JMLR: Workshop and Conference Proceedings 260:-, 2024 ACML 2024

Supplementary material for "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 this technical supplement, we provide further details concerning the penalized EM algorithm for a mixture of pdRCON models, and numerical performance with the quantile values of metric scores in the case of a balanced mixture proportion $\mathbf{w} = (0.5, 0.5)$ for two sub-populations.

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

S1. Technical details of the penalized EM algorithm

In this section, we provide a more technical detail of EM algorithm which alternates between (E-step) for computation of the conditional expectation of the penalized complete log-likelihood with current values of parameters, and (M-step) for updating the parameters based on maximizing the conditional expectation computed in E-step, until convergence.

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

$$\tau_{nk}^{(t)} = \frac{w_k^{(t)} p(\mathbf{y_n} \mid \Theta_{\mathcal{G}_k}^{(t)})}{\sum_{l=1}^{K} w_l^{(t)} p(\mathbf{y_n} \mid \Theta_{\mathcal{G}_l}^{(t)})}.$$

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

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

We observe that (1) can be decomposed into independent expressions as

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

where

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

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

1. **Update mixture proportion $\mathbf{w}^{(t+1)}$.** Using the Lagrange multiplier η to constrain $\sum_{k=1}^{K} w_k = 1$, we arrive at the updating formula as

$$\widehat{w}_k^{(t+1)} = \underset{\mathbf{w}}{\text{argmax}} \left\{ \sum_{k=1}^{K} \sum_{n=1}^{N} \tau_{nk}^{(t)} \log w_k - \eta \Big(\sum_{k=1}^{K} w_k - 1 \Big) \right\}$$

By taking the partial derivative of the Lagrange function with respect to w_k and setting it to 0, we get

$$\frac{\sum_{n=1}^{N} \tau_{nk}^{(t)}}{w_k} - \eta = 0 \text{ so that } \widehat{w}_k^{(t+1)} = \frac{\sum_{n=1}^{N} \tau_{nk}^{(t)}}{\eta}.$$

Since $\sum_{k=1}^{K} w_k = 1$, we find the multiplier η as

$$\sum_{k=1}^{K} w_k - 1 = \sum_{k=1}^{K} \frac{\sum_{n=1}^{N} \tau_{nk}^{(t)}}{\eta} - 1 = \sum_{n=1}^{N} \frac{\sum_{k=1}^{K} \tau_{nk}^{(t)}}{\eta} - 1 = \sum_{n=1}^{N} \frac{1}{\eta} - 1 = 0$$

which implies that $\eta = N$. Hence $\widehat{w}_k^{(t+1)} = N_k^{(t)}/N$ with $N_k^{(t)} = \sum_{n=1}^{N} \tau_{nk}^{(t)}$.

2. **Update models' parameters $\mathbf{\Theta}_{\mathcal{G}}$.** The second term of (1) can be written as

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

where $S_k^{(t)} = \sum_{n=1}^{N} \tau_{nk}^{(t)} \mathbf{y}_n^T \mathbf{y}_n / N_k^{(t)}$ is denoted as a weighted sample covariance for $k \in \{1, \dots, K\}$ and $tr(\cdot)$ is the trace of a square matrix, i.e. the sum of elements on the main diagonal entries. As shown in (2), 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 $k \in \{1, \ldots, K\}$,

$$\widehat{\Theta}_{\mathcal{G}_k}^{(t+1)} = \operatorname{argmin}\Big\{ -N_k^{(t)}\Big[\log\det(\Theta_{\mathcal{G}_k}) - tr\big(S_k^{(t)}\Theta_{\mathcal{G}_k}\big)\Big] + \lambda_k^{[1]}\|\Theta_{\mathcal{G}_k}\|_1 + \lambda_k^{[2]}\Big(\|\Theta_{\mathcal{G}_k}^{LL} - \Theta_{\mathcal{G}_k}^{RR}\|_1 + \|\Theta_{\mathcal{G}_k}^{LR} - \Theta_{\mathcal{G}_k}^{RL}\|_1\Big)\Big\}. \tag{3}$$

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. More specifically, the optimization problem in (3) is equivalent to ADMM form with respect to Θ_k and X_k by minimizing

$$-N_k^{(t)}\Big(\log\det\Theta_{\mathcal{G}_k} - tr(S_k^{(t)}\Theta_{\mathcal{G}_k})\Big) + \lambda_k^{[1]}\|X_k\|_1 + \lambda_k^{[2]}\Big(\|X_k^{LL} - X_k^{RR}\|_1 + \|X_k^{LR} - X_k^{RL}\|_1\Big) \tag{4}$$

subject to $\Theta_{\mathcal{G}_k} - X_k = 0$.

The ADMM algorithm for (4) is implemented by iterating the following steps:

(a) $\Theta_{\mathcal{G}_k}^{(l+1)} = \operatorname{argmin} -N_k^{(t)}\Big(\log\det\Theta_{\mathcal{G}_k} - tr(S_k^{(t)}\Theta_{\mathcal{G}_k})\Big) + \frac{\alpha}{2}\|\Theta_{\mathcal{G}_k} - X_k^{(l)} + U_k^{(l)}\|_F^2$

(b) $X_k^{(l+1)} = \operatorname{argmin}\Big(\lambda_k^{[1]}\|X_k\|_1 + \lambda_k^{[2]}\|X_k^{LL} - X_k^{RR}\|_1 + \frac{\alpha}{2}\|\Theta_{\mathcal{G}_k}^{(l+1)} - X_k + U_k^{(l)}\|_F^2\Big)$

(c) $U_k^{(l+1)} = U_k^{(l)} + \Theta_{\mathcal{G}_k}^{(l+1)} - X_k^{(l+1)}$

where $\|\cdot\|_F$ is the Frobenius norm, i.e. the square root of the sum of the squares of the matrix entries. Step (a) has an analytical solution which is given by

$$\Theta_{\mathcal{G}_k}^{(l+1)} = \mathbf{Q}D\mathbf{Q}^T, \tag{5}$$

where D is a diagonal matrix with the positive elements $d_p = \frac{N_k^{(t)}\left(\gamma_p + \sqrt{\gamma_p^2 + 4\alpha/N_k^{(t)}}\right)}{2\alpha}$ for $p \in \{1, \ldots, P\}$ with $\alpha > 0$, and $\mathbf{Q}$ is the $P \times P$ matrix whose p-th column is the eigenvector so that $QQ^T = Q^TQ = I$. Here, $\{\gamma_1, \ldots, \gamma_P\}$ are eigenvalues of the orthogonal eigenvalue decomposition of $\alpha\Big(X_k^{(l)} - U_k^{(l)}\Big)/N_k^{(t)} + S_k^{(t)}$.

We turn now to step (b) of the algorithm. The update $X_k^{(l+1)}$ for step (b) of the outer ADMM algorithm is thus the symmetric matrix such that $X_k^{(l+1)} = \mathcal{S}_{\beta_1}(x_{[\beta_1=0,\boldsymbol{\beta}_2]})$ with $\beta_1 = \lambda_k^{[1]}/\rho$, $\boldsymbol{\beta}_2 = \lambda_k^{[2]}/\rho \times \mathbf{1}_{2Q^2+Q}$ with $Q = P/2$, and $\mathcal{S}_\kappa(\cdot)$ is the soft thresholding operator defined as

$$\mathcal{S}_\kappa(a) = \begin{cases} a - \kappa, & \text{if } a > \kappa \\ 0, & \text{if } |a| \le \kappa \\ a + \kappa, & \text{if } a < -\kappa, \end{cases}$$

and $x_{[\beta_1=0,\boldsymbol{\beta}_2]}$ is the optimal solution of a generalized lasso problem which is updated until convergence by the following steps

(i) $x^{(m+1)} = (I + \rho' F^T F)^{-1} \Big[\Theta_{\mathcal{G}_k}^{(l)} + U_k^{(l)} + \rho' F^T (v^{(m)} - t^{(m)}) \Big]$

(ii) $v^{(m+1)} = \mathcal{S}_{\boldsymbol{\beta}_2/\rho'}(F x^{(m+1)} + t^{(m)})$

(iii) $t^{(m+1)} = t^{(m)} + F x^{(m+1)} - v^{(m+1)}$,

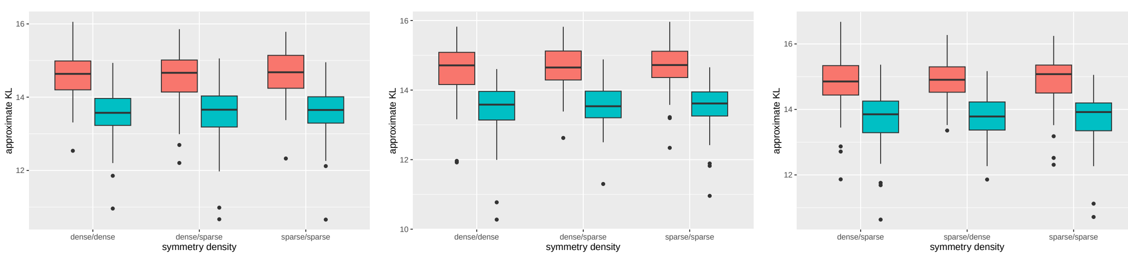
where F is a matrix encoding all (linear) equality constraints, i.e.

$$F = \begin{pmatrix} I_Q & -I_Q & 0_{QS} & 0_{QS} & 0_{QS} & 0_{QS} & 0_{QQ} \\ 0_{SQ} & 0_{SQ} & I_S & -I_S & 0_{SS} & 0_{SS} & 0_{SQ} \\ 0_{SQ} & 0_{SQ} & 0_{SS} & 0_{SS} & I_S & -I_Q S & 0_{SQ} \end{pmatrix}$$

where I and O are respectively the identity and zero matrix, with $S = Q(Q-1)/2$. For more details, see (Boyd et al., 2011, Section 6.4.1 and 6.6) and Ranciati and Roverato (2023).

S2. Additional details on the applications

This section provide the numerical performance of fused graphical lasso compared to the classical graphical lasso method to the mixture of pdRCON models when the (true) mixture proportions of two sub-populations are $\mathbf{w} = (0.5, 0.5)$.



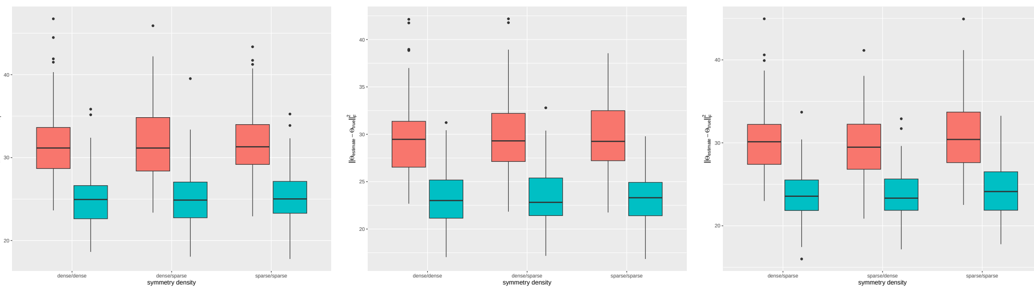
(a) (b) (c)

Figure S1: The quantile values of averaged Kullback-Leibler losses obtained from 100 replications of the graphical lasso method and fused graphical lasso (blue boxplot) for the two-components pdRCON models with the mixture proportion $\mathbf{w} = (0.5, 0.5)$. Subfigures (a), (b), and (c) show the results recorded for scenario A, scenario B, and scenario C, respectively, of the generated concentrations.

References

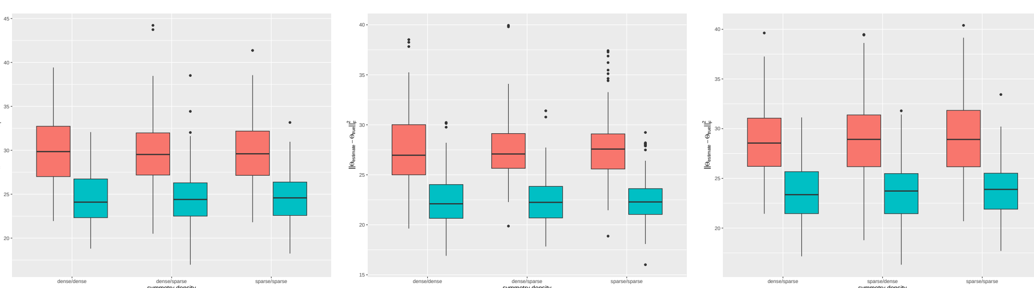
Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011. doi: https://doi.org/10.1561/2200000016.

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



(a) (b) (c)

Figure S2: The quantile values of Frobenius norm values of the difference between the true and estimated concentration matrices for sub-population $k = 1$, obtain from the graphical lasso (red boxplot) and fused graphical lasso method (blue boxplot). 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 $\mathbf{w} = (0.5, 0.5)$.



(a) (b) (c)

Figure S3: The quantile values of Frobenius norm values of the difference between the true and estimated concentration matrices for sub-population $k = 2$, obtain from the graphical lasso (red boxplot) and fused graphical lasso method (blue boxplot). 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 $\mathbf{w} = (0.5, 0.5)$.

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. doi: https://doi.org/10.1111/rssc.12514.