



Some results on the Gaussian Markov Random Field construction problem based on the use of invariant subgraphs

Juan Baz¹ · Irene Díaz² · Susana Montes¹ · Raúl Pérez-Fernández¹ 

Received: 1 October 2020 / Accepted: 1 February 2022 / Published online: 24 February 2022
© The Author(s) 2022

Abstract

The study of Gaussian Markov Random Fields has attracted the attention of a large number of scientific areas due to its increasing usage in several fields of application. Here, we consider the construction of Gaussian Markov Random Fields from a graph and a positive-definite matrix, which is closely related to the problem of finding the Maximum Likelihood Estimator of the covariance matrix of the underlying distribution. In particular, it is simultaneously required that the variances and the covariances between variables associated with adjacent nodes in the graph are fixed by the positive-definite matrix and that pairs of variables associated with non-adjacent nodes in the graph are conditionally independent given all other variables. The solution to this construction problem exists and is unique up to the choice of a vector of means. In this paper, some results focusing on a certain type of subgraphs (invariant subgraphs) and a representation of the Gaussian Markov Random Field as a Multivariate Gaussian Markov Random Field are presented. These results ease the computation of the solution to the aforementioned construction problem.

Keywords Gaussian Markov Random Field · Multivariate Gaussian Markov Random Field · Covariance matrix · Graph theory

Mathematics Subject Classification 60G60

This research has been partially supported by the Spanish Ministry of Science and Technology (TIN-2017-87600-P and PGC2018-098623-B-I00), the Research Foundation of Flanders (FWO17/PDO/160) and FICYT (FC-GRUPIN-IDI/2018/000176). Raúl Pérez-Fernández acknowledges the support of KERMIT, Department of Data Analysis and Mathematical Modelling, Ghent University, Belgium.

✉ Raúl Pérez-Fernández
perezfernandez@uniovi.es

¹ UNIMODE, Department of Statistics and O.R. and Mathematics Didactics, University of Oviedo, Calle Federico García Lorca 18, 33007 Oviedo, Spain

² Department of Computer Science, University of Oviedo, Campus de Viesques, 33004 Gijón, Spain

1 Introduction

A MRF is a random vector whose components are associated with the nodes of an undirected graph and the conditional distributions satisfy some properties (called Markov properties) over the structure of the graph. In this model, the dependence of two non-adjacent variables in the graph is explained in terms of the dependence between adjacent variables of the chains that connect these two non-adjacent variables (Rue and Held 2005). MRFs are a simple way to model spatial dependencies between the components of the random vector. For example, for modeling the evolution of a disease over some regions of a country, it is quite typical to consider an MRF over a graph in which two nodes are adjacent if the regions have a common border (see Section 4.4.2 in Rue and Held 2005).

The case in which the random vector has multivariate Gaussian distribution has attracted a lot of attention since; in such case, the MRF can be characterized by the null elements of the inverse of the covariance matrix. Furthermore, the modeling of a problem by means of a GMRF is, in general, computationally attractive, mainly because the inverse of the covariance matrix is positive-definite and typically sparse (Rue and Held 2005). The multivariate generalizations of GMRFs, called Multivariate Gaussian Markov Random Fields (MGMRFs), have also been widely studied in recent years (MacNab 2018).

In this paper, we follow the approach by Speed and Kiiveri (1986) in which they propose a GMRF construction problem over a graph. Formally, given a multivariate Gaussian distribution and an undirected graph, we search for another multivariate Gaussian distribution that keeps the same variances and the covariances between adjacent variables while satisfying the Markov properties. This problem is analogous to finding the Maximum Likelihood Estimator (MLE) of the covariance matrix of the GMRF given the sample covariance matrix of the distribution. If the GMRF model is reasonable, then this MLE will be similar to the sample covariance matrix but will also benefit from the computational advantages of having a sparse inverse. In this direction, several methods that allow to learn the underlying graph structure have been developed (see Furtlehner et al. 2021; Ferrer-Cid et al. 2021; Loftus et al. 2021 for some recent examples). Here, we solve this GMRF construction problem restricted to a certain type of subgraphs (which we will call invariant subgraphs) and then solve the whole GMRF construction problem by considering MGMRFs over forests. This will result in a significant reduction in computation time in comparison to directly solving the GMRF construction problem over the whole graph.

The remainder of the paper is organized as follows. In Sect. 2, we introduce the preliminary concepts necessary for presenting our results. The notions of invariant subgraph and complete separator are presented in Sect. 3. In Sect. 4, we define an MGMRF construction problem over a forest as a reduction in several GMRF construction problems over invariant subgraphs and we provide an easy formula for solving such construction problem. In Sect. 5, we study the complexity of the presented method and compare it with the original algorithms. Finally, in Sect. 6, we illustrate the presented results with three different examples and show the benefits of the considered approach. Some conclusions are presented in Sect. 7.

2 Preliminaries

2.1 Simple undirected finite graphs

A simple undirected finite graph $G = (V, E)$ consists of a couple of sets: a finite set of vertices (or nodes) V and a finite set of edges E , whose elements (i, j) are such that $i, j \in V$ and satisfy that $(i, j) \in E$ if and only if $(j, i) \in E$ and $(i, i) \notin E$ for any $i \in V$. The cardinal of V is called the order of the graph and is usually denoted by n and the cardinal of E is called the size of the graph and is denoted by m . The subgraph (of $G = (V, E)$) induced by $V' \subset V$ is $G' = (V', E')$ with $E' = \{(i, j) \in E \mid i, j \in V'\}$.

If $(i, j) \in E$, then i and j are said to be adjacent. The number of adjacent nodes to a node is called the degree of incidence of such node. A sequence of nodes (a_1, \dots, a_k) is called a chain from a_1 to a_k if $(a_i, a_{i+1}) \in E$ for all $i \in \{1, \dots, k-1\}$. The number of edges involved in a chain is called the length of the chain. If $a_1 = a_k$, then the chain is also called a cycle. Two chains that only share the first and the last node are called internally disjoint.

If there exists a chain between two nodes, both nodes are said to be connected. If there exist two internally disjoint chains between them, the nodes are called 2-connected. A graph is called connected if all pairs of nodes are connected and is called 2-connected if all pairs of nodes are 2-connected. A subgraph that is connected (respectively, 2-connected) and maximal with respect to this property is called a connected (respectively, 2-connected) component.

A graph is called complete if all pairs of nodes are adjacent. A graph that does not contain any cycle is called a forest and, if it is also connected, then it is called a tree.

A subset $V' \subset V$ is called a separator of $G = (V, E)$ if the subgraph induced by $V \setminus V'$ is not connected. Given three pairwise disjoint sets $V', V_1, V_2 \subset V$, such that V_1 and V_2 are non-empty, we say that V' separates V_1 and V_2 if any chain between a node in V_1 and a node in V_2 contains a node in V' .

Given two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, the Cartesian product of G_1 and G_2 is the graph $G_1 \square G_2 = (V, E)$ with $V = \{(u, v) \mid u \in V_1, v \in V_2\}$ such that (u_1, v_1) and (u_2, v_2) are adjacent if and only if it either holds that $u_1 = u_2$ and $(v_1, v_2) \in E_2$ or $v_1 = v_2$ and $(u_1, u_2) \in E_1$.

2.2 Multivariate Gaussian distribution

A continuous random vector \mathbf{X} has a multivariate Gaussian distribution if any linear combination of its components has a univariate Gaussian distribution (Mardia et al. 1979). Its joint density distribution has the following expression:

$$f(\mathbf{x}) = \frac{1}{\sqrt{|2\pi \Sigma|}} \exp\left(-\frac{(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})}{2}\right), \quad \forall \mathbf{x} \in \mathbb{R}^n,$$

where $\boldsymbol{\mu}$ is the mean vector and Σ is the covariance matrix. We denote the set of positive-definite matrices by \mathcal{P} and we require $\Sigma \in \mathcal{P}$. Given two subsets of indices A and B , we denote the corresponding submatrix of Σ by Σ_{AB} or simply by Σ_A in case $A = B$. We also denote $\mathbf{X} \setminus \mathbf{X}_A$ by \mathbf{X}_{-A} .

Given three continuous random vectors \mathbf{X} , \mathbf{Y} and \mathbf{Z} of dimensions n_X , n_Y and n_Z , respectively, with joint density function $f(\mathbf{x}, \mathbf{y}, \mathbf{z})$, \mathbf{X} and \mathbf{Y} are said to be conditionally independent given \mathbf{Z} (see, e.g., Rohatgi 1976) iff there exist $h : \mathbb{R}^{n_X+n_Z} \rightarrow [0, \infty]$ and $g : \mathbb{R}^{n_Y+n_Z} \rightarrow [0, \infty]$ such that $f(\mathbf{x}, \mathbf{y}, \mathbf{z}) = h(\mathbf{x}, \mathbf{z})g(\mathbf{y}, \mathbf{z})$, $\forall \mathbf{x} \in \mathbb{R}^{n_X}$, $\forall \mathbf{y} \in \mathbb{R}^{n_Y}$, $\forall \mathbf{z} \in \mathbb{R}^{n_Z}$. We denote the fact that \mathbf{X}_A and \mathbf{X}_B are conditionally independent given \mathbf{X}_C by $\mathbf{X}_A \perp \mathbf{X}_B \mid \mathbf{X}_C$.

In particular, the conditional independence structure of a Multivariate Gaussian distribution can be characterized by using Σ^{-1} , often referred to as the precision matrix.

Theorem 1 (Rue and Held 2005) *Let \mathbf{X} be a multivariate Gaussian random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ . For any $i \neq j$, it holds that $X_i \perp X_j \mid \mathbf{X}_{-\{i,j\}} \iff (\Sigma^{-1})_{ij} = 0$.*

2.3 The GMRF model

Let $G = (V, E)$ be a simple undirected finite graph where $V = \{1, \dots, n\}$ denotes the set of nodes and $E \subset V \times V$ denotes the set of edges. The neighborhood of a node i (i.e., the nodes that are adjacent to i) is denoted by $N(i)$. Given a random vector $\mathbf{X} = (X_1, \dots, X_n)$, the Markov properties are defined as follows:

- The pairwise Markov property: $X_i \perp X_j \mid \mathbf{X}_{-\{i,j\}}$ for any $i, j \in V$ such that $(i, j) \notin E$ and $i \neq j$.
- The local Markov property: $X_i \perp \mathbf{X}_{-\{i\} \cup N(i)} \mid \mathbf{X}_{N(i)}$ for any $i \in V$.
- The global Markov property: $\mathbf{X}_A \perp \mathbf{X}_B \mid \mathbf{X}_C$, for any pairwise disjoint $A, B, C \subset V$ with $A, B \neq \emptyset$ and where C separates A and B .

If \mathbf{X} is a multivariate Gaussian random vector, then the three properties above are equivalent (Rue and Held 2005) and if any of the properties above is satisfied, \mathbf{X} is called a Gaussian Markov Random Field (GMRF) over G . As a result of Theorem 1, given a GMRF, the Markov properties are characterized by the null elements of Σ^{-1} .

A multivariate version of the GMRFs can be also defined. Let $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ be a multivariate Gaussian random vector and $G = (V, E)$ with $V = \{1, \dots, n\}$. Similarly to a GMRF, any node can be associated with a random vector instead of a random variable. This results in the notion of Multivariate Gaussian Markov Random Field (MGMRF). In this case, the Markov properties are characterized by null submatrices of Σ^{-1} .

2.4 The GMRF construction problem

We focus on the construction of a GMRF over a graph when the covariances between adjacent variables are fixed. We start with a positive-definite matrix P of dimension $n \times n$ and we allow to change the non-fixed values in order to satisfy the Markov properties over the graph. Given the matrix P as initial data, the search for a matrix F that coincides with P for adjacent variables and whose inverse F^{-1} has zeros at the positions associated with non-adjacent variables is referred to as *the GMRF construction problem*.

Theorem 2 (Speed and Kiiveri 1986) *Let $P, R \in \mathcal{P}$ and $G = (V, E)$ be a graph. There exists a unique $F \in \mathcal{P}$ that satisfies:*

- $F_{ij} = P_{ij}$ if $(i, j) \in E$ or if $i = j$,
- $F_{ij}^{-1} = R_{ij}$ if $(i, j) \notin E$.

In particular, setting the matrix R in the theorem above as the identity matrix, the existence of a unique solution to the GMRF construction problem is assured. The generalization of this construction problem to MGMRFs is immediate, with also a unique solution. Some algorithms to compute an approximate solution to this problem are provided in Speed and Kiiveri (1986) and Wermuth and Scheidt (1977). We highlight the importance of finding the solution to this problem, since it is related to several widely studied problems.

Firstly, the problem above is equivalent to finding the MLE (Maximum Likelihood Estimation) of the covariance matrix of the GMRF given the sample covariance matrix (Dempster 1972; Xu et al. 2011; She and Tang 2019). More precisely, suppose that we have a sample of a GMRF and compute the sample covariance matrix. If we identify this sample covariance matrix with the matrix P in Theorem 2 and set R as the identity matrix, the matrix F corresponds to the MLE of the covariance matrix of the GMRF.

Proposition 1 *Let \mathbf{X} be a GMRF over G and consider a random sample of \mathbf{X} . If the matrix P in Theorem 2 is the resulting sample covariance matrix, then the matrix F in Theorem 2 is the Maximum Likelihood Estimation of the covariance matrix of \mathbf{X} .*

This result is relevant to applications such as those in Furtlehner et al. (2021), Ferrer-Cid et al. (2021) and Loftus et al. (2021). If the assumption of the data being sampled from a GMRF is reasonable, then the MLE is similar to the sample covariance matrix while also having a sparse inverse matrix. In order to decide whether a GMRF model is reasonable or not, we can use a likelihood ratio test or apply some penalty procedures (Banerjee et al. 2008).

Secondly, the problem above is also equivalent to finding the distribution that maximizes the differential entropy among all the random vectors with the variances and some of the covariances specified (since it maximizes the determinant of the associated covariance matrix Grone et al. 1984).

Proposition 2 *Let $P \in \mathcal{P}$, $G = (V, E)$ be a graph and consider the set of matrices $\mathcal{M} = \{A \in \mathcal{P} \mid A_{ij} = P_{ij} \text{ if } (i, j) \in E \text{ or } i = j\}$. The matrix F in Theorem 2 maximizes the determinant among all matrices in \mathcal{M} .*

This problem was also introduced by Dempster (1972) as a covariance selection model, which has been shown to be very useful for reducing the number of parameters in the estimation of the covariance matrix of a Multivariate Gaussian distribution (and, actually, of the exponential family) (Speed and Kiiveri 1986).

Note that there is a direct link between this problem and the construction of a GMRF from the marginal distribution of the cliques (complete subgraphs) of the graph by means of the positive definite completion of partial Hermitian matrices (Grone et al. 1984). In particular, the following result is a direct consequence of Theorem 2.

Proposition 3 *Let $G = (V, E)$ be a graph and \mathcal{C} be the set of cliques of G . Consider $\{\mathbf{X}_C \mid C \in \mathcal{C}\}$ to be a set of marginal multivariate normal distributions over the cliques of G satisfying that, for any $C_1, C_2 \in \mathcal{C}$, \mathbf{X}_{C_1} restricted to C_2 and \mathbf{X}_{C_2} restricted to C_1 are equally distributed. Let us denote by A the partial matrix such that $A_{i,j} = \text{Cov}(X_i, X_j)$ if there exists $C \in \mathcal{C}$ with $i, j \in C$ and all other elements unknown. It holds that a GMRF \mathbf{Y} over G such that $\mathbf{Y}_C = \mathbf{X}_C$ for all $C \in \mathcal{C}$ exists if and only if there exists a positive definite completion of A .*

Another well-known approach to the construction of a GMRF starts from the full conditionals (see Section 2.2.4 in Rue and Held 2005), which is equivalent to determining the mean vector and the inverse of the covariance matrix (also known as the precision matrix). The main difference with our approach is that it is necessary to make additional considerations to assure the positive definiteness of the precision matrix. In our case, given a positive definite matrix, which in real-life applications typically is the sample covariance matrix, the solution to the GMRF construction problem always exists and is unique. However, if we are working with a degenerate GMRF, in which at least one variable may be expressed as a linear combination of the other ones, working with the precision matrix is a better option (see Sect. 3 in Rue and Held 2005).

3 Invariant subgraphs and complete separators

3.1 Construction of invariant subgraphs from complete separators

Subvectors of a GMRF over a graph are not, in general, GMRFs over the subgraph induced by some of their components. For example, consider a GMRF over a tree with 3 nodes. The subgraph induced by the two nodes with degree of incidence 1 consists of two non-adjacent nodes. If the associated subvector were a GMRF over this subgraph, the associated variables would then be independent, a statement that is not true in general. For this very reason, we search for a type of subgraphs for which this property holds.

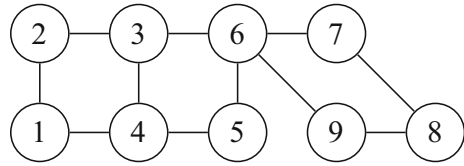
Definition 1 Let G be a graph. A subgraph G' of G is called an invariant subgraph (of G) if any GMRF over G restricted to G' is also a GMRF over G' .

The term invariant subgraph refers to the fact that, given the submatrix of the initial data matrix P associated with the invariant subgraph, the submatrix of the solution matrix is invariant to the rest of values of P . This is a very important property when we are interested in finding the MLE estimator of the covariance matrix associated with an invariant subgraph, since it states that we can restrict our attention to the associated variables rather than to all the components of the random vector.

The simplest invariant subgraphs are the complete ones, but we are interested in more general ones. In particular, we will make use of complete separators to find these invariant subgraphs.

Definition 2 Let $G = (V, E)$ be a graph. A subset $V' \subset V$ is a complete separator if it is a separator of G and the subgraph of G induced by V' is complete.

Fig. 1 Graph with complete separators



This type of subsets of nodes is also known in the literature as clique separators (Tarjan 1985; Coudert and Ducoffe 2018).

Example 1 In the graph of Fig. 1, the subgraph induced by $\{3, 4\}$ is a complete separator because it separates $\{1, 2\}$ and $\{5, 6, 7, 8, 9\}$ and, in addition, it is complete. Similarly, the subgraph induced by $\{6\}$ is a complete separator because it separates $\{1, 2, 3, 4, 5\}$ and $\{7, 8, 9\}$ and, in addition, it is complete.

The next theorem proves that, given a complete separator, two invariant subgraphs arise.

Theorem 3 Let \mathbf{X} be a GMRF over $G = (V, E)$ and $A, B, C \subset V$ be pairwise disjoint subsets satisfying $A \cup B \cup C = V$ and $A, B \neq \emptyset$. If C separates A and B and the subgraph induced by C is complete, then $\mathbf{X}_{A \cup C}$ is a GMRF over the subgraph induced by $A \cup C$ and $\mathbf{X}_{B \cup C}$ is a GMRF over the subgraph induced by $B \cup C$.

Proof We consider the global Markov property. For proving that $\mathbf{X}_{A \cup C}$ is a GMRF over the subgraph induced by $A \cup C$, it suffices to prove that, if $C' \subset A \cup C$ separates A' and B' in $A \cup C$, then C' separates A' and B' in G . Suppose that there exist $a' \in A'$ and $b' \in B'$ that are connected in $G \setminus C'$ but are not connected in $(A \cup C) \setminus C'$. We distinguish three cases:

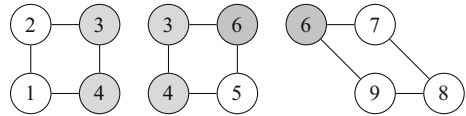
- (i) If $a', b' \in C \setminus C'$, since C is complete, then it holds that a' and b' are connected in $(A \cup C) \setminus C'$, a contradiction.
- (ii) If $a' \in A \setminus C'$ and $b' \in C \setminus C'$, then there exists a chain (a', \dots, b') between a' and b' of which at least an element is not contained in $(A \cup C) \setminus C'$. Therefore, there exists $b \in B$ with $b \in (a', \dots, b')$. Since C separates A and B , the fact that b and a' are connected implies that there exists $c \in C \setminus C'$ connected to a' . Since $C \setminus C'$ is complete, c is also connected to b' . The contradiction then follows from the fact that a' and b' are connected in $A \cup C$.
- (iii) If $a', b' \in A \setminus C'$, the contradiction is reached similarly to the previous case.

The proof for $B \cup C$ is identical to the one for $A \cup C$. □

It is concluded that, if we find a complete separator C that separates $A, B \neq \emptyset$, then the subgraphs induced by $A \cup C$ and $B \cup C$ are invariant subgraphs.

Example 2 The graph in Fig. 2 can be divided into three invariant subgraphs $\{1, 2, 3, 4\}$, $\{3, 4, 5, 6\}$ and $\{6, 7, 8, 9\}$ considering successively the complete separators $\{3, 4\}$ and $\{6\}$.

Fig. 2 Identification of invariant subgraphs of the graph in Fig. 1 using complete separators



3.2 Complete separators of low order

Finding complete separators and invariant subgraphs is not always an easy task. However, it is easy to characterize complete separators of order 0 and 1. A complete separator of order 0 arises whenever the graph is partitioned into two connected components. Complete separators of order 1 are linked to Menger’s Theorem.

Theorem 4 (Menger’s Theorem) (Menger 1927) *Let $G = (V, E)$ be a graph and $u, v \in V$ be such that $(u, v) \notin E$. The order of a minimal uv -separator of G is equal to the maximum number of internally disjoint chains that connect u and v in G .*

As a consequence of this theorem, it is concluded that we can find some invariant subgraphs just by studying disjoint chains.

Proposition 4 *Let $G = (V, E)$ be a graph and $G' = (V', E')$ be the subgraph of G induced by $V' \subseteq V$. If for every $u, v \in V'$ there exist two disjoint chains from u to v in G' and G' is maximal with respect to this property, then G' is an invariant subgraph.*

Proof If the maximal subgraph in which every two nodes are connected by two disjoint chains is the whole graph G , then the result follows. Otherwise, there exist two nodes connected by a maximum number of internally disjoint chains lower than 2. From Menger’s Theorem, these two nodes are separated by another node or by the empty set. In both cases we have a complete separator, thus two invariant subgraphs, Theorem 3. We can repeat the same process on each of the obtained invariant subgraphs several times, until all obtained invariant subgraphs are such that for every $u, v \in V'$ there exist two disjoint chains from u to v in G' and G' is maximal with respect to this property. □

Note that, if the order of all these subgraphs is greater than two, we obtain a decomposition into 2-connected components of the graph. It must be remarked that Menger’s Theorem cannot be used to study a similar case associated with 3 disjoint chains because the minimal separator may be formed by two nodes that are not adjacent, so it might not be a complete separator. As a conclusion, we expect to find complete separators in sparse graphs.

We end this section by stressing that the existence of complete separators of order 0 and 1 is crucial when studying the complete separators of the Cartesian product of two graphs $A \square B$, that can be seen as a two-dimensional graph. This is the case for the ladder graph, which is considered in Sect. 6.1. In particular, Theorem 2.2 in Anand et al. (2012) can be reformulated as follows:

Theorem 5 *Let $A \square B$ be the Cartesian product of two graphs A and B . It holds that $A \square B$ has a complete separator if and only if one of A and B is complete and the other one has a complete separator of order 0 or 1.*

4 Resolution of a MGMRF construction problem

4.1 Definition of the MGMRFs

From Theorem 3, we can find the elements of the submatrices $\Sigma_{A \cup C}$ and $\Sigma_{B \cup C}$ just by solving the GMRF construction problem over the associated subgraphs. The second step is then to find the rest of the values of Σ . Since for any $a \in A$ and $b \in B$ it holds that $(a, b) \notin E$, this can be done in an easy way by reducing the main problem to an MGMRF model. If $C \neq \emptyset$ we associate A, C, B with a 3-tree, being C associated with the node with degree of incidence 2 because \mathbf{X}_A and \mathbf{X}_B are conditionally independent given \mathbf{X}_C (since C separates A and B). Note that, if $C = \emptyset$, then the graph associated with the MGMRF only consists of two non-adjacent nodes, one for A and another one for B .

It is interesting to apply Theorem 3 iteratively whenever we identify a complete separator in $A \cup C$ or in $B \cup C$. In such case, we may solve several GMRF construction problems over the obtained invariant subgraphs.

4.2 Solving MGMRF over forests

The next step is to solve the MGMRF construction problem defined above, ultimately resulting in a solution to the initial GMRF construction problem. We will focus on the case of trees, keeping in mind that, since the covariance matrix of an MGMRF over a forest is a diagonal block matrix where the blocks are associated with the connected components of the forest, the provided results are also valid for solving MGMRF construction problems over forests. Firstly, we provide a lemma concerning the MGMRF construction problem over a tree with 3 nodes.

Lemma 1 *Let $\mathbf{X} = (\mathbf{X}_A, \mathbf{X}_C, \mathbf{X}_B)$ be an MGMRF over a 3-tree $G = (V, E)$ in which \mathbf{X}_C is associated with the node with degree of incidence 2. It holds that $\Sigma_{AB} = \Sigma_{AC} \Sigma_C^{-1} \Sigma_{CB}$.*

Proof By using the matrix formula to calculate the conditional distribution of $(\mathbf{X}_A, \mathbf{X}_B)$ given \mathbf{X}_C for a multivariate Gaussian distribution, we have that $\Sigma_{AB|C} = \Sigma_{AB} - \Sigma_{AC} \Sigma_C^{-1} \Sigma_{CB}$. The result then follows from applying that $\Sigma_{AB|C}$ is a null matrix. \square

We can repeat the 3-tree structure all over the tree to calculate the remaining elements of Σ by just operating with the matrices obtained from the solved GMRF construction problems over the invariant subgraphs.

Proposition 5 *Let \mathbf{X} be an MGMRF over a tree $G = (V, E)$. The covariance matrix between the subvectors associated with two non-adjacent nodes A and B is given by:*

$$\Sigma_{AB} = \Sigma_{K_1 K_2} \prod_{i=2}^{l-1} \Sigma_{K_i}^{-1} \Sigma_{K_i K_{i+1}}$$

where $K = (K_1, \dots, K_l)$ is the unique chain of length $l - 1$ from $K_1 = A$ to $K_l = B$.

Proof Since G is a tree, there exists a unique chain $K = (K_1, \dots, K_l)$ from $K_1 = A$ to $K_l = B$ (Kocay and Kreher 2016). The subvectors $(\mathbf{X}_A, \mathbf{X}_{K_2}, \mathbf{X}_{K_3}), (\mathbf{X}_A, \mathbf{X}_{K_3}, \mathbf{X}_{K_4}), \dots, (\mathbf{X}_A, \mathbf{X}_{K_{l-1}}, \mathbf{X}_B)$ are GMRFs over 3-trees. The result follows from applying $l - 2$ times Lemma 1. \square

Corollary 1 *Let \mathbf{X} be a GMRF over a tree $G = (V, E)$. Pearson's correlation coefficient between two variables is the product of Pearson's correlation coefficients of the adjacent variables in the unique chain that connects them.*

The case of tree graphs, although is one of the simplest cases, is interesting due to its applicability. For instance, Gaussian Markov chains are examples of GMRFs over a tree graph, which are closely related to the Kalman Filter (Kalman 1960), which has been used in many real-life applications (see, e.g., Auger et al. 2013). Moreover, any GMRF over a non-acyclic graph can be approximated by a GMRF over a tree graph, which is specially relevant to spatial pattern classification and image restoration problems (Wu and Doerschuk 1995). Regarding the construction of the matrix F in Theorem 2, from the corollary above, it is immediate to determine the covariance between two variables in linear time $O(n)$.

5 The proposed method and analysis of the computational complexity

As a conclusion of the previous results, a method to solve the GMRF construction problem by using invariant subgraphs may be defined. It can be summarized in the three following steps:

1. Decompose the graph into invariant subgraphs by using complete separators.
2. Solve the GMRF construction problem over every invariant subgraph.
3. Compute the solution for the whole graph.

For the second step, one of the original algorithms in Speed and Kiiveri (1986) and Wermuth and Scheidt (1977) is used to solve the GMRF construction problem over every invariant subgraph. These algorithms converge toward the solution by repeating a procedure iteratively. In each iteration, a loop over the non-adjacent nodes, the maximal cliques or the maximal cliques of the complementary graph, depending on the chosen algorithm, is performed.

Example 3 Consider the graph in Fig. 1 and the initial matrix P defined by $P_{ij} = 0.5$ if $i \neq j$ and $P_{ij} = 1$ if $i = j$, for any $i, j \in \{1, \dots, 9\}$. We recall the decomposition into invariant subgraphs of the graph provided in Fig. 2, corresponding to the subgraphs induced by $\{1, 2, 3, 4\}$, $\{3, 4, 5, 6\}$ and $\{6, 7, 8, 9\}$. The solution F_I of the GMRF construction problems restricted to each of the three invariant subgraphs coincides and is the following one:

$$F_I \approx \begin{pmatrix} 1 & 0.5 & 0.366 & 0.5 \\ 0.5 & 1 & 0.5 & 0.366 \\ 0.366 & 0.5 & 1 & 0.5 \\ 0.5 & 0.366 & 0.5 & 1 \end{pmatrix}$$

Therefore, some of the values of the solution matrix F are already known:

$$F \approx \begin{pmatrix} 1 & 0.5 & 0.366 & 0.5 & ? & ? & ? & ? & ? \\ 0.5 & 1 & 0.5 & 0.366 & 0.5 & ? & ? & ? & ? \\ 0.366 & 0.5 & 1 & 0.5 & 0.366 & ? & ? & ? & ? \\ 0.5 & 0.366 & 0.5 & 1 & 0.5 & 0.366 & ? & ? & ? \\ ? & ? & 0.366 & 0.5 & 1 & 0.5 & ? & ? & ? \\ ? & ? & 0.5 & 0.366 & 0.5 & 1 & 0.5 & 0.366 & 0.5 \\ ? & ? & ? & ? & ? & 0.5 & 1 & 0.5 & 0.366 \\ ? & ? & ? & ? & ? & 0.366 & 0.5 & 1 & 0.5 \\ ? & ? & ? & ? & ? & 0.5 & 0.366 & 0.5 & 1 \end{pmatrix}$$

All other elements can be computed by applying Proposition 5:

$$F \approx \begin{pmatrix} 1 & 0.5 & 0.366 & 0.5 & 0.268 & 0.232 & 0.116 & 0.085 & 0.116 \\ 0.5 & 1 & 0.5 & 0.366 & 0.232 & 0.268 & 0.134 & 0.098 & 0.134 \\ 0.366 & 0.5 & 1 & 0.5 & 0.366 & 0.5 & 0.25 & 0.183 & 0.25 \\ 0.5 & 0.366 & 0.5 & 1 & 0.5 & 0.366 & 0.183 & 0.134 & 0.183 \\ 0.268 & 0.232 & 0.366 & 0.5 & 1 & 0.5 & 0.25 & 0.183 & 0.25 \\ 0.232 & 0.268 & 0.5 & 0.366 & 0.5 & 1 & 0.5 & 0.366 & 0.5 \\ 0.116 & 0.134 & 0.25 & 0.183 & 0.25 & 0.5 & 1 & 0.5 & 0.366 \\ 0.085 & 0.098 & 0.183 & 0.134 & 0.183 & 0.366 & 0.5 & 1 & 0.5 \\ 0.116 & 0.134 & 0.25 & 0.183 & 0.25 & 0.5 & 0.366 & 0.5 & 1 \end{pmatrix}$$

5.1 Complexity of the proposed method

In order to study the complexity of the proposed method, it is necessary to examine the complexity of each one of the three sequential steps. In order to analyze the complexity we will assume that the number of adjacent nodes and the number of non-adjacent nodes grow with n^2 (see Tarjan 1985; Xu et al. 2011). This is a common assumption since the maximum number of adjacent nodes (and consequently the maximum number of non-adjacent nodes) in a graph of order n is $\frac{n(n-1)}{2}$. The same is assumed for the number of maximal cliques of the graph, whose maximum number is the number of edges. This will lead to a common complexity for the first and second cyclic algorithm in Speed and Kiiveri (1986) and the variant in Wermuth and Scheidt (1977). The complexity of the three steps involves the complexity of inverting an $n \times n$ matrix. Its value depends on the method applied but is of the order $O(n^{2+\epsilon})$, with $\epsilon > 0$ (Xu et al. 2011).

The first step of our method is based on decomposing the graph into invariant subgraphs by using complete separators, which is a purely graph-theoretical matter. This decomposition in terms of complete separators was considered in a classical paper (Tarjan 1985), where an algorithm for obtaining this decomposition is proposed. The authors prove that the computation time is $O(mn)$, being m the number of graph edges. In particular, their method is based on computing a minimal ordering of the graph (see Rose et al. 1976), and subsequently obtain the decomposition. More refined

results in Coudert and Ducoffe (2018) reduce the computation time to the complexity of multiplying an $n \times n$ matrix; thus, the complexity of this step is $O(n^{2+\epsilon})$.

For the second step, it suffices to apply one of the two cyclic algorithms in Speed and Kiiveri (1986) or the variant in Wermuth and Scheidt (1977) to all invariant subgraphs obtained in the previous step. Let us start by studying the complexity of solving the problem over the whole graph. In the three variants, an iteration of these algorithms visits all the cliques (second cyclic algorithm in Speed and Kiiveri 1986), the cliques of the complementary graph (first cyclic algorithm in Speed and Kiiveri 1986) or the edges of the complementary graph (algorithm in Wermuth and Scheidt 1977), which are of the order of m . For every element visited in the iteration, an inversion of an $n \times n$ matrix must be computed. Since the algorithm does not give the exact solution but converges toward this exact solution, the number of iterations may be determined by a maximum iteration number or a tolerance criterion. We conclude that the complexity of the algorithms is $O(mn^{2+\epsilon})$. A similar study of the complexity has been carried out in Xu et al. (2011), reaching the same conclusions. For the algorithm over the invariant subgraphs, it is convenient to consider the first cyclic algorithm in Speed and Kiiveri (1986) or the variant in Wermuth and Scheidt (1977), since they focus on the complementary of the graph. It is easy to see that two nodes $v, w \in V$ can be both contained in two different invariant subgraphs only if they are adjacent. Thus, considering an iteration over all the invariant subgraphs, the algorithms visit less cliques or edges in the complementary graph than if we consider the algorithm over the whole graph; thus, an upper bound will be m . In this case, we have to compute the inverse of the matrix associated with the invariant subgraph, with a maximum order of n . We conclude that the complexity for the second step has an upper bound of $O(mn^{2+\epsilon})$.

For the last step, we need to invert the matrices associated with the complete separators and compute a matrix multiplication. A loose upper bound for the complexity can be obtained if we suppose that there are n complete separators of maximum order (which is actually not possible). Since any complete separator is always contained in an invariant subgraph, an upper bound of their order is n . In this case, the complexity has an upper bound of $O(n^{3+\epsilon})$.

Thus, the complexity of each step of the sequential algorithm is, respectively, $O(n^{2+\epsilon})$, $O(mn^{2+\epsilon})$ and $O(n^{3+\epsilon})$. Since m grows with n^2 , then the total complexity is $O(mn^{2+\epsilon})$. Note that if we apply directly the algorithms in Speed and Kiiveri (1986) and Wermuth and Scheidt (1977), the complexity is also $O(mn^{2+\epsilon})$. We conclude that the addition of the first and third steps to the method does not increase the complexity.

5.2 Further comments on computational aspects

On the one hand, if there exists a complete separator of the graph, the dimensions of the matrices in the second step are smaller than those for the original algorithm. Since this step is the most computationally expensive, a slight decrease in the dimension of the matrix can result in a notable decrease in the computation time of the algorithm. Thus, we expect the computation time to be reduced, even for a big graph in which we can only separate a few nodes (possibly in the corners or the outskirts of the graph)

from the rest. The numerical results in Sects. 6.2 and 6.3 illustrate the improvement in the computation time for this case. In addition, the number of iterations in the algorithm used in step 2 is determined by a maximum number of iterations and a tolerance criterion. It seems reasonable that the number of iterations needed to reach the fixed tolerance criterion when we consider invariant subgraphs is smaller than in the general case.

On the other hand, if there does not exist any complete separator of the graph, no decomposition of the graph is reached and the second step consists in solving the GMRF construction problem over the whole graph, with the drawback of having computed step 1, losing time by trying to find complete separators that actually do not exist. This leads to an increase in the computation time with respect to the original algorithms. However, since the complexity of the first step is $O(n^{2+\epsilon})$, which is smaller than the complexity of solving the problem over the whole graph, $O(mn^{2+\epsilon})$, trying to find invariant subgraphs seems reasonable for graphs of large order.

We end this section by noticing that the decomposition step may be performed faster over some types of graphs. For instance, if the graph is planar then the complexity is $O(n)$ and if the graph has a bounded treewidth then the complexity is $O(n \log(n))$. We refer to (Sect. 6.1 in Coudert and Ducoffe 2018) for more information in this regard.

6 Examples

In this section, we consider three different scenarios in which the presented method simplifies the GMRF construction problem. Firstly, we consider a GMRF construction problem over the ladder graph. Secondly, we consider a generalization of the autoregressive model $AR(k)$ in which the distribution is not necessarily stationary. Finally, we consider a real-life graph representing the (peninsular) regions of Spain as the nodes and in which two nodes are adjacent if the regions have a common border. We use real data concerning the mean temperature over the years 2011–2015 to find the Maximum Likelihood Estimation of the covariance matrix.

6.1 The ladder graph

Consider the random vector $(X_1, \dots, X_n, Y_1, \dots, Y_n)$ following a multivariate Gaussian distribution. We search for a model where, for any $i \in \{1, \dots, n\}$, X_i is conditionally independent from all the other random variables but the three following ones: X_{i+1} , X_{i-1} and Y_i . The same holds for Y_i , being conditionally independent from all the other random variables but the three following ones: Y_{i+1} , Y_{i-1} and X_i . We can express the multivariate Gaussian distribution as a GMRF over a ladder graph, see Fig. 3. This structure can be interpreted as a pair of random variables studied over a period of time and measured at specific instants of time in a way such that the values of the random variables only depend on the value of the variable at the previous instant of time and the value of the other variable at the same instant of time. We assume that the variance of all variables is equal to 1, the correlation between X_i and X_{i-1} and between Y_i and

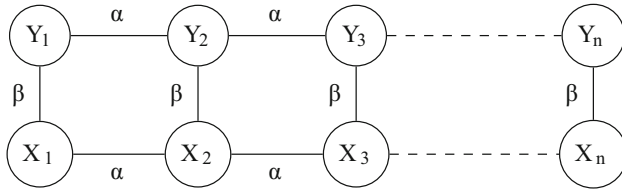


Fig. 3 Ladder graph associated with the considered GMRF

Y_{i-1} is constant and equals α for any $i \in \{2, \dots, n\}$ and the correlation between X_i and Y_i is constant and equals β for any $i \in \{1, \dots, n\}$.

Any set $\{X_i, Y_i\}$ with $i \in \{2, \dots, n - 1\}$ is a complete separator. Therefore, for any $i \in \{1, \dots, n - 1\}$, it holds that $\{X_i, Y_i, X_{i+1}, Y_{i+1}\}$ is an invariant subgraph. When solving the GMRF construction problem over $\{X_i, Y_i, X_{i+1}, Y_{i+1}\}$, we obtain the correlation between Y_i and X_{i+1} , denoted by $\phi(\alpha, x)$, which is equal to the correlation between X_i and Y_{i+1} . The value $\phi(\alpha, x)$ is the only real solution of the equation $\phi(\alpha, \beta)^3 - (\alpha^2 + \beta^2 + 1)\phi(\alpha, \beta) + 2\alpha\beta = 0$, obtained by inverting the covariance matrix and imposing the Markov properties. The following value is obtained:

$$\phi(\alpha, \beta) = \frac{\sqrt[3]{\sqrt{4(-3\alpha^2 - 3\beta^2 + 3) + 2916\beta^2\alpha^2} - 54\beta\alpha}}{3\sqrt[3]{2}} - \frac{\sqrt[3]{2}(-3\alpha^2 - 3\beta^2 + 3)}{3\sqrt[3]{\sqrt{4(-3\alpha^2 - 3\beta^2 + 3) + 2916\beta^2\alpha^2} - 54\beta\alpha}}.$$

The computation of $\phi(\alpha, \beta)$ is straightforward from the values of α and β . We may consider the MGMRF construction problem over the graph in Fig. 4, with $A_i = \{X_i, Y_i\}$ for any $i \in \{1, \dots, n\}$. From Proposition 5 and after diagonalizing the matrix on the right side, it is easy to verify that, we obtain that the covariance matrix between A_i and A_{i+d} is:

$$\begin{aligned} \Sigma_{A_i A_{i+d}} &= \begin{pmatrix} \alpha & \phi(\alpha, x) \\ \phi(\alpha, x) & \alpha \end{pmatrix} \left(\begin{pmatrix} 1 & x \\ x & 1 \end{pmatrix}^{-1} \begin{pmatrix} \alpha & \phi(\alpha, x) \\ \phi(\alpha, x) & \alpha \end{pmatrix} \right)^{d-1} \\ &= \frac{1}{2^d} \begin{pmatrix} \alpha & \phi(\alpha, x) \\ \phi(\alpha, x) & \alpha \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \left(\frac{\alpha - \phi(\alpha, x)}{1-x}\right)^{d-1} & 0 \\ 0 & \left(\frac{\alpha + \phi(\alpha, x)}{1+x}\right)^{d-1} \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \end{aligned}$$



Fig. 4 Graph associated with the MGMRF

This expression is easier to compute for any value of d than solving the GMRF construction problem over the whole graph, especially bearing in mind that the order of the graph is at least $2d$.

6.2 Non-stationary $AR(k)$ processes

In this example, we consider a generalization of the autoregression model $AR(k)$. The $AR(k)$ model considers a time series measured at specific instants of time that is stationary (i.e, it is invariant with respect to time translations) and such that the value at a certain instant of time only depends on the values of the k previous instants of time (Lindsey 2004). We consider a more general model, in which the time series is not necessarily required to be stationary. A non-stationary $AR(k)$ model may be understood as a GMRF in which each node is adjacent to the k preceding nodes and the k succeeding nodes (or by all preceding/succeeding nodes in case there are less than k preceding/succeeding nodes). Notice that any subset of k consecutive nodes is a complete separator of the graph. Thus, all subsets $\{1, \dots, k + 1\}, \{2, \dots, k + 2\}, \dots, \{n - k, \dots, n\}$ are invariant subgraphs.

For illustrative purposes, we consider a non-stationary $AR(k)$ with n variables for all combinations of $n \in \{70, 110, 150\}$ and $k \in \{3, 5, 10\}$. For setting the initial matrix, we choose a random matrix M of dimension $n \times n$ with elements ranging uniformly from 0 to 0.1. Subsequently, we compute the matrix $P = M^T + M + I_n$, where I_n is the identity matrix of dimension $n \times n$. The matrix $M + M^T$ is assured to be symmetric, whereas the addition of the term I_n was incorporated in order to make the matrix $M + M^T$ positive definite.

As we have discussed earlier on, the matrix F in Theorem 2 maximizes the determinant among all matrices verifying $F_{ij} = P_{ij}$ for any $(i, j) \in E$ or $i = j$ (Grone et al. 1984). Therefore, the evolution through time of the determinant of the matrix may be used as a tool for the comparison of the efficiency of the method. In particular, the faster the determinant increases, the faster we are approaching the solution to the GMRF construction problem.

We compute the determinant and the computation time (in seconds) of the algorithm presented by Wermuth and Scheidt (1977), both in case the decomposition into invariant subgraphs is considered and in case it is not considered for the aforementioned values of n and k . The result is illustrated in Fig. 5. It can be seen that for the method applied on the whole graph the determinant of the matrix converges smoothly toward the solution, whereas the here-presented method takes some initial time to decompose the graph into invariant subgraphs but attains the optimal solution immediately after this decomposition has been obtained. The difference between both methods seems to become larger for greater values of n .

In case the considered model is stationary, for the solution to the GMRF construction problem to be a stationary covariance matrix it is necessary for the initial matrix to be stationary. For this purpose, we may first estimate a stationary covariance matrix (see for instance Eldar et al. 2020) and then use the presented algorithm in order to impose the graph structure. Notice that in this case we can provide an explicit expression of the covariance between non-adjacent variables similarly as in Sect. 6.1.

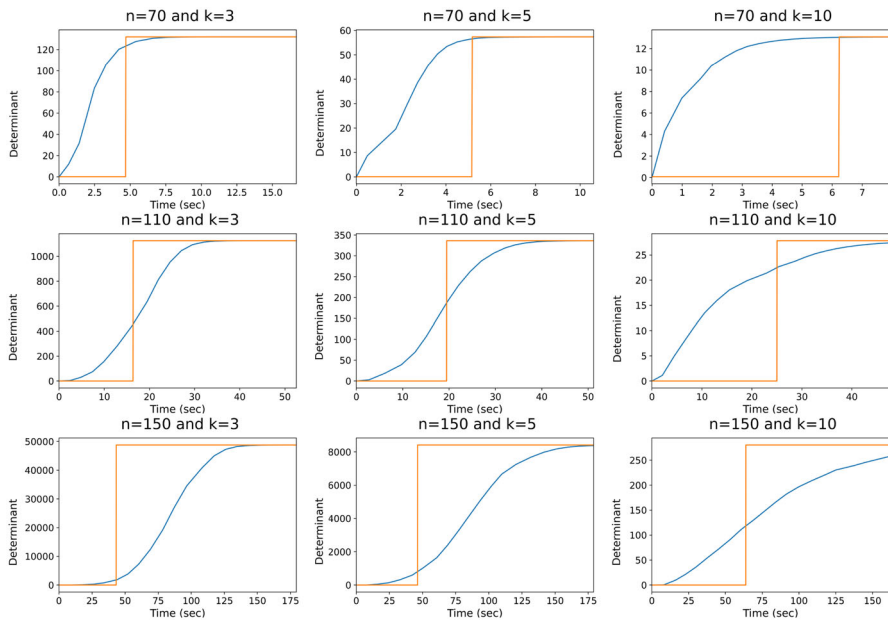


Fig. 5 Convergence speed of the determinant of the approximated MLE matrix for the classical algorithm (blue line) and the proposed method (orange line) for the computation of the MLE of the covariance matrix of the GMRF associated with the non-stationary $AR(k)$ process with $n \in \{70, 110, 150\}$ and $k \in \{3, 5, 10\}$. Bigger determinant implies less distance to exact the value of the MLE (color figure online)

6.3 A real-life graph

Let \mathbf{X} be a multivariate Gaussian random vector associated with the mean temperature over a year of the (peninsular) regions of Spain. A map of the (peninsular) regions of Spain may be found on the left-hand side of Fig. 6. A reasonable assumption may be to consider that the mean temperature of a region only depends directly on the mean temperatures of the regions that share a border with it, just as in the example in Section 4.4.2 of Rue and Held (2005). The associated graph is presented on the right-hand side of Fig. 6.

The decomposition of the graph into invariant subgraphs is represented by the following subsets of nodes: $\{1, 2, 8\}$, $\{2, 3, 8\}$, $\{3, 4, 8\}$, $\{4, 5, 6, 8, 9\}$, $\{6, 7, 13\}$, $\{8, 10, 12\}$ and $\{6, 8, 11, 12, 13, 14, 15\}$. Notice that the order of the whole graph is 15, but the largest order of all the invariant subgraphs is 7.

Next, we consider the mean temperature through the years 2011–2015, as provided by the Instituto Nacional de Estadística (INE 2016), taking as reference the meteorological observatories located at the capital of the regions (with the exception of Extremadura, for which we have considered Badajoz). From these data, we construct the sample covariance matrix and solve the GMRF construction problem, both considering and not considering the decomposition into invariant subgraphs. The comparison of the computation time (in s) and the value of the determinant of the matrix at the current iteration for the classical algorithm in Wermuth and Scheidt (1977) and our

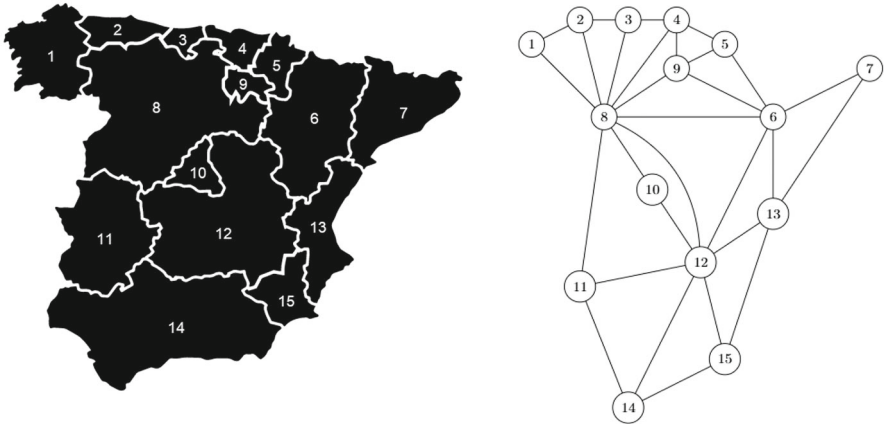


Fig. 6 Map of the (peninsular) regions of Spain (left) and associated graph (right)

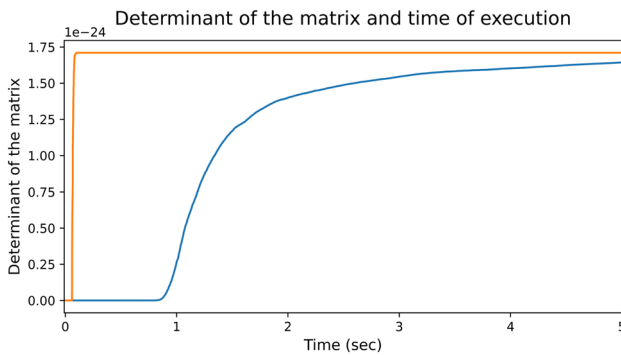


Fig. 7 Convergence speed of the determinant of the approximated MLE matrix for the classical algorithm (blue line) and the proposed method (orange line) for the computation of the MLE of the covariance matrix of the GMRF associated with the mean temperature of the regions of Spain. Bigger determinant implies less distance to the exact value of the MLE (color figure online)

proposed method is illustrated in Fig. 7. As can be observed, the convergence to the solution is faster in the case in which the graph is decomposed into invariant subgraphs.

We also want to stress that considering smaller regions (such as the provinces of Spain) makes it more difficult for complete separators to exist, mainly because in a planar graph any complete separator must have order 4 or less, as a consequence of Kuratowski's Theorem (Kuratowski 1930).

7 Conclusions

We have proposed a method to facilitate the search for the solution to the Gaussian Markov Random Field construction problem over a graph by independently considering several Gaussian Markov Random Field construction problems over invariant subgraphs and a Multivariate Gaussian Markov Random Field construction problem

over forests. We have provided a way to find these invariant subgraphs by identifying complete separators of the graph, and we have studied the easiest cases in which the complete separators are of order 0 and 1. In particular, we have illustrated the utility of this proposal by considering a Gaussian Markov Random Field construction problem over a ladder graph, a non-stationary $AR(k)$ process and a real-life problem regarding the mean temperature of the (peninsular) regions of Spain.

We have compared the complexity of the proposed method and the original algorithms in Speed and Kiiveri (1986) and Wermuth and Scheidt (1977), concluding that the decomposition of the graph in invariant subgraphs and the reconstruction of the solution for the initial problem do not increase the complexity. The numerical results illustrate that the proposed method has a lower computation time than the original algorithms for the considered examples. In case there does not exist a complete separator of the graph, both methods are equivalent, with the inconvenience of the here-presented method requiring to check the existence of complete separators at first. However, we believe that the cost of checking for the existence of complete separators is affordable, considering the potential gain. Furthermore, it should be borne in mind that the decomposition of the graph into complete separators only needs to be done once if several GMRF construction problems over the same graph are considered.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

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

- Anand BS, Balakrishnan K, Changat M, Peterin I (2012) Atoms and clique separators in graph products. *Appl Anal Discrete Math* 6:46–62
- Auger F, Hilairat M, Guerrero JM, Monmasson E, Orłowska-Kowalska T, Katsura S (2013) Industrial applications of the kalman filter: a review. *IEEE Trans Ind Electron* 60(12):5458–5471
- Banerjee O, El Ghaoui L, d'Aspremont A (2008) Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *J Mach Learn Res* 9:485–516
- Coudert D, Ducoffe G (2018) Revisiting decomposition by clique separators. *SIAM J Discrete Math* 32(1):682–694
- Dempster AP (1972) Covariance selection. *Biometrics* 28(1):157–175
- Eldar YC, Li J, Musco C, Musco C (2020) Sample efficient toeplitz covariance estimation. In: Proceedings of the fourteenth annual ACM-SIAM symposium on discrete algorithms. SIAM, pp 378–397
- Ferrer-Cid P, Barcelo-Ordinas JM, Garcia-Vidal J (2021) Graph learning techniques using structured data for iot air pollution monitoring platforms. *IEEE Internet Things J* 8:13652–13663
- Furtlehner C, Lasgouttes JM, Attanasi A, Pezzulla M, Gentile G (2021) Short-term forecasting of urban traffic using spatio-temporal markov field. *IEEE Trans Intell Transport Syst*. <https://doi.org/10.1109/TITS.2021.3096798>
- Grone R, Johnson CR, Sá EM, Wolkowicz H (1984) Positive definite completions of partial Hermitian matrices. *Linear Algebra Appl* 58:109–124

- INE (2016) Estadística de Vigilancia del Clima. <https://www.ine.es/daco/daco42/bme/c19.pdf>. Accessed 17 June 2021
- Kalman RE (1960) A new approach to linear filtering and prediction problems. *J Basic Eng* 82:35–45
- Kocay W, Kreher DL (2016) *Graphs, algorithms, and optimization*. CRC Press, Boca Raton
- Kuratowski K (1930) Sur le probleme des courbes gauches en topologie. *Fundam Math* 15(1):271–283
- Lindsey JK (2004) *Statistical analysis of stochastic processes in time*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge
- Loftus M, Hassouneh SAD, Yooseph S (2021) Bacterial associations in the healthy human gut microbiome across populations. *Sci Rep* 11(1):1–14
- MacNab YC (2018) Some recent work on multivariate Gaussian Markov random fields. *TEST* 27(3):497–541
- Mardia KV, Kent JT, Bibby JM (1979) *Multivariate analysis*. Academic Press INC, San Diego
- Menger K (1927) Zur allgemeinen kurventheorie. *Fundam Math* 10(1):96–115
- Rohatgi VK (1976) *Markov chains and stochastic stability*. Wiley, New York
- Rose DJ, Tarjan RE, Lueker GS (1976) Algorithmic aspects of vertex elimination on graphs. *SIAM J Comput* 5(2):266–283
- Rue H, Held L (2005) *Gaussian Markov random fields: theory and applications*. CRC Press, Boca Raton
- She Y, Tang S (2019) Iterative proportional scaling revisited: a modern optimization perspective. *J Comput Graph Stat* 28(1):48–60
- Speed TP, Kiiveri HT (1986) Gaussian Markov distributions over finite graphs. *Ann Stat* 14:138–150
- Tarjan RE (1985) Decomposition by clique separators. *Discrete Math* 55(2):221–232
- Wermuth N, Scheidt E (1977) Algorithm as 105: fitting a covariance selection model to a matrix. *J R Stat Soc: Ser C (Appl Stat)* 26(1):88–92
- Wu CH, Doerschuk PC (1995) Tree approximations to markov random fields. *IEEE Trans Pattern Anal Mach Intell* 17(4):391–402
- Xu PF, Guo J, He X (2011) An improved iterative proportional scaling procedure for gaussian graphical models. *J Comput Graph Stat* 20(2):417–431

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