arXiv:2401.05256v1 [math.ST] 10 Jan 2024

Tests of Missing Completely At Random based on sample covariance matrices

Alberto Bordino and Thomas B. Berrett

Department of Statistics, University of Warwick

Abstract

We study the problem of testing whether the missing values of a potentially high-dimensional dataset are Missing Completely at Random (MCAR). We relax the problem of testing MCAR to the problem of testing the compatibility of a sequence of covariance matrices, motivated by the fact that this procedure is feasible when the dimension grows with the sample size. Tests of compatibility can be used to test the feasibility of positive semi-definite matrix completion problems with noisy observations, and thus our results may be of independent interest. Our first contributions are to define a natural measure of the incompatibility of a sequence of correlation matrices, which can be characterised as the optimal value of a Semi-definite Programming (SDP) problem, and to establish a key duality result allowing its practical computation and interpretation. By studying the concentration properties of the natural plug-in estimator of this measure, we introduce novel hypothesis tests that we prove have power against all distributions with incompatible covariance matrices. The choice of critical values for our tests rely on a new concentration inequality for the Pearson sample correlation matrix, which may be of interest more widely. By considering key examples of missingness structures, we demonstrate that our procedures are minimax rate optimal in certain cases. We further validate our methodology with numerical simulations that provide evidence of validity and power, even when data are heavy tailed.

1 Introduction

Incomplete data are a common occurrence in almost all areas of statistical application, and the mechanisms leading to such data are diverse. For example, subjects in a survey may choose not to respond to certain questions, leading to missing values, or a practitioner may wish to combine data collected in different studies, where different variables were recorded in each. With incomplete data, traditional approaches become unreliable or even inapplicable, leading to a significant effect on the conclusions that can be drawn from the data. The most common approaches to dealing with missing values are to remove any incomplete observations, and thus to perform a complete-case analysis, or to replace any missing entry with a representative value, using an imputation method (e.g. Yates, 1933; van Buuren and Groothuis-Oudshoorn, 2011; Stekhoven and Bühlmann, 2011). However, the validity of such procedures, and the choice of an appropriate one, depends crucially on the mechanism that determines the missingness. Mechanisms have traditionally been classified as Missing Completely At Random (MCAR), Missing At Random (MAR) and Missing Not At

Random (MNAR) (e.g. Little and Rubin, 2002) according to the dependence structure between the variables themselves and their missingness, with such assumptions being required to link observations to targets of inference.

The typical formal setting is to suppose that we observe independent and identically distributed copies of a random object $\mathbf{X} \circ \Omega$, where $\mathbf{X}$ takes values in some product space $\mathcal{X} = \prod_{j=1}^{d} \mathcal{X}_j$, where Ω takes values in $\{0,1\}^d$ and where we define the operator $\circ$ by

$$(x \circ \omega)_j = \begin{cases} x_j & \text{if } \omega_j = 1, \\ \text{NA} & \text{if } \omega_j = 0. \end{cases}$$

The assumptions named above then control the dependence between the data $\mathbf{X}$ and the missingness indicator Ω. The simplest case of MCAR is when these are independent, denoted $\mathbf{X} \perp\!\!\!\perp \Omega$, which essentially means that the data we observe is representative of the population, even if it is incomplete. For example, consider the simple problem of estimating $\mathbb{E}[XY]$ from complete data $(X_1, Y_1), \dots, (X_N, Y_N)$ and incomplete data $(X_{N+1}, \text{NA}), \dots, (X_n, \text{NA})$. It is easy to see that under MCAR the complete case estimator $\widehat{\mu}^{CC} := N^{-1} \sum_{i=1}^{N} X_i Y_i$ is an unbiased estimator of $\mathbb{E}[XY]$. When MCAR or similar assumptions hold, we can often employ statistical methodologies that are easy to interpret and make good use of all incomplete data, and solid theoretical guarantees have been developed in various modern statistical problems such as high-dimensional regression (Loh and Wainwright, 2012), high-dimensional or sparse principal component analysis (Zhu et al., 2022; Elsener and van de Geer, 2018), classification (Cai and Zhang, 2019; Sell et al., 2023), and precision matrix and changepoint estimation (Follain et al., 2022). However, if MCAR does not hold, which is common in practice, alternative methods may be required.

Hypothesis tests can be used to guide practitioners in deciding whether or not missingness assumptions are reasonable. The goal of this work is to study the problem of testing the hypothesis of MCAR, which has been the subject of much research in the missing data literature. Most prior work has been developed within the context of parametric models. For example, Little (1988), works under the hypothesis that the data is Gaussian in the setting that all pairs of variable are observed together (see Section 5 for further details). Fuchs (1982) considers discrete data in the setting that a large number of complete cases are available. In both cases the methods are likelihood ratio tests, with the MLEs calculated using the EM algorithm (Dempster et al., 1977) and validity and power guarantees based on classical asymptotics. More recently, Berrett and Samworth (2023) provided a nonparametric formulation of the problem and methodology that was proved to be widely powerful under minimal assumptions. The key insight of Berrett and Samworth (2023) is to relate the problem of testing MCAR to the problem of testing *compatibility*. For $S \subseteq [d] := \{1, \dots, d\}$ denote by $\{\Omega = \mathbb{1}_S\}$ the event that X_j is observed if and only if $j \in S$, and write $\mathbb{S} = \{S : \mathbb{P}(\Omega = \mathbb{1}_S) > 0\}$ for the set of all possible observation patterns. Then, under MCAR, the distribution P_S of the observation $X_S | \{\Omega = \mathbb{1}_S\}$ is equal to the marginal distribution of the population distribution $\mathcal{L}(\mathbf{X})$ on $\mathcal{X}_S := \prod_{j \in S} \mathcal{X}_j$. Hence, if $P_{\mathbb{S}} := (P_S : S \in \mathbb{S})$ is *incompatible*, in the sense that there is no distribution P on $\mathcal{X}$ with marginal distribution P_S on $\mathcal{X}_S$ for all $S \in \mathbb{S}$, then the data cannot be MCAR. In fact, it is shown that this reasoning is tight in that it is not possible to rule out MCAR based on observations of $\mathbf{X} \circ \Omega$ if $P_{\mathbb{S}}$ is compatible. In general, fully testing the compatibility of a sequence of distributions requires us to look at complex interactions between the distributions, and methods for doing so will have sample complexity that is exponential in the dimension d. Our work aims to provide methods that are valid and powerful without strong assumptions while being effective as the dimension d grows.

Our methodology will be based on testing the compatibility of sequences of covariance matrices, which can be estimated consistently even for large d. Earlier studies have employed the covariance matrix to assess MCAR. As briefly discussed above, Little (1988) studied a likelihood ratio test of MCAR, effectively examining the homogeneity of means and covariances under the assumption of normality. However, Little expressed scepticism about its effectiveness unless the sample size is exceptionally large and the assumption of normality holds. This scepticism was further validated in simulations by Kim and Bentler (2002), who also developed a test for consistency of means and covariances based on generalised least squares. Both of these approaches work by comparing the sample covariance matrix associated to a given missingness pattern to the corresponding submatrix of an estimated complete covariance matrix. More recently, Jamshidian and Jalal (2010) developed k-sample tests of the equality of covariance matrices, given complete data, based on Hawkins' test (Hawkins, 1981). Using empirical evidence, they then argued that these tests could be combined with imputation techniques to test the homogeneity of covariance matrices calculated using incomplete data. These methodologies can be effective when the corresponding assumptions are met and when a complete covariance matrix can be consistently estimated.

Our method works by directly checking the compatibility of the observed sample covariance matrices, making no assumptions on the form of the underlying distributions and not requiring the estimation of a complete covariance matrix. In particular, this second point means that our test can be applied with any collection $\mathbb{S}$ of missingness patterns. More precisely, at the population level, we will consider $\Sigma_{\mathbb{S}} = (\Sigma_S : S \in \mathbb{S})$, a sequence of suitably-normalised covariance matrices Σ_S associated to the law of $\mathbf{X}_S|\{\Omega = \mathbb{1}_S\}$, and design a statistical test to check if $\Sigma_{\mathbb{S}}$ is compatible, meaning that each Σ_S can be obtained by marginalising a general $d \times d$ covariance matrix Σ, i.e. $(\Sigma)_S = \Sigma_S$. If MCAR holds then for each $S \in \mathbb{S}$ we must have $\mathrm{Cov}(\mathbf{X}_S|\Omega = \mathbb{1}_S) = (\mathrm{Cov}(\mathbf{X}))_S$, the block of the covariance matrix of $\mathbf{X}$ corresponding to the variables in S, so that the sequence $(\Sigma_S : S \in \mathbb{S})$ must be compatible. Hence, if we can reject the hypothesis $H_0 : \Sigma_{\mathbb{S}}$ *compatible*, then we can reject the hypothesis of MCAR. See Figure 1 for a pictorial summary of the key concepts so far.

More generally, one can consider the problem of testing the compatibility of moments of order $p \geq 1$ and, if it is found that these moments are incompatible, one can reject MCAR. For $p = 1$, this problem reduces to testing the compatibility of mean vectors, which essentially boils down to testing the equality of means. This has been studied in the statistical literature for nearly a century, and we refer to existing methods for solving this problem (e.g. Wilks, 1946; Little, 1988). In order to have power against a wider range of alternatives, while limiting the complexity of the testing procedure, we restrict attention in this work to the natural $p = 2$ problem. Here there are still various ways in which compatibility can fail. For example, we can rule out H_0 if $\Sigma_{\mathbb{S}}$ is *inconsistent*, in the sense that that there are two observation patterns $S_1, S_2 \in \mathbb{S}$ for which $(\Sigma_{S_1})_{S_1 \cap S_2} \neq (\Sigma_{S_2})_{S_1 \cap S_2}$, meaning that there exists a pair of variables whose covariance takes different values in different observation patterns. Testing the consistency of covariance matrices reduces to testing the equality of smaller covariance matrices, which has again been previously studied (e.g. Hawkins, 1981). The corresponding nonparametric problem of testing the consistency of distributions was studied by Li and Yu (2015); Spohn et al. (2021) in the context of testing MCAR. However, there exist $\Sigma_{\mathbb{S}}$ that are consistent but not compatible. As a concrete example, consider the case where $d = 3$, where $\mathbb{S} = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$,

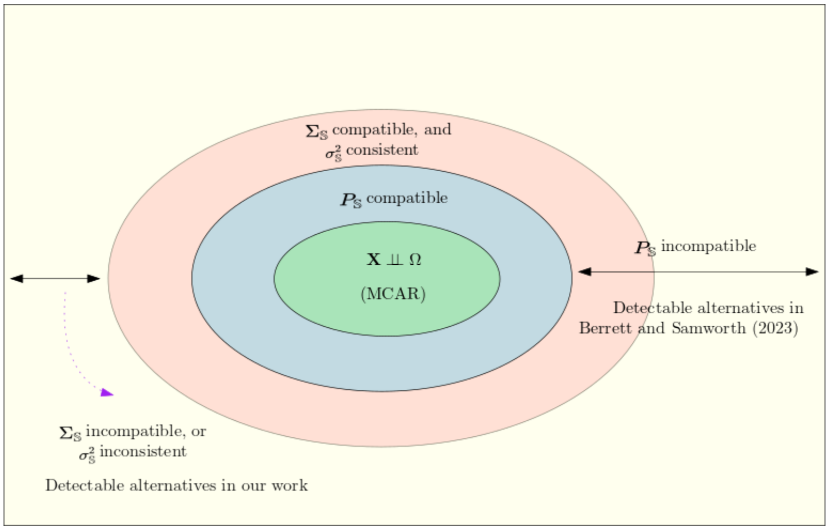

Figure 1: Our framework. We relax the methodology in Berrett and Samworth (2023) and consider covariance matrices instead of full distributions. The price we pay is to create an extra ring (red area) in which we cannot detect departure from H_0 just by looking at $\Sigma_{\mathbb{S}}$ or $\sigma^2_{\mathbb{S}}$, the sequence of variances. For example, if the sequence of third-moment tensors were inconsistent, but $\Sigma_{\mathbb{S}}$ were compatible and $\sigma^2_{\mathbb{S}}$ consistent, we would not be able to reject MCAR, although $P_{\mathbb{S}}$ would be incompatible.

and where

$$\Sigma_{ij} = \begin{pmatrix} 1 & \rho_{ij} \\ \rho_{ij} & 1 \end{pmatrix}$$

with $\rho_{23} = \rho_{13} = -\rho_{12} = \rho$. Then $\Sigma_{\mathbb{S}}$ is compatible if and only if $\rho \leq 1/2$, even though it is always consistent.

The above example is relatively simple because any pair of variables is observed together so that the full covariance matrix can be estimated. However, we can characterise compatibility for any $\mathbb{S}$ (see Proposition 2). While the compatibility of distributions can be characterised using linear programming (e.g. Kellerer, 1984), characterising the compatibility of covariance matrices requires ideas from semi-definite programming (SDP), which studies linear optimisation problems over spectrahedra (e.g. Blekherman et al., 2012; Vandenberghe and Boyd, 1996). If $\Sigma_{\mathbb{S}}$ is consistent, compatibility is equivalent to the feasibility of a positive semi-definite matrix completion problem, where we observe a partial symmetric matrix $A = (a_{ij})$ for positions (i, j) in a certain set of edges, and aim to construct a positive semi-definite completion of A. This problem is extensively studied owing to its widespread applications in diverse fields such as probability, statistics, systems engineering and geophysics; see, for example, Laurent (2009) and the references therein for an introduction to the topic. Statistical questions associated with such problems are relatively under-explored, though we mention the recent work Waghmare and Panaretos (2022) that provides estimated completions of covariance operators in settings where completions always exist. A distinct but related problem that has received more attention in the statistics literature is low-rank matrix completion. In particular, significant contributions (Candès and Recht, 2009; Candès and Tao, 2010; Recht, 2011) have been made in the realm of convex optimization, where a low-rank matrix is recovered from partial observations after introducing a nuclear norm penalty. In our work we make no low-rank assumptions and our main interest is in answering the question of whether or not a positive semi-definite completion exists.

We now briefly outline our main contributions. In Section 2 we define a numerical measure $R(\Sigma_{\mathbb{S}})$ of the incompatibility of a sequence of correlation matrices $\Sigma_{\mathbb{S}}$ and establish some key properties, including an interpretable and useful dual representation (Proposition 3). We then combine this index of incompatibility with a measure $V(\sigma^2_{\mathbb{S}})$ of the inconsistency of variances (e.g. Proposition 5), in order to introduce a numerical measure of the incompatibility of sequences of covariance matrices. In Section 3 we turn to the empirical estimation of such indices and the introduction of data-driven testing procedures. Under the non-singularity assumption that $\Sigma_S \succeq cI_S$ for all $S \in \mathbb{S}$, we first introduce an oracle test that relies on knowledge of $c > 0$ and give a result on its validity and power (Theorem 6). Next, we avoid this restriction by introducing a sample-splitting test that we prove to be valid (Proposition 8). The analysis of both tests is based on a novel concentration inequality for the spectral norm of the difference between the Pearson sample correlation matrix and its population version (Proposition 7). In Section 4 we study the performance of our oracle test in various examples and show that its separation rate is near-minimax optimal in some cases, while studying properties of the associated semi-definite programs. In Section 5 we validate our methodology in numerical experiments. Section 6 contains the proofs of our main results. The Appendix contains background and auxiliary results.

We conclude the introduction with some notation that is used throughout the paper. For $d \in \mathbb{N}$, we write $[d] := \{1, \cdots, d\}$. Given $a, b \geq 0$, we write $a \lesssim b$ to mean that there exists a universal constant $C > 0$ such that $a \leq Cb$. We use $a \wedge b$ for $\min\{a, b\}$, and $a \vee b$ for $\max\{a, b\}$. We will denote with $\mathbf{0}_d$ the null vector of dimension d, with $\mathbf{1}_d$ the all-one vector, with $\boldsymbol{e}_j$ the j-th element of the canonical basis of $\mathbb{R}^d$, with $\boldsymbol{O}_d$ the zero matrix of dimension d, and with I_d the identity matrix of dimension d. We will omit the subscript with the dimension d when it is clear from the context. For symmetric matrices A, B of dimension d, we write $A \succeq 0$ to mean that A is positive semi-definite, write $A \succeq B$ to mean that $A - B \succeq 0$, write $\mathrm{diag}(A)$ to indicate the diagonal matrix having the same diagonal as A, and $\mathrm{diag}(\boldsymbol{v})$ for a vector $\boldsymbol{v} = (v_1, \ldots, v_d)$ to indicate a diagonal matrix with diagonal elements equal to v_i. We will indicate the trace of A with $\mathrm{tr}(A)$, the determinant with either $|A|$ or $\det(A)$, and the minimum and maximum eigenvalues of A with $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$, respectively. We will use $\|\cdot\|_*$ for nuclear norm, or Schatten-1 norm of a matrix, $\|\cdot\|_2$ for the spectral norm, and $\|\cdot\|_F$ for the Frobenius norm. For random elements X, Y, we write $X \perp\!\!\!\perp Y$ to mean that X and Y are independent. For $\sigma > 0$, a random variable X with mean $\mu = \mathbb{E}[X]$ is said to be σ-subgaussian if

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \leq e^{\sigma^2\lambda^2/2} \quad \text{for all } \lambda \in \mathbb{R},$$

while, for $(\nu, \alpha) \in (0, \infty)^2$, it is said to be (ν, α)-subexponential if

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \leq e^{\frac{\nu^2\lambda^2}{2}} \quad \text{for all } |\lambda| < \frac{1}{\alpha}.$$

A random vector $\mathbf{X}$ in $\mathbb{R}^n$ is said to be σ-subgaussian if every one-dimensional projection, i.e. $v^T\mathbf{X}$ with $v \in \mathbb{R}^n$ and $\|v\| = 1$, is σ-subgaussian in the sense defined above.

2 Measure of incompatibility for covariance matrices

In this section we develop our index of the incompatibility of population covariance matrices. We check the compatibility of covariance matrices by checking the consistency of the variances and the compatibility of the correlation matrices separately. Standardising the covariance matrices is necessary to have a well-posed problem, and we choose to work with correlation matrices because the resulting compatibility measure is more tractable. Other standardisation are possible, though, each leading to a different measure of incompatibility, with different properties. In Appendix B we introduce another measure of incompatibility based on a different standardisation, analyse its properties, and derive a test of MCAR based on its estimation from data. Here, we standardise the sequence of covariance matrices, resulting in a sequence of correlation matrices $\Sigma_{\mathbb{S}}$, introduce a measure of incompatibility for $\Sigma_{\mathbb{S}}$, and develop some of its basic properties. We then define a measure of the inconsistency of variances, and combine the two for an overall measure of the incompatibility of sequences of covariance matrices. In order to do this, we must first introduce some basic algebraic objects for sequences of symmetric and positive semi-definite matrices. Our key notation is collected in Table 1 below.

Notation	**Definition**	**Meaning**
$\mathbb{S}$	A subset of the power set of $[d]$	Set of all missingness patterns
$\mathbb{S}_j$	$\{S \in \mathbb{S} : j \in S\}$	Set of all patterns that contain j
$\mathbb{S}_{jj'}$	$\{S \in \mathbb{S} : j, j' \in S\}$	Set of all patterns that contain (j, j')
$\mathcal{M} \equiv \mathcal{M}_d$	$\{X \in \mathbb{R}^{d\times d} : X = X^T\}$	Space of symmetric matrices
$\mathcal{M}_{\mathbb{S}}$	$\{(X_S : S \in \mathbb{S}) : X_S \in \mathcal{M}_{\lvert S\rvert}$ for all $S \in \mathbb{S}\}$	Space of sequences of symmetric matrices
$\langle X, Y\rangle$	$\mathrm{tr}(XY)$ for $X, Y \in \mathcal{M}$	Frobenius inner product
$\langle X_{\mathbb{S}}, Y_{\mathbb{S}}\rangle_{\mathbb{S}}$	$\sum_{S\in\mathbb{S}} \mathrm{tr}(X_S Y_S)$ for $X_{\mathbb{S}}, Y_{\mathbb{S}} \in \mathcal{M}_{\mathbb{S}}$	Sum of Frobenius inner products
$\mathcal{P}^*$	$\{\Sigma \in \mathcal{M} : \Sigma \succeq 0\}$	Cone of PSD matrices
$\mathcal{P}$	$\{\Sigma \in \mathcal{P}^* : \mathrm{diag}(\Sigma) = I_d\}$	Set of correlation matrices
$\Sigma_{\mathbb{S}} \succeq_{\mathbb{S}} 0$	$\Sigma_{\mathbb{S}} \succeq_{\mathbb{S}} 0$ if and only if $\Sigma_S \succeq 0$ for all $S \in \mathbb{S}$	Loewner order for sequences of matrices
$\mathcal{P}^*_{\mathbb{S}}$	$\{\Sigma_{\mathbb{S}} \in \mathcal{M}_{\mathbb{S}} : \Sigma_{\mathbb{S}} \succeq_{\mathbb{S}} 0\}$	Sequences of PSD matrices
$\mathcal{P}_{\mathbb{S}}$	$\{\Sigma_{\mathbb{S}} \in \mathcal{P}^*_{\mathbb{S}} : \mathrm{diag}(\Sigma_S) = I_{\lvert S\rvert}$ for all $S \in \mathbb{S}\}$	Sequence of correlation matrices
A	$A : \mathcal{M} \to \mathcal{M}_{\mathbb{S}}$ with $(AX)_S = (X_{jj'})_{j,j'\in S}$	Marginalisation operator on matrices
$\mathcal{P}^{0,*}_{\mathbb{S}}$	$\{A\Sigma : \Sigma \in \mathcal{P}^*\}$	Set of compatible sequences of PSD matrices
$\mathcal{P}^0_{\mathbb{S}}$	$\{A\Sigma : \Sigma \in \mathcal{P}\}$	Set of compatible sequences of correlation matrices
$\mathcal{Y}$	$\{\mathrm{diag}(v) : v \in \mathbb{R}^d$ and $\sum_{j=1}^d v_j = 0\}$	Space of diagonal matrices with null trace
$\boldsymbol{O}_{\mathbb{S}}$	$(O_{\lvert S\rvert} : S \in \mathbb{S})$	Sequence of zero matrices
$I_{\mathbb{S}}$	$(I_{\lvert S\rvert} : S \in \mathbb{S})$	Sequence of identity matrices

Table 1: Table of definitions commonly used in the main text.

Crucially, we say that an element of $\mathcal{P}^*_{\mathbb{S}}$ is *compatible* if and only if it is an element of $\mathcal{P}^{0,*}_{\mathbb{S}}$. In order to characterise compatibility, we first state a basic property of the linear operator A defined in Table 1 above.

Proposition 1. *The adjoint operator $A^* : \mathcal{M}_{\mathbb{S}} \to \mathcal{M}$ of A is given by*

$$(A^* X_{\mathbb{S}})_{jj'} = \sum_{S \in \mathbb{S}} \mathbb{1}_{\{j,j' \in S\}} (X_S)_{jj'} = \sum_{S \in \mathbb{S}_{jj'}} (X_S)_{jj'},$$

where we recall that $\mathbb{S}_{jj'} = \{S \in \mathbb{S} : j, j' \in S\} = \mathbb{S}_j \cap \mathbb{S}_{j'}$.

Now, the following proposition fully characterises compatibility in terms of the non-negativity of a collection of linear functionals.

Proposition 2. *For $\Sigma_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^*$ we have $\Sigma_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^{0,*}$ if and only if*

$$\langle X_{\mathbb{S}}, \Sigma_{\mathbb{S}} \rangle_{\mathbb{S}} \geq 0 \quad \textit{for all } X_{\mathbb{S}} \in \mathcal{M}_{\mathbb{S}} \textit{ satisfying } A^* X_{\mathbb{S}} \succeq 0.$$

The proof can be found in Section 6. This is an extension of well-known characterisation of the feasibility of positive semi-definite matrix completion (e.g. Laurent, 2009). Indeed, when $\Sigma_{\mathbb{S}}$ is consistent, we can show that $\langle X_{\mathbb{S}}, \Sigma_{\mathbb{S}} \rangle_{\mathbb{S}} = \langle A^* X_{\mathbb{S}}, \Sigma \rangle$, where Σ is the $d \times d$ symmetric matrix with $\Sigma_{jj'} = (\Sigma_S)_{jj'}$ for all $S \in \mathbb{S}_{jj'}$, if $\mathbb{S}_{jj'} \neq \emptyset$, and $\Sigma_{jj'} = 0$ if $\mathbb{S}_{jj'} = \emptyset$. Here Σ can be thought of as a partial matrix that is padded with zeros in unobserved positions. Since $A^* X_{\mathbb{S}}$ is also zero in these positions, the value of Σ there is arbitrary. Now our characterisation reduces to checking that $\langle A^* X_{\mathbb{S}}, \Sigma \rangle \geq 0$ for all $X_{\mathbb{S}} \in \mathcal{M}_{\mathbb{S}}$ satisfying $A^* X_{\mathbb{S}} \succeq 0$, which is equivalent to checking $\langle X, \Sigma \rangle \geq 0$ for all $X \in \mathcal{M}$ with $X_{jj'} = 0$ if $S_{jj'} = \emptyset$, which coincides with (4) in Laurent (2009).

Proposition 2 provides a characterisation of compatibility, but in order to assess the significance of departures from the null hypothesis and thus to define hypothesis tests, we will need a numerical measure of incompatibility. A natural way to do this is to minimise $\langle X_{\mathbb{S}}, \Sigma_{\mathbb{S}} \rangle_{\mathbb{S}}$ over a subset of $\{X_{\mathbb{S}} \in \mathcal{M}_{\mathbb{S}} : A^* X_{\mathbb{S}} \succeq 0\}$ that still characterises compatibility, but gives finite minimal values. First, observe that checking the compatibility of covariance matrices is equivalent to checking the consistency of the variances $\sigma_S^2 = \mathrm{diag}(\mathrm{Cov}(\mathbf{X}_S | \Omega = \mathbb{1}_S)$ for $S \in \mathbb{S}$, and the compatibility of the correlation matrices $\mathrm{Corr}(\mathbf{X}_S | \Omega = \mathbb{1}_S)$ for $S \in \mathbb{S}$. Now, whenever $\Sigma_{\mathbb{S}}$ is a sequence of correlation matrices we define

$$\begin{aligned} R(\Sigma_{\mathbb{S}}) &:= \sup\Big\{ -\frac{1}{d} \langle \Sigma_{\mathbb{S}}, X_{\mathbb{S}} \rangle_{\mathbb{S}} : X_{\mathbb{S}} + X_{\mathbb{S}}^0 \succeq_{\mathbb{S}} 0, A^* X_{\mathbb{S}} + Y \succeq 0 \text{ for some } Y \in \mathcal{Y} \Big\} \\ &= 1 - \frac{1}{d} \inf\{ \langle \Sigma_{\mathbb{S}}, Y_{\mathbb{S}} \rangle_{\mathbb{S}} : Y_{\mathbb{S}} \succeq_{\mathbb{S}} 0, A^* Y_{\mathbb{S}} + Y \succeq I_d \text{ for some } Y \in \mathcal{Y} \} \end{aligned} \tag{1}$$

where $X_{\mathbb{S}}^0 = (X_S^0 : S \in \mathbb{S})$, where $X_S^0 = \mathrm{diag}\left(1/\left|\mathbb{S}_j\right| : j \in S\right)$, and where $\mathcal{Y}$ is the set of diagonal $d \times d$ matrices with trace equal to zero. The objective function of this optimisation problem is a one-to-one mapping of the linear functional appearing in our characterisation of compatibility. Moreover, by choosing $Y = \boldsymbol{O}$ and noting that $X_{\mathbb{S}}^0 \succ_{\mathbb{S}} 0$, we can see that for any $X_{\mathbb{S}}$ that satisfies $A^* X_{\mathbb{S}} \succeq 0$, the sequence $\epsilon X_{\mathbb{S}}$ is feasible for $\epsilon > 0$ sufficiently small. Thus, by Proposition 2 we have that $R(\Sigma_{\mathbb{S}}) > 0$ whenever $\Sigma_{\mathbb{S}}$ is incompatible. On the other hand, when $\Sigma_{\mathbb{S}} = A\Sigma$ is compatible and $(X_{\mathbb{S}}, Y)$ is feasible we have $\langle \Sigma_{\mathbb{S}}, X_{\mathbb{S}} \rangle_{\mathbb{S}} = \langle \Sigma, A^* X_{\mathbb{S}} \rangle = \langle \Sigma, A^* X_{\mathbb{S}} + Y \rangle \geq 0$, where the second equality holds because Σ has a constant diagonal. Combining this with the observation that $X_{\mathbb{S}} = \boldsymbol{O}_{\mathbb{S}}$ is feasible, we see that $R(\Sigma_{\mathbb{S}}) = 0$ when $\Sigma_{\mathbb{S}}$ is compatible.

In the above argument we did not use the specific form of the lower bound $X_{\mathbb{S}} \succeq_{\mathbb{S}} -X_{\mathbb{S}}^0$ anywhere, and it would also have been possible to optimise over the restricted set of $X_{\mathbb{S}}$ that are feasible with $Y = \boldsymbol{O}$. The specific choice of the feasible set in the definition of $R(\Sigma_{\mathbb{S}})$ was made because it leads to an interpretable dual formulation. While there exist semi-definite programs for which strong duality does not hold, Slater's condition (see Appendix C for an introduction to the theory of semi-definite programming) is satisfied in our problem, so we do not encounter such issues. This is formalised in the result below.

Proposition 3. *For $\Sigma_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}$ we have*

$$\begin{aligned} R(\Sigma_{\mathbb{S}}) &= \inf\{\epsilon \in [0,1] : \Sigma_{\mathbb{S}} \in (1-\epsilon)\mathcal{P}^0_{\mathbb{S}} + \epsilon\mathcal{P}_{\mathbb{S}}\} \\ &= 1 - \frac{1}{d}\sup\{\operatorname{tr}(\Sigma) : A\Sigma \preceq_{\mathbb{S}} \Sigma_{\mathbb{S}},\, \Sigma_{11} = \ldots = \Sigma_{dd},\, \Sigma \succeq 0\}. \end{aligned} \tag{2}$$

This result shows that our measure of incompatibility $R(\Sigma_{\mathbb{S}})$ can be interpreted as the smallest amount of perturbation ϵ that a compatible sequence of correlation matrices must be corrupted by to result in the input sequence $\Sigma_{\mathbb{S}}$. It is immediate from this representation that $R(\Sigma_{\mathbb{S}})$ takes values in $[0,1]$. Moreover, it follows from Slater's condition that the optimal value of the dual problem is attained. Thus, writing $\lambda^* = 1 - R(\Sigma_{\mathbb{S}})$, we can write $\Sigma_{\mathbb{S}}$ as

$$\Sigma_{\mathbb{S}} = \lambda^* A\Sigma + (1-\lambda^*)\Sigma'_{\mathbb{S}}, \tag{3}$$

where $\Sigma \in \mathcal{P}$ and $\Sigma'_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}$. By the maximality of $\lambda^* = 1 - R(\Sigma_{\mathbb{S}})$, it must be the case that $R(\Sigma'_{\mathbb{S}}) = 1$. Indeed, if this were not the case, it would be possible to write $\Sigma'_{\mathbb{S}} = \lambda' A\Sigma' + (1-\lambda')\Sigma''_{\mathbb{S}}$ for some $\lambda' \in (0,1]$, which would contradict the fact that λ^* is optimal. This argument shows that, whenever $\mathbb{S}$ is such that there exists an incompatible sequence $\Sigma_{\mathbb{S}}$, the maximal value $R(\Sigma_{\mathbb{S}}) = 1$ is attainable, so that the quantity $R(\Sigma_{\mathbb{S}})$ is on an interpretable scale between compatibility at one extreme and maximal incompatibility at the other.

We remark that this dual interpretation of $R(\cdot)$ aligns with a similar representation of the incompatiblity of sequences of distributions defined by Berrett and Samworth (2023). In this earlier work it is shown that the incompatibility of sequences of distributions can be understood through linear programming techniques. Our work here, however, shows that we must consider the more complex problem of semi-definite programming to understand the incompatibility of sequences of covariance matrices. Despite this additional complexity, since Slater's condition is satisfied for our problem, the primal-dual interior point method has a computational complexity which is polynomial in the number of constraints and the dimension of the unknown square matrix (Section 6.4.1. of Nesterov and Nemirovskii (1994), Section 5.7. of Vandenberghe and Boyd (1996)). This ensures that $R(\Sigma_{\mathbb{S}})$ can be always computed efficiently without additional assumptions.

We conclude with some basic properties of $R(\cdot)$.

Proposition 4. *The following hold:*

(i) *$R(\cdot)$ is convex.*

(ii) *$R(\cdot)$ is continuous.*

(iii) *Suppose $\mathbb{S} \subseteq \mathbb{S}'$ and $\Sigma_{\mathbb{S}} \subseteq_{\mathbb{S}} \Sigma_{\mathbb{S}'}$, where the inclusion $\subseteq_{\mathbb{S}}$ means that every correlation matrix in $\Sigma_{\mathbb{S}}$ is also in $\Sigma_{\mathbb{S}'}$. Then $R(\Sigma_{\mathbb{S}}) \leq R(\Sigma_{\mathbb{S}'})$.*

It is interesting to observe that property (iii) says that R is monotone with respect to the inclusion operator, so that additional information can only make a sequence less compatible.

Having established the main properties of our measures of the incompatiblity of correlation matrices, we now turn to the simpler problem of measuring the inconsistency of individual variances. There could be cases for which variances are not consistent across different patterns, but a test based on $R(\cdot)$ alone would fail to reject the null hypothesis. In order to take into account such deviations from the null, we define an analogous test statistic for the consistency of the variances. To this aim, writing $\sigma^2_{\mathbb{S}} = (\sigma^2_S : S \in \mathbb{S})$ for the collection of individual variances, we fix our units of measurement such that

$$\bar{\mathrm{av}}_j(\sigma^2_{\mathbb{S}}) := |\mathbb{S}_j|^{-1} \sum_{S \in \mathbb{S}_j} \sigma^2_{S,j} = 1,$$

for all $j \in [d]$. This is a natural constraint, analogous to the standardisation of variables in complete-data problems, that does not remove information that may be present in the individual variances. For such $\sigma^2_{\mathbb{S}}$, define

$$V(\sigma^2_{\mathbb{S}}) := 1 - \min_{j \in [d]} \min_{S \in \mathbb{S}_j} \sigma^2_{S,j} = \max_{j \in [d]} \max_{S \in \mathbb{S}_j} \left(1 - \sigma^2_{S,j}\right).$$

Under the hypothesis $\bar{\mathrm{av}}_j(\sigma^2_{\mathbb{S}}) = 1$ for all $j \in [d]$, it is clear that $V(\sigma^2_{\mathbb{S}}) = 0$ if and only if $\sigma^2_{S,j} = 1$ for all $j \in [d]$, $S \in \mathbb{S}_j$. On the other hand, we have $V(\sigma^2_{\mathbb{S}}) > 0$, if and only if there exists at least one variance strictly less than 1. It is clear from the definition that V is bounded by one, and that this extreme value is attainable when $\mathbb{S}$ is non-trivial and there exists j such $\sigma^2_{S,j} = 0$ for some $S \in \mathbb{S}_j$. The following result gives a dual representation for $V(\sigma^2_{\mathbb{S}})$, providing justification for our specific measure of inconsistency.

Proposition 5. *If* $\bar{\mathrm{av}}_j(\sigma^2_{\mathbb{S}}) = 1$ *for all* $j \in [d]$, *then*

$$V(\sigma^2_{\mathbb{S}}) = \inf\left\{\epsilon \in [0,1] : \sigma^2_{\mathbb{S}} = (1-\epsilon) A_V \mathbf{1}_d + \epsilon {\sigma'}^2_{\mathbb{S}} \text{ with } \bar{\mathrm{av}}_j({\sigma'}^2_{\mathbb{S}}) = 1 \text{ for all } j \in [d]\right\},$$

where $(A_V \sigma^2)_S = (\sigma^2_k)_{k \in S}$.

This result gives a dual representation for $V(\sigma^2_{\mathbb{S}})$, which is analogous to Proposition 3 and leads to a similar interpretation, based on the idea of finding the smallest perturbation to make the sequence consistent.

Combining our measures of the incompatibility of correlation matrices and the inconsistency of variances, we define an overall measure of the incompatibility of sequences of covariance matrices by

$$T = R(\Sigma_{\mathbb{S}}) + V(\sigma^2_{\mathbb{S}}).$$

It is easy to see that $T = 0$ if and only if the sequence of correlation matrices $\Sigma_{\mathbb{S}}$ is compatible and the sequence of variances $\sigma^2_{\mathbb{S}}$ is consistent, and that $T \leq 2$.

3 Statistical tests

Having introduced our population-level measure of incompatibility, in this section we design a testing procedure based on the estimation of T to test compatibility with finite amounts of data. For each $S \in \mathbb{S}$ we assume that we have access to an independent sample

$$X_{S,1}, \ldots, X_{S,n_S} \overset{\text{i.i.d.}}{\sim} P_S$$

for some sample size n_S and some distribution P_S with correlation matrix Σ_S and vector of variances σ_S^2. We write $\widehat{\Sigma}_S$ and $\widehat{\sigma}_S^2$ for the sample correlation matrix and sample variances, respectively, of $X_{S,1}, \ldots, X_{S,n_S}$ for each $S \in \mathbb{S}$, and write $\widehat{\Sigma}_{\mathbb{S}} = (\widehat{\Sigma}_S : S \in \mathbb{S})$ and $\widehat{\sigma}_{\mathbb{S}}^2 = (\widehat{\sigma}_S^2 : S \in \mathbb{S})$ for the sequences collecting these estimators. Our approach is based on analysing the concentration properties of the plug-in estimates $R(\widehat{\Sigma}_{\mathbb{S}})$ and $V(\widehat{\sigma}_{\mathbb{S}}^2)$ and thus introducing suitable critical values. The analysis of $R(\widehat{\Sigma}_{\mathbb{S}})$, in particular, is challenging, as it is defined as the optimal value of a semi-definite program with an unbounded feasible set. In fact, without further assumptions, it is not possible to restrict attention to a compact feasible set. On the other hand, most statistical techniques for the analysis of suprema of empirical processes require feasible sets to be totally bounded so that, for example, covering arguments can be applied.

Fortunately, under the assumption that $\Sigma_{\mathbb{S}} \succ_{\mathbb{S}} 0$, our dual problem (2) is strictly feasible and hence Slater's condition implies that the optimal value is attained in the primal problem (1). This assumption is reasonable in many areas of application, and similar assumptions of invertibility have been used frequently in the literature (Meinshausen and Bühlmann, 2006; Cai et al., 2011). In fact, if we assume the stronger condition that $\Sigma_{\mathbb{S}} \succeq_{\mathbb{S}} cI_{\mathbb{S}}$ for some $c > 0$, we will see that the optimal value is always attained in a compact set whose size depends on c. Indeed, the strict feasibility of the dual problem (2) implies that there exists $X_{\mathbb{S}} \in \mathcal{M}_{\mathbb{S}}$ such that $X_{\mathbb{S}} + X_{\mathbb{S}}^0 \succeq_{\mathbb{S}} 0$ and $R(\Sigma_{\mathbb{S}}) = -d^{-1}\langle X_{\mathbb{S}}, \Sigma_{\mathbb{S}}\rangle_{\mathbb{S}}$. This in turn implies that

$$\sum_{S \in \mathbb{S}} \|X_S + X_S^0\|_* = \langle X_{\mathbb{S}} + X_{\mathbb{S}}^0, I_{\mathbb{S}}\rangle_{\mathbb{S}} \leq \frac{1}{c}\langle X_{\mathbb{S}} + X_{\mathbb{S}}^0, \Sigma_{\mathbb{S}}\rangle_{\mathbb{S}} = \frac{d}{c}\{1 - R(\Sigma_{\mathbb{S}})\} \leq \frac{d}{c},$$

so that we have a bound on the sum of the nuclear norms of the matrices in the sequence $X_{\mathbb{S}} + X_{\mathbb{S}}^0$. In finding the optimal value of the primal problem (1), then, we may restrict attention to

$$\mathcal{F}_c := \left\{X_{\mathbb{S}} \in \mathcal{M}_{\mathbb{S}} : X_{\mathbb{S}} + X_{\mathbb{S}}^0 \succeq_{\mathbb{S}} 0, \sum_{S \in \mathbb{S}} \|X_S + X_S^0\|_* \leq d/c, A^* X_{\mathbb{S}} + Y \succeq 0 \text{ for some } Y \in \mathcal{Y}\right\}, \tag{4}$$

which is compact.

Before moving on to describe how to construct a statistical test under this new assumption, we give a brief discussion of the norm on $\mathcal{M}_{\mathbb{S}}$ defined by

$$\|X_{\mathbb{S}}\|_{*,\mathbb{S}} := \sum_{S \in \mathbb{S}} \|X_S\|_*,$$

which reduces to $\sum_{S \in \mathbb{S}} \mathrm{tr}(X_S)$ in case that $X_{\mathbb{S}} \succeq_{\mathbb{S}} 0$. For each $S \in \mathbb{S}$, the nuclear norm $\|X_S\|_*$ can be thought of as the ℓ_1 norm applied to the eigenvalues of X_S. As these are then summed to give $\|X_{\mathbb{S}}\|_{*,\mathbb{S}}$, we see that $\|\cdot\|_{*,\mathbb{S}}$ can be thought of as an ℓ_1 norm on $\mathcal{M}_{\mathbb{S}}$. It is easy to see that the dual norm of $\|\cdot\|_{*,\mathbb{S}}$ with respect to the inner product $\langle \cdot, \cdot \rangle_{\mathbb{S}}$ is

$$\|X_{\mathbb{S}}\|_{2,\mathbb{S}} := \max_{S \in \mathbb{S}} \|X_S\|_2,$$

where $\|\cdot\|_2$ is the usual spectral norm of a matrix. This follows after writing the sequence of matrices in block-diagonal form, and allows us to derive the following generalisation of Holder's inequality in the space of matrix sequences,

$$|\langle X_{\mathbb{S}}, Y_{\mathbb{S}}\rangle_{\mathbb{S}}| \leq \|X_{\mathbb{S}}\|_{*,\mathbb{S}}\|Y_{\mathbb{S}}\|_{2,\mathbb{S}} = \sum_{S\in\mathbb{S}} \|Y_S\|_* \cdot \max_{S\in\mathbb{S}} \|X_S\|_2. \tag{5}$$

This inequality will be used in the proof of the following result, which provides valid critical values for the test statistic $\widehat{T} = R(\widehat{\Sigma}_{\mathbb{S}}) + V(\widehat{\sigma}^2_{\mathbb{S}})$ and gives conditions under which the resulting test has large power.

Theorem 6. *Suppose we observe $\mathbf{X}_{S,1}, \ldots, \mathbf{X}_{S,n_S} \overset{i.i.d.}{\sim} P_S$ for each $S \in \mathbb{S}$ independently, where each P_S is ν-subgaussian, with the sequence of variances $\sigma^2_{\mathbb{S}}$ satisfying $\bar{\mathrm{av}}_j(\sigma^2_{\mathbb{S}}) = 1$ for all $j \in [d]$, and the sequence of population correlation matrices $\Sigma_{\mathbb{S}}$ satisfying $\Sigma_{\mathbb{S}} \succeq_{\mathbb{S}} cI_{\mathbb{S}}$ for $c > 0$. Define $C_\alpha = \max\{C^{(R)}_\alpha, C^{(V)}_\alpha\}$, with*

$$C^{(V)}_\alpha = C_0\nu^2\sqrt{\frac{\log\left(\sum_{j\in[d]} |\mathbb{S}_j|/\alpha\right)}{\min_{S\in\mathbb{S}} n_S}},$$

and

$$C^{(R)}_\alpha = \frac{1}{c}\max_{S\in\mathbb{S}}\Bigg\{C_1\frac{\nu^2}{\sigma^2_{\min}}\sqrt{\frac{|S| + \log(|\mathbb{S}|/\alpha)}{n_S}} \vee \frac{|S| + \log(|\mathbb{S}|/\alpha)}{n_S} + C_2\frac{\nu^4}{\sigma^4_{\min}}\sqrt{\frac{|S|\log(|S||\mathbb{S}|/\alpha)}{n_S}} + C_3\frac{\nu^4}{\sigma^4_{\min}}\left(\sqrt{\frac{|S| + \log(|\mathbb{S}|/\alpha)}{n_S}} \vee \frac{|S| + \log(|\mathbb{S}|/\alpha)}{n_S}\right)\sqrt{\frac{|S|\log(|S|/\alpha)}{n_S}}\Bigg\},$$

for universal constants $C_0, C_1, C_2, C_3 > 0$. Then, for C_0, C_1, C_2, C_3 and universal constant $K > 0$ chosen sufficiently large, for all $\alpha \in (0,1)$ such that $\sigma^4_{\min}n \geq K\nu^4\log(\max_{S\in\mathbb{S}}|S|/\alpha)$, the test that rejects H_0 if and only if $\widehat{T} = R(\widehat{\Sigma}_{\mathbb{S}}) + V(\widehat{\sigma}^2_{\mathbb{S}}) \geq C_\alpha$ has Type I error bounded by α. Moreover, for $\beta \in (0,1)$ such that $\sigma^4_{\min}n \geq K\nu^4\log(\max_{S\in\mathbb{S}}|S|/\beta)$, if $T > C_\alpha + C_\beta$, then $\mathbb{P}(\widehat{T} \leq C_\alpha) \leq \beta$.

In proving this result we give concentration inequalities for the random quantities $R(\widehat{\Sigma}_{\mathbb{S}})$ and $V(\widehat{\sigma}^2_{\mathbb{S}})$. The analysis of $R(\widehat{\Sigma}_{\mathbb{S}})$ is crucially based on the fact that, under H_0 and in light of the inequality (5), we can control the oscillation $|R(\widehat{\Sigma}_{\mathbb{S}}) - R(\Sigma_{\mathbb{S}})|$ using $\max_{S\in\mathbb{S}}\|\widehat{\Sigma}_S - \Sigma_S\|_2$, where the Σ_S are the Pearson population correlation matrices and $\widehat{\Sigma}_S$ are the corresponding Pearson sample correlation matrices. To this end, we derive the following tail bound for the spectral norm $\|\widehat{P} - P\|_2$, where P is the population correlation matrix and $\widehat{P}$ is the sample correlation matrix of complete data, which may be of independent interest.

Proposition 7. *Suppose we observe an i.i.d sample $\mathbf{X}_1, \cdots, \mathbf{X}_n \sim \mathbf{X}$, where $\mathbf{X}$ is a ν-subgaussian random vector in $\mathbb{R}^d$, and let Σ and $\widehat{\Sigma}$ be the population and sample covariance matrices, respectively. Let $P = D^{-1/2}\Sigma D^{-1/2}$ be the population correlation matrix, where $D = \mathrm{diag}(\Sigma)$, and $\widehat{P} = \widehat{D}^{-1/2}\widehat{\Sigma}\widehat{D}^{-1/2}$ be the sample correlation matrix, where $\widehat{D} = \mathrm{diag}(\widehat{\Sigma})$. Then, there exist universal constants $C_1, C_2, C_3, K > 0$ such that, for every $t \in [0,1]$ such that $\sigma^4_{\min}n \geq K\nu^4\log(d/t)$ we have*

$$\|\widehat{P} - P\|_2 \leq C_1\frac{\nu^2}{\sigma^2_{\min}}\left(\sqrt{\frac{d + \log(1/t)}{n}} \vee \frac{d + \log(1/t)}{n}\right) + C_2\frac{\nu^4}{\sigma^4_{\min}}\sqrt{\frac{d\log(d/t)}{n}} + C_3\frac{\nu^4}{\sigma^4_{\min}}\left(\sqrt{\frac{d + \log(1/t)}{n}} \vee \frac{d + \log(1/t)}{n}\right)\sqrt{\frac{d\log(d/t)}{n}}$$

with probability $\geq 1-t$*, where* $\sigma_{\min}^2 := \min_{j\in[d]} \Sigma_{jj}$.

First, observe that the dependence on $1/\sigma_{\min}^2$ is reasonable, as the smaller the minimum variance the more problematic the normalisation matrix $D^{-1/2}$. Fortunately, since we work under the assumption that $\Sigma_{\mathbb{S}} \succeq_{\mathbb{S}} cI_{\mathbb{S}}$, we have that $\sigma_{\min}^2 \geq c$ when we apply this result. Second, observe that if $t > 2\exp\{-n+2d\log 3\}$, i.e. $n \gtrsim d$, the subgaussian regime prevails, and we obtain that

$$\|\widehat{P} - P\|_2 \lesssim \sqrt{\frac{d\log d}{n}}$$

in probability. Similar rates, with logarithmic factors, were found in high-dimensional covariance matrix estimation with missing observations (Lounici, 2014), sample covariance matrix estimator of reduced effective rank population matrices (Bunea and Xiao, 2015), concentration of the adjacency matrix and of the Laplacian in random graphs (Oliveira, 2010), and in the statistical analysis of latent generalized correlation matrix estimation in transelliptical distribution (Han and Liu, 2017). In particular, using the additional assumption that the data is generated according to a transelliptical distribution, Han and Liu (2017) gave an estimator $\widehat{K}$ based on Kendall's tau and proved that

$$\|\widehat{K} - P\|_2 \lesssim \sqrt{\frac{r(P)\log(d)}{n}},$$

where $r(P) := \mathrm{tr}(P)/\|P\|_2$ is the effective dimension of P. This is analogous to the bound given in Proposition 7, where we have an extra factor of $\nu^2/\sigma_{\min}^2$, which can be interpreted as the condition number and might lead to a suboptimal bound when it is large, and the ambient dimension d in place of the intrinsic dimension $r(P)$. This would improve the bound sensibly in the case of an approximately low-rank correlation matrix, but in the worst case the bounds have the same rates.

As well as providing a critical value for our test, Theorem 6 also gives upper bounds on the minimax separation rate for this testing problem. When $c > 0$ and $\nu > 0$ are fixed, our analysis gives an upper bound on the minimax rate of the order

$$2C_{1/4} \lesssim \max_{S\in\mathbb{S}} \sqrt{\frac{|S|\log(|S||\mathbb{S}|)}{n_S}}.$$

whenever $n_S \gtrsim |S|$ for all $S \in \mathbb{S}$. This is our main regime of interest, and we see in our examples in Section 4 below that reliable testing is only possible when sample sizes are large compared with dimensions, up to logarithmic factors.

We now illustrate the behaviour of this bound in certain examples where the expression for C_α can be simplified. The corresponding upper bounds on the minimax separation rate will be complemented by lower bounds in Section 4 to follow.

Example 1. *In the* d*-cycle example, with* $d \geq 3$ *and* $\mathbb{S} = \{\{1,2\},\{2,3\},\ldots,\{d-1,d\},\{d,1\}\}$*, we have* $|S| = 2$ *for all* $S \in \mathbb{S}$ *and* $|\mathbb{S}| = d$ *so that*

$$C_\alpha \lesssim \max_{S\in\mathbb{S}} \sqrt{\frac{\log(d/\alpha)}{n_S}} = \sqrt{\frac{\log(d/\alpha)}{\min_{S\in\mathbb{S}} n_S}}.$$

If $n_S = n$ for all $S \in \mathbb{S}$, this reduces to

$$C_\alpha \lesssim \sqrt{\frac{\log(d/\alpha)}{n}},$$

and by considering the sub-problem of testing the consistency of variances, we will see in Theorem 9 that this upper bound is optimal in this case. Our results reveal that, in this specific example, testing compatibility is no harder than testing consistency, up to constant factors.

Example 2. *Consider the block-3-cycle $\mathbb{S} = \{[2d], [d] \cup ([3d] \setminus [2d]), [3d] \setminus [d]\}$, with $d \geq 1$. Then*

$$C_\alpha \lesssim \max_{S \in \mathbb{S}} \sqrt{\frac{d \log(d/\alpha)}{n_S}} = \sqrt{\frac{d \log(d/\alpha)}{\min_{S \in \mathbb{S}} n_S}}.$$

As before, if $n_S = n$ for all $S \in \mathbb{S}$, this reduces to

$$C_\alpha \lesssim \sqrt{\frac{d \log(d/\alpha)}{n}}.$$

We prove the minimax optimality, up to logarithmic factors, of this rate in Theorem 14. In particular, this shows that the optimal separation rates for this testing problem are not signficantly faster than the optimal rates for the estimation of $\Sigma_{\mathbb{S}}$ with operator norm loss.

Example 3. *Consider the example $\mathbb{S} = \{S_{(-1)}, \cdots, S_{(-d)}\}$, with $d \geq 2$, where $S_{(-i)} = \{1, \ldots, i-1, i+1, \ldots, d\}$. This corresponds to the setting where all observations have a single missing value. In this case, $|S_{(-i)}| = d - 1$, and $|\mathbb{S}| = d$, so that*

$$C_\alpha \lesssim \max_{S \in \mathbb{S}} \sqrt{\frac{d \log(d/\alpha)}{n_S}} = \sqrt{\frac{d \log(d/\alpha)}{\min_{S \in \mathbb{S}} n_S}}.$$

The test we developed in Theorem 6 depends on the unknown quantity $c > 0$, which quantifies the level of positive-definiteness of $\Sigma_{\mathbb{S}}$ by requiring $\Sigma_{\mathbb{S}} \succeq_{\mathbb{S}} cI_{\mathbb{S}}$. As a result, the practical implementation of this method would be difficult. In the remainder of this section we relax this assumption using the well-known technique of sample splitting (e.g. Cox, 1975; Moran, 1973) to calculate a critical value that depends on the data. We randomly divide the data into two non-overlapping sets, $\mathcal{X}_1$ and $\mathcal{X}_2$, and then perform the so-called two steps of *hunting* and *testing*: sample $\mathcal{X}_1$ is used to select a test statistic from a large class of potential test statistics, which is then applied to $\mathcal{X}_2$ to produce the final test statistic. This has the advantage of being easier to calibrate, due to the fact that $\mathcal{X}_1 \perp\!\!\!\perp \mathcal{X}_2$. This approach has been used for a variety of problems, such as testing the location of multiple samples (Cox, 1975), constructing conformal prediction intervals (Lei et al., 2018; Solari and Djordjilović, 2022), goodness-of-fit testing (Janková et al., 2020), conditional mean independence testing (Scheidegger et al., 2022; Lundborg et al., 2022), and conducting inference that is agnostic to the asymptotic regime (Kim and Ramdas, 2024).

In our problem we use sample splitting as follows. Using the data in $\mathcal{X}_1$ we form the sample correlation matrices $\widehat{\Sigma}_{\mathbb{S}}^{(1)}$ then calculate the value of our incompatibility index $R(\widehat{\Sigma}_{\mathbb{S}}^{(1)})$ and retain the optimal primal solution $\widehat{X}_{\mathbb{S}}^{(1)}$, which satisfies $R(\widehat{\Sigma}_{\mathbb{S}}^{(1)}) = -d^{-1}\langle \widehat{X}_{\mathbb{S}}^{(1)}, \widehat{\Sigma}_{\mathbb{S}}^{(1)} \rangle_{\mathbb{S}}$. Now, using the data in $\mathcal{X}_2$ we form the sample correlation matrices $\widehat{\Sigma}_{\mathbb{S}}^{(2)}$ and calculate the statistic

$$\widehat{R}_2 := -\frac{1}{d}\langle \widehat{X}_{\mathbb{S}}^{(1)}, \widehat{\Sigma}_{\mathbb{S}}^{(2)} \rangle_{\mathbb{S}}.$$

We then take our final test statistic to be $V(\widehat{\sigma}_{\mathbb{S}}^2) + \widehat{R}_2$, where $\widehat{\sigma}_{\mathbb{S}}^2$ is calculated using the entire sample. Analytically, the advantage of using this approach is that the statistic $\widehat{R}_2$ is much easier to understand than the statistic $R(\widehat{\Sigma}_{\mathbb{S}})$ used for the oracle test above. Indeed, the independence of $\widehat{X}_{\mathbb{S}}^{(1)}$ and $\widehat{\Sigma}_{\mathbb{S}}^{(2)}$ means that it is enough to give concentration inequalities for linear functionals of $\widehat{\Sigma}_{\mathbb{S}}^{(2)}$, rather than for the supremum $R(\widehat{\Sigma}_{\mathbb{S}})$ over a class of linear functionals. The following result introduces a critical value for this test statistic, which is proved to result in a valid test.

Proposition 8. *Suppose we observe $\mathbf{X}_{S,1}, \ldots, \mathbf{X}_{S,n_S} \stackrel{i.i.d.}{\sim} P_S$ for each $S \in \mathbb{S}$ independently, where each P_S is ν-subgaussian, with the sequence of variances $\sigma_{\mathbb{S}}^2$ satisfying $\bar{\mathrm{av}}_j(\sigma_{\mathbb{S}}^2) = 1$ for all $j \in [d]$, and the sequence of population correlation matrices $\Sigma_{\mathbb{S}}$ satisfying $\Sigma_{\mathbb{S}} \succ_{\mathbb{S}} 0$. Define $C_\alpha(\mathcal{X}_1) = \max\{C_\alpha^{(R)}(\mathcal{X}_1), C_\alpha^{(V)}\}$, with*

$$C_\alpha^{(V)} = C_0 \nu^2 \sqrt{\frac{\log\left(\sum_{j\in[d]} |\mathbb{S}_j|/\alpha\right)}{\min_{S\in\mathbb{S}} n_S}},$$

and

$$C_\alpha^{(R)}(\mathcal{X}_1) = \frac{\|\widehat{X}_{\mathbb{S}}^{(1)}\|_{*,\mathbb{S}}}{d} \max_{S\in\mathbb{S}} \Bigg\{ C_1 \frac{\nu^2}{\sigma_{\min}^2} \sqrt{\frac{|S| + \log(|\mathbb{S}|/\alpha)}{n_S}} \vee \frac{|S| + \log(|\mathbb{S}|/\alpha)}{n_S} + C_2 \frac{\nu^4}{\sigma_{\min}^4} \sqrt{\frac{|S|\log(|S||\mathbb{S}|/\alpha)}{n_S}} + C_3 \frac{\nu^4}{\sigma_{\min}^4} \left(\sqrt{\frac{|S| + \log(|\mathbb{S}|/\alpha)}{n_S}} \vee \frac{|S| + \log(|\mathbb{S}|/\alpha)}{n_S} \right) \sqrt{\frac{|S|\log(|S|/\alpha)}{n_S}} \Bigg\},$$

for some universal constants $C_0, C_1, C_2, C_3 > 0$. Then, for C_0, C_1, C_2, C_3 and universal constant $K > 0$ chosen sufficiently large, for all $\alpha \in (0,1)$ such that $\sigma_{\min}^4 n \geq K\nu^4 \log(\max_{S\in\mathbb{S}} |S|/\alpha)$, the test that rejects H_0 if and only if $V(\widehat{\sigma}_{\mathbb{S}}^2) + \widehat{R}_2 \geq C_\alpha(\mathcal{X}_1)$ has Type I error bounded by α.

This test gives a random threshold $C_\alpha(\mathcal{X}_1)$ which does not depend on c, and it is of the order of

$$C_\alpha(\mathcal{X}_1) \lesssim \frac{\|\widehat{X}_{\mathbb{S}}^{(1)}\|_{*,\mathbb{S}}}{d} \max_{S\in\mathbb{S}} \sqrt{\frac{|S|\log(|S||\mathbb{S}|/\alpha)}{n_S}}$$

whenever $C_\alpha^{(R)} \geq C_\alpha^{(V)}$. Comparing this to the critical value used for the oracle test of Theorem 6, observe that $\|\widehat{X}_{\mathbb{S}}^{(1)}\|_{*,\mathbb{S}}/d$ takes the place of $1/c$. In this regard, notice that for $X_{\mathbb{S}} \in \mathcal{F}_c$ we have

$$\frac{\|X_{\mathbb{S}} + X_{\mathbb{S}}^0\|_{*,\mathbb{S}}}{d} \leq \frac{1}{c},$$

by definition of $\mathcal{F}_c$. This provides some justification for thinking of $\|\widehat{X}_{\mathbb{S}}^{(1)}\|_{*,\mathbb{S}}/d$ as an estimated lower bound on $1/c$ without the need of estimating the spectrum of the Σ_S.

To conclude, we should point out that these procedures suffer from the problem of replicability, due to the fact that $C_\alpha(\mathcal{X}_1)$ is random, implying that, for example, splitting the data in two different ways could potentially lead to contradictory conclusions. Moreover, it is likely that splitting the sample reduces the power of our tests, since we are not making full use of the information in the data. These issues were tackled in Guo and Shah (2023), where rank-transformed subsampling was introduced to increase the power of k-fold sample splitting.

4 Optimality and examples

In this section, we assess the optimality of the oracle test given in Theorem 6, under the settings of Examples 1 and 2, i.e. when $\mathbb{S} = \{\{1,2\},\{2,3\},\ldots,\{d-1,d\},\{d,1\}\}$, the d-cycle, and when $\mathbb{S} = \{[2d],[d]\cup([3d]\setminus[2d]),[3d]\setminus[d]\}$, the block 3-cycle. These collections $\mathbb{S}$ provide examples where our methodology is provably near rate-optimal. For a given dimension d, these two examples further demonstrate the range of optimal rates that can arise for different collections $\mathbb{S}$. Assuming for simplicity that $n_S = n$ for all $S \in \mathbb{S}$, we will see that the optimal rate in the d-cycle case is $\{\log(d)/n\}^{1/2}$, while for the block 3-cycle it is $(d/n)^{1/2}$ up to logarithmic factors. Together, these results show that the structure of $\mathbb{S}$ can have a significant effect on the difficulty of the problem.

We will characterise the optimality of a testing procedure using the minimax framework, where we aim at finding the smallest separation between the null and the alternative hypotheses such that there exists a test that can distinguish between H_0 and H_1 up to a given level of error. More precisely, given $\rho > 0$, we are interested in testing

$$H_0 : R(\Sigma_{\mathbb{S}}) + V(\sigma^2_{\mathbb{S}}) = 0 \quad \text{vs.} \quad H_1 : R(\Sigma_{\mathbb{S}}) + V(\sigma^2_{\mathbb{S}}) > \rho,$$

and our goal is to find the smallest value of ρ such that there exists a test with uniform error control. Write $\Psi \equiv \Psi_{\mathbb{S}}(n_{\mathbb{S}})$ for the set of all tests, that is measurable functions of the data $(X_{S,i} : S \in \mathbb{S}, i \in [n_S])$ taking values in $\{0,1\}$. Write $\mathcal{P}_{\mathbb{S}}(0)$ for the set of all sequences of distributions on $(\mathbb{R}^S : S \in \mathbb{S})$ such that the associated correlation matrices and variances satisfy $R(\Sigma_{\mathbb{S}}) + V(\sigma^2_{\mathbb{S}}) = 0$, and write $\mathcal{P}_{\mathbb{S}}(\rho)$ for the set of all sequences of distributions on $(\mathbb{R}^S : S \in \mathbb{S})$ such that the associated correlation matrices and variances satisfy $R(\Sigma_{\mathbb{S}}) + V(\sigma^2_{\mathbb{S}}) > \rho$. Given a sequence of distributions $P_{\mathbb{S}} = (P_S : S \in \mathbb{S})$ on $(\mathbb{R}^S : S \in \mathbb{S})$ and a sequence of sample sizes $n_{\mathbb{S}} = (n_S : S \in \mathbb{S})$, we write $P_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}$ for the distribution of the entire dataset $(X_{S,i} : S \in \mathbb{S}, i \in [n_S])$ when each observation is independent and $X_{S,i} \sim P_S$ for each $i \in [n_S]$ and $S \in \mathbb{S}$. For a fixed $\eta \in (0,1)$ we may then define the minimax separation to be

$$\rho^* \equiv \rho^*_{\mathbb{S}}(n_{\mathbb{S}},\eta) := \inf\left\{\rho > 0 : \inf_{\varphi\in\Psi}\left(\sup_{P_{\mathbb{S},0}\in\mathcal{P}_{\mathbb{S}}(0)} P_{\mathbb{S},0}^{\otimes n_{\mathbb{S}}}(\varphi = 1) + \sup_{P_{\mathbb{S},1}\in\mathcal{P}_{\mathbb{S}}(\rho)} P_{\mathbb{S},1}^{\otimes n_{\mathbb{S}}}(\varphi = 0)\right) \le \eta\right\}$$

In our analysis we take $\eta = 3/4$, but this is an arbitrary choice and any constant value in $(0,1)$ would result in the same qualitative behaviour. In common with previous work on minimax testing, we prove lower bounds on ρ^* by constructing suitable (prior) distributions μ_0, μ_1 whose support is contained in $\mathcal{P}_{\mathbb{S}}(0), \mathcal{P}_{\mathbb{S}}(\rho)$, respectively. In our proofs it will be sufficient to consider mean-zero Gaussian distributions with suitable priors over their covariance matrices. Having chosen these priors we can bound the minimal error probability by writing

$$\sup_{P_{\mathbb{S},0}\in\mathcal{P}_{\mathbb{S}}(0)} P_{\mathbb{S},0}^{\otimes n_{\mathbb{S}}}(\varphi=1) + \sup_{P_{\mathbb{S},1}\in\mathcal{P}_{\mathbb{S}}(\rho)} P_{\mathbb{S},1}^{\otimes n_{\mathbb{S}}}(\varphi=0) \geq \mathbb{E}_{\mu_0} P_{\mathbb{S},0}^{\otimes n_{\mathbb{S}}}(\varphi=1) + \mathbb{E}_{\mu_1} P_{\mathbb{S},1}^{\otimes n_{\mathbb{S}}}(\varphi=0)$$
$$\geq 1 - \mathrm{TV}\big(\mathbb{E}_{\mu_0} P_{\mathbb{S},0}^{\otimes n_{\mathbb{S}}}, \mathbb{E}_{\mu_1} P_{\mathbb{S},1}^{\otimes n_{\mathbb{S}}}\big),$$

where $\mathbb{E}_{\mu_i} P_{\mathbb{S},i}^{\otimes n_{\mathbb{S}}}$ denotes the mixture distribution of the dataset resulting from generating $P_{\mathbb{S},i} \sim \mu_i$ then, conditionally on $P_{\mathbb{S},i}$, generating the data. Then, the idea behind this method of finding a lower bound on ρ^* is to find priors μ_0, μ_1 with the largest separation possible such that no test can successfully distinguish between $\mathbb{E}_{\mu_0} P_{\mathbb{S},0}^{\otimes n_{\mathbb{S}}}$ and $\mathbb{E}_{\mu_1} P_{\mathbb{S},1}^{\otimes n_{\mathbb{S}}}$.

4.1 Cycles

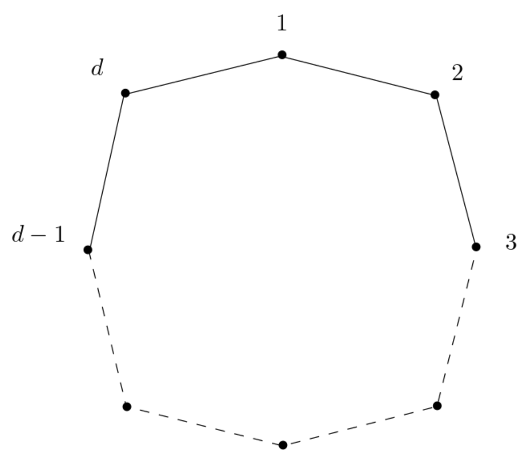

Figure 2: Graph associated to the d-cycle $\mathbb{S}_d = \{\{1,2\},\ldots,\{d,1\}\}$.

Recall that we refer to $\mathbb{S}$ as a d-cycle when $\mathbb{S} = \mathbb{S}_d := \{\{1,2\},\ldots,\{d,1\}\}$. This can be illustrated by Figure 2, where an edge in the graph represents an element of $\mathbb{S}$, so that two nodes are connected if and only if the corresponding variables are simultaneously observed. In this subsection, additions in subscripts of the form $(j, j+1)$ for $j \in [d]$ are intended modulo d, where d in the size of the cycle, so that $(0,1)$ and $(d, d+1)$ are equivalent to $(d,1)$. Our main statistical result in this subsection is following minimax lower bound.

Theorem 9. *Let $\mathbb{S} = \mathbb{S}_d$ for $d \geq 3$, with sample sizes $n_{\mathbb{S}} = (n_1,\ldots,n_d)$. There exists a universal constant $c_1 > 0$ such that*

$$\rho^* \geq c_1 \sqrt{\frac{\log d}{\min_{j\in[d]} n_j}}.$$

This result shows that the power guarantees for the oracle test given in Theorem 6 are optimal up to constants, in the case of a d-cycle. Combined with this upper bound, our construction of the prior distributions μ_0, μ_1 in the proof of Theorem 9 shows that testing the consistency of the variances $\sigma_{\mathbb{S}}^2$, i.e. testing $H_{0,j} : \sigma^2_{\{j-1,j\},j} = \sigma^2_{\{j,j+1\},j}$ for each $j = 1,\ldots,d$, captures the essential statistical difficulty of the problem.

In the remainder of this subsection we explore the properties of $R(\cdot)$ in the cycle example. In particular, we provide explicit expression in simple cases, we discuss the meaning of maximal incompatibility and we prove a result showing that $R(\cdot)$ is bounded below by the maxima of suitable linear functions if $\Sigma_{\mathbb{S}_d}$ is not too singular. Here we write $\Sigma_{\mathbb{S}_d} := (\Sigma_{\{1,2\}}, \cdots, \Sigma_{\{d,1\}})$ for a sequence of 2×2 correlation matrices with

$$\Sigma_{j,j+1} = \begin{pmatrix} 1 & \rho_{j,j+1} \\ \rho_{j,j+1} & 1 \end{pmatrix}.$$

Our next result shows that singular matrices can be removed from $\Sigma_{\mathbb{S}_d}$ when $d \geq 4$, without affecting the value of $R(\cdot)$, reducing the length of the cycle.

Proposition 10. *Fix $d \geq 3$ and $k \geq 1$. Let $\Sigma_{\mathbb{S}_{d+k}}$ be a $(d+k)$-cycle with correlations $(\rho_{\{1,2\}}, \ldots, \rho_{\{d+k,1\}})$ such that $|\rho_{j,j+1}| = 1$ for all $j \in \{d+1, \cdots, d+k\}$ and let $\Sigma_{\mathbb{S}_d}$ represent a d-cycle with correlations $(\bar{\rho}_{\{1,2\}}, \ldots, \bar{\rho}_{\{d,1\}})$ such that $\bar{\rho}_{j,j+1} = \rho_{j,j+1}$, for all $j \in [d-1]$ and*

$$\begin{cases} \bar{\rho}_{d,1} = \rho_{d,d+1} \text{ if } \prod_{j=d+1}^{d+k} \rho_{j,j+1} = 1 \\ \bar{\rho}_{d,1} = -\rho_{d,d+1} \text{ if } \prod_{j=d+1}^{d+k} \rho_{j,j+1} = -1. \end{cases}$$

Then we have $R(\Sigma_{\mathbb{S}_{d+k}}) = R(\Sigma_{\mathbb{S}_d})$.

This reduction applies when the correlations associated to an edge belonging to the path from node $d+1$ to node 1 are either $+1$ or -1. In this setting, we are allowed to identify node 1 with node $d+1$ in such a way that the incompatibility measure of the red d-cycle $\Sigma_{\mathbb{S}_d}$ in Figure 3 is the same as the one of the original $(d+k)$-cycle $\Sigma_{\mathbb{S}_{d+k}}$. This is to be expected, as $\rho_{j,j+1} = \pm 1$ means that variables j and $j+1$ can be identified, up to change in scale, and the dimensionality of the problem can be reduced. Clearly, this result is invariant under cyclic permutations of the nodes' labels.

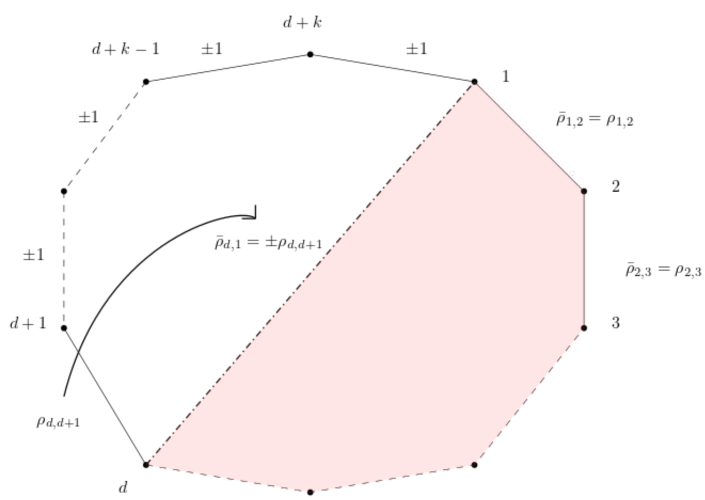

Figure 3: Illustration of Proposition 10. All the correlations associated to an edge belonging to the path from node $d+1$ to node 1 are either $+1$ or -1, and are such that $\prod_{i=d+1}^{d+k} \rho_{i,i+1} = \pm 1$. The new d-cycle $\Sigma_{\mathbb{S}_d}$ in red has the same correlations as the $(d+k)$-cycle $\Sigma_{\mathbb{S}_{d+k}}$, except for $\bar{\rho}_{d,1}$ corresponding to the new edge $\{1, d\}$, which is equal to $\pm\rho_{d,d+1}$.

We now give some explicit expressions for $R(\cdot)$ in special cases and discuss a case for which $R(\Sigma_{\mathbb{S}}) =$

1, meaning that $\Sigma_{\mathbb{S}}$ is maximally incompatible. It will be convenient for the rest of the subsection to reparametrise the correlations as $\rho_j = \cos\theta_j$, with $\theta_j \in [0, \pi]$.

Example 4. *If there are $\theta_1, \theta_2 \in [0,\pi]$ such that $\theta_1 \geq \theta_2$ and*

$$\Sigma_{\mathbb{S}_3} = \left\{ \begin{pmatrix} 1 & \cos\theta_1 \\ \cos\theta_1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & \cos\theta_2 \\ \cos\theta_2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right\},$$

then $R(\Sigma_{\mathbb{S}_3}) = (\cos\theta_2 - \cos\theta_1)/2$. In particular, setting $\theta_2 = 0$ we see that if

$$\Sigma_{\mathbb{S}_3} = \left\{ \begin{pmatrix} 1 & \cos\theta_1 \\ \cos\theta_1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right\},$$

then $R(\Sigma_{\mathbb{S}_3}) = \sin^2(\theta_1/2)$. Moreover, assuming without loss of generality that at most one correlation is negative, as justified in Proposition 11 below, for a general 3-cycle $\Sigma_{\mathbb{S}_3}$ we have $R(\Sigma_{\mathbb{S}_3}) = 1$ if and only if

$$\Sigma_{\mathbb{S}_3} := \left\{ \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right\}.$$

These results can be extended to a general d using Proposition 10.

Related to the last claim of Example 4, another important property for a d-cycle is that we can always assume without loss of generality that at most one θ_i is larger than $\pi/2$, which is equivalent to having at most one negative ρ_i, without changing the value of $R(\cdot)$. This is shown in the following result.

Proposition 11. *Consider the d-cycle with $\mathbb{S}_d = \{\{1,2\},\dots,\{d,1\}\}$ and $\Sigma_{\mathbb{S}_d} := (\Sigma_{\{1,2\}}, \cdots, \Sigma_{\{d,1\}})$, where the correlations $\rho_j = \cos\theta_j$ are uniquely determined by $0 \leq \theta_1, \dots, \theta_d \leq \pi$. Then, there exists another sequence of angles $0 \leq \tilde{\theta}_1, \dots, \tilde{\theta}_d \leq \pi$ with at most one $\tilde{\theta}_i$ larger than $\pi/2$ such that the corresponding d-cycle $\tilde{\Sigma}_{\mathbb{S}_d} := (\tilde{\Sigma}_{\{1,2\}}, \cdots, \tilde{\Sigma}_{\{d,1\}})$ satisfies $R(\tilde{\Sigma}_{\mathbb{S}_d}) = R(\Sigma_{\mathbb{S}_d})$.*

As the proof in Section 6 shows, this new cycle $\tilde{\Sigma}_{\mathbb{S}_d}$ is obtained after changing some X_i into $-X_i$ in such a way that at most $\cos\theta_1$ is negative.

The last result we present on the d-cycle gives an explicit lower bound for R in the case that $\Sigma_{\mathbb{S}_d}$ is incompatible. This is related to the results of Barrett et al. (1993) characterising exactly when the partial correlation matrix

$$\Sigma_{partial} = \begin{pmatrix} 1 & \cos\theta_1 & * & \cdots & \cos\theta_d \\ \cos\theta_1 & 1 & \cos\theta_2 & \cdots & * \\ * & \cos\theta_2 & 1 & \dots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \cos\theta_d & * & * & \cdots & 1 \end{pmatrix}$$

has a positive semi-definite completion. Barrett et al. (1993) shows that this is the case if and only if

$$\sum_{j \in K} \theta_j \leq (|K| - 1)\pi + \sum_{j \notin K} \theta_j \tag{6}$$

for all $K \subseteq [d]$ with $|K|$ odd. Remarkably, this shows that the parametrisation $\rho_{j,j+1} = \cos\theta_j$ allows us to characterise the feasibility of positive semi-definite matrix completion in terms of a finite number of linear inequalities. If we know that $0 \leq \theta_d \leq \theta_{d-1} \leq \ldots \leq \theta_1 \leq \pi$ then this reduces to checking

$$\sum_{j=1}^{k} \theta_j \leq (k-1)\pi + \sum_{j=k+1}^{d} \theta_j$$

for all odd $k \in [d]$. Furthermore, if $0 \leq \theta_1, \ldots, \theta_d \leq \pi$ with at most one θ_j larger than $\pi/2$, then $\Sigma_{partial}$ has a positive semi-definite completion if and only if

$$2 \max_{j \in [d]} \theta_j \leq \sum_{j=1}^{d} \theta_j.$$

Proposition 11 shows that we can always work under this setting, so that the problem of whether $\Sigma_{partial}$ has a PSD completion or not is determined by one condition only, namely $2\max_{j\in[d]}\theta_j \leq \sum_{j=1}^{d}\theta_j$. This is a novel contribution *per se*, since it is not present in Barrett et al. (1993). It is interesting to see that Barrett's characterisation enables us to derive a more explicit expression for $R(\Sigma_{\mathbb{S}_d})$ in the case of a d-cycle.

Proposition 12. *Let Σ be the optimum solution to the dual problem (2), and let $\boldsymbol{\varphi}^* = (\varphi_1^*, \ldots, \varphi_d^*)$ be such that $\Sigma_{j,j+1} = \cos\varphi_j^*$ for each $j \in [d]$. Then:*

(i) $R(\Sigma_{\mathbb{S}_d}) = |\rho_j - (1 - R(\Sigma_{\mathbb{S}_d}))\cos\varphi_j^*|,$ *for all $j \in [d]$;*

(ii) $\boldsymbol{\varphi}^* = (\varphi_1^*, \ldots, \varphi_d^*)$ *is unique, and* $\boldsymbol{\varphi}^*(\theta_1, \ldots, \theta_d)$ *is continuous for varying* $(\theta_1, \ldots, \theta_d) \in [0,\pi]^d$*;*

(iii) if $\theta_1 = \max_{j\in[d]}\theta_j$*, with* $\theta_2, \ldots, \theta_d \leq \pi/2$*, then*

$$1 - R(\Sigma_{\mathbb{S}_d}) = \frac{1 - \epsilon_j \cos\theta_j}{1 - \epsilon_j \cos\varphi_j^*}, \quad \textit{for all } j \in [d],$$

where $\boldsymbol{\epsilon}_d = (\epsilon_1, \ldots, \epsilon_d) = (-1, +\mathbf{1}_{d-1})$*. Also,* $\varphi_1^* = \sum_{j=2}^{d}\varphi_j^*$*.*

Observe that part (ii) only says that the entries of Σ corresponding to the cycle pattern are unique, not the whole Σ itself. Indeed, given the unique optimal $\boldsymbol{\varphi}^*$, there may exist infinitely many positive semi-definite completions. In fact, if A is a partial symmetric matrix admitting a positive semi-definite completion, then there exists a unique positive semi-definite completion with maximum determinant (see Theorem 2 in Grone et al. (1984)). Also, as a sanity check, observe that the optimal choice of the signs in part (iii) makes $\Sigma'_{\mathbb{S}}$ in (3) as incompatible as possible, in accordance with the dual representation given in Proposition 3. For a general sequence of angles $(\theta_1, \ldots, \theta_d)$, it is sufficient to use the transformation given in Proposition 11 to reduce to the case where at most one angle is larger than $\pi/2$, choose $\boldsymbol{\epsilon}_d$ as outlined above, and perform the inverse transformation to obtain the signs for the original $(\theta_1, \ldots, \theta_d)$. As an immediate corollary of this, it is easy to see that, under the same set of hypotheses, we have

$$1 - R(\Sigma_{\mathbb{S}_d}) = \frac{1 + \cos\theta_1}{1 + \cos\varphi_1^*},$$

where φ_1^* is the solution of

$$\begin{cases} \varphi_1^* = \sum_{j=2}^d \varphi_j^* \\ \\ \cos\varphi_j^* = 1 - \frac{1-\cos\theta_j}{1+\cos\theta_1}(1+\cos\varphi_1^*), \qquad \text{for all } j \in \{2,\dots,d\}. \end{cases}$$

This is a relatively explicit expression for $R(\cdot)$ for a general d-cycle.

Our final result in this subsection shows that, provided not too many of our input matrices are close to being singular, $R(\Sigma_{\mathbb{S}_d})$ can be bounded below by a finite maximum of linear functionals that is zero if and only if $\Sigma_{\mathbb{S}_d}$ is compatible. This lower bound constitutes another sanity check for our measure $R(\cdot)$, since the quantities appearing in the lower bound are a natural quantitive version of the qualitative conditions given in Barrett et al. (1993) to check whether the partial matrix $\Sigma_{partial}$ defined above admits a PSD completion.

Proposition 13. *Consider the d-cycle with $\mathbb{S} = \mathbb{S}_d = \{\{1,2\},\dots,\{d,1\}\}$ and suppose that*

$$\Sigma_{\{j,j+1\}} = \begin{pmatrix} 1 & \cos\theta_j \\ \cos\theta_j & 1 \end{pmatrix}.$$

Assume further that there exist $c > 0$ and two indices $k, j \in [d]$ such that $1-\rho_j^2 \geq c, 1-\rho_k^2 \geq c$, so that $\Sigma_{\{j,j+1\}}$ and $\Sigma_{\{k,k+1\}}$ are bounded away from singularity. Then, whenever $\Sigma_{\mathbb{S}_d}$ is incompatible, we have

$$R(\Sigma_{\mathbb{S}_d}) \geq c' \max_{\substack{K\subseteq[d] \\ |K| \ odd}} \left(\sum_{i\in K} \theta_i - (|K|-1)\pi - \sum_{i\in K^c} \theta_i \right),$$

where $c' > 0$ depends only on c.

The proof of Proposition 13 can be found in Section 6. First, observe that this lower bound reduces to

$$c' \left(\theta_1 - \sum_{i=2}^d \theta_i \right)_+,$$

in the case that $\theta_1 = \max_{j\in[d]} \theta_j$ and $\theta_2, \dots, \theta_d \leq \pi/2$, which we have already argued that we may assume without loss of generality. Furthermore, as a sanity check, the simple explicit expressions found in Example 3, in which we have seen that

$$R\left(\left\{ \begin{pmatrix} 1 & \cos\theta_1 \\ \cos\theta_1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & \cos\theta_2 \\ \cos\theta_2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \dots, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right\} \right) = (\cos\theta_2 - \cos\theta_1)/2,$$

is in accordance with Proposition 13, since $\cos\theta_2 - \cos\theta_1 \gtrsim_c \theta_1 - \theta_2$ when θ_1, θ_2 are bounded away from $\{0, \pi\}$.

4.2 Block cycles

So far, we studied with particular care the case of a d-cycle, which is a relatively simple high-dimensional setting, since it is a collection of d two-dimensional distributions. We now describe an evolution of this setting, where we consider a block-matrix version of the 3-cycle. In this case the number of variables per missingness pattern is large and we will see that the minimax separation rates are correspondingly much larger than in the d-cycle, though the number of variables is of the same order.

Theorem 14. *Let $\mathbb{S} = \{[2d], [d] \cup ([3d] \setminus [2d]), [3d] \setminus [d]\}$ for some $d \geq 1$. Writing $n_{\mathbb{S}} = (n_1, n_2, n_3)$ for the sample sizes within each pattern, there exists a universal constant $c_1 > 0$ such that*

$$\rho^* \geq c_1 \sqrt{\frac{d}{(n_1 \wedge n_2) \log^4(ed)}}$$

whenever $n_1 \wedge n_2 \geq d/2$.

This result shows that, up to logarithmic factors in d, the minimax separation rates for this testing problem are the same as the minimax estimation rates for estimating $\Sigma_{\mathbb{S}}$ in the operator norm distance. This is related to the fact that $R(\Sigma_{\mathbb{S}})$ is a non-smooth functional of $\Sigma_{\mathbb{S}}$. Indeed, the following result shows that we can construct examples of $\Sigma_{\mathbb{S}}$ such that $R(\Sigma_{\mathbb{S}})$ can be bounded below using the function $x \mapsto \max(0, x)$; see below for more discussion of the relevant literature.

Proposition 15. *Consider the set of patterns $\mathbb{S} = \{[2d], [d] \cup ([3d] \setminus [2d]), [3d] \setminus [d]\}$ for some $d \geq 1$, and suppose that*

$$\Sigma_{\mathbb{S}} = \left\{ \begin{pmatrix} I_d & P \\ P^T & I_d \end{pmatrix}, \begin{pmatrix} I_d & -P \\ -P^T & I_d \end{pmatrix}, \begin{pmatrix} I_d & \beta I_d \\ \beta I_d & I_d \end{pmatrix} \right\},$$

for some $P \in \mathbb{R}^{d \times d}$ such that $\|P\|_2 \leq 1$ and some $\beta \in [0, 1]$. Then:

(i) $R(\Sigma_{\mathbb{S}}) = 0$ if and only if $\|P\|_2^2 \leq \frac{1-\beta}{2}$,

(ii) $R(\Sigma_{\mathbb{S}}) \geq \frac{3}{4d} \sum_{j=1}^{d} (\sigma_j^2(P) - \frac{1-\beta}{2})_+$, where $\sigma_j(P)$ is the j-th singular value of P.

This shows that, for $\Sigma_{\mathbb{S}}$ of the form above, we can relate our testing problem to the problem of testing whether the vector of squared singular values of P belongs to the orthant $(-\infty, 0]^d$, or is separated from it in the ℓ_1 distance. In a Gaussian location model a similar problem, measuring separation with the ℓ_2 distance, was considered by Blanchard et al. (2018), and part of our lower bound construction is inspired by this work. However, the consideration of singular values of matrices rather than Gaussian means means that new technical tools are required. In this regard, the techniques of Thépaut and Verzelen (2021), who consider the estimation of quantities of the form $\sum_{j=1}^{d} \sigma_j(P)^q$ for $q > 0$, are useful. We also mention that such problem are related to the estimation of ℓ_1 distances, for which good references include Cai and Low (2011) and Jiao et al. (2016).

5 Numerical studies

The tests introduced in Theorem 6 and Proposition 8 give finite-sample Type-I error control over $\mathcal{P}_{\mathbb{S}}^0$, the parameter space associated to the null hypothesis. However, these procedures can be too conservative in some examples of interest, and have the further downside of depending on the unknown subgaussian variance proxy ν. For these reasons, we propose Monte-Carlo versions of our tests, that can be applied without any knowledge of unknown parameters, and we compare it with Little's test (Little, 1988). In Section 5.1 we introduce a bootstrap method that checks the incompability of sample correlation matrices, while in Section 5.2 we extend this method to also check the consistency of samples means and variances.

Little's test can be applied when all pairs of variables are observed together, so that the EM algorithm (Dempster et al., 1977) can be applied to find estimators $\widehat{\mu}$ and $\widehat{\Lambda}$ of the mean and covariance matrix of the data under the null hypothesis of MCAR. Little's test is a generalised likelihood ratio test whose validity is is based on the assumption that the data $(X_{S,i} : S \in \mathbb{S}, i \in [n_S])$ are Gaussian. Writing $\tilde{\Lambda} = n_{\text{tot}}\widehat{\Lambda}/(n_{\text{tot}}-1)$, where $n_{\text{tot}} = \sum_{S\in\mathbb{S}} n_S$, define

$$d^2 = \sum_{S\in\mathbb{S}} n_S(\bar{X}_S - \widehat{\mu}_{|S})\tilde{\Lambda}_{|S}^{-1}(\bar{X}_S - \widehat{\mu}_{|S}),$$

$$d^2_{\text{cov}} = \sum_{S\in\mathbb{S}} n_S[\text{tr}(\widehat{\Sigma}_S\widehat{\Lambda}_{|S}^{-1}) - |S| - \log|\widehat{\Sigma}_S| + \log|\widehat{\Lambda}_{|S}|],$$

and

$$d^2_{\text{aug}} = d^2 + d^2_{\text{cov}}.$$

Then, under MCAR, d^2_{aug} converges in law to a χ^2-distribution with

$$f = \sum_{S\in\mathbb{S}} \frac{1}{2}|S|(|S|+3) - \frac{1}{2}d(d+3),$$

degrees of freedom by Wilks' theorem. Based on these asymptotic results, Little's test rejects MCAR if and only if $d^2_{\text{aug}} > \chi^2_f(1-\alpha)$, where $\chi^2_f(1-\alpha)$ is such that $\mathbb{P}(W \geq \chi^2_f(1-\alpha)) = \alpha$, and where W is χ^2-distributed with f degrees of freedom. Similarly, one can define a test based on d^2_{cov} that ignores the means and only considers the partial covariance matrices, which converges to a χ^2-distribution with

$$f' = \sum_{S\in\mathbb{S}} \frac{1}{2}|S|(|S|+1) - \frac{1}{2}d(d+1)$$

degrees of freedom by Wilks' theorem. In the next section, we will compare our bootstrap method with these two versions of Little's test, one based on d^2_{aug}, the other on d^2_{cov}. This is because in Section 5.1 departures from the null are due to an incompatible sequence of correlation matrices, while the means are assumed to be constant, hence we do not want to give ourselves a clear advantage by comparing only with d^2_{aug}.

5.1 Correlation matrices and simulations for d-cycles

In this section, we design a bootstrap version of our test, and compare it with Little's test in detecting departures from MCAR due to an incompatible sequence of correlation matrices $\Sigma_{\mathbb{S}}$. Recall that we write $\widehat{\Sigma}_{\mathbb{S}} = (\widehat{\Sigma}_S)_{S\in\mathbb{S}}$, where $\widehat{\Sigma}_S$ is the sample correlation matrix of the data $X_S := (X_{S,i}, i \in [n_S])$ for $S \in \mathbb{S}$. From Proposition 3 we can write $\widehat{\Sigma}_{\mathbb{S}} = (1 - R(\widehat{\Sigma}_{\mathbb{S}}))\widehat{Q}_{\mathbb{S}} + R(\widehat{\Sigma}_{\mathbb{S}})\widehat{\Sigma}'_{\mathbb{S}}$, where $\widehat{Q}_{\mathbb{S}}$ can be thought as the closest compatible sequence of correlation matrices to $\widehat{\Sigma}_{\mathbb{S}}$, and can be computed at the same time as the test statistic $R^{(0)} := R(\widehat{\Sigma}_{\mathbb{S}})$. We then transform the original data by calculating $\tilde{X}_S := \widehat{Q}_S^{1/2}\widehat{\Sigma}_S^{-1/2}\,\text{scale}(X_S)$ for all $S \in \mathbb{S}$, where scale$(\cdot)$ represents the step of standardising the original data. This transformation means that the sample correlation matrices of $\tilde{X}_{\mathbb{S}} := (\tilde{X}_S : S \in \mathbb{S})$ are compatible. Fixing $B \in \mathbb{N}$, for each $b \in [B]$ and $S \in \mathbb{S}$ we generate $\tilde{X}_S^{(b)}$ as a nonparametric bootstrap sample from $\tilde{X}_S$ and calculate the sample correlation matrix $\widehat{\Sigma}_{S,b} = \mathrm{Cor}(\tilde{X}_S^{(b)})$. Then, for each $b \in [B]$ we compute the corresponding test statistic $R^{(b)} := R(\widehat{\Sigma}_{S,b} : S \in \mathbb{S})$. Finally, we reject H_0 at a significance level $\alpha \in (0,1)$ if and only if $1 + \sum_{i=1}^{B} \mathbb{1}\{R^{(b)} \geq R^{(0)}\} \geq \alpha(1+B)$.

We compare this test with Little's procedure in the settings given in Theorem 9, namely in the case of a d-cycle. For our first settings, we set $n_{\mathbb{S}} = (n_S)_{S\in\mathbb{S}} = (200, \ldots, 200)$, and simulate $X_{\{j,j+1\},i} \overset{\text{i.i.d.}}{\sim} N(\mathbf{0}_2, \Sigma_{\{j,j+1\}})$ for $i \in [200]$ and $j \in [d]$, where

$$\Sigma_{\mathbb{S}_d} = \left\{ \begin{pmatrix} 1 & \cos\theta_1 \\ \cos\theta_1 & 1 \end{pmatrix}, \ldots, \begin{pmatrix} 1 & \cos\theta_d \\ \cos\theta_d & 1 \end{pmatrix} \right\},$$

for certain values of $\theta_1, \ldots, \theta_d \in [0, \pi]$, and compare our bootstrap test with Little's procedure. Here, we repeat the the experiment $M = 200$ times, and report the average decision as an estimate of the power function. This makes sense only for $d = 3$, while for $d \geq 4$ there exists at least one pair of variables that are never observed together, making the EM algorithm to estimate $\widehat{\Lambda}$ inapplicable. As for the case $d = 3$, Figure 4 shows two different simulations, with different values of $(\theta_1, \theta_2, \theta_3)$, where our test performs very similarly to Little's tests. As stated above, for $d \geq 4$ Little's test cannot be applied, while our test remains

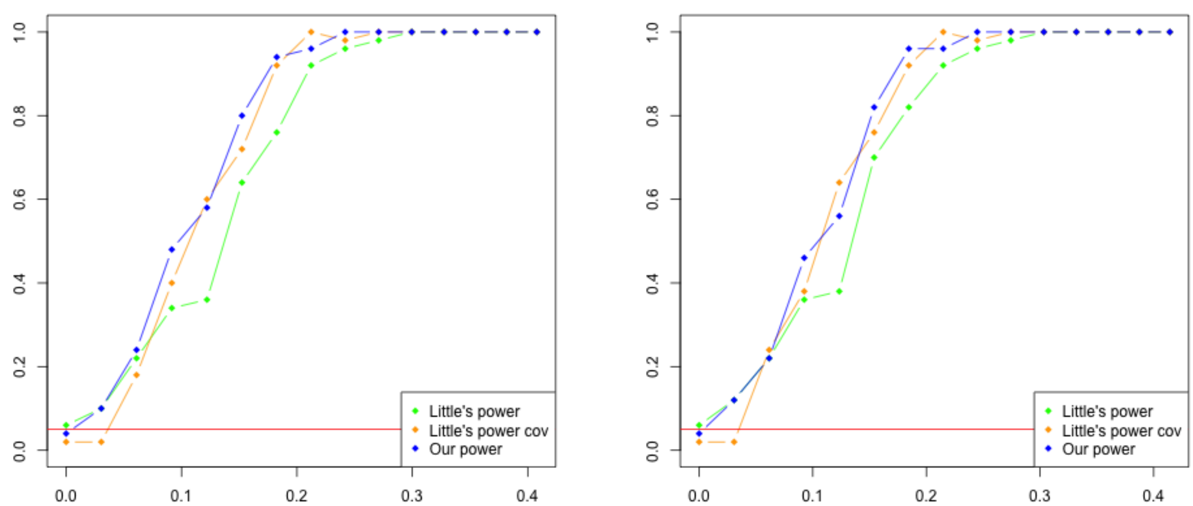

Figure 4: Simulation of the power functions of our method with $B = 99$ (blue), Little's method based on d^2_{aug} (green), and Little's method based on d^2_{cov} (orange), with Gaussian data. In each example, we vary $\theta_1 \in [\theta_2+\theta_3, (\theta_2+\theta_3+\pi)/2]$, with (θ_2, θ_3) equal to $(\pi/3, \pi/6)$ (left), $(\pi/4, \pi/4)$ (right). For each of this setting, we repeat the experiment $M = 200$ times, and report the average decision. The nominal level $\alpha = 0.05$ in red.

valid since there are no constraints on $\mathbb{S}$. In Figure 5 below, we show the power function of our bootstrap test in the case of a d-cycle, with $d \in \{100, 200\}$, with $\theta_2 = \ldots = \theta_d = \frac{\pi}{2(d-1)}$, and varying θ_1 in $[\pi/2, 5\pi/8]$. We repeat the procedure $M = 100$ times, and report the average decision as an estimate of the power function.

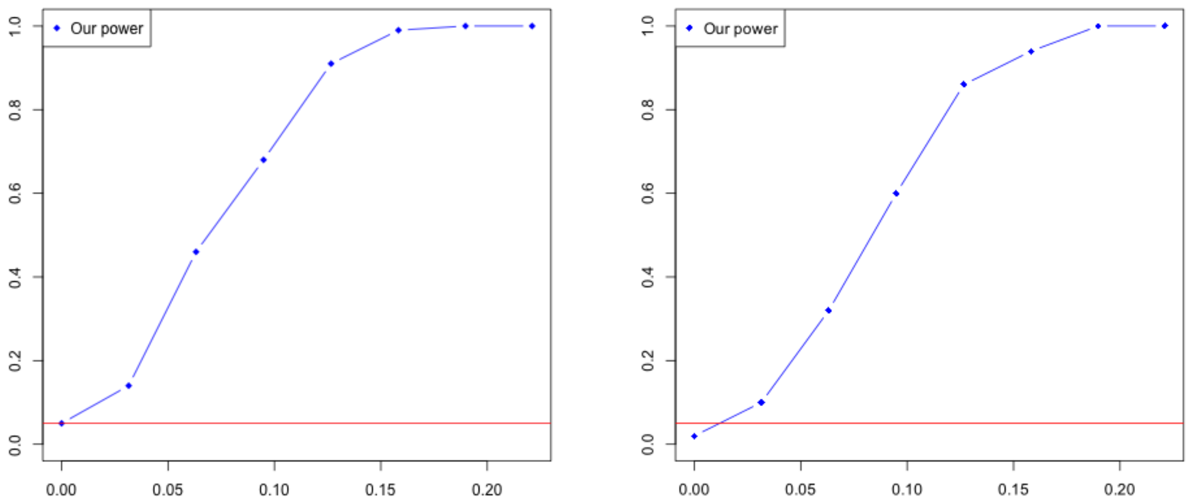

Figure 5: Simulation of the power functions of our method with $B = 99$(blue) for $d = 100$ (left), and $d = 200$ (right), with Gaussian data. In each example, we fix $\theta_2 = \ldots = \theta_d = \frac{\pi}{2(d-1)}$, and vary θ_1 in $[\pi/2, 5\pi/8]$. For each of this setting, we repeat the experiment $M = 100$ times, and report the average decision. The nominal level $\alpha = 0.05$ in red.

Our simulations so far have used Gaussian data, so that Little's test is valid. We now repeat our simulations with a heavy-tailed data distribution in order to assess the robustness of the methods. To this aim, we consider again a 3-cycle, and generate $X_{\{j,j+1\},i} \overset{\text{i.i.d.}}{\sim} \log N(\mathbf{0}_2, \Sigma_{\{j,j+1\}})$ for all $i \in [200], j \in [3]$, where $\log N(\mathbf{0}_2, \Sigma_{\{j,j+1\}})$ stands for the log-normal distribution, meaning that if $Y \sim \log N(\mathbf{0}_2, \Sigma_{\{j,j+1\}})$ then $Y_i = e^{Z_i}$, with $Z \sim N(\mathbf{0}_2, \Sigma_{\{j,j+1\}})$. Figure 6 below shows the analogue of Figure 4, in the sense that the parameters $(\theta_1, \theta_2, \theta_3)$ are the same, but we generated artificial data from a multivariate log-normal distribution rather than a Gaussian distribution. It is interesting to see that Little's test does not have Type-I error control. On the other hand, our test succeeds in controlling the Type-I error and, although being conservative, its power increases as $\Sigma_{\mathbb{S}_3}$ becomes more incompatible.

5.2 Omnibus approach

We now aim at designing a test that is able to detect departures from MCAR due to inconsistent means and variances as well as incompatible correlation matrices. To this aim, we define a new bootstrap test that checks both compatibility of $\Sigma_{\mathbb{S}}$, and consistency of $\sigma^2_{\mathbb{S}}$ and $\mu_{\mathbb{S}}$, the sequence of means along the different patterns in $\mathbb{S}$. We will use as test statistics estimators of $T = R(\Sigma_{\mathbb{S}}) + V(\sigma^2_{\mathbb{S}}) + M(\mu_{\mathbb{S}})$, where

$$M(\mu_{\mathbb{S}}) = \max_{j \in [d]} \max_{S_1, S_2 \in \mathbb{S}_j} |\mu_{S_1,j} - \mu_{S_2,j}|.$$

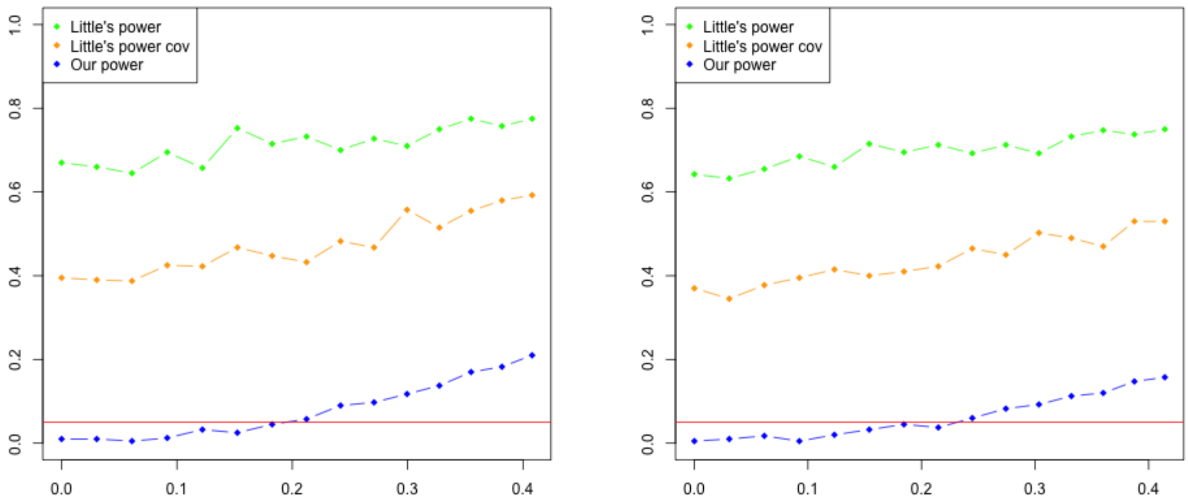

Figure 6: Simulation of the power functions of our method with $B = 99$(blue), Little's method based on d^2_{aug} (green), and Little's method based on d^2_{cov} (orange), with log-normal data. In each example, we vary $\theta_1 \in [\theta_2 + \theta_3, (\theta_2 + \theta_3 + \pi)/2]$, with (θ_2, θ_3) equal to $(\pi/3, \pi/6)$ (left), $(\pi/4, \pi/4)$ (right). For each of this setting, we repeat the experiment $M = 200$ times, and report the average decision. The nominal level $\alpha = 0.05$ in red.

Observe that $M(\mu_{\mathbb{S}}) = 0$ if and only if $\mu_{\mathbb{S}}$ is consistent. This results in Algorithm 1, which is implemented in the R-package `MCARtest` (Berrett et al., 2022).

Algorithm 1 MCAR bootstrap test checking compatibility of correlation matrices, and consistency of means and variances

---

Given data $X_{\mathbb{S}}$, compute $\widehat{\sigma}^2_{\mathbb{S}} = \text{Var}\, X_{\mathbb{S}}$, and rescale it such that $\text{av}_j(\widehat{\sigma}^2_{\mathbb{S}}) = 1$ for all $j \in [d]$; i.e. replace X_j with $X_j/\sqrt{\text{av}_j(\widehat{\sigma}^2_{\mathbb{S}})}$ for all $j \in [d]$.

Compute $\widehat{\mu}_{\mathbb{S}} = \mathbb{E}X_{\mathbb{S}}$, $\widehat{M}_{\mathbb{S}} = \text{Cov}\, X_{\mathbb{S}} = \text{diag}(\sigma^2_{\mathbb{S}})^{1/2} \cdot \Sigma_{\mathbb{S}} \cdot \text{diag}(\sigma^2_{\mathbb{S}})^{1/2}$.

Compute $T^{(0)} = R(\widehat{\Sigma}_{\mathbb{S}}) + V(\widehat{\sigma}^2_{\mathbb{S}}) + M(\widehat{\mu}_{\mathbb{S}})$, and compute at the same time the dual decomposition $\widehat{\Sigma}_{\mathbb{S}} = (1 - R(\widehat{\Sigma}_{\mathbb{S}}))\widehat{Q}_{\mathbb{S}} + R(\widehat{\Sigma}_{\mathbb{S}})\widehat{\Sigma}'_{\mathbb{S}}$.

Rotate the original data $X_{\mathbb{S}}$, i.e. for all $S \in \mathbb{S}$, for all $i \in [n_S]$ **do** $\tilde{X}_{S,i} = \widehat{Q}_S^{1/2}\widehat{M}_S^{-1/2}(X_{S,i} - \widehat{\mu}_S + \widehat{\mu}_{|S})$, where $\widehat{\mu}_j = |\mathbb{S}_j|^{-1}\sum_{S \in \mathbb{S}_j} \mu_{S,j}$.

for $b \in [B]$ **do**

  For all $S \in \mathbb{S}$, let $\tilde{X}^{(b)}_{S,i}$ be a nonparametric bootstrap sample from $\tilde{X}_{S,i}$, for $i \in [n_S]$.

  Compute $\widehat{\mu}_{\mathbb{S},b} = \mathbb{E}X^{(b)}_{\mathbb{S}}$, $\widehat{M}_{\mathbb{S},b} = \text{Cov}\, X^{(b)}_{\mathbb{S}} = \text{diag}(\widehat{\sigma}^2_{\mathbb{S},b})^{1/2} \cdot \widehat{\Sigma}_{\mathbb{S},b} \cdot \text{diag}(\widehat{\sigma}^2_{\mathbb{S},b})^{1/2}$.

  Compute $T^{(b)} = R(\widehat{\Sigma}_{\mathbb{S},b}) + V(\widehat{\sigma}^2_{\mathbb{S},b}) + M(\widehat{\mu}_{\mathbb{S},b})$.

end for

Reject H_0 if and only if $1 + \sum_{i=1}^{B} \mathbb{1}\{T^{(b)} \geq T^{(0)}\} \geq \alpha(1 + B)$.

As before, we compare this algorithm with Little's test based on d^2_{aug}. In this section we generate complete artificial data according to various distributions, and then delete entries using the R package `missMethods`

(Rockel, 2020). MCAR data are generated with the function `delete_MCAR`, where each entry of the data matrix is deleted independently of the others with probability $p \in (0, 1)$. Deviations from the null are generated by partitioning the columns in two groups, group A where the missing values are generated, and group B which determines the missingness mechanism, with two different mechanisms being considered. First, `delete_MAR_1_to_x` sets threshold values, splits the rows into two further groups depending on whether columns in group B have values greater or smaller than the threshold, and deletes some entries in columns in group A in a such a way that the probability for a value to be missing in group A divided by the probability for a value to be missing in group B equals 1 divided by `x`, with `x` to be specified as an input parameter. Second, `delete_MAR_rank` deletes each entry in a column of group A with probability proportional to the rank of the same row in the corresponding column of group B. For further details on these functions, and other methods to generate MCAR, MAR, MNAR data, refer to Santos et al. (2019). These three functions were also chosen in the numerical analysis of a test of MCAR based on U-statistics in Aleksić (2023).

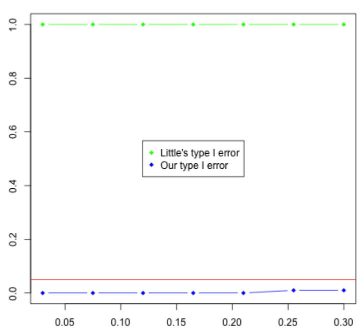

Figure 7: Type-I error under MCAR data generate with `delete_MCAR(p)`, for varying probability p of having a missing value. Data from Clayton copula with parameter 1 and log-normal margins. $B = 99, M = 300$.

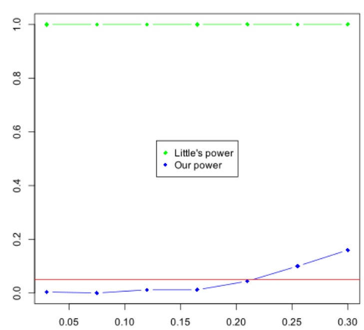

Figure 8: Power function under MAR data generate with `delete_MAR_1_to_x(p, x = 9)`, for varying probability p of having a missing value. Data from the same Clayton coupla. $B = 99, M = 300$.

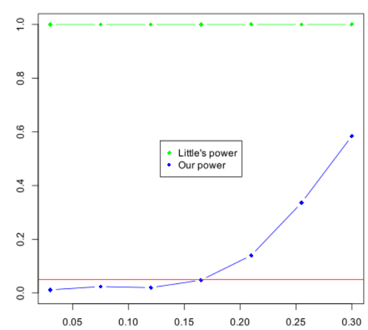

Figure 9: Power function under MAR data generate with `delete_MAR_rank(p)`, for varying probability p of having a missing value. Data from the same Clayton coupla. $B = 99, M = 300$.

For Figures 7, 8, 9, we generated 5-dimensional datasets of sample size $n = 1000$ distributed according to a Clayton copula, with parameter 1 and log-normal margins, using the function `mvdc` from the R-package `copula` Hofert et al. (2020). For Figure 7 we deleted the first two variables with `delete_MCAR(p)` for different values of $p \in \{0.03, 0.06, \ldots, 0.3\}$, in order to get an artificial setting coming from the null. For each p, we repeat the simulation 300 times, and report the average Type-I error. Alternatives to the null were generated using `delete_MAR_1_to_x`, with $x = 9$, for Figure 8, and `delete_MAR_rank` for Figure 9. Again, for each p, we repeat the simulations 300 times, and report the average power. The simulations show that Little's test is not able to recognise MCAR in this setting, and rejects the null hypothesis with high probability. On the other hand, Algorithm 1 has good control of the Type-I error, although being a little conservative, and its power increases as the missingness probability p gets bigger and the effective sample sizes for the incomplete patterns increase. Our test performs slightly better in the case of a 3-dimensional dataset of sample size $n = 1000$ distributed according to a Clayton copula, with parameter 1 and chi-squared margins. The results of these simulations are shown in Figures 10, 11, 12. In this case, we delete the first two columns, while the third one is always complete. Here our method retains Type I error control and is more powerful than in the previous settings, while Little's test does not have good control of the Type I error.

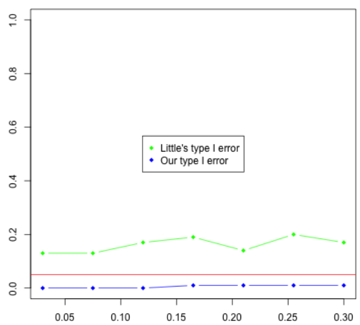

Figure 10: Type-I error under MCAR data generate with `delete_MCAR(p)`, for varying probability p of having a missing value. Data from Clayton copula with parameter 1 and chi-squared margins. $B = 99, M = 300$.

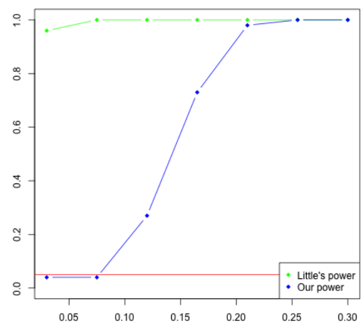

Figure 11: Power function under MAR data generate with `delete_MAR_1_to_x(p, x = 9)`, for varying probability p of having a missing value. Data from the same Clayton coupla. $B = 99, M = 300$.

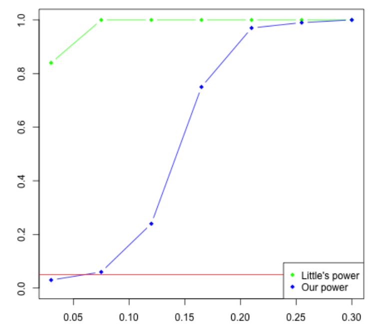

Figure 12: Power function under MAR data generate with `delete_MAR_rank(p,` for varying probability p of having a missing value. Data from the same Clayton coupla. $B = 99, M = 300$.

6 Proofs

6.1 Proofs for Section 2

Proof of Proposition 1. For any $X \in \mathcal{M}$ and $X_{\mathbb{S}} \in \mathcal{M}_{\mathbb{S}}$ we have

$$\begin{aligned}\langle AX, X_{\mathbb{S}} \rangle_{\mathbb{S}} &= \sum_{S \in \mathbb{S}} \sum_{j,j' \in S} ((AX)_S)_{jj'} (X_S)_{jj'} = \sum_{S \in \mathbb{S}} \sum_{j,j' \in S} X_{jj'} (X_S)_{jj'} \\ &= \sum_{j,j'=1}^{d} X_{jj'} \sum_{S \in \mathbb{S}} \mathbb{1}_{j,j' \in S} (X_S)_{jj'} = \langle X, A^* X_{\mathbb{S}} \rangle,\end{aligned}$$

as claimed. □

Proof of Proposition 2. The strategy is to use a semi-definite programming version of Farkas' lemma. This is well known in the relevant literature, but we provide a statement and short proof for completeness; see Proposition 25 in Appendix C. First, rewrite the matrix completion problem

$$\text{find } \Sigma \in \mathcal{M} \text{ such that } \begin{cases} \Sigma_{jj'} = (\Sigma_S)_{jj'}, \forall S \in \mathbb{S}_{jj'} \\ \Sigma \succcurlyeq 0 \end{cases}$$

as

$$\text{find } \Sigma \in \mathcal{M} \text{ such that } \begin{cases} \langle \Sigma, E_{jj'} \rangle = (\Sigma_S)_{jj'}, \forall S \in \mathbb{S}_{jj'} \\ \Sigma \succcurlyeq 0 \end{cases} \tag{7}$$

where $E_{jj'} = (\boldsymbol{e}_j \boldsymbol{e}_{j'}^T + \boldsymbol{e}_{j'} \boldsymbol{e}_j^T)/2$ and $\boldsymbol{e}_j$ is the j-th column vector of the standard orthonormal basis of $\mathbb{R}^d$. In order to apply the semi-definite version of Farkas' lemma we transform our problem so that the equality constraints have zero on the right-hand side. To this end, define

$$H_j := \begin{pmatrix} 0 & \boldsymbol{e}_j^T/2 \\ \boldsymbol{e}_j/2 & \boldsymbol{O} \end{pmatrix} \text{ and } G_{S,jj'} := \begin{pmatrix} -(\Sigma_S)_{jj'} & \boldsymbol{0}^T \\ \boldsymbol{0} & E_{jj'} \end{pmatrix},$$

and consider the completion problem

$$\text{find } \tilde{\Sigma} \in \mathcal{M} \text{ such that } \begin{cases} \langle \tilde{\Sigma}, H_j \rangle = 0, \forall j \in [d] \\ \langle \tilde{\Sigma}, G_{S,jj'} \rangle = 0, \forall S \in \mathbb{S}_{jj'} \\ \tilde{\Sigma} \succcurlyeq 0. \end{cases} \tag{8}$$

The condition $\langle \tilde{\Sigma}, H_j \rangle = 0, \forall j \in [d]$ forces $\tilde{\Sigma}$ to be in block diagonal form

$$\tilde{\Sigma} := \begin{pmatrix} \gamma_{0,0} & \boldsymbol{0}^T \\ \boldsymbol{0} & \Sigma \end{pmatrix}.$$

Now, observe that (7) has a solution if and only if (8) has a non-zero solution. Indeed, for every solution Σ_0 of (7), then $\operatorname{diag}(1, \Sigma_0)$ is a solution of (8). On the other hand, suppose that $\tilde{\Sigma}_0 = \operatorname{diag}(\gamma_{0,0}, \Sigma_0) \neq \boldsymbol{O}$ is a solution of (8). This implies that $\gamma_{0,0} \neq 0$, otherwise $0 = \langle \tilde{\Sigma}_0, G_{S,jj'} \rangle = -\gamma_{0,0}(\Sigma_S)_{jj'} + \Sigma_{jj'} = \Sigma_{jj'}$, which would imply $\tilde{\Sigma}_0 = \boldsymbol{O}$. Being $\gamma_{0,0} \neq 0$, we can rescale the bigger block in $\tilde{\Sigma}_0$ by $\gamma_{0,0}$, i.e. $\Sigma_0 =: \gamma_{0,0} G$, and get $0 = \langle \tilde{\Sigma}_0, G_{S,jj'} \rangle = -\gamma_{0,0}(\Sigma_S)_{jj'} + \gamma_{0,0} G_{jj'} = -(\Sigma_S)_{jj'} + G_{jj'}$, which shows that G is a solution of (7). This further implies that we can assume without loss of generality that $\gamma_{0,0} = 1$ when (8) admits a non-zero solution. Now, by Proposition 25, we know that (8) has a non-zero solution $\tilde{\Sigma} = \operatorname{diag}(1, \Sigma)$ if and only if

$$\sum_{S \in \mathbb{S}} \sum_{j,j' \in S} (X_S)_{jj'} G_{S,jj'} = \sum_{S \in \mathbb{S}} \begin{pmatrix} -\langle \Sigma_S, X_S \rangle & \boldsymbol{0}^T \\ \boldsymbol{0} & \frac{1}{2} X_S \end{pmatrix} = \begin{pmatrix} -\langle \Sigma_{\mathbb{S}}, X_{\mathbb{S}} \rangle & \boldsymbol{0}^T \\ \boldsymbol{0} & \frac{1}{2} A^* X_{\mathbb{S}} \end{pmatrix} \not\succ 0,$$

for all sequences of matrices $X_{\mathbb{S}}$, not necessarily PSD. Now, this block matrix is positive definite if and only if both $A^* X_{\mathbb{S}} \succ 0$ and $\langle \Sigma_{\mathbb{S}}, X_{\mathbb{S}} \rangle < 0$. Hence, (8) has a non-zero solution if and only if $\langle \Sigma_{\mathbb{S}}, X_{\mathbb{S}} \rangle \geq 0$ for all $X_{\mathbb{S}}$ such that $A^* X_{\mathbb{S}} \succ 0$, and the claim follows. □

Proof of Proposition 3. Weak duality, i.e. LHS ≤ RHS, always holds for SDPs (see Blekherman et al. (2012)), but we include a short proof for the sake of completeness. In fact, for any $\Sigma_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}$, we can rewrite

$$\inf\{\epsilon \in [0,1] : \Sigma_{\mathbb{S}} \in (1-\epsilon)\mathcal{P}_{\mathbb{S}}^0 + \epsilon \mathcal{P}_{\mathbb{S}}\} \tag{9}$$

as

$$\inf\{\epsilon \in [0,1] : \Sigma_{\mathbb{S}} \in (1-\epsilon)\mathcal{P}_{\mathbb{S}}^0 + \epsilon \mathcal{P}_{\mathbb{S}}\} = 1 - \sup\{\epsilon \in [0,1] : \Sigma_{\mathbb{S}} \in \epsilon \mathcal{P}_{\mathbb{S}}^0 + (1-\epsilon)\mathcal{P}_{\mathbb{S}}\}$$
$$= 1 - \frac{1}{d}\sup\{\mathrm{tr}(\Sigma) : \Sigma \in \mathcal{P}^*, \Sigma_{\mathbb{S}} - A\Sigma \succeq_{\mathbb{S}} 0, \Sigma_{11} = \ldots = \Sigma_{dd}\}.$$

Now, for any $Y_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^*$ such that $A^*Y_{\mathbb{S}} + Y \succeq I_d$ for some $Y \in \mathcal{Y}$, and any $\Sigma \in \mathcal{P}^*$ such that $\Sigma_{\mathbb{S}} - A\Sigma \succeq_{\mathbb{S}} 0$, we have

$$\mathrm{tr}(\Sigma) = \langle I_d, \Sigma\rangle = -\langle A^*Y_{\mathbb{S}} + Y - I_d, \Sigma\rangle + \langle A^*Y_{\mathbb{S}} + Y, \Sigma\rangle \leq \langle A^*Y_{\mathbb{S}}, \Sigma\rangle + \langle Y, \Sigma\rangle$$
$$= \langle A^*Y_{\mathbb{S}}, \Sigma\rangle = \langle Y_{\mathbb{S}}, A\Sigma\rangle_{\mathbb{S}} = \langle Y_{\mathbb{S}}, \Sigma_{\mathbb{S}}\rangle_{\mathbb{S}} - \langle Y_{\mathbb{S}}, \Sigma_{\mathbb{S}} - A\Sigma\rangle_{\mathbb{S}} \leq \langle Y_{\mathbb{S}}, \Sigma_{\mathbb{S}}\rangle_{\mathbb{S}}.$$

This shows that (9) is lower bounded by

$$1 - \frac{1}{d}\inf\{\langle Y_{\mathbb{S}}, \Sigma_{\mathbb{S}}\rangle_{\mathbb{S}} : Y_{\mathbb{S}} \in \mathcal{P}_{\mathbb{S}}^*, A^*Y_{\mathbb{S}} + Y \succeq I_d\}. \tag{10}$$

Weak duality follows upon noting that $A^*X_{\mathbb{S}}^0 = I_d$ and $\langle X_{\mathbb{S}}^0, \Sigma_{\mathbb{S}}\rangle_{\mathbb{S}} = d$ and setting $X_{\mathbb{S}} = Y_{\mathbb{S}} - X_{\mathbb{S}}^0$. This is not surprising, as we already mentioned that weak duality always holds for SDP problems.

We will now prove strong duality for this problem. Our strategy is to write our primal and dual problems in standard form and check Slater's condition for the primal problem (10). We already mentioned that (9) can be written as

$$1 - \frac{1}{d}\sup\{\mathrm{tr}(\Sigma) : \Sigma \in \mathcal{P}^*, \Sigma_{11} = \ldots = \Sigma_{dd}, \Sigma_{\mathbb{S}} - A\Sigma \succeq_{\mathbb{S}} 0\}.$$

We now write this maximisation problem in standard form by introducing variables $(Z_S : S \in \mathbb{S}) = \Sigma_{\mathbb{S}} - A\Sigma \in \mathcal{P}_{\mathbb{S}}^*$. Enumerating $\mathbb{S}$ as $\{S_1, \ldots, S_m\}$, we instead optimise over block-diagonal matrices of the form

$$X = \begin{pmatrix} \Sigma & 0 & \cdots & 0 \\ 0 & Z_{S_1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Z_{S_m} \end{pmatrix}$$

For such X our constraints are equivalent to $X \succeq 0$,

$$\langle E_{jj} - E_{11}, X\rangle = 0 \quad \text{for } j = 2, \ldots, d$$

and

$$\langle E_{jj'} + E_{S,jj'}, X\rangle = (\Sigma_S)_{jj'} \quad \text{for } S \in \mathbb{S} \text{ and } j, j' \in S,$$

where $E_{jj'} = (\boldsymbol{e}_j\boldsymbol{e}_{j'}^T + \boldsymbol{e}_{j'}\boldsymbol{e}_j^T)/2$ is the binary symmetric matrix of the same dimension as X with its only non-zero entries being in the (j,j')-th and (j',j)-th positions of the top left block, and where $E_{S,jj'} = (\boldsymbol{e}_{S,j}\boldsymbol{e}_{S,j'}^T + \boldsymbol{e}_{S,j'}\boldsymbol{e}_{S,j}^T)/2$ is the binary symmetric matrix of the same dimension as X with its only non-zero entries being in the (j,j')-th and (j',j)-th positions of the block occupied by Z_S in X. Write C for the diagonal matrix of the same dimension as X with I_d in the top left block, and all other entries equal to zero.

It is now possible to write

$$\begin{aligned}
&\sup\{\operatorname{tr}(\Sigma) : \Sigma \in \mathcal{P}^*, \quad \Sigma_{11} = \ldots = \Sigma_{dd}, \quad \Sigma_{\mathbb{S}} - A\Sigma \succeq_{\mathbb{S}} 0\} \\
&\quad = \sup\{\langle C, X\rangle : X \succeq 0 \text{ is block diagonal}, \langle E_{jj} - E_{11}, X\rangle = 0 \text{ for } j = 2, \ldots, d, \\
&\qquad\qquad \langle E_{jj'} + E_{S,jj'}, X\rangle = (\Sigma_S)_{jj'} \quad \text{for } S \in \mathbb{S} \text{ and } j, j' \in S\}, \qquad (11)
\end{aligned}$$

so that our dual problem (9) is now in standard form. Our primal problem (10) is put into standard form by writing

$$\begin{aligned}
&\inf\{\langle \Sigma_{\mathbb{S}}, Y_{\mathbb{S}}\rangle : A^* Y_{\mathbb{S}} + Y \succeq I_d, \quad Y_{\mathbb{S}} \succeq_{\mathbb{S}} 0, \quad Y_{\mathbb{S}} \in \mathcal{M}_{\mathbb{S}}, \quad Y \in \mathcal{Y}\} \\
&\quad = \inf\Bigg\{\sum_{S\in\mathbb{S}}\sum_{j,j'\in S} (\Sigma_S)_{jj'} y_{S,jj'} : \sum_{S\in\mathbb{S}}\sum_{j,j'\in S} y_{S,jj'} E_{jj'} + \sum_{j=2}^{d} y_{jj}(E_{jj} - E_{11}) \succeq I_d, \\
&\qquad\qquad \sum_{S\in\mathbb{S}}\sum_{j,j'\in S} y_{S,jj'} E_{S,jj'} \succeq 0 \text{ for all } S \in \mathbb{S}, \quad y_{jj}, y_{S,jj'} \in \mathbb{R} \text{ for all } S, j, j'\Bigg\}. \qquad (12)
\end{aligned}$$

With the problems written in standard form, it is now clear that (11) is the dual problem associated to (12); see Theorem 3.1 in Vandenberghe and Boyd (1996). Observe further that the primal problem is strictly feasible since $Y_{\mathbb{S}} = X^0_{\mathbb{S}}$ satisfies the linear constraints with Y equal to the zero matrix. Hence, by standard duality results (Theorem 2.15 in Blekherman et al. (2012), Theorem 3.1 in Vandenberghe and Boyd (1996)), we have that

$$\begin{aligned}
&\sup\{\langle C, X\rangle : X \succeq 0 \text{ is block diagonal}, \langle E_{jj} - E_{11}, X\rangle = 0 \text{ for } j = 2, \ldots, d, \\
&\qquad\qquad \langle E_{jj'} + E_{S,jj'}, X\rangle = (\Sigma_S)_{jj'} \quad \text{for } S \in \mathbb{S} \text{ and } j, j' \in S\} \\
&= \inf\Bigg\{\sum_{S\in\mathbb{S}}\sum_{j,j'\in S} (\Sigma_S)_{jj'} y_{S,jj'} : \sum_{S\in\mathbb{S}}\sum_{j,j'\in S} y_{S,jj'} E_{jj'} + \sum_{j=2}^{d} y_{jj}(E_{jj} - E_{11}) \succeq I_d, \\
&\qquad\qquad \sum_{S\in\mathbb{S}}\sum_{j,j'\in S} y_{S,jj'} E_{S,jj'} \succeq 0 \text{ for all } S \in \mathbb{S}, \quad y_{jj}, y_{S,jj'} \in \mathbb{R} \text{ for all } S, j, j'\Bigg\},
\end{aligned}$$

and the result follows. □

Proof of Proposition 4. (i) Convexity follows easily from basic properties of the supremum. Indeed, consider $\tilde{\Sigma}_{\mathbb{S}} := \lambda\Sigma^{(1)}_{\mathbb{S}} + (1-\lambda)\Sigma^{(2)}_{\mathbb{S}}$ with $\lambda \in [0, 1]$. Observe that R is well defined at $\tilde{\Sigma}_{\mathbb{S}}$, as the convex combination of correlation matrices is still a correlation matrix. Then, for all $\lambda \in [0, 1]$,

$$\begin{aligned}
R(\tilde{\Sigma}_{\mathbb{S}}) &= \sup\Big\{-\frac{1}{d}\langle X_{\mathbb{S}}, \tilde{\Sigma}_{\mathbb{S}}\rangle : X_{\mathbb{S}} + X^0_{\mathbb{S}} \succeq_{\mathbb{S}} 0, A^* X_{\mathbb{S}} + Y \succeq 0 \text{ for some } Y \in \mathcal{Y}\Big\} \\
&= \sup\Big\{-\frac{1}{d}\langle X_{\mathbb{S}}, \lambda\Sigma^{(1)}_{\mathbb{S}} + (1-\lambda)\Sigma^{(2)}_{\mathbb{S}}\rangle : X_{\mathbb{S}} + X^0_{\mathbb{S}} \succeq_{\mathbb{S}} 0, A^* X_{\mathbb{S}} + Y \succeq 0 \text{ for some } Y \in \mathcal{Y}\Big\} \\
&\leq \lambda \sup\Big\{-\frac{1}{d}\langle X_{\mathbb{S}}, \Sigma^{(1)}_{\mathbb{S}}\rangle : X_{\mathbb{S}} + X^0_{\mathbb{S}} \succeq_{\mathbb{S}} 0, A^* X_{\mathbb{S}} + Y \succeq 0 \text{ for some } Y \in \mathcal{Y}\Big\} \\
&\qquad + (1-\lambda) \sup\Big\{-\frac{1}{d}\langle X_{\mathbb{S}}, \Sigma^{(2)}_{\mathbb{S}}\rangle : X_{\mathbb{S}} + X^0_{\mathbb{S}} \succeq_{\mathbb{S}} 0, A^* X_{\mathbb{S}} + Y \succeq 0 \text{ for some } Y \in \mathcal{Y}\Big\}
\end{aligned}$$

$$= \lambda R(\Sigma_{\mathbb{S}}^{(1)}) + (1-\lambda) R(\Sigma_{\mathbb{S}}^{(2)}),$$

and the convexity of $R(\cdot)$ follows.

(ii) R acts on $\mathcal{P}_{\mathbb{S}}$, which is the space of correlation matrices over the patterns $\mathbb{S}$. Now, the spectrahedron of all correlation matrices of dimension p,

$$\mathcal{E}_p = \left\{ \left(x_1, \ldots, x_{\binom{p}{2}}\right) \in \mathbb{R}^{\binom{p}{2}} : \Sigma_{\boldsymbol{x}} = \begin{pmatrix} 1 & x_1 & \cdots & x_{p-1} \\ x_1 & 1 & \cdots & x_{2p-3} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p-1} & x_{2p-3} & \cdots & 1 \end{pmatrix} \succeq 0 \right\},$$

is called the *elliptope*, and identifies a closed subset of $\mathbb{R}^{\binom{p}{2}}$. This follows from the fact that the symmetry condition $\Sigma_{\boldsymbol{x}} = \Sigma_{\boldsymbol{x}}^T$ defines a linear subspace of $\mathbb{R}^p$ of dimension $\binom{p}{2}$, while the PSD condition $v^T \Sigma_{\boldsymbol{x}} v \geq 0$ for all $v \in \mathbb{R}^p$ defines a closed subset of $\mathbb{R}^{\binom{p}{2}}$, which is a convex cone. For further insights, refer to Laurent and Poljak (1996). This implies that, for every pattern $\mathbb{S}$, $\mathcal{P}_{\mathbb{S}}$ can be identified with a closed subspace of $\mathbb{R}^s$, where $s = \sum_{S \in \mathbb{S}} \binom{|S|}{2}$. The continuity of R follows from the fact that every convex function that is finite on $\mathbb{R}^s$ is necessarily continuous (see Corollary 10.1.1. in Rockafellar (1970)).

To prove (iii), we will make use of the fact that the dual characterisation allows us to express $R(\Sigma_{\mathbb{S}'})$ as

$$1 - \frac{1}{d'} \sup\{\operatorname{tr}(\Sigma) : \Sigma \in \mathcal{P}^*, \Sigma_{11} = \ldots = \Sigma_{d'd'}, \Sigma_{\mathbb{S}'} - A_{\mathbb{S}'}\Sigma \succeq_{\mathbb{S}'} 0\},$$

where $d' = |\cup_{S \in \mathbb{S}'} S|$. Now, let $\tilde{\Sigma}$ be an optimal feasible matrix for $\Sigma_{\mathbb{S}'}$, where all the diagonal elements of $\tilde{\Sigma}$ are the equal to each other by definition of R. Then, if we consider the restriction of $\tilde{\Sigma}$ on $\cup_{S \in \mathbb{S}} S$, call it $\tilde{\Sigma}_{|\mathbb{S}}$, it is clear that $\Sigma_{\mathbb{S}} - A_{\mathbb{S}} \tilde{\Sigma}_{|\mathbb{S}} \succeq_{\mathbb{S}} 0$, since $\Sigma_{\mathbb{S}'} - A_{\mathbb{S}'}\Sigma \succeq_{\mathbb{S}'} 0$ and $\Sigma_{\mathbb{S}} \subseteq \Sigma_{\mathbb{S}'}$ by hypothesis, while $\tilde{\Sigma}_{|\mathbb{S}} \succeq 0$ follows again by Cauchy's interlacing theorem. Hence, calling $d = |\cup_{S \in \mathbb{S}} S|$, for every $\tilde{\Sigma}$ that is optimal for $\Sigma_{\mathbb{S}'}$, we can construct a feasible $\tilde{\Sigma}_{|\mathbb{S}}$ for $\Sigma_{\mathbb{S}}$ such that $1 - \operatorname{tr}(\tilde{\Sigma}_{|\mathbb{S}})/d = R(\Sigma_{\mathbb{S}'})$, which completes the proof. □

Proof of Proposition 5. Let $\sigma_{\mathbb{S}}^2$ be a nonnegative sequence such that $\bar{\text{av}}_j(\sigma_{\mathbb{S}}^2) = 1$ for all $j \in [d]$. Using, for the third equality, the facts that $A_V \mathbf{1}_d$ also satisfies these properties and that $\bar{\text{av}}$ is linear, we have that

$$\begin{aligned} 1 - V(\sigma_{\mathbb{S}}^2) &= 1 - \inf\left\{\epsilon \in [0,1] : \sigma_{\mathbb{S}}^2 = (1-\epsilon)A_V \mathbf{1}_d + \epsilon {\sigma'}_{\mathbb{S}}^2 \text{ with } \bar{\text{av}}_j({\sigma'}_{\mathbb{S}}^2) = 1 \text{ for all } j \in [d]\right\} \\ &= \sup\left\{\epsilon \in [0,1] : \sigma_{\mathbb{S}}^2 = \epsilon A_V \mathbf{1}_d + (1-\epsilon){\sigma'}_{\mathbb{S}}^2 \text{ with } \bar{\text{av}}_j({\sigma'}_{\mathbb{S}}^2) = 1 \text{ for all } j \in [d]\right\} \\ &= \sup\left\{\epsilon \in [0,1] : \epsilon \leq \min_{j \in [d]} \min_{S \in \mathbb{S}_j} \sigma_{S,j}^2\right\} = \min_{j \in [d]} \min_{S \in \mathbb{S}_j} \sigma_{S,j}^2, \end{aligned}$$

as claimed. □

6.2 Proofs for Section 3

Proof of Theorem 6. We are interested in finding $C_\alpha \in (0,1)$ such that $\forall \alpha \in (0,1)$

$$\mathbb{P}_{H_0}\left(\widehat{T} \geq C_\alpha\right) = \mathbb{P}_{H_0}\left(R(\widehat{\Sigma}_{\mathbb{S}}) + V(\widehat{\sigma}_{\mathbb{S}}^2) \geq C_\alpha\right) \leq \alpha.$$

First, observe that

$$\mathbb{P}_{H_0}\left(R(\widehat{\Sigma}_{\mathbb{S}}) + V(\widehat{\sigma}^2_{\mathbb{S}}) \geq C_\alpha\right) \leq \mathbb{P}_{H_0}\left(R(\widehat{\Sigma}_{\mathbb{S}}) \geq C_\alpha/2\right) + \mathbb{P}_{H_0}\left(V(\widehat{\sigma}^2_{\mathbb{S}}) \geq C_\alpha/2\right),$$

and let us concentrate on the first term. For simplicity we replace our assumption that $\Sigma_{\mathbb{S}} \succeq_{\mathbb{S}} cI_{\mathbb{S}}$ by the assumption that $\Sigma_{\mathbb{S}} \succeq_{\mathbb{S}} 2cI_{\mathbb{S}}$. We have

$$\mathbb{P}_{H_0}\left(R(\widehat{\Sigma}_{\mathbb{S}}) \geq C_\alpha/2\right) \leq \mathbb{P}_{H_0}\left(R(\widehat{\Sigma}_{\mathbb{S}}) \geq C_\alpha/2,\, \widehat{\Sigma}_{\mathbb{S}} \succeq_{\mathbb{S}} cI_{\mathbb{S}}\right) + 1 - \mathbb{P}_{H_0}\left(\widehat{\Sigma}_{\mathbb{S}} \succeq_{\mathbb{S}} cI_{\mathbb{S}}\right). \tag{13}$$

Since $\Sigma_{\mathbb{S}} \succeq_{\mathbb{S}} 2cI_{\mathbb{S}}$ by assumption, we may bound the second part of (13) by writing

$$\begin{aligned}
1 = \mathbb{P}_{H_0}\left(\Sigma_{\mathbb{S}} \succeq_{\mathbb{S}} 2cI_{\mathbb{S}}\right) &= \mathbb{P}_{H_0}\left(\Sigma_{\mathbb{S}} - \widehat{\Sigma}_{\mathbb{S}} + \widehat{\Sigma}_{\mathbb{S}} \succeq_{\mathbb{S}} 2cI_{\mathbb{S}}\right) \\
&\leq \mathbb{P}_{H_0}\left(\Sigma_{\mathbb{S}} - \widehat{\Sigma}_{\mathbb{S}} \succeq_{\mathbb{S}} cI_{\mathbb{S}}\right) + \mathbb{P}_{H_0}\left(\widehat{\Sigma}_{\mathbb{S}} \succeq_{\mathbb{S}} cI_{\mathbb{S}}\right) \\
&\leq \mathbb{P}_{H_0}\left(\|\widehat{\Sigma}_{\mathbb{S}} - \Sigma_{\mathbb{S}}\|_{2,\mathbb{S}} \geq c\right) + \mathbb{P}_{H_0}\left(\widehat{\Sigma}_{\mathbb{S}} \succeq_{\mathbb{S}} cI_{\mathbb{S}}\right).
\end{aligned}$$

This implies that

$$1 - \mathbb{P}_{H_0}\left(\widehat{\Sigma}_{\mathbb{S}} \succeq_{\mathbb{S}} cI_{\mathbb{S}}\right) \leq \mathbb{P}_{H_0}\left(\|\widehat{\Sigma}_{\mathbb{S}} - \Sigma_{\mathbb{S}}\|_{2,\mathbb{S}} \geq c\right).$$

Now, define

$$\bar{\operatorname{tr}}(X_{\mathbb{S}}) = \sum_{j=1}^{d} |\mathbb{S}_j|^{-1} \sum_{S \in \mathbb{S}_j} (X_S)_{jj}$$

and observe that, for all $X_{\mathbb{S}} \in \mathcal{M}_{\mathbb{S}}$, we have $\langle X^0_{\mathbb{S}}, X_{\mathbb{S}}\rangle_{\mathbb{S}} = \bar{\operatorname{tr}}(X_{\mathbb{S}})$. See Proposition 20 in Appendix B for a proof of this fact. Using the arguments leading up to (4) above, the first term on the right-hand side of (13) can be written as

$$\mathbb{P}_{H_0}\left(R(\widehat{\Sigma}_{\mathbb{S}}) \geq C_\alpha/2,\, \widehat{\Sigma}_{\mathbb{S}} \succeq_{\mathbb{S}} cI_{\mathbb{S}}\right) = \mathbb{P}_{H_0}\left(\sup_{X_{\mathbb{S}} \in \mathcal{F}_c} -\frac{1}{d}\langle X_{\mathbb{S}}, \widehat{\Sigma}_{\mathbb{S}}\rangle_{\mathbb{S}} \geq C_\alpha/2,\, \widehat{\Sigma}_{\mathbb{S}} \succeq_{\mathbb{S}} cI_{\mathbb{S}}\right),$$

where $\mathcal{F}_c = \{X_{\mathbb{S}} + X^0_{\mathbb{S}} \succeq_{\mathbb{S}} 0, A^*X_{\mathbb{S}} + Y \succeq 0 \text{ for some } Y \in \mathcal{Y}, \langle X_{\mathbb{S}} + X^0_{\mathbb{S}}, 2cI_{\mathbb{S}}\rangle_{\mathbb{S}} \leq d\}$. Discarding the condition $A^*X_{\mathbb{S}} + Y \succeq 0$ for some $Y \in \mathcal{Y}$ and enlarging our feasible to $\tilde{\mathcal{F}}_c := \{X_{\mathbb{S}} + X^0_{\mathbb{S}} \succeq_{\mathbb{S}} 0, \langle X_{\mathbb{S}} + X^0_{\mathbb{S}}, 2cI_{\mathbb{S}}\rangle_{\mathbb{S}} \leq d\}$, we have

$$\begin{aligned}
\mathbb{P}_{H_0}\left(R(\widehat{\Sigma}_{\mathbb{S}}) \geq C_\alpha/2,\, \widehat{\Sigma}_{\mathbb{S}} \succeq_{\mathbb{S}} cI_{\mathbb{S}}\right) &= \mathbb{P}_{H_0}\left(\widehat{R} - R \geq C_\alpha/2,\, \widehat{\Sigma}_{\mathbb{S}} \succeq_{\mathbb{S}} cI_{\mathbb{S}}\right) \leq \mathbb{P}_{H_0}\left(|\widehat{R} - R| \geq C_\alpha/2,\, \widehat{\Sigma}_{\mathbb{S}} \succeq_{\mathbb{S}} cI_{\mathbb{S}}\right) \\
&\leq \mathbb{P}_{H_0}\left(\sup_{X_{\mathbb{S}} \in \tilde{\mathcal{F}}_c} \left|-\frac{1}{d}\langle X_{\mathbb{S}}, \widehat{\Sigma}_{\mathbb{S}} - \Sigma_{\mathbb{S}}\rangle_{\mathbb{S}}\right| \geq C_\alpha/2\right) \\
&= \mathbb{P}_{H_0}\left(\sup_{X_{\mathbb{S}} \in \tilde{\mathcal{F}}_c} \left|\langle X_{\mathbb{S}} + X^0_{\mathbb{S}}, \widehat{\Sigma}_{\mathbb{S}} - \Sigma_{\mathbb{S}}\rangle_{\mathbb{S}} - (\bar{\operatorname{tr}}(\widehat{\Sigma}_{\mathbb{S}}) - d)\right| \geq d \cdot C_\alpha/2\right) \\
&\leq \mathbb{P}_{H_0}\left(\|\widehat{\Sigma}_{\mathbb{S}} - \Sigma_{\mathbb{S}}\|_{2,\mathbb{S}} \cdot \sup_{X_{\mathbb{S}} \in \tilde{\mathcal{F}}_c} \|X_{\mathbb{S}} + X^0_{\mathbb{S}}\|_{*,\mathbb{S}} \geq d \cdot C_\alpha/2\right) \\
&\leq \mathbb{P}_{H_0}\left(\|\widehat{\Sigma}_{\mathbb{S}} - \Sigma_{\mathbb{S}}\|_{2,\mathbb{S}} \cdot d/c \geq d \cdot C_\alpha/2\right)
\end{aligned}$$

$$= \mathbb{P}_{H_0}\left(\|\widehat{\Sigma}_{\mathbb{S}} - \Sigma_{\mathbb{S}}\|_{2,\mathbb{S}} \geq c \cdot C_\alpha/2\right),$$

where we used Holder's inequality for sequences of matrices, and the fact that $\bar{\text{tr}}(\widehat{\Sigma}_{\mathbb{S}}) = d$, since $\widehat{\Sigma}_{\mathbb{S}}$ is a sequence of sample correlation matrices. Putting all the pieces together, we have

$$\begin{aligned}
\mathbb{P}_{H_0}\left(\widehat{T} \geq C_\alpha\right) &\leq \mathbb{P}_{H_0}\left(\|\widehat{\Sigma}_{\mathbb{S}} - \Sigma_{\mathbb{S}}\|_{2,\mathbb{S}} \geq c \cdot C_\alpha/2\right) + \mathbb{P}_{H_0}\left(\|\widehat{\Sigma}_{\mathbb{S}} - \Sigma_{\mathbb{S}}\|_{2,\mathbb{S}} \geq c\right) + \mathbb{P}_{H_0}\left(V(\widehat{\sigma}^2_{\mathbb{S}}) \geq C_\alpha/2\right) \\
&\leq 2\mathbb{P}_{H_0}\left(\|\widehat{\Sigma}_{\mathbb{S}} - \Sigma_{\mathbb{S}}\|_{2,\mathbb{S}} \geq c \cdot C_\alpha/2\right) + \mathbb{P}_{H_0}\left(V(\widehat{\sigma}^2_{\mathbb{S}}) \geq C_\alpha/2\right),
\end{aligned}$$

since $\mathbb{P}(X \geq x_1) + \mathbb{P}(X \geq x_2) \leq 2\mathbb{P}(X \geq \min\{x_1, x_2\})$. Hence, in order to bound this probability above by α, it is sufficient to find C_α such that

$$\max\left\{\mathbb{P}_{H_0}\left(\|\widehat{\Sigma}_{\mathbb{S}} - \Sigma_{\mathbb{S}}\|_{2,\mathbb{S}} \geq c \cdot C_\alpha/2\right), \mathbb{P}_{H_0}\left(V(\widehat{\sigma}^2_{\mathbb{S}}) \geq C_\alpha/2\right)\right\} \leq \alpha/3. \tag{14}$$

As for the first term inside the maximum, we have

$$\begin{aligned}
\mathbb{P}_{H_0}\left(\|\widehat{\Sigma}_{\mathbb{S}} - \Sigma_{\mathbb{S}}\|_{2,\mathbb{S}} \geq c \cdot C_\alpha/2\right) &\leq \mathbb{P}_{H_0}\left(\max_{S \in \mathbb{S}} \|\widehat{\Sigma}_S - \Sigma_S\|_2 \geq c \cdot C_\alpha/2\right) \\
&\leq \sum_{S \in \mathbb{S}} \mathbb{P}_{H_0}\left(\|\widehat{\Sigma}_S - \Sigma_S\|_2 \geq c \cdot C_\alpha/2\right) \leq |\mathbb{S}| \cdot \max_{S \in \mathbb{S}} \mathbb{P}_{H_0}\left(\|\widehat{\Sigma}_S - \Sigma_S\|_2 \geq c \cdot C_\alpha/2\right).
\end{aligned}$$

Hence, calling

$$\begin{aligned}
C_t(S) := C_1 \frac{\nu^2}{\sigma^2_{\min}}\left(\sqrt{\frac{|S| + \log(1/t)}{n}} \vee \frac{|S| + \log(1/t)}{n}\right) + C_2 \frac{\nu^4}{\sigma^4_{\min}}\sqrt{\frac{|S|\log(|S|/t)}{n}} \\
+ C_3 \frac{\nu^4}{\sigma^4_{\min}}\left(\sqrt{\frac{|S| + \log(1/t)}{n}} \vee \frac{|S| + \log(1/t)}{n}\right)\sqrt{\frac{|S|\log(|S|/t)}{n}}
\end{aligned}$$

for all $S \in \mathbb{S}$, with $C_1, C_2, C_3 > 0$ sufficiently big universal constants, it is immediate to see using Proposition 7 that it is sufficient to take

$$C_\alpha \geq \frac{2}{c} \max_{S \in \mathbb{S}} C_{\alpha/3|\mathbb{S}|}(S), \tag{15}$$

in order to have $\mathbb{P}_{H_0}\left(\|\widehat{\Sigma}_{\mathbb{S}} - \Sigma_{\mathbb{S}}\|_{2,\mathbb{S}} \geq c \cdot C_\alpha/2\right) \leq \alpha/3$. As for the second term in the maximum in (14), since $V(\sigma^2_{\mathbb{S}}) = 0$ under the null,

$$\begin{aligned}
\mathbb{P}_{H_0}\left(V(\widehat{\sigma}^2_{\mathbb{S}}) \geq C_\alpha/2\right) &= \mathbb{P}_{H_0}\left(V(\widehat{\sigma}^2_{\mathbb{S}}) - V(\sigma^2_{\mathbb{S}}) \geq C_\alpha/2\right) \leq \mathbb{P}_{H_0}\left(|V(\widehat{\sigma}^2_{\mathbb{S}}) - V(\sigma^2_{\mathbb{S}})| \geq C_\alpha/2\right) \\
&= \mathbb{P}_{H_0}\left(|\min_{j \in [d]} \min_{S \in \mathbb{S}_j} \widehat{\sigma}^2_{S,j} - \min_{j \in [d]} \min_{S \in \mathbb{S}_j} \sigma^2_{S,j}| \geq C_\alpha/2\right) \\
&\leq \mathbb{P}_{H_0}\left(\max_{j \in [d]} \max_{S \in \mathbb{S}_j} |\widehat{\sigma}^2_{S,j} - \sigma^2_{S,j}| \geq C_\alpha/2\right) \\
&\leq \sum_{j \in [d]} \sum_{S \in \mathbb{S}_j} \mathbb{P}_{H_0}\left(|\widehat{\sigma}^2_{S,j} - \sigma^2_{S,j}| \geq C_\alpha/2\right) \\
&\leq \left(\sum_{j \in [d]} |\mathbb{S}_j|\right) \max_{j \in [d]} \max_{S \in \mathbb{S}_j} \mathbb{P}_{H_0}\left(|\widehat{\sigma}^2_{S,j} - \sigma^2_{S,j}| \geq C_\alpha/2\right).
\end{aligned}$$

Now, the standard Chernoff method gives, for all $j \in [d]$, for all $S \in \mathbb{S}_j$,

$$\mathbb{P}_{H_0}\left(|\widehat{\sigma}^2_{S,j} - \sigma^2_{S,j}| \geq C_\alpha/2\right) \leq \exp\left\{-\frac{n_S C_\alpha^2}{256\nu^4}\right\} \leq \exp\left\{-\frac{\min_{S\in\mathbb{S}} n_S C_\alpha^2}{256\nu^4}\right\},$$

so that $\mathbb{P}_{H_0}\left(V(\widehat{\sigma}^2_{\mathbb{S}}) \geq C_\alpha/2\right) \leq \alpha/3$ is satisfied if

$$\left(\sum_{j\in[d]} |\mathbb{S}_j|\right) \exp\left\{-\frac{\min_{S\in\mathbb{S}} n_S C_\alpha^2}{256\nu^4}\right\} \leq \frac{\alpha}{3}.$$

Hence, it is sufficient to take

$$C_\alpha \geq 16\nu^2 \sqrt{\frac{\log\left(3\sum_{j\in[d]} |\mathbb{S}_j|/\alpha\right)}{\min_{S\in\mathbb{S}} n_S}}. \tag{16}$$

In order to satisfy both (15) and (16) at the same time, it is sufficient to take the maximum between the two right-hand sides, and the statement follows. □

Proof of Proposition 7. First, observe that a generic element of $\widehat{P} - P$, can be written as

$$\widehat{P}_{ij} - P_{ij} = \frac{\widehat{\sigma}_{ij}}{\widehat{\sigma}_i \widehat{\sigma}_j} - \frac{\sigma_{ij}}{\sigma_i\sigma_j} = \left(\frac{1}{\widehat{\sigma}_i\widehat{\sigma}_j} - \frac{1}{\sigma_i\sigma_j}\right)\widehat{\sigma}_{ij} + \frac{\widehat{\sigma}_{ij} - \sigma_{ij}}{\sigma_i\sigma_j}.$$

This implies that $\widehat{P} - P$ can be written as

$$\widehat{P} - P = W \circ \widehat{\Sigma} + \tilde{W} \circ (\widehat{\Sigma} - \Sigma),$$

where $\circ$ stands for the matrix pointwise product, also known as Hadamard product, and $W, \tilde{W}$ satisfy $W_{ij} = 1/\widehat{\sigma}_i\widehat{\sigma}_j - 1/\sigma_i\sigma_j$ and $\tilde{W}_{ij} = 1/\sigma_i\sigma_j$. We will now bound the operator norm of this difference using the following facts about Hadamard products. First, as shown in (3.7.12) of Johnson (1989), if $A, B \in \mathcal{M}$ and $A \succeq 0$ then we have

$$\|A \circ B\|_2 \leq \max_{j\in[d]} |A_{jj}| \|B\|_2.$$

Second, for arbitrary $A, B \in \mathcal{M}$ and $v = \sum_{i=1}^d v_i \mathbf{e}_i \in \mathbb{R}^d$, we have

$$\begin{aligned}
\|(A\circ B)v\| &= \|(A\circ B)\sum_{i=1}^d v_i \boldsymbol{e}_i\| \leq \sum_{i=1}^d |v_i| \|(A\circ B)e_i\| \\
&= \sum_{i=1}^d |v_i| \|(A_{1i}B_{1i}, \dots, A_{di}B_{di})^T\| \leq \sum_{i=1}^d |v_i| \left(\max_{j\in[d]} |A_{ji}|\right) \|(B_{1i}, \dots, B_{di})^T\| \\
&\leq \left(\max_{i,j\in[d]} |A_{ij}|\right) \sum_{i=1}^d |v_i| \|B\boldsymbol{e}_i\| \leq \left(\max_{i,j\in[d]} |A_{ij}|\right) \sum_{i=1}^d |v_i| \|B\|_2 \leq \sqrt{d}\left(\max_{i,j\in[d]} |A_{ij}|\right) \|B\|_2 \|v\|,
\end{aligned}$$

where $\{\mathbf{e}_i\}_{i\in[d]}$ is the canonical basis for $\mathbb{R}^d$. Since $\tilde{W} \succeq 0$ these facts imply that

$$\begin{aligned}
&\|\widehat{P} - P\|_2 \leq \|W \circ \widehat{\Sigma}\|_2 + \|\tilde{W} \circ (\widehat{\Sigma} - \Sigma)\|_2 \leq \|W \circ \widehat{\Sigma}\|_2 + \left(\max_{j\in[d]} |\tilde{W}_{jj}|\right) \|\widehat{\Sigma} - \Sigma\|_2 \\
&\leq \sqrt{d}\left(\max_{i,j\in[d]} |W_{ij}|\right) \|\widehat{\Sigma}\|_2 + \left(\max_{j\in[d]} |\tilde{W}_{jj}|\right) \|\widehat{\Sigma} - \Sigma\|_2 \\
&\leq \sqrt{d}\left(\max_{i,j\in[d]} |W_{ij}|\right) \|\Sigma\|_2 + \sqrt{d}\left(\max_{i,j\in[d]} |W_{ij}|\right) \|\widehat{\Sigma} - \Sigma\|_2 + \left(\max_{j\in[d]} |\tilde{W}_{jj}|\right) \|\widehat{\Sigma} - \Sigma\|_2 \\
&= \sqrt{d}\left(\max_{i,j\in[d]} \left|\frac{1}{\widehat{\sigma}_i\widehat{\sigma}_j} - \frac{1}{\sigma_i\sigma_j}\right|\right) \|\Sigma\|_2 + \sqrt{d}\left(\max_{i,j\in[d]} \left|\frac{1}{\widehat{\sigma}_i\widehat{\sigma}_j} - \frac{1}{\sigma_i\sigma_j}\right|\right) \|\widehat{\Sigma} - \Sigma\|_2 + \left(\max_{j\in[d]} \left|\frac{1}{\sigma_j^2}\right|\right) \|\widehat{\Sigma} - \Sigma\|_2.
\end{aligned}$$

For the second and final terms, Proposition 28 in Appendix D ensures that

$$\mathbb{P}\left(\|\widehat{\Sigma} - \Sigma\|_2 \geq x\right) \leq 2 \cdot 9^d \exp\left\{-n\frac{x}{16\nu^2} \wedge \left(\frac{x}{16\nu^2}\right)^2\right\},$$

and inverting this bound, we get that

$$\|\widehat{\Sigma} - \Sigma\|_2 \leq C_1\nu^2\sqrt{\frac{d + \log(4/t)}{n}} \vee \frac{d + \log(4/t)}{n},$$

with probability $\geq 1 - t/2$, for a universal constant $C_1 > 0$ sufficiently big. In this regard, we remark that universal constants might change from line to line, but we will use the same notation to ease the presentation. As for the first and second terms, writing $\sigma^2_{\min}$ for the smallest component of $(\sigma_1^2, \ldots, \sigma_d^2)$, observe that

$$\begin{aligned}
\max_{i,j\in[d]} \left|\frac{1}{\widehat{\sigma}_i\widehat{\sigma}_j} - \frac{1}{\sigma_i\sigma_j}\right| &= \max_{i,j\in[d]} \frac{1}{\sigma_i\sigma_j}\left|\frac{\sigma_i\sigma_j}{\widehat{\sigma}_i\widehat{\sigma}_j} - 1\right| \leq \frac{1}{\sigma^2_{\min}} \max_{i,j\in[d]} \left|\frac{\sigma_i\sigma_j}{\widehat{\sigma}_i\widehat{\sigma}_j} - 1\right| \\
&= \frac{1}{\sigma^2_{\min}} \max_{i,j\in[d]} \left|\frac{\sigma_i\sigma_j}{\widehat{\sigma}_i\widehat{\sigma}_j} - \frac{\sigma_i}{\widehat{\sigma}_i} + \frac{\sigma_i}{\widehat{\sigma}_i} - 1\right| \\
&\leq \frac{1}{\sigma^2_{\min}} \left(\max_{i,j\in[d]} \left|\frac{\sigma_j}{\widehat{\sigma}_j}\left(\frac{\sigma_i}{\widehat{\sigma}_i} - 1\right)\right| + \max_{i\in[d]} \left|\frac{\sigma_i}{\widehat{\sigma}_i} - 1\right|\right) \\
&= \frac{1}{\sigma^2_{\min}} \left(\max_{i\in[d]} \left|\frac{\sigma_i}{\widehat{\sigma}_i} - 1\right| \max_{j\in[d]} \left|\frac{\sigma_j}{\widehat{\sigma}_j}\right| + \max_{i\in[d]} \left|\frac{\sigma_i}{\widehat{\sigma}_i} - 1\right|\right) \\
&= \frac{1}{\sigma^2_{\min}} \left(\|D^{1/2}\widehat{D}^{-1/2} - I\|_2 \|D^{1/2}\widehat{D}^{-1/2}\|_2 + \|D^{1/2}\widehat{D}^{-1/2} - I\|_2\right),
\end{aligned}$$

which implies that, in order to control $\max_{i,j\in[d]} |1/\widehat{\sigma}_i\widehat{\sigma}_j - 1/\sigma_i\sigma_j|$, it is enough to control $\|D^{1/2}\widehat{D}^{-1/2} - I\|_2$. To this aim, first observe that, for all $t \in [0, 1]$,

$$\mathbb{P}(|\sigma_j^2/\widehat{\sigma}_j^2 - 1| > t) \leq \mathbb{P}(|\widehat{\sigma}_j^2/\sigma_j^2 - 1| > t/2).$$

Indeed,

$$\begin{aligned}
\mathbb{P}\left(|\sigma_j^2/\widehat{\sigma}_j^2 - 1| > t\right) &= \mathbb{P}\left(\left\{\sigma_j^2/\widehat{\sigma}_j^2 > 1 + t\right\} \cup \left\{\sigma_j^2/\widehat{\sigma}_j^2 < 1 - t\right\}\right) \\
&\leq \mathbb{P}\left(\sigma_j^2/\widehat{\sigma}_j^2 > 1 + t\right) + \mathbb{P}\left(\sigma_j^2/\widehat{\sigma}_j^2 < 1 - t\right)
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{P}\left(\widehat{\sigma}_j^2/\sigma_j^2 < (1+t)^{-1}\right) + \mathbb{P}\left(\widehat{\sigma}_j^2/\sigma_j^2 > (1-t)^{-1}\right) \\
&= \mathbb{P}\left(\widehat{\sigma}_j^2 - \sigma_j^2 < -\sigma_j^2 t(1+t)^{-1}\right) + \mathbb{P}\left(\widehat{\sigma}_j^2 - \sigma_j^2 > \sigma_j^2 t(1-t)^{-1}\right) \\
&= \mathbb{P}\left(\widehat{\sigma}_j^2/\sigma_j^2 - 1 < -t(1+t)^{-1}\right) + \mathbb{P}\left(\widehat{\sigma}_j^2/\sigma_j^2 - 1 > t(1-t)^{-1}\right),
\end{aligned}
$$

and since $t(1+t)^{-1}, t(1-t)^{-1} \geq t/2$ for $t \in [0,1]$, we can conclude

$$
\mathbb{P}\left(|\sigma_j^2/\widehat{\sigma}_j^2 - 1| > t\right) \leq \mathbb{P}(|\widehat{\sigma}_j^2/\sigma_j^2 - 1| > t/2).
$$

This is helpful, since we know how to control $|\widehat{\sigma}_j^2/\sigma_j^2 - 1|$. Indeed, since the $\mathbf{X}_i$ are ν-subgaussian random vectors, the X_{ij} for varying $i \in [n]$ are i.i.d. ν-subgaussian random variables, hence the X_{ij}^2 for varying $i \in [n]$ are i.i.d. subexponential with parameters $(4\sqrt{2}\nu^2, 4\nu^2)$ (see Proposition 27 in the Appendix). For every $j \in [d]$ and every $t > 0$, we can therefore use the standard Chernhoff method for subexponential random variables (see Proposition 26 in the Appendix) to see that

$$
\mathbb{P}\left(|\widehat{\sigma}_j^2/\sigma_j^2 - 1| > t\right) = \mathbb{P}\left(\left|\frac{1}{n}\sum_{i\in[n]}(X_{ij}^2 - \mathbb{E}[X_{ij}^2])\right| > \sigma_j^2 t\right) \leq \exp\left\{-n\sigma_j^4 t^2/(64\nu^4)\right\}
$$

if $0 \leq \sigma_j^2 t \leq 8\nu^2$. In particular, this remains valid for $0 \leq t \leq 8$, being $\sigma_j^2 \leq \nu^2$, and since we are interested in small values of t, we are allowed to focus just on this subgaussian regime in our analysis. Inverting the previous bound, we have that

$$
\begin{aligned}
1 - \mathbb{P}\left(|\sigma_j^2/\widehat{\sigma}_j^2 - 1| \leq \frac{16\nu^2}{\sigma_{\min}^2}\sqrt{\frac{\log(4d/t)}{n}}, \forall j \in [d]\right) &= \mathbb{P}\left(\exists j \in [d] : |\sigma_j^2/\widehat{\sigma}_j^2 - 1| \leq \frac{16\nu^2}{\sigma_{\min}^2}\sqrt{\frac{\log(4d/t)}{n}}\right) \\
\leq \sum_{j=1}^{d}\mathbb{P}\left(|\sigma_j^2/\widehat{\sigma}_j^2 - 1| \leq \frac{16\nu^2}{\sigma_j^2}\sqrt{\frac{\log(4d/t)}{n}}\right) &\leq \sum_{j=1}^{d}\mathbb{P}\left(|\widehat{\sigma}_j^2/\sigma_j^2 - 1| \leq \frac{8\nu^2}{\sigma_j^2}\sqrt{\frac{\log(4d/t)}{n}}\right) \leq t.
\end{aligned}
$$

Now, following similar lines as in Oliveira (2010), since for any $x \in [-3/4, 3/4]$ we have $|\sqrt{1+x} - 1| \leq x$ by the mean value theorem, if we assume $9\sigma_{\min}^4 n \geq 1024\nu^4 \log(4d/t)$ and take $x = \sigma_j^2/\widehat{\sigma}_j^2 - 1$, we also have

$$
\mathbb{P}\left(|\sigma_j/\widehat{\sigma}_j - 1| \leq \frac{16\nu^2}{\sigma_{\min}^2}\sqrt{\frac{\log(4d/t)}{n}}, \forall j \in [d]\right) \geq 1 - t,
$$

which leads to

$$
\mathbb{P}\left(\|D^{1/2}\widehat{D}^{-1/2} - I\|_2 \leq \frac{16\nu^2}{\sigma_{\min}^2}\sqrt{\frac{\log(4d/t)}{n}}\right) \geq 1 - t.
$$

It follows that

$$
\begin{aligned}
&\|D^{-1/2}\widehat{D}^{1/2} - I\|_2\|D^{-1/2}\widehat{D}^{1/2}\|_2 + \|D^{-1/2}\widehat{D}^{1/2} - I\|_2 \\
&\qquad \leq \frac{16\nu^2}{\sigma_{\min}^2}\sqrt{\frac{\log(4d/t)}{n}}\left(1 + \frac{16\nu^2}{\sigma_{\min}^2}\sqrt{\frac{\log(4d/t)}{n}}\right) + \frac{16\nu^2}{\sigma_{\min}^2}\sqrt{\frac{\log(4d/t)}{n}}
\end{aligned}
$$

with probability $\geq 1 - t$, and since for any $x \in [0, 1/2]$ we have $x(1+x) + x \leq 5x/2$, assuming $\sigma_{\min}^4 n \geq$

$1024\nu^4 \log(4d/t)$, we obtain

$$\|D^{-1/2}\widehat{D}^{1/2} - I\|_2 \|D^{-1/2}\widehat{D}^{1/2}\|_2 + \|D^{-1/2}\widehat{D}^{1/2} - I\|_2 \leq \frac{40\nu^2}{\sigma_{\min}^2}\sqrt{\frac{\log(4d/t)}{n}}$$

with probability $\geq 1-t$. This allows to conclude that

$$\|D^{-1/2}\widehat{D}^{1/2} - I\|_2 \|D^{-1/2}\widehat{D}^{1/2}\|_2 + \|D^{-1/2}\widehat{D}^{1/2} - I\|_2 \leq \frac{40\nu^2}{\sigma_{\min}^2}\sqrt{\frac{\log(8d/t)}{n}} \leq C_2 \frac{\nu^2}{\sigma_{\min}^2}\sqrt{\frac{\log(d/t)}{n}}$$

with probability $\geq 1-t/2$, for a universal constant $C_2 > 0$ sufficiently big. This implies that

$$\begin{aligned}
&\|\widehat{P} - P\|_2 \\
&\leq \left(\max_{j\in[d]}\left|\frac{1}{\sigma_j^2}\right|\right)\|\widehat{\Sigma}-\Sigma\|_2 + \sqrt{d}\left(\max_{i,j\in[d]}\left|\frac{1}{\widehat{\sigma}_i\widehat{\sigma}_j} - \frac{1}{\sigma_i\sigma_j}\right|\right)\|\Sigma\|_2 + \sqrt{d}\left(\max_{i,j\in[d]}\left|\frac{1}{\widehat{\sigma}_i\widehat{\sigma}_j} - \frac{1}{\sigma_i\sigma_j}\right|\right)\|\widehat{\Sigma}-\Sigma\|_2 \\
&\leq \frac{1}{\sigma_{\min}^2}\|\widehat{\Sigma}-\Sigma\|_2 + \frac{\sqrt{d}}{\sigma_{\min}^2}\left(\|D^{1/2}\widehat{D}^{-1/2} - I\|_2\|D^{1/2}\widehat{D}^{-1/2}\|_2 + \|D^{1/2}\widehat{D}^{-1/2} - I\|_2\right) + \\
&\quad + \frac{\sqrt{d}}{\sigma_{\min}^2}\left(\|D^{1/2}\widehat{D}^{-1/2} - I\|_2\|D^{1/2}\widehat{D}^{-1/2}\|_2 + \|D^{1/2}\widehat{D}^{-1/2} - I\|_2\right)\|\widehat{\Sigma}-\Sigma\|_2 \\
&\leq C_1\frac{\nu^2}{\sigma_{\min}^2}\left(\sqrt{\frac{d+\log(4/t)}{n}} \vee \frac{d+\log(4/t)}{n}\right) + C_2\frac{\nu^2}{\sigma_{\min}^4}\|\Sigma\|_2\sqrt{\frac{d\log(d/t)}{n}} \\
&\quad + C_3\frac{\nu^4}{\sigma_{\min}^4}\left(\sqrt{\frac{d+\log(4/t)}{n}} \vee \frac{d+\log(4/t)}{n}\right)\sqrt{\frac{d\log(d/t)}{n}} \\
&\leq C_1\frac{\nu^2}{\sigma_{\min}^2}\left(\sqrt{\frac{d+\log(1/t)}{n}} \vee \frac{d+\log(1/t)}{n}\right) + C_2\frac{\nu^4}{\sigma_{\min}^4}\sqrt{\frac{d\log(d/t)}{n}} \\
&\quad + C_3\frac{\nu^4}{\sigma_{\min}^4}\left(\sqrt{\frac{d+\log(1/t)}{n}} \vee \frac{d+\log(1/t)}{n}\right)\sqrt{\frac{d\log(d/t)}{n}}
\end{aligned}$$

with probability $1-t$, due to the fact that $\|\Sigma\|_2 \leq \nu^2$. □

Proof of Proposition 8. We aim at finding a random threshold $C_\alpha(\mathcal{X}_1)$ depending on the first part of the data $\mathcal{X}_1$ such that $\mathbb{P}_{H_0}\left(V(\widehat{\sigma}^2_{\mathbb{S}}) + \widehat{R}_2 \geq C_\alpha(\mathcal{X}_1)\right) \leq \alpha$, and since

$$\mathbb{P}_{H_0}\left(V(\widehat{\sigma}^2_{\mathbb{S}}) + \widehat{R}_2 \geq C_\alpha(\mathcal{X}_1)\right) \leq \mathbb{P}_{H_0}\left(V(\widehat{\sigma}^2_{\mathbb{S}}) \geq C_\alpha(\mathcal{X}_1)/2\right) + \mathbb{P}_{H_0}\left(\widehat{R}_2 \geq C_\alpha(\mathcal{X}_1)\right),$$

we can control the first term as in Theorem 6, while for the second one it is sufficient to choose $C_\alpha(\mathcal{X}_1)$ such that

$$\mathbb{P}_{H_0}\left(\widehat{R}_2 \geq C_\alpha(\mathcal{X}_1)/2\right) = \mathbb{P}_{H_0}\left(-\frac{1}{d}\langle \widehat{X}^{(1)}_{\mathbb{S}}, \widehat{\Sigma}^{(2)}_{\mathbb{S}}\rangle_{\mathbb{S}} \geq C_\alpha(\mathcal{X}_1)/2\right) \leq \alpha/2.$$

Using Holder's inequality for sequences of matrices, we have

$$\mathbb{P}_{H_0}\left(-\frac{1}{d}\langle \widehat{X}^{(1)}_{\mathbb{S}}, \widehat{\Sigma}^{(2)}_{\mathbb{S}}\rangle_{\mathbb{S}} \geq C_\alpha(\mathcal{X}_1)/2\right) \leq \mathbb{P}_{H_0}\left(-\frac{1}{d}\langle \widehat{X}^{(1)}_{\mathbb{S}}, \widehat{\Sigma}^{(2)}_{\mathbb{S}} - \Sigma_{\mathbb{S}}\rangle_{\mathbb{S}} \geq C_\alpha(\mathcal{X}_1)/2\right)$$

$$\leq \mathbb{P}_{H_0}\left(\left|\langle \widehat{X}^{(1)}_{\mathbb{S}}, \widehat{\Sigma}^{(2)}_{\mathbb{S}} - \Sigma_{\mathbb{S}}\rangle_{\mathbb{S}}\right| \geq dC_\alpha(\mathcal{X}_1)/2\right)$$
$$\leq \mathbb{P}_{H_0}\left(\|\widehat{X}^{(1)}_{\mathbb{S}}\|_{*,\mathbb{S}}\|\widehat{\Sigma}^{(2)}_{\mathbb{S}} - \Sigma_{\mathbb{S}}\|_{2,\mathbb{S}} \geq dC_\alpha(\mathcal{X}_1)/2\right).$$

Thus, if we choose $C_\alpha(\mathcal{X}_1) := 2\epsilon\|\widehat{X}^{(1)}_{\mathbb{S}}\|_{*,\mathbb{S}}/d$, with $\epsilon > 0$, it remains to choose ϵ such that

$$\mathbb{P}_{H_0}\left(\|\widehat{\Sigma}^{(2)}_{\mathbb{S}} - \Sigma_{\mathbb{S}}\|_{2,\mathbb{S}} \geq \epsilon\right) = \mathbb{P}_{H_0}\left(\max_{S\in\mathbb{S}}\|\widehat{\Sigma}^{(2)}_{S} - \Sigma_{S}\|_{2} \geq \epsilon\right) \leq \alpha/2.$$

This can be done again using Proposition 7, and the result follows. □

6.3 Proofs for Section 4.1

Proof of Theorem 9. For the d-cycle, our measure of consistency of the variances is

$$V(\Sigma_{\mathbb{S}}) = 1 - \min_{j\in[d]}\min_{S\in\{\{j-1,j\},\{j,j+1\}\}}\sigma^2_{S,j}.$$

We will show that testing

$$H_0 : V(\Sigma_{\mathbb{S}}) = 0 \quad \text{vs.} \quad H_1(\rho) : V(\Sigma_{\mathbb{S}}) > \rho,$$

requires at least a separation of the order $\sqrt{\log d/n}$, and since $T = V(\sigma^2_{\mathbb{S}}) + R(\Sigma_{\mathbb{S}})$, the statement would follow. Formally, referring to the same $\mathcal{P}_{\mathbb{S}}(0)$ and $\mathcal{P}_{\mathbb{S}}(\rho)$ defined in Section 4, this corresponds to assuming that $\Sigma_{\mathbb{S}}$ is always compatible, and constructing prior distributions just on $\{V(\sigma^2_{\mathbb{S}}) = 0\}$ and$\{V(\sigma^2_{\mathbb{S}}) > \rho\}$. We construct a lower bound to show that the minimax separation in this case is at least $c_1\sqrt{\log d/\min_{i\in[d]} n_i}$, where $c_1 > 0$ is a universal constant. Let

$$P_0 = \left(N^{\otimes n_1}\left(\mathbf{0}_2, \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right), \dots, N^{\otimes n_d}\left(\mathbf{0}_2, \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right)\right),$$

and

$$P_j = \left(N^{\otimes n_1}\left(\mathbf{0}_2, \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right), \dots, N^{\otimes n_j}\left(\mathbf{0}_2, \begin{pmatrix}1 & 0\\ 0 & 1-\delta\end{pmatrix}\right),\right.$$
$$\left. N^{\otimes n_{j+1}}\left(\mathbf{0}_2, \begin{pmatrix}1+\delta & 0\\ 0 & 1\end{pmatrix}\right), \dots, N^{\otimes n_d}\left(\mathbf{0}_2, \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right)\right),$$

for $j \in [d]$, and $\delta > 0$. It is clear that each P_j lies in $H_1(\delta)$, and that

$$\frac{dP_j}{dP_0}((x_{1,1}, y_{1,1}), \dots, (x_{1,n_1}, y_{1,n_1}), (x_{2,1}, y_{2,1}), \dots, (x_{d,n_d}, y_{d,n_d}))$$
$$= \prod_{h\in[n_j]}\frac{1}{(1-\delta)^{1/2}}\exp\left\{-\frac{\delta}{2(1-\delta)}y^2_{j,h}\right\}\prod_{h\in[n_{j+1}]}\frac{1}{(1+\delta)^{1/2}}\exp\left\{\frac{\delta}{2(1+\delta)}x^2_{j+1,h}\right\}.$$

Now, using the same strategy outlined in Section 4, it is enough to control the Total Variation distance

$$4\,\mathrm{TV}\left\{P_0, \frac{1}{d}\sum_{j=1}^{d} P_j\right\}^2 \le \chi^2\left(P_0, \frac{1}{d}\sum_{j=1}^{d} P_j\right) = \int \frac{\{\frac{1}{d}\sum_{j=1}^{d} dP_j\}^2}{dP_0} - 1$$

$$= \frac{1}{d^2}\sum_{j_1,j_2=1}^{d}\int \frac{dP_{j_1}dP_{j_2}}{dP_0} - 1 = \frac{1}{d^2}\sum_{j_1,j_2=1}^{d}\int \frac{dP_{j_1}}{dP_0}\frac{dP_{j_2}}{dP_0}dP_0 - 1.$$

Now, it is easy to see that if $j_2 \notin \{j_1-1, j_1, j_1+1\}$, then $\int \frac{dP_{j_1}}{dP_0}\frac{dP_{j_2}}{dP_0}dP_0 = 1$. This happens to be the case also when $j_2 = j_1 \pm 1$, since, for $j_2 = j_1 - 1 =: j-1$, we have

$$\int \frac{dP_{j_1}}{dP_0}\frac{dP_{j_2}}{dP_0}dP_0 = \int (1-\delta)^{-n_{j-1}/2}(1-\delta^2)^{-n_j/2}(1+\delta)^{-n_{j+1}/2}\prod_{h\in[n_{j-1}]}\exp\left\{-\frac{\delta}{2(1-\delta)}y_{j-1,h}^2\right\}$$

$$\times \prod_{h\in[n_j]}\exp\left\{\frac{\delta}{2(1+\delta)}x_{j,h}^2 - \frac{\delta}{2(1-\delta)}y_{j,h}^2\right\}\prod_{h\in[n_{j+1}]}\exp\left\{\frac{\delta}{2(1+\delta)}x_{j+1,h}^2\right\}dP_0$$

$$= \int \prod_{h\in[n_{j-1}]}\frac{1}{2\pi(1-\delta)^{1/2}}\exp\left\{-\frac{1}{2}\begin{pmatrix}x_{j-1,h}\\ y_{j-1,h}\end{pmatrix}^T\begin{pmatrix}1 & 0\\ 0 & \frac{1}{1-\delta}\end{pmatrix}\begin{pmatrix}x_{j-1,h}\\ y_{j-1,h}\end{pmatrix}\right\}d\boldsymbol{x}_{j-1}d\boldsymbol{y}_{j-1}$$

$$\times \int \prod_{h\in[n_j]}\frac{1}{2\pi(1-\delta^2)^{1/2}}\exp\left\{-\frac{1}{2}\begin{pmatrix}x_{j,h}\\ y_{j,h}\end{pmatrix}^T\begin{pmatrix}\frac{1}{1+\delta} & 0\\ 0 & \frac{1}{1-\delta}\end{pmatrix}\begin{pmatrix}x_{j,h}\\ y_{j,h}\end{pmatrix}\right\}d\boldsymbol{x}_{j}d\boldsymbol{y}_{j}$$

$$\times \int \prod_{h\in[n_{j+1}]}\frac{1}{2\pi(1+\delta)^{1/2}}\exp\left\{-\frac{1}{2}\begin{pmatrix}x_{j+1,h}\\ y_{j+1,h}\end{pmatrix}^T\begin{pmatrix}\frac{1}{1+\delta} & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}x_{j+1,h}\\ y_{j+1,h}\end{pmatrix}\right\}d\boldsymbol{x}_{j+1}d\boldsymbol{y}_{j+1},$$

which is equal to 1. Similarly, if $j_1 = j_2 = j$,

$$\int \frac{dP_{j_1}}{dP_0}\frac{dP_{j_2}}{dP_0}dP_0 = \int (1-\delta)^{-n_j}(1+\delta)^{-n_{j+1}}\prod_{h\in[n_j]}\exp\left\{-\frac{\delta}{1-\delta}y_{j,h}^2\right\}\prod_{h\in[n_{j+1}]}\exp\left\{\frac{\delta}{1+\delta}x_{j+1,h}^2\right\}dP_0$$

$$= (1-\delta)^{-n_j}(1+\delta)^{-n_{j+1}}\int \prod_{h\in[n_j]}\frac{1}{2\pi}\exp\left\{-\frac{1}{2}\begin{pmatrix}x_{j,h}\\ y_{j,h}\end{pmatrix}^T\begin{pmatrix}1 & 0\\ 0 & \frac{1+\delta}{1-\delta}\end{pmatrix}\begin{pmatrix}x_{j,h}\\ y_{j,h}\end{pmatrix}\right\}d\boldsymbol{x}_{j}d\boldsymbol{y}_{j}$$

$$\times \int \prod_{h\in[n_{j+1}]}\frac{1}{2\pi}\exp\left\{-\frac{1}{2}\begin{pmatrix}x_{j+1,h}\\ y_{j+1,h}\end{pmatrix}^T\begin{pmatrix}\frac{1-\delta}{1+\delta} & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}x_{j,h}\\ y_{j,h}\end{pmatrix}\right\}d\boldsymbol{x}_{j+1}d\boldsymbol{y}_{j+1}$$

$$= (1-\delta^2)^{-n_j/2}(1-\delta^2)^{-n_{j+1}/2}\int \prod_{h\in[n_j]}\frac{1}{2\pi}\sqrt{\frac{1+\delta}{1-\delta}}\exp\left\{-\frac{1}{2}\begin{pmatrix}x_{j,h}\\ y_{j,h}\end{pmatrix}^T\begin{pmatrix}1 & 0\\ 0 & \frac{1+\delta}{1-\delta}\end{pmatrix}\begin{pmatrix}x_{j,h}\\ y_{j,h}\end{pmatrix}\right\}d\boldsymbol{x}_{j}d\boldsymbol{y}_{j}$$

$$\times \int \prod_{h\in[n_{j+1}]}\frac{1}{2\pi}\sqrt{\frac{1-\delta}{1+\delta}}\exp\left\{-\frac{1}{2}\begin{pmatrix}x_{j+1,h}\\ y_{j+1,h}\end{pmatrix}^T\begin{pmatrix}\frac{1-\delta}{1+\delta} & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}x_{j,h}\\ y_{j,h}\end{pmatrix}\right\}d\boldsymbol{x}_{j+1}d\boldsymbol{y}_{j+1}$$

$$= (1-\delta^2)^{-n_j/2}(1-\delta^2)^{-n_{j+1}/2} = (1-\delta^2)^{-(n_j+n_{j+1})/2}.$$

It follows that

$$4\,\mathrm{TV}\left\{P_0, \frac{1}{d}\sum_{j=1}^{d} P_j\right\}^2 \le \frac{1}{d^2}\sum_{j_1,j_2=1}^{d}\int \frac{dP_{j_1}}{dP_0}\frac{dP_{j_2}}{dP_0}dP_0 - 1$$
$$= \frac{1}{d^2}\sum_{j_1,j_2=1}^{d} \mathbb{1}_{j_1=j_2}(1-\delta^2)^{-(n_{j_1}+n_{j_1+1})/2} + \mathbb{1}_{j_1\neq j_2}$$
$$= \frac{1}{d^2}\sum_{j=1}^{d}(1-\delta^2)^{-(n_j+n_{j+1})/2} - \frac{1}{d}$$
$$\le \frac{1}{d^2}\sum_{j=1}^{d}\exp\{-(n_j+n_{j+1})\delta^2/2\} - \frac{1}{d},$$

from which we see that $\mathrm{TV}\left\{P_0, \frac{1}{d}\sum_{j=1}^{d} P_j\right\} \le 1/2$ if $\delta \le \sqrt{2\log(1+d)/(n_j+n_{j+1})}$ for all $j \in [d]$. The above bound on the total variation distance demonstrates that we may choose $\delta = \sqrt{\log(1+d)/\min_j n_j}$, and hence that we have

$$\rho^* \ge \delta = \left\{\frac{\log(1+d)}{\min_j n_j}\right\}^{1/2},$$

as claimed. □

Proof of Proposition 10. We will prove the result using the dual characterisation, which allows expressing $R(\Sigma_{\mathbb{S}_d})$ as

$$1 - \frac{1}{d}\sup\{\mathrm{tr}(\Sigma) : \Sigma \in \mathcal{P}^*, \Sigma_{11} = \cdots = \Sigma_{dd}, \Sigma_{\mathbb{S}_d} - A\Sigma \succeq_{\mathbb{S}_d} 0\}.$$

Suppose that $\prod_{j=d+1}^{d+k}\rho_{j,j+1} = 1$. We will show that $R(\Sigma_{\mathbb{S}_{d+k}}) = R(\Sigma_{\mathbb{S}_d})$ by proving both $R(\Sigma_{\mathbb{S}_{d+k}}) \ge R(\Sigma_{\mathbb{S}_d})$ and $R(\Sigma_{\mathbb{S}_{d+k}}) \le R(\Sigma_{\mathbb{S}_d})$. As for the first of these, for every $\tilde{\Sigma}$ optimal for $\Sigma_{\mathbb{S}_{d+k}}$, we will show that $\Sigma = \tilde{\Sigma}_{|[d]}$ is feasible for $\Sigma_{\mathbb{S}_d}$. Now, $\Sigma \succeq 0$ since $\tilde{\Sigma} \succeq 0$, and $\Sigma_{11} = \ldots = \Sigma_{dd}$ by definition of $\tilde{\Sigma}$. As for $\Sigma_{\mathbb{S}_d} - A_{\mathbb{S}_d}\Sigma \succeq_{\mathbb{S}} 0$, observe that $\Sigma_{\mathbb{S}_d}$ contains exactly the first d matrices in $\Sigma_{\mathbb{S}_d}$, but $\mathbb{S}_d$ contains just $d-1$ patterns of $\Sigma_{\mathbb{S}_{d+k}}$. This is due to the fact that $\mathbb{S}_d$ has $\{d,1\}$ in place of $\{d,d+1\}$, which prevents us from employing Proposition 4 (ii). Nonetheless, observe that $\tilde{\Sigma}_{1,d} = \tilde{\Sigma}_{d,d+1}$, due to the fact that $\tilde{\Sigma}_{j,j+1} = \rho_{j,j+1}\tilde{\Sigma}_{11} = \pm\tilde{\Sigma}_{11}$ for all $j \in \{d+1,\ldots,d+k\}$. Indeed, $\Sigma_{\mathbb{S}_{d+k}} - A_{\mathbb{S}_{d+k}}\tilde{\Sigma} \succeq_{\mathbb{S}_{d+k}} 0$ implies that

$$\begin{pmatrix} 1 & \pm 1 \\ \pm 1 & 1 \end{pmatrix} - \begin{pmatrix} \tilde{\Sigma}_{11} & \tilde{\Sigma}_{j,j+1} \\ \tilde{\Sigma}_{j,j+1} & \tilde{\Sigma}_{11} \end{pmatrix} \succeq 0$$

for all $j \in \{d+1,\ldots,d+k\}$, which can be satisfied if and only if $\tilde{\Sigma}_{j,j+1} = \pm\tilde{\Sigma}_{11} = \rho_{j,j+1}\tilde{\Sigma}_{11}$, since we must also have $|\tilde{\Sigma}_{j,j+1}| \le \tilde{\Sigma}_{11}$ in order to have $\tilde{\Sigma} \succeq 0$. The fact that $\tilde{\Sigma}_{j,j+1} = \rho_{j,j+1}\tilde{\Sigma}_{11}$ for all $j \in \{d+1,\ldots,d+k\}$ implies that $\mathrm{Var}(X_{j+1} - \tilde{\Sigma}_{j+1,j+2}X_{j+2}) = 0$, for all $j \in \{d,\cdots,d+k-1\}$, since $A_{\mathbb{S}_{d+k}}\tilde{\Sigma}$ is compatible. By induction, this gives $\mathrm{Var}(X_{d+1} - \prod_{j=d+1}^{d+k}\tilde{\Sigma}_{j,j+1}X_1) = 0$ by which

$$\tilde{\Sigma}_{1,d} = \frac{1}{\prod_{j=d+1}^{d+k}\tilde{\Sigma}_{j,j+1}}\tilde{\Sigma}_{d,d+1} = \frac{1}{\prod_{j=d+1}^{d+k}\rho_{j,j+1}}\tilde{\Sigma}_{d,d+1} = \tilde{\Sigma}_{d,d+1}.$$

Since $\tilde{\Sigma}_{1,d} = \tilde{\Sigma}_{d,d+1}$, we know that $\Sigma_{\mathbb{S}_{d+k}} - A_{\mathbb{S}_{d+k}}\tilde{\Sigma} \succeq_{\mathbb{S}_{d+k}} 0$ implies that $\Sigma_{\mathbb{S}_d} - A_{\mathbb{S}_d}\Sigma \succeq_{\mathbb{S}} 0$.

To show the reverse inequality, consider an optimal Σ for $\Sigma_{\mathbb{S}_d}$ coming from the dual formulation above, and define

$$\tilde{\Sigma} := \begin{pmatrix} \Sigma & B \\ B^T & U \end{pmatrix},$$

where

$$U := \Sigma_{11} \begin{pmatrix} U_{11} & \cdots & U_{1k} \\ \vdots & & \vdots \\ U_{k1} & \cdots & U_{kk} \end{pmatrix}$$

is such that $U = U^T$, $U_{ii} = 1$ for $i \in [k]$, $U_{i,i+1} = \rho_{d+i,d+i+1}$ for $i \in [k-1]$, $U_{1k} = U_{k1}$ is either $+1$ or -1 to make this $(k-1)$-cycle completable, and the other entries are again $+1$ or -1 to make the cycle consistent; and

$$B^T := \begin{pmatrix} B_{11} & \cdots & B_{1d} \\ \vdots & & \vdots \\ B_{k1} & \cdots & B_{kd} \end{pmatrix}$$

is such that $B_{ij} = \Sigma_{1j} \cdot U_{i1}$ for $i \in [k], j \in [d]$. If such a $\tilde{\Sigma}$ is feasible for $\Sigma_{\mathbb{S}_{d+k}}$, then the result would follow from the fact that $R(\Sigma_{\mathbb{S}_{d+k}}) \leq 1 - \operatorname{tr}(\tilde{\Sigma})/(d+k) = 1 - \Sigma_{11} = R(\Sigma_{\mathbb{S}_d})$. The condition $\Sigma_{\mathbb{S}_{d+k}} - A_{\mathbb{S}_{d+k}}\tilde{\Sigma} \succeq_{\mathbb{S}_{d+k}} 0$ is implied by $\Sigma_{\mathbb{S}_d} - A\Sigma \succeq_{\mathbb{S}_d} 0$, which is satisfied by hypothesis, and $(1-\Sigma_{11})\Sigma_{\{i,i+1\}} \succeq 0$ for $i \in \{d+1, \cdots, d+k-1\}$, which is again satisfied since $\Sigma_{11} \in [0,1]$ and $\Sigma_{\{i,i+1\}} \succeq 0$. Moreover, being a symmetric block matrix, $\tilde{\Sigma}$ is positive semi-definite if and only if $\Sigma \succeq 0$, which is true by hypothesis, $U - B^T\Sigma^\dagger B \succeq 0$, and $(I - \Sigma\Sigma^\dagger)B = 0$, where $\Sigma^\dagger$ is the Moore-Penrose inverse of Σ. As for the first of these last two conditions, observe that the (k,h)-th entry of $B^T\Sigma^\dagger B$ is given by

$$(B^T\Sigma^\dagger B)_{kh} = U_{k1}\Sigma_1^T\Sigma^\dagger\Sigma_1 U_{h1} = U_{k1}U_{h1}\Sigma_1^T\Sigma^\dagger\Sigma_1 = U_{kh}\Sigma_1^T\Sigma^\dagger\Sigma_1,$$

where Σ_1 is the first column of Σ. What is left to prove is to check that $\Sigma_1^T\Sigma^\dagger\Sigma_1 = \operatorname{tr}(\Sigma_1^T\Sigma^\dagger\Sigma_1) = \Sigma_{11}$ and, to this aim, we will use the limit characterisation of the pseudoinverse (see pag. 19 in Albert (1972)), which allows writing $\Sigma^\dagger$ as $\lim_{\delta\to 0}\Sigma^T(\Sigma\Sigma^T + \delta^2 I)^{-1}$. With this in mind, and calling λ_i the eigenvalues of Σ, and $\boldsymbol{v}_i$ the associated orthonormal eigenvectors,

$$\begin{aligned}
\operatorname{tr}(\Sigma_1^T\Sigma^\dagger\Sigma_1) &= \operatorname{tr}(\Sigma^\dagger\Sigma_1\Sigma_1^T) = \operatorname{tr}\left(\lim_{\delta\to 0}\Sigma^T(\Sigma\Sigma^T + \delta^2 I)^{-1}\Sigma_1\Sigma_1^T\right) \\
&= \lim_{\delta\to 0}\operatorname{tr}\left((\Sigma\Sigma^T + \delta^2 I)^{-1}\Sigma_1\Sigma_1^T\Sigma^T\right) = \lim_{\delta\to 0}\operatorname{tr}\left(\sum_{i=1}^d \frac{1}{\lambda_i^2+\delta^2}\boldsymbol{v}_i\boldsymbol{v}_i^T\Sigma_1\Sigma_1^T\Sigma^T\right) \\
&= \lim_{\delta\to 0}\sum_{i=1}^d \frac{1}{\lambda_i^2+\delta^2}\operatorname{tr}\left(\boldsymbol{v}_i\boldsymbol{v}_i^T\Sigma_1\Sigma_1^T\Sigma^T\right) = \lim_{\delta\to 0}\sum_{i,j=1}^d \frac{\lambda_j}{\lambda_i^2+\delta^2}\operatorname{tr}\left(\boldsymbol{v}_j\boldsymbol{v}_j^T\boldsymbol{v}_i\boldsymbol{v}_i^T\Sigma_1\Sigma_1^T\right) \\
&= \lim_{\delta\to 0}\sum_{i=1}^d \frac{\lambda_i}{\lambda_i^2+\delta^2}\|\boldsymbol{v}_i\boldsymbol{v}_i^T\Sigma_1\|_2^2 = \lim_{\delta\to 0}\sum_{i=1}^d \frac{\lambda_i}{\lambda_i^2+\delta^2}\|\boldsymbol{v}_i\boldsymbol{v}_i^T\sum_{j=1}^d \lambda_j v_{j1}\boldsymbol{v}_j\|_2^2
\end{aligned}$$

$$= \lim_{\delta\to 0} \sum_{i=1}^{d} \frac{\lambda_i^3}{\lambda_i^2+\delta^2} \|v_{i1}\boldsymbol{v}_i\|_2^2 = \sum_{i=1}^{d} \lambda_i v_{i1}^2 = \Sigma_{11}.$$

The other condition can be checked easily using again the limit characterisation of $\Sigma^\dagger$ and the spectral decomposition of Σ. This concludes the proof for the case $\prod_{j=d+1}^{d+k} \rho_{j,j+1} = +1$. On the other hand, if $\prod_{j=d+1}^{d+k} \rho_{j,j+1} = -1$, $R(\Sigma_{\mathbb{S}_{d+k}}) \geq R(\Sigma_{\mathbb{S}_d})$ follows after noticing that, if $\tilde{\Sigma}$ is optimal for $\Sigma_{\mathbb{S}_{d+k}}$, we must have $\tilde{\Sigma}_{d,d+1} = -\tilde{\Sigma}_{d,1}$. As for $R(\Sigma_{\mathbb{S}_{d+k}}) \leq R(\Sigma_{\mathbb{S}_d})$, the proof follows the exact same line as the one above, with the only exception that B_{ij} should now be defined as $-\Sigma_{1j} \cdot U_{i1}$ for $i \in [k], j \in [d]$. ☐

Proof of Example 4. Start by considering a 3-cycle. In the first case, the optimal Σ of the dual representation

$$R(\Sigma_{\mathbb{S}_3}) = 1 - \frac{1}{d} \sup\{\mathrm{tr}(\Sigma) : \Sigma \in \mathcal{P}^*, \Sigma_{\mathbb{S}} - A\Sigma \succeq_{\mathbb{S}} 0, \Sigma_{11} = \Sigma_{22} = \Sigma_{33}\}$$

must be of the form

$$\begin{pmatrix} \lambda & x & \lambda \\ x & \lambda & y \\ \lambda & y & \lambda \end{pmatrix}$$

for some $\lambda \in [0,1]$ and some $x, y \in [-\lambda, \lambda]$ in order to satisfy $\Sigma_{\mathbb{S}} - A\Sigma \succeq_{\mathbb{S}} 0$. Furthermore, since $\det(\Sigma) = -\lambda(x-y)^2$, we must have $x = y$ in order to satisfy $\Sigma \succeq 0$. It follows that

$$\begin{aligned} R(\Sigma_{\mathbb{S}_3}) &= 1 - \sup\{\lambda \in [0,1] : 1-\lambda \geq \max\{|x-\rho_1|, |x-\rho_2|\},\ \text{with } x \in [-\lambda, \lambda]\} \\ &= \inf\{\epsilon \in [0,1] : \epsilon \geq \max\{|x-\rho_1|, |x-\rho_2|\}, \epsilon \leq 1 - |x|\} \\ &= \inf\{\max\{|x-\rho_1|, |x-\rho_2|\} \in [0,1] : \max\{|x-\rho_1|, |x-\rho_2|\} \leq 1 - |x|\} \end{aligned}$$

which is equal to $(\cos\theta_2 - \cos\theta_1)/2$. As for the second case with $d = 3$, setting $\rho_2 = 1$ in the above we see that if

$$\Sigma_{\mathbb{S}_3} = \left\{ \begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \right\},$$

then $R(\Sigma_{\mathbb{S}_3}) = (1-\rho_1)/2 = (1-\cos\theta_1)/2 = \sin^2(\theta_1/2)$. Plugging in $\theta_1 = \pi$ gives the sufficiency part of the second statement. As for the necessity part, Proposition 12 (i) implies that it is necessary that $|\rho_i| = 1$ for all $i \in [d]$ for R to be 1. ☐

Proof of Proposition 11. It is easy to see that we can always transform the original d-cycle into a new one where at most $\cos\theta_1$ is negative by changing some X_j into $-X_j$. To see why, let $\theta = (\theta_1, \ldots, \theta_d)$ be such that $\theta_j = \mathbb{1}\{\rho_{j,j+1} \geq 0\}$, and observe that, if $\theta_{j-1} = 0$ and $\theta_j = 1$, changing the sign of X_j corresponds to switching θ_{j-1} with θ_j. Hence, it is easy to see that we can switch signs to some variables in order to reach a configuration of θ in which all the zeros are at the beginning, and all the ones at the end. It is now sufficient to couple the zeros starting from the end, and switch sign to make it both one, to get $\theta = (\theta_1, \mathbf{1}_{d-1})$, where $\theta_1 = +1$ if the number of original $\rho_{j,j+1}$ is even, and zero otherwise. As a by-product, this also shows that we can always assume without loss of generality that at most $\cos\theta_1$ is negative. Now, let $\tilde{\Sigma}_{\mathbb{S}_d}$ be this new d-cycle: what we want to show is that $R(\tilde{\Sigma}_{\mathbb{S}_d}) = R(\Sigma_{\mathbb{S}_d})$, and, in order to do so, we will show that we can construct feasible $\tilde{X}_{\mathbb{S}}$ and $\tilde{\Sigma}$ for primal and dual problems of $\tilde{\Sigma}_{\mathbb{S}_d}$ which lead to the same target values, using the optimal $X_{\mathbb{S}}$ and Σ for $\Sigma_{\mathbb{S}_d}$. Starting from the dual problem, let M be a diagonal matrix such that $M_{jj} = -1$ if X_j was replaced with $-X_j$, and $+1$ otherwise. Then, it is easy to see $\tilde{\Sigma} = M\Sigma M$ has the same trace as Σ, and it is feasible for $\tilde{\Sigma}_{\mathbb{S}_d}$: indeed, $\tilde{\Sigma} \succeq 0$ since it has the same spectrum as Σ, being similar matrices, and $\tilde{\Sigma}_{\mathbb{S}_d} - A\tilde{\Sigma} \succeq_{\mathbb{S}_d} 0$ because for every $j \in [d]$ we have

$$\begin{aligned}
\tilde{\Sigma}_{\{j,j+1\}} - \tilde{\Sigma}_{|\{j,j+1\}} &= \begin{pmatrix} 1 & \tilde{\rho}_j \\ \tilde{\rho}_j & 1 \end{pmatrix} - \begin{pmatrix} \tilde{\Sigma}_{j,j} & \tilde{\Sigma}_{j,j+1} \\ \tilde{\Sigma}_{j,j+1} & \tilde{\Sigma}_{j+1,j+1} \end{pmatrix} \\
&= \begin{pmatrix} 1 & M_{j,j}M_{j+1,j+1}\rho_j \\ M_{j,j}M_{j+1,j+1}\rho_j & 1 \end{pmatrix} - \begin{pmatrix} \Sigma_{j,j} & M_{j,j}M_{j+1,j+1}\Sigma_{j,j+1} \\ M_{j,j}M_{j+1,j+1}\Sigma_{j,j+1} & \Sigma_{j+1,j+1} \end{pmatrix} \\
&= \begin{pmatrix} 1 - \Sigma_{j,j} & M_{j,j}M_{j+1,j+1}(\rho_j - \Sigma_j) \\ M_{j,j}M_{j+1,j+1}(\rho_j - \Sigma_j) & 1 - \Sigma_{j+1,j+1} \end{pmatrix} \succeq 0,
\end{aligned}$$

since $|M_{j,j}M_{j+1,j+1}(\rho_j - \Sigma_j)| = |\rho_j - \Sigma_j| \leq 1 - \Sigma_{j,j} = 1 - \Sigma_{j+1,j+1}$, due to the fact that Σ is feasible for $\Sigma_{\mathbb{S}}$. As for the primal problem, it is sufficient to define $\tilde{X}_{\mathbb{S}} = AM \cdot X_{\mathbb{S}} \cdot AM$, where $\cdot$ acts pointwise, which essentially consists in changing the signs of the off-diagonal entries of $X_{\mathbb{S}}$ according to M. Let

$$X_{\{j,j+1\}} = \begin{pmatrix} x_{j,11} & x_{j,12} \\ x_{j,21} & x_{j,22} \end{pmatrix} \text{ and } \tilde{X}_{\{j,j+1\}} = \begin{pmatrix} \tilde{x}_{j,11} & \tilde{x}_{j,12} \\ \tilde{x}_{j,21} & \tilde{x}_{j,22} \end{pmatrix} = \begin{pmatrix} x_{j,11} & M_{j,j}M_{j+1,j+1}x_{j,12} \\ M_{j,j}M_{j+1,j+1}x_{j,21} & x_{j,22} \end{pmatrix},$$

for all $j \in [d]$. It is easy to show that $\tilde{X}_{\mathbb{S}}$ is feasible, and clearly leads to $\langle \tilde{X}_{\mathbb{S}}, \tilde{\Sigma}_{\mathbb{S}} \rangle_{\mathbb{S}} = \langle X_{\mathbb{S}}, \Sigma_{\mathbb{S}} \rangle_{\mathbb{S}}$ since, for a generic pattern $j \in [d]$, we have

$$\begin{aligned}
\langle \tilde{X}_{\{j,j+1\}}, \tilde{\Sigma}_{\{j,j+1\}} \rangle &= \langle \begin{pmatrix} \tilde{x}_{j,11} & \tilde{x}_{j,12} \\ \tilde{x}_{j,21} & \tilde{x}_{j,22} \end{pmatrix}, \begin{pmatrix} 1 & \tilde{\rho}_j \\ \tilde{\rho}_j & 1 \end{pmatrix} \rangle \\
&= \langle \begin{pmatrix} x_{j,11} & M_{j,j}M_{j+1,j+1}x_{j,12} \\ M_{j,j}M_{j+1,j+1}x_{j,21} & x_{j,22} \end{pmatrix}, \begin{pmatrix} 1 & M_{j,j}M_{j+1,j+1}\rho_j \\ M_{j,j}M_{j+1,j+1}\rho_j & 1 \end{pmatrix} \rangle \\
&= x_{j,11} + x_{j,22} + M_{j,j}^2 M_{j+1,j+1}^2 x_{j,12}\rho_j + M_{j,j}^2 M_{j+1,j+1}^2 x_{j,21}\rho_j \\
&= x_{j,11} + x_{j,22} + x_{j,12}\rho_j + x_{j,21}\rho_j \\
&= \langle \begin{pmatrix} x_{j,11} & x_{j,12} \\ x_{j,21} & x_{j,22} \end{pmatrix}, \begin{pmatrix} 1 & \rho_j \\ \rho_j & 1 \end{pmatrix} \rangle = \langle X_{\{j,j+1\}}, \Sigma_{\{j,j+1\}} \rangle.
\end{aligned}$$

This completes the proof. □

Proof of Proposition 12. (i) We may suppose without loss of generality that $|\rho_i| \neq 1$ as, otherwise, we may perform the reduction given in Proposition 10. Possibly, this reduces the d-cycle to a 3-cycle: if there are no more correlations equal to ± 1, then we proceed, otherwise we know R exactly thanks to Example 4 and we can check that the claim holds. Now, calling $M_{i,i+1} = \cos\varphi_i$, we have

$$\begin{aligned}
\lambda^* = 1 - R(\Sigma_{\mathbb{S}}) &= \frac{1}{d}\sup\{\mathrm{tr}(\Sigma) : \Sigma \succeq 0, \Sigma_{11} = \ldots = \Sigma_{dd}, \Sigma_{\mathbb{S}'} - A\Sigma \succeq_{\mathbb{S}'} 0\} \\
&= \sup\{\lambda : M \succeq 0, M_{11} = \ldots = M_{dd} = 1, 1 - \lambda \geq |\rho_i - \lambda\cos\varphi_i| \text{ for all } i \in [d]\}
\end{aligned}$$

$$= \sup \left\{ \lambda : \lambda \leq \min_{i \in [d]} \min \left\{ \frac{1-\rho_i}{1-\cos\varphi_i}, \frac{1+\rho_i}{1+\cos\varphi_i} \right\}, \right.$$
$$\left. \sum_{i \in K} \varphi_i \leq (|K|-1)\pi + \sum_{i \notin K} \varphi_i \text{ for all } K \subseteq [d] \text{ with } |K| \text{ odd} \right\}.$$

Now, this implies that

$$\frac{1}{\lambda^*} = \inf \left\{ z : z \geq \max_{i \in [d]} \max \left\{ \frac{1-\cos\varphi_i}{1-\rho_i}, \frac{1+\cos\varphi_i}{1+\rho_i} \right\}, \right.$$
$$\left. \sum_{i \in K} \varphi_i \leq (|K|-1)\pi + \sum_{i \notin K} \varphi_i \text{ for all } K \subseteq [d] \text{ with } |K| \text{ odd} \right\}.$$

Calling $g(\varphi_i) = (1-\cos\varphi_i)/(1-\rho_i)$ and $h(\varphi_i) = (1+\cos\varphi_i)/(1+\rho_i)$ for all $i \in [d]$, this is a linearly constrained finite minimax problem (see Chapter 2 in Polak (2012)), namely

$$\frac{1}{\lambda^*} = \min \max_{i \in [d]} \max\{g(\varphi_i), h(\varphi_i)\}$$

under the 2^{d-1} linear constraints

$$\sum_{i \in K} \varphi_i \leq (|K|-1)\pi + \sum_{i \notin K} \varphi_i \text{ for all } K \subseteq [d] \text{ with } |K| \text{ odd},$$

which is equivalent to

$$\begin{aligned} &\text{minimise } z \\ &\text{subject to } g(\varphi_i) \leq z \text{ for all } i \in [d], \\ &\qquad h(\varphi_i) \leq z \text{ for all } i \in [d], \\ &\qquad \sum_{i \in K} \varphi_i \leq (|K|-1)\pi + \sum_{i \notin K} \varphi_i \text{ for all } K \subseteq [d] \text{ with } |K| \text{ odd}. \end{aligned}$$

As a result, every optimal solution $(\varphi_1^*, \ldots, \varphi_d^*)$ must satisfy the Karush–Kuhn–Tucker (KKT) conditions (see Chapter 5 of Boyd and Vandenberghe (2004), Chapter 28-30 of Rockafellar (1970))

$$\begin{aligned} &(i) \quad \left(\frac{\lambda_i}{1-\rho_i} - \frac{\lambda_{i+d}}{1+\rho_i} \right) \sin\varphi_i = \sum_{\substack{|K| odd \\ i \in K}} \mu_K - \sum_{\substack{|K| odd \\ i \in K^c}} \mu_K, \text{ for all } i \in [d], \\ &(ii) \quad \lambda_i \geq 0, \lambda_{d+i} \geq 0, \text{ for all } i \in [d], \\ &(iii) \quad \sum_{i=1}^{d} (\lambda_i + \lambda_{d+i}) = 1, \\ &(iv) \quad \lambda_i (g(\varphi_i) - \max_{i \in [d]} \max\{g(\varphi_i), h(\varphi_i)\}) = 0, \text{ for all } i \in [d], \\ &(v) \quad \lambda_{d+i} (h(\varphi_i) - \max_{i \in [d]} \max\{g(\varphi_i), h(\varphi_i)\}) = 0, \text{ for all } i \in [d], \\ &(vi) \quad g(\varphi_i) \leq \max_{i \in [d]} \max\{g(\varphi_i), h(\varphi_i)\}), \text{ for all } i \in [d], \end{aligned}$$

$$(vii) \quad h(\varphi_i) \le \max_{i\in[d]} \max\{g(\varphi_i), h(\varphi_i)\}), \text{ for all } i \in [d],$$

$$(viii) \quad \mu_K \ge 0, \text{ for all } K \subseteq [d] \text{ with } |K| \text{ odd},$$

$$(ix) \quad \sum_{i\in K} \varphi_i \le (|K|-1)\pi + \sum_{i\notin K} \varphi_i \text{ for all } K \subseteq [d] \text{ with } |K| \text{ odd},$$

$$(x) \quad \mu_K \left(\sum_{i\in K} \varphi_i - (|K|-1)\pi - \sum_{i\notin K} \varphi_i \right) = 0, \text{ for all } K \subseteq [d] \text{ with } |K| \text{ odd}.$$

Now, observe that conditions (iv) and (v) imply that, for all $i \in [d]$, either $g(\varphi_i)$ or $h(\varphi_i)$ reaches the maximum, meaning that the minimal $1/\lambda^*$ is equal to this common value. Indeed, if the original d-cycle is completable, this statement is trivial, since we must have $|\rho_i - \cos\varphi_i^*| = 0$. This is the only case in which we can have

$$\frac{1-\rho_i}{1-\cos\varphi_i} = \frac{1+\rho_i}{1+\cos\varphi_i} = 1,$$

meaning that when the d-cycle is incompatible, then either $\max_{i\in[d]} \max\{g(\varphi_i), h(\varphi_i)\} - g(\varphi_i) > 0$ or $\max_{i\in[d]} \max\{g(\varphi_i), h(\varphi_i)\} - h(\varphi_i) > 0$. Indeed, if $R > 0$, either λ_i or λ_{d+i} must be equal to zero since either $g(\varphi_i)$ or $h(\varphi_i)$ has a strictly positive gap from $\max_{i\in[d]} \max\{g(\varphi_i), h(\varphi_i)\}$. If both $\lambda_i = 0$ and $\lambda_{i+d} = 0$, we would have

$$\sum_{\substack{|K| odd \\ i\in K}} \mu_K = \sum_{\substack{|K| odd \\ i\in K^c}} \mu_K,$$

which is a contradiction due to the fact that there exists a unique $\mu_K \neq 0$. To prove the existence part, observe that if $\mu_K = 0$ for all $K \subseteq [d]$ with $|K|$ odd, then we would have

$$\left(\frac{\lambda_i}{1-\rho_i} - \frac{\lambda_{i+d}}{1+\rho_i} \right) \sin\varphi_i = 0$$

for all $i \in [d]$, and since there exists at least a j such that $\lambda_j + \lambda_{d+j} > 0$ due to (iii), this would imply that $\varphi_j \in \{0, \pi\}$, which leads to $\theta_j \in \{0, \pi\}$, which is excluded from our analysis. To prove the uniqueness part, suppose there exists another $[d] \supseteq M \neq K$, with $|M|$ odd, such that

$$\begin{cases} \sum_{i\in K} \varphi_i = (|K|-1)\pi + \sum_{i\notin K} \varphi_i \\ \sum_{i\in M} \varphi_i = (|M|-1)\pi + \sum_{i\notin M} \varphi_i, \end{cases}$$

hence summing these equalities gives

$$2 \left(\sum_{i\in K\cap M} \varphi_i - \sum_{i\in K^c\cap M^c} \varphi_i \right) = (|K| + |M| - 2)\pi.$$

Now, if we suppose that $K^c \cap M^c = \varnothing$, meaning that $K \cup M = [d]$, it is easy to show that $2|K\cap M| \le |K| + |M| - 2$. Indeed, $|K \cap M| \le |K| \wedge |M|$, with equality if and only if $M \subseteq K$ (or viceversa): in this case we must have $|K| \geq |M| + 2$, otherwise they would be equal, hence $2|K \cap M| \leq 2(|K| \wedge |M|) = 2|M|$ while $|K|+|M|-2 \geq |M|+2+|M|-2 = 2|M|$. If the equality is not reached, $2|K\cap M| \leq 2(|K|\wedge|M|-1) = 2|M|-2$, while $|K| + |M| - 2 \geq |M| + |M| - 2 = 2|M| - 2$. This shows that $2|K \cap M| \leq |K| + |M| - 2$, which implies that the equality above can be verified only if $\varphi_i = \pi$ for all $i \in K \cap M$, which is excluded from our analysis. Furthermore, if $K^c \cap M^c \neq \varnothing$, this is even worse unless $\varphi_i = 0$ for all $i \in K^c \cap M^c$, which is again excluded from our analysis. This completes the proof of the fact that for all $i \in [d]$, if $R > 0$, exactly one between λ_i and λ_{d_i} is greater than zero. As a corollary, we have that the optimal $(\varphi_1^*, \dots, \varphi_d^*)$ satisfies

$$1 - \lambda^* = |\rho_i - \lambda^* \cos\varphi_i^*|, \text{ for all } i \in [d],$$

as required.

(ii) The primal set is strictly feasible, hence we know that R is attained in the dual set, which is enough to prove existence. As for uniqueness, suppose there exists two optimal Σ_1, Σ_2 such that

$$\begin{cases} \Sigma_{\mathbb{S}} = \lambda^* A\Sigma_1 + (1-\lambda^*)\Sigma'_{\mathbb{S}} \\ \Sigma_{\mathbb{S}} = \lambda^* A\Sigma_2 + (1-\lambda^*)\Sigma''_{\mathbb{S}}. \end{cases}$$

This implies that for all $\mu \in (0, 1)$

$$\Sigma_{\mathbb{S}} = \lambda^* A(\mu\Sigma_1 + (1-\mu)\Sigma_2) + (1-\lambda^*)(\mu\Sigma'_{\mathbb{S}} + (1-\mu)\Sigma''_{\mathbb{S}}),$$

meaning that $\mu\Sigma_1 + (1-\mu)\Sigma_2$ is optimal. By the optimality of Σ_1 and Σ_2 we must have that $\Sigma'_{\mathbb{S}}$ and $\Sigma''_{\mathbb{S}}$ are maximally incompatible, which means they must all be singular, as stated in Example 4. Now, observe that if there exists $i \in [d]$ such that $\Sigma'_{\{i,i+1\}} \neq \Sigma''_{\{i,i+1\}}$,

$$\mu\Sigma'_{\{i,i+1\}} + (1-\mu)\Sigma''_{\{i,i+1\}} = \begin{pmatrix} 1 & \pm(2\mu - 1) \\ \pm(2\mu-1) & 1 \end{pmatrix},$$

which means that $\mu\Sigma'_{\mathbb{S}} + (1-\mu)\Sigma''_{\mathbb{S}}$ can never be maximally incompatible since $\mu \in (0, 1)$. This implies that $\Sigma'_{\mathbb{S}} = \Sigma''_{\mathbb{S}}$, which in turn implies that $\boldsymbol{\varphi}_1^* = \boldsymbol{\varphi}_2^*$. As for the continuity of $\boldsymbol{\varphi}^*(\theta_1, \dots, \theta_d)$, observe that $1 - \lambda^* = |\rho_j - \lambda^* \cos\varphi_j^*|$, for all $j \in [d]$ in point (i) means that there exist $\{\epsilon_j = \pm 1\}_{j\in[d]}$ such that

$$\lambda^* = \frac{1 - \epsilon_j \cos\theta_j}{1 - \epsilon_j \cos\varphi_j^*}, \quad \text{ for all } j \in [d].$$

Now, let $\left\{\boldsymbol{\theta}^{(n)} = (\theta_{1,n}, \dots, \theta_{d,n})\right\}_{n\in\mathbb{N}} \longrightarrow \boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$, and consider the associated sequence of optimal $\left\{\boldsymbol{\varphi}_n^* = (\varphi_{1,n}^*, \dots, \varphi_{d,n}^*)\right\}_{n\in\mathbb{N}}$, meaning that

$$\lambda_n^* = \frac{1 - \epsilon_{j,n} \cos\theta_{j,n}}{1 - \epsilon_{j,n} \cos\varphi_{j,n}^*}, \quad \text{ for all } j \in [d].$$

Taking the limit on both sides, since λ^* is continuous due to Proposition 4 (ii), we get that

$$\lambda^* = \frac{1 \pm \cos\theta_j}{1 \pm \cos\left(\lim_n \varphi^*_{j,n}\right)}, \qquad \text{for all } j \in [d].$$

This shows that the limit $\lim_n \varphi^*_{j,n}$ exists, and by uniqueness (i), we can conclude that $\lim_n \varphi^*_{j,n} = \varphi^*_j$, showing that $\boldsymbol{\varphi}$ is continuous.

As for (iii), supposing without loss of generality that $\theta_1 = \max_{i\in[d]} \theta_i$, with at most $\theta_1 > \pi/2$, observe that incompatibility is equivalent to having $\theta_1 - \sum_{i=2}^d \theta_i > 0$, hence in order to make λ^* as big as possible we should choose $\rho_j = \lambda^* \cos\varphi^*_j + (1-\lambda^*)$ for all $j \in \{2,\dots,d\}$, and $\rho_1 = \lambda^*\cos\varphi^*_1 - (1-\lambda^*)$. This would imply that the optimal choice of signs for a general d-cycle is $\epsilon_d = (-1, +\mathbf{1}^T_{d-1})$, and this turns out to be true indeed. To see why, start by considering the case $d = 3$, and observe that from (ii) we know that there exists a unique $K \subseteq [3]$ with $|K|$ odd such that

$$\sum_{\substack{|K| odd\\ i\in K}} \mu_K = \sum_{\substack{|K| odd\\ i\in K^c}} \mu_K.$$

The possible values of K are $\{1\},\{2\},\{3\}$ and $\{1,2,3\}$, and these are associated to the vectors of signs $(-1,1,1)$,$(1,-1,1)$, $(1,1,-1)$ and $(-1,-1,-1)$, respectively. Hence, in order to prove the statement it is necessary and sufficient to show that $K = \{1\}$ leads to the optimal λ^*, meaning that $\lambda^*_1 \geq \lambda^*_2$ and $\lambda^*_1 \geq \lambda^*_4$, where

$$\lambda^*_1 = \frac{1+\cos\theta_1}{1+\cos\varphi^*_1} = \frac{1-\cos\theta_2}{1-\cos\varphi^*_2} = \frac{1-\cos\theta_3}{1-\cos(\varphi^*_1-\varphi^*_2)},$$

$$\lambda^*_2 = \frac{1-\cos\theta_1}{1-\cos\tilde\varphi^*_1} = \frac{1+\cos\theta_2}{1+\cos\tilde\varphi^*_2} = \frac{1-\cos\theta_3}{1-\cos(\tilde\varphi^*_2-\tilde\varphi^*_1)},$$

$$\lambda^*_4 = \frac{1+\cos\theta_1}{1+\cos\tilde\varphi^*_1} = \frac{1+\cos\theta_2}{1+\cos\tilde\varphi^*_2} = \frac{1+\cos\theta_3}{1+\cos(2\pi-\tilde\varphi^*_1-\tilde\varphi^*_2)}.$$

Now, for $\lambda^*_1 < \lambda^*_2$ to be true it is necessary to have

$$\begin{cases} \cos(\varphi^*_1-\varphi^*_2) < \cos(\tilde\varphi^*_1-\tilde\varphi^*_2) \\ \\ \cos\tilde\varphi^*_1 > 1 - \frac{1-\cos\theta_1}{1+\cos\theta_1}(1+\cos\varphi^*_1) \\ \\ \cos\tilde\varphi^*_2 < -1 + \frac{1+\cos\theta_2}{1-\cos\theta_2}(1-\cos\varphi^*_2), \end{cases}$$

with $(\varphi^*_1,\varphi^*_2),(\tilde\varphi^*_1,\tilde\varphi^*_2) \in [0,\pi]^2$ that need to simultaneously satisfy

$$\begin{cases} \cos\varphi^*_2 = 1 - \frac{1-\cos\theta_2}{1+\cos\theta_1}(1+\cos\varphi^*_1) \\ \\ \cos\tilde\varphi^*_1 = 1 - \frac{1-\cos\theta_1}{1+\cos\theta_2}(1+\cos\tilde\varphi^*_2), \end{cases} \quad \text{and} \quad \begin{cases} \varphi^*_1 \leq \theta_1, \qquad \varphi^*_2 \geq \theta_2 \\ \\ \tilde\varphi^*_1 \geq \theta_1, \qquad \tilde\varphi^*_2 \leq \theta_2, \end{cases}$$

to ensure $R(\Sigma_{\mathbb{S}_3}) \in [0,1]$. This system of inequalities has no solution in $(\tilde\varphi^*_1,\tilde\varphi^*_2)$ for fixed $(\varphi^*_1,\varphi^*_2)$ and

$\theta_1 > \theta_2$. The same reasoning shows that $\lambda_1^* < \lambda_4^*$ can never be satisfied as well, showing that the optimal choice of signs for $d = 3$ is indeed $\epsilon_3 = (-1, +1, +1)$. For general d, it is sufficient to proceed by induction: indeed, suppose that $\epsilon_j = (-1, +\mathbf{1}_{j-1}^T)$ for all $j \in \{3, \dots, d-1\}$, and consider $\lambda^* = \lambda^*(\theta_d)$ as a function of θ_d, for fixed $\theta_1, \dots, \theta_{d-1}$. This function is continuous over $[0, \theta_1 - \sum_{i=2}^{d-1} \theta_i)$, because is the restriction of $\lambda^* = 1 - R(\Sigma_{\mathbb{S}_d})$, which is continuous by Proposition 4 (ii), onto the last coordinate. Now, $\lambda^*(\theta_d)$ uniquely identifies a vector of signs for varying $\theta_d \in [0, \theta_1 - \sum_{i=2}^{d-1} \theta_i)$, call it $\epsilon(\theta_d)$, taking values in $\{+1, -1\}^d$. This vector is unique because we supposed the cycle to be incompatible, hence either $g(\varphi_i)$ or $h(\varphi_i)$ in the KKT conditions has a strictly positive optimal gap, so that there exists a unique $\mu_K \neq 0$. We will show that this vector is constant for all $\theta_d \in [0, \theta_1 - \sum_{i=2}^{d-1} \theta_i)$, that is to say that each component of $\epsilon(\theta_d)$ is continuous in $\theta_d \in [0, \theta_1 - \sum_{i=2}^{d-1} \theta_i)$. Indeed, consider without loss of generality the first component of $\epsilon(\theta_d)$, and suppose by contradiction that $\epsilon(\theta_d)_1$ is not continuous in $\tilde{\theta}_d$. This implies that there exists a sequence of angles $\{\theta_{d,n}\}$ converging to $\tilde{\theta}_d$ such that

$$\lim_{n \to +\infty} \epsilon(\theta_{d,n})_1 = \epsilon_{\lim} = -\epsilon(\tilde{\theta}_d)_1.$$

Without loss of generality, assume $\epsilon_{\lim} = +1$ and $\epsilon(\tilde{\theta}_d)_1 = -1$. But we must have by continuity

$$\begin{aligned}\frac{1 + \cos\theta_1}{1 + \cos\varphi_1^*} &= \frac{1 - \epsilon(\tilde{\theta}_d)_1 \cos\theta_1}{1 - \epsilon(\tilde{\theta}_d)_1 \cos\varphi_1^*} = \lambda^*(\theta_d) = \lim_{n \to +\infty} \lambda^*(\theta_{d,n}) = \lim_{n \to +\infty} \frac{1 - \epsilon(\theta_{d,n})_1 \cos\theta_1}{1 - \epsilon(\theta_{d,n})_1 \cos\varphi_{1,n}^*} \\ &= \frac{1 - \cos\theta_1 \lim_{n \to +\infty} \epsilon(\theta_{d,n})_1}{1 - \cos\left(\lim_{n \to +\infty} \varphi_{1,n}^*\right) \lim_{n \to +\infty} \epsilon(\theta_{d,n})_1} = \frac{1 - \epsilon_{\lim} \cos\theta_1}{1 - \epsilon_{\lim} \cos\varphi_1^*} = \frac{1 - \cos\theta_1}{1 - \cos\varphi_1^*},\end{aligned}$$

where $\cos\varphi_{1,n}^*$ and $\cos\varphi_1^*$ are the $(1,2)$-th entries of the optimal matrix of the dual in $\theta_{d,n}$ and $\tilde{\theta}_d$, respectively. This implies that $\lambda^*(\tilde{\theta}_d)$ admits both representations, one with the plus sign, and one with the minus sign, and this can happen only if the cycle is compatible, which cannot be the case for $\tilde{\theta}_d \in [0, \theta_1 - \sum_{i=2}^{d-1} \theta_i)$ since $\theta_1 > \sum_{i=2}^{d} \theta_i$. This means that $\epsilon(\theta_d)_j$ is continuous for all $j \in [d]$ for varying $\theta_d \in [0, \theta_1 - \sum_{i=2}^{d-1} \theta_i)$, which implies that the vector $\epsilon(\theta_d)$ is constant on $[0, \theta_1 - \sum_{i=2}^{d-1} \theta_i)$, so that the behaviour of $\epsilon(\theta_d)$ is uniquely determined by $\epsilon(0)$. But we do know that

$$\epsilon(0) = (\epsilon_{d-1}, \epsilon(0)_d) = (-1, +\mathbf{1}_{d-2}^T, \epsilon(0)_d)$$

due to Proposition 10 and the induction step: this, together with the fact that $\epsilon(0)_d = +1$ in order to make $\Sigma'_{\mathbb{S}}$ maximally incompatible, completes the proof.

□

Proof of Proposition 13. We will prove the statement by induction, with base cases $d = 3$, and $d = 4$:

$d = 3$ Suppose without loss of generality that θ_1, θ_2 are bounded away from singularity. Also, assume without loss of generality that $\theta_2, \theta_3 \leq \pi/2$, using Proposition 11, so that incompatibility means $\theta_1 > \theta_2 + \theta_3$. We will prove the base case

$$R(\Sigma_{\mathbb{S}_3}) \gtrsim \theta_2 - \theta_1 - \theta_3.$$

by showing that

$$R(\Sigma_{\mathbb{S}_3}) \geq \frac{\theta_1 - \theta_2 - \theta_3}{\theta_1 - \theta_2} \frac{\cos\theta_2 - \cos\theta_1}{2},$$

and since $\cos\theta_2 - \cos\theta_1 \gtrsim_c \theta_1 - \theta_2$ being bounded away from singularity, the result would follow. Now, fix arbitrary θ_1, θ_2 satisfying the hypothesis of the statement, and suppose $\theta_1 - \theta_2 \leq \pi/2$. Observe that for $\theta_3 = 0$ and $\theta_3 = \theta_1 - \theta_2$ the lower bound is satisfied with equality sign due to Example 3 and Barrett's characterisation (6), respectively. Now, call

$$h = \frac{\theta_1 - \theta_2 - \theta_3}{\theta_1 - \theta_2},$$

and observe that the thesis is equivalent to

$$\lambda^* = 1 - R(\Sigma_{\mathbb{S}_d}) \leq 1 - \frac{h}{2}(\cos\theta_2 - \cos\theta_1),$$

Now, thanks to the KKT representation of the optimal λ^*, in order to have $\lambda^* > 1 - h(\cos\theta_2 - \cos\theta_1)/2$ we must have

$$\begin{cases} \cos\theta_1 < \cos\varphi_1^* < \frac{\cos\theta_1 + h(\cos\theta_2 - \cos\theta_1)/2}{1 - h(\cos\theta_2 - \cos\theta_1)/2} \\ \cos\theta_2 > \cos\varphi_2^* > \frac{\cos\theta_2 - h(\cos\theta_2 - \cos\theta_1)/2}{1 - h(\cos\theta_2 - \cos\theta_1)/2} \\ \cos\theta_3 > \cos\varphi_3^* > \frac{\cos\theta_3 - h(\cos\theta_2 - \cos\theta_1)/2}{1 - h(\cos\theta_2 - \cos\theta_1)/2}, \end{cases}$$

with $\varphi_1^* = \varphi_2^* + \varphi_3^*$ due to Proposition 12 (iii). We see numerically that this system of inequalities can never be satisfied for $\theta_3 \in (0, \theta_1 - \theta_2)$. Finally, taking into account all the possible ways in which a generic 3-cycle can be reduced to a 3-cycle with at most one negative correlation, as stated in Proposition 11, we get

$$R(\Sigma_{\mathbb{S}_3}) \gtrsim \max(\theta_1 - \theta_2 - \theta_3, \theta_2 - \theta_1 - \theta_3, \theta_3 - \theta_1 - \theta_2, \theta_1 + \theta_2 + \theta_3 - 2\pi).$$

$d = 4$ Suppose without loss of generality that one of the two angles bounded away from singularity is θ_1, with $\theta_1 > \theta_2 + \theta_3 + \theta_4$, and $\theta_2, \theta_3, \theta_4 \in [0, \pi/2]$. As shown in Figure 13, there are two possible cases: the first one (on the left) is when the two angles bounded away from singularity are adjacent, and the second one (on the right) when they are opposite to each other. We will use the following lemma:

Lemma 16. *Consider the d-cycle with $\mathbb{S}_d = \{\{1,2\}, \dots, \{d,1\}\}$ and $\Sigma_{\mathbb{S}_d} := (\Sigma_{\{1,2\}}, \cdots, \Sigma_{\{d,1\}})$. Then, for every optimal Σ of the dual problem, i.e. $\Sigma_{\mathbb{S}_d} = \lambda^* A\Sigma + (1-\lambda^*)\Sigma'_{\mathbb{S}_d}$, and for every $d \geq 4$,*

$$R(\Sigma_{\mathbb{S}_d}) \geq R(B_{\mathbb{S}_{d-1}}(\phi)) + R(E_{\mathbb{S}_3}(\phi)), \quad \forall \phi \in [\lambda^*\Sigma_{1,d-1} - R, \lambda^*\Sigma_{1,d-1} + R],$$

where $B_{\mathbb{S}_{d-1}}(\phi) = (\Sigma_{\{1,2\}}, \cdots, \Sigma_{\{d-2,d-1\}}, \Sigma_{\{d-1,1\}}(\phi))$, $E_{\mathbb{S}_3}(\phi) = (\Sigma_{\{d-1,d\}}, \Sigma_{\{d,1\}}, \Sigma_{\{d-1,1\}}(\phi))$ and $\Sigma_{\{d-1,1\}}(\phi)$ is the 2×2 correlation matrix with off-diagonal entries equal to ϕ.

In the first case, suppose we add an edge between $(2,4)$ with correlation $\cos(\theta_3 + \theta_4)$. We first show that this is a valid choice of ϕ to invoke Proposition 16. In this regard, observe that $R(E_{\mathbb{S}_3}(\phi)) = 0$ for all $\phi \in [\cos(\theta_{d-1} + \theta_d), \cos(\theta_{d-1} - \theta_d)]$, hence, since we proved $R(E_{\mathbb{S}_{d-1}}(\lambda^*\Sigma_{1,d-1} + R)) = 0$ in the

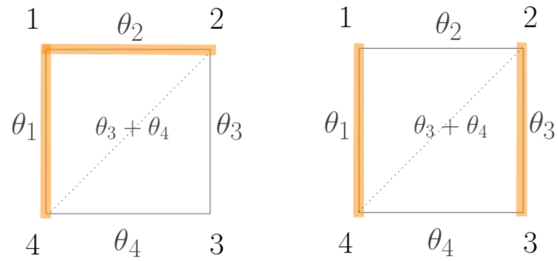

Figure 13: The two possible configurations of the two angles bounded away from singularity when $d = 4$. On the left, the two angles are adjacent, while on the right they are opposite to each other.

proof of the lemma above, we must have $\cos(\theta_{d-1}+\theta_d) \le \lambda^*\Sigma_{1,d-1}+R$. Similarly, $\cos(\theta_1 - \sum_{i=2}^{d-2}\theta_i) \ge \lambda^*\Sigma_{1,d-1} - R$. This, together with the fact that $\cos(\theta_1 - \sum_{i=2}^{d-2}\theta_i) \le \cos(\theta_{d-1}+\theta_d)$ since $\theta_1 > \sum_{i=2}^{d}\theta_i$, allows concluding that $\lambda^*\Sigma_{1,d-1} - R \le \cos(\theta_1 - \sum_{i=2}^{d-2}\theta_i) < \cos(\theta_{d-1}+\theta_d) \le \lambda^*\Sigma_{1,d-1}+R$. Now, Proposition 16 ensures that

$$R(\theta_1,\theta_2,\theta_3,\theta_4) \ge R(\theta_1,\theta_2,\theta_3+\theta_4),$$

and since θ_1, θ_2 are bounded away from singularity, we can employ the lower bound we found for $d = 3$, and conclude

$$R(\Sigma_{\mathbb{S}_4}) \ge \frac{\theta_1-\theta_2-(\theta_3+\theta_4)}{\theta_1-\theta_2}\frac{\cos\theta_2-\cos\theta_1}{2},$$

which gives the desired result. In the second case, we can proceed in the same way as before, and get

$$R(\Sigma_{\mathbb{S}_4}) \ge \frac{\theta_1-\theta_2-(\theta_3+\theta_4)}{\theta_1-(\theta_3+\theta_4)}\frac{\cos(\theta_3+\theta_4)-\cos\theta_1}{2}.$$

Now, if $\sin^2(\theta_3+\theta_4) \ge c$ we are done, otherwise, θ_4 must be bounded away from singularity. Indeed, since $\theta_3, \theta_4 \in [0, \pi/2]$, and $\sin^2(\theta_3) \ge c$ by hypothesis, in order to have $\sin^2(\theta_3+\theta_4) < c$ we must have $\sin^2\theta_4 \ge 1-c$. Now, since we can assume that c is small enough, say $c \le 1/2$, we conclude $\sin^2\theta_4 \ge 1-c > c$. This implies that θ_4 is bounded away from singularity, and since it is adjacent to θ_1, we can proceed as in the first case to get the desired result.

$d \ge 5$ Suppose again without loss of generality that θ_1 is bounded away from singularity, and call θ_j the other one. Now, since $d \ge 5$, we can find $k \ne 1, j$ such that θ_k, θ_{k+1} are not necessarily assumed to be bounded away from singularity. Then, proceeding as before, thanks to Proposition 16, we have

$$R(\theta_1,\dots,\theta_d) \ge R(\theta_1,\dots,\theta_{k-1},\theta_k+\theta_{k+1},\theta_{k+2},\dots,\theta_d),$$

so that the induction step gives immediately that

$$\begin{aligned} R(\theta_1,\dots,\theta_d) &\ge R(\theta_1,\dots,\theta_{k-1},\theta_k+\theta_{k+1},\theta_{k+2},\dots,\theta_d) \\ &\ge c'\left(\theta_1-(\theta_k+\theta_{k+1})-\sum_{i\ne 1,j,k,k+1}\theta_i\right) = c'\left(\theta_1-\sum_{i=2}^{d}\theta_i\right), \end{aligned}$$

where c' is a constant depending on c only. Finally, taking into account all the possible ways in which a generic d-cycle can be reduced to a d-cycle with at most one negative correlation, as stated in Proposition 11, we get

$$R(\Sigma_{\mathbb{S}_d}) \geq c' \max_{\substack{K\subseteq[d]\\|K| \text{ odd}}} \left(\sum_{i\in K} \theta_i - (|K|-1)\pi - \sum_{i\in K^c} \theta_i \right),$$

where $c' > 0$ depends only on c.

□

Proof of Lemma 16. Suppose without loss of generality that $\theta_d \geq \theta_{d-1}$, and that $\theta_1 = \max_{i\in[d]} \theta_i$, with $\theta_2, \ldots, \theta_d \leq \pi/2$. Let $R \equiv R(\Sigma_{\mathbb{S}_d})$, and let

$$\Sigma_{\mathbb{S}_d} = (1-R)A\Sigma + R\Sigma'_{\mathbb{S}} = \lambda^* A\Sigma + (1-\lambda^*)\Sigma'_{\mathbb{S}}$$

be a (not necessarily unique) dual representation of $\Sigma_{\mathbb{S}_d}$, and denote by $\Sigma_{1,d-1}$ the entry $(1, d-1)$ of Σ. We will prove the statement in three steps:

1. $R(B_{\mathbb{S}_{d-1}}(\lambda^*\Sigma_{1,d-1} + R)) \leq R(\Sigma_{\mathbb{S}_d})$ and $R(E_{\mathbb{S}_3}(\lambda^*\Sigma_{1,d-1} + R)) = 0$,
2. $R(E_{\mathbb{S}_3}(\lambda^*\Sigma_{1,d-1} - R)) \leq R(\Sigma_{\mathbb{S}_d})$ and $R(B_{\mathbb{S}_{d-1}}(\lambda^*\Sigma_{1,d-1} - R)) = 0$,
3. $\Xi(\phi) := R(B_{\mathbb{S}_{d-1}}(\phi)) + R(E_{\mathbb{S}_3}(\phi))$ is convex for all $\phi \in [-1, 1]$.

1. As for the fact that $R(E_{\mathbb{S}_3}(\lambda^*\Sigma_{1,d-1} + R)) = 0$ observe that

$$\begin{aligned}
A^*E_{\mathbb{S}_3}(\lambda^*\Sigma_{1,d-1} + R) - I_3 &= \begin{pmatrix} 1 & \lambda^*\Sigma_{1,d-1} + R & \rho_d \\ \lambda^*\Sigma_{1,d-1} + R & 1 & \rho_{d-1} \\ \rho_d & \rho_{d-1} & 1 \end{pmatrix} = \\
&= \begin{pmatrix} 1 & \lambda^*\Sigma_{1,d-1} + R & \lambda^*\Sigma_{1,d} + R \\ \lambda^*\Sigma_{1,d-1} + R & 1 & \lambda^*\Sigma_{d-1,d} + R \\ \lambda^*\Sigma_{1,d} + R & \lambda^*\Sigma_{d-1,d} + R & 1 \end{pmatrix} \\
&= \lambda^* \begin{pmatrix} 1 & \Sigma_{1,d-1} & \Sigma_{1,d} \\ \Sigma_{1,d-1} & 1 & \Sigma_{d-1,d} \\ \Sigma_{1,d} & \Sigma_{d-1,d} & 1 \end{pmatrix} + (1-\lambda^*) \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix},
\end{aligned}$$

where the second equality follows from the optimal choice of signs given in Proposition 12 (iii) under the hypothesis $\theta_1 = \max_{i\in[d]} \theta_i$, with $\theta_2, \ldots, \theta_d \leq \pi/2$. This implies $R(E_{\mathbb{S}_3}(\lambda^*\Sigma_{1,d-1} + R)) = 0$ since $A^*E_{\mathbb{S}_3}(\lambda^*\Sigma_{1,d-1}+R)-I_3$, which is the 3×3 correlation matrix whose 2×2 marginals are precisely those in $E_{\mathbb{S}_3}(\lambda^*\Sigma_{1,d-1} + R)$, is PSD being the sum of two PSD matrices. As for $R(B_{\mathbb{S}_{d-1}}(\lambda^*\Sigma_{1,d-1} + R)) \leq R(\Sigma_{\mathbb{S}_d})$, observe that, if $\Sigma_{\mathbb{S}_d} = \lambda^* A\Sigma + (1-\lambda^*)\Sigma'_{\mathbb{S}_d}$, then

$$B_{\mathbb{S}_{d-1}}(\lambda^*\Sigma_{1,d-1} + R) = \lambda^* A\Sigma_{|(-d)} + (1-\lambda^*)\Sigma''_{B_{\mathbb{S}_{d-1}}},$$

where

$$\Sigma''_{B_{\mathbb{S}_{d-1}}} = (\Sigma'_{\{1,2\}}, \ldots, \Sigma'_{\{d-1,d-2\}}, \mathbf{1}_2\mathbf{1}_2^T),$$

which is maximally incompatible. To see why, observe that $\Sigma'_{\mathbb{S}_d}$ is maximally incompatible by definition of the dual representation, and since $\theta_1 = \max_{i\in[d]} \theta_i$, with $\theta_2, \ldots, \theta_d \leq \pi/2$, Proposition 12 (iii) ensures that

$$\Sigma'_{\mathbb{S}_d} = (-\mathbf{1}_2\mathbf{1}_2^T, +\mathbf{1}_2\mathbf{1}_2^T, \ldots, +\mathbf{1}_2\mathbf{1}_2^T),$$

which leads to

$$\Sigma''_{B_{\mathbb{S}_{d-1}}} = (-\mathbf{1}_2\mathbf{1}_2^T, +\mathbf{1}_2\mathbf{1}_2^T, \ldots, +\mathbf{1}_2\mathbf{1}_2^T).$$

This shows that $\Sigma_{|(-d)}$ is feasible for $B_{\mathbb{S}_{d-1}}(\lambda^*\Sigma_{1,d-1} + R)$, and implies that $R(B_{\mathbb{S}_{d-1}}(\lambda^*\Sigma_{1,d-1} + R)) \leq R(\Sigma_{\mathbb{S}_d})$.

2. The arguments in the proof above can be followed *mutatis mutandis* to show that $R(E_{\mathbb{S}_3}(\lambda^*\Sigma_{1,d-1} - R)) \leq R(\Sigma_{\mathbb{S}_d})$ and $R(B_{\mathbb{S}_{d-1}}(\lambda^*\Sigma_{1,d-1} - R)) = 0$.

3. In order to show that $R(\Sigma_{\mathbb{S}_d}) \geq R(B_{\mathbb{S}_{d-1}}(\phi)) + R(E_{\mathbb{S}_3}(\phi)), \forall \phi \in I$, we will make use of the fact that R is convex and continuous, as stated in Proposition 4 (i) (ii), i.e.

$$R\left(\mu\Sigma_{\mathbb{S}}^{(1)} + (1-\mu)\Sigma_{\mathbb{S}}^{(2)}\right) \leq \mu R\left(\Sigma_{\mathbb{S}}^{(1)}\right) + (1-\mu)R\left(\Sigma_{\mathbb{S}}^{(2)}\right), \text{ for all } \mu \in [0,1].$$

Now, define

$$\Xi(\phi) = R(B_{\mathbb{S}_{d-1}}(\phi)) + R(E_{\mathbb{S}_3}(\phi)), \text{ for all } \phi \in I = [-1,1].$$

It is easy to see that $\Xi(\phi)$ is convex in I since, for all $\phi_1, \phi_2 \in I$, for all $\mu \in [0,1]$,

$$\begin{aligned}\Xi(\mu\phi_1 + (1-\mu)\phi_2) &= R(B_{\mathbb{S}_{d-1}}(\mu\phi_1 + (1-\mu)\phi_2) + R(E_{\mathbb{S}_3}(\mu\phi_1 + (1-\mu)\phi_2))\\ &= R(\mu B_{\mathbb{S}_{d-1}}(\phi_1) + (1-\mu)B_{\mathbb{S}_{d-1}}(\phi_2)) + R(\mu E_{\mathbb{S}_3}(\phi_1) + (1-\mu)E_{\mathbb{S}_3}(\phi_2))\\ &\leq \mu R(B_{\mathbb{S}_{d-1}}(\phi_1)) + (1-\mu)R(B_{\mathbb{S}_{d-1}}(\phi_2)) + \mu R(E_{\mathbb{S}_3}(\phi_1)) + (1-\mu)R(E_{\mathbb{S}_3}(\phi_2))\\ &= \mu\Xi(\phi_1) + (1-\mu)\Xi(\phi_2).\end{aligned}$$

This, implies that, for all $\mu \in [0,1]$,

$$\begin{aligned}R &\geq \mu R(E_{\mathbb{S}_{d-1}}(\lambda^*\Sigma_{1,d-1} - R)) + (1-\mu)R(B_{\mathbb{S}_{d-1}}(\lambda^*\Sigma_{1,d-1} - R))\\ &= \mu\Xi(\lambda^*\Sigma_{1,d-1} - R) + (1-\mu)\Xi(\lambda^*\Sigma_{1,d-1} + R) \geq \Xi(\lambda^*\Sigma_{1,d-1} + 1 - 2\mu R)\\ &=: \Xi(\phi) = R(B_{\mathbb{S}_{d-1}}(\phi)) + R(E_{\mathbb{S}_3}(\phi)),\end{aligned}$$

for all $\phi \in [\lambda^*\Sigma_{1,d-1} - R, \lambda^*\Sigma_{1,d-1} + R]$, as claimed. For general angles $(\theta_1, \ldots, \theta_d)$, it is sufficient to perform the transformation outlined in Proposition 11, find ϕ and I as above, and perform the inverse transformation.

As we can see from Figure 14, this reduction corresponds to adding an edge in correspondence to $\{1, d-1\}$, so that the d-cycle $\Sigma_{\mathbb{S}_d}$ is divided into two smaller cycles, $B_{\mathbb{S}_d}(\phi)$ of dimension $d-1$, and $E_{\mathbb{S}_d}(\phi)$ of dimension 3. The result ensures the possibility of adding a correlation $\rho_{1,d-1} = \phi$ for the edge $\{1, d-1\}$ to make $B_{\mathbb{S}_d}(\phi)$ and $E_{\mathbb{S}_d}(\phi)$ maximally compatible, or better, at least as compatible as the original d-cycle, since $R(\Sigma_{\mathbb{S}_d}) \geq R(B_{\mathbb{S}_d}(\phi)) + R(E_{\mathbb{S}_d}(\phi))$. □

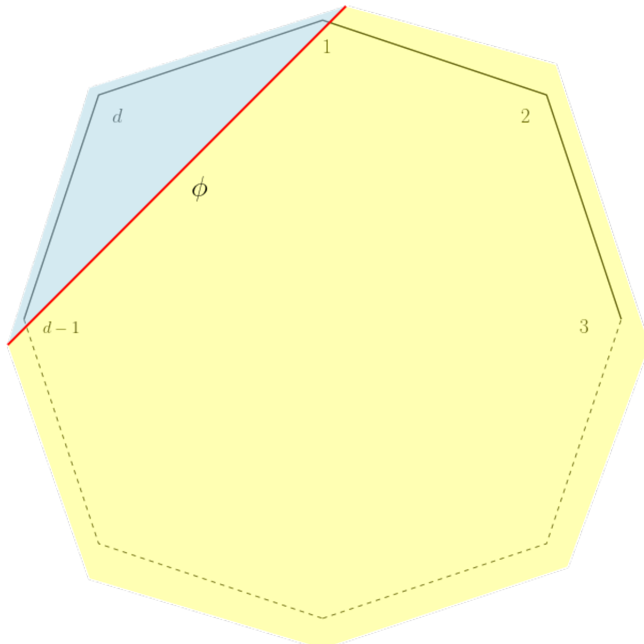

Figure 14: Illustration of Proposition 16. We split the original d-cycle into two smaller cycles, adding the extra edge $\{1, d-1\}$ with associated correlation ϕ. We end up with a $(d-1)$-cycle $B_{\mathbb{S}_{d-1}}(\phi)$ in yellow, and a 3-cycle $E_{\mathbb{S}_3}(\phi)$ in blue, such that $R(\Sigma_{\mathbb{S}_d}) \geq R(B_{\mathbb{S}_{d-1}}(\phi)) + R(E_{\mathbb{S}_3}(\phi))$ for all $\phi \in [\lambda^* \Sigma_{1,d-1} - R, \lambda^* \Sigma_{1,d-1} + R]$.

6.4 Proofs for Section 4.2

Proof of Theorem 14. We prove the result by considering the two cases $d \geq 42$ and $d < 42$ separately. For the first of these, as in the proof of Theorem 9, we focus on a proper subset of the testing problem: in this case, we construct distributions with consistent sequences of covariance matrices, so that our hypotheses reduce to statements about the compatibility of the associated sequences of correlation matrices. Formally, we look at

$$H_0 : R(\Sigma_{\mathbb{S}}) = 0 \quad \text{vs.} \quad H_1(\rho) : R(\Sigma_{\mathbb{S}}) > \rho,$$

for fixed $\rho > 0$, and we aim at finding the smallest of such ρ's for which we can have non-trivial power. Again, referring to the same $\mathcal{P}_{\mathbb{S}}(0)$ and $\mathcal{P}_{\mathbb{S}}(\rho)$ defined in Section 4, this corresponds to assuming that $\sigma_{\mathbb{S}}^2$ is always consistent, and constructing prior distributions just on $\{P_{\mathbb{S}} : R(\Sigma_{\mathbb{S}}) = 0\}$ and$\{P_{\mathbb{S}} : R(\Sigma_{\mathbb{S}}) > \rho\}$. We specialise $\Sigma_{\mathbb{S}}$ to be

$$\Sigma_{\mathbb{S}} = \left\{ \begin{pmatrix} I_d & P \\ P^T & I_d \end{pmatrix}, \begin{pmatrix} I_d & -P \\ -P^T & I_d \end{pmatrix}, \begin{pmatrix} I_d & \beta I_d \\ \beta I_d & I_d \end{pmatrix} \right\},$$

and since $R(\Sigma_{\mathbb{S}}) \geq \frac{3}{4d} \sum_{j=1}^{d} (\sigma_j^2(P) - \frac{1-\beta}{2})_+$ by Proposition 15, it is sufficient to study the testing problems

$$H_0' : \sum_{j=1}^{d} (\theta_i)_+ \leq 0 \quad \text{vs.} \quad H_1'(\rho') : \sum_{j=1}^{d} (\theta_i)_+ > \rho',$$

where $\theta_i = \sigma_j^2(P) - \frac{1-\beta}{2}$, find the smallest ρ' for which we have non-trivial power, and use the relationship $\rho = \frac{3}{4d}\rho'$. More precisely, focusing on the latter testing problem, we want to lower bound the minimax testing risk

$$\rho^*(n_{\mathbb{S}}, \eta) := \inf \left\{ \rho > 0 : \exists \varphi_{\mathbb{S}} \in \Psi_{\mathbb{S}} : \sup_{P_{\mathbb{S},0} \in \mathcal{P}_{\mathbb{S}}(0)} P_{\mathbb{S},0}^{\otimes n_{\mathbb{S}}}(\varphi_{\mathbb{S}} = 1) + \sup_{P_{\mathbb{S},1} \in \mathcal{P}_{\mathbb{S}}(\rho)} P_{\mathbb{S},1}^{\otimes n_{\mathbb{S}}}(\varphi_{\mathbb{S}} = 0) \le \eta \right\}$$

where $\mathcal{P}_{\mathbb{S}}(0) = \{P_{\mathbb{S}} : (\mathrm{Corr}(P_{\mathbb{S}}), \mathrm{Var}(P_{\mathbb{S}})) = (\Sigma_{\mathbb{S}}, \mathbf{1}_{\mathbb{S}}) \text{ and } R(\Sigma_{\mathbb{S}}) = 0\}$, $\mathcal{P}_{\mathbb{S}}(\rho) = \{P_{\mathbb{S}} : (\mathrm{Corr}(P_{\mathbb{S}}), \mathrm{Var}(P_{\mathbb{S}})) = (\Sigma_{\mathbb{S}}, \mathbf{1}_{\mathbb{S}}) \text{ and } R(\Sigma_{\mathbb{S}}) > \rho\}$, and $\Psi_{\mathbb{S}}$ is the set of sequence of tests coherent with $\mathbb{S}$. To this aim, we start by defining two prior distributions μ_0, μ_1 for P. First, there exist two measures ν_0, ν_1 with matching moments up to the M-th order such that

I. $\mathrm{supp}(\nu_0) \subseteq [-b, 0], \quad \mathrm{supp}(\nu_1) \subseteq [-b, 0] \cup \left\{\frac{b}{4M^2}\right\}$

II. $\nu_1\left(\left\{\frac{b}{4M^2}\right\}\right) \ge \frac{1}{2}$

III. $\forall k \in \{0, 1, \ldots, M\} : \int z^k \nu_0(\,\mathrm{d}z) = \int z^k \nu_1(\,\mathrm{d}z)$.

This is proved in Juditsky and Nemirovski (2002) using ideas from the theory of best polynomial approximation. A different, but closely related version, was proved in Cai and Low (2011) using similar techniques. Such prior distributions have been extensively used in the minimax literature in the last decade, and led to optimal, or nearly-optimal, lower bounds in many problems of interest such as optimal estimation of nonsmooth functionals (Cai and Low, 2011; Jiao et al., 2016; Thépaut and Verzelen, 2021), testing MCAR in a fully nonparametric setting (Berrett and Samworth, 2023), and testing convex hypothesis (Blanchard et al., 2018). Let $\mathcal{U}(d)$ denote the (normalised) Haar measure over the Lie group of orthogonal matrices $SO(d) = \{U \in \mathbb{R}^{d,d} : U^T U = UU^T = I_d\}$, and let ν_0, ν_1 the distributions with matching moments up to the order M defined above. Calling δ_0 the Dirac measure in zero, we define μ_i to be the distribution of $P = U^T \Lambda U$, where $U \sim \mathcal{U}(d)$, and $\Lambda = \mathrm{diag}(\sigma_{1:d})$, with $\sigma_{1:d} \sim \nu_i^{\otimes \lceil d/2 \rceil} \otimes \delta_0^{\otimes (d - \lceil d/2 \rceil)}$, for $i \in \{0, 1\}$. Observe now that the support of μ_1 also contains elements in $(\mathcal{P}_{\mathbb{S}}(\rho))^C$. In order to overcome this, we will consider the conditional measure $\mu_1|\xi$, where ξ is the event

$$\xi = \left\{ \sum_{i=1}^d \mathbb{1}_{\{\mu_i = b/4M^2\}} \ge \frac{d}{3} \right\},$$

which ensures that $\mu_1|\xi$ is supported on the alternative. Now, given P, we use the shorthand

$$N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}} \equiv N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}(P) = \left(N^{\otimes n_1}\left(0, \begin{pmatrix} I_d & P \\ P^T & I_d \end{pmatrix}\right), N^{\otimes n_2}\left(0, \begin{pmatrix} I_d & -P \\ -P^T & I_d \end{pmatrix}\right), N^{\otimes n_3}\left(0, \begin{pmatrix} I_d & \beta I_d \\ \beta I_d & I_d \end{pmatrix}\right) \right).$$

The marginal distribution of the data when P is generated according to $\mu_1|\xi$ is then given by the mixture distribution

$$\mathbb{E}_{\mu_1|\xi} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}} = \left(\mathbb{E}_{\mu_1|\xi} N^{\otimes n_1}\left(0, \begin{pmatrix} I_d & P \\ P^T & I_d \end{pmatrix}\right), \mathbb{E}_{\mu_1|\xi} N^{\otimes n_2}\left(0, \begin{pmatrix} I_d & -P \\ -P^T & I_d \end{pmatrix}\right), N^{\otimes n_3}\left(0, \begin{pmatrix} I_d & \beta I_d \\ \beta I_d & I_d \end{pmatrix}\right) \right).$$

Similarly, the marginal distribution of the data when P is generated according to μ_i is then given by the mixture distribution

$$\mathbb{E}_{\mu_i} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}} = \left(\mathbb{E}_{\mu_i} N^{\otimes n_1}\left(0, \begin{pmatrix} I_d & P \\ P^T & I_d \end{pmatrix}\right), \mathbb{E}_{\mu_i} N^{\otimes n_2}\left(0, \begin{pmatrix} I_d & -P \\ -P^T & I_d \end{pmatrix}\right), N^{\otimes n_3}\left(0, \begin{pmatrix} I_d & \beta I_d \\ \beta I_d & I_d \end{pmatrix}\right)\right),$$

for $i \in \{0,1\}$. For every test sequence $\varphi_{\mathbb{S}} \in \Psi_{\mathbb{S}}$, and for prior distributions $\mu_0, \mu_1|\xi$, we can bound the total error probability as

$$\begin{aligned}
\mathcal{R}(n_{\mathbb{S}}, \rho') &= \sup_{P_{\mathbb{S},0} \in \mathcal{P}_{\mathbb{S}}(0)} P_{\mathbb{S},0}^{\otimes n_{\mathbb{S}}}(\varphi_{\mathbb{S}} = 1) + \sup_{P_{\mathbb{S},1} \in \mathcal{P}_{\mathbb{S}}(\rho)} P_{\mathbb{S},1}^{\otimes n_{\mathbb{S}}}(\varphi_{\mathbb{S}} = 0) \\
&\geq \mathbb{E}_{\mu_0} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}(\varphi_{\mathbb{S}} = 1) + \mathbb{E}_{\mu_1|\xi} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}(\varphi_{\mathbb{S}} = 0) \\
&= \mathbb{E}_{\mu_0} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}(\varphi_{\mathbb{S}} = 1) + 1 - \frac{\mathbb{E}_{\mu_1} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}(\{\varphi_{\mathbb{S}} = 1\} \cap \xi)}{\mu_1(\xi)} \\
&\geq \mathbb{E}_{\mu_0} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}(\varphi_{\mathbb{S}} = 1) + 1 - \frac{10}{9} \mathbb{E}_{\mu_1} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}(\varphi_{\mathbb{S}} = 1) \\
&\geq \mathbb{E}_{\mu_0} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}(\varphi_{\mathbb{S}} = 1) + \frac{10}{9} \mathbb{E}_{\mu_1} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}(\varphi_{\mathbb{S}} = 0) - \frac{1}{9} \\
&\geq \mathbb{E}_{\mu_0} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}(\varphi_{\mathbb{S}} = 1) + \mathbb{E}_{\mu_1} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}(\varphi_{\mathbb{S}} = 0) - \frac{1}{9} \\
&\geq 1 - \mathrm{TV}\left(\mathbb{E}_{\mu_0} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}, \mathbb{E}_{\mu_1} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}\right) - \frac{1}{9}.
\end{aligned}$$

The second inequality follows from Hoeffding's inequality, which ensures that for all $d \geq 42$,

$$\mu_1(\xi) \geq \frac{9}{10},$$

since $\mu_1(\{b/4M^2\}) \geq 1/2$ by II. This shows that it is now sufficient to control the total variation distance between the marginals of $N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}$ with respect to the unconditional priors μ_0, μ_1 by finding $b/4M^2$ such that $\mathrm{TV}\left(\mathbb{E}_{\mu_0} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}, \mathbb{E}_{\mu_1} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}\right) \leq 1/2 - 1/9$. This would imply that $\mathcal{R}(n_{\mathbb{S}}, \rho') \geq 1/2$, and would lead to

$$\rho' = \frac{d}{3} \frac{b}{4M^2},$$

where the extra $d/3$ factor comes from conditioning on the event ξ. Hence, let us now focus on controlling $\mathrm{TV}\left(\mathbb{E}_{\mu_0} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}, \mathbb{E}_{\mu_1} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}\right)$. We have

$$\begin{aligned}
&\mathrm{TV}\left(\mathbb{E}_{\mu_0} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}, \mathbb{E}_{\mu_1} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}\right) \\
&\quad = \mathrm{TV}\left\{ \left(\mathbb{E}_{\mu_0} N^{\otimes n_1}\left(0, \begin{pmatrix} I_d & P \\ P^T & I_d \end{pmatrix}\right), \mathbb{E}_{\mu_0} N^{\otimes n_2}\left(0, \begin{pmatrix} I_d & -P \\ -P^T & I_d \end{pmatrix}\right), N^{\otimes n_3}\left(0, \begin{pmatrix} I_d & \beta I_d \\ \beta I_d & I_d \end{pmatrix}\right)\right), \right. \\
&\qquad\qquad \left. \left(\mathbb{E}_{\mu_1} N^{\otimes n_1}\left(0, \begin{pmatrix} I_d & P \\ P^T & I_d \end{pmatrix}\right), \mathbb{E}_{\mu_1} N^{\otimes n_2}\left(0, \begin{pmatrix} I_d & -P \\ -P^T & I_d \end{pmatrix}\right), N^{\otimes n_3}\left(0, \begin{pmatrix} I_d & \beta I_d \\ \beta I_d & I_d \end{pmatrix}\right)\right) \right\} \\
&\quad = \mathrm{TV}\left\{ \left(\mathbb{E}_{\mu_0} N^{\otimes n_1}\left(0, \begin{pmatrix} I_d & P \\ P^T & I_d \end{pmatrix}\right), \mathbb{E}_{\mu_0} N^{\otimes n_2}\left(0, \begin{pmatrix} I_d & -P \\ -P^T & I_d \end{pmatrix}\right)\right), \right.
\end{aligned}$$

$$\left(\mathbb{E}_{\mu_1} N^{\otimes n_1}\left(0, \begin{pmatrix} I_d & P \\ P^T & I_d \end{pmatrix}\right), \mathbb{E}_{\mu_1} N^{\otimes n_2}\left(0, \begin{pmatrix} I_d & -P \\ -P^T & I_d \end{pmatrix}\right)\right)\right\}$$

$$\leq \mathrm{TV}\left\{\mathbb{E}_{\mu_0} N^{\otimes n_1}\left(0, \begin{pmatrix} I_d & P \\ P^T & I_d \end{pmatrix}\right), \mathbb{E}_{\mu_1} N^{\otimes n_1}\left(0, \begin{pmatrix} I_d & P \\ P^T & I_d \end{pmatrix}\right)\right\}$$

$$+ \mathrm{TV}\left\{\mathbb{E}_{\mu_0} N^{\otimes n_2}\left(0, \begin{pmatrix} I_d & -P \\ -P^T & I_d \end{pmatrix}\right), \mathbb{E}_{\mu_1} N^{\otimes n_2}\left(0, \begin{pmatrix} I_d & -P \\ -P^T & I_d \end{pmatrix}\right)\right\}.$$

Dealing with such μ_0, μ_1 is not straightforward, due to the presence of the integrals with respect to the Haar measure. Nonetheless, following similar ideas as in Thépaut and Verzelen (2021), we upper bound the total variance distance above using the following two lemmata, where we suppose P to be symmetric.

Lemma 17. *Let P be symmetric, with spectral decomposition $P = U^T \Lambda U$. Let $\mathcal{U}(d)$ denote the (normalised) Haar measure over the Lie group of orthogonal matrices $SO(d) = \{U \in \mathbb{R}^{d,d} : U^T U = UU^T = I_d\}$, and let ν_0, ν_1 the distributions with matching moments up to the order M defined above. Denote by μ_i the distribution of $U^T \Lambda U$, where $U \sim \mathcal{U}(d)$, and $\Lambda = \mathrm{diag}(\sigma_{1:d})$, with $\sigma_{1:d} \sim \nu_i^{\otimes \lceil d/2 \rceil} \otimes \delta_0^{\otimes (d - \lceil d/2 \rceil)}$. Then*

$$\mathrm{TV}\left\{\mathbb{E}_{\mu_0} N^{\otimes n}\left(0, \begin{pmatrix} I_d & P \\ P^T & I_d \end{pmatrix}\right), \mathbb{E}_{\mu_1} N^{\otimes n}\left(0, \begin{pmatrix} I_d & P \\ P^T & I_d \end{pmatrix}\right)\right\}$$

$$\leq \lceil d/2 \rceil \, \mathrm{TV}\left\{\mathbb{E}_{\tilde{\pi}_0}\left\{N^{\otimes n}\left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & \eta u u^T \\ \eta u u^T & I_d \end{pmatrix}\right)\right\}, \mathbb{E}_{\tilde{\pi}_1}\left\{N^{\otimes n}\left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & \eta' u' u'^T \\ \eta' u' u'^T & I_d \end{pmatrix}\right)\right\}\right\},$$

where $\tilde{\pi}_0$ (resp. $\tilde{\pi}_1$) is the distribution of $\eta u u^T$ (resp. $\eta' u' u'^T$), where $\eta \sim \nu_0$ (resp. $\eta' \sim \nu_1$) and $u = (u_{d'}, \mathbf{0}_{d-d'}^T)$ (resp. $u' = (u'_{d'}, \mathbf{0}_{d-d'}^T)$) is such that $u_{d'}$ (resp $u'_{d'}$) is a uniform sample from the d'-dimensional sphere $\mathcal{S}^{d'-1} = \{x \in \mathbb{R}^{d'} : \|x\|_2 = 1\}$, with $d' = d + 1 - \lceil d/2 \rceil$.

Lemma 18. *With the same notation as above, then*

$$\mathrm{TV}^2\left\{\mathbb{E}_{\tilde{\pi}_0}\left\{N^{\otimes n}\left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & \eta u u^T \\ \eta u u^T & I_d \end{pmatrix}\right)\right\}, \mathbb{E}_{\tilde{\pi}_1}\left\{N^{\otimes n}\left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & \eta' u' u'^T \\ \eta' u' u'^T & I_d \end{pmatrix}\right)\right\}\right\}$$

$$\leq \sum_{k=M+1}^{\infty} \binom{k+n-1}{n-1} \mathbb{E}[u_1^{2k}] \left(\int \eta^k [\nu_0(d\eta) - \nu_1(d\eta)]\right)^2.$$

Applying these lemmata, it follows that

$$\mathrm{TV}\left\{\mathbb{E}_{\mu_0} N^{\otimes n_1}\left(0, \begin{pmatrix} I_d & P \\ P^T & I_d \end{pmatrix}\right), \mathbb{E}_{\mu_1} N^{\otimes n_1}\left(0, \begin{pmatrix} I_d & P \\ P^T & I_d \end{pmatrix}\right)\right\}$$

$$\leq \lceil d/2 \rceil \, \mathrm{TV}\left\{\mathbb{E}_{\tilde{\pi}_0}\left\{N^{\otimes n_1}\left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & \eta u u^T \\ \eta u u^T & I_d \end{pmatrix}\right)\right\}, \mathbb{E}_{\tilde{\pi}_1}\left\{N^{\otimes n_1}\left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & \eta' u' u'^T \\ \eta' u' u'^T & I_d \end{pmatrix}\right)\right\}\right\}$$

$$\leq \lceil d/2 \rceil \sqrt{\sum_{k=M+1}^{\infty} \binom{k+n_1-1}{n_1-1} \mathbb{E}[u_1^{2k}] \left(\int \eta^k [\nu_0(d\eta) - \nu_1(d\eta)]\right)^2},$$

where the first inequality comes from Lemma 17, and the second from Lemma 18. Here u_1 is the first coordinate of a uniform random vector in the d'-dimensional unit sphere, and ν_0, ν_1 are the distributions with matching moments up the order M defined above. Now, observe that $u_1^2 \overset{d}{=} Z_1^2 / \sum_{i=1}^{d'} Z_i^2$ where $Z_i \overset{\text{i.i.d.}}{\sim} N(0,1)$, due to the fact that the standard normal distribution is isotropic. Hence $u_1^2 \sim \text{Beta}(\frac{d'-1}{2}, \frac{1}{2})$ since if $X \sim \chi^2(\alpha)$ and $Y \sim \chi^2(\beta)$ are independent, then $\frac{X}{X+Y} \sim \text{Beta}\left(\frac{\alpha}{2}, \frac{\beta}{2}\right)$. It follows that

$$\begin{aligned}\mathbb{E}[u_1^{2k}] &= \int_{-1}^{1} \frac{u^{2k}(1-u^2)^{\frac{d'-3}{2}}}{B(\frac{d'-1}{2}, \frac{1}{2})} du = \int_0^1 \frac{v^{k-\frac{1}{2}}(1-v)^{\frac{d'-1}{2}-1}}{B(\frac{d'-1}{2}, \frac{1}{2})} dv \\ &= \frac{B(\frac{d'-1}{2}, k+\frac{1}{2})}{B(\frac{d'-1}{2}, \frac{1}{2})} = \frac{\Gamma(k+\frac{1}{2})}{\Gamma(\frac{1}{2})} \frac{\Gamma(\frac{d'}{2})}{\Gamma(\frac{d'}{2}+k)}.\end{aligned}$$

Moreover

$$\left(\int \eta^k[\nu_0(d\eta) - \nu_1(d\eta)]\right)^2 \le \left(b^k\left(1 + \frac{1}{4^k M^{2k}}\right)\right)^2 \le 4b^{2k}.$$

If we choose $b^2 = \frac{d'}{4n}$, we have

$$\begin{aligned}&\sum_{k=M+1}^{\infty} \binom{k+n-1}{n-1} \mathbb{E}[u_1^{2k}] \left(\int \eta^k[\nu_0(d\eta) - \nu_1(d\eta)]\right)^2 \\ &\quad \le 4 \sum_{k=M+1}^{\infty} \binom{k+n-1}{n-1} \frac{\Gamma(k+\frac{1}{2})}{\Gamma(\frac{1}{2})} \frac{\Gamma(\frac{d'}{2})}{\Gamma(\frac{d'}{2}+k)} b^{2k} \\ &\quad = 4 \sum_{k=M+1}^{\infty} \frac{\Gamma(n+k)}{\Gamma(n)\Gamma(k+1)} \frac{\Gamma(k+\frac{1}{2})}{\Gamma(\frac{1}{2})} \frac{\Gamma(\frac{d'}{2})}{\Gamma(\frac{d'}{2}+k)} b^{2k} \\ &\quad = 4 \sum_{k=M+1}^{\infty} \frac{\Gamma(k+\frac{1}{2})}{\Gamma(k+1)\Gamma(\frac{1}{2})} \frac{\Gamma(\frac{d'}{2})}{\Gamma(\frac{d'}{2}+k)} \frac{\Gamma(n+k)}{\Gamma(n)} \frac{(d')^k}{4^k n^k} \\ &\quad = 4 \sum_{k=M+1}^{\infty} \frac{\Gamma(k+\frac{1}{2})}{\Gamma(k+1)\Gamma(\frac{1}{2})} \frac{(\frac{d'}{2})^k \Gamma(\frac{d'}{2})}{\Gamma(\frac{d'}{2}+k)} \frac{\Gamma(n+k)}{n^k \Gamma(n)} \frac{1}{2^k} \\ &\quad \le 4 \sum_{k=M+1}^{\infty} \frac{(\frac{d'}{2})^k \Gamma(\frac{d'}{2})}{\Gamma(\frac{d'}{2}+k)} \frac{\Gamma(n+k)}{n^k \Gamma(n)} 2^{-k} \le 4 \cdot 2^{-(M+1)} = 2^{-(M-1)}.\end{aligned}$$

The second inequality follows from the fact that $\frac{\Gamma(k+\frac{1}{2})}{\Gamma(k+1)\Gamma(\frac{1}{2})} \le 1$, while the third one follows from the fact that

$$\frac{(\frac{d}{2})^k \Gamma(\frac{d}{2})}{\Gamma(\frac{d}{2}+k)} \frac{\Gamma(n+k)}{n^k \Gamma(n)} \le 1.$$

Indeed, writing ψ for the digamma function, the function $x \mapsto \log \frac{a^k \Gamma(a)}{\Gamma(a+k)}$ has derivative $\psi(a) - \psi(a+k) + k/a > 0$, and is therefore increasing. Thus, whenever $n \ge d/2 \ge d'/2$ the inequality follows. Summing up, if we set $b/4M^2 = d/\{4(n_1 \wedge n_2)\}$, we have that

$$\begin{aligned}\text{TV}\left(\mathbb{E}_{\mu_0} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}, \mathbb{E}_{\mu_1} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}\right) &\le \text{TV}\left\{\mathbb{E}_{\mu_0} N^{\otimes n_1}\left(0, \begin{pmatrix} I_d & P \\ P^T & I_d \end{pmatrix}\right), \mathbb{E}_{\mu_1} N^{\otimes n_1}\left(0, \begin{pmatrix} I_d & P \\ P^T & I_d \end{pmatrix}\right)\right\} \\ &\quad + \text{TV}\left\{\mathbb{E}_{\mu_0} N^{\otimes n_2}\left(0, \begin{pmatrix} I_d & -P \\ -P^T & I_d \end{pmatrix}\right), \mathbb{E}_{\mu_1} N^{\otimes n_2}\left(0, \begin{pmatrix} I_d & -P \\ -P^T & I_d \end{pmatrix}\right)\right\}\end{aligned}$$

$$\leq 2\lceil d/2\rceil 2^{-\frac{M-1}{2}} \leq (d+1)2^{-\frac{M-1}{2}},$$

which is upper bounded by $1/2 - 1/9$ if and only if

$$M > 2\frac{\log(d+1) - \log(1/2-1/9)}{\log 2} + 1.$$

Hence, this shows that

$$\frac{b}{4M^2} \gtrsim \sqrt{\frac{d}{(n_1 \wedge n_2)\log^4 d}}$$

is sufficient to have $\mathrm{TV}\left(\mathbb{E}_{\mu_0} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}, \mathbb{E}_{\mu_1} N_{\mathbb{S}}^{\otimes n_{\mathbb{S}}}\right) \leq 1/2 - 1/9$, which implies that $\mathcal{R}(n_{\mathbb{S}}, \rho') = \mathcal{R}(n_{\mathbb{S}}, \frac{d}{3}\frac{b}{4M^2}) \geq 1/2$. This allows us to conclude that

$$\rho^* \geq \frac{3}{4d}\rho' = \frac{b}{16M^2} \gtrsim \sqrt{\frac{d}{(n_1 \wedge n_2)\log^4 d}}.$$

We finally turn to the simpler case $d < 42$. It is sufficient to work as in the proof of Theorem 9, and show that testing the consistency of the variances represents the essential difficulty of the problem. More precisely, we will show that testing

$$H_0 : V(\Sigma_{\mathbb{S}}) = 0 \quad \text{vs.} \quad H_1(\rho) : V(\Sigma_{\mathbb{S}}) > \rho,$$

requires at least a separation of the order $\sqrt{1/(n_1 \wedge n_2)}$, and since $T = V(\sigma_{\mathbb{S}}^2) + R(\Sigma_{\mathbb{S}})$, the statement will follow. To this aim, we bound the total error probability by choosing

$$P_{\mathbb{S},0}^{\otimes n_{\mathbb{S}}} = \left(N^{\otimes n_1}(\mathbf{0}_{2d}, I_{2d}), N^{\otimes n_2}(\mathbf{0}_{2d}, I_{2d}), N^{\otimes n_3}(\mathbf{0}_{2d}, I_{2d})\right) \in \mathcal{P}_{\mathbb{S}}(0),$$

and

$$P_{\mathbb{S},1}^{\otimes n_{\mathbb{S}}} = \left(N^{\otimes n_1}(\mathbf{0}_{2d}, \mathrm{diag}(1+\delta, 1\ldots,1)), N^{\otimes n_2}(\mathbf{0}_{2d}, \mathrm{diag}(1-\delta, 1\ldots,1)), N^{\otimes n_3}(\mathbf{0}_{2d}, I_{2d})\right) \in \mathcal{P}_{\mathbb{S}}(\delta).$$

We have

$$\begin{aligned}
\mathcal{R}(n_{\mathbb{S}}, \delta) &= \inf_{\varphi_{\mathbb{S}}}\left\{\sup_{P_{\mathbb{S},0}\in\mathcal{P}_{\mathbb{S}}(0)} P_{\mathbb{S},0}^{\otimes n_{\mathbb{S}}}(\varphi - \mathbb{S} = 1) + \sup_{P_{\mathbb{S},1}\in\mathcal{P}_{\mathbb{S}}(\delta)} P_{\mathbb{S},1}^{\otimes n_{\mathbb{S}}}(\varphi - \mathbb{S} = 0)\right\} \geq 1 - \mathrm{TV}\big(P_{\mathbb{S},0}^{\otimes n_{\mathbb{S}}}, P_{\mathbb{S},1}^{\otimes n_{\mathbb{S}}}\big) \\
&\geq \big(1 - \mathrm{TV}\big(N^{\otimes n_1}(\mathbf{0}_{2d}, I_{2d}), N^{\otimes n_1}(\mathbf{0}_{2d}, \mathrm{diag}(1+\delta, 1\ldots,1))\big)\big) - \frac{1}{2} \\
&+ \big(1 - \mathrm{TV}\big(N^{\otimes n_2}(\mathbf{0}_{2d}, I_{2d}), N^{\otimes n_2}(\mathbf{0}_{2d}, \mathrm{diag}(1-\delta, 1\ldots,1))\big)\big) - \frac{1}{2}.
\end{aligned}$$

Now, if $P_0 := N^{\otimes n_1}(\mathbf{0}_{2d}, I_{2d}), P_1 := N^{\otimes n_1}(\mathbf{0}_{2d}, \mathrm{diag}(1+\delta, 1\ldots,1))$,

$$4\,\mathrm{TV}(P_0, P_1)^2 \leq \chi^2(P_0, P_1) = \int \left(\frac{dP_1}{dP_0}\right)^2 dP_0 - 1$$

$$= \frac{1}{(1+\delta)^{n_1}} \int \left(\prod_{i=1}^{n_1} \exp\left\{ \frac{\delta}{2(1+\delta)} x_{i1}^2 \right\} \right)^2 dP_0 - 1$$
$$= \frac{1}{(1+\delta)^{n_1}} \int \prod_{i=1}^{n_1} \exp\left\{ \frac{\delta}{1+\delta} x_{i1}^2 \right\} dP_0 - 1$$
$$= \frac{1}{(1+\delta)^{n_1}} \int \prod_{i=1}^{n_1} (2\pi)^{-d} \exp\left\{ -\frac{1}{2} x^T \operatorname{diag}\left(\frac{1-\delta}{1+\delta}, 1, \ldots, 1 \right) x \right\} dx - 1$$
$$= (1-\delta^2)^{-n_1/2} - 1 \le e^{n_1 \delta^2/2} - 1,$$

from which we see that $\mathrm{TV}(P_0, P_1) \le 1/4$ if $\delta \le \sqrt{2\log(5/4)/n_1}$. The same holds true for $Q_0 := N^{\otimes n_2}(\mathbf{0}_{2d}, I_{2d}), Q_1 := N^{\otimes n_2}(\mathbf{0}_{2d}, \operatorname{diag}(1-\delta, 1\ldots, 1))$, which shows that $\mathcal{R}(n_{\mathbb{S}}, \delta) \ge 1/2$ if $\delta = \sqrt{\frac{2\log(5/4)}{n_1 \wedge n_2}}$. The above bound on the total variation distance demonstrates that we may choose $\delta = \sqrt{2\log(5/4)/n_1 \wedge n_2}$, and hence that we have

$$\rho^* \ge \delta = \sqrt{\frac{2\log(5/4)}{n_1 \wedge n_2}},$$

as claimed. □

Proof of Lemma 17. Let $\varphi(z)$ denote the density of the d-dimensional Gaussian law with respect to the Lebesgue measure. By the triangle inequality we have that

$$\mathrm{TV}\left\{ \mathbb{E}_{\mu_0} \left\{ N^{\otimes n} \left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & U^T \Lambda U \\ U^T \Lambda U & I_d \end{pmatrix} \right) \right\}, \mathbb{E}_{\mu_1} \left\{ N^{\otimes n} \left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & U^T \Lambda U \\ U^T \Lambda U & I_d \end{pmatrix} \right) \right\} \right\}$$
$$\le \sum_{j=0}^{\lceil d/2 \rceil - 1} \mathrm{TV}\left\{ \mathbb{E}_{\pi_j} \left\{ N^{\otimes n} \left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & U^T \Lambda U \\ U^T \Lambda U & I_d \end{pmatrix} \right) \right\}, \mathbb{E}_{\pi_{j+1}} \left\{ N^{\otimes n} \left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & U^T \Lambda U \\ U^T \Lambda U & I_d \end{pmatrix} \right) \right\} \right\},$$

where π_j is distribution of $U^T \Lambda U$, where $U \sim \mathcal{U}(d)$ is common for all π_j, while $\Lambda = \operatorname{diag}(\sigma_{1:d})$, with $\sigma_{1:d} \sim \nu_0^{\otimes(\lceil d/2 \rceil - j)} \otimes \nu_1^{\otimes j} \otimes \delta_0^{\otimes \lfloor d/2 \rfloor}$, for $j \in \{0, \ldots, \lceil d/2 \rceil - 1\}$. Observe that $\pi_0 = \mu_0$ and $\pi_{\lceil d/2 \rceil} = \mu_1$, so that this inequality essentially interpolates μ_0 and μ_1 with $\lceil d/2 \rceil$ intermediate measures such that, for every $j \in \{0, \ldots, \lceil d/2 \rceil - 1\}$, π_j differs from π_{j+1} only for the distribution of σ_j in Λ. Now, consider a generic $j \in \{0, \ldots, \lceil d/2 \rceil - 1\}$ and define $S := \{1, \ldots, \lceil d/2 \rceil - j - 1, \lceil d/2 \rceil - j + 1, \ldots, \lceil d/2 \rceil\}$. We will show that we can bound each term of the summation above by

$$\mathrm{TV}\left\{ \mathbb{E}_{\tilde{\pi}_0} \left\{ N^{\otimes n} \left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & \eta u u^T \\ \eta u u^T & I_d \end{pmatrix} \right) \right\}, \mathbb{E}_{\tilde{\pi}_1} \left\{ N^{\otimes n} \left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & \eta' u' u'^T \\ \eta' u' u'^T & I_d \end{pmatrix} \right) \right\} \right\},$$

with $\tilde{\pi}_0, \tilde{\pi}_1$ defined in the statement, and this would conclude the proof. To this aim, observe that if

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & U^T \Lambda U \\ U^T \Lambda U & I_d \end{pmatrix} \right),$$

then

$$\begin{cases} X|Y \sim N(U^T\Lambda UY, I_d - U^T\Lambda^2 U) \\ \\ Y \sim N(\mathbf{0}_d, I_d). \end{cases}$$

This allows us to write

$$\begin{aligned} TV_0 &:= \mathrm{TV}\left\{\mathbb{E}_{\pi_j}\left\{N^{\otimes n}\left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & U^T\Lambda U \\ U^T\Lambda U & I_d \end{pmatrix}\right)\right\}, \mathbb{E}_{\pi_{j+1}}\left\{N^{\otimes n}\left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & U^T\Lambda U \\ U^T\Lambda U & I_d \end{pmatrix}\right)\right\}\right\} \\ &= \int \prod_{i=1}^n \varphi(y_i)\Bigg|\mathbb{E}_{\pi_j}\left\{\prod_{i=1}^n |I - U^T\Lambda^2 U|^{-1/2}\varphi((I - U^T\Lambda^2 U)^{-1/2}(x_i - U^T\Lambda U y_i))\right\} \\ &\qquad - \mathbb{E}_{\pi_{j+1}}\left\{\prod_{i=1}^n |I - U^T\Lambda^2 U|^{-1/2}\varphi((I - U^T\Lambda^2 U)^{-1/2}(x_i - U^T\Lambda U y_i))\right\}\Bigg| dxdy, \end{aligned}$$

where $dx = dx_1 \dots dx_n$, and similarly for dy. Now, let U_S be the restriction of U to its columns in S. By definition of π_j (resp. π_{j+1}), we can write $U^T\Lambda U$ as $U^T\Lambda U = \eta uu^T + U_S^T \operatorname{diag}(\sigma_{-j})U_S$ where $\sigma_{-j} \sim \nu_0^{\otimes(\lceil d/2\rceil - j - 1)} \otimes \delta_0 \otimes \nu_1^{\otimes j} \otimes \delta_0^{\otimes d - \lceil d/2\rceil}$, where δ_0 is the Dirac measure in 0. Write π for the distribution of $(U_S, \operatorname{diag}(\sigma_{-j}))$ and f_0 (resp. f_1) for the conditional distribution of (u, η) given $(U_S, \operatorname{diag}(\sigma_{-j}))$: this is given by $\eta \sim \nu_0$ (resp. ν_1), while $u|U_S$ is sampled uniformly from $\mathcal{S}^{d-1} \cap U_S^{\perp}$, i.e. the intersection between the d-dimensional unit sphere $\mathcal{S}^{d-1} := \{x \in \mathbb{R}^d : \|x\|_2 = 1\}$ and the orthogonal complement of the columns spanned by U_S. First, observe that $\dim(\mathcal{S}^{d-1} \cap U_S^{\perp}) = d + 1 - \lceil d/2\rceil$. Secondly, observe that for every measurable function h,

$$\mathbb{E}h(P) = \int h(P)\pi_j(dP) = \int h(P) f_0(du, d\eta)\pi\left(dU_S, d\sigma_{-j}\right),$$

and similarly for π_{j+1}. This allows to bound the TV distance above one step further as

$$\begin{aligned} TV_0 &\le \int \prod_{i=1}^n \varphi(y_i)\Bigg|\int \prod_{i=1}^n |I - U^T\Lambda^2 U|^{-1/2}\varphi((I - U^T\Lambda^2 U)^{-1/2}(x_i - U^T\Lambda U y_i) f_0(du, d\eta) \\ &\quad - \int \prod_{i=1}^n |I - U'^T\Lambda'^2 U'|^{-1/2}\varphi((I - U'^T\Lambda'^2 U')^{-1/2}(x_i - U'^T\Lambda' U' y_i) f_1(du', d\eta')\Bigg|\pi\left(dU_S, d\sigma_{-j}\right) dxdy \\ &= \int \prod_{i=1}^n \varphi(y_i)\Bigg(\int \Bigg|\int \prod_{i=1}^n |I - U^T\Lambda^2 U|^{-1/2}\varphi((I - U^T\Lambda^2 U)^{-1/2}(x_i - U^T\Lambda U y_i) f_0(du, d\eta) \\ &\quad - \int \prod_{i=1}^n |I - U'^T\Lambda'^2 U'|^{-1/2}\varphi((I - U'^T\Lambda'^2 U')^{-1/2}(x_i - U'^T\Lambda' U' y_i) f_1(du', d\eta')\Bigg| dx\Bigg)\pi\left(dU_S, d\sigma_{-j}\right) dy, \end{aligned}$$

where in the first step we used Jensen's inequality, bringing the common π outside the absolute value, while in the last step we used Fubini-Tonelli theorem with positive integrand. Consider now the innermost integral

$$\int \Bigg|\int \prod_{i=1}^n |I - U^T\Lambda^2 U|^{-1/2}\varphi((I - U^T\Lambda^2 U)^{-1/2}(x_i - U^T\Lambda U y_i) f_0(du, d\eta)$$

$$
\begin{aligned}
&- \int \prod_{i=1}^{n} |I - U'^T \Lambda'^2 U'|^{-1/2} \varphi((I - U'^T \Lambda'^2 U')^{-1/2} (x_i - U'^T \Lambda' U' y_i) f_1(du', d\eta') \Big| dx \\
&= \int \Big| \int \prod_{i=1}^{n} |I - U^T \Lambda^2 U|^{-1/2} \varphi((I - U^T \Lambda^2 U)^{-1/2} (x_i - \eta u u^T y_i - \sum_{k \neq j} \sigma_k u_k u_k^T y_i) f_0(du, d\eta) \\
&\quad - \int \prod_{i=1}^{n} |I - U'^T \Lambda'^2 U'|^{-1/2} \varphi((I - U'^T \Lambda'^2 U')^{-1/2} (x_i - \eta' u' u'^T y_i - \sum_{k \neq j} \sigma_k u_k u_k^T y_i) f_1(du', d\eta') \Big| dx
\end{aligned}
$$

for fixed U_S, σ_{-j}, y. This can be simplified to

$$
\begin{aligned}
\int \Big| \int \prod_{i=1}^{n} &|I - U^T \Lambda^2 U|^{-1/2} \varphi((I - U^T \Lambda^2 U)^{-1/2} (x_i - \eta u u^T y_i)) f_0(du, d\eta) \\
&- \int \prod_{i=1}^{n} |I - U'^T \Lambda'^2 U'|^{-1/2} \varphi((I - U'^T \Lambda'^2 U')^{-1/2} (x_i - \eta' u' u'^T y_i)) f_1(du', d\eta') \Big| dx
\end{aligned}
$$

after the change of variable $x_i' = x_i - \sum_{k \neq j} \sigma_k u_k u_k^T y_i$, for all $i \in [n]$. Now, observe that, under f_0, we have

$$
(I - U^T \Lambda^2 U)^{-1/2} = \frac{1}{\sqrt{1 - \eta^2}} u u^T + \sum_{i \neq j} \frac{1}{\sqrt{1 - \sigma_i^2}} u_i u_i^T,
$$

which yields

$$
|I - U^T \Lambda^2 U|^{-1/2} = \frac{1}{\sqrt{1 - \eta^2} \prod_{i \neq j} \sqrt{1 - \sigma_i^2}};
$$

similarly under for $(I - U'^T \Lambda'^2 U')^{-1/2}$ under π_1, with η', u' in place of η, u. Perform the change of variables

$$
x_i = \left(I - \sum_{k \neq j} \sigma_k^2 u_k u_k^T \right)^{1/2} z_i,
$$

for all $i \in [n]$, whose Jacobian is

$$
\prod_{i=1}^{n} \left| I - \sum_{k \neq j} \sigma_k^2 u_k u_k^T \right|^{1/2} = \prod_{i=1}^{n} \prod_{k \neq j} \sqrt{1 - \sigma_k^2}.
$$

We get

$$
\begin{aligned}
&\int \Big| \int \prod_{i=1}^{n} \frac{1}{\sqrt{1 - \eta^2}} \varphi\left(\left(\sum_{i \neq j} u_i u_i^T + \frac{1}{\sqrt{1 - \eta^2}} u u^T \right) z_i - \frac{\eta}{\sqrt{1 - \eta^2}} u u^T y_i \right) f_0(du, d\eta) \\
&\quad - \int \prod_{i=1}^{n} \frac{1}{\sqrt{1 - \eta'^2}} \varphi\left(\left(\sum_{i \neq j} u_i u_i^T + \frac{1}{\sqrt{1 - \eta'^2}} u' u'^T \right) z_i - \frac{\eta'}{\sqrt{1 - \eta'^2}} u' u'^T y_i \right) f_1(du', d\eta') \Big| dx \\
&= \int \Big| \int \prod_{i=1}^{n} \frac{1}{\sqrt{1 - \eta^2}} \varphi\left(\left(\sum_{i \neq j} u_i u_i^T + \frac{1}{\sqrt{1 - \eta^2}} u u^T \right) (z_i - \eta u u^T y_i) \right) f_0(du, d\eta)
\end{aligned}
$$

$$- \int \prod_{i=1}^{n} \frac{1}{\sqrt{1-\eta'^2}} \varphi \left(\left(\sum_{i \neq j} u_i u_i^T + \frac{1}{\sqrt{1-\eta'^2}} u' u'^T \right) (z_i - \eta' u' u'^T y_i) \right) f_1(du', d\eta') \Big| dx$$

$$= \int \Big| \int \prod_{i=1}^{n} \frac{1}{\sqrt{1-\eta^2}} \varphi \left(\left(I - \eta^2 u u^T \right)^{-1/2} (z_i - \eta u u^T y_i) \right) f_0(du, d\eta)$$

$$- \int \prod_{i=1}^{n} \frac{1}{\sqrt{1-\eta'^2}} \varphi \left(\left(I - \eta'^2 u' u'^T \right)^{-1/2} (z_i - \eta' u' u'^T y_i) \right) f_1(du', d\eta') \Big| dx$$

$$= \int \Big| \int \prod_{i=1}^{n} \left| I - \eta^2 u u^T \right|^{-1/2} \varphi \left(\left(I - \eta^2 u u^T \right)^{-1/2} (z_i - \eta u u^T y_i) \right) f_0(du, d\eta)$$

$$- \int \prod_{i=1}^{n} \left| I - \eta'^2 u' u'^T \right|^{-1/2} \varphi \left(\left(I - \eta'^2 u' u'^T \right)^{-1/2} (z_i - \eta' u' u'^T y_i) \right) f_1(du', d\eta') \Big| dx$$

$$= \mathrm{TV} \left\{ \mathbb{E}_{f_0} \{ N^{\otimes n} (\eta u u^T y, I - (\eta u u^T)(\eta u u^T)^T) \}, \mathbb{E}_{f_1} \{ N^{\otimes n} (\eta' u' u'^T y, I - (\eta' u' u'^T)(\eta' u' u'^T)^T) \} \right\}.$$

Coming back to the initial TV distance we wish to bound, we get that

$$\mathrm{TV} \left\{ \mathbb{E}_{\pi_j} \left\{ N^{\otimes n} \left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & U^T \Lambda U \\ U^T \Lambda U & I_d \end{pmatrix} \right) \right\}, \mathbb{E}_{\pi_{j+1}} \left\{ N^{\otimes n} \left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & U^T \Lambda U \\ U^T \Lambda U & I_d \end{pmatrix} \right) \right\} \right\}$$

$$\leq \int \mathrm{TV} \left\{ \mathbb{E}_{f_0} \{ N^{\otimes n} (\eta u u^T y, I - (\eta u u^T)(\eta u u^T)^T) \}, \mathbb{E}_{f_1} \{ N^{\otimes n} (\eta' u' u'^T y, I - (\eta' u' u'^T)(\eta' u' u'^T)^T) \} \right\}$$

$$\varphi(y) \pi \left(dU_S, d\sigma_{-j} \right) dy$$

$$= \int \mathrm{TV} \left\{ \mathbb{E}_{f_0} \left\{ N^{\otimes n} \left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & \eta u u^T \\ \eta u u^T & I_d \end{pmatrix} \right) \right\}, \mathbb{E}_{f_1} \left\{ N^{\otimes n} \left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & \eta' u' u'^T \\ \eta' u' u'^T & I_d \end{pmatrix} \right) \right\} \right\} \pi \left(dU_S, d\sigma_{-j} \right)$$

$$= \mathrm{TV} \left\{ \mathbb{E}_{\tilde{\pi}_0} \left\{ N^{\otimes n} \left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & \eta u u^T \\ \eta u u^T & I_d \end{pmatrix} \right) \right\}, \mathbb{E}_{\tilde{\pi}_1} \left\{ N^{\otimes n} \left(\mathbf{0}_{2d}, \begin{pmatrix} I_d & \eta' u' u'^T \\ \eta' u' u'^T & I_d \end{pmatrix} \right) \right\} \right\},$$

where $\tilde{\pi}_0$ (resp. $\tilde{\pi}_1$) is the distribution of $\eta u u^T$ (resp. $\eta' u' u'^T$), where $\eta \sim \nu_0$ (resp. $\eta' \sim \nu_1$) and u (resp. u') is sampled uniformly from a d'-dimensional unit sphere embedded in $\mathbb{R}^d$, with $d' = d + 1 - \lceil d/2 \rceil$. Now, since the Gaussian distribution is invariant under orthogonal transformation, we might assume that $u = (u_{d'}, \mathbf{0}_{d-d'}^T)$, with $u_{d'}$ uniformly sampled from the d'-dimensional sphere $\mathcal{S}^{d'-1}$, and the result follows. ☐

Proof of Lemma 18. Consider

$$\int \prod_{i=1}^{n} \varphi(y_i) \Big| \int \prod_{i=1}^{n} \frac{1}{\sqrt{1-\eta^2}} \varphi \left(\left(I - \eta^2 u u^T \right)^{-1/2} (z_i - \eta u u^T y_i) \right) [\tilde{\pi}_0(du, d\eta) - \tilde{\pi}_1(du, d\eta)] \Big| dz dy,$$

and observe that

$$\left(I - \eta^2 u u^T \right)^{-1} = I + \frac{\eta^2}{1-\eta^2} u u^T.$$

Hence,

$$\varphi \left(\left(I - \eta^2 u u^T \right)^{-1/2} (z_i - \eta u u^T y_i) \right) = (2\pi)^{-\frac{d}{2}} \exp \left\{ -\frac{1}{2} (z_i - \eta u u^T y_i)^T \left(I + \frac{\eta^2}{1-\eta^2} u u^T \right) (z_i - \eta u u^T y_i) \right\}$$

$$= (2\pi)^{-\frac{d}{2}} \exp\left\{ -\frac{1}{2} z_i^T z_i - \frac{1}{2}\frac{\eta^2}{1-\eta^2} z_i^T uu^T z_i - \frac{1}{2}\eta^2 y_i^T uu^T y_i - \frac{1}{2}\frac{\eta^4}{1-\eta^2} y_i^T uu^T y_i \right.$$
$$\left. + \eta z_i^T uu^T y_i + \frac{\eta^3}{1-\eta^2} z_i^T uu^T y_i \right\}$$
$$= (2\pi)^{-\frac{d}{2}} \exp\left\{ -\frac{1}{2} z_i^T z_i - \frac{1}{2}\frac{\eta^2}{1-\eta^2} z_i^T uu^T z_i - \frac{1}{2}\frac{\eta^2}{1-\eta^2} y_i^T uu^T y_i + \frac{\eta}{1-\eta^2} z_i^T uu^T y_i \right\}$$
$$= \varphi(z_i) \exp\left\{ -\frac{1}{2}\frac{\eta^2}{1-\eta^2} z_i^T uu^T z_i - \frac{1}{2}\frac{\eta^2}{1-\eta^2} y_i^T uu^T y_i + \frac{\eta}{1-\eta^2} z_i^T uu^T y_i \right\}$$
$$= \varphi(z_i) \exp\left\{ -\frac{1}{2} \langle \begin{pmatrix} \frac{\eta^2}{1-\eta^2} uu^T & -\frac{\eta}{1-\eta^2} uu^T \\ -\frac{\eta}{1-\eta^2} uu^T & \frac{\eta^2}{1-\eta^2} uu^T \end{pmatrix}, \begin{pmatrix} z_i \\ y_i \end{pmatrix} \begin{pmatrix} z_i \\ y_i \end{pmatrix}^T \rangle \right\} =: \varphi(z_i) g(\eta, u, z_i, y_i).$$

Hence,

$$TV_1 := \int \prod_{i=1}^n \varphi(y_i) \left| \int \prod_{i=1}^n \frac{1}{\sqrt{1-\eta^2}} \varphi\left(\left(I - \eta^2 uu^T\right)^{-1/2} (z_i - \eta uu^T y_i) \right) [\tilde{\pi}_0(du, d\eta) - \tilde{\pi}_1(du, d\eta)] \right| dxdy$$
$$= \int \prod_{i=1}^n \varphi(z_i)\varphi(y_i) \left| \int \prod_{i=1}^n \frac{1}{\sqrt{1-\eta^2}} g(\eta, u, z_i, y_i) [\tilde{\pi}_0(du, d\eta) - \tilde{\pi}_1(du, d\eta)] \right| dzdy$$
$$\leq \sqrt{ \int \prod_{i=1}^n \varphi(z_i)\varphi(y_i) \left[\int \prod_{i=1}^n \frac{1}{\sqrt{1-\eta^2}} g(\eta, u, z_i, y_i) [\tilde{\pi}_0(du, d\eta) - \tilde{\pi}_1(du, d\eta)] \right]^2 dzdy },$$

where we used Cauchy-Schwartz inequality in the last step. Thus, it follows that

$$TV_1^2 \leq \int \prod_{i=1}^n \varphi(z_i)\varphi(y_i) \sum_{k=0,1} \sum_{j=0,1} (-1)^{k+j}$$
$$\left(\int \int \prod_{i=1}^n \frac{1}{\sqrt{1-\eta^2}\sqrt{1-\eta'^2}} g(\eta, u, z_i, y_i) g(\eta', u', z_i, y_i) \tilde{\pi}_k(du, d\eta) \tilde{\pi}_j(du', d\eta') \right) dzdy$$
$$= \sum_{k=0,1} \sum_{j=0,1} (-1)^{k+j} \int \int (1-\eta^2)^{-n/2} (1-\eta'^2)^{-n/2}$$
$$\left(\int \prod_{i=1}^n g(\eta, u, z_i, y_i) g(\eta', u', z_i, y_i) \varphi(z_i)\varphi(y_i) dzdy \right) \tilde{\pi}_k(du, d\eta) \tilde{\pi}_j(du', d\eta')$$
$$= \sum_{k=0,1} \sum_{j=0,1} (-1)^{k+j} \int \int (1-\eta^2)^{-n/2} (1-\eta'^2)^{-n/2}$$
$$\left(\prod_{i=1}^n \int g(\eta, u, z_i, y_i) g(\eta', u', z_i, y_i) \varphi(z_i)\varphi(y_i) dz_i dy_i \right) \tilde{\pi}_k(du, d\eta) \tilde{\pi}_j(du', d\eta'),$$

where in the second equality we used Fubini-Tonelli's theorem to change the order of integration, and Fubini's theorem to factorise independent integrands in the last one. Let us consider a generic

$$\int g(\eta, u, z_i, y_i) g(\eta', u', z_i, y_i) \varphi(z_i)\varphi(y_i) dz_i dy_i,$$

bearing in mind that $u = (u_{d'}, \mathbf{0}_{d-d'}^T)$, $u' = (u'_{d'}, \mathbf{0}_{d-d'}^T)$, with $u_{d'}, u'_{d'}$ being independent and uniform samples from the d'-dimensional unit sphere, where $d' = d + 1 - \lceil d/2 \rceil$. We have

$$\begin{aligned}
&\int g(\eta, u, z_i, y_i) g(\eta', u', z_i, y_i) \varphi(z_i)\varphi(y_i) dz_i dy_i \\
&= \int (2\pi)^{-d} \exp\left\{ -\frac{1}{2} \langle \begin{pmatrix} \frac{\eta^2}{1-\eta^2} uu^T & -\frac{\eta}{1-\eta^2} uu^T \\ -\frac{\eta}{1-\eta^2} uu^T & \frac{\eta^2}{1-\eta^2} uu^T \end{pmatrix} + \right. \\
&\qquad\qquad \left. + \begin{pmatrix} \frac{\eta'^2}{1-\eta'^2} u'u'^T & -\frac{\eta'}{1-\eta'^2} u'u'^T \\ -\frac{\eta'}{1-\eta'^2} u'u'^T & \frac{\eta'^2}{1-\eta'^2} u'u'^T \end{pmatrix} + \begin{pmatrix} I_d & \boldsymbol{O}_d \\ \boldsymbol{O}_d & I_d \end{pmatrix}, \begin{pmatrix} z_i \\ y_i \end{pmatrix} \begin{pmatrix} z_i \\ y_i \end{pmatrix}^T \rangle \right\} \\
&= \int (2\pi)^{-d} \exp\left\{ -\frac{1}{2} \langle \begin{pmatrix} I_d + \frac{\eta^2}{1-\eta^2} uu^T + \frac{\eta'^2}{1-\eta'^2} u'u'^T & -\frac{\eta}{1-\eta^2} uu^T - \frac{\eta'}{1-\eta'^2} u'u'^T \\ -\frac{\eta}{1-\eta^2} uu^T - \frac{\eta'}{1-\eta'^2} u'u'^T & I_d + \frac{\eta^2}{1-\eta^2} uu^T \frac{\eta'^2}{1-\eta'^2} u'u'^T \end{pmatrix}, \begin{pmatrix} z_i \\ y_i \end{pmatrix} \begin{pmatrix} z_i \\ y_i \end{pmatrix}^T \rangle \right\} \\
&=: \int (2\pi)^{-d} \exp\left\{ -\frac{1}{2} \langle K, \begin{pmatrix} z_i \\ y_i \end{pmatrix} \begin{pmatrix} z_i \\ y_i \end{pmatrix}^T \rangle \right\} = |K|^{-1/2}.
\end{aligned}$$

Now K takes the form

$$K = \begin{pmatrix} I_d + \frac{\eta^2}{1-\eta^2} uu^T + \frac{\eta'^2}{1-\eta'^2} u'u'^T & -\frac{\eta}{1-\eta^2} uu^T - \frac{\eta'}{1-\eta'^2} u'u'^T \\ -\frac{\eta}{1-\eta^2} uu^T - \frac{\eta'}{1-\eta'^2} u'u'^T & I_d + \frac{\eta^2}{1-\eta^2} uu^T \frac{\eta'^2}{1-\eta'^2} u'u'^T \end{pmatrix}.$$

It is straightforward to show that

$$|K| = \frac{(1 - (u^T u')^2 \eta\eta')^2}{(1-\eta^2)(1-\eta'^2)} \stackrel{d}{=} \frac{(1 - u_1^2 \eta\eta')^2}{(1-\eta^2)(1-\eta'^2)},$$

but, since it requires some lengthy algebraic computations, we defer its proof to Lemma 19 below. Now, it follows that

$$\begin{aligned}
TV_1^2 \leq& \sum_{k=0,1} \sum_{j=0,1} (-1)^{k+j} \int\int (1-\eta^2)^{-n/2} (1-\eta'^2)^{-n/2} \\
&\qquad \left(\prod_{i=1}^n \int g(\eta, u, z_i, y_i) g(\eta', u', z_i, y_i) \varphi(z_i)\varphi(y_i) dz_i dy_i \right) \tilde{\pi}_k(du, d\eta) \tilde{\pi}_j(du', d\eta') \\
&= \sum_{k=0,1} \sum_{j=0,1} (-1)^{k+j} \int\int (1-\eta^2)^{-n/2} (1-\eta'^2)^{-n/2} |K|^{-n/2} \tilde{\pi}_k(du, d\eta) \tilde{\pi}_j(du', d\eta') \\
&= \sum_{k=0,1} \sum_{j=0,1} (-1)^{k+j} \int\int \frac{1}{(1-u_1^2 \eta\eta')^n} \tilde{\pi}_k(du, d\eta) \tilde{\pi}_j(du', d\eta') \\
&= \sum_{h=0}^{\infty} \sum_{k=0,1} \sum_{j=0,1} (-1)^{k+j} \int\int \binom{h+n-1}{n-1} u_1^{2h} \eta^h \eta'^h \tilde{\pi}_0(du, d\eta) \tilde{\pi}_1(du', d\eta') \\
&= \sum_{k=0}^{\infty} \binom{k+n-1}{n-1} \mathbb{E}[u_1^{2k}] \left(\int \eta^k [\nu_0(d\eta) - \nu_1(d\eta)] \right)^2 \\
&= \sum_{k=M+1}^{\infty} \binom{k+n-1}{n-1} \mathbb{E}[u_1^{2k}] \left(\int \eta^k [\nu_0(d\eta) - \nu_1(d\eta)] \right)^2,
\end{aligned}$$

since ν_0, ν_1 share the first M moments. □

Lemma 19. *Let*

$$K = \begin{pmatrix} I_d + \frac{\eta^2}{1-\eta^2} uu^T + \frac{\eta'^2}{1-\eta'^2} u'u'^T & -\frac{\eta}{1-\eta^2} uu^T - \frac{\eta'}{1-\eta'^2} u'u'^T \\ -\frac{\eta}{1-\eta^2} uu^T - \frac{\eta'}{1-\eta'^2} u'u'^T & I_d + \frac{\eta^2}{1-\eta^2} uu^T \frac{\eta'^2}{1-\eta'^2} u'u'^T \end{pmatrix},$$

where u, u' are d-dimensional unit vectors. Then,

$$|K| = \frac{(1-(u^T u')^2 \eta\eta')^2}{(1-\eta^2)(1-\eta'^2)}$$

Proof. Let $\alpha = \eta^2/(1-\eta^2), \alpha' = \eta'^2/(1-\eta'^2), \beta = -\eta/(1-\eta^2), \beta' = -\eta'/(1-\eta'^2)$. We aim at finding $|K|$, where

$$K = \begin{pmatrix} I_d + \alpha uu^T + \alpha' u'u'^T & \beta uu^T + \beta' u'u'^T \\ \beta uu^T + \beta' u'u'^T & I_d + \alpha uu^T \alpha u'u'^T \end{pmatrix}.$$

First, observe that by Schur's complement

$$\begin{aligned} |K| = &\left| I_d + \alpha uu^T + \alpha' u'u'^T \right| \\ &\times \left| I_d + \alpha uu^T + \alpha' u'u'^T - \left(\beta uu^T + \beta' u'u'^T\right) \left(I_d + \alpha uu^T + \alpha' u'u'^T\right)^{-1} \left(\beta uu^T + \beta' u'u'^T\right) \right|, \end{aligned}$$

and that we may assume without loss of generality that $u' = \boldsymbol{e}_1$. Indeed, let R be any orthogonal matrix in $\mathbb{R}^{d,d}$, and consider Ru, Ru' in place of u, u' respectively. Then

$$\begin{aligned} \left| I_d + \alpha Ruu^T R^T + \alpha' Ru'u'^T R^T \right| = \left| R(I_d + \alpha uu^T + \alpha' u'u'^T) R^T \right| = \left| R \right| \left| I_d + \alpha uu^T + \alpha' u'u'^T \right| \left| R^T \right| \\ \left| I_d + \alpha uu^T + \alpha' u'u'^T \right|, \end{aligned}$$

and it is easy to check that the same happens for

$$\begin{aligned} &\Big| I_d + \alpha Ruu^T R^T + \alpha' Ru'u'^T R^T \\ &- \left(\beta Ruu^T R^T + \beta' Ru'u'^T R^T\right) \left(I_d + \alpha Ruu^T R^T + \alpha' Ru'u'^T R^T\right)^{-1} \left(\beta Ruu^T R^T + \beta' Ru'u'^T R^T\right) \Big|. \end{aligned}$$

This is not necessary for the proof, but it helps with the notation, and also explains why $u^T u' \overset{d}{=} u_1$ when u and u' are sampled as described when we apply the result. Now, for all $\alpha, \alpha' \in \mathbb{R}$,

$$\begin{aligned} \left(I + \alpha uu^T + \alpha' \boldsymbol{e}_1 \boldsymbol{e}_1^T\right)^{-1} = I - &\frac{\alpha}{1+\alpha - \frac{\alpha\alpha'}{1+\alpha'}\gamma^2} uu^T - \frac{\alpha'}{1+\alpha' - \frac{\alpha\alpha'}{1+\alpha'}\gamma^2} \boldsymbol{e}_1 \boldsymbol{e}_1^T + \\ &+ \gamma \frac{\alpha'}{1+\alpha'} \frac{\alpha}{1+\alpha - \frac{\alpha\alpha'}{1+\alpha'}\gamma^2} \boldsymbol{e}_1 u^T + \gamma \frac{\alpha}{1+\alpha} \frac{\alpha'}{1+\alpha' - \frac{\alpha\alpha'}{1+\alpha'}\gamma^2} u \boldsymbol{e}_1^T, \end{aligned}$$

where $\gamma = \boldsymbol{e}_1^T u$, and

$$I + \alpha u u^T + \alpha' \boldsymbol{e}_1 \boldsymbol{e}_1^T - \left(\beta u u^T + \beta' \boldsymbol{e}_1 \boldsymbol{e}_1^T\right)\left(I + \alpha u u^T + \alpha' \boldsymbol{e}_1 \boldsymbol{e}_1^T\right)^{-1}\left(\beta u u^T + \beta' \boldsymbol{e}_1 \boldsymbol{e}_1^T\right)$$
$$= I + \frac{\gamma^2\eta^2\eta'^2}{1-\gamma^2\eta^2\eta'^2}(uu^T + \boldsymbol{e}_1\boldsymbol{e}_1^T) - \frac{\gamma\eta\eta}{1-\gamma^2\eta^2\eta'^2}(\boldsymbol{e}_1 u^T + u\boldsymbol{e}_1^T).$$

Calling $x = (\gamma^2\eta^2\eta'^2)/(1-\gamma^2\eta^2\eta'^2)$ and $c = \gamma\eta\eta'$, we thus have

$$|K| = \left|I_d + \alpha u u^T + \alpha' \boldsymbol{e}_1\boldsymbol{e}_1^T\right|\left|I_d + x u u^T + x\boldsymbol{e}_1\boldsymbol{e}_1^T - \frac{x}{c}\begin{pmatrix} | & | \\ u & \boldsymbol{e}_1 \\ | & | \end{pmatrix}\begin{pmatrix} — & \boldsymbol{e}_1 & — \\ — & u & — \end{pmatrix}\right|.$$

In order to compute these determinants, we will repeatedly make use of the fact that, if A is an invertible $n \times n$ matrix, U, V are $n \times m$ matrices, then

$$|A + uu^T| = |I_m + V^T A^{-1} U||A|.$$

If $A = I_n$, this is commonly referred as the Weinstein–Aronszajn identity. Now,

$$\left|I_d + \alpha u u^T + \alpha' \boldsymbol{e}_1\boldsymbol{e}_1^T\right| = \left|I_d + \alpha u u^T\right|\left|1 + \alpha' \boldsymbol{e}_1^T(I - \alpha u u^T/(1+\alpha))\boldsymbol{e}_1\right| = (1+\alpha)\left(1 + \alpha' - \frac{\alpha\alpha'}{1+\alpha}\gamma^2\right),$$

and

$$\left|I_d + x u u^T + x\boldsymbol{e}_1\boldsymbol{e}_1^T - \frac{x}{c}\begin{pmatrix} | & | \\ u & \boldsymbol{e}_1 \\ | & | \end{pmatrix}\begin{pmatrix} — & \boldsymbol{e}_1 & — \\ — & u & — \end{pmatrix}\right|$$
$$= \left|I_d + x u u^T + x\boldsymbol{e}_1\boldsymbol{e}_1^T\right|\left|I - \frac{x}{c}\begin{pmatrix} — & \boldsymbol{e}_1 & — \\ — & u & — \end{pmatrix}(I_d + xuu^T + x\boldsymbol{e}_1\boldsymbol{e}_1^T)^{-1}\begin{pmatrix} | & | \\ u & \boldsymbol{e}_1 \\ | & | \end{pmatrix}\right|$$
$$= (1+x)(1+x-\frac{x^2}{1+x}\gamma^2)\left|I - \frac{x}{c}\begin{pmatrix} — & \boldsymbol{e}_1 & — \\ — & u & — \end{pmatrix}\right.$$
$$\left.\times\left(I - \frac{x}{1+x-\frac{x^2}{1+x}\gamma^2}(uu^T + \boldsymbol{e}_1\boldsymbol{e}_1^T) + \gamma\frac{x}{1+x}\frac{x}{1+x-\frac{x^2}{1+x}\gamma^2}\begin{pmatrix} | & | \\ u & \boldsymbol{e}_1 \\ | & | \end{pmatrix}\begin{pmatrix} — & \boldsymbol{e}_1 & — \\ — & u & — \end{pmatrix}\right)\begin{pmatrix} | & | \\ u & \boldsymbol{e}_1 \\ | & | \end{pmatrix}\right|$$
$$= (1+x)(1+x-\frac{x^2}{1+x}\gamma^2)\left|\begin{pmatrix} 1 - \frac{x}{c}(\gamma + 2\tau_1\gamma + \tau_2(1+\gamma^2)) & -\frac{x}{c}(1 + \tau_1(1+\gamma^2) + 2\tau_2\gamma) \\ -\frac{x}{c}(1 + \tau_1(1+\gamma^2) + 2\tau_2\gamma) & 1 - \frac{x}{c}(\gamma + 2\tau_1\gamma + \tau_2(1+\gamma^2)) \end{pmatrix}\right|,$$

where $\tau_1 = -x/(1+x-\frac{x^2}{1+x}\gamma^2), \tau_2 = \gamma x\tau_1/(1+x)$. Putting all the pieces together,

$$|K| = (1+\alpha)\left(1+\alpha' - \frac{\alpha\alpha'}{1+\alpha}\gamma^2\right)(1+x)\left(1+x-\frac{x^2}{1+x}\gamma^2\right)$$
$$\times\left((1 - \frac{x}{c}(\gamma + 2\tau_1\gamma + \tau_2(1+\gamma^2)))^2 - \frac{x^2}{c^2}(1+\tau_1(1+\gamma^2) + 2\tau_2\gamma))^2\right),$$

and substituting the expressions of $\alpha, \alpha', x, c, \tau_1, \tau_2$ as functions of η, η', γ gives

$$|K| = \frac{(1-(u^T u')^2 \eta\eta')^2}{(1-\eta^2)(1-\eta'^2)},$$

as claimed. □

Proof of Proposition 15. We start by proving the first statement. Since $\Sigma_{\mathbb{S}}$ is consistent, we have that

$$\Sigma_{\mathbb{S}} \text{ is compatible} \quad \text{if and only if} \quad \begin{pmatrix} I_d & P & -P \\ P^T & I_d & \beta I_d \\ -P^T & \beta I_d & I_d \end{pmatrix} \succeq 0$$

$$\text{if and only if} \quad \begin{pmatrix} I_d & \beta I_d \\ \beta I_d & I_d \end{pmatrix} - \begin{pmatrix} P^T P & -P^T P \\ -P^T P & P^T P \end{pmatrix} \succeq 0,$$

where the second equivalence follows by standard properties of Schur complements. However, we can see that

$$\begin{aligned}
\inf\left\{ \begin{pmatrix} x \\ y \end{pmatrix}^T \left\{ \begin{pmatrix} I_d & \beta I_d \\ \beta I_d & I_d \end{pmatrix} - \begin{pmatrix} P^T P & -P^T P \\ -P^T P & P^T P \end{pmatrix} \right\} \begin{pmatrix} x \\ y \end{pmatrix} : x, y \in \mathbb{R}^d \right\}& \\
= \inf\left\{ \|x-y\|_2^2 + 2(1+\beta)x^T y - \|P(x-y)\|_2^2 : x, y \in \mathbb{R}^d \right\}& \\
= \inf\left\{ \|v\|_2^2 + 2(1+\beta)(v+y)^T y - \|Pv\|_2^2 : v, y \in \mathbb{R}^d \right\}& \\
= \inf\left\{ \frac{1-\beta}{2}\|v\|_2^2 - v^T P^T P v : v \in \mathbb{R}^d \right\}& \\
= \inf\left\{ \left(\frac{1-\beta}{2} - \|P\|_2^2 \right) v^2 : v \in [0, \infty) \right\}&,
\end{aligned}$$

where the third equality follows on noting that the minimising choice of y is given by $-v/2$. It is now clear that $\Sigma_{\mathbb{S}}$ is compatible if and only if $\|P\|_2^2 \le (1-\beta)/2$, as claimed.

As for the second part of the statement, let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_d$ the orthonormal eigenvectors or $P^T P$ with eigenvalues $\lambda_1 \ge \ldots \ge \lambda_d$, and let L be the maximal l such that $\lambda_l \ge (1-\beta)/2$. For $l \in [L]$, define

$$X_{\mathbb{S}}^{(l)} = \frac{c}{4}\left\{ \begin{pmatrix} 2P\boldsymbol{v}_l\boldsymbol{v}_l^T P^T & -2P\boldsymbol{v}_l\boldsymbol{v}_l^T \\ -2\boldsymbol{v}_l\boldsymbol{v}_l^T P^T & \boldsymbol{v}_l\boldsymbol{v}_l^T/2 \end{pmatrix}, \begin{pmatrix} 2P\boldsymbol{v}_l\boldsymbol{v}_l^T P^T & 2P\boldsymbol{v}_l\boldsymbol{v}_l^T \\ 2\boldsymbol{v}_l\boldsymbol{v}_l^T P^T & \boldsymbol{v}_l\boldsymbol{v}_l^T/2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{v}_l\boldsymbol{v}_l^T/2 & -\boldsymbol{v}_l\boldsymbol{v}_l^T \\ -\boldsymbol{v}_l\boldsymbol{v}_l^T & \boldsymbol{v}_l\boldsymbol{v}_l^T/2 \end{pmatrix} \right\},$$

with $0 < c \le 5/6 + \sqrt{73}/6$, and define $X_{\mathbb{S}} = \sum_{l=1}^L X_{\mathbb{S}}^{(l)}$. We first show that $X_{\mathbb{S}}$ is a feasible solution for our primal optimisation problem. We have

$$A^* X_{\mathbb{S}} = \frac{c}{4}\sum_{l=1}^L \begin{pmatrix} 4P\boldsymbol{v}_l\boldsymbol{v}_l^T P^T & -2P\boldsymbol{v}_l\boldsymbol{v}_l^T & 2P\boldsymbol{v}_l\boldsymbol{v}_l^T \\ -2\boldsymbol{v}_l\boldsymbol{v}_l^T P^T & \boldsymbol{v}_l\boldsymbol{v}_l^T & -\boldsymbol{v}_l\boldsymbol{v}_l^T \\ 2\boldsymbol{v}_l\boldsymbol{v}_l^T P^T & -\boldsymbol{v}_l\boldsymbol{v}_l^T & \boldsymbol{v}_l\boldsymbol{v}_l^T \end{pmatrix} = \frac{c}{4}\sum_{l=1}^L \begin{pmatrix} 2P\boldsymbol{v}_l \\ -\boldsymbol{v}_l \\ \boldsymbol{v}_l \end{pmatrix} \begin{pmatrix} 2P\boldsymbol{v}_l \\ -\boldsymbol{v}_l \\ \boldsymbol{v}_l \end{pmatrix}^T \succeq 0,$$

and since $X_{\mathbb{S}}^{(0)} = \frac{1}{2}(I_{2d}, I_{2d}, I_{2d})$,

$$X_{\mathbb{S}} + X_{\mathbb{S}}^{(0)} = \frac{1}{2}\left(\begin{pmatrix} I_d + cP(\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T)P^T & -cP(\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T) \\ -c(\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T)P^T & I_d + \frac{c}{4}\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T \end{pmatrix}, \right.$$
$$\left. \begin{pmatrix} I_d + cP(\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T)P^T & +cP(\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T) \\ +c(\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T)P^T & I_d + \frac{c}{4}\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T \end{pmatrix}, \begin{pmatrix} I_d + \frac{c}{4}\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T & -\frac{c}{2}\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T \\ -\frac{c}{2}\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T & I_d + \frac{c}{4}\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T \end{pmatrix}\right).$$

It remains to show that $X_{\mathbb{S}} + X_{\mathbb{S}}^{(0)} \succeq_{\mathbb{S}} 0$. Now, as for the first component of $X_{\mathbb{S}} + X_{\mathbb{S}}^{(0)}$, observe that the bottom-right block

$$I_d + \frac{c}{4}\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T \succeq 0,$$

and it is invertible due to the fact that $\|c\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T/4\|_2 \leq c/4 < 1$, since the $\boldsymbol{v}_l$'s are orthonormal. The inverse is

$$\begin{aligned}\left(I_d + \frac{c}{4}\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T\right)^{-1} &= \left(I_d - \left(-\frac{c}{4}\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T\right)\right)^{-1} = \sum_{k=0}^{\infty}(-1)^k\left(\frac{c}{4}\right)^k\left(\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T\right)^k \\ &= I_d + \sum_{k=1}^{\infty}(-1)^k\left(\frac{c}{4}\right)^k\left(\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T\right)^k = I_d + \left(\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T\right)\sum_{k=1}^{\infty}(-1)^k\left(\frac{c}{4}\right)^k \\ &= I_d + \left(\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T\right)\left(\frac{1}{1+c/4} - 1\right) = I_d - \frac{c}{c+4}\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T,\end{aligned}$$

where the fourth equality comes from the fact that $\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T$ is idempotent again by the orthonormality of the $\boldsymbol{v}_l$'s. Hence, the first component of $X_{\mathbb{S}} + X_{\mathbb{S}}^{(0)}$ is positive semidefinite if and only if

$$I_d + cP\left(\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T\right)P^T \succeq c^2 P\left(\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T\right)\left(I_d - \frac{c}{4+c}\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T\right)\left(\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T\right)P^T,$$

which is equivalent to

$$I_d \succeq \left(\frac{4c^2}{4+c} - c\right)P\left(\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T\right)P^T = \left(\frac{4c^2}{4+c} - c\right)P\left(\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T\right)^2 P^T,$$

which is satisfied if and only if $4c^2/(4+c) - c \leq 1$, due to the fact that $\|P\left(\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T\right)^2 P^T\|_2 = \|P\left(\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T\right)\|_2^2 \leq 1$ again by orthonormality. This implies that the first component of $X_{\mathbb{S}} + X_{\mathbb{S}}^{(0)} \succeq_{\mathbb{S}}$ is PSD if and only if $0 < c \leq 5/6 + \sqrt{73}/6$, and of course the same is true for the second component of $X_{\mathbb{S}} + X_{\mathbb{S}}^{(0)}$. As for the third component, using an analogous idea, it is easy to show that it is positive semidefinite if and only if

$$I_d \succeq \left(\frac{c^2}{4+c} - \frac{c}{4}\right)\sum_{l=1}^{L} \boldsymbol{v}_l \boldsymbol{v}_l^T,$$

which is satisfied if and only if $0 < c \leq 4$. Summing up, this shows that $X_{\mathbb{S}}$ is feasible for $0 < c \leq 5/6+\sqrt{73}/6$, and leads to

$$\begin{aligned} R(\Sigma_{\mathbb{S}}) &\geq -\frac{c}{3d}\sum_{l=1}^{L}\begin{pmatrix}2P\boldsymbol{v}_l\\ -\boldsymbol{v}_l\\ \boldsymbol{v}_l\end{pmatrix}\begin{pmatrix}I_d & P & -P\\ P & I_d & \beta I_d\\ -P^T\beta I_d & I_d & \end{pmatrix}\begin{pmatrix}2P\boldsymbol{v}_l\\ -\boldsymbol{v}_l\\ \boldsymbol{v}_l\end{pmatrix}^T \\ &= \frac{c}{3d}\sum_{l=1}^{L}\left(\lambda_l - \frac{1-\beta}{2}\right) = \frac{c}{3d}\sum_{l=1}^{d}\left(\lambda_l - \frac{1-\beta}{2}\right)_+ \\ &= \frac{c}{3d}\sum_{l=1}^{d}\left(\sigma_l^2(P)^2 - \frac{1-\beta}{2}\right)_+ > \frac{3}{4d}\sum_{l=1}^{d}\left(\sigma_l^2(P)^2 - \frac{1-\beta}{2}\right)_+, \end{aligned}$$

since $5/6+\sqrt{73}/6 > 9/4$. □

References

Arthur Albert. Regression and the Moore-Penrose pseudoinverse. *Mathematics in science and engineering*, 94, 1972.

Danijel Aleksić. A novel test of Missing Completely at Random: U-statistics-based approach. *arXiv preprint arXiv:2310.19189*, 2023.

Wayne Barrett, Charles R Johnson, and Pablo Tarazaga. The real positive definite completion problem for a simple cycle. *Linear Algebra Appl.*, 192:3–31, 1993.

Thomas B Berrett and Richard J Samworth. Optimal nonparametric testing of Missing Completely At Random, and its connections to compatibility. *Ann. Statist.*, 51(5):2170–2193, 2023.

Thomas B. Berrett, Alberto Bordino, Danat Duisenbekov, Sean Jaffe, and Richard J. Samworth. *MCARtest: Optimal nonparametric testing of Missing Completely At Random*, 2022. URL https://cran.r-project.org/web/packages/MCARtest/index.html. R package version 1.2.

Gilles Blanchard, Alexandra Carpentier, and Maurilio Gutzeit. Minimax Euclidean separation rates for testing convex hypotheses in $\mathbb{R}^d$. *Electronic Journal of Statistics*, 12(2):3713 – 3735, 2018. doi: 10.1214/18-EJS1472. URL https://doi.org/10.1214/18-EJS1472.

Grigoriy Blekherman, Pablo A. Parrilo, and Rekha R. Thomas. *Semidefinite Optimization and Convex Algebraic Geometry*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2012. doi: 10.1137/1.9781611972290. URL https://epubs.siam.org/doi/abs/10.1137/1.978161197229.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, March 2004. ISBN 0521833787. URL https://www.cambridge.org/gb/universitypress/subjects/statistics-probability/optimization-or-and-risk/convex-optimization?format=HB.

Florentina Bunea and Luo Xiao. On the sample covariance matrix estimator of reduced effective rank population matrices, with applications to fPCA. *Bernoulli*, 21(2), may 2015. URL https://doi.org/10.3150%2F14-bej602.

T Tony Cai and Mark G Low. Testing composite hypotheses, Hermite polynomials and optimal estimation of a nonsmooth functional. *Ann. Statist.*, 39:1012–1041, 2011.

T Tony Cai and Linjun Zhang. High-dimensional linear discriminant analysis: Optimality, adaptive algorithm, and missing data. *J. Roy. Statist. Soc. Ser. B*, 81(4):675–705, 2019.

Tony Cai, Weidong Liu, and Xi Luo. A Constrained l_1 Minimization Approach to Sparse Precision Matrix Estimation. *Journal of the American Statistical Association*, 106(494):594–607, 2011. ISSN 01621459. URL http://www.jstor.org/stable/41416395.

Emmanuel J. Candès and Terence Tao. The Power of Convex Relaxation: Near-Optimal Matrix Completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080, 2010. doi: 10.1109/TIT.2010.2044061.

Emmanuel J. Candès and Benjamin Recht. Exact Matrix Completion via Convex Optimization. *Foundations of Computational Mathematics*, 9(6):717–772, Dec 2009. ISSN 1615-3375. Funding by NSF.

D. R. Cox. A note on data-splitting for the evaluation of significance levels. *Biometrika*, 62:441–444, 1975.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1977.tb01600.x.

Andreas Elsener and Sara van de Geer. Sparse spectral estimation with missing and corrupted measurements, 2018. URL https://arxiv.org/abs/1811.10443.

Bertille Follain, Tengyao Wang, and Richard J. Samworth. High-dimensional Changepoint Estimation with Heterogeneous Missingness. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84 (3):1023–1055, 07 2022. ISSN 1369-7412. doi: 10.1111/rssb.12540. URL https://doi.org/10.1111/rssb.12540.

Camil Fuchs. Maximum likelihood estimation and model selection in contingency tables with missing data. *J. Amer. Statist. Assoc.*, 77:270–278, 1982.

Robert Grone, Charles R. Johnson, Eduardo M. Sá, and Henry Wolkowicz. Positive definite completions of partial Hermitian matrices. *Linear Algebra and its Applications*, 58:109–124, 1984. ISSN 0024-3795. doi: https://doi.org/10.1016/0024-3795(84)90207-6. URL https://www.sciencedirect.com/science/article/pii/0024379584902076.

F Richard Guo and Rajen D Shah. Rank-transformed subsampling: inference for multiple data splitting and exchangeable p-values. *arXiv preprint arXiv:2301.02739*, 2023.

Fang Han and Han Liu. Statistical analysis of latent generalized correlation matrix estimation in transelliptical distribution. *Bernoulli*, 23(1), feb 2017. doi: 10.3150/15-bej702. URL https://doi.org/10.3150%2F15-bej702.

Douglas M. Hawkins. A new test for multivariate normality and homoscedasticity. *Technometrics*, 23(1): 105–110, 1981. ISSN 0040-1706. doi: 10.1080/00401706.1981.10486244.

Marius Hofert, Ivan Kojadinovic, Martin Maechler, and Jun Yan. *copula: Multivariate Dependence with Copulas*, 2020. URL https://CRAN.R-project.org/package=copula. R package version 1.0-0.

Mortaza Jamshidian and Siavash Jalal. Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data. *Psychometrika*, 75:649–674, 2010.

Jana Janková, Rajen D. Shah, Peter Bühlmann, and Richard J. Samworth. Goodness-of-fit Testing in High Dimensional Generalized Linear Models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(3):773–795, 05 2020. ISSN 1369-7412. doi: 10.1111/rssb.12371. URL https://doi.org/10.1111/rssb.12371.

Jiantao Jiao, Yanjun Han, and Tsachy Weissman. Minimax Estimation of the L_1 Distance. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 750–754, 2016. doi: 10.1109/ISIT.2016.7541399.

C.R. Johnson. *Matrix Theory and Applications*. American Mathematical Soc., 1989. ISBN 9780821867464. URL https://books.google.co.uk/books?id=EOflkS9ruxcC.

Anatoli Juditsky and Arkadi Nemirovski. On nonparametric tests of positivity/monotonicity/convexity. *The Annals of Statistics*, 30(2):498 – 527, 2002. doi: 10.1214/aos/1021379863. URL https://doi.org/10.1214/aos/1021379863.

Hans G Kellerer. Duality theorems for marginal problems. *Z. Wahrscheinlichkeit.*, 67:399–432, 1984.

L. Khachiyan and L. Porkolab. Computing integral points in convex semi-algebraic sets. In *Proceedings 38th Annual Symposium on Foundations of Computer Science*, pages 162–171, 1997. doi: 10.1109/SFCS.1997.646105.

Ilmun Kim and Aaditya Ramdas. Dimension-agnostic inference using cross U-statistics. *Bernoulli*, 30(1): 683–711, 2024. doi: 10.3150/23-BEJ1613. URL https://doi.org/10.3150/23-BEJ1613.

Kevin H Kim and Peter M Bentler. Tests of homogeneity of means and covariance matrices for multivariate incomplete data. *Psychometrika*, 67:609–623, 2002.

Monique Laurent. Matrix Completion Problems. *Encyclopedia of Optimization*, 3:221–229, 2009.

Monique Laurent and Svatopluk Poljak. On the Facial Structure of the Set of Correlation Matrices. *SIAM Journal on Matrix Analysis and Applications*, 17(3):530–547, 1996. doi: 10.1137/0617031. URL https://doi.org/10.1137/0617031.

Jing Lei, Max G'Sell, Alessandro Rinaldo, Ryan J. Tibshirani, and Larry Wasserman. Distribution-Free Predictive Inference for Regression. *Journal of the American Statistical Association*, 113(523):1094–1111, 2018. doi: 10.1080/01621459.2017.1307116. URL https://doi.org/10.1080/01621459.2017.1307116.

Jun Li and Yao Yu. A nonparametric test of missing completely at random for incomplete multivariate data. *Psychometrika*, 80:707–726, 2015.

R.J.A. Little and D.B. Rubin. *Statistical analysis with missing data*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. Wiley, 2002. ISBN 9780471183860. URL http://books.google.com/books?id=aYPwAAAAMAAJ.

Roderick JA Little. A test of Missing Completely at Random for multivariate data with missing values. *J. Amer. Statist. Assoc.*, 83:1198–1202, 1988.

Po-Ling Loh and Martin J Wainwright. High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *Ann. Statist.*, 40(3):1637–1664, 2012.

Karim Lounici. High-dimensional covariance matrix estimation with missing observations. *Bernoulli*, 20(3): 1029 – 1058, 2014. doi: 10.3150/12-BEJ487. URL https://doi.org/10.3150/12-BEJ487.

L. Lovász. *Semidefinite Programs and Combinatorial Optimization*, pages 137–194. Springer New York, New York, NY, 2003. ISBN 978-0-387-22444-2. URL https://doi.org/10.1007/0-387-22444-0_6.

Anton Rask Lundborg, Ilmun Kim, Rajen D. Shah, and Richard J. Samworth. The Projected Covariance Measure for assumption-lean variable significance testing, 2022.

Nicolai Meinshausen and Peter Bühlmann. High-Dimensional Graphs and Variable Selection with the Lasso. *The Annals of Statistics*, 34(3):1436–1462, 2006. ISSN 00905364. URL http://www.jstor.org/stable/25463463.

P. A. P. Moran. Dividing a Sample into Two Parts. A Statistical Dilemma. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 35(3):329–333, 1973. ISSN 0581572X. URL http://www.jstor.org/stable/25049883.

Yurii Nesterov and Arkadii Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. Society for Industrial and Applied Mathematics, 1994. doi: 10.1137/1.9781611970791. URL https://epubs.siam.org/doi/abs/10.1137/1.9781611970791.

Roberto Imbuzeiro Oliveira. Concentration of the adjacency matrix and of the Laplacian in random graphs with independent edges. *arXiv preprint arXiv:0911.0600*, 2010.

Elijah Polak. *Optimization*. Springer Science and Business Media, 12 2012.

Motakuri V. Ramana. An exact duality theory for semidefinite programming and its complexity implications. *Mathematical Programming*, 77:129–162, 1997.

Benjamin Recht. A Simpler Approach to Matrix Completion. *J. Mach. Learn. Res.*, 12:3413–3430, dec 2011. ISSN 1532-4435.

R. Tyrrell Rockafellar. *Convex analysis*. Princeton Mathematical Series. Princeton University Press, Princeton, N. J., 1970.

Tobias Rockel. *missMethods: Methods for Missing Data*, 2020. URL https://CRAN.R-project.org/package=missMethods. R package version 0.2.0.

Miriam Seoane Santos, Ricardo Cardoso Pereira, Adriana Fonseca Costa, Jastin Pompeu Soares, João Santos, and Pedro Henriques Abreu. Generating Synthetic Missing Data: A Review by Missing Mechanism. *IEEE Access*, 7:11651–11667, 2019. doi: 10.1109/ACCESS.2019.2891360.

Cyrill Scheidegger, Julia Hörrmann, and Peter Bühlmann. The Weighted Generalised Covariance Measure. *J. Mach. Learn. Res.*, 23:273:1–273:68, 2022. URL http://jmlr.org/papers/v23/21-1328.html.

Torben Sell, Thomas B Berrett, and Timothy I Cannings. Nonparametric classification with missing data. *arXiv preprint arXiv:2305.11672*, 2023.

Aldo Solari and Vera Djordjilović. Multi split conformal prediction. *Statistics & Probability Letters*, 184(C), 2022. doi: 10.1016/j.spl.2022.109395. URL https://ideas.repec.org/a/eee/stapro/v184y2022ics0167715222000177.html.

Meta-Lina Spohn, Jeffrey Näf, Loris Michel, and Nicolai Meinshausen. PKLM: A flexible MCAR test using Classification. *arXiv preprint arXiv:2109.10150*, 2021.

Daniel J. Stekhoven and Peter Bühlmann. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1):112–118, 10 2011. ISSN 1367-4803. doi: 10.1093/bioinformatics/btr597. URL https://doi.org/10.1093/bioinformatics/btr597.

Solène Thépaut and Nicolas Verzelen. Optimal estimation of Schatten norms of a rectangular matrix. *arXiv preprint arXiv:2111.13551*, 2021.

Stef van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45:1–67, 2011.

Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996. doi: 10.1137/1038003. URL https://doi.org/10.1137/1038003.

Roman Vershynin. *High-Dimensional Probability*. Cambridge University Press, 2019. URL https://www.math.uci.edu/~rvershyn/papers/HDP-book/HDP-book.pdf.

Kartik G Waghmare and Victor M Panaretos. The completion of covariance kernels. *The Annals of Statistics*, 50(6):3281–3306, 2022.

Martin J Wainwright. *High-dimensional Statistics: A Non-asymptotic Viewpoint*. Cambridge University Press, 2019.

S. S. Wilks. Sample Criteria for Testing Equality of Means, Equality of Variances, and Equality of Covariances in a Normal Multivariate Distribution. *The Annals of Mathematical Statistics*, 17(3):257 – 281, 1946. doi: 10.1214/aoms/1177730940. URL https://doi.org/10.1214/aoms/1177730940.

F. Yates. The Analysis of Replicated Experiments When the Field Results Are Incomplete. *Empirical Journal of Experimental Agriculture 1 (2): 129–42*, 1933.

Ziwei Zhu, Tengyao Wang, and Richard J. Samworth. High-dimensional principal component analysis with heterogeneous missingness. *Journal of the Royal Statistical Society Series B*, 84(5):2000–2031, November 2022. doi: 10.1111/rssb.12550. URL https://ideas.repec.org/a/bla/jorssb/v84y2022i5p2000-2031.html.

Appendices

Appendix A contains further properties of our measure of incompatibility R that were not investigated in the main body of the text. Appendix B contains another oracle test based on a different measure of incompatibility, which acts on covariance matrices normalised in such a way to have fixed scale. Appendix C contains auxiliary results in Semi-definite Programming, while classical tail bounds are contained in Appendix D.

Appendix A Further properties of $R(\cdot)$

We analyse two further examples, where the missingness patterns $\mathbb{S}$ is more complex. These examples are interesting *per se*, but we have decided not to include them in the main body because they would have disrupted the flow of the presentation. We start from the case where we observe all possible patterns of cardinality $d-1$, and nothing else.

Example 5. *Consider the set of patterns* $\mathbb{S} = \{S_{(-1)}, \cdots, S_{(-d)}\}$*, with* $d \geq 2$*, where* $S_{(-i)} = \{1, \ldots, i-1, i+1, \ldots, d\}$*. We show how* $R(\Sigma_{\mathbb{S}})$ *can be lower-bounded by the maximal inconsistency, or, more precisely,*

$$R(\Sigma_{\mathbb{S}}) \geq \frac{1}{2} \max_{i>j} \max_{\substack{k>h \\ k,h \neq i,j}} |\rho_{ij}^{(k)} - \rho_{ij}^{(h)}| =: \Theta,$$

where $\rho_{ij}^{(k)}$ *is the correlation between* X_i *and* X_j *for the pattern* $S_{(-k)}$*, for* $k \in [d] \setminus \{i, j\}$*.*

Proof of Example 5. In order to prove the statement, suppose the maximum is $|\rho_{ij}^{(k)} - \rho_{ij}^{(h)}|$, and consider $X_{\mathbb{S}} = dY_{\mathbb{S}} - X_{\mathbb{S}}^{(0)}$, where $X_{\mathbb{S}}^{(0)} = \frac{1}{d-1}(I_{d-1}, \ldots, I_{d-1})$ and

$$Y_{\mathbb{S}} = (0, \ldots, \underbrace{A_1}_{h}, 0, \ldots, 0, \underbrace{A_2}_{k}, 0, \ldots, 0),$$

with

$$(A_1)_{\tilde{i},\tilde{j}} = \begin{cases} 1/4 & \text{if } (\tilde{i}, \tilde{j}) \in \{(i,i), (j,j), (i,j), (j,i)\} \\ 0 & \text{otherwise,} \end{cases}$$

and

$$(A_2)_{\tilde{i},\tilde{j}} = \begin{cases} 1/4 & \text{if } (\tilde{i}, \tilde{j}) \in \{(i,i), (j,j)\} \\ -1/4 & \text{if } (\tilde{i}, \tilde{j}) \in \{(i,j), (j,i)\} \\ 0 & \text{otherwise.} \end{cases}$$

Then, provided $X_{\mathbb{S}}$ is feasible, we get precisely that

$$R(\Sigma_{\mathbb{S}}) \geq -\frac{1}{d}\langle X_{\mathbb{S}}, \Sigma_{\mathbb{S}}\rangle_{\mathbb{S}} = \frac{1}{2} \max_{i>j} \max_{\substack{k>h \\ k,h\neq i,j}} |\rho_{ij}^{(k)} - \rho_{ij}^{(h)}|.$$

All is left to prove is that $X_{\mathbb{S}}$ is indeed feasible: $X_{\mathbb{S}} + X_{\mathbb{S}}^{(0)} = dY_{\mathbb{S}} \succeq_{\mathbb{S}} 0$, and $A^* X_{\mathbb{S}}$ is diagonal with trace zero, hence we can choose $Y = -A^* X_{\mathbb{S}} \in \mathcal{Y}$ in the primal characterisation so that $A^* X_{\mathbb{S}} + Y = \boldsymbol{O} \succeq 0$.

□

Observe that the same is true in the case where we also have a complete case pattern, i.e. $\mathbb{S} = \{[d], S_{(-1)}, \cdots, S_{(-d)}\}$, with $d \geq 2$, meaning that using the same strategy we can control R with the maximal inconsistency. Related to this, it would be interesting to know if there is a case in which the incompatibility value Θ controls $R(\Sigma_{\mathbb{S}})$ both from above and below, meaning that Θ fully characterises $R(\Sigma_{\mathbb{S}})$. In this regard, we have the following:

Example 6. *Consider* $\mathbb{S} = \{S_{(-1)}, \cdots, S_{(-d)}\}$ *and* $\Sigma_{\mathbb{S}} = (I_{d-1}, \ldots, I_{d-1}, A)$, *where*

$$A = \begin{pmatrix} 1 & \epsilon_1/2 & 0 & 0 & \cdots & 0 \\ \epsilon_1/2 & 1 & \epsilon_2/2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \\ & & & & & \\ \vdots & & & \ddots & \ddots & \ddots \\ & & & & & \\ 0 & \cdots & 0 & 0 & \epsilon_{d-1}/2 & 1 \end{pmatrix},$$

with $\epsilon_i \in [-1, 1]$. *Then,*

$$R(\Sigma_{\mathbb{S}}) = \Theta = \frac{1}{2} \max_{i \in [d-1]} |\epsilon_i|.$$

Proof of Example 6. If we consider

$$\Sigma = \begin{pmatrix} 1-\Theta & \epsilon_1/4 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \epsilon_1/4 & 1-\Theta & \epsilon_2/4 & \cdots & 0 & 0 & 0 & 0 \\ \ddots & \ddots & \ddots & & & & & \\ & & & & & & & \\ & & & \ddots & \ddots & \ddots & & \\ 0 & \cdots & 0 & \cdots & \epsilon_{d-2}/4 & 1-\Theta & \epsilon_{d-1}/4 & 0 \\ 0 & \cdots & 0 & 0 & \cdots & \epsilon_{d-1}/4 & 1-\Theta & 0 \\ 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & 1-\Theta \end{pmatrix} \in \mathbb{R}^{d,d},$$

if Σ were feasible we would be able to conclude $R(\Sigma_{\mathbb{S}}) = \frac{1}{2}\max_{i\in[d]} |\epsilon_i|$ being

$$\Theta \geq R(\Sigma_{\mathbb{S}}) \geq \Theta = \frac{1}{2} \max_{i \in [d-1]} |\epsilon_i|.$$

All is left to prove is that Σ is feasible. First, $\Sigma \succeq 0$ since it is diagonally dominant, being $1-\max_{i\in[d]}|\epsilon_i|/2 \in [1/2,1]$ and $\epsilon_i/4 \in [-1/4,1/4]$. Finally, a generic element in $\Sigma_{\mathbb{S}} - A\Sigma$ is given by

$$\begin{pmatrix} \max_{i\in[d]}|\epsilon_i|/2 & \alpha_1/4 & 0 & 0 & \cdots & 0 \\ \alpha_1/4 & \max_{i\in[d]}|\epsilon_i|/2 & \alpha_2/4 & 0 & \cdots & 0 \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & 0 & \alpha_{d-1}/4 & \max_{i\in[d]}|\epsilon_i|/2 \end{pmatrix} \in \mathbb{R}^{d-1,d-1},$$

where $\alpha_i \in \{\pm\epsilon_i, 0\}$. This is again diagonally dominant since $\max_{i\in[d]}|\epsilon_i|/2 \geq |\alpha_j|/4 + |\alpha_{j+1}|/4$ for all $j \in [d-1]$, by definition of the maximum. □

This example is particularly important since it clearly shows that, in this case, testing compatibility is at least as hard as testing consistency. Indeed, Θ is a pointwise measure of consistency, and equals 0 if and only if $\Sigma_{\mathbb{S}}$ is consistent. Nonetheless, the equality $R(\Sigma_{\mathbb{S}}) = \Theta$ holds for a very specific subclass of $\Sigma_{\mathbb{S}}$, while, in general, there could be cases for which $R(\Sigma_{\mathbb{S}}) > 0$, while $\Theta = 0$.

Example 7. *Consider $\mathbb{S} = \{[d-2] \cup \{d-1\}, [d-2] \cup \{d\}\}$. Call S_1 and S_2 the two patterns, respectively, and suppose we observe the sequence of correlation matrices given by $\Sigma_{\mathbb{S}} = (\Sigma_{S_1}, \Sigma_{S_2})$. If we call*

$$\tilde{\Sigma} = (\Sigma_{S_2})_{|[d-2]} - (\Sigma_{S_1})_{|[d-2]},$$

where $(\Sigma_{S_i})_{|[d-2]}$ is the restriction of Σ_{S_i} on the set $[d-2]$, for $i \in \{1,2\}$, then

$$R(\Sigma_{\mathbb{S}}) \geq \frac{1}{2d}\|\tilde{\Sigma}\|_*,$$

where $\|\cdot\|_*$ is the nuclear norm, also known as the Schatten-1 norm.

Proof of Example 7. Define

$$X_{\mathbb{S}} = \left(\begin{pmatrix} X & \mathbf{0}_{d-2} \\ \mathbf{0}_{d-2}^T & 0 \end{pmatrix}, \begin{pmatrix} -X & \mathbf{0}_{d-2} \\ \mathbf{0}_{d-2}^T & 0 \end{pmatrix} \right),$$

where $X \in \mathbb{R}^{d-2,d-2}$ and $\|X\|_2 \leq 1/2$. Observe that this choice of $X_{\mathbb{S}}$ is feasible since $A^* X_{\mathbb{S}} = \boldsymbol{O} \succeq 0$, and

$$X_{\mathbb{S}} + X_{\mathbb{S}}^0 = \left(\begin{pmatrix} X + \frac{1}{2}I_{d-2} & \mathbf{0}_{d-2} \\ \mathbf{0}_{d-2}^T & 1 \end{pmatrix}, \begin{pmatrix} -X + \frac{1}{2}I_{d-2} & \mathbf{0}_{d-2} \\ \mathbf{0}_{d-2}^T & 1 \end{pmatrix} \right) \succeq_{\mathbb{S}} 0,$$

since $\|X\|_2 \leq 1/2$. It follows that

$$R(\Sigma_{\mathbb{S}}) \geq \sup_{\substack{X=X^T \\ \|X\|_2 \leq 1/2}} -\frac{1}{d}\langle X_{\mathbb{S}}, \Sigma_{\mathbb{S}} \rangle_{\mathbb{S}} = \sup_{\substack{X=X^T \\ \|X\|_2 \leq 1/2}} \frac{1}{d}\langle X, \tilde{\Sigma} \rangle = \frac{1}{2d}\|\tilde{\Sigma}\|_*,$$

where $\|\cdot\|_*$ is the nuclear norm and equality follows since the spectral norm and the nuclear norm are dual with respect to the Frobenius inner product. $\square$

Appendix B Another test under trace normalisation

In the main body we were dealing with the incompatibility measure R, which acts on correlation matrices, normalised in a such a way that diagonal elements are all equal to one. Nonetheless, other standardisations are possible, and these lead to different compatibility measures. In this section, we will define another measure of compatibility $\tilde{R}(\cdot)$, study its properties, and use it to define a testing procedure. Similarly to Table 1 in the main body, refer to Table 2 for all the new algebraic definitions needed in this section.

Notation	Definition	Meaning
$\bar{\mathrm{tr}} : \mathcal{M}_{\mathbb{S}} \to \mathbb{R}$	$\bar{\mathrm{tr}}(X_{\mathbb{S}}) = \sum_{j=1}^{d} \lvert\mathbb{S}_j\rvert^{-1} \sum_{S\in\mathbb{S}_j} (X_S)_{jj}$	Generalisation of the trace such that, if $\Sigma_{\mathbb{S}}$ is compatible, then $\bar{\mathrm{tr}}(\Sigma_{\mathbb{S}})$ is equal to the trace of the underlying true covariance matrix
$\tilde{\mathcal{P}}$	$\{\Sigma \in \mathcal{P}^* : \mathrm{tr}(\Sigma) = d\}$	Set of PSD matrices with fixed scale
$\tilde{\mathcal{P}}_{\mathbb{S}}$	$\{\Sigma_{\mathbb{S}} \in \mathcal{P}^*_{\mathbb{S}} : \bar{\mathrm{tr}}(\Sigma_{\mathbb{S}}) = d\}$	Sequences of PSD matrices with scale fixed
$\tilde{\mathcal{P}}^0_{\mathbb{S}}$	$\{A\Sigma : \Sigma \in \tilde{\mathcal{P}}\}$	Same as $\mathcal{P}^0_{\mathbb{S}}$, but Σ has fixed scale

Table 2: Table with all the definitions needed in Appendix B.

The linear operator $\bar{\mathrm{tr}}$ satisfies the following:

Proposition 20. *The following hold:*

(i) If we define $X^0_{\mathbb{S}} \in \mathcal{M}_{\mathbb{S}}$ by taking X^0_S to be the diagonal matrix with $(X^0_S)_{jj} = 1/\lvert\mathbb{S}_j\rvert$, we have

$$\bar{\mathrm{tr}}(X_{\mathbb{S}}) = \langle X_{\mathbb{S}}, X^0_{\mathbb{S}} \rangle_{\mathbb{S}}$$

for all $X_{\mathbb{S}} \in \mathcal{M}_{\mathbb{S}}$.

(ii) Suppose that $X_{\mathbb{S}}$ is consistent, meaning that $(X_{S_1})_{jj'} = (X_{S_2})_{jj'}$ whenever $S_1, S_2 \in \mathbb{S}_{jj'}$, and write X^{partial} for the incomplete $d \times d$ matrix with $(X^{\mathrm{partial}})_{jj'} = (X_S)_{jj'}$ for any $S \in \mathbb{S}_{jj'}$. Then

$$\bar{\mathrm{tr}}(X_{\mathbb{S}}) = \mathrm{tr}(X^{\mathrm{partial}}) \quad \text{and} \quad \langle X_{\mathbb{S}}, Y_{\mathbb{S}} \rangle_{\mathbb{S}} = \langle X^{\mathrm{partial}}, A^* Y_{\mathbb{S}} \rangle$$

for any $Y_{\mathbb{S}} \in \mathcal{M}_{\mathbb{S}}$.

Proof of Proposition 20. Now for any $X_{\mathbb{S}} \in \mathcal{M}_{\mathbb{S}}$ we see that

$$\langle X_{\mathbb{S}}, X^0_{\mathbb{S}} \rangle_{\mathbb{S}} = \sum_{S\in\mathbb{S}} \sum_{j\in S} (X_S)_{jj} (X^0_S)_{jj} = \sum_{j=1}^{d} \lvert\mathbb{S}_j\rvert^{-1} \sum_{S\in\mathbb{S}} \mathbb{1}_{\{j\in S\}} (X_S)_{jj} = \bar{\mathrm{tr}}(X_{\mathbb{S}}),$$

proving property (i). The first part of (ii) can be seen immediately from the definition of $\bar{\mathrm{tr}}$. For the second part, write

$$\langle X_{\mathbb{S}}, Y_{\mathbb{S}} \rangle_{\mathbb{S}} = \sum_{S\in\mathbb{S}} \sum_{j,j'\in S} (X_S)_{jj'}(Y_S)_{jj'} = \sum_{j,j'=1}^{d} (X^{\mathrm{partial}})_{jj'} \sum_{S\in\mathbb{S}} \mathbb{1}_{\{j,j'\in S\}}(Y_S)_{jj'} = \langle X^{\mathrm{partial}}, A^*Y_{\mathbb{S}}\rangle.$$

□

Now, suppose that $\Sigma_{\mathbb{S}}$ is such that $\bar{\mathrm{tr}}(\Sigma_{\mathbb{S}}) = d$, where $\bar{\mathrm{tr}}(X_{\mathbb{S}}) = \sum_{j=1}^{d} |\mathbb{S}_j|^{-1} \sum_{S\in\mathbb{S}_j} (X_S)_{jj}$, with $\mathbb{S}_j := \{S \in \mathbb{S} : j \in S\}$, and define

$$\tilde{R}(\Sigma_{\mathbb{S}}) := \sup\biggl\{-\frac{1}{d}\langle X_{\mathbb{S}}, \Sigma_{\mathbb{S}}\rangle_{\mathbb{S}} : X_{\mathbb{S}} + X^0_{\mathbb{S}} \succeq_{\mathbb{S}} 0, A^*X_{\mathbb{S}} \succeq 0\biggr\}.$$

This new measure of incompatibility has the following dual representation:

Proposition 21. *For $\Sigma_{\mathbb{S}} \in \tilde{\mathcal{P}}_{\mathbb{S}}$ we have*

$$\tilde{R}(\Sigma_{\mathbb{S}}) = \inf\{\epsilon \in [0,1] : \Sigma_{\mathbb{S}} \in (1-\epsilon)\tilde{\mathcal{P}}^0_{\mathbb{S}} + \epsilon\tilde{\mathcal{P}}_{\mathbb{S}}\}.$$

Proof of Proposition 21. As in the proof of Proposition 3, the strategy is to write this optimisation problem

$$\sup\biggl\{-\frac{1}{d}\langle X_{\mathbb{S}}, \Sigma_{\mathbb{S}}\rangle_{\mathbb{S}} : X_{\mathbb{S}} + X^0_{\mathbb{S}} \succeq_{\mathbb{S}} 0, A^*X_{\mathbb{S}} \succeq 0\biggr\} \tag{17}$$

in standard SDP form, prove that the dual problem is precisely

$$1 - \frac{1}{d}\sup\{\mathrm{tr}(\Sigma) : \Sigma \in \mathcal{P}^*, \Sigma_{\mathbb{S}} - A\Sigma \succeq_{\mathbb{S}} 0\}, \tag{18}$$

and then show that Slater's condition is satisfied for the primal problem (17). Calling $Y_{\mathbb{S}} = X_{\mathbb{S}} + X^0_{\mathbb{S}}$, we have that

$$\begin{aligned}
&\sup\biggl\{ - \frac{1}{d}\langle X_{\mathbb{S}}, \Sigma_{\mathbb{S}}\rangle_{\mathbb{S}} : X_{\mathbb{S}} + X^0_{\mathbb{S}} \succeq_{\mathbb{S}} 0, A^*X_{\mathbb{S}} \succeq 0\biggr\} \\
&\quad = \sup\biggl\{-\frac{1}{d}\langle Y_{\mathbb{S}}, \Sigma_{\mathbb{S}}\rangle_{\mathbb{S}} + \frac{1}{d}\underbrace{\langle X^0_{\mathbb{S}}, \Sigma_{\mathbb{S}}\rangle_{\mathbb{S}}}_{=\bar{\mathrm{tr}}(\Sigma_{\mathbb{S}})=d} : Y_{\mathbb{S}} \succeq_{\mathbb{S}} 0, A^*Y_{\mathbb{S}} \succeq I_d\biggr\} \\
&\quad = 1 - \frac{1}{d}\inf\biggl\{\langle Y_{\mathbb{S}}, \Sigma_{\mathbb{S}}\rangle_{\mathbb{S}} : Y_{\mathbb{S}} \succeq_{\mathbb{S}} 0, A^*Y_{\mathbb{S}} - Z = I_d,\ \text{for some } Z \succeq 0\biggr\}.
\end{aligned}$$

We write this optimisation problem in standard SDP form as follows: enumerate $\mathbb{S}$ as $\{S_1, \ldots, S_m\}$, and define

$$X := \begin{pmatrix} Y_{S_1} & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & Y_{S_m} & 0 \\ 0 & \cdots & 0 & Z \end{pmatrix},$$

so that $\langle Y_{\mathbb{S}}, \Sigma_{\mathbb{S}}\rangle_{\mathbb{S}} = \langle X, C\rangle$, where

$$C := \begin{pmatrix} \Sigma_{S_1} & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \Sigma_{S_m} & 0 \\ 0 & \cdots & 0 & 0 \end{pmatrix}.$$

As for the constraints, they are equivalent to $X \succeq 0$ and $\langle X, A^{jj'}\rangle = \delta_{jj'}$, for $j, j' \in [d]$, with

$$A^{jj'} := \begin{pmatrix} E_{S_1,jj'} & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & E_{S_m,jj'} & 0 \\ 0 & \cdots & 0 & -E_{jj'} \end{pmatrix},$$

where $E_{jj'} = (\boldsymbol{e}_j \boldsymbol{e}_{j'}^T + \boldsymbol{e}_{j'} \boldsymbol{e}_j^T)/2$ is the symmetric matrix of the same dimension as Z with its only non-zero entries being in the (j, j')-th and (j', j)-th positions, and where $E_{S,jj'} = (\boldsymbol{e}_{S,j} \boldsymbol{e}_{S,j'}^T + \boldsymbol{e}_{S,j'} \boldsymbol{e}_{S,j}^T)/2$ is the symmetric matrix of the same dimension as Y_S with its only non-zero entries being in the (j, j')-th and (j', j)-th positions of Y_S. Then, the standard dual problem is

$$\begin{aligned} &\sup\Big\{ \sum_{j,j' \in [d]} \delta_{j,j'} Y_{j,j'} : C - \sum_{j,j' \in [d]} Y_{j,j'} A^{jj'} \succeq 0 \Big\} \\ &\quad = \sup\Big\{ \operatorname{tr}(Y) : \Sigma_{\mathbb{S}} - \frac{1}{2} A \left(Y + Y^T\right) \succeq_{\mathbb{S}} 0, \left(Y + Y^T\right) \succeq 0 \Big\} \\ &\quad = \sup\Big\{ \operatorname{tr}(W) : \Sigma_{\mathbb{S}} - AW \succeq_{\mathbb{S}} 0, W \succeq 0 \Big\}, \end{aligned}$$

where we made the substitution $W = \left(Y + Y^T\right)/2$ and used the fact that $\operatorname{tr}(W) = \operatorname{tr}(Y)/2 + \operatorname{tr}(Y^T)/2 = \operatorname{tr}(Y)$. This shows that (18) is the dual problem of (17). As in the proof of Proposition 3, the result follows upon noticing that the primal problem (17) is strictly feasible, since $Y_{\mathbb{S}} = X^0_{\mathbb{S}} \succ_{\mathbb{S}} 0$ is such that $A^* Y_{\mathbb{S}} = I_d \succeq_{\mathbb{S}} I_d$, which ensures that strong duality holds. □

As before, we can prove some properties for $\tilde{R}(\cdot)$

Proposition 22. *The following hold:*

(i) $\tilde{R}$ *is convex.*

(ii) $\tilde{R}$ *is continuos.*

(iii) If $\mathbb{S} \subseteq \mathbb{S}'$ *and* $\Sigma_{\mathbb{S}} \subseteq \Sigma_{\mathbb{S}'}$*, then* $\tilde{R}(\Sigma_{\mathbb{S}}) \leq d' \tilde{R}(\Sigma_{\mathbb{S}'})/d$*, where* $d' = \operatorname{card}(\cup_{S \in \mathbb{S}'} S)$ *and* $d = \operatorname{card}(\cup_{S \in \mathbb{S}} S)$*.*

Proof of Proposition 22. (i) and (ii) are essentially the same as in Proposition 4. To prove (iii), let $\tilde{X}^{(1)}_{\mathbb{S}}$ be a feasible point of $\{X_{\mathbb{S}} + X^0_{\mathbb{S}} \succeq_{\mathbb{S}} 0, A^* X_{\mathbb{S}} \succeq 0\}$, and define $\tilde{X}^{(2)}_{\mathbb{S}'} := (\tilde{X}^{(1)}_{\mathbb{S}}, \boldsymbol{O}, \cdots, \boldsymbol{O})$, where we added a compatible zero matrix $\boldsymbol{O}$ for every element in $\mathbb{S}^C \cap \mathbb{S}'$. Then, $\tilde{X}^{(2)}_{\mathbb{S}'} + X^0_{\mathbb{S}'} \succeq_{\mathbb{S}'} 0$ is equivalent to $\tilde{X}^{(1)}_{\mathbb{S}} + X^0_{\mathbb{S}} \succeq_{\mathbb{S}} 0$ and $X^0_{\mathbb{S}' \setminus \mathbb{S}} \succeq_{\mathbb{S}' \setminus \mathbb{S}} 0$, which are satisfied, while $A^*_{\mathbb{S}'} \tilde{X}^{(2)}_{\mathbb{S}'} = A^*_{\mathbb{S}} \tilde{X}^{(1)}_{\mathbb{S}} \succeq 0$, since $\tilde{X}^{(1)}_{\mathbb{S}}$ is feasible. Hence, $\tilde{X}^{(2)}_{\mathbb{S}'}$ is feasible for $\Sigma^{(2)}_{\mathbb{S}'}$, and the thesis follows from the fact that the normalising constant changes from $1/d$ to

$1/d'$. Observe that the dual representation given by Proposition 21 allows proving the statement differently. Indeed, let $\Sigma \succeq 0 \in \mathbb{R}^{d',d'}$ be such that $\mathrm{tr}(\Sigma) = d'$ and

$$\Sigma_{\mathbb{S}'} = (1-\lambda')A_{\mathbb{S}'}\Sigma' + \lambda'\tilde{\Sigma}_{\mathbb{S}'},$$

where $\tilde{\Sigma}_{\mathbb{S}'} \succeq_{\mathbb{S}'} 0$ and $\lambda' = R(\Sigma_{\mathbb{S}'})$. Then, since $\mathbb{S} \subseteq \mathbb{S}'$, we can automatically write also Σ_S in this form as

$$\Sigma_{\mathbb{S}} = (1-\lambda')A_{\mathbb{S}}\Sigma + \lambda'\tilde{\Sigma}_{\mathbb{S}},$$

where Σ results from deleting all rows and columns of Σ' associated to every element $i \notin \cup_{S\in\mathbb{S}'}S \setminus \cup_{S\in\mathbb{S}}S$, and is ensured to be non-negative definite by Cauchy interlacing theorem. Then, calling σ_i^2 the diagonal elements of $(1-\lambda')\Sigma'$,

$$\begin{aligned}\tilde{R}(\Sigma_{\mathbb{S}}) &\le 1 - \frac{1}{d}\sum_{i\in\cup_{S\in\mathbb{S}}S}\sigma_i^2 \le 1 - \frac{1}{d}\left(\sum_{i\in[d']}\sigma_i^2 - (d'-d)\right)\\ &= 1 - \frac{1}{d}\Big(d(1-\tilde{R}(\Sigma_{\mathbb{S}'})) - (d'-d)\Big) = \frac{d'}{d}\tilde{R}(\Sigma_{\mathbb{S}'}).\end{aligned}$$

□

We conclude this section with one last example, where we come back to the measure $\tilde{R}$ to show how complex can it be even for very simple settings.

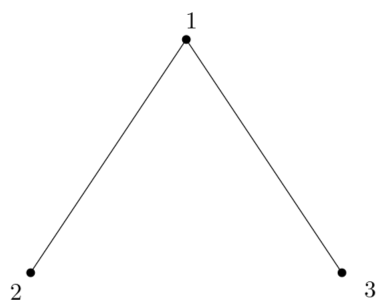

Figure 15: Graph associated to the pattern $\mathbb{S} = \{\{1,2\},\{1,3\}\}$.

Example 8. *Consider $\mathbb{S} = \{\{1,2\},\{1,3\}\}$, which is associated to the graph in Figure 15, and suppose without loss of generality that we observe*

$$\Sigma_{\{1,2\}} = \begin{pmatrix}\sigma_1^2 & \rho_{12}\sigma_1\sigma_2\\ \rho_{12}\sigma_1\sigma_2 & \sigma_2\end{pmatrix}, \quad \Sigma_{\{1,3\}} = \begin{pmatrix}\tilde{\sigma}_1^2 & \rho_{13}\tilde{\sigma}_1\sigma_2\\ \rho_{12}\tilde{\sigma}_1\sigma_2 & \sigma_3^2\end{pmatrix}$$

with $\tilde{\sigma}_1^2 \ge \sigma_1^2$. Let $\theta,\phi \in [0,\pi/2]$ be such that $\cos\theta = \sigma_1/\tilde{\sigma}_1$ and $\cos\phi = |\rho_{13}|$. Then we have

$$R(\Sigma_{\mathbb{S}}) = \frac{1}{6}(\tilde{\sigma}_1^2 - \sigma_1^2) + \frac{1}{3}\sigma_3^2\sin^2\big((\theta-\phi)_+\big).$$

Proof of Example 8. We prove this statement by giving an optimal choice of $X_{\mathbb{S}}$ for the primal problem and an optimal choice of Σ for the dual problem. It turns out that the optimal $X_{\mathbb{S}}$ is of the form

$$X_{\mathbb{S}} = \left(\begin{pmatrix}\lambda & 0\\ 0 & 0\end{pmatrix}, uu^T - \begin{pmatrix}\lambda & 0\\ 0 & 0\end{pmatrix}\right)$$

for $\lambda \in \mathbb{R}$ and $v \in \mathbb{R}^2$. Given $v \in \mathbb{R}^2$, we take $\lambda = 1/2 + 3v_1^2/(1+3v_2^2)$ as this is the maximal value for which $X_{\mathbb{S}} + X_{\mathbb{S}}^0 \succeq_{\mathbb{S}} 0$. It is clear that $A^* X_{\mathbb{S}} \succeq 0$, so this choice of λ always leads to a feasible $X_{\mathbb{S}}$. When $\phi \geq \theta$ we will simply take $\lambda = 1/2$ and $v = 0$ to recover the same feasible solution as for $\bar{R}$ and the simple lower bound $R(\Sigma_{\mathbb{S}}) \geq (1/6)(\tilde{\sigma}_1^2 - \sigma_1^2)$. When $\phi = 0$ (so that $|\rho_{13}| = 1$) we take $v = \mu(\sigma_3/\tilde{\sigma}_1, -\operatorname{sgn}(\rho_{13}))$ with $\mu \to \infty$ to see that

$$R(\Sigma_{\mathbb{S}}) \geq \sup_{\mu \geq 0}\left(\frac{1}{6} + \mu^2 \frac{\sigma_3^2/\tilde{\sigma}_1^2}{1+3\mu^2}\right)(\tilde{\sigma}_1^2 - \sigma_1^2) = (1/6)(\tilde{\sigma}_1^2 - \sigma_1^2) + (1/3)\sigma_3^2(1 - \sigma_1^2/\tilde{\sigma}_1^2),$$

which matches our claim. When $\theta > \phi > 0$ we choose

$$v = \sqrt{\frac{\sin(\theta-\phi)}{\cos(\theta)\sin(\phi)}}\begin{pmatrix}(\sigma_3/\tilde{\sigma}_1)\cos(\theta-\phi)\\ -\operatorname{sgn}(\rho_{13})\cos(\theta)\end{pmatrix}.$$

Using trigonometric identities, it can be seen that $\lambda = 1/2 + (\sigma_3^2/\tilde{\sigma}_1^2)\frac{\sin(\theta-\phi)\cos(\theta-\phi)}{\sin(\theta)\cos(\theta)}$ and

$$\begin{aligned}
&R(\Sigma_{\mathbb{S}}) \geq -(1/3)\langle X_{\mathbb{S}}, \Sigma_{\mathbb{S}}\rangle_{\mathbb{S}}\\
&= (1/3)\big\{\lambda\tilde{\sigma}_1^2\sin^2(\theta) - v_1^2\tilde{\sigma}_1^2 - 2v_1v_2\tilde{\sigma}_1\sigma_3\cos(\phi)\operatorname{sgn}(\rho_{13}) - v_2^2\sigma_3^2\big\}\\
&= \frac{1}{6}\tilde{\sigma}_1^2\sin^2(\theta) + \frac{1}{3}\sigma_3^2\frac{\sin(\theta-\phi)}{\cos(\theta)\sin(\phi)}\big\{\cos(\theta-\phi)\sin(\theta)\sin(\phi) - \cos^2(\theta-\phi)\\
&\qquad + 2\cos(\theta-\phi)\cos(\theta)\cos(\phi) - \cos^2(\theta)\big\}\\
&= \frac{1}{6}\tilde{\sigma}_1^2\sin^2(\theta) + \frac{1}{3}\sigma_3^2\sin^2(\theta-\phi).
\end{aligned}$$

We have now provided the required lower bound in all cases, and turn to the upper bound through the dual problem. Start first with the case that $\phi \geq \theta$. Then $\tilde{\sigma}_1|\rho_{13}| \leq \sigma_1$ so that

$$\Sigma = \begin{pmatrix}\sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \sigma_1\sigma_3\frac{\tilde{\sigma}_1\rho_{13}}{\sigma_1}\\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \sigma_2\sigma_3\rho_{12}\frac{\tilde{\sigma}_1\rho_{13}}{\sigma_1}\\ \sigma_1\sigma_3\frac{\tilde{\sigma}_1\rho_{13}}{\sigma_1} & \sigma_2\sigma_3\rho_{12}\frac{\tilde{\sigma}_1\rho_{13}}{\sigma_1} & \sigma_3^2\end{pmatrix}$$

is a valid covariance matrix. We have

$$\Sigma_{\mathbb{S}} - A\Sigma = \left(0, \begin{pmatrix}\tilde{\sigma}_1^2 - \sigma_1^2 & 0\\ 0 & 0\end{pmatrix}\right) \succeq 0$$

so Σ is feasible. Thus, when $\phi \geq \theta$, we have

$$R(\Sigma_{\mathbb{S}}) \leq 1 - \frac{1}{3}\operatorname{tr}(\Sigma) = \frac{\tilde{\sigma}_1^2 + \sigma_1^2}{6} + \frac{\sigma_2^2 + \sigma_3^2}{3} - \frac{1}{3}(\sigma_1^2 + \sigma_2^2 + \sigma_3^2) = \frac{1}{6}(\tilde{\sigma}_1^2 - \sigma_1^2)$$

as required. When $\phi < \theta$ we consider

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \sigma_1\sigma_3\cos(\theta-\phi)\,\mathrm{sgn}(\rho_{13}) \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \sigma_2\sigma_3\rho_{12}\cos(\theta-\phi)\,\mathrm{sgn}(\rho_{13}) \\ \sigma_1\sigma_3\cos(\theta-\phi)\,\mathrm{sgn}(\rho_{13}) & \sigma_2\sigma_3\rho_{12}\cos(\theta-\phi)\,\mathrm{sgn}(\rho_{13}) & \sigma_3^2\cos^2(\theta-\phi) \end{pmatrix},$$

which is a covariance matrix so $\Sigma \succeq 0$. Clearly $(A\Sigma)_{\{1,2\}} = \Sigma_{\{1,2\}}$. It follows from trigonometric identities that

$$\begin{aligned} \tilde{\sigma}_1\sigma_3\rho_{13} - \sigma_1\sigma_3\cos(\theta-\phi)\,\mathrm{sgn}(\rho_{13}) &= \tilde{\sigma}_1\sigma_3\,\mathrm{sgn}(\rho_{13})\{\cos(\phi) - \cos(\theta)\cos(\theta-\phi)\} \\ &= \tilde{\sigma}_1\sigma_3\,\mathrm{sgn}(\rho_{13})\sin(\theta)\sin(\theta-\phi) \end{aligned}$$

so that

$$\Sigma_{\{1,3\}} - (A\Sigma)_{\{1,3\}} = \begin{pmatrix} \tilde{\sigma}_1^2\sin^2(\theta) & \tilde{\sigma}_1\sigma_3\,\mathrm{sgn}(\rho_{13})\sin(\theta)\sin(\theta-\phi) \\ \tilde{\sigma}_1\sigma_3\,\mathrm{sgn}(\rho_{13})\sin(\theta)\sin(\theta-\phi) & \sigma_3^2\sin^2(\theta-\phi) \end{pmatrix},$$

which is a covariance matrix so is positive semi-definite. Thus Σ is feasible and when $\phi < \theta$ we have

$$\begin{aligned} R(\Sigma_{\mathbb{S}}) \le 1 - \frac{1}{3}\mathrm{tr}(\Sigma) &= \frac{\tilde{\sigma}_1^2 + \sigma_1^2}{6} + \frac{\sigma_2^2 + \sigma_3^2}{3} - \frac{1}{3}\{\sigma_1^2 + \sigma_2^2 + \sigma_3^2\cos^2(\theta-\phi)\} \\ &= \frac{1}{6}(\tilde{\sigma}_1^2 - \sigma_1^2) + \frac{1}{3}\sin^2(\theta-\phi), \end{aligned}$$

as required. □

Now, the goal of this subsection is to develop an analogous oracle test for the measure $\tilde{R}$, under the usual hypothesis of $\Sigma_{\mathbb{S}} \succeq_{\mathbb{S}} cI_{\mathbb{S}}$, with $c > 0$. In this case, the maximum is attained in the set

$$\mathcal{H}_c := \{X_{\mathbb{S}} + X^0_{\mathbb{S}} \succeq_{\mathbb{S}} 0, A^*X_{\mathbb{S}} \succeq 0, \langle X_{\mathbb{S}} + X^0_{\mathbb{S}}, cI_{\mathbb{S}}\rangle_{\mathbb{S}} \le d\},$$

hence the only difference with $\mathcal{F}_c$ is that $A^*X_{\mathbb{S}} + Y \succeq 0$ for some $Y \in \mathcal{Y}$ is substituted by $A^*X_{\mathbb{S}} \succeq 0$. Hence, since in the previous subsection we discarded the condition $A^*X_{\mathbb{S}} + Y \succeq 0$ for some $Y \in \mathcal{Y}$, if we now discard $A^*X_{\mathbb{S}} \succeq 0$, all the previous steps remain valid for controlling $\mathbb{P}_{H_0}\left(\tilde{R}(\widehat{\Sigma}_{\mathbb{S}}) \ge C_\alpha\right)$, so that we can again reduce this problem to bounding $\max_{S\in\mathbb{S}} \|\widehat{\Sigma}_S - \Sigma_S\|_2$, with the only difference that now Σ_S are the population covariance matrices and $\widehat{\Sigma}_S$ are the corresponding sample covariance matrices, which makes the problem slightly easier in light of standard concentration inequalities (e.g. Theorem 6.5 in Wainwright (2019)). In this regard, repeating the same steps that lead to the proof of Theorem 6, we can prove the following result, which gives the right separation to test compatibility based on $\tilde{R}$. Of course, this could be generalised easily to include a test for the consistency of the variances based on $V(\sigma^2_{\mathbb{S}})$.

Proposition 23. *Suppose we observe* $\mathbf{X}_{S,1}, \ldots, \mathbf{X}_{S,n_S} \overset{i.i.d.}{\sim} P_S, \forall S \in \mathbb{S}$ *independently, where each* P_S *is* ν*-subgaussian with* $\nu \gg 1$*, with the sequence of population covariance matrices* $\Sigma_{\mathbb{S}}$ *satisfying* $\bar{\mathrm{tr}}(\Sigma_{\mathbb{S}}) = d$*, and* $\Sigma_{\mathbb{S}} \succeq_{\mathbb{S}} cI_{\mathbb{S}}$*, for a given* $c > 0$*. Let* $\widehat{\Sigma}_{\mathbb{S}}$ *be the sequence of sample covariance matrix associated to each pattern* $S \in \mathbb{S}$*,* $n_{\mathbb{S}}$ *the sequence of sample sizes, and suppose that also* $\widehat{\Sigma}_{\mathbb{S}}$ *are normalised so that* $\bar{\mathrm{tr}}(\widehat{\Sigma}_{\mathbb{S}}) = d$*. Then,* *for all $\alpha \in (0,1)$, the test that rejects H_0 if and only if $\widehat{R} \geq C_\alpha$ has Type I error bounded by α, where*

$$C_\alpha = \frac{C_1 \nu^2}{c} \max_{S \in \mathbb{S}} \sqrt{\frac{|S| + \log(|\mathbb{S}|/\alpha)}{n_S}} \vee \frac{|S| + \log(|\mathbb{S}|/\alpha)}{n_S},$$

and $C_1 > 0$ is a universal constant. Moreover, for $\beta \in (0,1)$, if $R(\Sigma_{\mathbb{S}}) > C_\alpha + C_\beta$, then $\mathbb{P}(\widehat{R} \leq C_\alpha) \leq \beta$.

The proof is essential analogous to the one of Theorem 6, except for the fact that now we used a standard concentration inequality for covariance matrices (see Proposition 28 in Appendix D) in place of Proposition 7. Also, observe that the separation rate in this case is slightly better than the one we found in Theorem 6, being of the order of

$$C_\alpha \lesssim \max_{S \in \mathbb{S}} \sqrt{\frac{|S| + \log(|\mathbb{S}|/\alpha)}{n_S}},$$

under $n_S \gtrsim |S|$ for all $|S| \in \mathbb{S}$, which is necessary to have a consistent test. As far as the drawbacks are concerned, notice that here we need to normalise the sample covariance matrix a priori, so that $\bar{\operatorname{tr}}(\widehat{\Sigma}_{\mathbb{S}}) = d$, which is somehow annoying. What is even more disturbing is the hypothesis that the subgaussian proxy ν^2 needs to be significantly bigger than one, due to the fact that for a ν-subgaussian random variable X we have $\mathrm{Var}[X] \leq \nu^2$. Hence, the hypothesis $\nu^2 \gg 1$ is necessary to have a little flexibility in the variances, while still satisfying $\bar{\operatorname{tr}}(\widehat{\Sigma}_{\mathbb{S}}) = d$. There is no reason to assume that $\nu^2 \gg 1$, so that this is another point in favour of the incompatibility measure R. As before, Proposition 23 can be used to derive a test based on $\tilde{R}$ which uses sample splitting. Repeating the steps which lead to Proposition 8, we can prove the following:

Proposition 24. *Suppose we observe $\mathbf{X}_{S,1}, \ldots, \mathbf{X}_{S,n_S} \overset{i.i.d.}{\sim} P_S, \forall S \in \mathbb{S}$ independently, where each P_S is ν-subgaussian with $\nu \gg 1$, and that $\Sigma_{\mathbb{S}} \succ_{\mathbb{S}} 0$. Then, we partition the data into two parts, $\mathcal{X}_1$ and $\mathcal{X}_2$, and use $\mathcal{X}_1$ to find the optimal $\widehat{X}^{(1)}_{\mathbb{S}}$ based on the estimate $\widehat{\Sigma}^{(1)}_{\mathbb{S}}$, and $\mathcal{X}_2$ to produce the independent estimates $\widehat{\Sigma}^{(2)}_{\mathbb{S}}$ and find the sample sizes n_S. For $\alpha \in (0,1)$, define $C_\alpha(\mathcal{X}_1) > 0$ by*

$$C_\alpha(\mathcal{X}_1) := \epsilon \|\widehat{X}^{(1)}_{\mathbb{S}}\|_{*,\mathbb{S}}/d,$$

where ϵ is such that

$$\epsilon := C_1 \nu^2 \max_{S \in \mathbb{S}} \sqrt{\frac{|S| + \log(|\mathbb{S}|/\alpha)}{n_S}} \vee \frac{|S| + \log(|\mathbb{S}|/\alpha)}{n_S}.$$

Then, the test that rejects H_0 if and only if $-\frac{1}{d}\langle \widehat{X}^{(1)}_{\mathbb{S}}, \widehat{\Sigma}^{(2)}_{\mathbb{S}} \rangle_{\mathbb{S}} \geq C_\alpha(\mathcal{X}_1)$ has Type I error bounded by α.

This test leads to a testing separation rate of the order of

$$C_\alpha(\mathcal{X}_1) \lesssim \frac{\|\widehat{X}^{(1)}_{\mathbb{S}}\|_{*,\mathbb{S}}}{d} \max_{S \in \mathbb{S}} \sqrt{\frac{|S| + \log(|\mathbb{S}|/\alpha)}{n_S}},$$

with high probability. As in Proposition 6, also these tests can be slightly modified to include a term that checks consistency of the variances, based on the population measure of inconsistency $V(\sigma^2_{\mathbb{S}})$.

Appendix C Auxiliary results in SDP

Semi-definite programs are linear optimisation problems over spectrahedra, i.e. sets of the form

$$S = \left\{ (x_1, \ldots, x_m) \in \mathbb{R}^m : A_0 + \sum_{i=1}^{m} A_i x_i \succeq 0 \right\},$$

for some given symmetric matrices $A_0, A_1, \ldots, A_m$. An SDP problem in standard primal form is written as

$$\begin{cases} \text{minimize} & \langle C, X \rangle \\ \text{subject to} & X \succeq 0 \text{ and } \langle A_i, X \rangle = b_i, \quad i \in [m], \end{cases}$$

where C, A_i are given symmetric matrices, and b_i are given scalars. For every semi-definite program in primal form, there is another associated SDP, called the dual problem, that can be stated as

$$\begin{cases} \text{maximize} & b^T y \\ \text{subject to} & \sum_{i=1}^{m} A_i y_i \preceq C, \end{cases}$$

where $b = (b_1, \ldots, b_m)$, and $y = (y_1, \ldots, y_m)$ are the dual decision variables. As in linear programming, the so-called weak duality holds, meaning that if X and y are any two feasible solutions of the primal and dual problems respectively, we have

$$\langle C, X \rangle - b^T y = \langle C, X \rangle - \sum_{i=1}^{m} y_i \langle A_i, X \rangle = \left\langle C - \sum_{i=1}^{m} A_i y_i, X \right\rangle \geq 0.$$

Unfortunately, the equality is not always satisfied in general (see Example 2.14. in Blekherman et al. (2012)), but under some mild conditions, strong duality holds. One of such conditions is Slater's condition, where either the primal or the dual problem is required to be strictly feasible, meaning that there exists either $X \succ 0$ for the primal problem satisfying $\langle A_i, X \rangle = b_i$, for $i \in [m]$, or y for the dual satisfying $\sum_{i=1}^{m} A_i y_i \prec C$. If this is the case, it can be shown that strong duality holds (Theorem 2.15. in Blekherman et al. (2012), Theorem 3.1. in Vandenberghe and Boyd (1996)). Furthermore, if the primal is strictly feasible, then the dual optimum is attained, and viceversa. In the proof, we show that it is possible to define R as the optimal value of an SDP problem written in primal form, find its dual and show that Slater's condition is satisfied. This, apart from enabling us to prove Proposition 3, ensures that R can be computed explicitly using standard SDP libraries, which are available for almost all programming languages. As for the computational cost, for SDP problems in their general setting, without extra assumptions like strict complementarity, no polynomial-time algorithms are known, and there are examples of SDPs for which every solution needs exponential space (Khachiyan and Porkolab, 1997). Moreover, Ramana (1997) showed that SDP lies either in the intersection of NP and co-NP, or outside the union of NP and co-NP, and nothing better than this is known. Luckily, if Slater's condition is satisfied, like in our case, then the primal-dual interior point method has a computational complexity which is polynomial in the number of constraints and the dimension of the unknown square matrix (Section 6.4.1. of Nesterov and Nemirovskii (1994), Section 5.7. of Vandenberghe and Boyd (1996)), which ensures that R can be always computed efficiently without additional assumptions.

Finally, we recall Farkas' lemma for SDP problems, and its proof, following Lemma 6.3.2 in Lovász (2003).

Proposition 25 (Farkas' lemma for Semi-definite Programming)**.** *Let $A_1, \ldots, A_n$ be symmetric $m \times m$ matrices. The system*

$$x_1 A_1 + \ldots + x_n A_n \succ 0$$

has no solution in $x_1, \ldots, x_n$ if and only if there exists a symmetric matrix $Y \neq 0$ such that

$$\begin{cases} \langle A_1, Y \rangle = 0 \\ \langle A_2, Y \rangle = 0 \\ \quad \vdots \\ \langle A_n, Y \rangle = 0 \\ Y \succeq 0. \end{cases}$$

Proof. The set $\mathcal{P}_m^*$ of $m \times m$ positive semi-definite matrices forms a closed convex cone. If

$$x_1 A_1 + \ldots + x_n A_n \succ 0$$

has no solution, then the linear subspace $\mathcal{L}$ of matrices of the form $x_1 A_1 + \ldots x_n A_n$ is disjoint from the interior of $\mathcal{P}_m^*$, which in turn implies that $\mathcal{L}$ is contained in a hyperplane that is disjoint from the interior of $\mathcal{P}_m^*$. This hyperplane can be described as $\{X \in \mathcal{P}_m^* : \langle Y, X \rangle = 0\}$ for a certain symmetric Y, where we may assume that $\langle Y, X \rangle \geq 0$ for every $X \in \mathcal{P}_m^*$. Then, since a matrix A is positive semi-definite if and only if $\langle A, B \rangle \geq 0$ for every positive semi-definite matrix B, we conclude that $Y \neq 0$, $Y \succeq 0$, and, since A_i belong to $\mathcal{L}$, that $\langle A_i, Y \rangle = 0$. □

Appendix D Technical inequalities

Proposition 26 (Tail bound for a sum of subexponential RVs)**.** *Consider an independent sequence $\{X_k\}_{k=1}^n$ of random variables, such that X_k has mean μ_k, and is sub-exponential with parameters (v_k, α_k). Then, $\sum_{k=1}^n (X_k - \mu_k)$ is sub-exponential with the parameters (v_*, α_*), where*

$$\alpha_* := \max_{k=1,\ldots,n} \alpha_k \quad \text{and} \quad v_* := \sqrt{\sum_{k=1}^n v_k^2},$$

and

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^n (X_k - \mu_k) \right| \geq t \right) \leq \begin{cases} 2e^{-\frac{n^2 t^2}{2v_*^2}} & \text{for } 0 \leq t \leq \frac{v_*^2}{n\alpha_*} \\ 2e^{-\frac{nt}{2\alpha_*}} & \text{for } t > \frac{v_*^2}{n\alpha_*}. \end{cases}$$

Proof. See Proposition 2.9. in Wainwright (2019). □

Proposition 27 (The square of a subgaussian is subexponential)**.** *If X is σ-subgaussian, then X^2 is subexponential with parameters $(\nu, \alpha) = (4\sqrt{2}\sigma^2, 4\sigma^2)$.*

Proof. Using the definitions of the Orlicz norm $\|\cdot\|_{\psi_1}$ and $\|\cdot\|_{\psi_2}$ (see Wainwright (2019); Vershynin (2019)), it is easy to prove that the product of two subgaussian RVs is subexponential (Lemma 2.7.7. in Vershynin (2019)), and that X is subgaussian if and only if X^2 is subexponential (Lemma 2.7.6. in Vershynin (2019)). As for its subexponential parameters, assuming WLOG that X has mean zero, we know that

$$\mathbb{E}\left[e^{\lambda X}\right] \leq e^{\frac{1}{2}\lambda^2\sigma^2}, \quad \text{for all } \lambda \in \mathbb{R}.$$

Our goal is to find a similar bound for the moment generating function of X^2, and, to this aim, we will make use of the fact that the moments of X are bounded as follows

$$\mathbb{E}\left[|X|^r\right] \leq r2^{r/2}\sigma^r\Gamma(r/2), \quad \text{for all } r > 0,$$

where $\Gamma(r)$ is the Gamma function. Now, calling $\mu = \mathbb{E}[X^2]$, by power series expansion and since $\Gamma(r) = (r-1)!$ for an integer r, we have

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda\left(X^2-\mu\right)}\right] &= 1 + \lambda\mathbb{E}\left[X^2-\mu\right] + \sum_{r=2}^{\infty}\frac{\lambda^r\mathbb{E}\left[\left(X^2-\mu\right)^r\right]}{r!} \\
&\leq 1 + \sum_{r=2}^{\infty}\frac{\lambda^r\mathbb{E}\left[|X|^{2r}\right]}{r!} \leq 1 + \sum_{r=2}^{\infty}\frac{\lambda^r 2r2^r\sigma^{2r}\Gamma(r)}{r!} \\
&= 1 + \sum_{r=2}^{\infty}\lambda^r 2^{r+1}\sigma^{2r} = 1 + \frac{8\lambda^2\sigma^4}{1-2\lambda\sigma^2}.
\end{aligned}$$

By making $|\lambda| \leq 1/4\sigma^2$, we have $1/\left(1-2\lambda\sigma^2\right) \leq 2$. Finally, since for every $\alpha \in \mathbb{R}$ it holds $1+\alpha \leq e^{\alpha}$, we have that the MGF of X^2 satisfies

$$\mathbb{E}\left[e^{\lambda\left(X^2-\mathbb{E}\left[X^2\right]\right)}\right] \leq e^{16\lambda^2\sigma^4}, \quad \text{for all } |\lambda| \leq 1/\left(4\sigma^2\right).$$

Thus, we obtained a bound for the moment generating function of the subexponential variable X^2, that is similar to that of subgaussian variables but holds only for a small range of λ. ☐

Proposition 28 (Concentration inequality for Covariance Matrices)**.** *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be an i.i.d sequence of σ-subgaussian random vectors with covariance matrix Σ and let $\widehat{\Sigma}_n := \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i^T$ be the sample covariance matrix. Then there exists a universal constant $C > 0$ such that, for $\delta \in (0,1)$, with probability at least $1-\delta$*

$$\|\widehat{\Sigma}_n - \Sigma\|_2 \leq C\sigma^2 \max\left\{\sqrt{\frac{d+\log(2/\delta)}{n}}, \frac{d+\log(2/\delta)}{n}\right\}.$$

Proof. We break the proof up into two steps: use a discretisation argument to reduce the problem to the task of computing the maximum of finitely many random variables, and then use standard concentration inequalities. Firstly, let $A \in S^{d\times d}$ and let N_ϵ be an ϵ-net of the d-dimensional sphere $\mathbb{S}^{d-1}$. Then

$$\|A\|_2 \leq \frac{1}{1-2\epsilon}\max_{y\in N_\epsilon}\left|y^T A y\right|.$$

Indeed, let $y \in N_\epsilon$ satisfy $\|x-y\| \leq \epsilon$. Then

$$\begin{aligned}\left|xAx - y^TAy\right| &= \left|x^TA(x-y) + y^TA(x-y)\right| \\ &\leq \left|x^TA(x-y)\right| + \left|y^TA(x-y)\right|\end{aligned}$$

Looking at $\left|x^TA(x-y)\right|$ we have

$$\begin{aligned}\left|x^TA(x-y)\right| &\leq \|A(x-y) \mid \| \|x\| \\ &\leq \|A\|_2 \underbrace{\|x-y\|}_{\leq \epsilon} \underbrace{\|x\|}_{=1} \\ &\leq \|A\|_2 \epsilon\end{aligned}$$

Applying the same argument to $\left|y^TA(x-y)\right|$ gives us $\left|x^Ax - y^TAy\right| \leq 2\epsilon\|A\|_2$. To complete the proof, we see that $\|A\|_2 = \max_{x \in \mathbb{S}^{d-1}} x^TAx \leq 2\epsilon\|A\|_2 + \max_{y \in N_\epsilon} y^TAy$. Rearranging the equation gives $\|A\|_2 \leq \frac{1}{1-2\epsilon} \max_{y \in N_\epsilon} y^TAy$ as desired. Then, if we apply this result to $\widehat{\Sigma}_n - \Sigma$ with $\epsilon = 1/4$ we have

$$\|\widehat{\Sigma}_n - \Sigma\|_2 \leq 2 \max_{v \in N_{1/4}} \left|v^T \left(\widehat{\Sigma}_n - \Sigma\right) v\right|$$

Additionally, we know that $\text{card}(N_{1/4}) \leq 9^d$ (see Lemma 5.7 and Example 5.8 in Wainwright (2019). From here, we can apply standard concentration tools to get

$$\begin{aligned}\mathbb{P}\left(\|\widehat{\Sigma}_n - \Sigma\|_2 \geq t\right) &\leq \mathbb{P}\left(\max_{v \in N_{1/4}} \left|v^T \left(\widehat{\Sigma}_n - \Sigma\right) v\right| \geq t/2\right) \\ &\leq \text{card}(N_{1/4}) \cdot \mathbb{P}\left(\left|v_i^T \left(\widehat{\Sigma}_n - \Sigma\right) v_i\right| \geq t/2\right),\end{aligned}$$

where v_i is a unit vector on the d-dimensional sphere. Now, $v_i^T \left(\widehat{\Sigma}_n - \Sigma\right) v_i$ can be rewritten as

$$\begin{aligned}v_i^T \left(\widehat{\Sigma}_n - \Sigma\right) v_i &= \frac{1}{n}\sum_{j=1}^n \left(v_i^T \mathbf{X}_j\right)^2 - \mathbb{E}\left[\left(v_i^T \mathbf{X}_j\right)^2\right] \\ &= \frac{1}{n}\sum_{j=1}^n Z_j - \mathbb{E}\left[Z_j\right],\end{aligned}$$

where the $Z_j - \mathbb{E}[Z_j]$ are independent subexponential of parameters $(\nu, \alpha) = \left(4\sqrt{2}\sigma^2, 4\sigma^2\right)$, since $v_i^T\mathbf{X}_j$ are σ-subgaussian by definition of subgaussian random vector. Applying the subexponential tail bound in Proposition 26 gives us

$$\mathbb{P}\left(\left|v_i^T \left(\widehat{\Sigma}_n - \Sigma\right) v_i\right| \geq t/2\right) \leq 2\exp\left\{-n \min\left\{\left(\frac{t}{16\sigma^2}\right)^2, \frac{t}{16\sigma^2}\right\}\right\}.$$

so that

$$\mathbb{P}\left(\|\widehat{\Sigma}_n - \Sigma\|_2 \geq t\right) \leq 2 \cdot 9^d \exp\left\{-n \min\left\{\left(\frac{t}{16\sigma^2}\right)^2, \frac{t}{16\sigma^2}\right\}\right\}.$$

Inverting the bound gives the desired result. For further reference, please refer to Chapter 3 in Wainwright (2019). □