The shape of partial correlation matrices

Richard Artner , Paul P. Wellingerhof , Ginette Lafit , Tim Loossens , Wolf Vanpaemel & Francis Tuerlinckx

Published online: 03 Sep 2020.

Submit your article to this journal ⤤

Article views: 66

View related articles ⤤

View Crossmark data ⤤

Taylor & Francis
Taylor & Francis Group

Check for updates

The shape of partial correlation matrices

Richard Artner[a]* (iD), Paul P. Wellingerhof[b]* (iD), Ginette Lafit[a] (iD), Tim Loossens[a] (iD), Wolf Vanpaemel[a‡] (iD), and Francis Tuerlinckx[a‡] (iD)

[a]KU Leuven - Faculty of Psychology and Educational Sciences, Leuven, Belgium; [b]Eberhard Karls Universität Tübingen - Department of Psychology, Tubingen, Germany

ABSTRACT

The correlational structure of a set of variables is often conveniently described by the pairwise partial correlations as they contain the same information as the Pearson correlations with the advantage of straightforward identifications of conditional linear independence. For mathematical convenience, multiple matrix representations of the pairwise partial correlations exist in the literature but their properties have not been investigated thoroughly. In this paper, we derive necessary and sufficient conditions for the eigenvalues of differently defined partial correlation matrices so that the correlation structure is a valid one. Equipped with these conditions, we will then emphasize the intricacies of algorithmic generations of correlation structures via partial correlation matrices. Furthermore, we examine the space of valid partial correlation structures and juxtapose it with the space of valid Pearson correlation structures. As these spaces turn out to be equal in volume for every dimension and equivalent with respect to rotation, a simple formula allows the creation of valid partial correlation matrices by the use of current algorithms for the generation and approximation of correlation matrices. Lastly, we derive simple conditions on the partial correlations for frequently assumed sparse structures.

1. Introduction

Partial correlation coefficients are commonly used across a wide variety of research fields, for instance, psychology and biology. Partial correlations are useful when dealing with conditional (linear) dependence relations, as in Gaussian graphical models (e.g., Anandkumar et al. 2012; Ha and Sun 2014; Epskamp et al. 2018). The fact that such partial correlation coefficients are much easier to interpret than the corresponding bivariate correlations makes them very appealing. The ease of interpretation is derived mainly from the property that a partial correlation of zero corresponds to conditional independence for jointly normally distributed variables (as shown by Baba, Shibata, and Sibuya 2004, Theorem 3, 663). More generally, a partial correlation of zero corresponds to conditional linear independence for every random vector whose covariance matrix exists (i.e., all marginal distributions have finite second moments). This convenient

---

CONTACT Richard Artner ✉ richard.artner@kuleuven.be KU Leuven - Faculty of Psychology and Educational Sciences, Leuven, 3000 Belgium
*Richard Artner and Paul P. Wellingerhof share first authorship.
‡Wolf Vanpaemel and Francis Tuerlinckx share last authorship.

property of partial correlations is shared with (standardized) multiple regression coefficients. However, unlike partial correlations which are bounded by the interval [−1, 1], (standardized) regression coefficients are unbounded, and that makes their magnitude much harder to interpret.

For a collection of p variables, the $\frac{p(p-1)}{2}$ bivariate correlations are often conveniently represented in a correlation matrix. This matrix arises by scaling the covariance matrix of the p variables, and its properties are well-known (see e.g., Rousseeuw and Molenberghs 1994). The set of the $\frac{p(p-1)}{2}$ partial correlations contains the same amount of information as the correlation matrix asking for a similar compact representation for them. Stemming from the fact that the partial correlation of a variable with itself is undefined, a partial correlation matrix has been defined in multiple different ways. However, the properties of such matrices have not yet been investigated, making the uncritical use of algorithms that generate random matrices potentially dangerous.

By deriving necessary and sufficient conditions for different definitions of the partial correlation matrix, we will show that the matrix representation allows for straightforward validity checks, and we discuss its implications with respect to the efficient and correct usage of random matrix algorithms. The derived matrix properties will further be used to derive restrictions on the partial correlations for frequently used graph structures. Lastly, the space of valid partial correlation structures will be examined and juxtaposed with valid correlation structures.

2. Partial correlation matrices

Let $X := (X_1, , ..., X_p)$ be any p-dimensional random vector with non singular covariance matrix $\Sigma = (\sigma_{ij})_{i,j=1,...,p}$. Without any loss of generality we assume that X is mean-centered to shorten subsequent notation. For two random variables $X_i, X_j \in X$ the mean square regression plane with respect to $X \backslash \{X_i, X_j\}$ is respectively uniquely defined as the following minimizer

$$\underset{\beta_k^i \in \mathbb{R} \forall k \in \{1,...,p\} \backslash \{i,j\}}{\arg\min} \mathbb{E}(\tilde{\epsilon}_i)^2 \text{ with } \tilde{\epsilon}_i := X_i - \sum_{k \neq i,j} \beta_k^i X_k \qquad (2.1)$$

$$\underset{\beta_k^j \in \mathbb{R} \forall k \in \{1,...,p\} \backslash \{i,j\}}{\arg\min} \mathbb{E}(\tilde{\epsilon}_j)^2 \text{ with } \tilde{\epsilon}_j := X_j - \sum_{k \neq i,j} \beta_k^j X_k \qquad (2.2)$$

The partial correlation $\rho_{X_i X_i \cdot X \backslash \{X_i, X_j\}}$ (which we will abbreviate as $\tilde{\rho}_{ij}$) of X_i and X_j with respect to $X \backslash \{X_i, X_j\}$ is then defined as the Pearson correlation of the random variables $\tilde{\epsilon}_i$ and $\tilde{\epsilon}_j$ (see Cramer 1946, Definition 23.4.1, 306) and equal to

$$\tilde{\rho}_{ij} := \rho(\tilde{\epsilon}_i, \tilde{\epsilon}_j) = \frac{\mathbb{E}(\tilde{\epsilon}_i \tilde{\epsilon}_j)}{\sqrt{\mathbb{E}(\tilde{\epsilon}_i^2)\mathbb{E}(\tilde{\epsilon}_j^2)}} \qquad (2.3)$$

since $\mathbb{E}(\tilde{\epsilon}_i) = \mathbb{E}(\tilde{\epsilon}_j) = 0$ as a result of Equations (2.1) and (2.2).

The following useful relation sometimes leads to confusion about the sign of the (estimated) partial correlation coefficient (Antti and Puntanen 1983):

$$-\tilde{\rho}_{ij} = \rho(\epsilon_i, \epsilon_j) = \frac{\mathbb{E}(\epsilon_i \epsilon_j)}{\sqrt{\mathbb{E}(\epsilon_i{}^2)\mathbb{E}(\epsilon_j{}^2)}} \, \forall i, j \in \{1, ..., p\}, i \neq j, \qquad (2.4)$$

with ϵ_i representing the residual of X_i when regressed on $X \backslash \{X_i\}$:

$$\underset{\beta_k^i \in \mathbb{R} \forall k \in \{1, ..., p\} \backslash \{i\}}{\arg \min} \quad \mathbb{E}(\epsilon_i)^2 \text{ with } \epsilon_i := X_i - \sum_{k \neq i} \beta_k^i X_k \qquad (2.5)$$

Definition 1. For a p-dimensional random vector $\boldsymbol{X}$ with non singular covariance matrix $\boldsymbol{\Sigma}$ let the partial correlation matrix $\tilde{\boldsymbol{R}}$ be defined as the negative scaled concentration matrix ($\boldsymbol{S}$):

$$\tilde{\boldsymbol{R}} := -\boldsymbol{D}_{\boldsymbol{\Sigma}^{-1}}^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{D}_{\boldsymbol{\Sigma}^{-1}}^{-1} =: -\boldsymbol{S}, \qquad (2.6)$$

with concentration matrix $\boldsymbol{\Sigma}^{-1} = (\sigma^{ij})_{i,j=1,...,p}$ and $\boldsymbol{D}_{\boldsymbol{\Sigma}^{-1}} := \left(\sqrt{\sigma^{ij} \delta_{ij}} \right)_{i,j=1,...,p}$ where δ_{ij} represents the Kronecker delta.

The justification of this definition stems from the fact that $\tilde{\boldsymbol{R}} = (\tilde{\rho}_{ij})_{i,j=1,...,p}$ then has the partial correlation coefficients $\tilde{\rho}_{ij}$ as its off-diagonal entries thereby mirroring the correlation matrix $\boldsymbol{R} = (\rho(X_i, X_j))_{i,j=1,...,p}$ (see Lewis and Styan 1981)[1] This relation between the partial correlations and the scaled concentration matrix is often stated in the case that $\boldsymbol{X}$ follows a multivariate normal distribution (e.g., Lauritzen 1996, 130), but it holds true for any distribution with regular covariance matrix that $\tilde{\boldsymbol{R}}_{ij} = \tilde{\rho}_{ij}$. Potentially this originates in the confusion of the partial correlation coefficient and the conditional correlation coefficient which happen to coincide only in case of multivariate normality (Baba, Shibata, and Sibuya 2004). The fact that $\rho(\epsilon_i, \epsilon_i) = 1$ further leads to $-\tilde{\boldsymbol{R}} = \boldsymbol{S} = (\rho(\epsilon_i, \epsilon_j))_{i,j=1,...,p}$ in light of Equation (2.4).

Any non degenerate covariance matrix $\boldsymbol{\Sigma}$ gives rise to one unique correlation matrix $\boldsymbol{R}$ (via scaling to unit variance) and one unique partial correlation matrix $\tilde{\boldsymbol{R}}$ (via Equation 2.6). The following two lemmata establishing the unique relation of $\boldsymbol{R}$ and $\tilde{\boldsymbol{R}}$ allowing us to define a matrix to be a valid partial correlation matrix if it gives rise to a valid correlation matrix.

Lemma 1. The partial correlation matrix $\tilde{\boldsymbol{R}}$ can also be computed from the correlation matrix $\boldsymbol{R}$ as

$$\tilde{\boldsymbol{R}} = f(\boldsymbol{R}) = -\boldsymbol{D}_{\boldsymbol{R}^{-1}}^{-1} \boldsymbol{R}^{-1} \boldsymbol{D}_{\boldsymbol{R}^{-1}}^{-1}. \qquad (2.7)$$

Proof. See Appendix. □

Lemma 2. (a) The correlation matrix $\boldsymbol{R}$ can be calculated from the partial correlation matrix $\tilde{\boldsymbol{R}}$ as

$$\boldsymbol{R} = g(\tilde{\boldsymbol{R}}) = \boldsymbol{D}_{\boldsymbol{S}^{-1}}^{-1} \boldsymbol{S}^{-1} \boldsymbol{D}_{\boldsymbol{S}^{-1}}^{-1}, \text{ with } \boldsymbol{S} = -\tilde{\boldsymbol{R}}. \qquad (2.8)$$

(b) The transformation function (2.7) is bijective and (2.8) its inverse.

Proof. See Appendix. □

The partial correlation matrix as defined in Equation (2.6) is used by some authors (e.g., Wong, Carter, and Kohn 2003; Stifanelli et al. 2013; Ha and Sun 2014). Other authors, however, work with partial correlation matrices that feature ones on the diagonal (e.g., Schäfer and Strimmer 2005; Marrelec et al. 2006; Lafit et al. 2019). Such matrices are, for instance, generated by three R libraries (*corpcor*, *ppcor*, and *GGMselect*) making a study of its properties useful.

Definition 2. For a p-dimensional random vector X with non singular covariance matrix Σ let the partial correlation matrix $\bar{R}$ be defined as

$$\bar{R} := \tilde{R} + 2I_p = -S + 2I_p \tag{2.9}$$

with $I_p = (\delta_{ij})_{i,j=1,\ldots,p}$ being the identity matrix of size p.

Simply adding and subtracting $2I_p$ in Equations (2.7) and (2.8), respectively, shows that $\bar{R}$, just like $\tilde{R}$, can be calculated via R and vice versa.

It now has to be established what properties a partial correlation matrix (e.g., $\tilde{R}, \bar{R}$) has to fulfill. Consider, for example, a three-variable case with partial correlations $\tilde{\rho}_{12\cdot3} = .7$ and $\tilde{\rho}_{13\cdot2} = .5$. As will subsequently be shown it can then readily be seen that it is impossible that $\tilde{\rho}_{23\cdot1}$ is equal to .6 by creating the following symmetric matrix

$$\bar{R}_c = \begin{bmatrix} 1 & .7 & .5 \\ .7 & 1 & .6 \\ .5 & .6 & 1 \end{bmatrix}$$

and calculating its eigenvalues ($\bar{\lambda}_1 = 2.2, \bar{\lambda}_2 = 0.5,$ and $\bar{\lambda}_3 = 0.3$ when rounded to one decimal).

Using Equation (2.8) for the proposed partial correlation matrix $\bar{R}_c$ would not result in a real 3×3 matrix since the square roots of the diagonal entries of

$$S_c^{-1} = \left[-(\bar{R}_c - 2I_p)\right]^{-1} \approx \begin{bmatrix} -1.23 & -1.92 & -1.77 \\ -1.92 & -1.44 & -1.83 \\ -1.77 & -1.83 & -0.98 \end{bmatrix},$$

which determine the diagonal matrix $D_{S_c^{-1}}$, are negative.

We will now derive necessary and sufficient conditions for the validity of the proposed correlation structure for both, $\tilde{R}$ and $\bar{R}$.

Proposition 1. (a) *If R is a valid non degenerate $p \times p$ correlation matrix (i.e., symmetric, all main diagonal elements equal to 1, off-diagonal values $\in [-1, 1]$, positive definite), the corresponding partial correlation matrix $\tilde{R}$ computed by Equation (2.7) is negative definite.*

(b) *If $\tilde{R}$ is a negative definite, symmetric $p \times p$ matrix with main diagonal elements all equal to -1 and off-diagonal values in $[-1, 1]$, the matrix computed by Equation (2.8) is a valid non degenerate correlation matrix R.*

Proof. (a) The solutions to the equation

$$Aw = \lambda w$$

are the eigenvectors $w \in \mathbb{R}^p$ (excluding the zero vector) and eigenvalues $\lambda \in \mathbb{R}$ of a square matrix A. If A is non singular, inversion yields

$$\frac{1}{\lambda} w = A^{-1} w. \tag{2.10}$$

Accordingly, the inverse of the matrix R (all $\lambda > 0$) is also positive definite (all $\frac{1}{\lambda} > 0$). This in turn implies that the diagonal values of R^{-1} are positive, because, for positive definite matrices, $z^T R^{-1} z > 0$ for all non zero $z \in \mathbb{R}^p$, including vectors like $z_0 = (1, 0, ..., 0)^T$.

Taking the square root of the diagonal entries of R^{-1} will therefore result in positive real numbers, which make up the diagonal elements of $D_{R^{-1}}$. Since $D_{R^{-1}}$ is a diagonal matrix, its diagonal entries are also its eigenvalues (because vectors like $w_0 = (1, 0, ..., 0)^T$ satisfy the equation $D_{R^{-1}} w = \lambda w$) and thus $D_{R^{-1}}$ must be positive definite, as well as $D_{R^{-1}}^{-1}$.

As is pointed out by Seber (2008, Definition 16.7 (2), 330), two square matrices A and B are congruent if there exists a non singular matrix M, such that

$$B = M^T A M.$$

Since $D_{R^{-1}}^{-1}$ is non singular and symmetric, R^{-1} and $D_{R^{-1}}^{-1} R^{-1} D_{R^{-1}}^{-1}$ are congruent. From Sylvester's Law of Inertia (Sylvester 1852) it then follows that the symmetric and congruent matrices R^{-1} and $D_{R^{-1}}^{-1} R^{-1} D_{R^{-1}}^{-1}$ have the same number of positive, negative and zero eigenvalues, that is, both of them are positive definite. Multiplying $D_{R^{-1}}^{-1} R^{-1} D_{R^{-1}}^{-1}$ by (-1) then reverses the sign of all its eigenvalues as a consequence of the eigenvalue equations (2.10) making $\tilde{R}$ negative definite.

(b) Since $\tilde{R}$ is negative definite all its eigenvalues are negative. Multiplying with (-1) then results in the positive definite matrix $S = -\tilde{R}$. Finally, by Sylvester's Law of Inertia, S^{-1} and $D_{S^{-1}}^{-1} S^{-1} D_{S^{-1}}^{-1} = R$ have the same amount of positive eigenvalues (all of them) making R positive definite. Hence, if $\tilde{R}$ is a partial correlation matrix the computation (2.8) results in a non degenerate correlation matrix R. ∎

Corollary 1. (a) *If R is a valid non degenerate $p \times p$ correlation matrix (i.e., symmetric, all main diagonal elements equal to 1, off-diagonal values $\in [-1, 1]$, positive definite), the corresponding partial correlation matrix $\bar{R} = -D_{R^{-1}}^{-1} R^{-1} D_{R^{-1}}^{-1} + 2I_p$ has eigenvalues $\bar{\lambda}_i < 2 \forall i = 1, ..., p$.*

(b) *If $\bar{R}$ is a symmetric $p \times p$ matrix with main diagonal elements all equal to $+1$, off-diagonal values in $[-1, 1]$, and all p eigenvalues $\bar{\lambda}_i$ less than 2, the matrix $R = D_{S^{-1}}^{-1} S^{-1} D_{S^{-1}}^{-1}$ with $S = -\bar{R} + 2I_p$ is a valid non degenerate correlation matrix.*

Proof. (a) Due to Proposition 1 we know that all eigenvalues of $\tilde{R} = \bar{R} - 2I_p$ are negative. Adding $2I_p$ to a matrix increases all its eigenvalues by 2, as mentioned by Strang (2016, 291). The reasoning behind this is simple: Take the two equations

$$2I_p w^* = 2w^*$$
$$A w = \lambda w,$$

where A is still a square matrix, and add them to

$$Aw + 2I_p w^* = \lambda w + 2w^*.$$

We can then set $w^* = w$, as all non zero vectors are eigenvectors of I_p, to get

$$(A + 2I_p)w = (\lambda + 2)w. \tag{2.11}$$

Consequently, $\bar{\lambda}_i < 2 \forall i = 1, ..., p$ holds for $\bar{R}$.

(b) Since all eigenvalues of $\bar{R}$ are smaller than 2, Proposition 1 can be applied to the negative definite matrix $\tilde{R} = \bar{R} - 2I_p$. ∎

We now established a set of necessary and sufficient conditions that a candidate matrix must fulfill to be a valid partial correlation matrix ($\tilde{R}$, or $\bar{R}$). The conditions of symmetry, all main diagonal elements equal to -1 (for $\tilde{R}$) or 1 (for $\bar{R}$), off-diagonal elements in $[-1, 1]$, and eigenvalues smaller than 0 (for $\tilde{R}$) or 2 (for $\bar{R}$) are necessary, since the partial correlation matrix corresponding to a valid correlation matrix by Equation (2.7) must fulfill these conditions (Proposition/Corollary 1, a). They are also sufficient, because the matrix resulting from Equation (2.8) then always constitutes a valid, non degenerate correlation matrix (Proposition/Corollary 1, b).

3. Implications

3.1. Algorithmic generation of (random) partial correlation structures

As there are multiple definitions for partial correlation matrices used in the literature, care for their respective properties should be taken. A partial correlation matrix defined as $\tilde{R} + I_p$, as done in Anandkumar et al. (2012), for instance, must have eigenvalues smaller than 1, as can be seen at once by looking at Corollary 1. In the case of random matrix generation, we advocate to thoroughly investigate and explicitly mention the restrictions the used algorithm has on the partial correlation structure. Schäfer and Strimmer (2005) and Tenenhaus et al. (2010), for example, both use an algorithm proposed in Schäfer and Strimmer (2005) to generate positive definite matrices with $+1$'s on the main diagonal and off-diagonal elements in $[-1, 1]$ to generate sparse partial correlation structures without offering a discussion of its properties.

A concise description of the algorithm proposed in Schäfer and Strimmer (2005) goes as follows:

1. Choose $\epsilon > 0, \eta \in [0, 1]$, and $p \in \mathbb{N}$ such that $\eta t \in \mathbb{N}$ for $t := \frac{p(p-1)}{2}$.
2. Generate the $p \times p$ zero matrix A.
3. Randomly select ηt of the t upper-diagonal elements of A and replace them with random draws from the continuous uniform distribution on the interval $[-1, 1]$.
4. Generate a symmetric matrix $M = (m_{ij})$ via $M = A + A^T$ and set its diagonal elements to $m_{jj} = \sum_{i=1}^{p} |m_{ij}| + \epsilon$.
5. Scale M via $D_M^{-1} M D_M^{-1} =: R$ with $D_M := \left(\sqrt{m_{ij} \delta_{ij}} \right)_{i,j=1,...,p}$ and δ_{ij} the Kronecker delta to receive a symmetric matrix R with $+1$'s on the main diagonal.

Due to its construction, M is symmetric and diagonally dominant with strictly positive elements on the main diagonal. From Gershgorin's circle theorem (see Gershgorin 1931, Satz II, 751) it then follows that M is positive definite. Hence, this algorithm generates positive definite matrices R with $+1$'s on the main diagonal and off-diagonal elements in $[-1, 1]$ regardless of the chosen parameters ϵ, η, and p. As we have shown in Corollary 1, the positive definiteness of matrices with $+1$'s on the main diagonal is not sufficient for valid partial correlation structures (all eigenvalues need to be smaller than 2). However, neither Schäfer and Strimmer (2005) nor Tenenhaus et al. (2010) prove that all eigenvalues of the resulting matrix R are smaller than 2. That being said, results from Monte Carlo simulations suggest that this algorithm indeed always produces valid partial correlation matrices $\bar{R}$.

As positive definiteness is not just not sufficient, it is also not necessary (an indefinite symmetric matrix with $+1$'s on the main diagonal and off-diagonal elements in $[-1,1]$ can represent a valid partial correlation structure), this algorithm restricts certain valid (sparse) structures. Moreover, it renders certain types of partial correlation structures disproportionately more likely as can be seen in Figure 1. In the case of $\eta = 0.02$ (top), the smallest eigenvalue is likely to be extremely close to zero and the largest eigenvalue is likely to be extremely close to two. In the case of $\eta = 0.05$ (bottom), we find that both the distribution of the smallest and the distribution of the largest eigenvalue are bimodal.
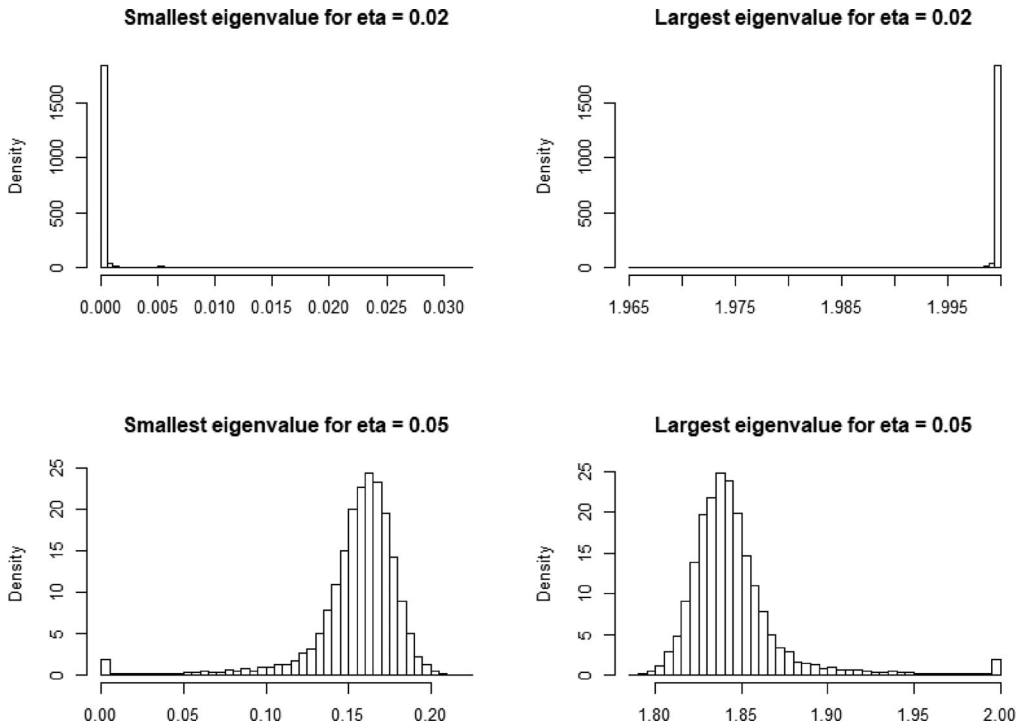
Figure 1. Distribution of the smallest and largest eigenvalue of 10000 PC matrices generated by the Algorithm proposed in Schäfer and Strimmer (2005) with $\epsilon = 0.0001$ for $p = 100$ and $\eta = 0.02$ (top) and $p = 100$ and $\eta = 0.05$ (bottom).

The following direct consequence of Proposition 1 may be useful in the development of new and better-understood algorithms.

Corollary 2. *Given a valid non degenerate correlation matrix* $\boldsymbol{R}$, *a valid partial correlation matrix* $\tilde{\boldsymbol{R}}^*$ *can easily be constructed by*

$$\tilde{\boldsymbol{R}}^* = -\boldsymbol{R}. \tag{3.1}$$

Proof. Multiplying $\boldsymbol{R}$ by (-1) renders $-\boldsymbol{R}$ negative definite and symmetric with a negative unit-diagonal and off-diagonal values in [–1, 1]. ∎

Corollary 2 tells us that changing the signs of all $\frac{p(p-1)}{2}$ Pearson correlations of a set of p variables will always result in a valid partial correlation structure for p random variables. Clearly, $\tilde{\boldsymbol{R}}^*$ is not the partial correlation matrix that structurally corresponds to the particular correlation matrix $\boldsymbol{R}$ it has been derived from. However, this transformation allows for the generation of valid partial correlation matrices by the use of existing algorithms for $\boldsymbol{\Sigma}$ or its scaled version $\boldsymbol{R}$ (e.g., Knol and ten Berge 1989; Lewandowski, Kurowicka, and Joe 2009; Pourahmadi and Wang 2015), where first a valid correlation matrix is generated and second a valid partial correlation matrix is obtained by simply switching the signs of all off-diagonal entries. Contrary to this, generating partial correlation matrices via Equation (2.7) is computationally more intricate as it involves matrix inversions.

As shown in the next section, the spaces of valid correlation and partial correlation matrices are of equal volume. Hence, the principles by which an algorithm generates valid correlation matrices must also apply to this minor extension with the only difference that the signs of the results are opposite.

3.2. The space of valid partial correlation structures

The set of conditions derived in Proposition 1 and Corollary 1 facilitate simple validity checks for conditional linear dependence structures. However, one may wonder how likely it is that a random, symmetric $p \times p$ matrix $\boldsymbol{M}_p$ with negative unit-diagonal and off-diagonal elements uniformly distributed between -1 and 1, constitutes a valid partial correlation matrix. This question has been investigated for correlation matrices by Böhm and Hornik (2014), who derived an exact analytic formula for the probability of $\boldsymbol{M}_p$ representing a valid $p \times p$ correlation matrix $\boldsymbol{R}$:

$$\mathbb{P}(\boldsymbol{M}_p \text{ is valid } \boldsymbol{R}) = \frac{2^{\sum_{i=1}^{p-1} i^2} \times \prod_{i=1}^{p-1} \left[B(\frac{i+1}{2}, \frac{i+1}{2}) \right]^i}{2^t}, \tag{3.2}$$

with Euler's beta function $B(x, y)$ and $t = \frac{p(p-1)}{2}$. This quotient consists of two volumes: The volume of "valid" column vectors $\in \mathbb{R}^t$ in the numerator (specified by Joe 2006) and the volume of the t-dimensional cube with edge length 2 in the denominator.

Proposition 2. *The probabilities given by Equation (3.2) also apply to the space of valid partial correlation matrices (i.e.,* $\mathbb{P}(\boldsymbol{M}_p \text{ is valid } \tilde{\boldsymbol{R}}) = \mathbb{P}(\boldsymbol{M}_p \text{ is valid } \boldsymbol{R}) \forall p \geq 2$).

Proof. From Corollary 2 we know that $\tilde{\boldsymbol{R}}^*$ is a valid partial correlation matrix. Representing the respective t off-diagonal elements of $\tilde{\boldsymbol{R}}^*$ in the form of a column vector, Equation (3.1) can be rewritten as

$$
\begin{bmatrix} \tilde{\rho}_{12} \\ \tilde{\rho}_{13} \\ \vdots \\ \tilde{\rho}_{(p-1)p} \end{bmatrix} = \begin{bmatrix} -1 & 0 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 \end{bmatrix} \begin{bmatrix} \rho_{12} \\ \rho_{13} \\ \vdots \\ \rho_{(p-1)p} \end{bmatrix}.
$$

This matrix transformation constitutes a reflection through the origin, with $-\boldsymbol{I}_t$ as the reflection matrix and it applies to all valid (partial) correlation matrices, since Equation (3.1) is clearly bijective. The absolute determinant of $-\boldsymbol{I}_t$ is one for every valid (partial) correlation matrix irrespective of t, hence, the preservation of volume. ∎

Proposition 2 tells us that the numerator in Equation (3.2) is equal for partial correlation matrices. Inspecting Equation (3.2) and the precise probabilities for the dimensions $p = 2$ to $p = 10$, on display in Table 1, reveals a severe curse of dimensionality.

The valid space for 3×3 correlation matrices has been plotted by Rousseeuw and Molenberghs (1994) and referred to as "elliptical tetrahedron." Figure 2 shows the preservation of volume as well as a point reflection through the origin for 3×3 partial correlation matrices. For $p > 3$, the point reflection through the origin holds too.

3.2.1. The importance of the determinant

The aim of this section is to highlight the usefulness of eigenvalues and the determinant when categorizing partial correlation structures. For every $p \times p$ correlation matrix $\boldsymbol{R}$ we know that all eigenvalues λ_i are positive and that they sum to p since $\sum_i^p \lambda_i = tr(\boldsymbol{R}) = \sum_i^p 1 = p$. Therefore, if one eigenvalue increases, at least one other eigenvalue must decrease. Since $\det(\boldsymbol{R}) = \prod_i^p \lambda_i$ it follows that $0 < \det(\boldsymbol{R}) \leq 1$ with the maximum of 1 being reached if and only if $\boldsymbol{R} = \boldsymbol{I}_p$ (i.e., if all correlations are zero).

The partial correlation matrix $\bar{\boldsymbol{R}}$, unlike $\boldsymbol{R}$, can be indefinite (i.e., at least two eigenvalues $\bar{\lambda}_i$ have opposite signs) and the bounds for $\det(\bar{\boldsymbol{R}})$ depend on p. Therefore it pays to work with $\tilde{\boldsymbol{R}}$ instead. All eigenvalues $\tilde{\lambda}_i$ of $\tilde{\boldsymbol{R}}$ are negative and $\sum_i^p \tilde{\lambda}_i = tr(\tilde{\boldsymbol{R}}) = \sum_i^p (-1) = -p$. Since $\det(\tilde{\boldsymbol{R}}) = \prod_i^p \tilde{\lambda}_i$ we further have $-1 \leq \det(\tilde{\boldsymbol{R}}) < 0$ for odd p and $0 < \det(\tilde{\boldsymbol{R}}) \leq 1$ for even p. Hence, $0 < |\det(\tilde{\boldsymbol{R}})| \leq 1$ for every p.

Table 1. A curse of dimensionality.

p	$\mathbb{P}(\boldsymbol{M}_p$ is valid $\tilde{\boldsymbol{R}})$
2	1.0000000000000000
3	0.6168502750680850
4	0.1827704518720251
5	0.0220044523675988
6	0.0009495201911854
7	0.0000132838387928
8	0.0000000554226343
9	0.0000000000641964
10	0.0000000000000194
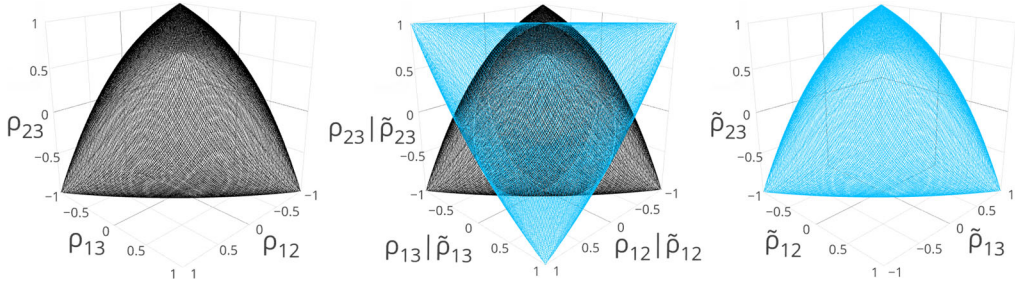
Figure 2. Spaces of valid 3×3 correlation (black) and partial correlation (blue) matrices. Note the rotation of the third graph, revealing the geometric equality of the two shapes.
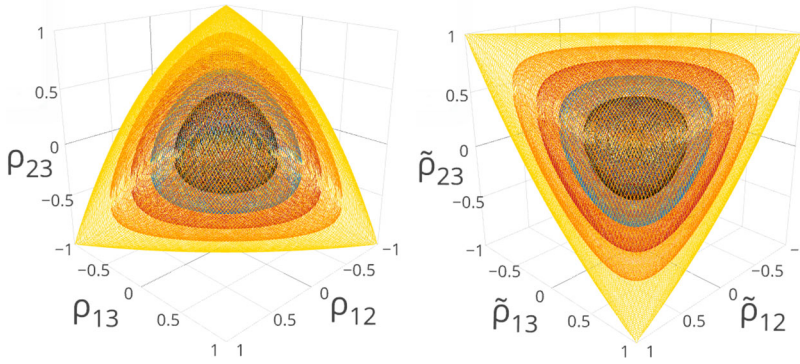

Figure 3. Subspaces with respect to the determinant for valid 3×3 correlation (left) and partial correlation (right) matrices.

Figure 3 displays subspaces for $\boldsymbol{R}$ and $\tilde{\boldsymbol{R}}$ for $p = 3$ defined via the inequalities $\det(\boldsymbol{R}) \geq x$ and $-\det(\tilde{\boldsymbol{R}}) \geq x$, respectively, for $x \in \{0.0, 0.2, 0.4, 0.6, 0.8\}$ (the corresponding colors are: yellow, orange, red, blue, black). For $x = 0$ the inequality generates the whole space of partial correlation structures (compare with Figure 2) because the first principal minor is always negative and the second principal minor is always positive. Hence, a negative determinant is equivalent to negative definiteness via Sylvester's criterion.

Figure 3 shows that $|det(\tilde{\boldsymbol{R}})|$ approaches zero as the Euclidean norm of the three-dimensional vector containing the partial correlations increases, although not at an equal rate in all directions. Due to its monotonic decrease when moving away from the origin, it can be seen as a measure of extremity – the higher the partial correlations, the lower the absolute determinant. For more information regarding the use of the determinant to describe the amount of multicollinearity, the reader is referred to Peña and Rodríguez (2003).

3.3. Sparse partial correlation structures

The obtained results allow the derivation for simple restrictions for partial correlations in sparse settings, which are commonly assumed in the field of graphical models. Three types of sparse graphical model structures are chain, ring, and star graphs (see Figure 4). In what follows, we will study the general case of $p \times p$ partial correlation
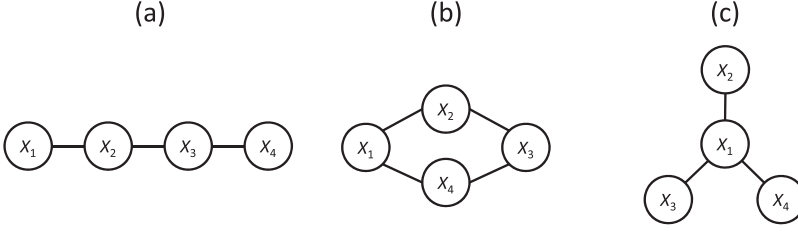
Figure 4. Some special graphical structures for four variables: (a) chain structure, (b) ring structure and (c) star structure. These three structures correspond to special partial correlation matrices.

matrices $\tilde{R}$ for these three graph types under the assumption that all partial correlations have the same value $\tilde{\rho}$ (with $|\tilde{\rho}| \in (0, 1)$).

3.3.1. Chain graphs

A chain graph for p variables corresponds to a tridiagonal partial correlation matrix $\tilde{R} = (\tilde{\rho}_{ij})_{i,j=1,...,p}$ with $\tilde{\rho}_{ij} = -1$ for $i = j$, $\tilde{\rho}_{ij} = \tilde{\rho}$ for $|i - j| = 1$, and $\tilde{\rho}_{ij} = 0$ for $|i - j| > 1$. The eigenvalues λ_j of such a matrix equal (Noschese, Pasquini, and Reichel 2013): $\lambda_j = -1 + 2|\tilde{\rho}| \cos\left(\frac{\pi j}{p+1}\right)$ for $j = 1, ..., p$. Due to Proposition 1 we have that $-1 + 2|\tilde{\rho}| \cos\left(\frac{\pi j}{p+1}\right) < 0$ for $j = 1, ..., p$ for any valid partial correlation matrix, from which it follows that

$$|\tilde{\rho}| < \frac{1}{2 \cos\left(\frac{\pi j}{p+1}\right)} \text{ for } j = 1, ..., p. \tag{3.3}$$

Because the argument of the cosine function runs from $0 < \frac{\pi}{p+1}$ to $\frac{\pi p}{p+1} < \pi$ and the cosine function is monotonically decreasing within that interval, we only have to evaluate the inequality at $\frac{\pi}{p+1}$. Hence, for $p = 3$, $|\tilde{\rho}| < \frac{1}{2 \cos\left(\frac{\pi}{4}\right)} = \frac{\sqrt{2}}{2}$. For larger values of p, we can use the small angle approximation to the cosine: $\cos(x) \approx 1 - \frac{x^2}{2}$. This yields that $|\tilde{\rho}| < \frac{1}{2 - \frac{\pi^2}{(p+1)^2}} - \epsilon_p, \epsilon_p > 0$. The larger p, the better the approximation (i.e., $\epsilon_p \to 0$ for $p \to \infty$) and for very large p, the upper bound basically becomes $\frac{1}{2}$.

3.3.2. Ring graphs

A ring graph of p variables corresponds to a symmetric circulant partial correlation matrix $\tilde{R} = (\tilde{\rho}_{ij})_{i,j=1,...,p}$ with $\tilde{\rho}_{ij} = -1$ for $i = j$, $\tilde{\rho}_{ij} = \tilde{\rho}$ for $|i - j| \in \{1, p - 1\}$, and $\tilde{\rho}_{ij} = 0$ for $1 < |i - j| < p - 1$. For such a matrix the eigenvalues are $\lambda_j = -1 + 2\tilde{\rho} \cos\left(\frac{2\pi j}{p}\right)$ for $j = 1, ..., p$. A valid partial correlation structure is then equivalent to $\tilde{\rho} \cos\left(\frac{2\pi j}{p}\right) < \frac{1}{2}$ for $j = 1, ..., p$ due to Proposition 1. Since these inequalities are trivially fulfilled if the two factors have opposite signs, we only have to consider the case of equal signs. Taking the absolute value on both sides of the inequalities we find:

$$|\tilde{\rho}| < \frac{1}{2}, \tag{3.4}$$

because $|\cos\left(\frac{2\pi j}{p}\right)| \leq 1$ for $j = 1, ..., p$. Since the bounds in (3.4) do not depend on the dimension p, valid ring structures can easily be created. The fact that the bounds on $\tilde{\rho}$ in (3.3) converge to $\frac{1}{2}$ makes intuitive sense since every chain graph is just a ring graph minus one edge. Whereas a long chain graph and a ring graph with the same $\tilde{\rho}$ are very similar, the missing edge makes a difference for small number of variables.

3.3.3. Star graphs

As a last special structure, assume a star graph where the first variable is equally strong connected to all other variables and no other edges exist (see Figure 4). In this case, $\tilde{R}$ has a symmetric arrowhead structure:

$$\tilde{R} = \begin{bmatrix} -1 & r \\ r^T & -I_{p-1} \end{bmatrix}.$$

with $r = \tilde{\rho} 1_{p-1}$ and 1_{p-1} the $(p-1)$-dimensional unit vector. The eigenvalues $\tilde{\lambda}$ of $\tilde{R}$ are the roots of the characteristic polynomial of $\tilde{R}$:

$$\tilde{\lambda} \in \mathbb{R} : \det\begin{bmatrix} (-1 - \tilde{\lambda})I_{p-1} & r \\ r^T & -1 - \tilde{\lambda} \end{bmatrix} = 0.$$

Using a classical result for the determinant of block matrices (e.g., Harville 1998, Theorem 13.3.8) we get

$$\det\begin{bmatrix} (-1 - \tilde{\lambda})I_{p-1} & r \\ r^T & -1 - \tilde{\lambda} \end{bmatrix} = \det\left[(-1 - \tilde{\lambda})I_{p-1}\right]\det\left[(-1 - \tilde{\lambda}) - r^T((-1 - \tilde{\lambda})I_{p-1})^{-1}r\right]$$

$$= (-1 - \tilde{\lambda})^{p-1}\det\left[(-1 - \tilde{\lambda}) - (-1 - \tilde{\lambda})^{-1}\tilde{\rho}^2(p - 1)\right]$$

$$= (-1 - \tilde{\lambda})^{p-2}\left[(-1 - \tilde{\lambda})^2 - \tilde{\rho}^2(p - 1)\right].$$

Root $\tilde{\lambda} = -1$ has multiplicity $p - 2$. Solving

$$(-1 - \tilde{\lambda})^2 - \tilde{\rho}^2(p - 1) = 0.$$

for $\tilde{\lambda}$ determines the remaining two roots:

$$\tilde{\lambda}_{1,2} = -1 \pm \tilde{\rho}\sqrt{p - 1}.$$

As a result of Proposition 1 it then holds that

$$|\tilde{\rho}| < \frac{1}{\sqrt{p - 1}}. \tag{3.5}$$

For star graphs, the upper limit for the absolute value of $\tilde{\rho}$ converges to zero with an increase in the number of variables, which makes them very restrictive in contrast to chain and ring structures.

4. Concluding remarks

Partial correlations and the corresponding matrices are used in different settings in applied statistics, such as Gaussian graphical models and directed acyclic graphs (DAGs). This article aimed at the clarification of common misconceptions about partial correlations stemming from a scattered literature. Through precise definitions of partial correlation matrices, we were able to derive their properties, which led to several implications.

A first important implication is that when generating random partial correlation structures, caution is advised, as only a careful examination of the characteristics of the used algorithm allows an assessment of the generalizability of conclusions drawn from the generated data. In particular, positive definiteness of the commonly used symmetric matrix with $+1$'s on the main diagonal and the partial correlations as its off-diagonal elements (see, e.g., Schäfer and Strimmer 2005; Tenenhaus et al. 2010) is neither necessary nor sufficient for the validity of the correlation structure. Generating only positive definite matrices $\bar{R}$ is particularly restrictive in non sparse and high-dimensional settings: The probability that a valid 5×5 matrix $\bar{R}$ is positive definite is approximately 6% if sampled uniformly over the whole valid space. For a valid 7×7 matrix it is less than 0.1% and for a valid 9×9 matrix it is already less than 0.0001%.[2] As algorithmic generation of valid partial correlation structures is useful in many domains future research on existing algorithms, particularly on elaborate algorithms that generate a random number of non zero partial correlations as used for the study of Erdös–Rényi graph models (Erdős and Rényi 1960), as well as the development of new algorithms, is much needed.

A second implication is that the spaces of valid correlation and partial correlation matrices have the same volume regardless of dimension since they are point reflections through the origin of one another. This relation allows the (random) generation of valid partial correlation structures by switching the sign of valid Pearson correlation structures generated by existing algorithms. However, the distributional properties of the resulting algorithm then need to be investigated thoroughly.

A final implication is that it allowed the derivation of simple restrictions on the partial correlations for certain commonly used types of sparse structures. Naturally, this work could be extended in the future to other types of sparse structures.

Notes

1. Strictly speaking Lewis and Styan (1981) only proved that $\tilde{R}_{1,2} = \tilde{\rho}_{12}$. However, by pre- and post-multiplying Σ with permutation matrices U^T and U such that the i-th and j-th row/column respectively become the 1-st and 2-nd row/column we find that $\tilde{R}_{i,j} = \tilde{\rho}_{ij}$ since $(U^T \Sigma U)_{12}^{-1} = U^T \Sigma^{-1} U_{12} = \Sigma_{ij}^{-1}$.
2. These results were generated by repeatedly generating random partial correlation matrices in d dimensions in R (R Core Team 2018) via *cor2pcor(rcorrmatrix(d))*.

ORCID

Richard Artner http://orcid.org/0000-0002-4515-5650
Paul P. Wellingerhof http://orcid.org/0000-0003-3318-7003
Ginette Lafit http://orcid.org/0000-0002-8227-128X
Tim Loossens http://orcid.org/0000-0002-7587-1922
Wolf Vanpaemel http://orcid.org/0000-0002-5855-3885
Francis Tuerlinckx http://orcid.org/0000-0002-1775-7654

References

Anandkumar, A., V. Y. Tan, F. Huang, and A. S. Willsky. 2012. High-dimensional Gaussian graphical model selection: Walk summability and local separation criterion. *Journal of Machine Learning Research* 13:2293–337.

Antti, K., and S. Puntanen. 1983. A connection between the partial correlation coefficient and the correlation coefficient of certain residuals. *Communications in Statistics - Simulation and Computation* 12 (5):639–41. doi:10.1080/03610918308812348.

Baba, K., R. Shibata, and M. Sibuya. 2004. Partial correlation and conditional correlation as measures of conditional independence. *Australian & New Zealand Journal of Statistics* 46 (4): 657–64. doi:10.1111/j.1467-842X.2004.00360.x.

Böhm, W., and K. Hornik. 2014. Generating random correlation matrices by the simple rejection method: Why it does not work. *Statistics & Probability Letters* 87:27–30. doi:10.1016/j.spl.2013.12.012.

Cramer, H. 1946. *Mathematical methods of statistics*, 500. Princeton, NJ: Princeton University Press.

Epskamp, S., L. J. Waldorp, R. Mõttus, and D. Borsboom. 2018. The Gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research* 53 (4):453–80. doi:10.1080/00273171.2018.1454823.

Erdős, P., and A. Rényi. 1960. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5 (1):17–60.

Gershgorin, S. A. 1931. Über die Abgrenzung der Eigenwerte einer Matrix. *Bulletin de L'Académie Des Sciences de l'URSS. Classe Des Sciences Mathématiques et na* 6:749–54.

Ha, M. J., and W. Sun. 2014. Partial correlation matrix estimation using ridge penalty followed by thresholding and re-estimation. *Biometrics* 70 (3):762–70. doi:10.1111/biom.12186.

Harville, D. A. 1998. *Matrix algebra from a statistician's perspective*. New York City, NY: Springer.

Joe, H. 2006. Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis* 97 (10):2177–89. doi:10.1016/j.jmva.2005.05.010.

Knol, D. L., and J. M. ten Berge. 1989. Least-squares approximation of an improper correlation matrix by a proper one. *Psychometrika* 54 (1):53–61. doi:10.1007/BF02294448.

Lafit, G., F. Tuerlinckx, I. Myin-Germeys, and E. Ceulemans. 2019. A partial correlation screening approach for controlling the false positive rate in sparse gaussian graphical models. *Scientific Reports* 9 (1):1–24. doi:10.1038/s41598-019-53795-x.

Lauritzen, S. L. 1996. *Graphical models*. Oxford, England: Clarendon Press.

Lewandowski, D., D. Kurowicka, and H. Joe. 2009. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis* 100 (9):1989–2001. doi:10.1016/j.jmva.2009.04.008.

Lewis, M. C., and G. P. Styan. 1981. Equalities and inequalities for conditional and partial correlation coefficients. In *Statistics and related topics: Proceedings of the International Symposium on Statistics and Related Topics: Ottawa, May 1980*, ed. M. Csorgo, D. A. Dawson, J. N. K. Rao, and A. K. Saleh, 57–65. Amsterdam: North–Holland.

Marrelec, G., A. Krainik, H. Duffau, M. Pélégrini-Issac, S. Lehéricy, J. Doyon, and H. Benali. 2006. Partial correlation for functional brain interactivity investigation in functional MRI. *Neuroimage* 32 (1):228–37. doi:10.1016/j.neuroimage.2005.12.057.

Noschese, S., L. Pasquini, and L. Reichel. 2013. Tridiagonal toeplitz matrices: Properties and novel applications. *Numerical Linear Algebra with Applications* 20 (2):302–26. doi:10.1002/nla.1811.

Peña, D., and J. Rodríguez. 2003. Descriptive measures of multivariate scatter and linear dependence. *Journal of Multivariate Analysis* 85 (2):361–74. doi:10.1016/S0047-259X(02)00061-1.

Pourahmadi, M., and X. Wang. 2015. Distribution of random correlation matrices: Hyperspherical parameterization of the cholesky factor. *Statistics & Probability Letters* 106: 5–12. doi:10.1016/j.spl.2015.06.015.

R Core Team. 2018. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rousseeuw, P. J., and G. Molenberghs. 1994. The shape of correlation matrices. *The American Statistician* 48 (4):276–79. doi:10.2307/2684832.

Schäfer, J., and K. Strimmer. 2005. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics (Oxford, England)* 21 (6):754–64. doi:10.1093/bioinformatics/bti062.

Seber, G. A. 2008. *A matrix handbook for statisticians*. Hoboken, NJ: John Wiley & Sons.

Stifanelli, P. F., T. M. Creanza, R. Anglani, V. C. Liuzzi, S. Mukherjee, F. P. Schena, and N. Ancona. 2013. A comparative study of covariance selection models for the inference of gene regulatory networks. *Journal of Biomedical Informatics* 46 (5):894–904. doi:10.1016/j.jbi.2013.07.002.

Strang, G. 2016. *Introduction to linear algebra*. Wellesley, MA: Wellesley-Cambridge Press.

Sylvester, J. J. 1852. XIX. A demonstration of the theorem that every homogeneous quadratic polynomial is reducible by real orthogonal substitutions to the form of a sum of positive and negative squares. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 4 (23):138–42. doi:10.1080/14786445208647087.

Tenenhaus, A., V. Guillemot, X. Gidrol, and V. Frouin. 2010. Gene association networks from microarray data using a regularized estimation of partial correlation based on PLS regression. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 7 (2):251–62. doi:10.1109/TCBB.2008.87.

Wong, F., C. K. Carter, and R. Kohn. 2003. Efficient estimation of covariance selection models. *Biometrika* 90 (4):809–30. doi:10.1093/biomet/90.4.809.

Appendix

A. Operators

For the subsequent formal proofs, we introduce the following operators.

A.1. Pth Hadamard product

The first operator works on (only positive) matrices and extends the mechanism of the "simple" Hadamard product (HP) (see Seber 2008, Definiton 11.12, p. 251). For two matrices $T, U \in \mathbb{R}_+^{n \times m}$ with $T = [t_{ik}]$ and $U = [u_{ik}]$, the HP is defined as $T \circ U = [t_{ik} u_{ik}]$, for $i = 1, ..., n$ and $k = 1, ..., m$. The said extension now describes a formalization of the pth HP of a matrix T with itself: $T^{\circ p} = [t_{ik}^p]$, where $p \in \mathbb{R}$. For example, taking the square root of each element would be expressed as $T^{\circ 1/2} = [t_{ik}^{1/2}]$. For readability, we will write this as $\sqrt{T}$.

A.2. $diag_X$ and $diag_C$

The other two operators are closely related. $\text{diag}_X(A)$ extracts the diagonal elements of a square matrix A, resulting in a vector $[a_{11}, ..., a_{nn}]^T$, with a_{ii} the ith diagonal entry of A, $i = 1, ..., n$. Oppositely, $\text{diag}_C(z)$ creates a diagonal matrix, which ith diagonal entry equals the value z_i of the vector $z \in \mathbb{R}^n$. For diagonal matrices, both operators are the inverse function of the other one.

A.3. Properties

The following properties apply to the operators $T^{\circ p}, \mathrm{diag}_X$ and diag_C, as they were defined above, for all $T, U \in \mathbb{R}_+^{n \times m}$, $v, z \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$ and diagonal matrices $D \in \mathbb{R}^{n \times n}$, respectively:

(P1.1) $(T \circ U)^{\circ p} = T^{\circ p} \circ U^{\circ p}$

(P1.2) $(T^{\circ p})^{\circ 1/p} = T$

(P1.3) $\mathrm{diag}_C(v)\mathrm{diag}_C(z) = \mathrm{diag}_C(z)\mathrm{diag}_C(v) = \mathrm{diag}_C(z \circ v)$

(P1.4) $\mathrm{diag}_X(AD) = \mathrm{diag}_X(DA) = \mathrm{diag}_X(D) \circ \mathrm{diag}_X(A)$

(P1.1) and (P1.2) readily follow from power laws:

$$
\begin{aligned}
(X \circ Y)^{\circ p} &= \left[(x_{ij}y_{ij})^p\right] \\
&= \left[x_{ij}^p y_{ij}^p\right] \\
&= X^{\circ p} \circ Y^{\circ p} \\
(X^{\circ p})^{\circ 1/p} &= \left[x_{ij}^p\right]^{\circ 1/p} \\
&= \left[(x_{ij}^p)^{1/p}\right] \\
&= X
\end{aligned}
$$

(P1.3) is true because the matrix product of two diagonal matrices does essentially the same operations as the HP of two vectors, as all off-diagonal entries are zero. The HP is clearly commutative, as are diagonal matrices with other diagonal matrices.

In the matrix product AD of (P1.4), D does not operate only on A's diagonal of course, instead, each element of the jth column of A gets multiplied by the ith diagonal element of D. Nevertheless (P1.4) always holds since

$$
\begin{aligned}
\mathrm{diag}_X(AD) &= [a_{11}d_1, ..., a_{nn}d_n]^T \\
&= [d_1 a_{11}, ..., d_n a_{nn}]^T \\
&= \mathrm{diag}_X(DA) \\
&= \mathrm{diag}_X(D) \circ \mathrm{diag}_X(A)
\end{aligned}
$$

with a_{jj} the jth diagonal element of A, $j = 1, ..., n$.

B. Formal definition of scaling matrices

For a square matrix M we define the scaling matrix D_M^{-1} as the inverse of

$$
D_M := \mathrm{diag}_C\left[\sqrt{\mathrm{diag}_X(M)}\right]
$$

C. Proof of Lemma 1

For a regular covariance matrix Σ the correlation matrix is defined as

$$
R := D_\Sigma^{-1} \Sigma D_\Sigma^{-1}, \tag{C.1}
$$

which can be rearranged to

$$
\Sigma = D_\Sigma R D_\Sigma.
$$

Inverting then yields

$$
\Sigma^{-1} = D_\Sigma^{-1} R^{-1} D_\Sigma^{-1}. \tag{C.2}
$$

Inserting Equation (C.2) in Equation (2.6) gives

$$\tilde{R} = -D_{\Sigma^{-1}}^{-1}D_{\Sigma}^{-1}R^{-1}D_{\Sigma}^{-1}D_{\Sigma^{-1}}^{-1},$$

which reduces to the proof to showing that

$$D_{R^{-1}}^{-1} = D_{\Sigma^{-1}}^{-1}D_{\Sigma}^{-1} = D_{\Sigma}^{-1}D_{\Sigma^{-1}}^{-1}, \tag{C.3}$$

From inserting the inverse of Equation (C.1) into the definition of $D_{R^{-1}}$, and from (P1.1) to (P1.4) and the commutativity of the HP, it follows that

$$D_{R^{-1}} = \text{diag}_C\left[\sqrt{\text{diag}_X(D_\Sigma\Sigma^{-1}D_\Sigma)}\right]$$

$$= \text{diag}_C\left[\sqrt{\text{diag}_X(\text{diag}_C\left[\sqrt{\text{diag}_X(\Sigma)}\right]\Sigma^{-1}\text{diag}_C\left[\sqrt{\text{diag}_X(\Sigma)}\right])}\right]$$

$$= \text{diag}_C\left[\sqrt{\text{diag}_X(\text{diag}_C\left[\sqrt{\text{diag}_X(\Sigma)}\right]\Sigma^{-1}\text{diag}_C\left[\sqrt{\text{diag}_X(\Sigma)}\right])}\right]$$

$$= \text{diag}_C\left[\sqrt{\text{diag}_X(\text{diag}_C\left[\sqrt{\text{diag}_X(\Sigma)}\right])\circ\text{diag}_X(\Sigma^{-1})\circ\text{diag}_X(\text{diag}_C\left[\sqrt{\text{diag}_X(\Sigma)}\right])}\right]$$

$$= \text{diag}_C\left[\sqrt{\text{diag}_X(\Sigma^{-1})\circ\text{diag}_X(\text{diag}_C\left[\sqrt{\text{diag}_X(\Sigma)}\right]\text{diag}_C\left[\sqrt{\text{diag}_X(\Sigma)}\right])}\right]$$

$$= \text{diag}_C\left[\sqrt{\text{diag}_X(\Sigma^{-1})\circ\text{diag}_X(\text{diag}_C\left[\sqrt{\text{diag}_X(\Sigma)}\circ\sqrt{\text{diag}_X(\Sigma)}\right])}\right]$$

$$= \text{diag}_C\left[\sqrt{\text{diag}_X(\Sigma^{-1})\circ\text{diag}_X(\Sigma)}\right]$$

$$= \text{diag}_C\left[\sqrt{\text{diag}_X(\Sigma^{-1})}\circ\sqrt{\text{diag}_X(\Sigma)}\right]$$

$$= \text{diag}_C\left[\sqrt{\text{diag}_X(\Sigma^{-1})}\right]\text{diag}_C\left[\sqrt{\text{diag}_X(\Sigma)}\right]$$

$$= D_{\Sigma^{-1}}D_\Sigma.$$

Inverting this relation then leads to Equation (C.3) since $D_{\Sigma^{-1}}D_\Sigma = D_\Sigma D_{\Sigma^{-1}}$ due to the commutativity of diagonal matrices. ∎

D. Proof of Lemma 2

(a): From Equations (2.6) and (2.7) we have

$$\tilde{R} = -S = -D_{R^{-1}}^{-1}R^{-1}D_{R^{-1}}^{-1}, \tag{D.1}$$

and it becomes clear that R can be computed from $\tilde{R}$ if $D_{R^{-1}}$ is known. In particular, we have

$$D_{R^{-1}}SD_{R^{-1}} = R^{-1}$$

which inverts to

$$R = D_{R^{-1}}^{-1}S^{-1}D_{R^{-1}}^{-1}. \tag{D.2}$$

18 😄 R. ARTNER ET AL.

From this, Equation (2.8) follows if we can show that

$$\boldsymbol{D}_{S^{-1}} = \boldsymbol{D}_{R^{-1}}. \tag{D.3}$$

Now, by using *(P1.1)* to *(P1.4)* once again, inserting a rearranged Equation (D.2) for $\boldsymbol{S}^{-1}$ and the fact that the diagonal of $\boldsymbol{R}$ is a vector of ones, we get

$$\begin{aligned}
\boldsymbol{D}_{S^{-1}} &= \mathrm{diag}_C\left[\sqrt{\mathrm{diag}_X(\boldsymbol{D}_{R^{-1}}\boldsymbol{R}\boldsymbol{D}_{R^{-1}})}\right]\\
&= \mathrm{diag}_C\left[\sqrt{\mathrm{diag}_X(\boldsymbol{D}_{R^{-1}})\circ\mathrm{diag}_X(\boldsymbol{R})\circ\mathrm{diag}_X(\boldsymbol{D}_{R^{-1}})}\right]\\
&= \mathrm{diag}_C\left[\mathrm{diag}_X(\boldsymbol{D}_{R^{-1}})\right]\\
&= \boldsymbol{D}_{R^{-1}}.
\end{aligned}$$

(b): Using Equation (D.3) we find that

$$\begin{aligned}
g(f(\boldsymbol{R})) &= g(-\boldsymbol{D}_{R^{-1}}^{-1}\boldsymbol{R}^{-1}\boldsymbol{D}_{R^{-1}}^{-1})\\
&= \boldsymbol{D}_{S^{-1}}^{-1}(-(-\boldsymbol{D}_{R^{-1}}^{-1}\boldsymbol{R}^{-1}\boldsymbol{D}_{R^{-1}}^{-1}))^{-1}\boldsymbol{D}_{S^{-1}}^{-1}\\
&= \boldsymbol{D}_{S^{-1}}^{-1}(\boldsymbol{D}_{R^{-1}}^{-1}\boldsymbol{R}^{-1}\boldsymbol{D}_{R^{-1}}^{-1})^{-1}\boldsymbol{D}_{S^{-1}}^{-1}\\
&= \boldsymbol{D}_{S^{-1}}^{-1}\boldsymbol{D}_{R^{-1}}\boldsymbol{R}\boldsymbol{D}_{R^{-1}}\boldsymbol{D}_{S^{-1}}^{-1}\\
&= \boldsymbol{R},
\end{aligned}$$

and conversely

$$\begin{aligned}
f(g(\tilde{\boldsymbol{R}})) &= f(\boldsymbol{D}_{S^{-1}}^{-1}(-\tilde{\boldsymbol{R}})^{-1}\boldsymbol{D}_{S^{-1}}^{-1})\\
&= -\boldsymbol{D}_{R^{-1}}^{-1}(\boldsymbol{D}_{S^{-1}}^{-1}(-\tilde{\boldsymbol{R}})^{-1}\boldsymbol{D}_{S^{-1}}^{-1})^{-1}\boldsymbol{D}_{R^{-1}}^{-1}\\
&= -\boldsymbol{D}_{R^{-1}}^{-1}\boldsymbol{D}_{S^{-1}}(-\tilde{\boldsymbol{R}})\boldsymbol{D}_{S^{-1}}\boldsymbol{D}_{R^{-1}}^{-1}\\
&= -(-\tilde{\boldsymbol{R}})\\
&= \tilde{\boldsymbol{R}} \qquad\qquad\qquad\blacksquare
\end{aligned}$$