusion lasso penalized MLE for a mixture of pdRCON models and the overlap between mixture components in the clustering algorithm.

Acknowledgement We would like to express our gratitude for the support provided by CSIRO Agriculture and Food in undertaking this research. Our thanks also go to Dr. Jia Liu for proofreading the initial version of this manuscript and to the three anonymous reviewers for their valuable and constructive feedback.

References

Dvir Aran, Roman Camarda, Justin Odegaard, Hyojung Paik, Boris Oskotsky, Gregor Krings, Andrei Goga, Marina Sirota, and Atul J Butte. Comprehensive analysis of normal adjacent to tumor transcriptomes. *Nature communications*, 8(1):1077, 2017.

Mélisande Blein-Nicolas, Emilie Devijver, Mélina Gallopin, and Emeline Perthame. Non-linear network-based quantitative trait prediction from biological data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, page qlae012, 2024.

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

Edward T Bullmore and Danielle S Bassett. Brain graphs: graphical models of the human brain connectome. *Annual review of clinical psychology*, 7(1):113–140, 2011.

Hsun-Hsien Chang and Michael McGeachie. Phenotype prediction by integrative network analysis of SNP and gene expression microarrays. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6849–6852. IEEE, 2011.

Patrick Danaher, Pei Wang, and Daniela M Witten. The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(2):373–397, 2014.

Arthur P Dempster. Covariance selection. *Biometrics*, pages 157–175, 1972.

Patrik D'haeseleer, Shoudan Liang, and Roland Somogyi. Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics*, 16(8):707–726, 2000.

Karoline Faust. Open challenges for microbial network construction and analysis. *The ISME journal*, 15:3111—-3118, 2021.

Rina Foygel and Mathias Drton. Extended Bayesian information criteria for Gaussian graphical models. *Advances in neural information processing systems*,

23, 2010. URL https://proceedings.neurips.cc/paper_files/paper/2010/file/072b030ba126b2f4b2374f342be9ed44-Paper.pdf.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

Taylor C Gibbons, David CH Metzger, Timothy M Healy, and Patricia M Schulte. Gene expression plasticity in response to salinity acclimation in threespine stickleback ecotypes from different salinity habitats. *Molecular ecology*, 26(10):2711–2725, 2017.

Thomas J Hardcastle and Krystyna A Kelly. Empirical Bayesian analysis of paired high-throughput sequencing data with a beta-binomial distribution. *BMC bioinformatics*, 14: 1–11, 2013.

Søren Højsgaard and Steffen L Lauritzen. Graphical Gaussian models with edge and vertex symmetries. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 70 (5):1005–1027, 2008.

Zachary D Kurtz, Christian L Müller, Emily R Miraldi, Dan R Littman, Martin J Blaser, and Richard A Bonneau. Sparse and compositionally robust inference of microbial ecological networks. *PLoS computational biology*, 11(5):e1004226, 2015.

Thomas Lartigue, Stanley Durrleman, and Stéphanie Allassonnière. Mixture of conditional Gaussian graphical models for unlabelled heterogeneous populations in the presence of co-factors. *SN Computer Science*, 2(6):466, 2021.

Steffen L Lauritzen. *Graphical models*, volume 17. Clarendon Press, 1996.

Kevin H Lee and Lingzhou Xue. Nonparametric finite mixture of Gaussian graphical models. *Technometrics*, 60(4):511–521, 2018.

Faming Liang and Bochao Jia. *Sparse Graphical Modeling for High Dimensional Data: A Paradigm of Conditional Independence Tests*. Chapman and Hall/CRC, 2023.

Liliana López-Kleine, Luis Leal, and Camilo López. Biostatistical approaches for the reconstruction of gene co-expression networks based on transcriptomic data. *Briefings in functional genomics*, 12(5):457–467, 2013.

Anani Lotsi and Ernst Wit. Sparse Gaussian graphical mixture model. *Afrika Statistika*, 11(2):1041–1059, 2016.

Shisong Ma, Qingqiu Gong, and Hans J Bohnert. An Arabidopsis gene network based on the graphical Gaussian model. *Genome research*, 17(11):1614–1625, 2007.

Henry E Miller and Alexander JR Bishop. Correlation analyzer: functional predictions from gene co-expression correlations. *BMC bioinformatics*, 22:1–19, 2021.

Saverio Ranciati and Alberto Roverato. On the application of Gaussian graphical models to paired data problems. *arXiv preprint arXiv:2307.14160*, 2023.

Saverio Ranciati, Alberto Roverato, and Alessandra Luati. Fused graphical lasso for brain networks with symmetries. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 70(5):1299–1322, 2021.

Xiaolan Rao and Richard A Dixon. Co-expression networks for plant biology: why and how. *Acta biochimica et biophysica Sinica*, 51(10):981–988, 2019.

Alberto Roverato and Dung Ngoc Nguyen. Model inclusion lattice of coloured Gaussian graphical models for paired data. In *International Conference on Probabilistic Graphical Models*, pages 133–144. PMLR, 2022. URL https://proceedings.mlr.press/v186/roverato22a.html.

Alberto Roverato and Dung Ngoc Nguyen. Exploration of the search space of Gaussian graphical models for paired data. *Journal of Machine Learning Research*, 25(92):1–41, 2024. URL http://jmlr.org/papers/v25/23-0295.html.

Rory Stark, Marta Grzelak, and James Hadfield. RNA sequencing: the teenage years. *Nature Reviews Genetics*, 20(11):631–656, 2019.

Remco Ursem, Yury Tikunov, Arnaud Bovy, Ralph Van Berloo, and Fred Van Eeuwijk. A correlation network approach to metabolic data analysis for tomato fruits. *Euphytica*, 161:181–193, 2008.

Beatriz Valcarcel, Peter Würtz, Nafisa-Katrin Seich al Basatena, Taru Tukiainen, Antti J Kangas, Pasi Soininen, Marjo-Riitta Järvelin, Mika Ala-Korpela, Timothy M Ebbels, and Maria de Iorio. A differential network approach to exploring differences between biological states: an application to prediabetes. *PLoS One*, 6(9):e24702, 2011.

Yingnan Wang, Yongxin Zhao, Yu Wang, Zitong Li, Baocheng Guo, and Juhä Merila. Population transcriptomics reveals weak parallel genetic basis in repeated marine and freshwater divergence in nine-spined sticklebacks. *Molecular Ecology*, 29(9):1642–1656, 2020.

Ming Yuan and Yi Lin. Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94(1):19–35, 2007.

Zihao Zheng, Stefan Hey, Talukder Jubery, Huyu Liu, Yu Yang, Lisa Coffey, Chenyong Miao, Brandi Sigmon, James C Schnable, Frank Hochholdinger, et al. Shared genetic control of root system architecture between Zea mays and sorghum bicolor. *Plant physiology*, 182 (2):977–991, 2020.

Hui Zhou, Wei Pan, and Xiaotong Shen. Penalized model-based clustering with unconstrained covariance matrices. *Electronic journal of statistics*, 3:1473, 2009.