LATTICE OF GAUSSIAN GRAPHICAL MODELS FOR PAIRED DATA WITH COMMON UNDIRECTED STRUCTURE

Dung Ngoc Nguyen [1] and Alberto Roverato[1]

[1] Department of Statistical Sciences, University of Padova, (e-mail: ngocdung.nguyen@unipd.it, alberto.roverato@unipd.it)

ABSTRACT: Typically, a model space embedded with a submodel order relationship has a lattice structure, called the model inclusion lattice. Recent works are related to the problems of joint learning of Gaussian graphical models suited for paired data, with exactly two dependent groups of variables. In this framework, it was shown that the model inclusion lattice does not satisfy the distributivity property, and this increases the complexity of procedures for the exploration of the search space. We consider a relevant subfamily of Gaussian graphical models for paired data represented by coloured graphs with common uncoloured structure. We show that this subfamily forms a proper sublattice of the family of Gaussian graphical models for paired data and that, within this sublattice, the distributivity property is satisfied. This can be exploited to improve efficiency in model search procedures.

KEYWORDS: coloured Gaussian graphical model, RCON model, distributivity.

1 Introduction

In the joint learning of multiple networks, recent works have considered the case of paired data, where the observations come from two dependent groups with the same variables, and every variable in the first group has a *homologous* variable in the second group. In this context, it is of interest to learn the similarities and differences between groups (Xie *et al.*, 2016; Ranciati *et al.*, 2021; Roverato & Nguyen, 2022; Zhang *et al.*, 2022; Roverato & Nguyen, 2023).

Let Y_V be a multivariate Gaussian random vector indexed by $V = \{1,\ldots,p\}$ with covariance matrix Σ and concentration matrix $\Sigma^{-1} = \Theta = (\theta_{ij})_{i,j\in V}$. An undirected graph $G = (V,E)$ consists of a set V of vertices and a set E of edges, which are unordered pairs of elements of V. In a *Gaussian graphical model* (GGM) for Y_V every missing edge of G implies that the corresponding entry of Θ is equal to zero (Lauritzen, 1996). A coloured version of G, denoted by $\mathcal{G} = (\mathcal{V},\mathcal{E})$, consists of a partition $\mathcal{V} = \{V_1,\ldots,V_v\}$ of V and a partition

$\mathcal{E} = \{E_1, \ldots, E_e\}$ of E, into colour classes. A coloured GGM (Højsgaard & Lauritzen, 2008) with coloured graph $\mathcal{G}$ is a GGM with additional symmetry restrictions on the parameters implied by the colouring of $\mathcal{G}$. More specifically, the parameters associated with vertices or edges belonging to a same colour class are restricted to be identical. One type of such restrictions are equalities between elements of the concentration matrix Θ, thereby identifying the family of RCON models.

In paired data problems, the random vector Y_V is partitioned into $Y_V = (Y_L, Y_R)^T$ with $|L| = |R| = p/2 = q$, and it is assumed, without loss of generality, that $L = \{1, \ldots, q\}$, $R = \{1', \ldots, q'\}$ with $i' = q + i$ for $i \in L$, and that for every $i \in L$, Y_i and $Y_{i'}$ form a pair of homologous variables.

Roverato & Nguyen, 2022 introduced a subfamily of RCON models specifically suited for paired data (PD-RCON models) where symmetries are implemented as equality constraints on the diagonal entries $\theta_{ii} = \theta_{i'i'}$ implied by the colour class $\{i, i'\}$ that we call a *twin-pairing* class. Similarly, for the symmetries of the off-diagonal entries that can be either $\theta_{ij} = \theta_{i'j'}$ implied by the twin-pairing class of edges $\{(i, j), (i', j')\}$ between groups, or $\theta_{ij'} = \theta_{ji'}$ implied by the twin-pairing class of edges $\{(i, j'), (j, i')\}$ across groups. Therefore, in the coloured graphs for paired data, $\mathcal{V}$ is divided into $\mathcal{V} = \mathcal{V}^{(t)} \cup \mathcal{V}^{(a)}$ that contains either twin-pairing classes of $\mathcal{V}^{(t)}$ consisting of a pair of homologous vertices, or atomic classes of $\mathcal{V}^{(a)}$ consisting of a single vertex. This is similar to colouring of edges with $\mathcal{E} = \mathcal{E}^{(t)} \cup \mathcal{E}^{(a)}$ where $\mathcal{E}^{(t)}$ contains twin-pairing classes made up of a pair of homologous edges between or across groups and $\mathcal{E}^{(a)}$ contains atomic classes made up of a single edge present on the graph.

2 Exploration of the search space

Typically, a model space is embedded with the *model inclusion*, or *submodel*, relationship resulting in a lattice structure, which is obtained by specifying the meet $\wedge$ and join $\vee$ operations between two models. These operations are used in structure learning of graphical models for the identification of neighbouring models, and it is important that they can be efficiently computed. This is the case of GGMs where the model inclusion coincides with the subset relation between edge sets. Formally, for two GGMs $G = (V, E_G)$ and $H = (V, E_H)$, G is a submodel of H, denoted by $G \preceq H$, if and only if $E_G \subseteq E_H$. Therefore, the family of GGMs is a lattice where the meet $G \wedge H$ and the join $G \vee H$ take particularly simple forms represented by graphs with the edge sets $E_G \cap E_H$ and $E_G \cup E_H$, respectively. The distributivity between these operations is thus satisfied, and we recall that distributivity is a useful property that facilitates the implementation of efficient procedures in lattices and has also been exploited in model selection (Edwards & Havránek, 1987; Davey & Priestley, 2002; Gehrmann, 2011).

Roverato & Nguyen, 2022 considered the family of PD-RCON models, denoted by $\mathcal{P}$, and showed that $\mathcal{P}$ forms a proper sublattice of RCON models under model inclusion and that, also for this sublattice, the distributivity property does not hold. Here, we notice that the family of PD-RCON models can be naturally split into equivalence classes. All the models in a same class have the same underlying uncoloured undirected graph, that is, they are obtained by imposing additional equality restrictions to a common GGM. For an undirected graph $G = (V, E)$ we denote by $\mathcal{P}_E$ the family of PD-RCON models represented by coloured graphs with a common uncoloured structure G. In the following, we show that such equivalence classes form a proper sublattice of $\mathcal{P}$ that is distributive.

Theorem 1 *Let $G = (V, E)$ be an undirected graph. The class of PD-RCON models with a common uncoloured structure $\mathcal{P}_E$, equipped with the model inclusion order $\preceq$, forms a distributive lattice where if $\mathcal{G}, \mathcal{H} \in \mathcal{P}_E$,*

(i) $\mathcal{G} \preceq \mathcal{H}$ *if and only if* $\mathcal{V}_{\mathcal{G}}^{(a)} \subseteq \mathcal{V}_{\mathcal{H}}^{(a)}$ *and* $\mathcal{E}_{\mathcal{G}}^{(a)} \subseteq \mathcal{E}_{\mathcal{H}}^{(a)}$,

(ii) the meet $(\mathcal{V}_{\wedge}, \mathcal{E}_{\wedge}) \in \mathcal{P}_E$ *can be computed as*

- *atomic classes:* $\mathcal{V}_{\wedge}^{(a)} = \mathcal{V}_{\mathcal{G}}^{(a)} \cap \mathcal{V}_{\mathcal{H}}^{(a)}, \quad \mathcal{E}_{\wedge}^{(a)} = \mathcal{E}_{\mathcal{G}}^{(a)} \cap \mathcal{E}_{\mathcal{H}}^{(a)}$
- *twin-pairing classes:* $\mathcal{V}_{\wedge}^{(t)} = \mathcal{V}_{\mathcal{G}}^{(t)} \cup \mathcal{V}_{\mathcal{H}}^{(t)}, \quad \mathcal{E}_{\wedge}^{(t)} = \mathcal{E}_{\mathcal{G}}^{(t)} \cup \mathcal{E}_{\mathcal{H}}^{(t)}$;

(iii) the join $(\mathcal{V}_{\vee}, \mathcal{E}_{\vee}) \in \mathcal{P}_E$ *can be computed as*

- *atomic classes:* $\mathcal{V}_{\vee}^{(a)} = \mathcal{V}_{\mathcal{G}}^{(a)} \cup \mathcal{V}_{\mathcal{H}}^{(a)}, \quad \mathcal{E}_{\vee}^{(a)} = \mathcal{E}_{\mathcal{G}}^{(a)} \cup \mathcal{E}_{\mathcal{H}}^{(a)}$
- *twin-pairing classes:* $\mathcal{V}_{\vee}^{(t)} = \mathcal{V}_{\mathcal{G}}^{(t)} \cap \mathcal{V}_{\mathcal{H}}^{(t)}, \quad \mathcal{E}_{\vee}^{(t)} = \mathcal{E}_{\mathcal{G}}^{(t)} \cap \mathcal{E}_{\mathcal{H}}^{(t)}$.

Proof. Point (i) follows from Proposition 2 of Roverato & Nguyen, 2022 because $E_{\mathcal{G}} = E_{\mathcal{H}} = E$. Furthermore, the meet and the join between $\mathcal{G}$ and $\mathcal{H}$ in (ii) and (iii) can be computed as described in Theorem 4 of Roverato & Nguyen, 2022, with $\tilde{\mathcal{E}}_{\mathcal{G}}^{(a)} = \mathcal{E}_{\mathcal{G}}^{(a)}$, $\tilde{\mathcal{E}}_{\mathcal{G}}^{(t)} = \mathcal{E}_{\mathcal{G}}^{(t)}$, $\tilde{\mathcal{E}}_{\mathcal{H}}^{(a)} = \mathcal{E}_{\mathcal{H}}^{(a)}$ and $\tilde{\mathcal{E}}_{\mathcal{H}}^{(t)} = \mathcal{E}_{\mathcal{H}}^{(t)}$; moreover, $E^* = \emptyset$ with $\mathcal{E}^{(a)}(E) \subseteq (\mathcal{E}_{\mathcal{G}}^{(a)} \cap \mathcal{E}_{\mathcal{H}}^{(a)})$, $\mathcal{E}_{\mathcal{G}}^{(t)}(E) = \mathcal{E}_{\mathcal{G}}^{(t)}$, $\mathcal{E}_{\mathcal{H}}^{(t)}(E) = \mathcal{E}_{\mathcal{H}}^{(t)}$. All notations $\tilde{\mathcal{E}}_{\mathcal{G}}^{(\cdot)}$, $\tilde{\mathcal{E}}_{\mathcal{H}}^{(\cdot)}$, E^*, $\mathcal{E}^{(a)}(E)$, $\mathcal{E}_{\mathcal{G}}^{(t)}(E)$, and $\mathcal{E}_{\mathcal{H}}^{(t)}(E)$ are defined in Section 3 of Roverato & Nguyen, 2022.

3 Conclusions

We have shown that the family of PD-RCON models can be split into equivalence classes which form distributive lattices with respect to model inclusion. Future research work will concern the exploitation of this property to achieve more efficiency in model search procedures.

Acknowledgement. Financial support has been provided by the MUR – Progetti di Ricerca di Rilevante Interesse Nazionale (PRIN) grant 2022 SMNNKY.

References

DAVEY, BRIAN A, & PRIESTLEY, HILARY A. 2002. *Introduction to lattices and order*. Cambridge University Press.

EDWARDS, DAVID, & HAVRÁNEK, TOMÁŠ. 1987. A fast model selection procedure for large families of models. *Journal of the American Statistical Association*, **82**(397), 205–213.

GEHRMANN, HELENE. 2011. Lattices of graphical Gaussian models with symmetries. *Symmetry*, **3**(3), 653–679.

HØJSGAARD, SØREN, & LAURITZEN, STEFFEN L. 2008. Graphical Gaussian models with edge and vertex symmetries. *Journal of the Royal Statistical Society: Series B*, **70**(5), 1005–1027.

LAURITZEN, STEFFEN L. 1996. *Graphical models*. Oxford University Press.

RANCIATI, SAVERIO, ROVERATO, ALBERTO, & LUATI, ALESSANDRA. 2021. Fused graphical lasso for brain networks with symmetries. *Journal of the Royal Statistical Society: Series C*, **70**(5), 1299–1322.

ROVERATO, ALBERTO, & NGUYEN, DUNG NGOC. 2022. Model inclusion lattice of coloured Gaussian graphical models for paired data. *Pages 133–144 of:* SALMERÓN, ANTONIO, & RUMÍ, RAFAEL (eds), *Proceedings of the 11th International Conference on Probabilistic Graphical Models*. Proceedings of Machine Learning Research, vol. 186. PMLR.

ROVERATO, ALBERTO, & NGUYEN, DUNG NGOC. 2023. Exploration of the search space of Gaussian graphical models for paired data. *arXiv preprint arXiv:2303.05561*.

XIE, YUYING, LIU, YUFENG, & VALDAR, WILLIAM. 2016. Joint estimation of multiple dependent Gaussian graphical models with applications to mouse genomics. *Biometrika*, **103**(3), 493–511.

ZHANG, HONGMEI, HUANG, XIANZHENG, & ARSHAD, HASAN. 2022. Comparing Dependent Undirected Gaussian Networks. *Bayesian Analysis*, 1–26.