Statistics

A Journal of Theoretical and Applied Statistics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/gsta20

Gaussian-related undirected graphical models for circular variables

Anna Gottard & Agnese Panzera

To cite this article: Anna Gottard & Agnese Panzera (2024) Gaussian-related undirected graphical models for circular variables, Statistics, 58:5, 1224-1247, DOI: 10.1080/02331888.2024.2400179

To link to this article: https://doi.org/10.1080/02331888.2024.2400179

STATISTICS
2024, VOL. 58, NO. 5, 1224–1247
https://doi.org/10.1080/02331888.2024.2400179

Taylor & Francis
Taylor & Francis Group

Gaussian-related undirected graphical models for circular variables

Anna Gottard and Agnese Panzera

Department of Statistics, Computer Science, Applications, University of Florence, Florence, Italy

ABSTRACT

Graphical models are pivotal in studying the conditional independence structure of a set of random variables. Circular variables, arising in several contexts and fields, are characterized by periodicity. Models for studying the dependence/independence structure of circular variables are under-explored but of increasing interest. This paper delves into two multivariate circular distributions, the Wrapped Normal and the Inverse Stereographic Normal distribution as undirected graphical models. For each of these distributions, we study their key properties with respect to conditional independence and introduce specific classes of graphical models. The usefulness of the proposal is shown by modelling the conditional independence among dihedral angles that play a critical role in defining the three-dimensional structure and functionality of proteins. This can provide valuable insights, for instance, into the multiform protein folding understanding.

ARTICLE HISTORY
Received 23 April 2024
Accepted 23 August 2024

KEYWORDS
Conditional independence; multivariate circular distributions; protein folding problem; toroidal data

2020 MATHEMATICS SUBJECT CLASSIFICATIONS
62H22; 62H11

1. Introduction

A circular observation can be regarded as a point on the circumference of the unit circle, as a unit vector on the plane or as a unit complex number. Once both an origin and a sense of rotation have been chosen, a circular observation can be measured by an angle which, in radians, ranges from 0 to 2π or, equivalently, from $-\pi$ to π. Typical examples of circular data include bird migration directions from the point of release, directions of the winds and marine currents, angles in the polypeptide chains forming proteins, but also the time of the day when a given event occurs. As an example, see Figure 1 where circular data are the times of the day on a 24-h clock, with the origin at 24 (0 o'clock). The time 8 o'clock can be represented as the angle $2\pi/3$ in radians ranging from 0 to 2π. This can be simply achieved by multiplying the time by $2\pi/24$.

The special nature of circular data lies in their *periodicity*, i.e. a circular observation measured by an angle of θ radians corresponds, on the unit circle, henceforth defined as $\mathbb{T} = [-\pi, \pi)$, to a circular observation measured by an angle of $\theta + 2\pi k$ radians, where k is a whatever integer number.

CONTACT Anna Gottard anna.gottard@unifi.it Department of Statistics, Computer Science, Applications, University of Florence, V.le Morgagni, 59, Florence 50134, Italy

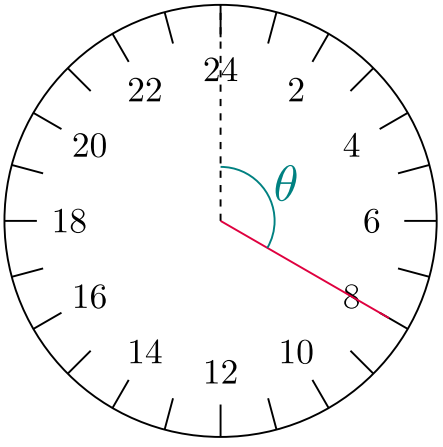


Figure 1. Example of the time of day depicted on a 24-h clock, where 8 o'clock corresponds to $\theta = \frac{2\pi}{3}$.

The periodicity sets apart circular statistics from standard methods designed for Euclidean data. For instance, consider finding the average of two angles measuring 0.1 radians and $2\pi - 0.1$ radians (close to zero on the unit circle). One would expect the average of these two angles to be zero. However, their arithmetic mean is π radians, which is on the opposite side of the unit circle from 0. The right way to average angles is to compute instead the *circular mean*, which is the arctangent of the ratio between the mean of the sines and the mean of the cosines of the angles. Circular statistics collects statistical methods suitable for dealing with circular data. Probability distributions of circular random variables can be defined according to various approaches. In both the *wrapping* and the *embedding* approaches, the circular distribution is linked, albeit in a different way, to a distribution on the real line. Specifically, the *wrapping* approach constructs circular random variables as the modulo 2π version of the real ones, with distributions obtained by wrapping their real counterparts around the circle. Conversely, the *embedding* approach obtains circular distributions as radial projections onto the unit circle of distributions on the real plane. A comprehensive account on circular statistics and distributions is provided, among others, by Jammalamadaka and Sengupta [1] and Mardia and Jupp [2]. Some recent advances are collected by Ley and Verdebout [3,4].

In some applications, the focus is on p angles, which can be regarded as points on the surface of a p-dimensional torus, $\mathbb{T}^p$, obtained by the p-fold Cartesian product of unit circles. A typical example of toroidal data arises in the three-dimensional protein structure, summarized by a sequence of angles. Despite the advances in circular statistics, both the distributions on $\mathbb{T}^p$ and the models involving p circular variables seem to be not deeply explored in the literature whenever $p > 2$.

In some fields of application, it might be of interest to study the relationship among random angles. Graphical models [see, for instance, 5] are a powerful probabilistic tool for analyzing conditional independences between random variables. This class of multivariate models expresses the conditional independence structure of p variables by a graph $\mathcal{G} = (V, E)$, where $V = \{1, \ldots, p\}$ is a set of nodes and $E \subset V \times V$ is a set of edges. The nodes in V represent the random variables, and the missing edges in E represent conditional independence statements. The associated joint probability distribution factorizes

according to the graph structure. A graphical model is called *undirected graph model* or *Markov network* whenever $\mathcal{G}$ is undirected, that is, when $E \subseteq \{\{i,j\} : i,j \in V, i \neq j\}$ is a set of unordered pairs. This model treats all variables on equal footing. The most commonly used undirected graphical model is the *concentration graph model* or *Gaussian graphical model* that assumes a joint Gaussian distribution for the p variables. In these models, learning the missing edges is equivalent to identifying the zero elements in the inverse of the covariance matrix.

Graphical models have been applied in a wide variety of research fields, such as systems engineering, pattern recognition, machine learning and artificial intelligence, healthcare, biology and computational biology, omics sciences, psychology, social sciences, and finance. Some interesting recent applications of graphical models can be found in Mencarini et al. [6], Giudici and Spelta [7], Warnick et al. [8], Ahelegbey et al. [9], Kotiang and Eslami [10], McNally [11], and Ren et al. [12], among many others.

Graphical models tailored for circular variables seem to be a topic of growing interest. The first proposal, up to our knowledge, is provided by Boomsma et al. [13]. They studied the sequence and the structure of a protein, by using two classes of models, combining Hidden Markov Models, viewed as Bayesian networks, with distributions defined on the sphere and on the bidimensional torus. In the framework of undirected graphical models, Razavian et al. [14] firstly studied conditional independence among angles having a p-variate von Mises distribution and represented it via a factor graph. Klein et al. [15] introduced a class of graphical models for data lying on $\mathbb{T}^p$, called *torus graphs*, to tackle the problem of identifying phase coupling among oscillatory signals recorded from multiple electrodes in the brain. These models are defined not on a specific distribution but as members of a full exponential family with pairwise interactions. In particular, Klein et al. [15] focused on some subfamilies of torus graphs, with specific types of associations, that are useful for their motivating application.

In this paper, we propose two novel classes of undirected graphical models for variables on $\mathbb{T}^p$, which are related, albeit in different manners, to the *classical* Gaussian graphical model. Specifically, the first class is based on the Wrapped Normal distribution, while the second is based on the Inverse Stereographic projected Normal distribution [16]. The Wrapped Normal distribution, despite the complexity of its distributional form, has received increasing interest in the current literature, mainly due to computational developments that make feasible the inference on its parameters. We show that this distribution does not admit conditional independence. Consequently, we propose a class of graphical models for its unwrapped counterpart, where conditional independence can occur. The idea of using unwrapped circular data has been exploited also by Greco et al. [17] for parameter estimation in clustering based on a mixture model of multivariate Wrapped Normal distributions and by Marques et al. [18] in regression models for spatially dependent circular response data based on wrapped Gaussian processes. The multivariate Inverse Stereographic projected Normal distribution is a relatively new multivariate distribution for toroidal data. This distribution assumes that the stereographic projection of a toroidal variable has a multivariate Gaussian distribution. Then, the probability density function of the inverse map differs from multivariate Gaussian density only for a factor which does not include any parameter. Consequently, this distribution shares the properties of the Gaussian graphical models. In addition, for this distribution, we invoke classical theory also to propose a class of related semi-parametric graphical models.

The flexibility to select from a broader class of distributions within the framework of graphical models is especially useful in practical applications. A key application of graphical models for circular variables that will be used in this paper as an illustrative example, concerns the protein folding problem. Traditionally, proteins were thought of as rigid structures, with conformation determining their function. This gave rise to the protein folding problem, a central issue in molecular biology that seeks to understand how a protein's amino acid sequence dictates its three-dimensional structure. While machine learning has made major advances in this topic, as indicated by reviews such as Wei and Zou [19] and Noé et al. [20], recent discoveries highlight that some proteins dynamically change their shape to perform their functions. The existence of such proteins suggests that the relationship between amino-acid sequence and structure is more complex than previously believed. Graphical models can provide useful insights into understanding this complex relationship.

The proposed class of models finds applications in many fields, in addition to the structural analysis of proteins presented here. In computer vision and robotics, these models can analyze the relationships among the directions of multiple moving robots or drones together with to other moving agents. Hence, graphical models for angles might assist in the development of systems for autonomous navigation, where discerning the directions of surrounding objects, that cannot be assumed in independent movement, is essential. Other promising fields of application derives from the use of wireless tracking devices, for monitoring human or animal movement. Here graphical models can help to study birds genetic migration patterns, but also to understand the human kinematics and kinetics. Graphical models might also provide valuable insights into the analysis of human posture to understand how to prevent back pain, neck pain, or postural instability.

The paper is organized as follows. Section 2 presents some properties of the Wrapped Normal related distributions and a subsequent class of graphical models. Section 3 introduces a class of graphical models for the Inverse Stereographic Normal distribution and a related semi-parametric version. Section 4 contains an illustrative example where the introduced models are employed to describe the relationships among angles in the structure of a shape-shifting protein. Section 5 concludes with some final considerations. The appendices collect the proofs of some results (Appendix 1), some further insights and some simulations for the Wrapped Normal model (Appendix 2), and a summary of the key properties of the considered distributions (Appendix 3).

2. Wrapped normal related graphical models

The *wrapping* approach is one of the main approaches for defining probability distributions on the circle and the torus. Circular random variables are obtained as the modulo 2π version of the real ones, with distributions obtained by wrapping their real counterparts around the circle. Therefore, any *linear* random variable X can be transformed into a circular variable Θ by reducing it by modulo 2π, i.e. $\Theta = X(\mathrm{mod} 2\pi)$. A vector of random angles $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_p)'$ has a Wrapped Normal (WN) distribution, $\boldsymbol{\Theta} \sim WN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with parameters $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\boldsymbol{\Sigma} \in \mathrm{Sym}_+^{p\times p}$, the set of symmetric positive definite matrices of order p, if $\boldsymbol{X} = \boldsymbol{\Theta} + 2\pi \boldsymbol{K}$ is such that $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{K}$ is a *latent* random vector taking values on the p-dimensional integer lattice $\mathbb{Z}^p$. Hence, $\boldsymbol{X}$ determines both $\boldsymbol{\Theta}$ and

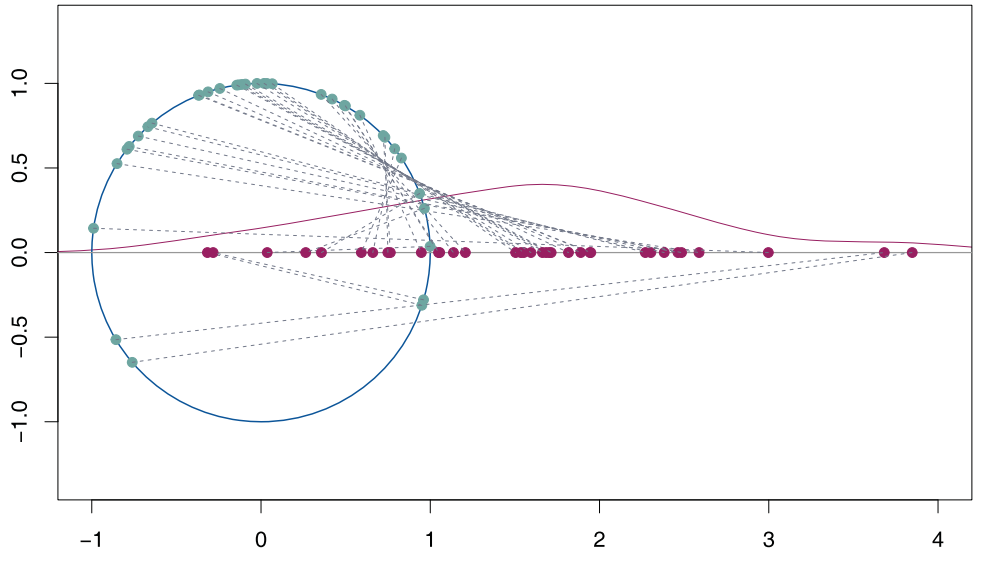


Figure 2. Visual representation of data from a WN distribution: purple points represent real-line data from a Normal distribution and teal points represent the corresponding points on the circle after mod2π operation.

$\boldsymbol{K}$ via modulo operation. For further details on this approach to define the WN distribution, see [21]. Examples of data generated by a univariate WN distribution are depicted in Figure 2.

Denoting by $f_{\boldsymbol{X}}$ the density function of $\boldsymbol{X}$, the joint density of $(\boldsymbol{\Theta}, \boldsymbol{K})$ is again $f_{\boldsymbol{X}}$, while the density of $\boldsymbol{\Theta}$, at $\boldsymbol{\theta} \in \mathbb{T}^p$, is obtained by marginalization as $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^p} f_{\boldsymbol{X}}(\boldsymbol{\theta} + 2\pi\boldsymbol{k})$. To the best of our knowledge, the properties of marginal and conditional independence of this distribution have not been deeply investigated. To initiate this exploration, we provide the interpretation of the parameters in the matrix $\boldsymbol{\Sigma}$ of a WN distribution via the following result which uses the complex form representation of $\boldsymbol{\Theta}$.

Result 2.1: *Let* $\boldsymbol{\Theta} \sim WN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, *and* $\boldsymbol{Z} = e^{\mathrm{i}\boldsymbol{\Theta}}$, *with* $\mathrm{i}^2 = -1$. *Then*

$$e^{-(\boldsymbol{\Sigma})_{ij}} = \frac{\mathbb{E}[Z_i Z_j]}{\mathbb{E}[Z_i]\mathbb{E}[Z_j]}.$$

Proof: See Appendix 1. ■

From Result 2.1, we can deduce that if $Z_i \perp\!\!\!\perp Z_j$, then $(\boldsymbol{\Sigma})_{ij} = 0$ and therefore $X_i \perp\!\!\!\perp X_j$. Conversely, this suggests that, if $X_i \perp\!\!\!\perp X_j$, then Z_i and Z_j are mean independent. To further delve into independence and conditional independence for Wrapped Normal random angles, consider the partition of $\boldsymbol{\Theta} \sim WN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ into the components $\boldsymbol{\Theta}_A$ and $\boldsymbol{\Theta}_B$ taking values respectively on $\mathbb{T}^q$ and $\mathbb{T}^{p-q}$, for integer $q < p$, and let $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ be partitioned accordingly, i.e. ,

$$\boldsymbol{\Theta} = \begin{pmatrix} \boldsymbol{\Theta}_A \\ \boldsymbol{\Theta}_B \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AB} \\ \boldsymbol{\Sigma}_{BA} & \boldsymbol{\Sigma}_{BB} \end{pmatrix}. \tag{1}$$

Then, the following results hold.

Result 2.2: *Given the vector of random angles* $\boldsymbol{\Theta} \sim WN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, for any non-empty subset* A *of* $\{1, \ldots, p\}$ *of cardinality* $q < p$*,*

$$\boldsymbol{\Theta}_A \sim WN_q(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{AA}).$$

Proof: See Appendix 1. ■

Result 2.2 assesses that the Wrapped Normal distribution is closed under marginalization. The next result deals with the conditional distribution of $\boldsymbol{\Theta}_A$ given the pair $(\boldsymbol{\Theta}_B, \boldsymbol{K}_B)$ which univocally determines the unwrapped variable $\boldsymbol{X}_B$, also providing insight on marginal independence of Wrapped Normal random angles.

Result 2.3: *Given the vector of random angles* $\boldsymbol{\Theta} \sim WN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, for disjoint non-empty subsets* A *and* B *of* $\{1, \ldots, p\}$ *with respective cardinalities* $q < p$ *and* $p-q$

$$\boldsymbol{\Theta}_A \mid \boldsymbol{\Theta}_B, \boldsymbol{K}_B \sim WN_q(\boldsymbol{\mu}_{A|B}, \boldsymbol{\Sigma}_{A|B}),$$

where $\boldsymbol{K}_B$ *is such that* $\boldsymbol{X}_B = \boldsymbol{\Theta}_B + 2\pi \boldsymbol{K}_B$*, with* $\boldsymbol{X}_B \sim N_{p-q}(\boldsymbol{\mu}_B, \boldsymbol{\Sigma}_B)$*, and*

$$\boldsymbol{\mu}_{A|B} = \boldsymbol{\mu}_A - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}(\boldsymbol{x}_B - \boldsymbol{\mu}_B), \quad \text{and} \quad \boldsymbol{\Sigma}_{A|B} = \boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\boldsymbol{\Sigma}_{BA}.$$

Moreover, $\boldsymbol{\Theta}_A \perp\!\!\!\perp \boldsymbol{\Theta}_B$ *iff* $\boldsymbol{\Sigma}_{AB}$ *is the null matrix* $\mathbf{0}_{q\times(p-q)}$*.*

Proof: See Appendix 1. ■

The next result deals with toroidal conditional distributions for jointly Wrapped Normal random angles.

Result 2.4: *Given the vector of random angles* $\boldsymbol{\Theta} \sim WN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, the conditional density of* $\boldsymbol{\Theta}_A \mid \boldsymbol{\Theta}_S$*, with* A*,* S *being disjoint subsets of* $\{1, \ldots, p\}$*, with cardinality* q *and* s *respectively,* $s \leq p - q$*, is*

$$f_{\boldsymbol{\Theta}_A|\boldsymbol{\Theta}_S}(\boldsymbol{\theta}_A \mid \boldsymbol{\theta}_S) = \sum_{\boldsymbol{k}_A \in \mathbb{Z}^q} \sum_{\boldsymbol{k}_S \in \mathbb{Z}^s} f_{\boldsymbol{X}_A|\boldsymbol{X}_S}(\boldsymbol{\theta}_A + 2\pi \boldsymbol{k}_A \mid \boldsymbol{\theta}_S + 2\pi \boldsymbol{k}_S) w_S(\boldsymbol{\theta}_S, \boldsymbol{k}_S)$$

where

$$w_S(\boldsymbol{\theta}_S, \boldsymbol{k}_S) = \frac{f_{\boldsymbol{X}_S}(\boldsymbol{\theta}_S + 2\pi \boldsymbol{k}_S)}{\sum_{\boldsymbol{k}_S \in \mathbb{Z}^s} f_{\boldsymbol{X}_S}(\boldsymbol{\theta}_S + 2\pi \boldsymbol{k}_S)}.$$

Proof: See Appendix 1. ■

Result 2.4 shows that the conditional toroidal density is obtained by wrapping a mixture of the corresponding real distributions. Consequently, the resulting conditional distribution is not a Wrapped Normal. This issue has been pointed out, in the different settings of toroidal diffusion processes, also by Garcýa-Portugués et al. [22].

Now, letting A, C and S be non-empty disjoint subsets of $\{1,\dots,p\}$, with respective cardinalities q, c and s such that $q+c+s \le p$, then

$$f_{\boldsymbol{\Theta}_A\boldsymbol{\Theta}_C|\boldsymbol{\Theta}_S}(\boldsymbol{\theta}_A,\boldsymbol{\theta}_C \mid \boldsymbol{\theta}_S) = \sum_{\boldsymbol{k}_A\in\mathbb{Z}^q}\sum_{\boldsymbol{k}_C\in\mathbb{Z}^c}\sum_{\boldsymbol{k}_S\in\mathbb{Z}^s} f_{\boldsymbol{X}_A\boldsymbol{X}_C|\boldsymbol{X}_S}(\boldsymbol{\theta}_A+2\pi\boldsymbol{k}_A,\boldsymbol{\theta}_C+2\pi\boldsymbol{k}_C \mid \boldsymbol{\theta}_S+2\pi\boldsymbol{k}_S)\cdot w_S(\boldsymbol{\theta}_S,\boldsymbol{k}_S).$$

If $\boldsymbol{\Theta}_A \perp\!\!\!\perp \boldsymbol{\Theta}_C \mid \boldsymbol{\Theta}_S$, then we should have $f_{\boldsymbol{\Theta}_A\boldsymbol{\Theta}_C|\boldsymbol{\Theta}_S}(\boldsymbol{\theta}_A,\boldsymbol{\theta}_C \mid \boldsymbol{\theta}_S) = f_{\boldsymbol{\Theta}_A|\boldsymbol{\Theta}_S}(\boldsymbol{\theta}_A \mid \boldsymbol{\theta}_S) f_{\boldsymbol{\Theta}_C|\boldsymbol{\Theta}_S}(\boldsymbol{\theta}_C \mid \boldsymbol{\theta}_S)$. Unfortunately, because of the sum over $\boldsymbol{k}_S$, there is no value or combination of values of the parameters ensuring conditional independence.

2.1. Unwrapped normal graphical model

In terms of graphical models, the main issue for the WN distribution is not the lack of closedness under conditioning but rather the lack of conditional independence relationships. The lack of value or combination of values of the parameters ensuring conditional independence can also be explained as the confounding due to the marginalization over the vector of winding numbers. See, among many, Wermuth [23] for the effect of marginalization on undirected graph models. This behaviour prevents this distribution from being a graphical model, apart from the case of complete or null graphs. However, it is straightforward to define a graphical model for the unwrapped variables $\boldsymbol{X} = \boldsymbol{\Theta} + 2\pi\boldsymbol{K}$.

Definition 2.1 (Unwrapped Normal graphical model): Let $\mathcal{G} = (V, E)$ be an undirected graph with set of vertices $V = \{1,\dots,p\}$ and set of undirected edges $E \subset V \times V$. The p-variate random variable $\boldsymbol{X} = \boldsymbol{\Theta} + 2\pi\boldsymbol{K}$ is an Unwrapped Normal graphical model with respect to $\mathcal{G}$ if $\boldsymbol{\Theta} \sim WN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and

$$\left(\boldsymbol{\Sigma}^{-1}\right)_{ij} = 0 \quad \text{for all} \quad \{i,j\} \notin E.$$

The unwrapped variables $\boldsymbol{X}$ have a multivariate Gaussian distribution whose conditional independence structure can be described by a *concentration graph* [5].

Notice that, while the distribution of $\boldsymbol{X}$ factorizes with respect to $\mathcal{G}$, the distribution of $\boldsymbol{\Theta}$ can be written as a *wrapped factorization*. This is the sum over $\boldsymbol{k}$ of the factorized distribution of $\boldsymbol{X}$. For instance, let A, C and S be sets forming a partition of the vertex set V, with respective cardinalities q, c and s. Assume that S separates A from C, that is, every path in $\mathcal{G}$ between a node in A and a node in C passes through elements in S, which is called the *separator* set. According to the global Markov property, $\boldsymbol{X}_A$ is independent of $\boldsymbol{X}_C$ given $\boldsymbol{X}_S$, with the wrapped factorization resulting in

$$\begin{aligned} f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) &= \sum_{\boldsymbol{k}_A\in\mathbb{Z}^q}\sum_{\boldsymbol{k}_C\in\mathbb{Z}^c}\sum_{\boldsymbol{k}_S\in\mathbb{Z}^s} f_{\boldsymbol{X}}(\boldsymbol{\theta}_A+2\pi\boldsymbol{k}_A,\boldsymbol{\theta}_C+2\pi\boldsymbol{k}_C,\boldsymbol{\theta}_S+2\pi\boldsymbol{k}_S) \\ &= \sum_{\boldsymbol{k}_S\in\mathbb{Z}^s}\sum_{\boldsymbol{k}_A\in\mathbb{Z}^q}\sum_{\boldsymbol{k}_C\in\mathbb{Z}^c} f_{\boldsymbol{X}_A|\boldsymbol{X}_S}(\boldsymbol{\theta}_A+2\pi\boldsymbol{k}_A \mid \boldsymbol{\theta}_S+2\pi\boldsymbol{k}_S) \\ &\quad \cdot f_{\boldsymbol{X}_C|\boldsymbol{X}_S}(\boldsymbol{\theta}_C+2\pi\boldsymbol{k}_C \mid \boldsymbol{\theta}_S+2\pi\boldsymbol{k}_S)\cdot f_{\boldsymbol{X}_S}(\boldsymbol{\theta}_S+2\pi\boldsymbol{k}_S). \end{aligned}$$

Consequently, in virtue of Results 2.3 and 2.4, if $\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_C \mid \boldsymbol{X}_S$, then $\boldsymbol{\Theta}_A \perp\!\!\!\perp \boldsymbol{\Theta}_C \mid \boldsymbol{X}_S$, or equivalently $\boldsymbol{\Theta}_A \perp\!\!\!\perp \boldsymbol{\Theta}_C \mid \boldsymbol{\Theta}_S, \boldsymbol{K}_S$, while it is not implied that $\boldsymbol{\Theta}_A \perp\!\!\!\perp \boldsymbol{\Theta}_C \mid \boldsymbol{\Theta}_S$.

When $(\boldsymbol{\Sigma})_{ii}$ is small enough for any $i \in \{1, \ldots, p\}$, then the distribution of $\boldsymbol{\Theta}$ can be safely approximated by a $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution [see, for instance, 21]. Moreover, if $(\boldsymbol{\Sigma}_S)_{ii}$ is small enough for any $i \in \{1, \ldots, s\}$, then

$$\|f_{\boldsymbol{\Theta}_A|\boldsymbol{\Theta}_S}(\boldsymbol{\theta}_A \mid \boldsymbol{\theta}_S) - f_{\boldsymbol{\Theta}_A|\boldsymbol{X}_S}(\boldsymbol{\theta}_A \mid \boldsymbol{x}_S)\|_\infty \to 0, \tag{2}$$

that is the distribution of $\boldsymbol{\Theta}_A \mid \boldsymbol{\Theta}_S$ can be approximated by the Wrapped Normal distribution of $\boldsymbol{\Theta}_A \mid \boldsymbol{X}_S$. In such a situation, it holds that $w_S(\boldsymbol{\theta}_S, \boldsymbol{k}_s)$ goes to 0 for any $\boldsymbol{k}_s \neq \boldsymbol{0}_s$. An example of the effect of the magnitude of variability of the separators on conditional independence is presented in Appendix 2.1.

Finally, whenever in addition also $(\boldsymbol{\Sigma}_{AC|S})_{ii}$ is small enough for any $i \in \{1, \ldots, q + c\}$, then

$$\|f_{\boldsymbol{\Theta}_A\boldsymbol{\Theta}_C|\boldsymbol{\Theta}_S}(\boldsymbol{\theta}_A, \boldsymbol{\theta}_C \mid \boldsymbol{\theta}_S) - f_{\boldsymbol{X}_A\boldsymbol{X}_C|\boldsymbol{X}_S}(\boldsymbol{\theta}_A, \boldsymbol{\theta}_C \mid \boldsymbol{\theta}_S)\|_\infty \to 0. \tag{3}$$

The condition for (2) is less stringent than the condition for (3), as it admits nodes that are not separators, such as, for instance, singletons, to have larger variance. However, the graph of the unwrapped variables approximately describes the conditional independence structure of the angles whenever the variances of the separators in $\mathcal{G}$ are small enough, that is (2) or (3) hold.

2.1.1. Some issues on structural learning

Regarding the estimation task for the Unwrapped Normal graphical model, we need to consider that the unwrapped variables are actually unobserved, as the vector $\boldsymbol{K}$ is unobservable. The concentration matrix driving the conditional independence structure of the Unwrapped Normal graphical model depends on the parameters shared by the distributions of $\boldsymbol{X}$ and $\boldsymbol{\Theta}$. It then can be estimated using the observed angles. As well known, the maximization of the complete likelihood is infeasible for the Wrapped Normal distribution because of the infinite sums. A simple possible strategy can be based on an approximate maximum likelihood approach by replacing the infinite sums with finite sums. In addition, a profile version can be obtained by plugging in the circular sample mean for $\boldsymbol{\mu}$, which is not a parameter of interest in this framework. For $p = 1$, Mardia and Jupp [2] suggests that the likelihood function can be adequately approximated simply by setting $k = 0$, but this approximation appears inadequate for $p > 1$. As discussed in Appendix 2.2, the approximation using the sum over $\boldsymbol{k} \in \{-1, 0, 1\}^p$ works quite well also for moderately large variances (2π). Conversely, the approximation using $\boldsymbol{k} = \boldsymbol{0}_p$ provides good estimates only for very small variances (for example, 0.0001 when $p = 5$). As a possible initial check, the sample circular variances [2], computed component-wise, can provide indicative values of the magnitude of the variances.

Further estimation techniques for the multivariate Wrapped Normal distribution has been proposed by Nodehi et al. [24], who introduce two new algorithms, respectively based on Expectation-Maximization and Classification Expectation–Maximization methods. A robust alternative of these procedures is provided by Saraceno et al. [25]. In their simulation study, Nodehi et al. [24] noted that for moderate p and variances smaller than $\pi^2/4$, the EM based algorithms perform similarly to the direct maximization of the approximate log-likelihood function.

Unfortunately, this estimation procedure does not scale efficiently to high-dimensional settings, as it requires evaluating each of the n data points across $|\mathcal{K}|^p$ configurations, where

$|\mathcal{K}|$ is the cardinality of the subset of $\mathbb{Z}$ used in the approximated likelihood function. For instance, for the data analyzed in Section 4, with $n = 80$, $p = 8$ and $\mathcal{K} = \{-1, 0, 1\}$, the computation involves a matrix of dimension 524,880 × 8. The procedure proposed by Nodehi et al. [24] is computationally more efficient and provides acceptable solutions in moderately higher dimensional settings.

Once estimated the parameters of the joint distribution, a typical inferential task in graphical models is learning the structure of the graph. For Unwrapped Normal graphical models, edge selection can be achieved by testing the off-diagonal elements of the inverse of the variance-covariance matrix via, for instance, Wald-type tests using the asymptotic standard errors from the Hessian matrix. The overall error rate for incorrect edge inclusion can be controlled by using, for example, the Holm correction.

An evaluation of this procedure through Monte Carlo simulations is reported in Appendix 2.3.

3. Inverse Stereographic Gaussian graph models

The primary drawback of Unwrapped Gaussian graphical models lies in their limited scalability in handling high-dimensional data. This section introduces a novel class of graphical models tailored for high-dimensional contexts. These models can utilize all the inferential procedures specific to concentration graph models, thereby enhancing their applicability in high-dimensional settings.

A vector of random angles $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_p)'$ has an Inverse Stereographic Normal (ISN) distribution, $\boldsymbol{\Theta} \sim ISN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\boldsymbol{\Sigma} \in \mathsf{Sym}_+^{p\times p}$, if the probability density function of $\boldsymbol{\Theta}$ at $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)' \in \mathbb{T}^p$ is

$$f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = (2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{u}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{u}-\boldsymbol{\mu})\right) \prod_{j=1}^{p} \frac{1}{1+\cos(\theta_j)}, \qquad (4)$$

where $\boldsymbol{u} = (u_1, \ldots, u_p)'$ with $u_j = \tan(\theta_j/2)$ being the *stereographic projection* of the vector $(\cos(\theta_j), \sin(\theta_j))'$. For further details, see Selvitella [16]. Examples of data generated by an ISN distribution when $p = 1$ are provided by Figure 3.

Given the stereographic projection $\boldsymbol{U} = \tan(\boldsymbol{\Theta}/2)$, if $\boldsymbol{\Theta} \sim \mathrm{ISN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\boldsymbol{U} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and the parameters of the ISN distribution are the location and scale parameters of the distribution of $\boldsymbol{U}$. For $\boldsymbol{\Theta} \sim \mathrm{ISN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, using partitions as in (1), it holds that

$$\boldsymbol{\Theta}_A \sim \mathrm{ISN}_q(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{AA}), \quad \text{and} \quad \boldsymbol{\Theta}_A \mid \boldsymbol{\Theta}_B \sim \mathrm{ISN}_q(\boldsymbol{\mu}_{A|B}, \boldsymbol{\Sigma}_{A|B}),$$

where

$$\boldsymbol{\mu}_{A|B} = \boldsymbol{\mu}_A - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}(\boldsymbol{u}_B - \boldsymbol{\mu}_B) \quad \text{and} \quad \boldsymbol{\Sigma}_{A|B} = \boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\boldsymbol{\Sigma}_{BA},$$

with $\boldsymbol{u}_B$ having as its jth element the tangent of the halved jth entry of $\boldsymbol{\theta}_B$. Consequently, if $\boldsymbol{\Sigma}_{AB} = \mathbf{0}_{q\times(p-q)}$, then $\boldsymbol{\Theta}_A \perp\!\!\!\perp \boldsymbol{\Theta}_B$. Moreover, the following result on conditional independence holds.

Result 3.1: *Let* $\boldsymbol{\Theta} \sim \mathrm{ISN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. *Consider* $\boldsymbol{\Theta}_A$, $\boldsymbol{\Theta}_C, \boldsymbol{\Theta}_S$ *where A, C and S are non-empty disjoint subsets of* $\{1, \ldots, p\}$ *with respective cardinalities q, c and s such that* $q + c + s \leq p$.

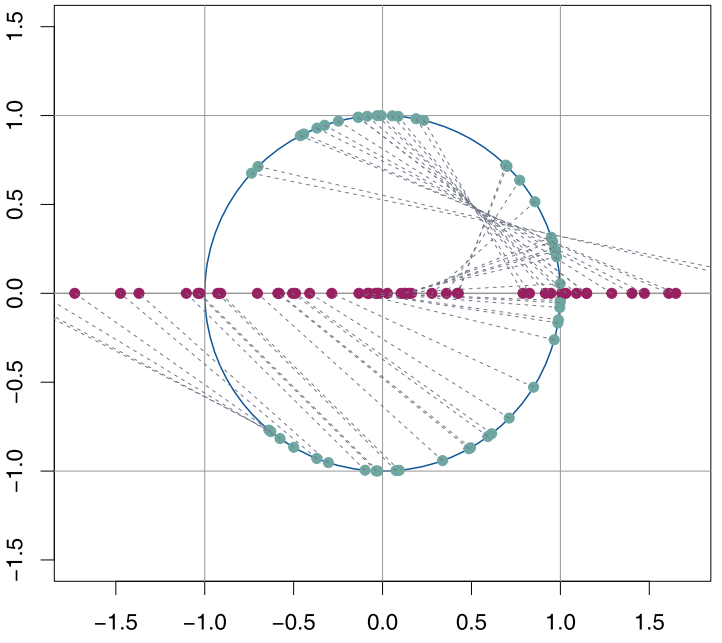


Figure 3. Visual representation of the stereographic projection: teal points represent points on the circle from an ISN distribution and purple points are the corresponding points projected on the real-line.

Then $\boldsymbol{\Theta}_A \perp\!\!\!\perp \boldsymbol{\Theta}_C \mid \boldsymbol{\Theta}_S$ *iff*

$$(\boldsymbol{\Sigma}_{\mathrm{ACS}})^{-1}_{AC} = \mathbf{0}_{q\times c}.$$

Proof: See Appendix 1. ■

When $\theta_j = -\pi$ for at least one j, $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ has removable singularities. A possible way to remove these singularities, suggested by Selvitell [16], is to put at 0 the values of the density at these points. For our purpose, it is convenient to replace the value of the density at singularity points with the value of the density when $\theta_j = -\pi + \epsilon_j$ with a non-zero, small enough ϵ_j. Using this latter convention, the resulting density is positive on $\mathbb{T}^p$, and the Hammersley–Clifford Theorem [5] can be invoked to get the following definition.

Definition 3.1 (Inverse Stereographic Gaussian graphical model): Let $\mathcal{G} = (V, E)$ be an undirected graph with set of vertices $V = \{1, \ldots, p\}$ and set of undirected edges $E \subset V \times V$. The vector of random angles $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_p)'$ is an Inverse Stereographic Gaussian graphical model with respect to $\mathcal{G}$ if $\boldsymbol{\Theta} \sim \mathrm{ISN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and

$$\left(\boldsymbol{\Sigma}^{-1}\right)_{ij} = 0 \quad \text{for all} \quad \{i, j\} \notin E.$$

The assumption of strict positivity of $f_{\boldsymbol{\Theta}}$ implies in addition that $\boldsymbol{\Theta}_A \perp\!\!\!\perp \boldsymbol{\Theta}_C \mid \boldsymbol{\Theta}_S$ whenever S separates A and C in $\mathcal{G}$ for all disjoint subsets A, C and S of V (global Markov property). Since the stereographic projection is a diffeomorphism, the factorization property for decomposable graphs directly follows by the respective property of the Gaussian

graphical model [see for details, 5]. Here, both the estimation and the edge selection can be carried out by using classical approaches for Gaussian graphical models as those described, among others, in Córdoba et al. [26]. A drawback of the ISN distribution is the lack of invariance under rotation. Thus, the definition of the model is conditional on the choice of the origin. Typically, this issue is avoided by assuming $\boldsymbol{\mu} = \mathbf{0}$. For further solutions, see [16], Section 7.

In many practical situations, the joint Gaussianity assumption could not be plausible. In the Euclidean setting, a possible way to overcome this issue is to resort to more flexible distributions, such as the Nonparanormal distribution [see 27]. We define a toroidal counterpart of the Nonparanormal distribution as follows.

Definition 3.2 (Inverse Stereographic Nonparanormal distribution): We say that the vector of random angles $\boldsymbol{\Theta} = (\Theta_1, \dots, \Theta_p)'$ has an Inverse Stereographic Nonparanormal distribution if there exists a set of functions $\boldsymbol{h} = \{h_1, \dots, h_p\}$ such that $\boldsymbol{h}(\boldsymbol{U}) = (h_1(U_1), \dots, h_p(U_p))'$, with $U_j = \tan(\Theta_j/2)$, satisfies $\boldsymbol{h}(\boldsymbol{U}) \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Notice that, if $\boldsymbol{h}(\boldsymbol{U}) \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\boldsymbol{U}$ has a Nonparanormal distribution, $\boldsymbol{U} \sim NPN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{h})$, and $\boldsymbol{\Theta}$ has an Inverse Stereographic Nonparanormal (ISNPN) distribution, $\boldsymbol{\Theta} \sim \text{ISNPN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{h})$, with density function, at $\boldsymbol{\theta} \in \mathbb{T}^p$,

$$f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{h}(\boldsymbol{u}) - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{h}(\boldsymbol{u}) - \boldsymbol{\mu})\right) \prod_{j=1}^{p} \frac{|h_j'(u_j)|}{1 + \cos(\theta_j)}. \quad (5)$$

As in the Euclidean case, to assure the identifiability of the above density function, we require that the transformation h_j preserve means and variances for all j, i.e. ,

$$\mathbb{E}[h_j(U_j)] = \mathbb{E}[U_j] = \mu_j, \quad \mathbb{V}[h_j(U_j)] = \mathbb{V}[U_j] = (\boldsymbol{\Sigma})_{jj}.$$

From (5), it can be seen that when $\boldsymbol{\Theta} \sim \text{ISNPN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{h})$, then $\Theta_i \perp\!\!\!\perp \Theta_j \mid \boldsymbol{\Theta}_{V/\{i,j\}}$ iff $(\boldsymbol{\Sigma}^{-1})_{ij} = 0$. This enables us to define the following class of models.

Definition 3.3 (Inverse Stereographic Nonparanormal graphical model): Let $\mathcal{G} = (V, E)$ be an undirected graph with set of vertices $V = \{1, \dots, p\}$ and set of undirected edges $E \subset V \times V$. The vector of random angles $\boldsymbol{\Theta} = (\Theta_1, \dots, \Theta_p)'$ is an Inverse Stereographic Nonparanormal graphical model with respect to $\mathcal{G}$ if $\boldsymbol{\Theta} \sim \text{ISNPN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and

$$\left(\boldsymbol{\Sigma}^{-1}\right)_{ij} = 0 \quad \text{for all} \quad \{i, j\} \notin E.$$

For the above class of models, the global Markov property is implied by using the same convention on the singularity points as for the Inverse Stereographic Gaussian graphical models. The estimation of $\boldsymbol{h}$ and $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ does not require *ad hoc* procedures as it can be carried out by using the approach proposed for the classical Nonparanormal graphical model in Liu et al. [27] over the realizations of U_j and $\widehat{h}(U_j)$, respectively.

The estimation and graph learning methods for both the Inverse Stereographic Normal models and their semi-parametric counterparts adhere to those used in classical Gaussian graphical models, thus we omit their discussion here. For instance, we suggest referring to

the graph learning techniques outlined in Mazumder and Hastie [28], Meinshausen and Bühlmann [29], Li and Maathuis [30] and the references therein, for efficient and scalable algorithms for graph learning.

4. An illustrative example for structural bioinformatics

This section presents the application of the proposed graphical models in structural bioinformatics, a field where accurately predicting the three-dimensional structure of a protein is crucial. In this context, understanding the conditional independence structure of a protein's dihedral angles can be a valid aid, especially for proteins which exhibit more than one conformation, like the Methionine–enkephalin (Menk).

Menk, a pentapeptide endogenous opioid, is predominantly found in the human central nervous system and gastrointestinal tract. Besides having an analgesic activity, Menk is involved in the control of respiratory, cardiovascular, gastrointestinal functions, and neuroendocrine regulation. Marcotte et al. [31] carried out two NMR experiments to measure Menk's three-dimensional structure in different model membrane systems, focusing on the impact of membrane composition on peptide conformation. They investigated zwitterionic (PC) bicelles and negatively charged bicelles (Bic/PG). The corresponding data sets, named 1PLW and 1PLX, respectively, are available at the RCSB Protein Data Bank (`https://www.rcsb.org`). In each of the two experiments, they collected $n = 80$ models of the Menk measured in fast-tumbling bicelles using multidimensional 1H NMR.

Here, we apply the proposed classes of models to the dihedral angles of the shape-changing Menk protein, primarily to illustrate the methodology rather than for detailed analysis. In the graphical models, each node represents a specific dihedral angle of an amino acid in the protein. As a motivation for this choice, consider the Ramachandran plot, a traditional visualization method for backbone dihedral angles ϕ and Ψ, typically without distinguishing the different amino acids. Figure 4 illustrates the Ramachandran plots of the dihedral angles ϕ and Ψ for Glycine residues in Menk, measured by Marcotte et al. [31]. Menk comprises two Glycine residues, named Gly1 and Gly2. Figure 4(b) presents the Ramachandran plot, distinguishing between Gly1 and Gly2 in both 1PLW and 1PLX experiments. In contrast, Figure 4(a) collectively shows all Glycine angles from the two residues and experiments. The comparison of the two plots indicates a varying dependence between the dihedral angles across amino acids and experimental conditions. This variation suggests that investigating the dependence structure among each dihedral angle of each amino acid could yield valuable insights.

4.1. Unwrapped Gaussian graph model for Menk

In this section, we separately study the data sets corresponding to the two experiments, assuming a p-variate Wrapped Normal distribution for the $p = 8$ dihedral angles forming the protein structure. For the estimation task, we employ the approach of direct maximization of the approximate likelihood outlined in Section 2.1.1.

Preliminary variance estimates, using sample circular variances, are relatively small, ranging from 0.001 to 0.036 for 1PLX and from 0.003 to 1.856 for 1PLW angles. As discussed in Section 2.1.1, these estimates indicate that the data are far from converging to

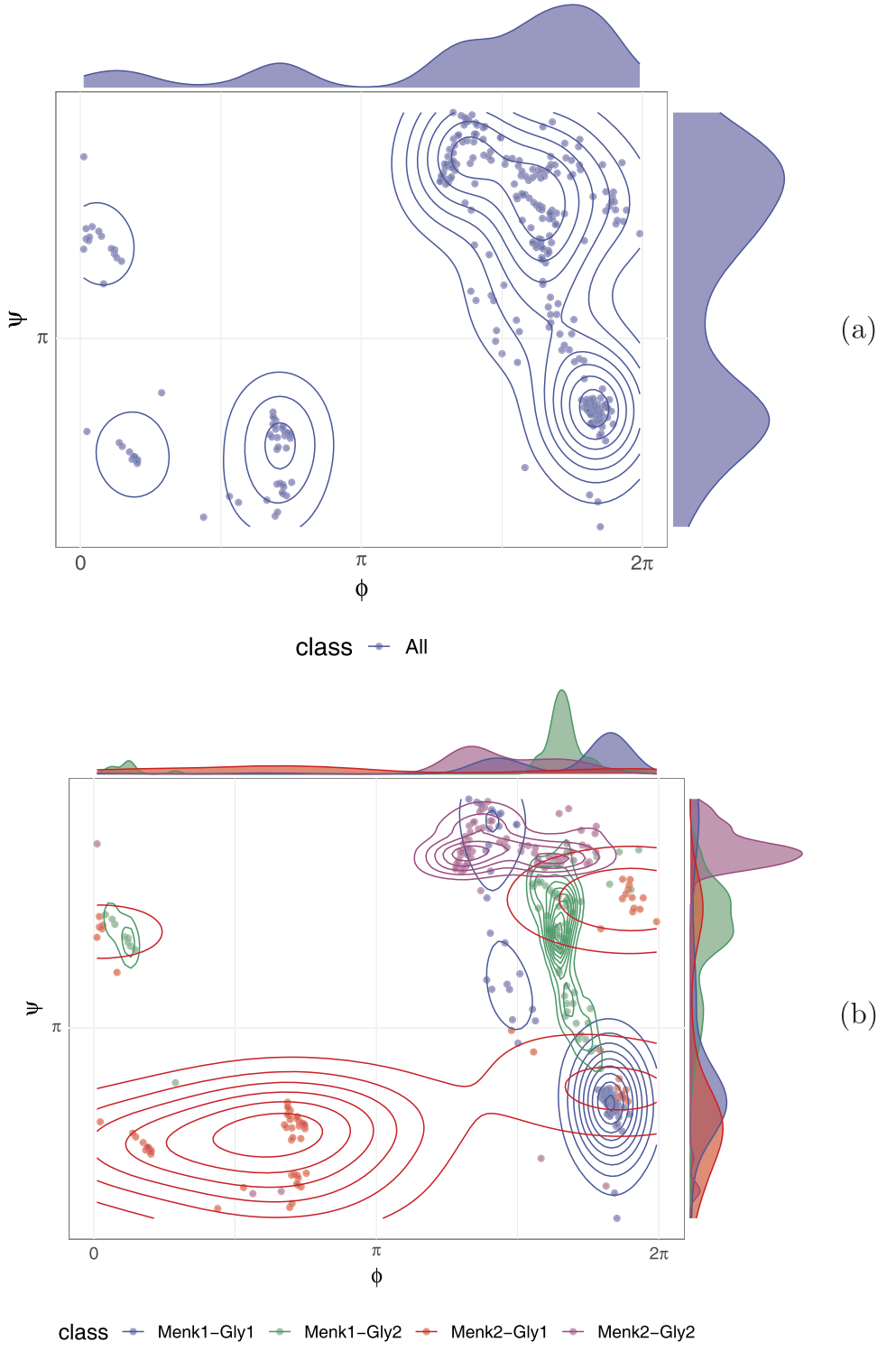
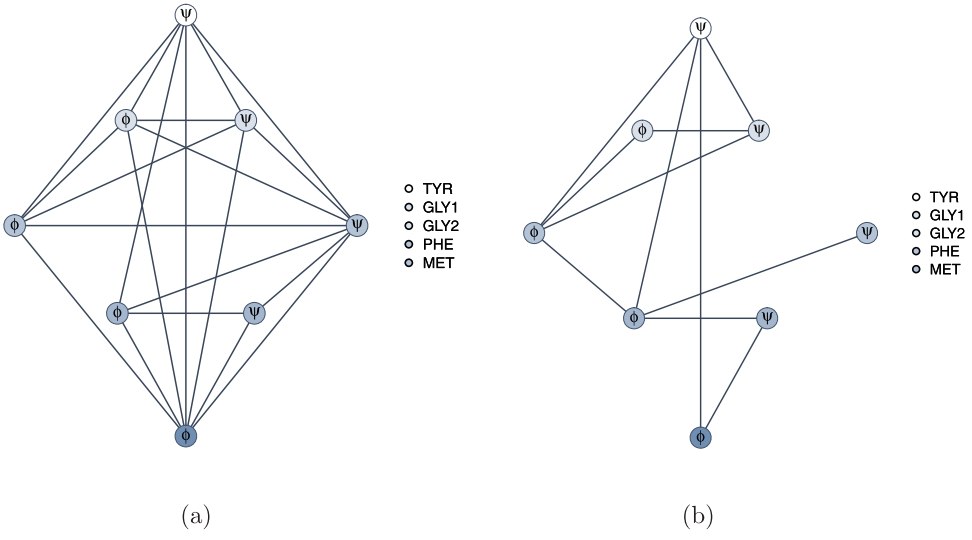


Figure 4. Ramachandran plots of (a) the dihedral angles of Glycine in Menk and (b) distinguishing the two amino acids (Gly1 and Gly2), and the two experiments (Menk1 = 1PLX, Menk2 = 1PLW).



(a) (b)

Figure 5. Selected Unwrapped graphs of the Menk angles for 1PLX (a) and for 1PLW (b).

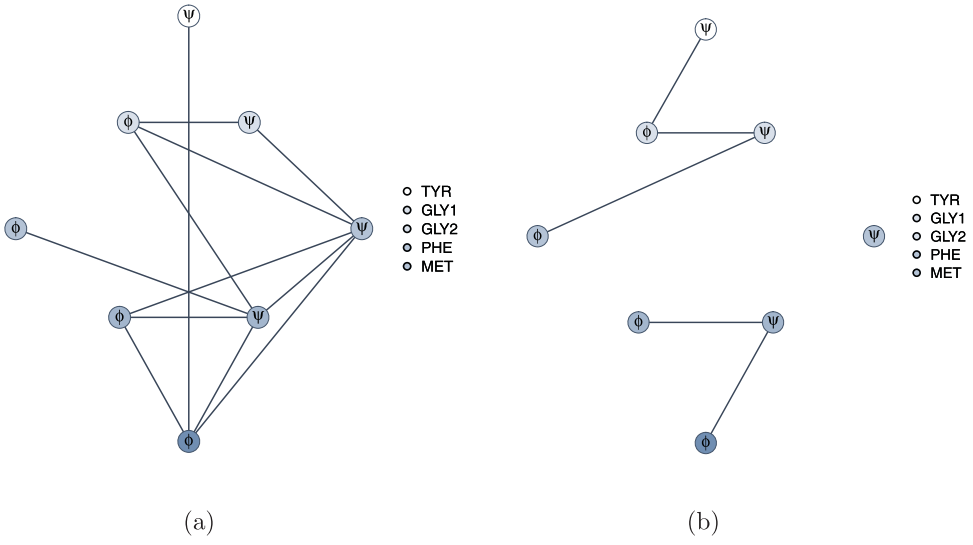
the Multivariate Normal distribution, but it is reasonable to truncate the winding numbers at ± 1. Edge selection is accomplished by individually testing each edge and applying the Holm correction to control the overall significance level, set at 0.05. The resulting unwrapped graphs are displayed in Figure 5. Nodes with the same colour belong to the same amino acid. These graphs seem to confirm the influence of membrane composition on peptide conformation. Only a few independences are detected for the 1PLX version of the Menk. In particular, it is interesting to note that Ψ_{GLY2} is an important hub in 1PLX as all the other angles depend on it. On the contrary, in 1PLW, Ψ_{GLY2} is independent of all the other angles given ϕ_{PHE}.

Under the assumed model, these graphs may be suggestive that Menk could adopt several conformations according to the membrane environments, not only in terms of means and variances of the dihedral angles, as noted by Marcotte et al. [31], but also in the independence structure among the angles. In particular, the conformation of 1PLW seems substantially less rigid than that of 1PLX.

4.2. Inverse stereographic nonparanormal graph model for Menk

In this section, we consider the Inverse Stereographic model. In particular, by applying the inverse stereographic projection to the collected data, the Shapiro–Wilks test rejects Normality both on the univariate and the joint distributions. Consequently, we assume here that the Menk dihedral angles have the ISNPN distribution defined in (5), whose transformation function $\boldsymbol{h}$ is estimated by using the same approach as in Liu et al. [27].

Among the possible ways to learn the ISNPN graphical model, we use maximum likelihood estimation with a SINful approach for edge selection [see, 32], with $\alpha = 0.1$. As an alternative, one could use, for instance, the graphical lasso [33] or its extensions [see, for instance, 28, and the references therein] to perform estimation and edge selection in a single step whenever the dimension of the protein is large.



(a) (b)

Figure 6. Selected Inverse Stereographic Nonparanormal graphs for 1PLX (a) and for 1PLW (b).

The resulting graphs are depicted in Figure 6. It is interesting to note that, under this model assumption, Ψ_{GLY2} remains the most important hub in 1PLX, as it has the largest number of connections. Conversely, in 1PLW, Ψ_{GLY2} is a singleton and is therefore independent of all the other angles both conditionally and marginally. The 1PLW structure is extremely sparse, with dependencies seemingly following the primary structure of the protein. Here $(\Psi_{\text{TYR}}, \phi_{\text{GLY1}}, \Psi_{\text{GLY1}}, \phi_{\text{GLY2}}) \perp\!\!\!\perp (\phi_{\text{PHE}}, \Psi_{\text{PHE}}, \phi_{\text{MET}})$, marginally. According to the Global Markov property, it is also true, for instance, that $\Psi_{\text{TYR}} \perp\!\!\!\perp \Psi_{\text{GLY1}} \mid \phi_{\text{GLY1}}$.

Despite the obvious disparities between the graphs selected under the semi-parametric ISNPN and those under the WN assumption, the conditional independence structure in the two experiments clearly likewise differs in terms of sparsity levels, as foreseen by Marcotte et al. [31], who noted that structural differences between the conformers of Menk are not surprising. They proved that Enkephalins are flexible pentapeptides and their conformation is dependent on their environment. Our analysis corroborates their results by showing that these conformations are distinct not only in their three-dimensional structures but also in the relationships among the angles formed by the conformations themselves. This result underscores the complexity of protein dynamics and highlights the importance of studying the conditional independence structure of the dihedral angles when analyzing the functional implications of conformational variability. Taking into consideration that existing algorithms for protein predictions, such as AlphaFold by DeepMind, can predict only static protein structures, our insight can help researchers to understand, predict and manipulate flexible protein structures. This understanding might be relevant for drug design and understanding disease mechanisms where protein interactions play a pivotal role.

5. Conclusions

The interest in the conditional independence structure of circular variables is shared by many scientific fields and has been strongly growing over the last years. For instance, in

biochemistry, the structure of molecules such as proteins, DNA, and RNA can be described in terms of angles. Graphical models whose nodes represent these angles may lead to crucial contributions in this field as well as in the understanding of protein structure. Despite their potential, graphical models for angular variables seem to be under-studied. This work introduces some classes of graphical models for angular multivariate data in some ways related to the Gaussian distribution. The proposed models enrich the classes of graphical models involving distribution in the families of wrapped and projected distributions. The previously defined models rely on distribution directly defined on the torus, according to the so-called intrinsic approach. The fact that both the distributions considered in this work are related to the Gaussian one links these new classes of models to the most popular and well-studied graphical models in the Euclidean setting. A summary of key properties related to graphical models of the distributions considered in this article is provided in Table A1 of Appendix 3.

The Unwrapped Normal graphical model represents the first class of undirected graphical models defined using the wrapping approach. While studying the conditional densities for the Wrapped Normal distribution, we show that the winding number in the conditional set prevents any proper factorization, making it impossible to specify conditional independence starting from the circular joint distribution. This issue rules out the definition of a related graphical model. To circumvent this limitation, we propose the Unwrapped Normal graphical model, which uses the distribution of the unwrapped variables. Although these unwrapped variables are unobserved, their conditional independence parameters can be estimated from the observed angles with Wrapped Normal distribution. However, the presence of doubly infinite summations in the density formulation makes the estimation task computational challenging and not scalable. Despite this well-known inferential problem, the Wrapped Normal remains an important member of the class of wrapped distributions, deserving consideration in this context.

The Inverse Stereographic Normal graphical model represents a further class of models related to the Gaussian distribution. To relax the normality assumption, we also propose a more flexible semi-parametric model called the Inverse Stereographic Nonparanormal graphical model. The most relevant aspect of both these classes of models is that they inherit the properties of the concentration graph models. Consequently, these models can benefit from all the broad and well-established literature on the Gaussian graphical models. This also includes all the *classical* procedures for parameter estimation and edge selection, even for high-dimensional settings. As a drawback, the Inverse Stereographic Normal distribution is not invariant with respect to the choice of origin. This problem stems directly from the projection, as the circle is homeomorphic to the real projective line but not to the real line. Consequently, the specification of the probabilistic model is conditional on the choice of the origin. One can set the origin according to subject matter considerations, or, as suggested by Selvitella [16], one can consider a richer family of distributions by including a translation parameter so that the origin can be data-driven.

The choice between the two proposed graphical models might be guided by practical and computational considerations or, preferably, via statistical tests on the underlying distributional assumptions. Specifically, classical tests on multivariate Gaussianity apply to the Inverse Stereographic case and its Nonparanormal version. However, for the Unwrapped Normal model, ad hoc goodness-of-fit tests are a topic of future research. This research

could build on the idea of SenGupta and Roy [34], adapting their approach to the multivariate Wrapped Normal distribution.

Acknowledgments

The authors would like to thank the Associate Editor and the Referee for their careful reading and for their comments that greatly improved the paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Italian Ministry of University and Research (MUR), Department of Excellence project 2023-2027 ReDS *Rethinking Data Science* – Department of Statistics, Computer Science, Applications – University of Florence. In addition, it benefited from financial support provided by the MUR-PRIN grant 2022 SMNNKY, CUP B53D23009470006 and by the European Union *NextGenerationEU*, National Recovery and Resilience Plan, Mission 4 Component 2 – Investment 1.5 – THE – Tuscany Health Ecosystem – ECS00000017 – CUP B83C22003920001.

References

[1] Jammalamadaka SR, Sengupta A. Topics in circular statistics. Vol. 5. Singapore: World Scientific; 2001.
[2] Mardia KV, Jupp PE. Directional statistics. Vol. 494. Chichester: Wiley; 2009.
[3] Ley C, Verdebout T. Modern directional statistics. Boca Raton: CRC Press; 2017.
[4] Ley C, Verdebout T. Applied directional statistics: modern methods and case studies. Boca Raton: CRC Press; 2018.
[5] Lauritzen SL. Graphical models. Vol. 17. Oxford: Clarendon Press; 1996.
[6] Mencarini L, Vignoli D, Gottard A. Fertility intentions and outcomes: implementing the theory of planned behavior with graphical models. Adv Life Course Res. 2015;23:14–28. doi: 10.1016/j.alcr.2014.12.004
[7] Giudici P, Spelta A. Graphical network models for international financial flows. J Bus Econ Stat. 2016;34(1):128–138. doi: 10.1080/07350015.2015.1017643
[8] Warnick R, Guindani M, Erhardt E, et al. A Bayesian approach for estimating dynamic functional network connectivity in fMRI data. J Am Stat Assoc. 2018;113(521):134–151. doi: 10.1080/01621459.2017.1379404
[9] Ahelegbey DF, Giudici P, Hadji-Misheva B. Latent factor models for credit scoring in P2P systems. Phys A: Stat Mech Appl. 2019;522:112–121. doi: 10.1016/j.physa.2019.01.130
[10] Kotiang S, Eslami A. A probabilistic graphical model for system-wide analysis of gene regulatory networks. Bioinformatics. 2020;36(10):3192–3199. doi: 10.1093/bioinformatics/btaa122
[11] McNally RJ. Network analysis of psychopathology: controversies and challenges. Annu Rev Clin Psychol. 2021;17(1):31–53. doi: 10.1146/clinpsy.2021.17.issue-1
[12] Ren Y, Osborne N, Peterson CB, et al. Bayesian varying-effects vector autoregressive models for inference of brain connectivity networks and covariate effects in pediatric traumatic brain injury. Hum Brain Mapp. 2024;45(10):e26763. doi: 10.1002/hbm.v45.10
[13] Boomsma W, Kent JT, Mardia KV, et al. Graphical models and directional statistics capture protein structure. Interdiscip Stat Bioinf. 2006;25:91–94.
[14] Razavian N, Kamisetty H, Langmead CJ. The von Mises graphical model: regularized structure and parameter learning. Technical Report CMU-CS-11-129, Carnegie Mellon University, 2011.
[15] Klein N, Orellana J, Brincat SL, et al. Torus graphs for multivariate phase coupling analysis. Ann Appl Stat. 2020;14(2):635–660. doi: 10.1214/19-AOAS1300

[16] Selvitella A. On geometric probability distributions on the torus with applications to molecular biology. Electron J Stat. 2019;13(2):2717–2763. doi: 10.1214/19-EJS1579
[17] Greco L, Inverardi PLN, Agostinelli C. Finite mixtures of multivariate wrapped normal distributions for model based clustering of *p*-torus data. J Comput Graph Stat. 2023;32(3):1215–1228. doi: 10.1080/10618600.2022.2128808
[18] Marques I, Kneib T, Klein N. A non-stationary model for spatially dependent circular response data based on wrapped Gaussian processes. Stat Comput. 2022;32(5):73. doi: 10.1007/s11222-022-10136-9
[19] Wei L, Zou Q. Recent progress in machine learning-based methods for protein fold recognition. Int J Mol Sci. 2016;17(12):2118. doi: 10.3390/ijms17122118
[20] Noé F, Fabritiis GD, Clementi C. Machine learning for protein folding and dynamics. Curr Opin Struct Biol. 2020;60:77–84. doi: 10.1016/j.sbi.2019.12.005
[21] Jona-Lasinio G, Gelfand A, Jona-Lasinio M. Spatial analysis of wave direction data using wrapped Gaussian processes. Ann Appl Stat. 2012;6(4):1478–1498. doi: 10.1214/12-AOAS576
[22] García-Portugués E, Sørensen M, Mardia KV, et al. Langevin diffusions on the torus: estimation and applications. Stat Comput. 2019;29(1):1–22. doi: 10.1007/s11222-017-9790-2
[23] Wermuth N. Probability distributions with summary graph structure. Bernoulli. 2011;17(3):845–879. doi: 10.3150/10-BEJ309
[24] Nodehi A, Golalizadeh M, Maadooliat M, et al. Estimation of parameters in multivariate wrapped models for data on a p-torus. Comput Stat. 2021;36(1):193–215. doi: 10.1007/s00180-020-01006-x
[25] Saraceno G, Agostinelli C, Greco L. Robust estimation for multivariate wrapped models. Metron. 2021;79(2):225–240. doi: 10.1007/s40300-021-00214-9
[26] Córdoba I, Bielza C, Larrañaga P. A review of Gaussian Markov models for conditional independence. J Stat Plan Inference. 2020;206:127–144. doi: 10.1016/j.jspi.2019.09.008
[27] Liu H, Lafferty J, Wasserman L. The nonparanormal: semiparametric estimation of high dimensional undirected graphs. J Mach Learn Res. 2009;10(10):2295–2328.
[28] Mazumder R, Hastie T. The graphical lasso: new insights and alternatives. Electron J Stat. 2012;6:2125. doi: 10.1214/12-EJS740
[29] Meinshausen N, Bühlmann P. Stability selection. J Roy Stat Soc Ser B: Stat Methodol. 2010;72(4):417–473. doi: 10.1111/j.1467-9868.2010.00740.x
[30] Li J, Maathuis MH. GGM knockoff filter: false discovery rate control for Gaussian graphical models. J Roy Stat Soc Ser B: Stat Methodol. 2021;83(3):534–558. doi: 10.1111/rssb.12430
[31] Marcotte I, Separovic F, Auger M, et al. A multidimensional 1H NMR investigation of the conformation of methionine–enkephalin in fast-tumbling bicelles. Biophys J. 2004;86(3):1587–1600. doi: 10.1016/S0006-3495(04)74226-5
[32] Drton M, Perlman MD. A sinful approach to Gaussian graphical model selection. J Stat Plan Inference. 2008;138(4):1179–1200. doi: 10.1016/j.jspi.2007.05.035
[33] Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics. 2008;9(3):432–441. doi: 10.1093/biostatistics/kxm045
[34] SenGupta A, Roy M. A characteristic function based circular distribution family and its goodness of fit: the flexible wrapped Linnik family. J Appl Stat. 2023;1–18. doi: 10.1080/02664763.2023.2283689

Appendices

Appendix 1 Proofs

Proof of Result 2.1

Since the mth moment of the complex variable Z_j corresponds to the characteristic function of the unwrapped variable $X_j \sim N(\mu_j, (\boldsymbol{\Sigma})_{jj})$ evaluated at the integer m, it holds that

$$\mathbb{E}[Z_j] = \mathbb{E}\left[e^{\mathrm{i}X_j}\right] = e^{\mathrm{i}\mu_j - (\boldsymbol{\Sigma})_{jj}/2} = e^{\mathrm{i}\mu_j}\, e^{-(\boldsymbol{\Sigma})_{jj}/2}. \tag{A1}$$

Incidentally, notice that because of $e^{\mathrm{i}\mu_j} = \cos(\mu_j) + \mathrm{i}\sin(\mu_j)$, it holds that $\|\mathbb{E}[Z_j]\| = e^{-(\boldsymbol{\Sigma})_{jj}/2}$. Consequently, $\|\mathbb{E}[Z_j]\|$ plays the role of the concentration parameter for Θ_j and it depends on the concentration of X_j. Furthermore, the first mixed moment between Z_i and Z_j amounts to the characteristic function of $(X_i + X_j) \sim N(\mu_i + \mu_j, (\boldsymbol{\Sigma})_{ii} + (\boldsymbol{\Sigma})_{jj} + 2(\boldsymbol{\Sigma})_{ij})$, that is

$$\mathbb{E}[Z_i Z_j] = \mathbb{E}\left[e^{\mathrm{i}(X_i+X_j)}\right] = e^{\mathrm{i}(\mu_i+\mu_j)-\frac{1}{2}\left((\boldsymbol{\Sigma})_{ii}+(\boldsymbol{\Sigma})_{jj}+2(\boldsymbol{\Sigma})_{ij}\right)} = e^{\mathrm{i}(\mu_i+\mu_j)} e^{-(\boldsymbol{\Sigma})_{ii}/2} e^{-(\boldsymbol{\Sigma})_{jj}/2} e^{-(\boldsymbol{\Sigma})_{ij}}. \quad \text{(A2)}$$

Then, combining (A1) with (A2) leads to the result.

Proof of Result 2.2

Let $B = \{1, \ldots, p\} \setminus A$, then the density of $\boldsymbol{\Theta} \sim WN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ satisfies

$$\begin{aligned} f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) &= \sum_{\boldsymbol{k}\in\mathbb{Z}^p} f_X(\boldsymbol{\theta} + 2\pi\boldsymbol{k}) \\ &= \sum_{\boldsymbol{k}_A\in\mathbb{Z}^q} \sum_{\boldsymbol{k}_B\in\mathbb{Z}^{p-q}} (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \\ &\qquad \exp\left\{-\frac{1}{2}\begin{pmatrix} \boldsymbol{\theta}_A - \boldsymbol{\mu}_A + 2\pi\boldsymbol{k}_A \\ \boldsymbol{\theta}_B - \boldsymbol{\mu}_B + 2\pi\boldsymbol{k}_B \end{pmatrix}' \boldsymbol{\Sigma}^{-1} \begin{pmatrix} \boldsymbol{\theta}_A - \boldsymbol{\mu}_A + 2\pi\boldsymbol{k}_A \\ \boldsymbol{\theta}_B - \boldsymbol{\mu}_B + 2\pi\boldsymbol{k}_B \end{pmatrix}\right\} \\ &= \sum_{\boldsymbol{k}_A\in\mathbb{Z}^q} \sum_{\boldsymbol{k}_B\in\mathbb{Z}^{p-q}} f_{\boldsymbol{X}_A\boldsymbol{X}_B}\left(\boldsymbol{\theta}_A + 2\pi\boldsymbol{k}_A, \boldsymbol{\theta}_B + 2\pi\boldsymbol{k}_B\right). \end{aligned}$$

Therefore, the partition of $\boldsymbol{\Theta}$ into $\boldsymbol{\Theta}_A$ and $\boldsymbol{\Theta}_B$ corresponds to a similar partition of $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as

$$\begin{pmatrix} \boldsymbol{X}_A \\ \boldsymbol{X}_B \end{pmatrix} \sim N_p\left(\begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AB} \\ \boldsymbol{\Sigma}_{BA} & \boldsymbol{\Sigma}_{BB} \end{pmatrix}\right).$$

Now, the marginal distribution of $\boldsymbol{\Theta}_A$ can be obtained, by interchanging the integral operator with the sums, as

$$\begin{aligned} f_{\boldsymbol{\Theta}_A}(\boldsymbol{\theta}_A) &= \int_{\mathbb{T}^{p-q}} \sum_{\boldsymbol{k}_A\in\mathbb{Z}^a} \sum_{\boldsymbol{k}_B\in\mathbb{Z}^b} f_{\boldsymbol{X}_A\boldsymbol{X}_B}(\boldsymbol{\theta}_A + 2\pi\boldsymbol{k}_A, \boldsymbol{\theta}_B + 2\pi\boldsymbol{k}_B)\,\mathrm{d}\boldsymbol{\theta}_B \\ &= \sum_{\boldsymbol{k}_A\in\mathbb{Z}^q} \sum_{\boldsymbol{k}_B\in\mathbb{Z}^{p-q}} \int_{\mathbb{T}^{p-q}} f_{\boldsymbol{X}_A\boldsymbol{X}_B}(\boldsymbol{\theta}_A + 2\pi\boldsymbol{k}_A, \boldsymbol{\theta}_B + 2\pi\boldsymbol{k}_B)\,\mathrm{d}\boldsymbol{\theta}_B. \end{aligned}$$

Consequently, by concatenation of integrals

$$\begin{aligned} f_{\boldsymbol{\Theta}_A}(\boldsymbol{\theta}_A) &= \sum_{\boldsymbol{k}_A\in\mathbb{Z}^q} \int_{\mathbb{R}^{p-q}} f_{\boldsymbol{X}_A\boldsymbol{X}_B}(\boldsymbol{\theta}_A + 2\pi\boldsymbol{k}_A, \boldsymbol{u})\,\mathrm{d}\boldsymbol{u} \\ &= \sum_{\boldsymbol{k}_A\in\mathbb{Z}^q} f_{\boldsymbol{X}_A}(\boldsymbol{\theta}_A + 2\pi\boldsymbol{k}_A). \end{aligned}$$

Using the fact that $\boldsymbol{X}_A \sim N_q(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{AA})$, the result follows. □

Proof of Result 2.3

As $\boldsymbol{X}_B = \boldsymbol{\Theta}_B + 2\pi\boldsymbol{K}_B$, the density of $(\boldsymbol{\Theta}_A, \boldsymbol{K}_A) \mid \boldsymbol{\Theta}_B, \boldsymbol{K}_B$ is

$$f_{\boldsymbol{\Theta}_A,\boldsymbol{K}_A|\boldsymbol{\Theta}_B,\boldsymbol{K}_B}(\boldsymbol{\theta}_A, \boldsymbol{k}_A \mid \boldsymbol{\theta}_B, \boldsymbol{k}_B) = f_{\boldsymbol{X}_A|\boldsymbol{X}_B}(\boldsymbol{\theta}_A + 2\pi\boldsymbol{k}_A \mid \boldsymbol{\theta}_B + 2\pi\boldsymbol{k}_B).$$

Hence, as $\boldsymbol{X}_A \mid \boldsymbol{X}_B \sim N_q(\boldsymbol{\mu}_{A|B}, \boldsymbol{\Sigma}_{A|B})$, it holds that

$$f_{\boldsymbol{\Theta}_A|\boldsymbol{\Theta}_B,\boldsymbol{K}_B}(\boldsymbol{\theta}_A \mid \boldsymbol{\theta}_B, \boldsymbol{k}_B) = \sum_{\boldsymbol{k}_A\in\mathbb{Z}^q} f_{\boldsymbol{X}_A|\boldsymbol{X}_B}(\boldsymbol{\theta}_A + 2\pi\boldsymbol{k}_A \mid \boldsymbol{\theta}_B + 2\pi\boldsymbol{k}_B).$$

The above density is a (conditional) wrapped Normal density sharing its parameters with the distribution of $\boldsymbol{X}_A \mid \boldsymbol{X}_B$, i.e. ,$\boldsymbol{\Theta}_A \mid \boldsymbol{X}_B \sim WN_q(\boldsymbol{\mu}_{A|B}, \boldsymbol{\Sigma}_{A|B})$.

For the independence result, notice that as $f_{\boldsymbol{\Theta}_A|\boldsymbol{\Theta}_B,\boldsymbol{K}_B}(\boldsymbol{\theta}_A \mid \boldsymbol{\theta}_B, \boldsymbol{k}_B) = f_{\boldsymbol{\Theta}_A}(\boldsymbol{\theta}_A)$ whenever

$$f_{\boldsymbol{X}_A|\boldsymbol{X}_B}(\boldsymbol{\theta}_A + 2\pi\boldsymbol{k}_A \mid \boldsymbol{\theta}_B + 2\pi\boldsymbol{k}_B) = f_{\boldsymbol{X}_A}(\boldsymbol{\theta}_A + 2\pi\boldsymbol{k}_A),$$

then $\boldsymbol{\Theta}_A \perp\!\!\!\perp \boldsymbol{\Theta}_B$ iff $\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B$, i.e. ,iff $\boldsymbol{\Sigma}_{AB} = \boldsymbol{0}_{q\times(p-q)}$. □

Proof of Result 2.4

As a consequence of Result 2.2, if $\boldsymbol{\Theta} \sim WN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, for disjoint subsets A and S of $\{1, \ldots, p\}$ with respective cardinalities q and s, both $(\boldsymbol{\Theta}_A, \boldsymbol{\Theta}_S)$ and $\boldsymbol{\Theta}_S$ have Wrapped Normal distribution, and then, the density of $\boldsymbol{\Theta}_A \mid \boldsymbol{\Theta}_S$ is the ratio

$$f_{\boldsymbol{\Theta}_A|\boldsymbol{\Theta}_S}(\boldsymbol{\theta}_A \mid \boldsymbol{\theta}_S) = \frac{\sum_{\boldsymbol{k}_A\in\mathbb{Z}^q}\sum_{\boldsymbol{k}_S\in\mathbb{Z}^s} f_{\boldsymbol{X}_A\boldsymbol{X}_S}(\boldsymbol{\theta}_A + 2\pi\boldsymbol{k}_A, \boldsymbol{\theta}_S + 2\pi\boldsymbol{k}_S)}{\sum_{\boldsymbol{k}_S\in\mathbb{Z}^s} f_{\boldsymbol{X}_S}(\boldsymbol{\theta}_S + 2\pi\boldsymbol{k}_S)},$$

where $(\boldsymbol{X}_A, \boldsymbol{X}_S)$ and $\boldsymbol{X}_S$ have Normal distribution. The result comes writing the joint density of $\boldsymbol{X}_A$ and $\boldsymbol{X}_S$ at the numerator as the product of the conditional density of $\boldsymbol{X}_A \mid \boldsymbol{X}_S$ and the marginal density of $\boldsymbol{X}_S$,

$$f_{\boldsymbol{\Theta}_A|\boldsymbol{\Theta}_S}(\boldsymbol{\theta}_A \mid \boldsymbol{\theta}_S) = \frac{\sum_{\boldsymbol{k}_A\in\mathbb{Z}^q}\sum_{\boldsymbol{k}_S\in\mathbb{Z}^s} f_{\boldsymbol{X}_A|\boldsymbol{X}_S}(\boldsymbol{\theta}_A + 2\pi\boldsymbol{k}_A \mid \boldsymbol{\theta}_S + 2\pi\boldsymbol{k}_S) f_{\boldsymbol{X}_S}(\boldsymbol{\theta}_S + 2\pi\boldsymbol{k}_S)}{\sum_{\boldsymbol{k}_S\in\mathbb{Z}^s} f_{\boldsymbol{X}_S}(\boldsymbol{\theta}_S + 2\pi\boldsymbol{k}_S)},$$

and setting

$$w_S(\boldsymbol{\theta}_S, \boldsymbol{k}_S) = \frac{f_{\boldsymbol{X}_S}(\boldsymbol{\theta}_S + 2\pi\boldsymbol{k}_S)}{\sum_{\boldsymbol{k}_S\in\mathbb{Z}^s} f_{\boldsymbol{X}_S}(\boldsymbol{\theta}_S + 2\pi\boldsymbol{k}_S)}.$$

□

Proof of Result 3.1

The proof directly follows by noting that the density function in Equation (4) differs from the density of $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ by the multiplicative factor $\prod_{j=1}^{p}(1 + \cos(\theta_j))^{-1}$. Then, the conditional independence parameter for $\boldsymbol{\Theta} \sim ISN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the same as $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and the two distributions factorize together. Therefore, the conditional independence property is the same as the Gaussian distribution. □

Appendix 2 On the wrapped normal distribution

A.1 Remarks on conditional independence in WN distribution

For a random vector of angles with Wrapped Normal distribution, if the unwrapped graph is able to suggest the conditional independence structure of the random angles, then the conditional distribution of an angle, given its neighbours, should be very close to the conditional distribution of the same angle given all the other angles. For instance, let $\boldsymbol{\Theta} = (\Theta_1, \Theta_2, \Theta_3)' \sim WN_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with

$$\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} j/2\pi & -0.5 & -0.5 \\ -0.5 & 4/\pi & 0 \\ -0.5 & 0 & 4/\pi \end{pmatrix}$$

where j varies from 1 to 1000. The corresponding unwrapped variables X_1, X_2, X_3 follow a three-variate Normal distribution satisfying $X_2 \perp\!\!\!\perp X_3 \mid X_1$. Therefore, in the unwrapped graph, X_1 separates X_2 and X_3. Different values of j imply different magnitudes of the variance, ranging from $2\pi/1000$ to 2π.

Figure A1 represents the Kullback–Leibler divergence (KLD), when the variance σ_1^2 is varying, between the distributions $f_{\Theta_2|\Theta_1\Theta_3}(\theta_2|\theta_1\theta_3)$, of Θ_2 given all the other variables, and $f_{\Theta_2|\Theta_1}(\theta_2|\theta_1)$,

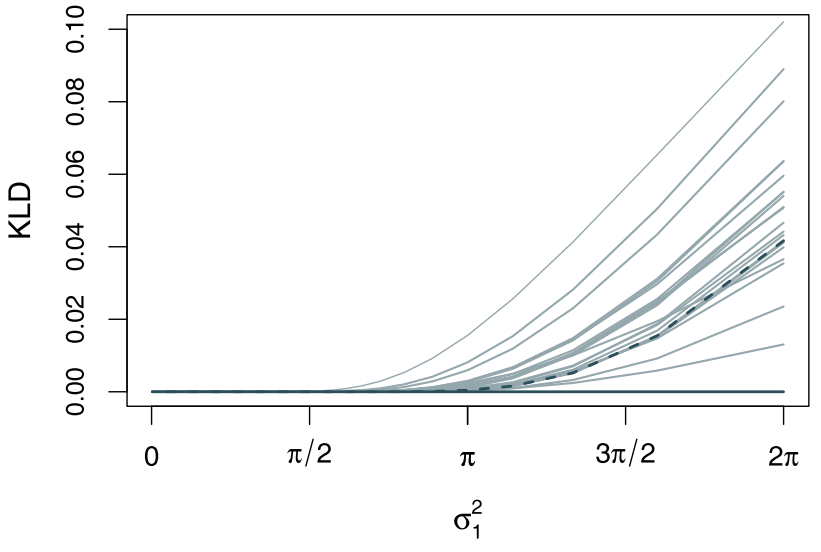


Figure A1. Kullback–Leibler divergence of the conditional distribution of $\Theta_2 \mid \Theta_1, \Theta_3$ and of $\Theta_2 \mid \Theta_1$, when $X_2 \perp\!\!\!\perp X_3 \mid X_1$ for a set of randomly chosen values of the conditional variables. The bold solid line corresponds to $(\theta_1 = 0, \theta_3 = 0)$, the dashed line corresponds to $(\theta_1 = \pi, \theta_3 = \pi)$.

conditional only on the neighbour as suggested by the unwrapped graph. The values of the conditioning set are randomly chosen. This example confirms that, when the variance σ_1^2 is small enough, the divergence from the two distributions is indistinguishable from zero. In the considered case, this seems to occur for $\sigma_1^2 < \pi/2$. According to this type of investigation, when the variance of the separators is small enough, the joint distribution over the angles can be simplified according to the unwrapped graph and the dependence induced by the winding numbers is negligible.

A.2 Remarks on the approximation of the likelihood function for the WN distribution

Let $\boldsymbol{\Theta} \sim WN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and $\boldsymbol{X} = \boldsymbol{\Theta} + 2\pi \boldsymbol{K}$. The concentration matrix driving the conditional independence structure of the Unwrapped Normal graphical model depends on the parameters shared by the distributions of $\boldsymbol{X}$ and $\boldsymbol{\Theta}$, and it would be desirable to estimate it using the observed angles. To make a feasible inference for the Wrapped Normal distribution, one can rely on an accurate approximation that avoids the infinite sum by appropriately truncating the winding numbers.

Consider the vector of thresholds for the winding numbers $\boldsymbol{k}^* = k^* \mathbf{1}_p$, with $k^* \in \mathbb{Z}$ and $\mathbf{1}_p$ being the p-dimensional vector of ones. The approximation we are going to consider replaces the doubly infinite summation in the WN distribution with the finite sum from $-\boldsymbol{k}^*$ to $\boldsymbol{k}^*$. The crucial point for likelihood-based inference on the multivariate Wrapped Normal distribution is whether this approximation affects the maximum of the likelihood function. In particular, when learning a graphical model, the interest is on the estimate of the precision matrix $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$, encoding the conditional independence structure of the variables. For this purpose, we conducted a small simulation study, generating data from a trivariate Wrapped Normal distribution with $\boldsymbol{\mu} = (\pi, \pi, \pi)'$, covariance matrix $\boldsymbol{\Sigma}$ and subsequent precision matrix $\boldsymbol{\Omega}$ and partial correlation matrix $\boldsymbol{R}$ respectively given by

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1.0 & 2.0 & 0.8 \\ 2.0 & 5.0 & 2.0 \\ 0.8 & 2.0 & 1.0 \end{pmatrix} \quad \boldsymbol{\Omega} = \begin{pmatrix} 5.0 & -2.0 & 0.0 \\ -2.0 & 1.8 & -2.0 \\ 0.0 & -2.0 & 5.0 \end{pmatrix} \quad \boldsymbol{R} = \begin{pmatrix} 1.00 & 0.67 & 0.00 \\ 0.67 & 1.00 & 0.67 \\ 0.00 & 0.67 & 1.00 \end{pmatrix}.$$

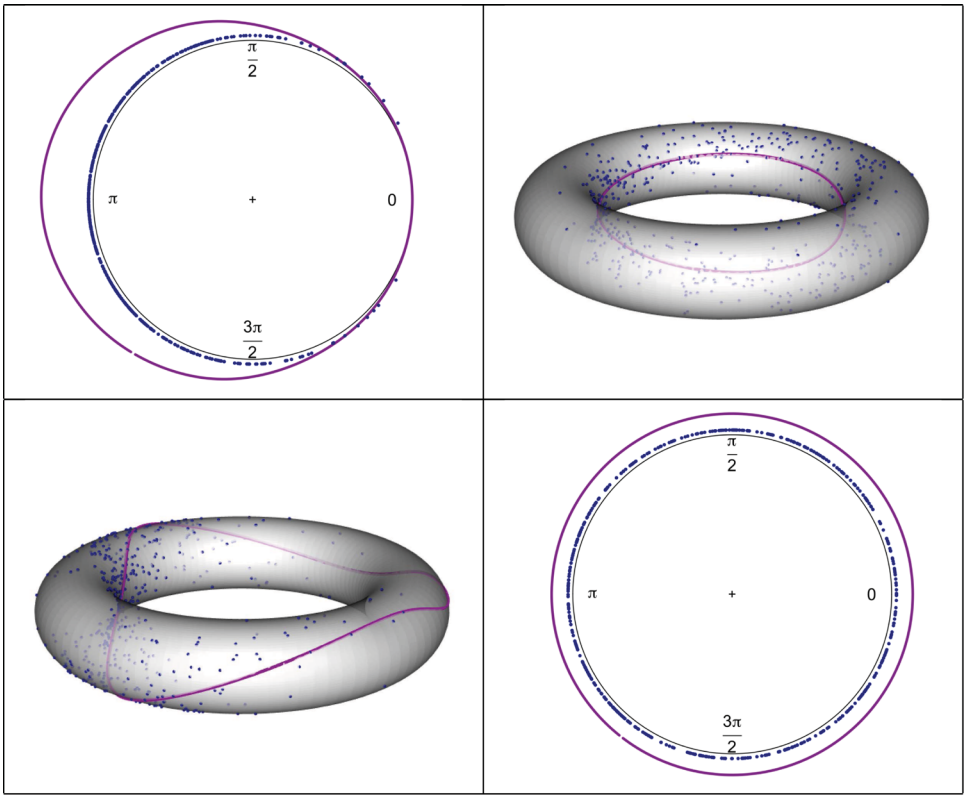


Figure A2. Scatter-plot matrix for Θ_1 and Θ_2 with kernel density estimate on the main diagonal and kernel regression function estimate on the off-diagonal elements.

The reported scenario is then characterized by the independence $X_1 \perp\!\!\!\perp X_3 \mid X_2$, with a larger variance of the separator X_2 and moderate variances for the remaining nodes. As an example of data generated from this distribution, Figure A2 depicts the portion of the scatter-plot matrix concerning Θ_1 and Θ_2. The characteristics of the distribution of Θ_3 are, as a matter of fact, similar to those of Θ_1.

We consider the profile likelihood approach described in Section 2, and report in Figure A3 the estimates of the off-diagonal entries of $\boldsymbol{R}$ for some values of k^*. We generate 100 samples with two sample sizes, $n = 500$ and $n = 5000$. The Monte Carlo experiment suggests that the approximation with the Gaussian distribution ($k^* = 0$) is unsatisfactory. At the same time, the maximum likelihood estimates remain constant for the values of κ^* greater than zero. This result was also confirmed for other scenarios, where the values of the main diagonal entries of $\boldsymbol{\Sigma}$ varied from 0.1 to 10. In addition, comparing the results from the two different sample sizes, it seems that the maximum likelihood consistency is not affected by the approximation.

A.3 Monte Carlo simulations for unwrapped graphical models

To investigate the ability of the procedure for parameter estimation and graph recovery in Unwrapped Gaussian graphical models, we are reporting here the results of a small simulation study. We set two scenarios by generating data from two different four-dimensional WN distributions for different sample sizes ($n = 30, 50, 100, 150, 250$). Both the data-generating distributions have mean

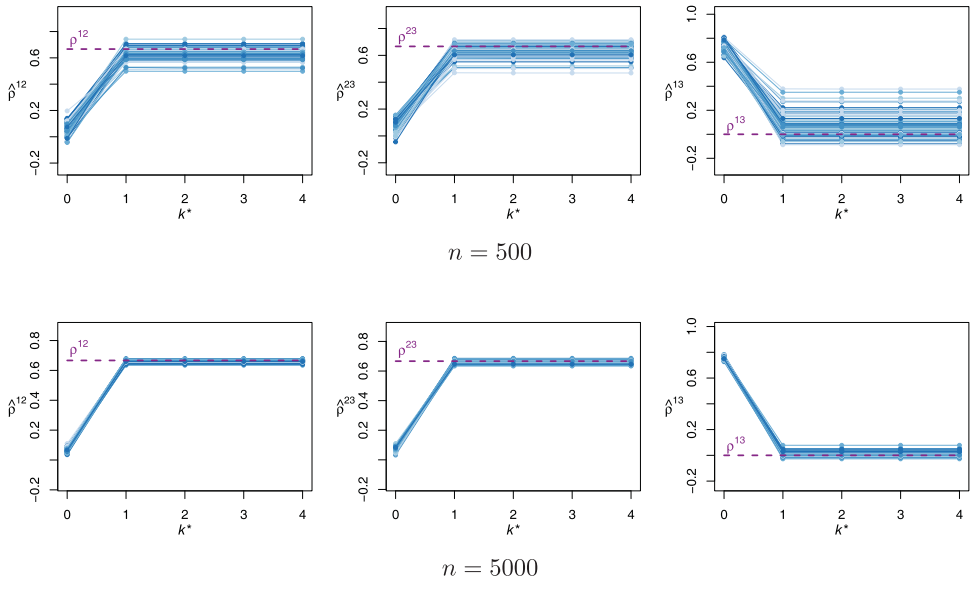


Figure A3. Estimated partial correlation coefficients with k^* varying from 0 to 4 over 100 simulations for $n = 500$ (top panels) and $n = 5000$ (bottom panels).

vector $\boldsymbol{\mu} = (\pi, \pi, \pi, \pi)'$, and covariance matrices

$$\boldsymbol{\Sigma_1} = \begin{pmatrix} 0.79 & 0.39 & 0.20 & 0.10 \\ 0.39 & 0.98 & 0.49 & 0.25 \\ 0.20 & 0.49 & 1.03 & 0.52 \\ 0.10 & 0.25 & 0.52 & 1.04 \end{pmatrix} \quad \boldsymbol{\Sigma_2} = \begin{pmatrix} 0.79 & 0.20 & 0.05 & 0.01 \\ 0.20 & 0.83 & 0.21 & 0.05 \\ 0.05 & 0.21 & 0.84 & 0.21 \\ 0.01 & 0.05 & 0.21 & 0.84 \end{pmatrix}$$

yielding to the respective partial correlation matrices

$$\boldsymbol{\Omega_1} = \begin{pmatrix} 1 & 0.40 & 0 & 0 \\ 0.40 & 1 & 0.40 & 0 \\ 0 & 0.40 & 1 & 0.45 \\ 0 & 0 & 0.45 & 1 \end{pmatrix} \quad \boldsymbol{\Omega_2} = \begin{pmatrix} 1 & 0.24 & 0 & 0 \\ 0.24 & 1 & 0.24 & 0 \\ 0 & 0.24 & 1 & 0.24 \\ 0 & 0 & 0.24 & 1 \end{pmatrix}.$$

In the first scenario, there is a stronger dependence on the connected variables, while in the second one, such dependence is weaker. Variances have similar amounts in both scenarios, varying between $\pi/4$ and $\pi/3$. This results in larger partial correlation coefficients for the first scenario with respect to the second one. To evaluate the effectiveness of the graph learning procedure, we use ROC curves to show the true positive rate (TPR) versus the false positive rate (FPR) at various threshold settings. Here

$$\text{TPR} = \frac{TP}{TP + FN} \quad \text{and} \quad \text{FPR} = \frac{FP}{TN + FP}$$

where TP (TN, respectively) represents the number of correctly identified edges (missing edges, resp.), while *FP* (*FN*, respectively) stands for the number of falsely identified edges (missing edges, resp.). Monte Carlo ROC curves for the considered scenarios are depicted in Figure A4 for the considered sample sizes. In the first scenario, where the dependence is stronger, the procedure performs reasonably well, even for moderate sample sizes. As expected, the performances worsen for weaker dependence, requiring larger samples to provide satisfactory results, while when n is 30, the actual performance is only slightly above the random guess.

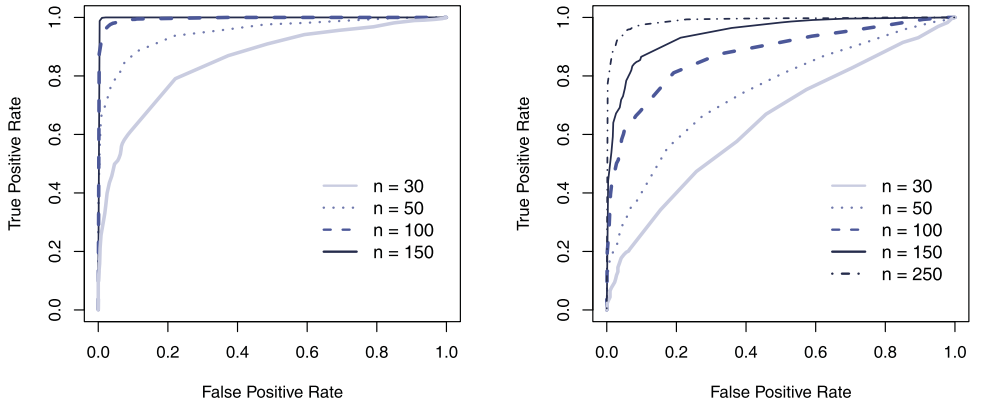


Figure A4. Monte Carlo ROC curves over 250 simulation runs for data simulated in different sample sizes from Wrapped Normal distribution with covariance matrix $\mathbf{\Sigma_1}$ (left panel) and $\mathbf{\Sigma_2}$ (right panel) with $n = 30, 50, 100, 150, 250$.

Appendix 3 Summary table

Table A1. Key properties of Wrapped Normal and Inverse Stereographic Normal distributions.

Distribution	$\boldsymbol{\Theta} \sim WN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$\boldsymbol{\Theta} \sim \text{ISN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
Type of distribution	Wrapped	Projected
Link to $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$\boldsymbol{\Theta} = \boldsymbol{X} (\text{mod} 2\pi)$	$\boldsymbol{X} = \tan(\boldsymbol{\Theta}/2)$
Marginalization	$\boldsymbol{\Theta}_A \sim WN_q(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{AA})$	$\boldsymbol{\Theta}_A \sim \text{ISN}_q(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{AA})$
Conditioning	$\boldsymbol{\Theta}_A \mid \boldsymbol{\Theta}_B \not\sim WN$	$\boldsymbol{\Theta}_A \mid \boldsymbol{\Theta}_B \sim \text{ISN}_q(\boldsymbol{\mu}_{A\|B}, \boldsymbol{\Sigma}_{A\|B})$
Conditional indep. parameter	None	$(\boldsymbol{\Sigma}^{-1})_{ij}$
Related Graphical Models	Unwrapped GM	ISN GM & ISNPN GM