Model selection in the space of Gaussian models invariant by symmetry

Piotr Graczyk, Hideyuki Ishi, Bartosz Kołodziejek, Hélène Massam

Citation	Annals of Statistics. 50(3): 1747-1774
Published	2022-06
Type	Journal Article
Textversion	Publisher
Highlights	◇多次元データの対称性をベイズ統計によって効果的に探索する数学的方法を開発 ◇ベイズ統計のネックとされていた積分計算の困難さに対し、正確な公式を導出 ◇遺伝子解析などの応用が期待
Description	この論文では，ある置換群の作用に関して不変な平均ゼロ多変量ガウス分布の統計モデルを考える．有限群の実数体上の表現論を利用して，そういったモデルの共分散行列の最尤推定量と，精度行列のディアコニス・イルヴィサカ共役事前分布の正規化公式の解析的表示を導出した．これを用いて，置換群不変な多変量ガウス分布モデル（完全 RCOP モデルと呼ぶ）のベイズ式モデル選択を実行した．これらの結果を，４次元データの簡単な例や，巡回群に限定した 100 次元データのモデル選択といった高次元のいくつかの例によって説明した．
Supplementary material	Supplementary material is available at https://doi.org/10.1214/22-AOS2174.
Rights	© 2022 Institute of Mathematical Statistics. The Version of Record can be accessed online at https://doi.org/10.1214/22-AOS2174.
DOI	10.1214/22-AOS2174
Funding	●JST さきがけ「正定値対称行列の数理に関する革新的新技術」(2016.10.5～2020.3.31) ●科研費（基盤研究（C）16K05174）「凸錐上の調和解析とその応用」(2016.4.1～2021.3.31) ●科研費（基盤研究（C）20K03657）「凸錐上の解析と幾何の新展開」(2020.4.1～2024.3.31)

Piotr Graczyk. Hideyuki Ishi. Bartosz Kołodziejek. Hélène Massam. "Model selection in the space of Gaussian models invariant by symmetry." Ann. Statist. 50 (3) 1747 - 1774, June 2022.
https://doi.org/10.1214/22-AOS2174

‘遺伝子解析などへの応用が期待　ベイズ統計によってデータの対称性を探索する方法を開発’ 大阪公立大学. https://www.omu.ac.jp/info/research_news/entry-01905.html. (参照 2022-08-26).

News

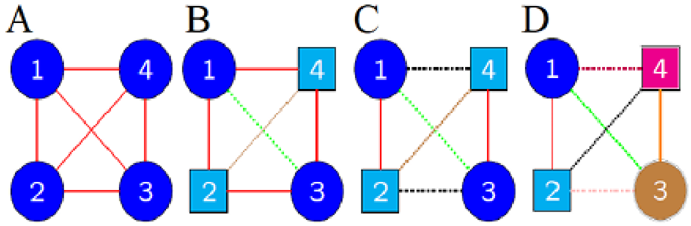
＜概要＞

国際研究グループは、群の表現論という純粋数学分野の知識を活用し、多次元データの対称性をベイズ統計の技法によって探索する方法を開発しました。

ベイズ統計は、計算機の性能向上や、人工知能への活用の可能性から、近年一層の脚光を浴びており、2022 年 7 月には、文部科学省が策定した「2030 年に向けた数理科学の展開－数理科学への期待と重要課題－」内の、AI×数理科学の分野で取り上げられています（https://www.mext.go.jp/b_menu/houdou/kagaku/2022/mext_01067.html）。
ベイズ統計には複雑な積分計算が必要とされるため、多くの場合、近似計算で処理されています。本研究では、未だ見つけられていなかった正確な積分公式を導出することができました。今後、先行研究に比して探索の精度が上がり、遺伝子解析などの注目分野においてさらなる応用が期待できます。

＜研究の内容＞

多次元データの対称性を考える際によく使われる例として、G. P. Frets が計測した 25 の家族の『①長男の頭の長さ ②長男の頭の幅 ③次男の頭の長さ ④次男の頭の幅』というデータを扱うとします。このデータにおいて、仮に①と③、②と④をそれぞれ入れ替えたとき、分布の構造が変化しないかどうか（対称性があるかどうか）を計算によって導き出します。そのようなデータの対称性は次のような色付きのグラフで表現されます（図）。



A：①，②，③，④の全ての入れ替えに対して構造が変わらないというモデル。
B：①と③，②と④の入れ替えのそれぞれに関して変わらないというモデル。
C：①と③，②と④を同時に入れ替える（グラフを１８０度回転する）という操作で変わらないというモデル。
D は対称性を持たないというモデル。

構造を変えない入れ替えで移りあう頂点や辺を同じ色や形にすることで、対称性を表現しています。このような統計モデルは色付きグラフィカルモデルとよばれています。

上述の設定では、4 つのデータの置換に関して 22 のパターンが考えられます。ベイズ統計の考え方では、まずそれら 22 個は同じ可能性をもつと想定し（事前確率）、25 家族のデータを組み入れた後でどのように可能性が変化するかを計算し（事後確率），事後確率

	が一番大きいパターンが最も確からしさが高いと判断します。この事後確率の計算には、複雑な積分計算が必要ですが、本研究ではそのための正確な積分公式を導出できました。さらに、データが 10 個になると数百万、18 個だと数百京（けい）というように、データ数が増えれば増えるほど、膨大な数のパターンから最適なものを選択することになるため、問題を厳密に解くのが困難になりますが、本研究ではそのような場合でも有効な近似解を求めるアルゴリズムの開発に成功しました。 **＜期待される効果・今後の展開＞** データの対称性はさまざまなモデルで普遍的にみられるもので、対称性の存在が確認されれば、そのデータの構造を表示するのに必要なパラメータの数、そしてパラメータを定めるために必要なサンプルの数も何分の一にも削減できます。今後、本研究成果を応用することで、遺伝子解析において、染色体が別の場所でも同じ機能をもつような部位を発見すること等が期待できます。

The Annals of Statistics
2022, Vol. 50, No. 3, 1747–1774
https://doi.org/10.1214/22-AOS2174

MODEL SELECTION IN THE SPACE OF GAUSSIAN MODELS INVARIANT BY SYMMETRY

BY PIOTR GRACZYK[1,a], HIDEYUKI ISHI[2,b], BARTOSZ KOŁODZIEJEK[3,c] AND HÉLÈNE MASSAM[4,d]

[1]*LAREMA, UFR Sciences, Université d'Angers,* [a]*piotr.graczyk@univ-angers.fr*
[2]*Department of Mathematics, Osaka City University,* [b]*hideyuki@sci.osaka-cu.ac.jp*
[3]*Faculty of Mathematics and Information Science, Warsaw University of Technology,* [c]*b.kolodziejek@mini.pw.edu.pl*
[4]*Department of Mathematics and Statistics, York University,* [d]*massamh@mathstat.yorku.ca*

We consider multivariate centered Gaussian models for the random variable $Z = (Z_1, \dots, Z_p)$, invariant under the action of a subgroup of the group of permutations on $\{1, \dots, p\}$. Using the representation theory of the symmetric group on the field of reals, we derive the distribution of the maximum likelihood estimate of the covariance parameter Σ and also the analytic expression of the normalizing constant of the Diaconis–Ylvisaker conjugate prior for the precision parameter $K = \Sigma^{-1}$. We can thus perform Bayesian model selection in the class of complete Gaussian models invariant by the action of a subgroup of the symmetric group, which we could also call complete RCOP models. We illustrate our results with a toy example of dimension 4 and several examples for selection within cyclic groups, including a high-dimensional example with $p = 100$.

1. Introduction.

1.1. *Motivations and applications.* Let $V = \{1, \dots, p\}$ be a finite index set and let $Z = (Z_1, \dots, Z_p)$ be a multivariate random variable following a centered Gaussian model $\mathrm{N}_p(0, \Sigma)$. Let $\mathfrak{S}_p$ denote the symmetric group on V, that is, the group of all permutations on $\{1, \dots, p\}$ and let Γ be a subgroup of $\mathfrak{S}_p$. A centered Gaussian model is said to be invariant under the action of Γ if for all $g \in \Gamma$, $g \cdot \Sigma \cdot g^\top = \Sigma$ (here we identify a permutation g with its permutation matrix).

Given n data points $Z^{(1)}, \dots, Z^{(n)}$ from a Gaussian distribution, our aim in this paper is to do Bayesian model selection within the class of models invariant by symmetry, that is, invariant under the action of some subgroup Γ of $\mathfrak{S}_p$ on V. Given the data, our aim is therefore to identify the subgroup $\Gamma \subset \mathfrak{S}_p$ such that the model invariant under Γ has the highest posterior probability. We achieve this goal by constructing a Markov chain on the space of models and using the Metropolis–Hastings algorithm.

There are many alternative ways of doing model search in modern statistics on big data sets, both frequentist and Bayesian. Bayesian model selection methods (cf. Ghosal and van der Vaart (2017), Chapter 10) are widely used in practice, thanks to the possibility of using a prior knowledge on the model and to their rigorous mathematical bases. Moreover, in Bayesian approach the Metropolis–Hastings algorithm is naturally applicable and generally accepted.

Our work can be viewed as a special case of colored graphical Gaussian models (the underlying graph is complete so we do not impose conditional independence structure), that is,

Received April 2020; revised January 2022.
MSC2020 subject classifications. Primary 62H99, 62F15; secondary 20C35.
Key words and phrases. Colored graph, conjugate prior, covariance selection, invariance, permutation symmetry.

statistical graphical models with additional symmetries (equality constraints on the precision or correlation matrix). Such models were introduced into modern exploratory analysis of data in the seminal paper Højsgaard and Lauritzen (2008), as a powerful tool of dimension reduction in unsupervised learning; cf. Maathuis et al. ((2018), Chapter 9.8). A preponderant role is given in Højsgaard and Lauritzen (2008) to the RCOP models studied in our paper, thanks to their most tractable structure and interpretability through symmetries among the variables. One of motivations of this paper is to address the task stated in Højsgaard and Lauritzen ((2008), page 1025): *For the models to become widely applicable, it is mandatory to develop algorithms for model identification which are robust, reliable and transparent*.

For high-dimensional data, Gaussian models, which have symmetries and are graphical, allow statisticians to reduce the dimension of a model. In genetics, such models can be used to identify genes having the same function or groups of genes having similar interactions. Below we mention some of the studies in which our model could find potential application.

In Højsgaard and Lauritzen (2008), gene expression signatures for p53 mutation status in 58 breast cancer samples consisting of 150 genes were investigated and interpreted. We apply our algorithm to this data in the Supplementary Material; see Section 6 in Graczyk et al. (2022) We claim that our model selection procedure can be used as an exploratory tool. Assuming that the variables are all on some common scale, the proposed algorithm can be run to look for potential hidden symmetries between the variables.

It is worth to underline that one of the characteristics of our model is the lack of scale invariance. We point out below that there are many examples where our model can still be applied. For example, the data from gene expression are on the same scale in the sense that they are results of experiments of the same type, measured in the same gauges. Similar situation appears generally for omic data sets in proteomics and metabolomics. For more details see, for example, the monograph Frommlet, Bogdan and Ramsey (2016). In Sobczyk et al. ((2020), Section 6.2 TCGA Breast Cancer Data), genetic information in tumoral tissues DNA that are involved in gene expression are measured from messenger sequencing by the RNASeq method and they are all on the same scale, as they are the numbers of transcripts in a sample. In clinical epidemiology and medicine, one often uses scales combined into scores to classify outcome; see, for example, Toyoda et al. (2022), Missio et al. (2019). Range of values of such scores are often similar, even though not formally tested statistically to be so. In the paper, Descatha et al. (2007) the normalization or nonnormalization of data did not influence their statistical interpretation.

Moreover, we argue that it is natural to expect certain symmetries in the data from gene expression. Namely, expression of a given gene is triggered by binding the transcription factors to the gene transcription factor binding sites. The transcription factors are the proteins produced by other genes, say regulatory genes. In the gene network, there are often many genes triggered by the same regulatory genes and it makes sense to assume that their relative expressions depend on the abundance of proteins of the regulatory genes (i.e., gene expressions) in a similar way.

In Gao and Massam (2015), 12,625 neutrophil gene expressions were monitored with imposed symmetry constraints to the graphical modeling. The paper (Li, Gao and Massam (2020)) contains a study of the structure of colored graphs applied to a flow cytometry data set on signaling networks of human immune system cells, which consists of 7466 measurements on 11 phosphorylated proteins.

A very recent application of graphical models with symmetries to fMRI real data on brain networks is proposed in Ranciati, Roverato and Luati (2021). An impressive number of recent applications of graphical models to real data analysis is listed in the recent monograph Maathuis et al. ((2018), Chapters 19, 20, 21) and includes genetics, genomics, molecular systems biology and forensic analysis; cf. also the books Roverato (2017) for medical and Li (2009) for image data applications.

Finally, let us mention that colored graphical models provide interesting examples of exponential algebraic varieties and algebraic exponential families, for example, Toeplitz matrices Michałek et al. (2016); see also Davies and Marigliano (2021). The recent algebro-geometric approach to graphical models and Gaussian Bayesian networks is being developed intensely Maathuis et al. ((2018), Chapter 3).

1.2. *Contribution of the paper and relations to previous work.* In this subsection, we carefully describe and position this paper in the context of previous research.

Theory of invariant normal models (with the so-called lattice conditional independencies (Andersson and Madsen (1998), Madsen (2000)), which are not considered in the present paper) was developed by the Danish school. History regarding this subject is nicely presented in Andersson and Madsen (1998), where the reader can also find references to earlier works dealing with particular symmetry models such as, for example, the circular symmetry model of Olkin and Press (1969) that we will consider further (Section 5). Among others, Andersson, Brøns, Jensen, Madsen and Perlman developed a fairly complete theory of MLE $\hat{\Sigma}$ of the covariance matrix in invariant normal models, however, the problems considered in our paper are very different.

These works were concentrating on the derivation of statistical properties of the maximum likelihood estimate of Σ and on testing the hypothesis that models were of a particular type. In particular, to the best of our knowledge the Danish school never considered any model search in the context of invariant normal models. When the model space is very big (and this is the usual case of our framework), then it is impossible to perform simultaneous tests for all possible models. Despite the computation problems, there is also even bigger issue due to multiple comparisons problem.

Just like the classical papers mentioned above, the fundamental algebraic tool we use in this work is the irreducible decomposition theorem for the matrix representation of the group Γ, which in turn means that, through an adequate change of basis, any matrix X in $\mathcal{Z}_\Gamma$, the space of symmetric matrices invariant under the subgroup Γ of $\mathfrak{S}_p$, can be written in a block diagonal form. The following result is a reformulation of an observation made in Andersson ((1975), 4.6–4.8).

THEOREM 1. *Fix a permutation subgroup $\Gamma \subset \mathfrak{S}_p$. Then there exist constants $L \in \mathbb{N}$, $(k_i, d_i, r_i)_{i=1}^L$ and orthogonal matrix U_Γ such that if $X \in \mathcal{Z}_\Gamma$, that is, $X \in \mathrm{Sym}(p; \mathbb{R})$ and*

$$X_{ij} = X_{\sigma(i)\sigma(j)} \quad (\sigma \in \Gamma, i, j \in \{1, \dots, p\}),$$

then

$$X = U_\Gamma \cdot \begin{pmatrix} M_{\mathbb{K}_1}(x_1) \otimes I_{k_1/d_1} & & & \\ & M_{\mathbb{K}_2}(x_2) \otimes I_{k_2/d_2} & & \\ & & \ddots & \\ & & & M_{\mathbb{K}_L}(x_L) \otimes I_{k_L/d_L} \end{pmatrix} \cdot U_\Gamma^\top, \tag{1}$$

where $M_{\mathbb{K}_i}(x_i)$ is a real matrix representation of an $r_i \times r_i$ Hermitian matrix x_i with entries in $\mathbb{K}_i = \mathbb{R}, \mathbb{C}$ or $\mathbb{H}$, $i = 1, \dots, L$, and $A \otimes B$ denotes the Kronecker product of matrices A and B.

Elements of $(k_i, d_i, r_i)_{i=1}^L$ are integer constants called structure constants that we will define later. At this point, we note that k_i/d_i are also integers and $d_i = \dim_{\mathbb{R}} \mathbb{K}_i \in \{1, 2, 4\}$. The mappings $M_{\mathbb{K}_i} \colon \mathrm{Herm}(r_i; \mathbb{K}_i) \to \mathrm{Sym}(d_i r_i; \mathbb{R})$ are defined in Section 2.2. As was already observed in Jensen (1988), the space $\mathcal{Z}_\Gamma$ equipped with a Jordan product and trace inner product forms a Euclidean Jordan algebra. Thus, (1) can be understood as a decomposition of $\mathcal{Z}_\Gamma$ into Euclidean simple Jordan algebras. Theorem 1 is the existence result and actual computation of structure constants and the orthogonal matrix U_Γ is in general a hard technical task. A complete proof of Theorem 1 can be found in the Supplementary Material (Graczyk et al. (2022)). We tried to ensure that our arguments are concrete and should be easier to understand for the reader who is not familiar with representation theory.

The main novel results of the paper are:

(a) new Bayesian model selection procedure within Gaussian models invariant by a permutation subgroup, Section 4.1,

(b) explicit formulas for Gamma integrals, normalizing constants of densities of Diaconis–Ylvisaker conjugate prior for K and of the MLE of Σ on $\mathcal{P}_\Gamma$, Bayes factors, which are necessary for performing a), Theorems 8 and 9 in Section 3,

(c) efficient algorithm for finding a decomposition (1) when the subgroup Γ is cyclic, Theorems 5 and 6 in Section 2.4,

(d) simulations that visualize the performance of the method in low and high-dimensional examples, Section 4.2, Section 5 and Section 4 of the Supplementary Material (Graczyk et al. (2022)).

Ad (a). We are aware of three papers which concern model selection in the space of colored graphical model, namely Gehrmann (2011), Massam, Li and Gao (2018), Li, Gao and Massam (2020).

In Gehrmann (2011), the author used the lattice structure of the colored graphical model classes and applied Edwards–Havránek model selection procedure to $p = 4$ and $p = 5$ examples, admitted that applying this method to high-dimensional problems requires additional work.

Both papers (Massam, Li and Gao (2018) and Li, Gao and Massam (2020)) used Bayesian methods and allow for model selection in the space of RCON models (which is a superclass of RCOP models introduced in Højsgaard and Lauritzen (2008)) and for arbitrary graphs describing conditional independencies in a vector. Such generality comes at a certain cost: as the authors were not able to compute normalizing constants for such general models, they had to approximate these constants or bypass the problem (which comes with a significant increase in computational complexity): we quote a few lines from these articles that describe the situation well.

- Massam, Li and Gao (2018): *However, just as sampling schemes for the G-Wishart distribution are not recommended for computation of (normalizing constant) $I_G(\delta, D)$ and model selection in higher dimensions, our sampling scheme is not recommended for computing (normalizing constant) $I_G(\delta, D)$ in high dimensions.*
- Li, Gao and Massam (2020): *The model G^* with an additional edge is then compared to the current model G using the Bayes factor (...), which itself is computed with the help of the double reversible jump MCMC algorithm. (...) We thus avoid computing these quantities which are the usual computational stumbling blocks in graphical Gaussian model selection.*

Our approach to the Bayesian model selection is much simpler as we were able to compute normalizing constants of Diaconis–Ylvisaker conjugate priors for K.

Ad (b). We note that a general form of a density of the MLE under our assumptions was already written in Andersson (1975) and in more explicit form in Andersson and Madsen (1998). However, an explicit expression for the normalizing constant of density of $\hat{\Sigma}$ or Diaconis–Ylvisaker conjugate prior was not the object of interest of the Danish school and it is crucial for the Bayes paradigm and the Bayesian model selection.

Still, there are certain results in their numerous works that can be compared with our formulas. In particular, Andersson and Madsen ((1998), equation (A.4)) gives a formula for $\mathbb{E}[\mathrm{Det}(\hat{\Sigma})^{\alpha}]$, which is consistent with our results. Indeed, after substitution of $(d_i, nk_i/d_i, r_i)_i$ for $(d_\mu, n_\mu, p_\mu)_\mu$, the right-hand side of their formula coincides with $2^{\alpha p} n^{-\alpha p} \Gamma_{\mathcal{P}_\Gamma}(\alpha + n/2)/\Gamma_{\mathcal{P}_\Gamma}(n/2)$ in our notation (see Theorem 8). Further, in Andersson, Brøns and Jensen ((1983), Section 8) explicit formula for normalizing constants of the density of eigenvalues of $\hat{\Sigma}$ is given. However, as distribution of eigenvalues of a random matrix does not determine the distribution of this matrix, our formulas do not follow from these results.

In some very special cases, normalizing constants for Diaconis–Ylvisaker conjugate prior are given in Massam, Li and Gao (2018).

Ad (c). In order to compute normalizing constants in our model, one needs to know explicit decomposition (1), that is, the structure constants and the orthogonal matrix U_Γ. The same issue can be seen in Jensen ((1988), Theorem 1), which is the existence result (like our Theorem 1) and does not give the answer how should one proceed to find such decomposition. In order for this theory to be applied, we proved that when Γ is a cyclic subgroup, then we can efficiently find explicit decomposition (1) for arbitrary p. This practical aspect of our work has not been addressed before. To our knowledge, our paper is the first one to identify a nontrivial class of subgroups for which all objects can be calculated explicitly.

For a moment, let us consider the more general situation of Gaussian graphical models with conditional independence structure encoded by a noncomplete graph G. Then one can introduce symmetry restrictions (RCOP) by requiring that the precision matrix K is invariant under some subgroup Γ of $\mathfrak{S}_p$. However, when G is not complete, not all subgroups are suited to the problem. In such cases, one has to require that Γ belongs to the automorphism group $\mathrm{Aut}(G)$ of G. If a graph G is sparse, then $\mathrm{Aut}(G)$ may be very small and it is natural to expect that the vast majority of subgroups of $\mathrm{Aut}(G)$ are actually cyclic. Moreover, finding the structure constants for a general group is much more expensive and in some situations it may not be worth to consider the problem in its full generality. We consider our work as a first step toward the rigorous analytical treatment of Bayesian model selection in the space of graphical Gaussian models invariant under the action of $\Gamma \subset \mathfrak{S}_p$ when conditional independencies are allowed.

Moreover, we offer here a new heuristic approach to colored graphical models using our "full graph" approach. It was already observed in Højsgaard and Lauritzen (2008) that the color pattern of the covariance matrix and the precision matrix are the same (i.e., they belong to the space $\mathcal{Z}_\Gamma$). The same applies to the off-diagonal elements of the partial correlation matrix. Our procedure allows one to find the color pattern of the covariance matrix. Since our model does not suppose any preliminary conditional independence structure, the corresponding graph is complete and there are no zeros in the partial correlation matrix. However, if the true graph is not complete, it is natural to expect from the model that similar entries of the partial correlation matrix (in particular those which are close to 0) are colored in the same way. Thus, to recover the true graph we may threshold the values of the partial correlation matrix. More precisely, we choose a threshold $\alpha > 0$ and we construct a colored graph G by maintaining the color pattern previously found and requiring that for $i \neq j$,

$$i \sim j \quad \text{if and only if} \quad \frac{|k_{ij}|}{\sqrt{k_{ii}k_{jj}}} > \alpha,$$

where $K = (k_{ij})_{i,j}$ is the precision matrix. Resulting graph G is in general not complete and the corresponding space of admissible covariance matrices is still invariant under the action of the subgroup found by our procedure; thus we obtain a RCOP model. We applied this approach to a real data example in Section 4 of the Supplementary Material.

There are also several recent papers which use a version of Theorem 1. The subject of Soloveychik, Trushin and Wiesel (2016) is estimation of complex covariance matrices in complex random vectors in non-Gaussian models invariant under the action of a fixed permutation subgroup; see also De Maio et al. (2016). We remark that the argument of Soloveychik, Trushin and Wiesel (2016) is based on representation theory over complex number fields, and as was noticed by them, the fundamental structure theorem is much simpler than Theorem 1 because of the difference between the representation theory over $\mathbb{C}$ and $\mathbb{R}$. In Shah and Chandrasekaran (2012), the authors consider the real case and sub-Gaussian model for which they establish rates of convergence of an estimator of Σ, empirical covariance matrix regularized by the action of a known permutation subgroup.

1.3. *Outline of the paper.* Let us consider the following example, which shows how Theorem 1 works.

EXAMPLE 1. For $p=3$ and $\Gamma=\mathfrak{S}_3$, the space of symmetric matrices X invariant under Γ, that is, such that $X_{ij}=X_{\sigma(i)\sigma(j)}$ for all $\sigma\in\Gamma$, is

$$\mathcal{Z}_\Gamma=\left\{\begin{pmatrix} a & b & b\\ b & a & b\\ b & b & a\end{pmatrix}; a,b\in\mathbb{R}\right\}.$$

The decomposition (1) yields $U_\Gamma:=(v_1,\ v_2,\ v_3)\in \mathrm{O}(3)$ with

$$v_1:=\begin{pmatrix}1/\sqrt{3}\\ 1/\sqrt{3}\\ 1/\sqrt{3}\end{pmatrix},\qquad v_2:=\begin{pmatrix}\sqrt{2/3}\\ -1/\sqrt{6}\\ -1/\sqrt{6}\end{pmatrix},\qquad v_3:=\begin{pmatrix}0\\ 1/\sqrt{2}\\ -1\sqrt{2}\end{pmatrix},$$

and

$$\begin{pmatrix} a & b & b\\ b & a & b\\ a & a & b\end{pmatrix}=U_\Gamma\cdot\begin{pmatrix} a+2b & & \\ & a-b & \\ & & a-b\end{pmatrix}\cdot U_\Gamma^\top.$$

Here, $L=2$, $k_1/d_1=1$, $k_2/d_2=2$, $\mathbb{K}_1=\mathbb{K}_2=\mathbb{R}$, $d_1=d_2=1$.

We see immediately in the example above that, following the decomposition (1), the trace $\mathrm{Tr}[X]=a+2b+2(a-b)$ and the determinant $\mathrm{Det}(X)=(a+2b)(a-b)^2$ can be readily obtained. Similarly, using (1) allows us to easily obtain $\mathrm{Det}(X)$ and $\mathrm{Tr}[X]$ in general.

In Section 3, we will see that having the explicit formulas for $\mathrm{Det}(X)$ and $\mathrm{Tr}[X]$, in turn, allows us to derive the analytic expression of the Gamma function on $\mathcal{P}_\Gamma=\mathcal{Z}_\Gamma\cap\mathrm{Sym}^+(p;\mathbb{R})$, defined as

$$\Gamma_{\mathcal{P}_\Gamma}(\lambda):=\int_{\mathcal{P}_\Gamma}\mathrm{Det}(X)^\lambda e^{-\mathrm{Tr}[X]}\varphi_\Gamma(X)\,\mathrm{d}X,$$

where $\varphi_\Gamma(X)\,\mathrm{d}X$ is the invariant measure on $\mathcal{P}_\Gamma$ (see Definition 10 and Proposition 7) and $\mathrm{d}X$ denotes the Euclidean measure on the space $\mathcal{Z}_\Gamma$ with the trace inner product.

With our results, we can derive the analytic expression of the normalizing constant $I_\Gamma(\delta,D)$ of the Diaconis–Ylvisaker conjugate prior on $K=\Sigma^{-1}$ with density, with respect to the Euclidean measure on $\mathcal{Z}_\Gamma$, equal to

$$f(K;\delta,D)=\frac{1}{I_\Gamma(\delta,D)}\mathrm{Det}(K)^{(\delta-2)/2}e^{-\frac{1}{2}\mathrm{Tr}[K\cdot D]}\mathbf{1}_{\mathcal{P}_\Gamma}(K)$$

for appropriate values of the scalar hyperparameter δ and the matrix hyperparameter $D\in\mathcal{P}_\Gamma$. By analogy with the G-Wishart distribution, defined in the context of the graphical Gaussian models, Markov with respect to an undirected graph G on the cone P_G of positive definite matrices with zero entry (i, j) whenever there is no edge between the vertices i and j in G, (see Maathuis et al. (2018)), we can call the distribution with density $f(K; \delta, D)$, the RCOP-Wishart (RCOP is the name coined in Højsgaard and Lauritzen (2008) for graphical Gaussian models with restrictions generated by permutation symmetry). It is important to note here that if Σ is in $\mathcal{P}_\Gamma$, so is $K = \Sigma^{-1}$ so that K can also be decomposed according to (1). Equipped with all these results, we compute the Bayes factors comparing models pairwise and perform model selection. We will indicate in Section 4 how to travel through the space of subgroups of the symmetric group.

In Section 3, we also derive the distribution of the maximum likelihood estimate (henceforth abbreviated MLE) of Σ and show that for $n \geq \max_{i=1,\dots,L}\{r_i d_i/k_i\}$ it has a density equal to

$$\frac{\operatorname{Det}(X)^{n/2} e^{-\frac{1}{2}\operatorname{Tr}[X\cdot\Sigma^{-1}]}}{\operatorname{Det}(2\Sigma)^{n/2}\Gamma_{\mathcal{P}_\Gamma}(\frac{n}{2})}\varphi_\Gamma(X)\mathbf{1}_{\mathcal{P}_\Gamma}(X).$$

Clearly, the key to computing the Gamma integral on $\mathcal{P}_\Gamma$, the normalizing constant $I_\Gamma(\delta, D)$ or the density of the MLE of Σ is, for each $\Gamma \subset \mathfrak{S}_p$, to obtain the block diagonal matrix with diagonal block entries $M_{\mathbb{K}_i}(x_i) \otimes I_{k_i/d_i}$, $i = 1, \dots, L$, in the decomposition (1). In principle, we have to derive the invariant measure φ_Γ and find the structure constants $(k_i, d_i, r_i)_{i=1}^L$. This goal can be achieved by constructing an orthogonal matrix U_Γ and using (1). However, doing so for every Γ visited during the model selection process is computationally heavy.

We will show that for small to moderate dimensions, we can obtain the structure constants as well as the expression of $\operatorname{Det}(X)$ and $\varphi_\Gamma(X)$ without having to compute U_Γ. Indeed, as indicated in Lemma 4, for any $X \in \mathcal{P}_\Gamma$, $\operatorname{Det}(X)$ admits a unique irreducible factorization of the form

$$\operatorname{Det}(X) = \prod_{i=1}^{L} \operatorname{Det}\big(M_{\mathbb{K}_i}(x_i)\big)^{k_i/d_i} = \prod_{j=1}^{L} f_j(X)^{a_j} \quad (X \in \mathcal{Z}_\Gamma), \tag{2}$$

where each a_j is a positive integer, each $f_j(X)$ is an irreducible polynomial of $X \in \mathcal{Z}_\Gamma$, and $f_i \neq f_j$ if $i \neq j$. The constants k_i, d_i, r_i are obtained by identification of the two expressions of $\operatorname{Det}(X)$ in (2). Factorization of a homogeneous polynomial $\operatorname{Det}(X)$ can be performed using standard software such as either MATHEMATICA or PYTHON.

Due to computational complexity, for bigger dimensions, it is difficult to obtain the irreducible factorization of $\operatorname{Det}(X)$. For special cases such as the case where the subgroup Γ is a cyclic group, we give (Section 2.4) a simple construction of the matrix U_Γ, and thus, for any dimension p, we can do model selection in the space of models invariant under the action of a cyclic group. We argue that restriction to cyclic groups is not as limiting as it may look. The formula for the number of different colorings $c_p = \#\{\mathcal{Z}_\Gamma; \Gamma \subset \mathfrak{S}_p\}$ for given p is unknown. Obviously, it is bounded from above by the number of all subgroups of $\mathfrak{S}_p$, because different subgroups may produce the same coloring (e.g., in Example 1 we have $\mathcal{Z}_{\mathfrak{S}_3} = \mathcal{Z}_{\langle(1,2,3)\rangle}$). On the other hand, it is known (see Lemma 15) that c_p is bounded from below by the number of distinct cyclic subgroups, which grows rapidly with p (see OEIS[1] sequence A051625). In particular, for $p = 18$,[2] we have $c_p \in (7.1 \cdot 10^{14}, 7.6 \cdot 10^{18})$; see also Table 1. The lower bound for c_p indicates that the colorings obtained from cyclic subgroups form a rich subfamily of all possible colorings.

The procedure to do model selection will be described in Section 4 and we will illustrate this procedure with Frets' data (see Frets (1921)) and several examples for selection within

[1]The On-Line Encyclopedia of Integer Sequences, https://oeis.org/.

[2]The number of subgroups of $\mathfrak{S}_p$ is unknown for $p > 18$; see Holt (2010) and OEIS sequence A005432.

cyclic groups, including a high-dimensional example with $p = 100$ (Section 5) and a real data example (Miller et al. (2005)) with $p = 150$ in Section 4 of the Supplementary Material.

2. Preliminaries and structure constants. In this section, we present methods to calculate the structure constants of a decomposition given in Theorem 1. Additions to this section can be found in Section 3 of the Supplementary Material (Graczyk et al. (2022)).

2.1. *Notation.* Let $\mathrm{Mat}(n, m; \mathbb{R})$, $\mathrm{Sym}(n; \mathbb{R})$ denote the linear spaces of real $n \times m$ matrices and symmetric real $n \times n$ matrices, respectively. Let $\mathrm{Sym}^+(n; \mathbb{R})$ be the cone of symmetric positive definite real $n \times n$ matrices. $A^\top$ denotes the transpose of a matrix A. Det and Tr denote the usual determinant and trace in $\mathrm{Mat}(n, n; \mathbb{R})$.

For $A \in \mathrm{Mat}(m, n; \mathbb{R})$ and $B \in \mathrm{Mat}(m', n'; \mathbb{R})$, we denote by $A \oplus B$ the matrix $\left(\begin{smallmatrix} A & O \\ O & B \end{smallmatrix}\right) \in \mathrm{Mat}(m + m', n + n'; \mathbb{R})$, and by $A \otimes B$ the Kronecker product of A and B. For a positive integer r, we write $B^{\oplus r}$ for $I_r \otimes B \in \mathrm{Mat}(rm', rn'; \mathbb{R})$.

Let p denote the fixed number of vertices of a graph and let $\mathfrak{S}_p$ denote the symmetric group. We write permutations in cycle notation, meaning that $(i_1, i_2, \dots, i_n)$ maps i_j to i_{j+1} for $j = 1, \dots, r - 1$ and i_n to i_1. By $\langle \sigma_1, \dots, \sigma_k \rangle$, we denote the group generated by permutations $\sigma_1, \dots, \sigma_k$. The composition (product) of permutations $\sigma, \sigma' \in \mathfrak{S}_p$ will be denoted by $\sigma \circ \sigma'$.

DEFINITION 2. For a subgroup $\Gamma \subset \mathfrak{S}_p$, we define the space of symmetric matrices invariant under Γ, or the vector space of colored matrices,

$$\mathcal{Z}_\Gamma := \{x \in \mathrm{Sym}(p; \mathbb{R}); x_{ij} = x_{\sigma(i)\sigma(j)} \text{ for all } \sigma \in \Gamma\},$$

and the cone of positive definite matrices valued in $\mathcal{Z}_\Gamma$,

$$\mathcal{P}_\Gamma := \mathcal{Z}_\Gamma \cap \mathrm{Sym}^+(p; \mathbb{R}).$$

We note that the same colored space and cone can be generated by two different subgroups: in Example 1, the subgroup $\Gamma' = \langle (1, 2, 3) \rangle$ generated by the permutation $\sigma = (1, 2, 3)$ is such that $\Gamma' \neq \Gamma$ but $\mathcal{Z}_{\Gamma'} = \mathcal{Z}_\Gamma$. Let us define

$$\Gamma^* = \{\sigma^* \in \mathfrak{S}_p; x_{ij} = x_{\sigma^*(i)\sigma^*(j)} \text{ for all } x \in \mathcal{Z}_\Gamma\}.$$

Clearly, Γ is a subgroup of Γ^* and Γ^* is the unique largest subgroup of $\mathfrak{S}_p$ such that $\mathcal{Z}_{\Gamma^*} = \mathcal{Z}_\Gamma$ or, equivalently, such that the Γ^*- and Γ- orbits in $\{\{v_1, v_2\}; v_i \in V, i = 1, 2\}$ are the same. The group Γ^* is called the 2^*-closure of Γ. The group Γ is said to be 2^*-closed if $\Gamma = \Gamma^*$. Subgroups which are 2^*-closed are in bijection with the set of colored spaces. These concepts have been investigated in Wielandt (1969), Siemons (1982) along with a generalization to regular colorings in Siemons (1983). The combinatorics of 2^*-closed subgroups is very complicated and little is known in general (Graham, Grötschel and Lovász (1995), page 1502). In particular, the number of such subgroups is not known, but brute-force search for small p indicates that this number is much less than the number of all subgroups of $\mathfrak{S}_p$ (see Table 1). Even though cyclic subgroups of $\mathfrak{S}_p$ are in general not 2^*-closed, each cyclic group corresponds to a different coloring (see Lemma 15).

For a permutation $\sigma \in \mathfrak{S}_p$, denote its matrix by

$$R(\sigma) := \sum_{i=1}^{p} E_{\sigma(i)i}, \tag{3}$$

where E_{ab} is the $p \times p$ matrix with 1 in the (a, b)-entry and 0 in other entries. The condition $x_{\sigma(i)\sigma(j)} = x_{ij}$ is then equivalent to $R(\sigma) \cdot x \cdot R(\sigma)^\top = x$. Consequently,

$$\mathcal{Z}_\Gamma = \{x \in \mathrm{Sym}(p; \mathbb{R}); R(\sigma) \cdot x \cdot R(\sigma)^\top = x \text{ for all } \sigma \in \Gamma\}. \tag{4}$$

DEFINITION 3. Let $\pi_\Gamma\colon \mathrm{Sym}(p;\mathbb{R}) \to \mathcal{Z}_\Gamma$ be the orthogonal projection on $\mathcal{Z}_\Gamma$, that is, the linear map such that for any $x \in \mathrm{Sym}(p;\mathbb{R})$ the element $\pi_\Gamma(x) \in \mathcal{Z}_\Gamma$ is uniquely determined by

$$\mathrm{Tr}[x \cdot y] = \mathrm{Tr}\big[\pi_\Gamma(x) \cdot y\big] \quad (y \in \mathcal{Z}_\Gamma). \tag{5}$$

In view of (4), it is clear that

$$\pi_\Gamma(x) = \frac{1}{|\Gamma|}\sum_{\sigma\in\Gamma} R(\sigma)\cdot x \cdot R(\sigma)^\top \tag{6}$$

satisfies the above definition. Here, $|\Gamma|$ denotes the order of Γ.

2.2. *$\mathcal{Z}_\Gamma$ as a Jordan algebra.* To derive analytic expression for Gamma-like functions on $\mathcal{P}_\Gamma$ it is convenient to see $\mathcal{Z}_\Gamma$ as a Euclidean Jordan algebra and $\mathcal{P}_\Gamma$ as the corresponding symmetric cone. This fact was already observed in Jensen (1988). We recall here the fundamentals of Jordan algebras; cf. Faraut and Korányi (1994). A Euclidean Jordan algebra is a Euclidean space $\mathcal{A}$ (endowed with the scalar product denoted by $\langle\cdot,\cdot\rangle$) equipped with a bilinear mapping (product)

$$\mathcal{A}\times\mathcal{A} \ni (x,y) \mapsto x\bullet y \in \mathcal{A}$$

such that for all x, y, z in $\mathcal{A}$:

(i) $x\bullet y = y \bullet x$,
(ii) $x\bullet((x\bullet x)\bullet y) = (x\bullet x)\bullet(x\bullet y)$,
(iii) $\langle x, y\bullet z\rangle = \langle x\bullet y, z\rangle$.

A Euclidean Jordan algebra is said to be simple if it is not a Cartesian product of two Euclidean Jordan algebras of positive dimensions. We have the following result.

PROPOSITION 2. *The Euclidean space $\mathcal{Z}_\Gamma$ with inner product $\langle x,y\rangle = \mathrm{Tr}[x\cdot y]$ and the Jordan product*

$$x\bullet y = \frac{1}{2}(x\cdot y + y\cdot x), \tag{7}$$

is a Euclidean Jordan algebra. This algebra is generally nonsimple.

PROOF. Since $\mathcal{Z}_\Gamma$ is a subset of the Euclidean Jordan algebra $\mathrm{Sym}(p;\mathbb{R})$, if it is endowed with Jordan product (7), conditions (i)–(iii) are automatically satisfied. Moreover, characterization (4) of $\mathcal{Z}_\Gamma$ implies that the Jordan product is closed in $\mathcal{Z}_\Gamma$, that is, $R(\sigma)\cdot(x\bullet y) = (x\bullet y)\cdot R(\sigma)$ for all $x, y \in \mathcal{Z}_\Gamma$ and $\sigma\in\Gamma$. The result follows. $\square$

Up to linear isomorphism, there are only five kinds of Euclidean simple Jordan algebras. Let $\mathbb{K}$ denote the set of either the real numbers $\mathbb{R}$, the complex ones $\mathbb{C}$ or the quaternions $\mathbb{H}$. Let us write $\mathrm{Herm}(r;\mathbb{K})$ for the space of $r\times r$ Hermitian matrices valued in $\mathbb{K}$. Then $\mathrm{Sym}(r;\mathbb{R})$, $r\geq 1$, $\mathrm{Herm}(r;\mathbb{C})$, $r\geq 2$, $\mathrm{Herm}(r;\mathbb{H})$, $r\geq 2$ are the first three kinds of Euclidean simple Jordan algebras and they are the only ones that will concern us. The determinant and trace in Jordan algebras $\mathrm{Herm}(r;\mathbb{K})$ will be denoted by det and tr (see Faraut and Korányi (1994), page 29), respectively, so that they can be easily distinguished from the determinant and trace in $\mathrm{Mat}(n,n;\mathbb{R})$ which we denote by Det and Tr.

To each Euclidean Jordan algebra $\mathcal{A}$, one can attach the set $\overline{\Omega}$ of Jordan squares, that is, $\overline{\Omega} = \{x\bullet x; x\in\mathcal{A}\}$. The interior Ω of $\overline{\Omega}$ is a symmetric cone, that is, it is self-dual and homogeneous. We say that Ω is irreducible if it is not the Cartesian product of two convex cones. One can prove that an open convex cone is symmetric and irreducible if and only if it is the symmetric cone Ω of some Euclidean simple Jordan algebra. Each simple Jordan algebra corresponds to a symmetric cone. The first three kinds of irreducible symmetric cones are thus, the symmetric positive definite real matrices $\mathrm{Sym}^+(r;\mathbb{R})$ for $r \geq 1$, complex Hermitian positive definite matrices $\mathrm{Herm}^+(r;\mathbb{C})$ and quaternionic Hermitian positive definite matrices $\mathrm{Herm}^+(r;\mathbb{H})$, $r \geq 2$.

It follows from Definition 2 and Proposition 2 that $\mathcal{P}_\Gamma$ is a symmetric cone. In Faraut and Korányi ((1994), Proposition III.4.5), it is stated that any symmetric cone is a direct sum of irreducible symmetric cones. As it will turn out, only three out of the five kinds of irreducible symmetric cones may appear in this decomposition.

Moreover, we will want to represent the elements of the symmetric cones in their real symmetric matrix representations. So, we recall that both $\mathrm{Herm}(r;\mathbb{C})$ and $\mathrm{Herm}(r;\mathbb{H})$ can be realized as real symmetric matrices, but of bigger dimension. For $z = a + bi \in \mathbb{C}$, define $M_\mathbb{C}(z) = \left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$. The function $M_\mathbb{C}$ is a matrix representation of $\mathbb{C}$. Similarly, any $r \times r$ complex matrix can be realized as a $(2r) \times (2r)$ real matrix by setting the correspondence

$$\mathrm{Mat}(r,r;\mathbb{C}) \ni (z_{i,j})_{1\leq i,j\leq r} \simeq \big(M_\mathbb{C}(z_{i,j})\big)_{1\leq i,j\leq r} \in \mathrm{Mat}(2r,2r;\mathbb{R}),$$

that is, an (i, j)-entry of a complex matrix is replaced by its 2×2 real matrix representation. Note that $M_\mathbb{C}$ maps the space $\mathrm{Herm}(r;\mathbb{C})$ of Hermitian matrices into the space $\mathrm{Sym}(2r;\mathbb{R})$ of symmetric matrices. For example,

$$M_\mathbb{C}\begin{pmatrix} a & c-di \\ c+di & b \end{pmatrix} = \begin{pmatrix} a & 0 & c & d \\ 0 & a & -d & c \\ c & -d & b & 0 \\ d & c & 0 & b \end{pmatrix}.$$

Moreover, by direct calculation one sees that

$$\mathrm{Det}\begin{pmatrix} a & 0 & c & d \\ 0 & a & -d & c \\ c & -d & b & 0 \\ d & c & 0 & b \end{pmatrix} = \det\begin{pmatrix} a & c-di \\ c+di & b \end{pmatrix}^2.$$

It can be shown that, in general,

$$\text{(8)} \qquad \mathrm{Det}\big(M_\mathbb{C}(Z)\big) = \big[\det(Z)\big]^2 \quad \text{and} \quad \mathrm{Tr}\big[M_\mathbb{C}(Z)\big] = 2\,\mathrm{tr}[Z] \quad \big(Z \in \mathrm{Herm}(r;\mathbb{C})\big).$$

Similarly, quaternions can be realized as a 4×4 matrix:

$$a + bi + cj + dk \simeq \begin{pmatrix} a+bi & -c+di \\ c+di & a-bi \end{pmatrix} \simeq \begin{pmatrix} a & -b & -c & -d \\ b & a & d & -c \\ c & -d & a & b \\ d & c & -b & a \end{pmatrix}.$$

Then quaternionic $r \times r$ matrices are realized as $(4r) \times (4r)$ real matrices. Thus, $M_\mathbb{H}$ maps $\mathrm{Herm}(r;\mathbb{H})$ into $\mathrm{Sym}(4r;\mathbb{R})$. Moreover, it is true that

$$\text{(9)} \qquad \mathrm{Det}\big(M_\mathbb{H}(Z)\big) = \big[\det(Z)\big]^4 \quad \text{and} \quad \mathrm{Tr}\big[M_\mathbb{H}(Z)\big] = 4\,\mathrm{tr}[Z] \quad \big(Z \in \mathrm{Herm}(r;\mathbb{H})\big).$$

2.3. *Determining the structure constants and invariant measure on $\mathcal{P}_\Gamma$.* As mentioned in the Introduction, in order to derive the analytic expression of the Gamma-like functions on $\mathcal{P}_\Gamma$, we need the structure constants $(k_i, d_i, r_i)_{i=1}^L$ as well as the invariant measure φ_Γ. However, due to Proposition 7 below, $\varphi_\Gamma(X)$ is expressed in terms of the polynomials $\det(x_i)$, where $x_i \in \mathrm{Herm}(r_i;\mathbb{K}_i)$, $i = 1, \ldots, L$, coming from decomposition (1). These can be derived from the decomposition of $\mathcal{Z}_\Gamma$. Let us note that the constants $(d_i)_i$ and $(k_i)_i$ depend only on the group Γ, while r_i depend on a particular representation of Γ, which is R.

In view of decomposition (1), for $X \in \mathcal{Z}_\Gamma$, define $\phi_i(X) = x_i \in \mathrm{Herm}(r_i; \mathbb{K}_i)$ for $i = 1, \dots, L$.

COROLLARY 3. *For $X \in \mathcal{Z}_\Gamma$, one has*

$$\mathrm{Det}(X) = \prod_{i=1}^{L} \det\bigl(\phi_i(X)\bigr)^{k_i}. \tag{10}$$

PROOF. By (1), we have

$$\begin{aligned}\mathrm{Det}(X) &= \prod_{i=1}^{L} \mathrm{Det}\bigl(M_{\mathbb{K}_i}(x_i) \otimes I_{k_i/d_i}\bigr) = \prod_{i=1}^{L} \mathrm{Det}\bigl(M_{\mathbb{K}_i}(x_i)\bigr)^{k_i/d_i} \\ &= \prod_{i=1}^{L} \bigl[\det(x_i)^{d_i}\bigr]^{k_i/d_i} = \prod_{i=1}^{L} \det(x_i)^{k_i},\end{aligned}$$

whence follows the formula. We have used (8) and (9) for the third equality above. □

LEMMA 4. *Assume that $\Gamma \subset \mathfrak{S}_p$ and that $(k_i, d_i, r_i)_{i=1}^{L}$ are the structure constants corresponding to $\mathcal{Z}_\Gamma$. Assume that we have an irreducible factorization*

$$\mathrm{Det}(X) = \prod_{j=1}^{L'} f_j(X)^{a_j} \quad (X \in \mathcal{Z}_\Gamma), \tag{11}$$

where each a_j is a positive integer, each $f_j(X)$ is an irreducible polynomial of $X \in \mathcal{Z}_\Gamma$, and $f_i \neq f_j$ if $i \neq j$.

Then, $L = L'$, for each j there exists unique i such that $f_j(X)^{a_j} = \det(\phi_i(X))^{k_i}$, and:

(a) $k_i = a_j$,

(b) *r_i is the degree of $f_j(X) = \det(\phi_i(X))$,*

(c) *If $r_i > 1$, then d_i can be calculated from $r_i + d_i r_i(r_i - 1)/2 = \mathrm{rank}(P_j)$, where P_j is the linear operator defined by $\mathcal{Z}_\Gamma \ni x \mapsto P_j(x) = E_j \bullet x \in \mathcal{Z}_\Gamma$ and $E_j \in \mathcal{Z}_\Gamma$ is the gradient of $f_j(X)$ at $X = I_p$.*

REMARK 4. If $r_i = 1$, the determination of d_i is not needed for writing the block decomposition of $\mathcal{Z}_\Gamma$, since in this case $\mathbb{R} = \mathrm{Herm}(1; \mathbb{R}) = \mathrm{Herm}(1; \mathbb{C}) = \mathrm{Herm}(1; \mathbb{H})$ and, if k_i is divisible by 2 or by 4, we have $M_{\mathbb{K}_i}(x_i) \otimes I_{k_i/d_i} = x_i I_{k_i}$.

PROOF OF LEMMA 4. Since the determinant polynomial of a simple Jordan algebra is always irreducible Upmeier ((1986), Lemma 2.3(1)), comparing (10) and (11), we obtain $L = L'$, and that, for each j, there exists i such that $f_j(X)^{a_j} = \det(\phi_i(X))^{k_i}$. From this follows, also (a) and (b).

Observe that $r_i + d_i r_i(r_i - 1)/2 = \dim_{\mathbb{R}} \mathrm{Herm}(r_i; \mathbb{K}_i)$. Point (c) follows from the fact that P_j coincides with the projection $\bigoplus_{i=1}^{L} \mathrm{Herm}(r_i; \mathbb{K}_i) \to \mathrm{Herm}(r_i; \mathbb{K}_i)$. □

The practical significance of the method proposed in this lemma is that neither representation theory nor group theory is used. It is a strong advantage when we consider colorings corresponding to a large number of different groups, for which finding structure constants is very complicated.

REMARK 5. The factorization of multivariate polynomials over an algebraic number field can be done, for example, in PYTHON (see *sympy.polys.polytools.factor*) or in MATHEMATICA (see *Factor*). However, in order to make use of Lemma 4, one has to perform a factorization over the real number field. It turns out that the previously listed tools can be used for this purpose by selecting an appropriate *Extension* parameter. Indeed, in our setting, the irreducible factorization over the real number field coincides with the one over the real cyclotomic field

$$\mathbb{Q}\left[\zeta+\frac{1}{\zeta}\right]=\left\{\sum_{k=0}^{\varphi_E(M)/2-1} q_k\left(\zeta+\frac{1}{\zeta}\right)^k; q_k\in\mathbb{Q}, k=0,1,\dots,\varphi_E(M)/2-1\right\},$$

where ζ is the primitive Mth root $e^{2\pi i/M}$ of unity with M being the least common multiple of the orders of elements $\sigma\in\Gamma$, and $\varphi_E(M)$ is the number of positive integers up to M that are relatively prime to M Serre ((1977), Section 12.3).

An example showing the utility of Lemma 4 can be found in the Supplementary Material Graczyk et al. ((2022), Section 3.1).

2.4. *Finding structure constants and construction of the orthogonal matrix U_Γ when Γ is cyclic.* We now show that, when the group Γ is generated by one permutation $\sigma\in\mathfrak{S}_p$, the orthogonal matrix U_Γ can be constructed explicitly, and we obtain the structure constants r_i, k_i and d_i easily.

Let us consider the Γ-orbits in $\{1,2,\dots,p\}$. Let $\{i_1,\dots,i_C\}$ be a complete system of representatives of the Γ-orbits, and for each $c=1,\dots,C$, let p_c be the cardinality of the Γ-orbit through i_c. The order N of Γ equals the least common multiple of $p_1,p_2,\dots,p_C$ and one has $\Gamma=\{\mathrm{id},\sigma,\sigma^2,\dots,\sigma^{N-1}\}$. In what follows, we treat 0 as a multiple of N.

THEOREM 5. *Let $\Gamma=\langle\sigma\rangle$ be a cyclic group of order N. For $\alpha=0,1,\dots,\lfloor\frac{N}{2}\rfloor$, set*

$$r^*_\alpha=\#\{c\in\{1,\dots,C\};\alpha\, p_c \text{ is a multiple of } N\},$$

$$d^*_\alpha=\begin{cases}1 & (\alpha=0 \text{ or } N/2),\\ 2 & (\text{otherwise}).\end{cases}$$

Then we have $L=\#\{\alpha;r^*_\alpha>0\}$, $r=(r^*_\alpha;r^*_\alpha>0)$ and $k=d=(d^*_\alpha;r^*_\alpha>0)$.

Note that, r^*_0 equals the number C of cycles in a decomposition of a permutation.

EXAMPLE 6. Let us consider $\sigma=(1,2,3)(4,5)(6)\in\mathfrak{S}_6$. The three Γ-orbits are $\{1,2,3\}$, $\{4,5\}$ and $\{6\}$. Set $i_1=1$, $i_2=4$, $i_3=6$. Then $p_1=3$, $p_2=2$, $p_3=1$. We have $N=6$. We count $r^*_0=3$, $r^*_1=0$, $r^*_2=1$, $r^*_3=1$, so that $r=(3,1,1)$. Since $d=(1,2,1)$, we have $\mathcal{Z}_\Gamma\simeq\mathrm{Sym}(3;\mathbb{R})\oplus\mathrm{Herm}(1;\mathbb{C})\oplus\mathrm{Sym}(1;\mathbb{R})$.

For $c=1,\dots,C$, define $v^{(c)}_1,\dots,v^{(c)}_{p_c}\in\mathbb{R}^p$ by

$$v^{(c)}_1:=\sqrt{\frac{1}{p_c}}\sum_{k=0}^{p_c-1}e_{\sigma^k(i_c)},$$

$$v^{(c)}_{2\beta}:=\sqrt{\frac{2}{p_c}}\sum_{k=0}^{p_c-1}\cos\left(\frac{2\pi\beta k}{p_c}\right)e_{\sigma^k(i_c)}\quad(1\le\beta<p_c/2),$$

$$v_{2\beta+1}^{(c)} := \sqrt{\frac{2}{p_c}} \sum_{k=0}^{p_c-1} \sin\left(\frac{2\pi\beta k}{p_c}\right) e_{\sigma^k(i_c)} \quad (1 \le \beta < p_c/2),$$

$$v_{p_c}^{(c)} := \sqrt{\frac{1}{p_c}} \sum_{k=0}^{p_c-1} \cos(\pi k) e_{\sigma^k(i_c)} \quad (\text{if } p_c \text{ is even}).$$

THEOREM 6. *The orthogonal matrix U_Γ from Theorem 1 can be obtained by arranging column vectors $\{v_k^{(c)}\}$, $1 \le c \le C$, $1 \le k \le p_c$ in the following way: we put $v_k^{(c)}$ earlier than $v_{k'}^{(c')}$ if:*

(i) $\frac{[k/2]}{p_c} < \frac{[k'/2]}{p_{c'}}$, *or*
(ii) $\frac{[k/2]}{p_c} = \frac{[k'/2]}{p_{c'}}$ *and $c < c'$, or*
(iii) $\frac{[k/2]}{p_c} = \frac{[k'/2]}{p_{c'}}$ *and $c = c'$ and k is even and k' is odd.*

Proofs of the above results are presented in the Supplementary Material. We shall see there that $R(\sigma)$ acts on the 2-dimensional space spanned by $v_{2\beta}^{(c)}$ and $v_{2\beta+1}^{(c)}$ as a rotation with the angle $2\pi\beta/p_c$, $1 \le \beta < p_c/2$. The condition (i) means that the angle for $v_k^{(c)}$ is smaller than the one for $v_{k'}^{(c')}$.

EXAMPLE 7. We continue Example 6. According to Theorem 6,

$$U_\Gamma = \left(v_1^{(1)}, v_1^{(2)}, v_1^{(3)}, v_2^{(1)}, v_3^{(1)}, v_2^{(2)}\right)$$
$$= \begin{pmatrix} 1/\sqrt{3} & 0 & 0 & \sqrt{2/3} & 0 & 0 \\ 1/\sqrt{3} & 0 & 0 & -\sqrt{1/6} & 1/\sqrt{2} & 0 \\ 1/\sqrt{3} & 0 & 0 & -\sqrt{1/6} & -1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} & 0 & 0 & 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 0 & 0 & 0 & -1/\sqrt{2} \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

Then we have

$$U_\Gamma^\top \cdot R(\sigma^k) \cdot U_\Gamma = \begin{pmatrix} I_3 \otimes B_0(\sigma^k) & & \\ & B_2(\sigma^k) & \\ & & B_3(\sigma^k) \end{pmatrix},$$

where $B_0(\sigma^k) = 1$, $B_2(\sigma^k) = \mathrm{Rot}(\frac{2\pi k}{3}) \in \mathrm{GL}(2; \mathbb{R})$ and $B_3(\sigma^k) = (-1)^k$. Here, $\mathrm{Rot}(\theta)$ denotes the rotation matrix $\left(\begin{smallmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{smallmatrix}\right)$ for $\theta \in \mathbb{R}$.

The block diagonal decomposition of $\mathcal{Z}_\Gamma$ is

$$U_\Gamma^\top \cdot \mathcal{Z}_\Gamma \cdot U_\Gamma = \left\{ \begin{pmatrix} x_1 & & \\ & x_2 I_2 & \\ & & x_3 \end{pmatrix} ; x_1 \in \mathrm{Sym}(3; \mathbb{R}), x_2, x_3 \in \mathbb{R} \right\}.$$

REMARK 8. In the cyclic case, we have $k = d$ and so the formula (1) holds without the Kronecker product terms. Since $d_i \in \{1, 2\}$, the quaternionic case never occurs.

3. Gamma integrals and normalizing constants.

3.1. *Gamma integrals on irreducible symmetric cones.* Let Ω be one of the first three kinds of irreducible symmetric cones, that is, $\Omega = \mathrm{Herm}^+(r;\mathbb{K})$, where $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}, \mathbb{H}\}$. As before, determinant and trace on corresponding Euclidean Jordan algebras are denoted by det and tr. Then we have the relation

$$\dim\Omega = r + \frac{r(r-1)}{2}d,$$

where $d = 1$ if $\mathbb{K} = \mathbb{R}$, $d = 2$ if $\mathbb{K} = \mathbb{C}$ and $d = 4$ if $\mathbb{K} = \mathbb{H}$.

Recall that Euclidean measure is the volume measure induced by the Euclidean metric. Let $m(\mathrm{d}x)$ denote the Euclidean measure associated with the Euclidean structure defined on $\mathcal{A} = \mathrm{Herm}(r;\mathbb{K})$ by $\langle x, y\rangle = \mathrm{tr}[x \bullet y] = \mathrm{tr}[x \cdot y]$. The Gamma integral

$$\Gamma_\Omega(\lambda) := \int_\Omega \det(x)^\lambda e^{-\mathrm{tr}[x]} \det(x)^{-\dim\Omega/r} m(\mathrm{d}x)$$

is finite if and only if $\lambda > \frac{1}{2}(r-1)d = \dim\Omega/r - 1$ and in such case

$$\Gamma_\Omega(\lambda) = (2\pi)^{(\dim\Omega - r)/2}\Gamma(\lambda)\Gamma(\lambda - d/2)\cdots\Gamma\bigl(\lambda - (r-1)d/2\bigr). \tag{12}$$

Moreover, one has

$$\int_\Omega \det(x)^\lambda e^{-\mathrm{tr}[x\cdot y]} \det(x)^{-\dim\Omega/r} m(\mathrm{d}x) = \Gamma_\Omega(\lambda)\det(y)^{-\lambda} \tag{13}$$

for any $y \in \Omega$.

The measure $\mu_\Omega(\mathrm{d}x) = \det(x)^{-\dim\Omega/r} m(\mathrm{d}x)$ is invariant in the following sense. Let $\mathrm{G}(\Omega)$ be the linear automorphism group of Ω, that is, the set $\{g \in \mathrm{GL}(\mathcal{A}); g\,\Omega = \Omega\}$, where $\mathcal{A}$ is the associated Euclidean Jordan algebra. Then the measure μ_Ω is a $\mathrm{G}(\Omega)$-invariant measure in the sense that for any Borel measurable set B one has

$$\mu_\Omega(g^{-1}B) = \mu_\Omega(B) \quad \bigl(g \in \mathrm{G}(\Omega)\bigr).$$

3.2. *Gamma integrals on the cone $\mathcal{P}_\Gamma$.* We endow the space $\mathcal{Z}_\Gamma$ with the scalar product

$$\langle x, y\rangle = \mathrm{Tr}[x \cdot y] \quad (x, y \in \mathcal{Z}_\Gamma).$$

Let $\mathrm{d}X$ denote the Euclidean measure on the Euclidean space $(\mathcal{Z}_\Gamma, \langle\cdot,\cdot\rangle)$. Let us note that this normalization is not important in the Bayesian model selection procedure as there we always consider quotients of integrals.

EXAMPLE 9. Consider $p = 3$ and $\Gamma = \mathfrak{S}_3$. The space $\mathcal{Z}_\Gamma$ is 2-dimensional and it consists of matrices of the form (see Example 1)

$$X = \begin{pmatrix} a & b & b \\ b & a & b \\ b & b & a \end{pmatrix}$$

for $a, b \in \mathbb{R}$. Since $\|X\|^2 = \mathrm{Tr}[X^2] = 3a^2 + 6b^2 = v^\top v$ with $v^\top = (\sqrt{3}a, \sqrt{6}b)$, we have $\mathrm{d}X = \sqrt{3}\sqrt{6}\,\mathrm{d}a\,\mathrm{d}b = 3\sqrt{2}\,\mathrm{d}a\,\mathrm{d}b$.

Generally, if m_i denotes the Euclidean measure on $\mathcal{A}_i := \mathrm{Herm}(r_i;\mathbb{K}_i)$ with the inner product defined from the Jordan algebra trace (recall (8) and (9)), then (1) implies that for $X \in \mathcal{Z}_\Gamma$ we have

$$\|X\|^2 = \langle X, X\rangle = \sum_{i=1}^L \frac{k_i}{d_i}\mathrm{Tr}\bigl[M_{\mathbb{K}_i}(x_i)^2\bigr] = \sum_{i=1}^L k_i\,\mathrm{tr}\bigl[(x_i)^2\bigr],$$

which implies that

$$\mathrm{d}X = \prod_{i=1}^{L} (\sqrt{k_i})^{\dim \Omega_i} m_i(\mathrm{d}x_i) = e^{B_\Gamma} \prod_{i=1}^{L} m_i(\mathrm{d}x_i), \tag{14}$$

where

$$B_\Gamma := \frac{1}{2} \sum_{i=1}^{L} (\dim \Omega_i)(\log k_i). \tag{15}$$

DEFINITION 10. Let $\mathrm{G}(\mathcal{P}_\Gamma) = \{g \in \mathrm{GL}(p; \mathbb{R}); g\mathcal{P}_\Gamma = \mathcal{P}_\Gamma\}$ be the linear automorphism group of $\mathcal{P}_\Gamma$. We define the $\mathrm{G}(\mathcal{P}_\Gamma)$-invariant measure $\varphi_\Gamma(X)\,\mathrm{d}X$ by

$$\varphi_\Gamma(X) = e^{B_\Gamma} \left(\prod_{i=1}^{L} \frac{1}{\Gamma_{\Omega_i}(\dim \Omega_i / r_i)} \right) \int_{\mathcal{P}_\Gamma^*} e^{-\operatorname{Tr}[X \cdot Z]} \,\mathrm{d}Z,$$

where $\mathcal{P}_\Gamma^* = \{y \in \mathcal{Z}_\Gamma; \operatorname{Tr}[y \cdot x] > 0, \forall x \in \overline{\mathcal{P}_\Gamma} \setminus \{0\}\}$ is the dual cone of $\mathcal{P}_\Gamma$.

PROPOSITION 7. *We have*

$$\varphi_\Gamma(X) = \prod_{i=1}^{L} \det\big(\phi_i(X)\big)^{-\dim \Omega_i / r_i}. \tag{16}$$

The proofs of the following results of this section can be found in the Supplementary Material (Graczyk et al. (2022)).

DEFINITION 11. The Gamma function of $\mathcal{P}_\Gamma$ is defined by the following integral:

$$\Gamma_{\mathcal{P}_\Gamma}(\lambda) := \int_{\mathcal{P}_\Gamma} \operatorname{Det}(X)^\lambda e^{-\operatorname{Tr}[X]} \varphi_\Gamma(X) \,\mathrm{d}X, \tag{17}$$

whenever it converges.

THEOREM 8. *The integral* (17) *converges if and only if*

$$\lambda > \max_{i=1,\dots,L} \left\{ \frac{(r_i - 1)d_i}{2k_i} \right\} \tag{18}$$

and, for these values of λ, *we have*

$$\Gamma_{\mathcal{P}_\Gamma}(\lambda) = e^{-A_\Gamma \lambda + B_\Gamma} \prod_{i=1}^{L} \Gamma_{\Omega_i}(k_i \lambda), \tag{19}$$

where Γ_{Ω_i} *is given in* (12), B_Γ *in* (15) *and*

$$A_\Gamma := \sum_{i=1}^{L} r_i\, k_i \log k_i. \tag{20}$$

Moreover, if $Y \in \mathcal{P}_\Gamma$ *and* (18) *holds true, then*

$$\int_{\mathcal{P}_\Gamma} \operatorname{Det}(X)^\lambda e^{-\operatorname{Tr}[Y \cdot X]} \varphi_\Gamma(X) \,\mathrm{d}X = \Gamma_{\mathcal{P}_\Gamma}(\lambda) \operatorname{Det}(Y)^{-\lambda}. \tag{21}$$

We also have the following result.

THEOREM 9. *If $Y \in \mathcal{P}_\Gamma$ and*

$$\lambda > \max_{i=1,\dots,L}\left\{-\frac{1}{k_i}\right\},$$

then

$$\int_{\mathcal{P}_\Gamma} \mathrm{Det}(X)^\lambda e^{-\mathrm{Tr}[Y\cdot X]}\,\mathrm{d}X = e^{-A_\Gamma\lambda - B_\Gamma}\prod_{i=1}^{L}\Gamma_{\Omega_i}\left(k_i\,\lambda + \frac{\dim\Omega_i}{r_i}\right)\frac{\varphi_\Gamma(Y)}{\mathrm{Det}(Y)^\lambda}. \tag{22}$$

3.3. *RCOP-Wishart laws on $\mathcal{P}_\Gamma$.* Let $\Sigma \in \mathcal{P}_\Gamma \subset \mathrm{Sym}^+(p;\mathbb{R})$ and consider i.i.d. random vectors $Z^{(1)},\dots,Z^{(n)}$ following the $\mathrm{N}_p(0,\Sigma)$ distribution. Define $U_i = Z^{(i)}\cdot Z^{(i)\top}$, $i = 1,\dots,n$ and $U = \sum_{i=1}^n U_i$. We note that such model is clearly not invariant under changing the scale of variables: random vector $\mathrm{diag}(\underline{\alpha})\cdot Z^{(1)}$ for $\underline{\alpha} \in \mathbb{R}^p$ is in general not invariant under any permutation subgroup. Such issue is an immanent property of RCON models (a generalization of RCOP models) and was noticed already in Højsgaard and Lauritzen (2008). The authors recommend to keep all variables in the same units.

Our aim is to analyze the probability distribution of the random matrix

$$W_n = \pi_\Gamma(U) = \pi_\Gamma(U_1 + \cdots + U_n) = \pi_\Gamma(U_1) + \cdots + \pi_\Gamma(U_n).$$

In the rest of this section, we find n_0 such that for $n \ge n_0$ the random matrix W_n follows an absolutely continuous law, and we compute its density. Further, we extend the shape parameter to a continuous range and define the RCOP-Wishart law on $\mathcal{P}_\Gamma$.

We start with the following easy result.

LEMMA 10. *For any $\theta \in \mathrm{Sym}^+(p;\mathbb{R})$, we have*

$$\mathbb{E}e^{-\mathrm{Tr}[\theta\cdot\pi_\Gamma(U_1)]} = \mathrm{Det}\big(I_p + 2\,\Sigma\cdot\pi_\Gamma(\theta)\big)^{-1/2}.$$

PROOF. Using (5) repeatedly, we have

$$\mathrm{Tr}\big[\theta\cdot\pi_\Gamma(U_1)\big] = \mathrm{Tr}\big[\pi_\Gamma(\theta)\cdot\pi_\Gamma(U_1)\big] = \mathrm{Tr}\big[\pi_\Gamma(\theta)\cdot U_1\big].$$

The assertion follows from the usual multivariate Gauss integral. □

PROPOSITION 11. *The law of W_n is absolutely continuous on $\mathcal{P}_\Gamma$ if and only if*

$$n \ge n_0 := \max_{i=1,\dots,L}\left\{\frac{r_i d_i}{k_i}\right\}. \tag{23}$$

If $n \ge n_0$, then its density function with respect to $\mathrm{d}X$ is given by

$$\frac{\mathrm{Det}(X)^{n/2}e^{-\frac{1}{2}\mathrm{Tr}[X\cdot\Sigma^{-1}]}}{\mathrm{Det}(2\Sigma)^{n/2}\Gamma_{\mathcal{P}_\Gamma}(\frac{n}{2})}\varphi_\Gamma(X)\mathbf{1}_{\mathcal{P}_\Gamma}(X). \tag{24}$$

PROOF. With $\lambda = n/2$, condition (18) becomes

$$n > \max_{i=1,\dots,L}\left\{\frac{(r_i-1)d_i}{k_i}\right\}.$$

Since the quotient k_i/d_i is an integer, the last condition is equivalent to (23).

In view of Lemma 10, it is enough to show that W_n has density (24) if and only if for any $\theta \in \mathcal{P}_\Gamma$,

$$\mathbb{E}e^{-\mathrm{Tr}[\theta\cdot W_n]} = \mathrm{Det}(I_p + 2\,\Sigma\cdot\theta)^{-n/2}. \tag{25}$$

This follows directly from (21). □

It is known that the MLE exists and is unique if and only if the sufficient statistic lies in the interior of its convex support; see Barndorff-Nielsen (2014). It is clear that if (23) is not satisfied, then the support of W_n is contained in the boundary of $\mathcal{P}_\Gamma$. Recall that the orthogonal projection π_Γ is given by (6).

COROLLARY 12. *The MLE of Σ exists if and only if the number of samples n satisfies* (23). *If it exists, it is given by*

$$\hat{\Sigma} = \frac{1}{n}\pi_\Gamma(U_1 + \cdots + U_n).$$

The above result has been already proven in Andersson ((1975), Theorem 5.9) (see also Andersson and Madsen (1998), Sections A.3, A.4).

REMARK 12. If $U = (U_1 + \cdots + U_n)/n$ is positive definite, then U can be regarded as an empirical covariance matrix. The Kullback–Leibler divergence between $\mathrm{N}_p(0, U)$ and $\mathrm{N}_p(0, \Sigma)$ is equal to $\frac{1}{2}\{\log\det\Sigma + \operatorname{tr} U\Sigma^{-1} - \log\det(U) - p\}$, which is obviously minimized by the MLE $\hat{\Sigma}$. Therefore, Corollary 12 implies that $\pi_\Gamma(U)$ is the Kullback–Leibler projection of U onto $\mathcal{P}_\Gamma$ (Goutis and Robert (1998)). We note that the KL projection in general is not linear, whereas π_Γ clearly is.

Let us recall that the MLE of Σ in the standard normal model exists if and only if $n \geq p$. We recover this case for $\Gamma = \{\mathrm{id}\}$, since then we have $L = 1$, $r_1 = p$ and $k_1 = d_1 = 1$.

When $n < n_0$, the law of W_n is singular, and it can be described as a direct product of the singular Wishart laws on the irreducible symmetric cones Ω_i; see, for example, Hassairi and Lajmi (2001).

DEFINITION 13. Let $\eta > \max\{(r_i - 1)\frac{d_i}{k_i}; i = 1, \ldots, L\}$ and $\Sigma \in \mathcal{P}_\Gamma$. The RCOP–Wishart law $W^\Gamma_{\eta,\Sigma}$ is defined by its density

$$W^\Gamma_{\eta,\Sigma}(\mathrm{d}X) = \frac{\operatorname{Det}(X)^{\eta/2} e^{-\frac{1}{2}\operatorname{Tr}[X\cdot\Sigma^{-1}]}}{\operatorname{Det}(2\Sigma)^{\eta/2}\Gamma_{\mathcal{P}_\Gamma}(\frac{\eta}{2})}\varphi_\Gamma(X)\mathbf{1}_{\mathcal{P}_\Gamma}(X)\,\mathrm{d}X. \tag{26}$$

With this new notation, we see that if (23) is satisfied, then $W_n \sim W^\Gamma_{n,\Sigma}$.

LEMMA 13. *The Jacobian of the transformation*

$$\mathcal{P}_\Gamma \ni X \mapsto X^{-1} \in \mathcal{P}_\Gamma$$

equals $\varphi_\Gamma(X^{-1})^2$.

Proof of the lemma can be found in the Supplementary Material. By this lemma, we obtain another useful formula for the invariant measure, namely

$$\varphi_\Gamma(X) = \operatorname{Det}_{\mathrm{End}}(\mathbb{P}_X)^{-1/2} \quad (X \in \mathcal{Z}_\Gamma),$$

where $\operatorname{Det}_{\mathrm{End}}$ is the determinant in the space of endomorphisms of $\mathcal{Z}_\Gamma$ and for any $X \in \mathcal{Z}_\Gamma$ by $\mathbb{P}_X$ we denote the linear map on $\mathcal{Z}_\Gamma$ to itself defined by $\mathbb{P}_X Y = X \cdot Y \cdot X$. Lemma 13 gives also the following result.

PROPOSITION 14. *Let $W \sim W^\Gamma_{\eta,\Sigma}$ with $\eta > \max\{(r_i - 1)d_i/k_i; i = 1, \ldots, L\}$ and $\Sigma \in \mathcal{P}_\Gamma$. Then its inverse $Y = W^{-1}$ has density*

$$\frac{\operatorname{Det}(Y)^{-\eta/2} e^{-\frac{1}{2}\operatorname{Tr}[Y^{-1}\cdot\Sigma^{-1}]}}{\operatorname{Det}(2\Sigma)^{\eta/2}\Gamma_{\mathcal{P}_\Gamma}(\frac{\eta}{2})}\varphi_\Gamma(Y)\mathbf{1}_{\mathcal{P}_\Gamma}(Y).$$

3.4. *The Diaconis–Ylvisaker conjugate prior for K.* The Diaconis–Ylvisaker conjugate prior (Diaconis and Ylvisaker (1979)) for the canonical parameter $K = \Sigma^{-1}$ is given by

$$f(K; \delta, D) = \frac{1}{I_\Gamma(\delta, D)} \operatorname{Det}(K)^{(\delta-2)/2} e^{-\frac{1}{2}\operatorname{Tr}[K\cdot D]} \mathbf{1}_{\mathcal{P}_\Gamma}(K),$$

for hyperparameters $\delta > 2\max\{1 - 1/k_i; i = 1, \dots, L\}$ and $D \in \mathcal{P}_\Gamma$. By (22), the normalizing constant is equal to

$$I_\Gamma(\delta, D) = e^{-A_\Gamma(\delta-2)/2 - B_\Gamma} \prod_{i=1}^{L} \Gamma_{\Omega_i}\left(k_i \frac{\delta-2}{2} + \frac{\dim \Omega_i}{r_i}\right) \frac{\varphi_\Gamma(\frac{1}{2}D)}{\operatorname{Det}(\frac{1}{2}D)^{(\delta-2)/2}}, \tag{27}$$

where A_Γ, B_Γ and φ_Γ are given in (20), (15) and (16).

We note that despite the fact that the choice of hyperparameters is not scale invariant, statisticians usually take $\delta = 3$ and $D = I_p$; see, for example, Massam, Li and Gao (2018).

4. Model selection. Bayesian model selection on all colored spaces seems at the moment intractable. This is due in great part to a poor combinatorial description of the colored spaces $\mathcal{Z}_\Gamma$. In particular, the number of such spaces, that is, $\#\{\mathcal{Z}_\Gamma; \Gamma \in \mathfrak{S}_p\}$ is generally unknown for large p. It was shown in Gehrmann (2011) that these colorings constitute a lattice with respect to the usual inclusion of subspaces. However, the structure of this lattice is rather complicated and is unobtainable for big p. This, in turn, does not allow to define a Markov chain with known transition probabilities on such colorings. Finally, the fundamental problem, which prevents us from doing Bayesian model selection on all colored spaces for arbitrary p is the following. In order to compute Bayes factors, one has to be able to find the structure constants $(k_i, d_i, r_i)_{i=1}^{L}$ for arbitrary subgroups of $\mathfrak{S}_p$. This is equivalent to finding irreducible representations over reals for an arbitrary finite group, which is very hard in general, although general algorithms have been developed for this issue (see Plesken and Souvignier (1996)).

In this section, we are making a step forward in the problem of model selection for colored models in two ways. In Section 4.1, we use the results of Section 2.4, to obtain the structure constants when we restrict our search to the space of colored models generated by a cyclic group, that is, when $\Gamma = \langle\sigma\rangle$ for $\sigma \in \mathfrak{S}_p$ and we propose a model selection procedure restricted to the cyclic colorings. In Section 4.2, we use Lemma 4 and Remark 5 to obtain the irreducible representations of $\mathcal{Z}_\Gamma$ and the structure constants by factorization of the determinant. We apply this technique to do model selection for the four-dimensional example given by Frets' data since, in that case, there are only 22 models and we can compute all the Bayes factors.

4.1. *Model selection within cyclic groups.* The smaller space of cyclic colorings has a much better combinatorial description. In particular, the following result can be proved.

LEMMA 15. *If $\mathcal{Z}_{\langle\sigma\rangle} = \mathcal{Z}_{\langle\sigma'\rangle}$ for some $\sigma, \sigma' \in \mathfrak{S}_p$, then $\langle\sigma\rangle = \langle\sigma'\rangle$.*

This result allows us to calculate the number of different colorings corresponding to cyclic groups, that is, the number of labeled cyclic subgroups of the symmetric group $\mathfrak{S}_p$, which can be found in OEIS, sequence A051625 (see the last column of Table 1).

We will present two applications of the Metropolis–Hastings algorithm. In the first one, the Markov chain will move on the space of cyclic groups. The drawback of this first approach is that we need to compute the proposal distribution g, whose computational complexity grows faster than quadratically as p increases (see (29)). In the second algorithm, we consider a larger state space $\mathfrak{S}_p$, which allows us to consider an easy proposal distribution. However, this comes at the cost of slower convergence of the posterior probabilities (see Theorem 16).

TABLE 1
Number of all subgroups of a symmetric group, number of their conjugacy classes, number of different colorings and a number of cyclic groups

p	#subgroups of $\mathfrak{S}_p$	#conjugacy classes of $\mathfrak{S}_p$	#different $\mathcal{Z}_\Gamma$	#cyclic groups
1	1	1	1	1
2	2	2	2	2
3	6	4	5	5
4	30	11	22	17
5	156	19	93	67
6	1455	56	739	362
7	11,300	96	4508	2039
8	151,221	296	?	14,170
9	1,694,723	554	?	109,694
10	29,594,446	1593	?	976,412
18	$\approx 7.6 \cdot 10^{18}$	$7.3 \cdot 10^6$	?	$\approx 7.1 \cdot 10^{14}$

4.1.1. *First approach.* Each cyclic subgroup Γ can be uniquely represented by a permutation, which is minimal in the lexicographic order within permutations generating Γ. Let $\nu(\Gamma) \in \mathfrak{S}_p$ be such a permutation, that is,

$$\nu(\Gamma) = \min\{\sigma \in \mathfrak{S}_p; \langle\sigma\rangle = \Gamma\}.$$

Define

$$c_t := \langle \nu(c_{t-1}) \circ x_t \rangle, \tag{28}$$

where c_0 is a fixed cyclic subgroup and $(x_t)_{t\in\mathbb{N}}$ is a sequence of i.i.d. random transpositions distributed uniformly, that is, $\mathbb{P}(x_t = \alpha) = 1/\binom{p}{2}$ for any $\alpha \in \mathcal{T} := \{(i, j) \in \mathfrak{S}_p\}$. Clearly, the sequence $(c_t)_t$ is a Markov chain. Its state space is the set of all cyclic subgroups of $\mathfrak{S}_p$. Moreover, the trivial subgroup $\{\mathrm{id}\}$ can be reached from any subgroup c_t (and vice versa) in a finite number of steps with positive probability. Thus the chain $(c_t)_t$ is irreducible. The proposal distribution in the Metropolis–Hastings algorithm is the conditional distribution of $c_t | c_{t-1}$. It is proportional to the number of possible transitions from c to c', that is,

$$g(c'|c) := \frac{\#\{(i, j) \in \mathfrak{S}_p; c' = \langle \nu(c) \circ (i, j)\rangle\}}{\binom{p}{2}}, \tag{29}$$

where c and c' are cyclic subgroups.

We follow the principles of Bayesian model selection for graphical models, presented, for example, in Maathuis et al. ((2018), Chapter 10, page 247). Let Γ be uniformly distributed on the set $\mathcal{C} := \{\langle\sigma\rangle; \sigma \in \mathfrak{S}_p\}$ of cyclic subgroups of $\mathfrak{S}_p$. We assume that $K|\{\Gamma = c\}$, $c \in \mathcal{C}$, follows the Diaconis–Ylvisaker conjugate prior distribution on $\mathcal{P}_c$ with hyperparameters δ and D, that is,

$$f_{K|\Gamma=c}(k) = \frac{1}{I_c(\delta, D)} \mathrm{Det}(k)^{(\delta-2)/2} e^{-\frac{1}{2}\mathrm{Tr}[D\cdot k]} \mathbf{1}_{\mathcal{P}_c}(k),$$

where the normalizing constant is given in (27). Suppose that $Z_1, \dots, Z_n$ given $\{K = k, \Gamma = c\}$ are i.i.d. $\mathrm{N}_p(0, k^{-1})$ random vectors with $k \in \mathcal{P}_c$. Then it is easily seen that we have

$$\mathbb{P}(\Gamma = c | Z_1, \dots, Z_n) \propto \frac{I_c(\delta + n, D + U)}{I_c(\delta, D)} \quad (c \in \mathcal{C}) \tag{30}$$

with $U = \sum_{i=1}^n Z_i \cdot Z_i^\top$. These derivations allow us to run the Metropolis–Hastings algorithm restricted to cyclic groups, as follows.

ALGORITHM 14. Starting from a cyclic group $C_0 \in \mathcal{C}$, repeat the following two steps for $t = 1, 2, \dots$:

1. Sample x_t uniformly from the set $\mathcal{T}$ of all transpositions and set $c' = \langle \nu(C_{t-1}) \circ x_t \rangle$;
2. Accept the move $C_t = c'$ with probability

$$\min\left\{1, \frac{I_{c'}(\delta+n, D+U) I_{C_{t-1}}(\delta, D)}{I_{c'}(\delta, D) I_{C_{t-1}}(\delta+n, D+U)} \frac{g(C_{t-1}|c')}{g(c'|C_{t-1})}\right\}.$$

If the move is rejected, set $C_t = C_{t-1}$.

4.1.2. *Second approach.* It is known that $\langle\sigma\rangle = \langle\sigma'\rangle$ if and only if $\sigma' = \sigma^k$ for some $k \in \beta(|\sigma|)$, where

$$\beta(n) = \{k \in \{1, \dots, n\}; k \text{ and } n \text{ are relatively prime}\} \tag{31}$$

and $|\sigma|$ denotes the order of σ. Let $\mathcal{C} = \{\langle\sigma\rangle; \sigma \in \mathfrak{S}_p\}$ denote the set of cyclic subgroups of $\mathfrak{S}_p$. For $c \in \mathcal{C}$, we define $\Phi(c) := \#\beta(|c|)$ and $\mathcal{C}_c := \{\sigma \in \mathfrak{S}_p; \langle\sigma\rangle = c\}$, the set of permutations, which generate the cyclic subgroup c. We have

$$\Phi(c) = \#\mathcal{C}_c \quad (c \in \mathcal{C}).$$

For $c \in \mathcal{C}$, we denote

$$\pi_c = \mathbb{P}(\Gamma = c|Z_1, \dots, Z_n),$$

which we want to approximate. In our model, we have (see (30))

$$\pi_c \propto \frac{I_c(\delta+n, D+U)}{I_c(\delta, D)} \quad (c \in \mathcal{C}). \tag{32}$$

In order to find $\pi = (\pi_c; c \in \mathcal{C})$ let us consider $\tilde{\pi} = (\tilde{\pi}_\sigma; \sigma \in \mathfrak{S}_p)$, a probability distribution on $\mathfrak{S}_p$ such that

$$\tilde{\pi}_\sigma \propto \frac{I_{\langle\sigma\rangle}(\delta+n, D+U)}{I_{\langle\sigma\rangle}(\delta, D)} \quad (\sigma \in \mathfrak{S}_p). \tag{33}$$

Since (32) and (33) imply that $\tilde{\pi}_\sigma \propto \pi_{\langle\sigma\rangle}$, we have

$$\tilde{\pi}_\sigma = \frac{\pi_{\langle\sigma\rangle}}{\sum_{c\in\mathcal{C}} \Phi(c)\pi_c} \quad (\sigma \in \mathfrak{S}). \tag{34}$$

As before, let $(x_t)_{t\in\mathbb{N}}$ be a sequence of i.i.d. random transpositions distributed uniformly on $\mathcal{T} = \{(i, j) \in \mathfrak{S}_p\}$. We define a random walk on $\mathfrak{S}_p$ by

$$s_{t+1} = s_t \circ x_{t+1} \quad (t = 0, 1, \dots).$$

Then $(s_t)_t$ is an irreducible Markov chain with symmetric transition probability

$$g(\sigma'|\sigma) = \begin{cases} \dfrac{1}{\binom{p}{2}} & \text{if } \sigma^{-1} \circ \sigma' \in \mathcal{T}, \\ 0 & \text{if } \sigma^{-1} \circ \sigma' \notin \mathcal{T}. \end{cases}$$

We note that $(\langle s_t\rangle)_t$ is not a Markov chain on the space of cyclic subgroups. Indeed, it can be shown that the necessary conditions for $(f(s_t))_t$ to be a Markov chain (see Burke and Rosenblatt (1958), equation (3)) are not satisfied for $f(\sigma) := \langle\sigma\rangle$ if $p > 4$. A remedy for this fact was introduced in (28). Indeed, the sequence $(\langle s_t\rangle)_t$ is very similar to the sequence $(c_t)_t$ defined previously. Both move along cyclic subgroups and their definitions are very similar. However, $(\langle s_t\rangle)_t$ is not a Markov chain, whereas $(c_t)_t$ is a Markov chain. We took care of this problem by using the minimal generator $\nu(\cdot)$ as in definition (28) of c_t.

We use the Metropolis–Hastings algorithm with the above proposal distribution to approximate $\tilde{\pi}$.

ALGORITHM 15. Starting from a permutation $\sigma_0 \in \mathfrak{S}_p$, repeat the following two steps for $t = 1, 2, \dots$:

1. Sample x_t uniformly from the set $\mathcal{T}$ of all transpositions and set $\sigma' = \sigma_{t-1} \circ x_t$;
2. Accept the move $\sigma_t = \sigma'$ with probability

$$\min\left\{1, \frac{I_{\langle\sigma'\rangle}(\delta+n, D+U) I_{\langle\sigma_{t-1}\rangle}(\delta, D)}{I_{\langle\sigma'\rangle}(\delta, D) I_{\langle\sigma_{t-1}\rangle}(\delta+n, D+U)}\right\}.$$

If the move is rejected, set $\sigma_t = \sigma_{t-1}$.

By the ergodicity of the Markov chain $(\sigma_t)_t$ constructed above, as the number of steps $T \to \infty$, we have

$$\frac{\sum_{t=1}^T \mathbf{1}_{\sigma=\sigma_t}}{T} \xrightarrow{\text{a.s.}} \tilde{\pi}_\sigma \quad (\sigma \in \mathfrak{S}_p). \tag{35}$$

This fact allows us to develop a scheme for approximating the posterior probability π.

THEOREM 16. *We have as $T \to \infty$,*

$$\frac{\frac{1}{\Phi(c)} \sum_{t=1}^T \mathbf{1}_{c=\langle\sigma_t\rangle}}{\sum_{t=1}^T \frac{1}{\Phi(\langle\sigma_t\rangle)}} \xrightarrow{\text{a.s.}} \pi_c \quad (c \in \mathcal{C}). \tag{36}$$

PROOF. Let us denote $n_\sigma^{(T)} = \sum_{t=1}^T \mathbf{1}_{\sigma=\sigma_t}$, $\sigma \in \mathfrak{S}_p$. We have $T = \sum_{\sigma\in\mathfrak{S}_p} n_\sigma^{(T)}$ and $n_\sigma^{(T)}/T \xrightarrow{\text{a.s.}} \tilde{\pi}_\sigma$. Moreover,

$$\begin{aligned}
\frac{\frac{1}{\Phi(c)} \sum_{t=1}^T \mathbf{1}_{c=\langle\sigma_t\rangle}}{\sum_{t=1}^T \frac{1}{\Phi(\langle\sigma_t\rangle)}} &= \frac{\frac{1}{\Phi(c)} \sum_{\sigma\in\mathcal{C}_c} n_\sigma^{(T)}}{\sum_{t=1}^T \sum_{\gamma\in\mathcal{C}} \frac{1}{\Phi(\gamma)} \mathbf{1}_{\gamma=\langle\sigma_t\rangle}} \\
&= \frac{\frac{1}{\Phi(c)} \sum_{\sigma\in\mathcal{C}_c} \frac{n_\sigma^{(T)}}{T}}{\sum_{\gamma\in\mathcal{C}} \frac{1}{\Phi(\gamma)} \sum_{\sigma\in\mathcal{C}_\gamma} \frac{n_\gamma^{(T)}}{T}} \\
&\xrightarrow{\text{a.s.}} \frac{\frac{1}{\Phi(c)} \sum_{\sigma\in\mathcal{C}_c} \tilde{\pi}_\sigma}{\sum_{\gamma\in\mathcal{C}} \frac{1}{\Phi(\gamma)} \sum_{\sigma\in\mathcal{C}_\gamma} \tilde{\pi}_\gamma}.
\end{aligned}$$

Finally, by (34) we have

$$\frac{1}{\Phi(c)} \sum_{\sigma\in\mathcal{C}_c} \tilde{\pi}_\sigma = \frac{\pi_c}{\sum_{\gamma\in\mathcal{C}} \Phi(\gamma)\pi_\gamma} \propto \pi_c,$$

which completes the proof. $\square$

In order to approximate the posterior probability π, we allowed the Markov chain to travel on the larger space $\mathfrak{S}_p$. In particular, each state $c \in \mathcal{C}$ was multiplied $\Phi(c) \geq 1$ times, where $\Phi(c)$ is the number of permutations generating c. This procedure should result in slower convergence to the stationary distribution in (36). By comparing with (35), we see that (36) can be interpreted as follows: let us assign to each cyclic subgroup c a weight $1/\Phi(c) \leq 1$. Then the denominator $N_T := \sum_{t=1}^T 1/\Phi(\langle\sigma_t\rangle)$ can be thought of as an "effective" number of steps and the numerator is the number of "effective" steps spent in state c. In general, for large T we expect $N_T \ll T$ (see an example in Section 5.2).

4.2. *Model selection for* $p=4$. Our numbering of colored models on four vertices is in accordance with Gehrmann ((2011), Figures 15 and 16, pages 674–675). However, we identify models by the largest group with the same coloring Γ^* rather than the smallest as in Gehrmann (2011). There are 30 different subgroups of $\mathfrak{S}_4$, which generate 22 different colored spaces. Up to conjugacy (renumbering of vertices), there are 8 different conjugacy classes. Within a conjugacy class, constants $(k_i, r_i, d_i)_{i=1}^L$ remain the same. Groups Γ_k^* for $k=1,\dots,17$ correspond to cyclic colorings.

We apply our results and methods in order to do Bayesian model selection for the celebrated example of Frets' heads, Frets (1921), Whittaker (1990). The head dimensions (length L_i and breadth B_i, $i=1,2$) of 25 pairs of first and second sons were measured. Thus we have $n=25$ and $p=4$. The following sample covariance matrix is obtained (we have $Z=(L_1,B_1,L_2,B_2)^\top$),

$$U=\sum_{i=1}^{n} Z^{(i)}\cdot Z^{(i)\top}=\begin{pmatrix}2287.04 & 1268.84 & 1671.88 & 1106.68\\ 1268.84 & 1304.64 & 1231.48 & 841.28\\ 1671.88 & 1231.48 & 2419.36 & 1356.96\\ 1106.68 & 841.28 & 1356.96 & 1080.56\end{pmatrix}.$$

We perform Bayesian model selection within all RCOP models, not just the ones corresponding to cyclic subgroups. In Table 2, we list all RCOP models on full graph with four vertices, along with corresponding structure constants. Structure constants remain the same within a conjugacy class, however, the invariant measure φ_Γ is always different. Since there are only

TABLE 2
Structure constants for all colorings with four vertices

Group	(k_i)	(r_i)	(d_i)
$\Gamma_1^*=\{\mathrm{id}\}$	(1)	(4)	(1)
$\Gamma_2^*=\langle(1,2)\rangle$	(1,1)	(3,1)	(1,1)
$\Gamma_3^*=\langle(1,3)\rangle$			
$\Gamma_4^*=\langle(1,4)\rangle$			
$\Gamma_5^*=\langle(2,3)\rangle$			
$\Gamma_6^*=\langle(2,4)\rangle$			
$\Gamma_7^*=\langle(3,4)\rangle$			
$\Gamma_8^*=\langle(1,2,3),(1,2)\rangle$	(1,2)	(2,1)	(1,1)
$\Gamma_9^*=\langle(1,2,4),(1,2)\rangle$			
$\Gamma_{10}^*=\langle(1,3,4),(1,3)\rangle$			
$\Gamma_{11}^*=\langle(2,3,4),(2,3)\rangle$			
$\Gamma_{12}^*=\langle(1,2)(3,4)\rangle$	(1,1)	(2,2)	(1,1)
$\Gamma_{13}^*=\langle(1,3)(2,4)\rangle$			
$\Gamma_{14}^*=\langle(1,4)(2,3)\rangle$			
$\Gamma_{15}^*=\langle(1,2,3,4),(1,3)\rangle$	(1,1,2)	(1,1,1)	(1,1,1)
$\Gamma_{16}^*=\langle(1,2,4,3),(1,4)\rangle$			
$\Gamma_{17}^*=\langle(1,3,2,4),(1,2)\rangle$			
$\Gamma_{18}^*=\langle(1,2),(3,4)\rangle$	(1,1,1)	(2,1,1)	(1,1,1)
$\Gamma_{19}^*=\langle(1,3),(2,4)\rangle$			
$\Gamma_{20}^*=\langle(1,4),(2,3)\rangle$			
$\Gamma_{21}^*=\langle(1,2)(3,4),(1,4)(2,3)\rangle$	(1,1,1,1)	(1,1,1,1)	(1,1,1,1)
$\Gamma_{22}^*=\mathfrak{S}_4$	(1,3)	(1,1)	(1,1)

TABLE 3
Posterior probabilities in Frets' heads for three best models, $\delta = 3$ and given D

D	Best model		2nd best		3rd best	
I_4	Γ_{22}^*	(95.2%)	Γ_{16}^*	(2.5%)	Γ_{17}^*	(1.3%)
$50I_4$	Γ_{19}^*	(33.8%)	Γ_{13}^*	(29.6%)	Γ_8^*	(13.3%)
$100I_4$	Γ_{13}^*	(39.6%)	Γ_{19}^*	(29.8%)	Γ_8^*	(7.2%)
$1000I_4$	Γ_1^*	(38.9%)	Γ_{13}^*	(10.5%)	Γ_3^*	(10.3%)

22 such models, we calculate all exact posterior probabilities. Table 2 and the invariant measures φ_Γ were obtained by using Lemma 4.

In Table 3, we summarize the results when $\delta = 3$ (a parameter of the prior distribution, Section 3.4), giving the three best coloring models with the highest posterior probability, for each given D. Results are very similar for $\delta = 10$ and the given values of D. For comparison, the three best models according to BIC are Γ_{19}^*, Γ_{13}^* and Γ_8^* with the BIC 834.5, 835.4 and 835.5, respectively.

For different values of $D = dI_4$, the only models that have highest posterior probability are the 4 models: $\Gamma_{22}^* = \mathfrak{S}_4$, $\Gamma_{19}^* = \langle(1,3),(2,4)\rangle$, $\Gamma_{13}^* = \langle(1,3)(2,4)\rangle$, $\Gamma_1^* = \{\mathrm{id}\}$. These four subgroups form a path in the Hasse diagram of subgroups of $\mathfrak{S}_4^*$, that is, $\Gamma_{22}^* \supset \Gamma_{19}^* \supset \Gamma_{13}^* \supset \Gamma_1^*$. Thus the four selected colorings, corresponding to the permutation groups are in some way consistent. Moreover, each of them has a good statistical interpretation. Let us interpret models Γ_{13}^* and Γ_{19}^*. Recall the enumeration of vertices $(1,2,3,4) = (L_1, B_1, L_2, B_2)$. The invariance with respect to the transposition $(1,3)$ means that L_1 is exchangeable with L_2 and, similarly, the invariance with respect to the transposition $(2,4)$ implies exchangeability of B_1 and B_2. Both together correspond to the fact that sons should be exchangeable in some way.

We observe that only the Γ_{22}^* model appeared in former attempts of model selection for Frets' heads data. It was considered in Massam, Li and Gao ((2018), Figure S7, page 28 of the Supplementary Material) with eleven other models. Note that the only complete RCOP model selected in Gehrmann (2011) (who used the Edwards–Havránek model selection procedure) among the 9 minimally accepted models on page 676 of her article is Γ_{10}^*, which is not selected by our exact Bayesian procedure for any choice of $D = d\,I_4$.

5. Simulations. Let the covariance matrix $\Sigma = (c_{ij})_{ij} \in \mathrm{Sym}^+(p;\mathbb{R})$ be the symmetric circulant matrix defined by

$$c_{ij} = \begin{cases} 1 - |i-j|/p, & i \neq j, \\ 1 + 1/p, & i = j. \end{cases}$$

It is easily seen that this matrix belongs to $\mathcal{P}_{\langle\sigma^*\rangle}$ with $\sigma^* = (1, 2, \dots, p-1, p)$.

5.1. *First approach.* For $p = 10$ and $n = 20$, we sampled $Z^{(1)}, \dots, Z^{(n)}$ from the $\mathrm{N}_p(0, \Sigma)$ distribution and obtained $U = \sum_{i=1}^n Z^{(i)} \cdot Z^{(i)\top}$ depicted in Figure 1(b).

We run the Metropolis–Hastings algorithm starting from the group $\langle\sigma_0\rangle = \{\mathrm{id}\}$ with hyperparameters $\delta = 3$ and $D = I_{10}$. After 1,000,000 steps, the five most visited states are given in the Table 4.

The Metropolis–Hastings (M-H) algorithm recovered the true pattern of the covariance matrix. The acceptance rate was 2.5% and the Markov chain visited 746 different cyclic groups. The acceptance rate can be increased by a suitable choice of the hyperparameters (e.g., for $D = 10I_{10}$ the acceptance rate is around 10%).

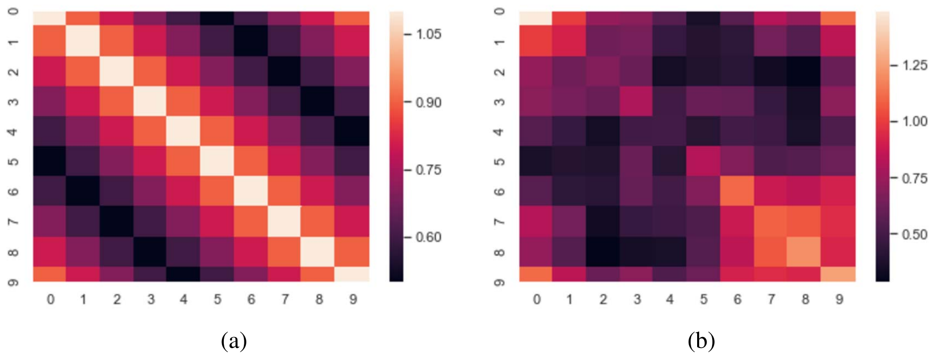


FIG. 1. *Heat map of matrix* Σ (*a*) *and matrix* U/n (*b*).

TABLE 4
Five most visited cyclic subgroups

Generator of a cyclic group	Number of visits
(1, 2, 3, 4, 5, 6, 7, 8, 9,10)	457,725
(1, 6, 2, 7)(3, 5, 9)(4, 8, 10)	110,677
(1, 6)(2, 7)(3, 5, 9)(4, 8, 10)	51,618
(1, 7)(2, 6)(3, 5, 9)(4, 8, 10)	40,895
(1, 2, 6, 7)(3, 5, 9)(4, 8, 10)	34,883

In order to grasp how randomness may influence results, we performed 100 simulations, where each time we sample $Z^{(1)}, \dots, Z^{(n)}$ from $\mathrm{N}_p(0, \Sigma)$ and we run M-H for 100,000 steps with the same parameters as before. In Table 5, we present how many times a given cyclic subgroup was most visited during these 100 simulations (second column). There were 53 distinct cyclic subgroups, which were most visited at least in one of the 100 simulations; below we present 10 such subgroups. The average acceptance rate is 1.4% (see the histogram in Figure 2). When we regard colorings as partitions of the set $V \cup E$ according to group orbit decomposition, the two colorings may be compared using the so-called adjusted Rand index (ARI, see Hubert and Arabie (1985)), a similarity measure comparing partitions, which takes values between -1 and 1, where 1 stands for perfect match and independent random labelings have score close to 0. In the third column of Table 5, we give the adjusted Rand index between the colorings generated by given cyclic subgroup and the true coloring.

We see that groups which were most visited by the Markov chain have positive ARI and the true pattern was recovered in a quarter of cases. We stress that even though the colorings

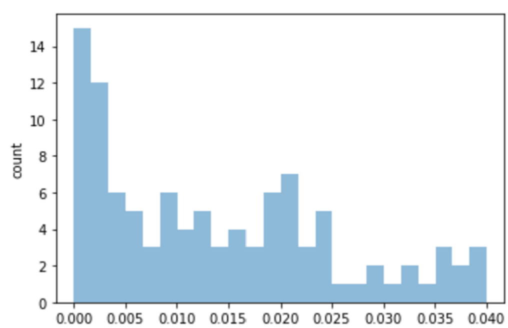


FIG. 2. *Histogram of acceptance rates in* 100 *simulations of Metropolis–Hastings algorithm.*

TABLE 5
Cyclic subgroups which were chosen by M-H algorithm most often

Generator of a cyclic group	#most visited	ARI
(1, 2, 3, 4, 5, 6, 7, 8, 9,10)	25	1.00
(1, 3, 5, 7, 9)(2, 4, 6, 8,10)	13	0.60
(1, 2, 4, 3, 5, 6, 7, 9, 8, 10)	3	0.43
(1, 2, 4, 3, 5, 6, 7, 8, 9, 10)	2	0.46
(1, 3, 2, 4, 5, 6, 8, 7, 9, 10)	2	0.43
(1, 3, 5, 9, 2, 6, 8, 10, 4, 7)	2	0.43
(1, 4, 3, 5, 2, 6, 9, 8, 10, 7)	2	0.35
(1, 4, 5, 7, 8)(2, 3, 6, 9, 10)	2	0.24
(1, 8, 10, 9)(2, 7)(3, 5, 4, 6)	2	0.19
(1, 2, 10, 3)(4, 9)(5, 8, 6, 7)	2	0.19

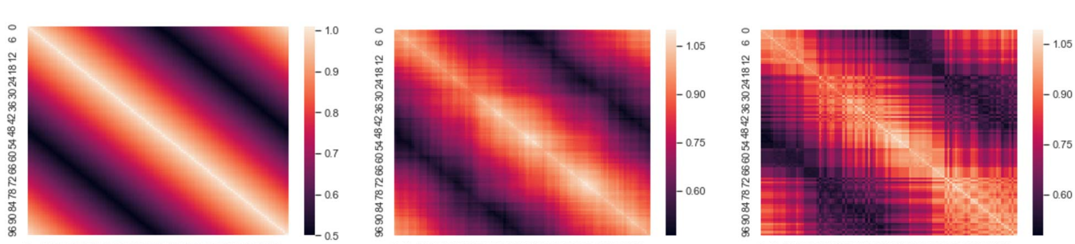
generated by $\langle(1,2,3,4,5,6,7,8,9,10)\rangle$ and $\langle(1,3,5,7,9)(2,4,6,8,10)\rangle$ are very similar, the distance between these subgroups is 9, that is, the Markov chain $(C_t)_t$ needs at least 9 steps to get from one subgroup to the other. We performed similar simulations for $n = p = 10$ and the results were only slightly worse: the true pattern was recovered in 18 out of 100 runs of the algorithm.

This indicates that the Markov chain may encounter many local maxima and one should always tune the hyper parameters in order to have higher acceptance rate or to allow the Markov chain $(C_t)_t$ to make bigger steps.

5.2. *Second approach.* We performed $T = 100{,}000$ steps of Algorithm 15 with $\sigma_0 = \mathrm{id}$, $p = 100$, $n = 200$, $\delta = 3$ and $D = I_{100}$. Let us note that for $p = 100$, there are about $4 \cdot 10^{155}$ cyclic subgroups and this is the number of models we consider in our model search.

We have used Theorem 16 to approximate the posterior probability distribution $(\pi_c; c \in \mathcal{C})$ (see (30)). The highest estimated posterior probability was obtained for $c^* := \langle\sigma^*\rangle$, where

$$\begin{aligned}\sigma^* = {} & (1,2,3,4)(6,8,15)(7,10,9)(11,16,12)(13,17,14)(18,19,20,22,21)(23,26)\\ & (24,42,28,44)(25,31,30,32)(27,34)(29,37)(33,45)(35,39,36,40)\\ & (38,47,41,48)(43,51,46,49)(50,52,53,54)(56,58,57)(59,66,67)\\ & (60,65,63)(61,62,64)(68,71,72,70,69)(73,93)(74,77)(75,98,81,100)\\ & (76,84,78,83)(79,85)(80,94,82,91)(86,92,87,90)(88,96,89,97)(95,99).\end{aligned}$$



(a) (b) (c)

FIG. 3. *Heat map of matrix Σ (a) and matrix U/n (b) and projection of U/n onto $\mathcal{Z}_{c^*}$.*

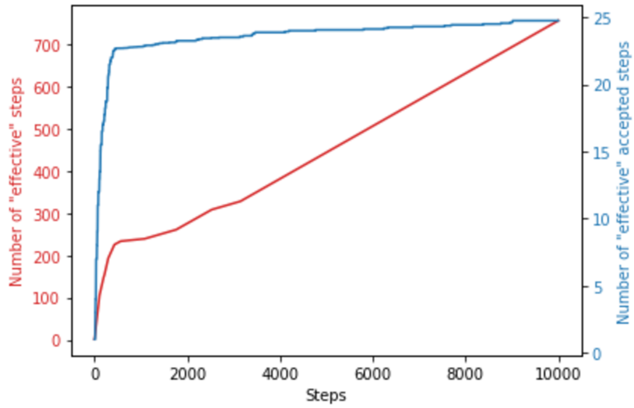


FIG. 4. *Number of "effective" steps (red) and number of "effective" accepted steps (blue).*

The order of c^* is $|c^*| = 60$ and $\Phi(c^*) = 16$. The estimate of the posterior probability π_{c^*} is equal to (recall (36))

$$\frac{\frac{1}{\Phi(c^*)}\sum_{t=1}^{T} \mathbf{1}_{c^*=\langle\sigma_t\rangle}}{\sum_{t=1}^{T}\frac{1}{\Phi(\langle\sigma_t\rangle)}} \approx \frac{2361.5}{6381.5} \approx 37\%.$$

The true covariance matrix Σ, the data matrix U/n and the projection $\Pi_{c^*}(U/n)$ are illustrated in Figure 3.

We visualize the performance of the algorithm on Figure 4. In red color, a sequence $(\sum_{t=1}^{k}\frac{1}{\Phi(\langle\sigma_t\rangle)})_k$ is depicted, which can be thought of as an "effective" number of steps of the algorithm (for an explanation, see the paragraph at the end of Section 4.1.2). In blue, we present a sequence $(\sum_{t=1}^{k}\frac{1}{\Phi(\langle\sigma_t\rangle)}\mathbf{1}_{\langle\sigma_t\rangle\neq\langle\sigma_{t-1}\rangle})_k$, which represents the number of weighted accepted steps, where the weight of the kth step equals $\frac{1}{\Phi(\langle\sigma_k\rangle)}$. We restricted the plot to steps $k = 1, \dots, 10{,}000$, because after 10,000 steps, the Markov chain $(\sigma_t)_{10{,}000\leq t\leq 100{,}000}$ changed its state only 9 times. For $k = 100{,}000$, the value of the blue curve is 25.75, while the value of red one is 6381.5.

The model suffers from poor acceptance rate, which could be improved by an appropriate choice of the hyperparameter D or by allowing the Markov chain to do bigger steps.

Acknowledgments. The authors would like to thank Steffen Lauritzen for his interest and encouragements. We also thank M. Bogdan from Wrocław University, A. Descatha from INSERM and Centre Hospitalier Universitaire Angers and V. Seegers from Institut de Cancerologie de l'Ouest Nantes for explaining the specific nature of medical and genetic data. The paper benefited from the comments of an anonymous referee to whom the authors are grateful.

Funding. The second author was supported by JSPS KAKENHI Grant Number 16K05174, 20K03657 and JST PRESTO.

The third author was supported by Grant 2016/21/B/ST1/00005 of the National Science Center, Poland.

The fourth author was supported by an NSERC Discovery Grant.

SUPPLEMENTARY MATERIAL

Supplement to "Model selection in the space of Gaussian models invariant by symmetry" (DOI: 10.1214/22-AOS2174SUPP; .pdf). Supplement contains proofs and examples.

We provide proofs of Theorems 1, 5, 6 along with a background on representation theory that is needed to understand proofs. Moreover, we present proofs of Proposition 7, Theorems 8 and 9, an example to Section 2.3, proof of Lemma 13 and the real data example considered in Miller et al. (2005) and Højsgaard and Lauritzen (2008).

REFERENCES

ANDERSSON, S. (1975). Invariant normal models. *Ann. Statist.* **3** 132–154. MR0362703

ANDERSSON, S. A., BRØNS, H. K. and JENSEN, S. T. (1983). Distribution of eigenvalues in multivariate statistical analysis. *Ann. Statist.* **11** 392–415. MR0696055 https://doi.org/10.1214/aos/1176346149

ANDERSSON, S. and MADSEN, J. (1998). Symmetry and lattice conditional independence in a multivariate normal distribution. *Ann. Statist.* **26** 525–572. MR1626059 https://doi.org/10.1214/aos/1028144848

BARNDORFF-NIELSEN, O. (2014). *Information and Exponential Families in Statistical Theory*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR3221776 https://doi.org/10.1002/9781118857281

BURKE, C. J. and ROSENBLATT, M. (1958). A Markovian function of a Markov chain. *Ann. Math. Stat.* **29** 1112–1122. MR0101557 https://doi.org/10.1214/aoms/1177706444

DAVIES, I. and MARIGLIANO, O. (2021). Coloured graphical models and their symmetries. *Matematiche* (*Catania*) **76** 501–515. MR4334901 https://doi.org/10.4418/2021.76.2.13

DE MAIO, A., ORLANDO, D., SOLOVEYCHIK, I. and WIESEL, A. (2016). Invariance theory for adaptive detection in interference with group symmetric covariance matrix. *IEEE Trans. Signal Process.* **64** 6299–6312. MR3562723 https://doi.org/10.1109/TSP.2016.2591502

DESCATHA, A., ROQUELAURE, Y., EVANOFF, B., NIEDHAMMER, I., CHASTANG, J. F., MARIOT, C., HA, C., IMBERNON, E., GOLDBERG, M. et al. (2007). Selected questions on biomechanical exposures for surveillance of upper-limb work-related musculoskeletal disorders. *Int. Arch. Occup. Environ. Health* **81** 1–8. https://doi.org/10.1007/s00420-007-0180-5

DIACONIS, P. and YLVISAKER, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7** 269–281. MR0520238

FARAUT, J. and KORÁNYI, A. (1994). *Analysis on Symmetric Cones*. *Oxford Mathematical Monographs*. The Clarendon Press, Oxford Univ. Press, New York. MR1446489

FRETS, G. P. (1921). Heredity of head form in man. *Genetica* **41** 193–400.

FROMMLET, F., BOGDAN, M. and RAMSEY, D. (2016). *Phenotypes and Genotypes*: *The Search for Influential Genes*. *Computational Biology* **18**. Springer, London. MR3443801 https://doi.org/10.1007/978-1-4471-5310-8

GAO, X. and MASSAM, H. (2015). Estimation of symmetry-constrained Gaussian graphical models: Application to clustered dense networks. *J. Comput. Graph. Statist.* **24** 909–929. MR3432922 https://doi.org/10.1080/10618600.2014.937811

GEHRMANN, H. (2011). Lattices of graphical Gaussian models with symmetries. *Symmetry* **3** 653–679. MR2845303 https://doi.org/10.3390/sym3030653

GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. *Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge Univ. Press, Cambridge. MR3587782 https://doi.org/10.1017/9781139029834

GOUTIS, C. and ROBERT, C. P. (1998). Model choice in generalised linear models: A Bayesian approach via Kullback–Leibler projections. *Biometrika* **85** 29–37. MR1627250 https://doi.org/10.1093/biomet/85.1.29

GRACZYK, P., ISHI, H., KOŁODZIEJEK, B. and MASSAM, H. (2022). Supplement to "Model selection in the space of Gaussian models invariant by symmetry." https://doi.org/10.1214/22-AOS2174SUPP

GRAHAM, R. L., GRÖTSCHEL, M. and LOVÁSZ, L., eds. (1995). *Handbook of Combinatorics*. *Vols*. 1, 2. Elsevier, Amsterdam; MIT Press, Cambridge, MA.

HASSAIRI, A. and LAJMI, S. (2001). Riesz exponential families on symmetric cones. *J. Theoret. Probab.* **14** 927–948. MR1860082 https://doi.org/10.1023/A:1012592618872

HØJSGAARD, S. and LAURITZEN, S. L. (2008). Graphical Gaussian models with edge and vertex symmetries. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 1005–1027. MR2530327 https://doi.org/10.1111/j.1467-9868.2008.00666.x

HOLT, D. F. (2010). Enumerating subgroups of the symmetric group. In *Computational Group Theory and the Theory of Groups*, *II*. *Contemp. Math.* **511** 33–37. Amer. Math. Soc., Providence, RI. MR2655292 https://doi.org/10.1090/conm/511/10041

HUBERT, L. and ARABIE, P. (1985). Comparing partitions. *J. Classification* **2** 193–218.

JENSEN, S. T. (1988). Covariance hypotheses which are linear in both the covariance and the inverse covariance. *Ann. Statist.* **16** 302–322. MR0924873 https://doi.org/10.1214/aos/1176350707

LI, S. Z. (2009). *Markov Random Field Modeling in Image Analysis*, 3rd ed. *Advances in Pattern Recognition*. Springer London, Ltd., London. MR2493908

LI, Q., GAO, X. and MASSAM, H. (2020). Bayesian model selection approach for coloured graphical Gaussian models. *J. Stat. Comput. Simul.* **90** 2631–2654. MR4145359 https://doi.org/10.1080/00949655.2020.1784175

MAATHUIS, M., DRTON, M., LAURITZEN, S. and WAINWRIGHT, M., eds. (2018). *Handbook of Graphical Models*. *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, FL. MR3889064

MADSEN, J. (2000). Invariant normal models with recursive graphical Markov structure. *Ann. Statist.* **28** 1150–1178. MR1810923 https://doi.org/10.1214/aos/1015956711

MASSAM, H., LI, Q. and GAO, X. (2018). Bayesian precision and covariance matrix estimation for graphical Gaussian models with edge and vertex symmetries. *Biometrika* **105** 371–388. MR3804408 https://doi.org/10.1093/biomet/asx084

MICHAŁEK, M., STURMFELS, B., UHLER, C. and ZWIERNIK, P. (2016). Exponential varieties. *Proc. Lond. Math. Soc.* (3) **112** 27–56. MR3458144 https://doi.org/10.1112/plms/pdv066

MILLER, L. D., SMEDS, J., GEORGE, J., VEGA, V. B., VERGARA, L., PLONER, A., PAWITAN, Y., HALL, P., KLAAR, S. et al. (2005). An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc. Natl. Acad. Sci. USA* **102** 13550–13555.

MISSIO, G., MORENO, D. H., DEMETRIO, F. N., SOEIRO-DE-SOUZA, M. G., FERNANDES, F. D. S., BARROS, V. B. and MORENO, R. A. (2019). A randomized controlled trial comparing lithium plus valproic acid versus lithium plus carbamazepine in young patients with type 1 bipolar disorder: The LICAVAL study. *Trials* **20** 608. https://doi.org/10.1186/s13063-019-3655-2

OLKIN, I. and PRESS, S. J. (1969). Testing and estimation for a circular stationary model. *Ann. Math. Stat.* **40** 1358–1373. MR0245139 https://doi.org/10.1214/aoms/1177697508

PLESKEN, W. and SOUVIGNIER, B. (1996). Constructing rational representations of finite groups. *Exp. Math.* **5** 39–47. MR1412953

RANCIATI, S., ROVERATO, A. and LUATI, A. (2021). Fused graphical lasso for brain networks with symmetries. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **70** 1299–1322. MR4347714 https://doi.org/10.1111/rssc.12514

ROVERATO, A. (2017). *Graphical Models for Categorical Data*. *SemStat Elements*. Cambridge Univ. Press, Cambridge. MR3751385 https://doi.org/10.1017/9781108277495

SERRE, J.-P. (1977). *Linear Representations of Finite Groups*. *Graduate Texts in Mathematics* **42**. Springer, New York. MR0450380

SHAH, P. and CHANDRASEKARAN, V. (2012). Group symmetry and covariance regularization. *Electron. J. Stat.* **6** 1600–1640. MR2988459 https://doi.org/10.1214/12-EJS723

SIEMONS, J. (1982). On partitions and permutation groups on unordered sets. *Arch. Math.* (*Basel*) **38** 391–403. MR0666910 https://doi.org/10.1007/BF01304806

SIEMONS, J. (1983). Automorphism groups of graphs. *Arch. Math.* (*Basel*) **41** 379–384. MR0731610 https://doi.org/10.1007/BF01371410

SOBCZYK, P., WILCZYNSKI, S., BOGDAN, M., GRACZYK, P., JOSSE, J., PANLOUP, F., SEEGERS, V. and STANIAK, M. (2020). VARCLUST: Clustering variables using dimensionality reduction. Preprint. Available at arXiv:2011.06501.

SOLOVEYCHIK, I., TRUSHIN, D. and WIESEL, A. (2016). Group symmetric robust covariance estimation. *IEEE Trans. Signal Process.* **64** 244–257. MR3432975 https://doi.org/10.1109/TSP.2015.2486739

TOYODA, K., YOSHIMURA, S., NAKAI, M., KOGA, M., SASAHARA, Y., SONODA, K., KAMIYAMA, K., YAZAWA, Y., KAWADA, S. et al. (2022). Twenty-year change in severity and outcome of ischemic and hemorrhagic strokes. *JAMA Neurol.* **79** 61–69. https://doi.org/10.1001/jamaneurol.2021.4346

UPMEIER, H. (1986). Jordan algebras and harmonic analysis on symmetric spaces. *Amer. J. Math.* **108** 1–25. MR0821311 https://doi.org/10.2307/2374466

WHITTAKER, J. (1990). *Graphical Models in Applied Multivariate Statistics*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, Chichester. MR1112133

WIELANDT, H. (1969). *Permutation Groups Through Invariant Relations and Invariant Functions*. *Lect. Notes Dept. Math.* Ohio St. Univ. Columbus.