Research Article

Numerical Reconstruction of the Covariance Matrix of a Spherically Truncated Multinormal Distribution

Filippo Palombi,[1,2] Simona Toti,[1] and Romina Filippini[1]

[1]*Istituto Nazionale di Statistica (ISTAT), Via Cesare Balbo 16, 00184 Rome, Italy*
[2]*Italian Agency for New Technologies, Energy and Sustainable Economic Development (ENEA),*
 Via Enrico Fermi 45, 00044 Frascati, Italy

Correspondence should be addressed to Filippo Palombi; filippo.palombi@enea.it

We relate the matrix $\mathfrak{S}_{\mathcal{B}}$ of the second moments of a spherically truncated normal multivariate to its full covariance matrix Σ and present an algorithm to invert the relation and reconstruct Σ from $\mathfrak{S}_{\mathcal{B}}$. While the eigenvectors of Σ are left invariant by the truncation, its eigenvalues are nonuniformly damped. We show that the eigenvalues of Σ can be reconstructed from their truncated counterparts via a fixed point iteration, whose convergence we prove analytically. The procedure requires the computation of multidimensional Gaussian integrals over an Euclidean ball, for which we extend a numerical technique, originally proposed by Ruben in 1962, based on a series expansion in chi-square distributions. In order to study the feasibility of our approach, we examine the convergence rate of some iterative schemes on suitably chosen ensembles of Wishart matrices. We finally discuss the practical difficulties arising in sample space and outline a regularization of the problem based on perturbation theory.

1. Introduction

It has been more than forty years since Tallis [1] worked out the moment-generating function of a normal multivariate $X \equiv \{X_k\}_{k=1}^{\nu} \sim \mathcal{N}_{\nu}(0, \Sigma)$, subject to the conditional event

$$X \in \mathscr{E}_{\nu}(\rho; \Sigma), \quad \mathscr{E}_{\nu}(\rho; \Sigma) \equiv \left\{ x \in \mathbb{R}^{\nu} : x^{\mathrm{T}} \Sigma^{-1} x \leq \rho \right\}. \quad (1)$$

The perfect match between the symmetries of the ellipsoid $\mathscr{E}_{\nu}(\rho; \Sigma)$ and those of $\mathcal{N}_{\nu}(0, \Sigma)$ allows for an exact analytic result, from which the complete set of multivariate truncated moments can be extracted upon differentiation. Consider, for instance, the matrix $\mathfrak{S}_{\mathscr{E}}(\rho; \Sigma)$ of the second truncated moments, expressing the covariances among the components of X within $\mathscr{E}_{\nu}(\rho; \Sigma)$. From Tallis' paper it turns out that

$$\mathfrak{S}_{\mathscr{E}}(\rho; \Sigma) = c_{\mathrm{T}}(\rho) \Sigma, \quad c_{\mathrm{T}}(\rho) \equiv \frac{F_{\nu+2}(\rho)}{F_{\nu}(\rho)}, \quad (2)$$

with F_{ν} denoting the cumulative distribution function of a χ^2-variable with ν degrees of freedom. Inverting (2)—so as to express Σ as a function of $\mathfrak{S}_{\mathscr{E}}$—is trivial, since $c_{\mathrm{T}}(\rho)$ is a scalar

damping factor independent of Σ. In this paper, we shall refer to such inverse relation as the *reconstruction* of Σ from $\mathfrak{S}_{\mathscr{E}}$. Unfortunately, life is not always so easy. In general, the effects produced on the expectation of functions of X by cutting off the probability density outside a generic domain $\mathscr{D} \subset \mathbb{R}^{\nu}$ can be hardly calculated in closed form, especially if the boundary of $\mathscr{D}$ is shaped in a way that breaks the ellipsoidal symmetry of $\mathcal{N}_{\nu}(0, \Sigma)$. Thus, for instance, unlike (2), the matrix of the second truncated moments is expected in general to display a nonlinear/nontrivial dependence upon Σ.

In the present paper, we consider the case where $\mathscr{D}$ is a ν-dimensional Euclidean ball with center in the origin and square radius ρ. Specifically, we discuss the reconstruction of Σ from the matrix $\mathfrak{S}_{\mathscr{B}}$ of the spherically truncated second moments. To this aim, we need to mimic Tallis' calculation, with (1) being replaced by the conditional event

$$X \in \mathscr{B}_{\nu}(\rho), \quad \mathscr{B}_{\nu}(\rho) \equiv \left\{ x \in \mathbb{R}^{\nu} : x^{\mathrm{T}} x \leq \rho \right\}. \quad (3)$$

This is precisely an example of the situation described in the previous paragraph: although $\mathscr{B}_{\nu}(\rho)$ has a higher degree of symmetry than $\mathscr{E}_{\nu}(\rho; \Sigma)$, still there is no possibility

of obtaining a closed-form relation between Σ and $\mathfrak{S}_{\mathscr{B}}$, since $\mathscr{B}_v(\rho)$ breaks the ellipsoidal symmetry of $\mathscr{N}_v(0, \Sigma)$: technically speaking, in this case we cannot perform any change of variable under the defining integral of the moment-generating function, which may reduce the dimensionality of the problem, as in Tallis' paper.

In spite of that, the residual symmetries characterizing the truncated distribution help simplify the problem in the following respects: (i) the reflection invariance of the whole setup still yields $\mathbb{E}[X_k \mid X \in \mathscr{B}_v(\rho)] = 0 \; \forall k$ and (ii) the rotational invariance of $\mathscr{B}_v(\rho)$ preserves the possibility of defining the principal components of the distribution just like in the unconstrained case. In particular, the latter property means that $\mathfrak{S}_{\mathscr{B}}$ and Σ share the same orthonormal eigenvectors. In fact, the reconstruction of Σ from $\mathfrak{S}_{\mathscr{B}}$ amounts to solving a system of nonlinear integral equations, having the eigenvalues $\lambda \equiv \{\lambda_k\}_{k=1}^v$ of Σ as unknown variables and the eigenvalues $\mu \equiv \{\mu_k\}_{k=1}^v$ of $\mathfrak{S}_{\mathscr{B}}$ as input parameters. In a lack of analytic techniques to evaluate exactly the integrals involved, we resort to a numerical algorithm, of which we investigate feasibility, performance, and optimization.

The paper is organized as follows. In Section 2, we describe a few examples illustrating the occurrence of spherical truncations in practical situations. In Section 3, we show that the aforementioned integral equations have the analytic structure of a fixed point vector equation; that is to say, $\lambda = T(\lambda)$. This suggests achieving the reconstruction of λ numerically via suitably chosen iterative schemes. In Section 4, we prove the convergence of the simplest of them by inductive arguments, the validity of which relies upon the monotonicity properties of ratios of Gaussian integrals over $\mathscr{B}_v(\rho)$. In Section 5, we review some numerical techniques for the computation of Gaussian integrals over $\mathscr{B}_v(\rho)$ with controlled systematic error. These are based on and extend a classic work by Ruben [2] on the distribution of quadratic forms of normal variables. For the sake of readability, we defer proofs of statements made in this section to the appendix. In Section 6, we report on our numerical experiences: since the simplest iterative scheme, namely, the Gauss–Jacobi iteration, is too slow for practical purposes, we investigate the performance of its improved version based on overrelaxation; as expected, we find that the latter has a higher convergence rate; yet it still slows down polynomially in $1/\rho$ as $\rho \to 0$ and exponentially in v as $v \to \infty$; in order to reduce the slowing down, we propose an acceleration technique, which boosts the higher components of the eigenvalue spectrum. A series of Monte Carlo simulations enables us to quantify the speedup. In Section 7 we discuss the problems arising when μ is affected by statistical uncertainty and propose a regularization technique based on perturbation theory. To conclude, we summarize our findings in Section 8.

2. Motivating Examples

Spherical truncations of multinormal distributions may characterize different kinds of experimental devices and may occur in various problems of statistical and convex analysis. In this section, we discuss two motivating examples.

2.1. A Two-Dimensional Gedanken Experiment in Classical Particle Physics. Consider the following ideal situation. An accelerator physicist prepares an elliptical beam of classical particles with Gaussian transversal profile. The experimenter knows *a priori* the spatial distribution of the beam, that is, the covariance matrix Σ of the two-dimensional coordinates of the particles on a plane orthogonal to their flight direction. We can assume with no loss of generality that the transversal coordinate system has origin at the maximum of the beam intensity and axes along the principal components of the beam; thus it holds $\Sigma = \mathrm{diag}(\lambda_1, \lambda_2)$. The beam travels straightforward until it enters a linear coaxial pipeline with circular profile, schematically depicted in Figure 1, where the beam is longitudinally accelerated. While the outer part of the beam is stopped by an absorbing wall, the inner part propagates within the pipeline. At the end of the beam flight the physicist wants to know if the transversal distribution of the particles is changed, due to unknown disturbance factors arisen within the pipeline. Accordingly, he measures again the spatial covariance matrix of the beam. Unfortunately, the absorbing wall has cut off the Gaussian tail, thus damping the covariance matrix and making it no more comparable to the original one. To perform such a comparison in the general case $\lambda_1 \neq \lambda_2$, the covariance matrix of the truncated circular beam has to go through the reconstruction procedure described in next sections.

2.2. A Multivariate Example: Connections to Compositional Data Analysis. Compositional Data Analysis (CoDA) has been the subject of a number of papers, pioneered by Aitchison [3] over the past forty years. As a methodology of statistical investigation, it finds application in all cases where the main object of interest is a multivariate with strictly positive continuous components to be regarded as portions of a total amount κ (the normalization $\kappa = 1$ is conventionally adopted in the mathematical literature). In other words, compositional variates belong to the κ-simplex:

$$\mathscr{S}_v = \{z \in \mathbb{R}_+^v : |z|_1 = \kappa\}, \quad v \geq 2, \tag{4}$$

with $|z|_1 = \sum_{k=1}^v z_k$ the taxi-cab norm of z, while compositions with different norms can be always projected onto $\mathscr{S}_v$ via the closure operator $\mathscr{C} \cdot x \equiv \{\kappa x_1/|x|_1, \ldots, \kappa x_v/|x|_1\}$. There are countless types of compositional data, whose analysis raises problems of interest for statistics [4], for example, geochemical data, balance sheet data, and election data. The simplex constraint induces a kind of dependency among the parts of a composition that goes beyond the standard concept of covariance. This invalidates many ordinary techniques of statistical analysis.

In order to measure distances on $\mathscr{S}_v$, Aitchison introduced a positive symmetric function $d_A : \mathscr{S}_v \times \mathscr{S}_v \to \mathbb{R}_+$, explicitly defined by

$$d_A(x, y) = \sqrt{\frac{1}{2v} \sum_{i,k=1}^v \left[\log\left(\frac{x_i}{x_k}\right) - \log\left(\frac{y_i}{y_k}\right) \right]^2}. \tag{5}$$

The Aitchison distance is a key tool in CoDA. It is scale invariant in both its first and second arguments; that is, it
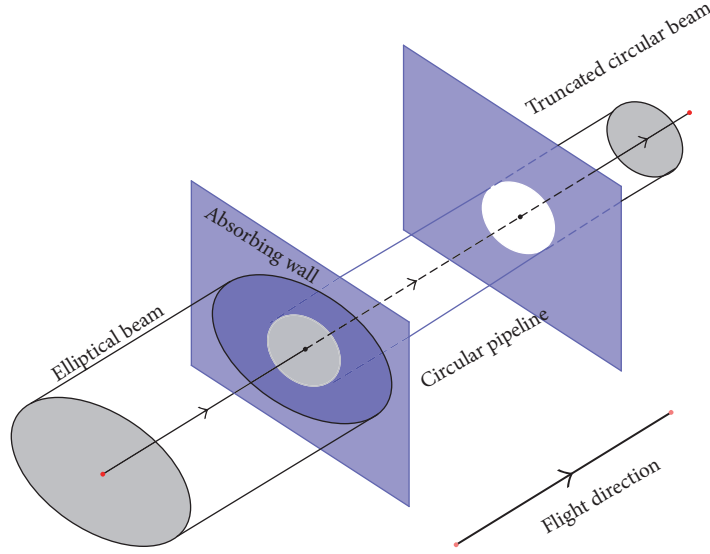
FIGURE 1: A classical particle beam with elliptical transversal profile is cut off upon entering a circular coaxial pipeline.

is left invariant by redefinitions $z \rightarrow \{\alpha z_1, \ldots, \alpha z_\nu\}$ with $\alpha \in \mathbb{R}_+$. Accordingly, its support can be extended to $\mathbb{R}_+^\nu \times \mathbb{R}_+^\nu$ by imposing

$$d_A(x, y) \equiv d_A(\mathscr{C} \cdot x, \mathscr{C} \cdot y), \quad x, y \in \mathbb{R}_+^\nu. \tag{6}$$

It was proved in [5] that d_A is a norm-induced metric on $\mathcal{S}_\nu$, provided the latter is given an appropriate normed vector space structure. Owing to the compositional constraint $|\cdot|_1 = \kappa$, it holds $\dim \mathcal{S}_\nu = \nu - 1$. Accordingly, the description of $\mathcal{S}_\nu$ in terms of ν components is redundant: an essential representation requires compositions to be properly mapped onto $(\nu-1)$-tuples. Among various possibilities, the Isometric Log-Ratio (ILR) transform introduced in [5] is the only known map of this kind leaving d_A invariant. More precisely, the ILR fulfills

$$d_A(x, y) = d_E(\mathrm{ilr}(x), \mathrm{ilr}(y)),$$

$$d_E(u, v) \equiv \sqrt{\sum_{k=1}^{\nu-1}(u_k - v_k)^2}. \tag{7}$$

It is known from [6] that if $X \sim \log \mathcal{N}_\nu(\mu, \Sigma)$ is a log-normal ν-variate, then $\mathscr{C} \cdot X \sim L_\nu(\mu', \Sigma')$ is a logistic-normal ν-variate (the reader should notice that in [6] the simplex is defined by $\mathcal{S}_\nu = \{z \in \mathbb{R}_+^\nu : |z|_1 < 1\}$; thus property 2.2 of [6] is here reformulated so as to take into account such difference), with a known relation between (μ, Σ) and (μ', Σ'). Analogously, it is not difficult to show that $\mathrm{ilr}(\mathscr{C} \cdot X) \sim \mathcal{N}_{\nu-1}(\mu'', \Sigma'')$ is a normal $(\nu-1)$-variate, with (μ'', Σ'') related to (μ', Σ') via the change of basis matrices derived in [5]. Just to sum up, it holds

$$X \sim \log \mathcal{N}_\nu(\mu, \Sigma) \Longrightarrow$$

$$\mathscr{C} \cdot X \sim L_\nu(\mu', \Sigma') \Longleftrightarrow \tag{8}$$

$$\mathrm{ilr}(\mathscr{C} \cdot X) \sim \mathcal{N}_{\nu-1}(\mu'', \Sigma'').$$

Now, suppose that (i) X fulfills (8) and has a natural interpretation as a composition, (ii) a representative set $\mathcal{D}_X$ of observations of X is given, and (iii) we wish to select from $\mathcal{D}_X$ those units which are compositionally closer to the center of the distribution, according to the Aitchison distance. To see that the problem is well posed, we first turn $\mathcal{D}_X$ into a set $\mathcal{D}_{\mathscr{C} \cdot X} \equiv \{y : y = \mathscr{C} \cdot x, x \in \mathcal{D}_X\}$ of compositional observations of $Y = \mathscr{C} \cdot X$. Then, we consider the special point $\mathrm{cen}[Y] = \mathscr{C} \cdot \exp\{\mathbb{E}[\ln Y]\}$, representing the center of the distribution of Y in a compositional sense: $\mathrm{cen}[Y]$ minimizes the expression $\mathbb{E}[d_A^2(Y, \mathrm{cen}[Y])]$ over $\mathcal{S}_\nu$ and fulfills $\mathrm{cen}[Y] = \mathrm{ilr}^{-1}(\mathbb{E}[\mathrm{ilr}(Y)])$, see [7]. By virtue of (8) this entails $\mathrm{ilr}(\mathrm{cen}[Y]) = \mathbb{E}[\mathrm{ilr}(Y)] = \mu''$. In order to select the observations which are closer to $\mathrm{cen}[Y]$, we set a threshold $\delta > 0$ and consider only those elements $y \in \mathcal{D}_{\mathscr{C} \cdot X}$ fulfilling $d_A^2(y, \mathrm{cen}[y]) < \delta$, with $\mathrm{cen}[y]$ being a sample estimator of $\mathrm{cen}[Y]$ on $\mathcal{D}_{\mathscr{C} \cdot X}$. Such selection rule operates a well-defined truncation of the distribution of Y. Moreover, in view of (7) and (8), we have

$$\mathbb{P}\left[d_A^2(Y, \mathrm{cen}[Y]) < \delta \mid Y \sim L_\nu(\mu', \Sigma')\right]$$

$$= \mathbb{P}\left[d_E^2(Z, \mu'') < \delta \mid Z \sim \mathcal{N}_{\nu-1}(\mu'', \Sigma'')\right], \tag{9}$$

with $Z = \mathrm{ilr}(\mathscr{C} \cdot X)$. As a consequence, we see that a compositional selection rule based on the Aitchison distance and (8) is equivalent to a spherical truncation of a multinormal distribution. Obviously, once Z has been spherically truncated, the covariance matrix of the remaining data is damped; thus an estimate of the full covariance matrix requires the reconstruction procedure described in next sections.

2.3. Compositional Outlier Detection via Forward Search Techniques. Outlier detection in CoDA is a practical problem where the considerations made in Section 2.2 find concrete application. Most of the established methods for outlier

detection make use of the Mahalanobis metric. This is however unfit to describe the distance between compositions. Log-ratio transforms allow to get rid of the compositional constraint and make it possible to apply standard statistical methods [8]. Here we consider an innovative approach, namely, the Forward Search Algorithm (FSA), introduced in [9] and thoroughly discussed in [10]. The FSA admits an elegant extension to CoDA, of which the covariance reconstruction is a key step. In the next few lines we sketch the original algorithm and outline its extension to CoDA.

2.3.1. Construction of the Signal. In its standard formulation the FSA applies to normal data. It assumes a dataset $\mathcal{D}_X = \{x^{(k)}\}_{k=1}^{N}$ with $x^{(k)} \in \mathbb{R}_+^{(\nu)}$ for $k = 1, \ldots, N$. The null hypothesis is that all the elements of $\mathcal{D}_X$ are independent realizations of a stochastic variable $X \sim \mathcal{N}(\mu_0, \Sigma_0)$. The FSA consists of a sequence of steps where data are recursively sorted. Along the recursion a signal is built and tested.

As a preliminary step, m_0 observations are randomly chosen from the bulk of $\mathcal{D}_X$. Let $S(m_0)$ be the set of these observations. $\mathcal{S}(m_0)$ is used to provide initial estimates $\mu(m_0)$, $\Sigma(m_0)$ of the true mean μ_0 and the true covariance matrix Σ_0, respectively. For $m = m_0 + 1, m_0 + 2, \ldots$, the $(m - m_0)$th step of the algorithm goes as follows:

  (i) sort the elements of $\mathcal{D}_X$ according to the increasing values of the square Mahalanobis distance:

$$d_m(x)^2 = [x - \mu(m-1)]^T \Sigma(m-1)^{-1} [x - \mu(m-1)]; \quad (10)$$

  (ii) take the mth element $x^{(m)}$ of the newly sorted dataset and regard $s_m = d_m(x^{(m)})^2$ as the $(m - m_0)$th component of the signal;

  (iii) let $\mathcal{S}(m)$ be the set of the first m observations of the newly sorted dataset;

  (iv) use $\mathcal{S}(m)$ to provide new estimates $\mu(m)$, $\Sigma(m)$ of the true mean and the true covariance matrix, respectively.

Notice that $\mathcal{S}(m)$ is a truncated dataset. Therefore, $\Sigma(m)$ must include the Tallis' correction factor, (2) with $\rho = s_m$. While the recursion proceeds, the inliers of $\mathcal{D}_X$ populate progressively $\mathcal{S}(m)$. The recursion stops at the $\overline{m}$th step, when the first outlier $x^{(\overline{m})}$ of $\mathcal{D}_X$ produces a statistically relevant discontinuity in the signal.

2.3.2. Statistical Test of the Signal. Statistical tests are needed to assess the relevance of discontinuities observed in the signal. At each step of the algorithm a new test is actually performed. Specifically, at the mth step, s_m is computed together with the lowest and highest values, respectively, $\delta s_{m,\alpha}$ and $\delta s_{m,1-\alpha}$, which are theoretically admissible for s_m under the null hypothesis at $(1 - \alpha)$ significance level for some α. These values are nothing but the α- and $(1 - \alpha)$-percentage points of s_m. Their envelopes for $m - m_0 > 0$ generate two curves that surround the signal when plotted versus m. More precisely, one curve lies below it and the other lies above, provided the null hypothesis is not broken. The violation of one of the curves by the signal is interpreted as the result of the entrance of an outlier into $\mathcal{S}(m)$. Although the distribution of s_m cannot be calculated in closed form, its percentage points are obtained from a general formula, first derived in [11], yielding

$$\delta s_{m,\alpha} = \left(\chi_\nu^2\right)^{-1} \left(\frac{m}{m + (N - m - 1)\, f_{2(N-m-1),2m;1-\alpha}} \right), \quad (11)$$

with $f_{a,b;\alpha}$ the α-percentage point of the Fisher distribution with parameters (a, b). Equation (11) holds with decent approximation, as confirmed by numerical simulations.

2.3.3. Extension of the Forward Search to CoDA. When $\mathcal{D}_X$ is a compositional dataset, it is unnatural to assume that its elements are realizations of a multivariate $X \sim \mathcal{N}(\mu_0, \Sigma_0)$. In this case the use of the FSE as outlined above does not make sense at all. Sometimes it is reasonable to assume $X \sim L_\nu(\mu_0, \Sigma_0)$, as first argued in [6]. In this case we can use the FSA to find outliers, provided that we first modify the algorithm in two respects:

  (i) we replace the Mahalanobis distance by the Aitchison one;

  (ii) we perform statistical tests consistently with the change of null hypothesis.

Specifically, at the mth step of the algorithm, we sort $\mathcal{D}_X$ according to the increasing values of the square Aitchison distance:

$$d_A(x)^2 = \frac{1}{2\nu} \sum_{i,k=1}^{\nu} \left[\log\left(\frac{x_i}{x_k}\right) - \log\left(\frac{(c_{m-1})_i}{(c_{m-1})_k}\right) \right]^2, \quad (12)$$

where $c_{m-1} = \text{cen}[y \mid y \in \mathcal{S}(m-1)]$ is the center of $\mathcal{S}(m-1)$. Analogously, given the mth element $x^{(m)}$ of the newly sorted dataset, we regard $s_m = d_A(x^{(m)})^2$ as the mth component of the signal. The percentage points of s_m are obtained from $\text{ilr}(\mathcal{S}(m))$ by using the probability correspondence established in (9). Since $\text{ilr}(\mathcal{S}(m))$ is a spherically truncated dataset, the estimate of the covariance matrix $\Sigma(m)$ derived from it must undergo the reconstruction procedure described in next sections.

2.4. General Covariance Reconstruction Problem. The examples discussed in the previous subsections are special cases of a more general inverse problem, namely, the reconstruction of the covariance matrix Σ of a normal multivariate X on the basis of the covariance matrix $\mathfrak{S}_\mathcal{D}$ of X truncated to some (convex) region $\mathcal{D}$. This is the simplest yet nontrivial inverse problem, which can be naturally associated with the normal distribution. The case $\mathcal{D} = \mathcal{B}_\nu(\rho)$ corresponds to a setup where theoretical and practical aspects of the problem can be investigated with relatively modest mathematical effort. It is certainly a well-defined framework where to study regularization techniques for nonlinear inverse problems in statistics, for which there is still much room for interesting work [12, 13].

3. Definitions and Setup

Let $X \in \mathbb{R}^v$ be a random vector with jointly normal distribution $\mathcal{N}_v(0, \Sigma)$ in $v \geq 1$ dimensions. The probability that X falls within $\mathcal{B}_v(\rho)$ is measured by the Gaussian integral

$$\alpha(\rho; \Sigma) \equiv \mathbb{P}\left[X \in \mathcal{B}_v(\rho)\right]$$
$$= \frac{1}{(2\pi)^{v/2} |\Sigma|^{1/2}} \int_{\mathcal{B}_v(\rho)} d^v x \exp\left\{-\frac{1}{2} x^T \Sigma^{-1} x\right\}. \quad (13)$$

Since Σ is symmetric positive definite, it has orthonormal eigenvectors $\Sigma v^{(i)} = \lambda_i v^{(i)}$. Let us denote by $R \equiv \{v_i^{(j)}\}_{i,j=1}^v$ the orthogonal matrix having these vectors as columns and by $\Lambda \equiv \text{diag}(\lambda) = R^T \Sigma R$ the diagonal counterpart of Σ. From the invariance of $\mathcal{B}_v(\rho)$ under rotations, it follows that α depends upon Σ just by way of λ. Accordingly, we rename the Gaussian probability content of $\mathcal{B}_v(\rho)$ as

$$\alpha(\rho; \lambda) \equiv \int_{\mathcal{B}_v(\rho)} d^v x \prod_{m=1}^v \delta(x_m, \lambda_m),$$
$$\delta(y, \eta) = \frac{1}{\sqrt{2\pi\eta}} \exp\left\{-\frac{y^2}{2\eta}\right\}. \quad (14)$$

Note that (14) is not sufficient to fully characterize the random vector X under the spherical constraint, for which we need to calculate the distribution law $\mathbb{P}[X \in A \mid X \in \mathcal{B}_v(\rho)]$ as a function of $A \subset \mathbb{R}^v$. Alternatively, we can describe X in terms of the complete set of its truncated moments

$$m_{k_1 \cdots k_v}(\rho; \Sigma) \equiv \mathbb{E}\left[X_1^{k_1} X \cdots \in X_v^{k_v} \mathcal{B}_v(\rho)\right],$$
$$\{k_i\}_{i=1,\ldots,v} \in \mathbb{N}^v. \quad (15)$$

As usual, these can be all obtained from the moment-generating function

$$\alpha m(t)$$
$$= \frac{1}{(2\pi)^{v/2} |\Sigma|^{1/2}} \int_{\mathcal{B}_v(\rho)} d^v x \exp\left\{t^T x - \frac{1}{2} x^T \Sigma^{-1} x\right\}, \quad (16)$$
$$t \in \mathbb{R}^v,$$

by differentiating the latter an arbitrary number of times with respect to the components of t; namely,

$$m_{k_1 \cdots k_v}(\rho; \Sigma) = \frac{\partial^{k_1 + \cdots + k_v} m(t)}{(\partial t_1)^{k_1} \cdots (\partial t_v)^{k_v}}\Bigg|_{t=0}. \quad (17)$$

It will be observed that $m(t)$ is in general not invariant under rotations of t. Therefore, unlike α, the moments $m_{k_1 \cdots k_v}$ depend effectively on both λ and R. For instance, for the matrix of the second moments $\mathfrak{S}_{\mathcal{B}} \equiv \{\partial^2 m/\partial t_i \partial t_j|_{t=0}\}_{i,j=1}^v$ such dependence amounts to

$$\alpha(\mathfrak{S}_{\mathcal{B}})_{ij} = \sum_{k,\ell=1}^v R_{ki} R_{\ell j} \int_{\mathcal{B}_v(\rho)} d^v x \, x_k x_\ell \prod_{m=1}^v \delta(x_m, \lambda_m). \quad (18)$$

By parity, the only nonvanishing terms in the above sum are those with $k = \ell$. Hence, it follows that Σ and $\mathfrak{S}_{\mathcal{B}}$ share R as a common diagonalizing matrix. In other words, if $M \equiv \text{diag}(\mu)$ is the diagonal matrix of the eigenvalues of $\mathfrak{S}_{\mathcal{B}}$, then $M = R^T \mathfrak{S}_{\mathcal{B}} R$. Moreover, μ_k is related to λ_k by

$$\mu_k = \lambda_k \frac{\alpha_k}{\alpha},$$
$$\alpha_k(\rho; \lambda) \equiv \int_{\mathcal{B}_v(\rho)} d^v x \frac{x_k^2}{\lambda_k} \prod_{m=1}^v \delta(x_m, \lambda_m), \quad (19)$$
$$k = 1, \ldots, v.$$

The ratios α_k/α are naturally interpreted as adimensional correction factors to the eigenvalues of Σ, so they play the same role as $c_T(\rho)$ in (2). However, α_k/α depends explicitly on the subscript k; thus each eigenvalue is damped differently from the others as a consequence of the condition $X \in \mathcal{B}_v(\rho)$.

Remark 1. In practical terms, (18)-(19) tell us that estimating the sample covariance matrix of $X \sim \mathcal{N}_v(0, \Sigma)$ from a spherically truncated population $\{x^{(m)}\}_{m=1}^M$, made of M independent units, via the classical estimator $Q_{ij} = (M - 1)^{-1} \sum_{m=1}^M (x_i^{(m)} - \tilde{x}_i^{(m)})(x_j^{(m)} - \tilde{x}_j^{(m)})$, being $\tilde{x} = M^{-1} \sum_{m=1}^M x^{(m)}$ the sample mean, yields a damped result. Nonetheless, the damping affects only the eigenvalues of the estimator, whereas its eigenvectors are left invariant.

3.1. Monotonicity Properties of Ratios of Gaussian Integrals. Equations (14) and (19) suggest introducing a general notation for the Gaussian integrals over $\mathcal{B}_v(\rho)$, under the assumption $\Sigma = \Lambda$. So, we define

$$\alpha_{k\ell m \cdots}(\rho; \lambda) \equiv \int_{\mathcal{B}_v(\rho)} d^v x \frac{x_k^2}{\lambda_k} \frac{x_\ell^2}{\lambda_\ell} \frac{x_m^2}{\lambda_m} \cdots \prod_{n=1}^v \delta(x_n, \lambda_n), \quad (20)$$

with each subscript q on the *left-hand side* addressing an additional factor of x_q^2/λ_q under the integral sign on the *right-hand side* (no subscripts means α). Several analytic properties of such integrals are discussed in [14]. In the following proposition, we lay emphasis on some issues concerning the monotonicity trends of the ratios α_k/α.

Proposition 2 (monotonicities). *Let $\lambda_{(k)} \equiv \{\lambda_i\}_{i=1,\ldots,v}^{i \neq k}$ denote the set of the full eigenvalues without λ_k. The ratios α_k/α fulfill the following properties:*

(p_1) $\lambda_k(\alpha_k/\alpha)(\rho; \lambda)$ is a monotonic increasing function of λ_k at fixed ρ and $\lambda_{(k)}$;

(p_2) $(\alpha_k/\alpha)(\rho; \lambda)$ is a monotonic decreasing function of λ_k at fixed ρ and $\lambda_{(k)}$;

(p_3) $(\alpha_k/\alpha)(\rho; \lambda)$ is a monotonic decreasing function of λ_i $(i \neq k)$ at fixed ρ and $\lambda_{(i)}$,

where an innocuous abuse of notation has been made on writing $(\alpha_k/\alpha)(\rho; \lambda)$ in place of $\alpha_k(\rho; \lambda)/\alpha(\rho; \lambda)$.

Proof. Let the symbol $\partial_k \equiv \partial/\partial\lambda_k$ denote a derivative with respect to λ_k. In order to prove property (p_1), we apply the

chain rule of differentiation to $\lambda_k \alpha_k / \alpha$ and then we pass ∂_k under the integral sign in $\partial_k \alpha$ and $\partial_k \alpha_k$. In this way, we obtain

$$
\begin{aligned}
\partial_k \left(\lambda_k \frac{\alpha_k}{\alpha} \right) &= \frac{1}{2} \left(\frac{\alpha_{kk}}{\alpha} - \frac{\alpha_k^2}{\alpha^2} \right) \\
&= \frac{1}{2\lambda_k^2} \left\{ \mathbb{E} \left[X_k^4 \mid X \in \mathcal{B}_\nu (\rho) \right] \right. \\
&\left. - \mathbb{E} \left[X_k^2 \mid X \in \mathcal{B}_\nu (\rho) \right]^2 \right\} = \frac{1}{2\lambda_k^2} \\
&\cdot \operatorname{var} \left(X_k^2 \mid X \in \mathcal{B}_\nu (\rho) \right) \geq 0.
\end{aligned}
\tag{21}
$$

Moreover, since the truncated marginal density of X_k^2 is positive within a set of nonzero measure in $\mathbb{R}$, the monotonic trend of $\lambda_k \alpha_k / \alpha$ in λ_k is strict.

Properties (p_2) and (p_3) are less trivial than (p_1). Indeed, the same reasoning as above now yields on the one hand

$$
\begin{aligned}
\lambda_k \partial_k \left(\frac{\alpha_k}{\alpha} \right) &= \partial_k \left(\lambda_k \frac{\alpha_k}{\alpha} \right) - \frac{\alpha_k}{\alpha} \\
&= \frac{1}{2\lambda_k^2} \left\{ \operatorname{var} \left(X_k^2 \mid X \in \mathcal{B}_\nu (\rho) \right) \right. \\
&\left. - 2\lambda_k \mathbb{E} \left[X_k^2 \mid X \in \mathcal{B}_\nu (\rho) \right] \right\} \leq 0,
\end{aligned}
\tag{22}
$$

and on the other

$$
\begin{aligned}
\lambda_i \partial_i \left(\frac{\alpha_k}{\alpha} \right) &= \frac{1}{2} \left(\frac{\alpha_{ik}}{\alpha} - \frac{\alpha_i \alpha_k}{\alpha^2} \right) \\
&= \frac{1}{2\lambda_i \lambda_k} \operatorname{cov} \left(X_i^2, X_k^2 \mid X \in \mathcal{B}_\nu (\rho) \right) \leq 0
\end{aligned}
\tag{23}
$$

$$(i \neq k).$$

Despite being not *a priori* evident, the *right-hand side* of both (22) and (23) is negative (and vanishes in the limit $\rho \to \infty$). The inequalities $\operatorname{var}(X_k^2) \leq 2\lambda_k \mathbb{E}[X_k^2]$ within Euclidean balls have been first discussed in [14], while the inequalities $\operatorname{cov}(X_j^2, X_k^2)$ within generalized Orlicz balls have been discussed in [15, 16] for the case where the probability distribution of X is flat instead of being normal. More recently, a complete proof of both inequalities has been given in [17]. Despite the technical difficulties in proving them, their meaning should be intuitively clear. The variance inequality quantifies the squeezing affecting X_k^2 as a consequence of the truncation (in the unconstrained case it would be $\operatorname{var}(X_k^2) = 2\lambda_k^2$). The covariance inequality follows from the opposition arising among the square components in proximity of the boundary of $\mathcal{B}_\nu(\rho)$. Indeed, if $X_j^2 \nearrow \rho$, then $X_k^2 \searrow 0 \; \forall k \neq j$ in order for X to stay within $\mathcal{B}_\nu(\rho)$. $\qquad \square$

3.2. Definition Domain of the Reconstruction Problem. A consequence of Proposition 2 is represented by the following.

Corollary 3. *Given ν, ρ, and λ, μ_k is bounded by*

$$
\frac{\rho}{r(\nu, \rho/2\lambda_k)} \leq \mu_k \leq \frac{\rho}{3},
$$
$$
r(\nu, z) \equiv (2\nu + 1) \frac{M(\nu, \nu + 1/2, z)}{M(\nu, \nu + 3/2, z)},
\tag{24}
$$

with M denoting the Kummer function; namely,

$$
M(a, b, z) = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{(a)_n}{(b)_n} z^n,
$$
$$
(x)_n \equiv \frac{\Gamma(x + n)}{\Gamma(x)}.
\tag{25}
$$

Proof. The upper bound of (24) corresponds to the value of μ_k in the ν-tuple limit $\lambda_k \to \infty$, $\lambda_{(k)} \to \{0, \ldots, 0\}$. This is indeed the maximum possible value allowed for μ_k according to properties (p_1) and (p_3) of Proposition 2. In order to perform this limit, we observe that

$$
\lim_{\eta \to 0^+} \delta(y, \eta) = \delta(y),
\tag{26}
$$

with the δ symbol on the *right-hand side* representing the Dirac delta function (the reader who is not familiar with the theory of distributions may refer, for instance, to [18] for an introduction). Accordingly,

$$
\lim_{\lambda_k \to \infty} \lim_{\lambda_{(k)} \to \{0, \ldots, 0\}} \mu_k = \frac{\int_{-\sqrt{\rho}}^{+\sqrt{\rho}} dx_k \, x_k^2}{\int_{-\sqrt{\rho}}^{+\sqrt{\rho}} dx_k} = \frac{\rho}{3}.
\tag{27}
$$

The lower bound corresponds instead to the value taken by μ_k as $\lambda_{(k)} \to \{\infty, \ldots, \infty\}$ and λ_k is kept fixed. In this limit, all the Gaussian factors in the probability density function except the kth one flatten to one. Hence,

$$
\begin{aligned}
&\lim_{\lambda_{(k)} \to \{\infty, \ldots, \infty\}} \mu_k \\
&= \lim_{\lambda_{(k)} \to \{\infty, \ldots, \infty\}} \frac{\int_{-\sqrt{\rho}}^{+\sqrt{\rho}} dx_k \, x_k^2 \delta(x_k, \lambda_k) \alpha^{(\nu-1)} \left(\rho - x_k^2; \lambda_{(k)} \right)}{\int_{-\sqrt{\rho}}^{+\sqrt{\rho}} dx_k \, \delta(x_k, \lambda_k) \alpha^{(\nu-1)} \left(\rho - x_k^2; \lambda_{(k)} \right)} \\
&= \frac{\int_{-\sqrt{\rho}}^{+\sqrt{\rho}} dx_k \, x_k^2 e^{-x_k^2/2\lambda_k} \left(\rho - x_k^2 \right)^{\nu-1}}{\int_{-\sqrt{\rho}}^{+\sqrt{\rho}} dx_k \, e^{-x_k^2/2\lambda_k} \left(\rho - x_k^2 \right)^{\nu-1}} \\
&= \rho \frac{\int_0^1 dx_k \, x_k^2 e^{-(\rho/2\lambda_k)x_k^2} \left(1 - x_k^2 \right)^{\nu-1}}{\int_0^1 dx_k \, e^{-(\rho/2\lambda_k)x_k^2} \left(1 - x_k^2 \right)^{\nu-1}}.
\end{aligned}
\tag{28}
$$

Numerator and denominator of the rightmost ratio are easily recognized to be integral representations of Kummer functions (see, e.g., [19, ch. 13]). $\qquad \square$

The upper bound of (24) can be sharpened, as clarified by the following.

Proposition 4 (bounds on the truncated moments). *Let v, ρ, and λ be given. If $\{i_1, \ldots, i_v\}$ is a permutation of $\{1, \ldots, v\}$ such that $\mu_{i_1} \leq \mu_{i_2} \leq \cdots \leq \mu_{i_v}$, then the following upper bounds hold:*

$$
\begin{aligned}
&(i) \quad \sum_{k=1}^{v} \mu_k \leq \rho; \\
&(ii) \quad \mu_{i_k} \leq \frac{\rho}{v-k+1},
\end{aligned}
\tag{29}
$$

$$
k = 1, \ldots, v.
$$

Proof. The overall upper bound on the sum of truncated moments follows from

$$
\begin{aligned}
\sum_{k=1}^{v} \mu_k &= \frac{1}{\alpha} \sum_{k=1}^{v} \lambda_k \alpha_k \\
&= \frac{1}{\alpha} \int_{\mathcal{B}_v(\rho)} \mathrm{d}^v x \left(\sum_{k=1}^{v} x_k^2 \right) \prod_{m=1}^{v} \delta\left(x_m, \lambda_m\right) \leq \rho.
\end{aligned}
\tag{30}
$$

At the same time, the sum can be split and bounded from below by

$$
\sum_{k=1}^{v} \mu_{i_k} = \sum_{k=1}^{n} \mu_{i_k} + \sum_{k=n+1}^{v} \mu_{i_k} \geq \sum_{k=1}^{n} \mu_{i_k} + (v-n)\, \mu_{i_{n+1}},
\tag{31}
$$

$$
n = 0, 1, \ldots, v - 1.
$$

The single upper bounds on the μ_k's are then obtained from (30)-(31). It will be noted that (29) *(ii)* is sharper than the upper bound of (24) only for $v > 3$ and $k < v - 2$. □

From now on, we shall assume, with no loss of generality, that the eigenvalues of Σ are increasingly ordered, namely, $0 < \lambda_1 \leq \cdots \leq \lambda_v$ (we can always permute the labels of the coordinate axes, so as to let this be the case). An important aspect related to the eigenvalue ordering is provided by the following.

Proposition 5 (eigenvalue ordering). *Let v, ρ, and λ be given. If $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_v$, then $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_v$ holds as well.*

Proof. In order to show that the spherical truncation does not violate the eigenvalue ordering, we make repeated use of the monotonicity properties of Proposition 2. Specifically, if $i < j$, then

$$
\mu_i = \lambda_i \frac{\alpha_i}{\alpha} \left(\rho; \left\{ \lambda_1, \ldots, \lambda_i, \ldots, \lambda_j, \ldots, \lambda_v \right\} \right)
$$

$$
\leq \lambda_j \frac{\alpha_i}{\alpha} \left(\rho; \left\{ \lambda_1, \ldots, \lambda_j, \ldots, \lambda_j, \ldots, \lambda_v \right\} \right)
$$

$$
\rightsquigarrow \text{increasing monotonicity of } \lambda_i \frac{\alpha_i}{\alpha}
$$

$$
= \lambda_j \frac{\alpha_j}{\alpha} \left(\rho; \left\{ \lambda_1, \ldots, \lambda_j, \ldots, \lambda_j, \ldots, \lambda_v \right\} \right)
$$

$$
\rightsquigarrow \text{exchange symmetry } i \longleftrightarrow j
$$

$$
\leq \lambda_j \frac{\alpha_j}{\alpha} \left(\rho; \left\{ \lambda_1, \ldots, \lambda_i, \ldots, \lambda_j, \ldots, \lambda_v \right\} \right)
$$

$$
\rightsquigarrow \text{decreasing monotonicity of } \frac{\alpha_j}{\alpha}
$$

$$
= \mu_j,
\tag{32}
$$

where the symbol "$\rightsquigarrow$" is used to explain where the inequality sign preceding it comes from and the "exchange symmetry" refers to the formal property of the one-index Gaussian integrals over $\mathcal{B}_v(\rho)$ to fulfill $\alpha_i(\rho; \{\lambda_1, \ldots, \lambda_i, \ldots, \lambda_j, \ldots, \lambda_v\}) = \alpha_j(\rho; \{\lambda_1, \ldots, \lambda_j, \ldots, \lambda_i, \ldots, \lambda_v\})$. □

Let us now focus on (19). They have to be solved in order to reconstruct λ from μ. Formally, if we introduce a family of truncation operators $\tau_\rho : \mathbb{R}_+^v \to \mathbb{R}_+^v$ (parametrically depending on ρ), such that

$$
\left(\tau_\rho \cdot \lambda \right)_k \equiv \lambda_k \frac{\alpha_k}{\alpha} \left(\rho; \lambda \right), \quad k = 1, \ldots, v,
\tag{33}
$$

then the reconstruction of λ from μ amounts to calculating $\lambda = \tau_\rho^{-1} \cdot \mu$. One should be aware that τ_ρ is not a surjective operator in view of Corollary 3 and Proposition 4. Therefore, τ_ρ^{-1} is only defined within a bounded domain $\mathcal{D}(\tau_\rho^{-1})$. If we define

$$
\mathcal{D}_0 = \left\{ \mu \in \mathbb{R}_+^v : \mu_1 \leq \cdots \leq \mu_v,\ \mu_k = \lambda_k \frac{\alpha_k}{\alpha} \text{ for } k \right.
\tag{34}
$$

$$
\left. = 1, \ldots, v, \text{ for some } \lambda \in \mathbb{R}_+^v \right\},
$$

then we have $\mathcal{D}(\tau_\rho^{-1}) = \{\mu : \mu = \sigma \cdot \mu_0 \text{ for some } \mu_0 \in \mathcal{D}_0, \ \sigma \in S_v\}$, where S_v is the set of permutations of v elements. From Proposition 4 we conclude that $\mathcal{D}_0 \subseteq H_v(\rho)$, being

$$
H_v(\rho) \equiv \left\{ x \in \mathbb{R}_+^v : x_k \leq \min\left\{ \frac{\rho}{3}, \frac{\rho}{v-k+1} \right\}, \sum_{k=1}^{v} x_k \right.
\tag{35}
$$

$$
\left. \leq \rho, \ \forall k \right\}.
$$

In fact, there are vectors $\mu \in \mathbb{R}_+^v$ fulfilling $\mu \in H_v(\rho)$ and $\mu \notin \mathcal{D}_0$; thus we conclude that $\mathcal{D}_0$ is a proper subset of $H_v(\rho)$. Numerical experiences based on the techniques discussed in the next sections show indeed that

$$
\mathcal{D}\left(\tau_\rho^{-1}\right) = \bigcap_{k=1}^{v} \left\{ \mu \in \mathbb{R}_+^v : \sum_{j \neq k}^{1 \cdots v} \mu_j + 3\mu_k \leq \rho \right\}.
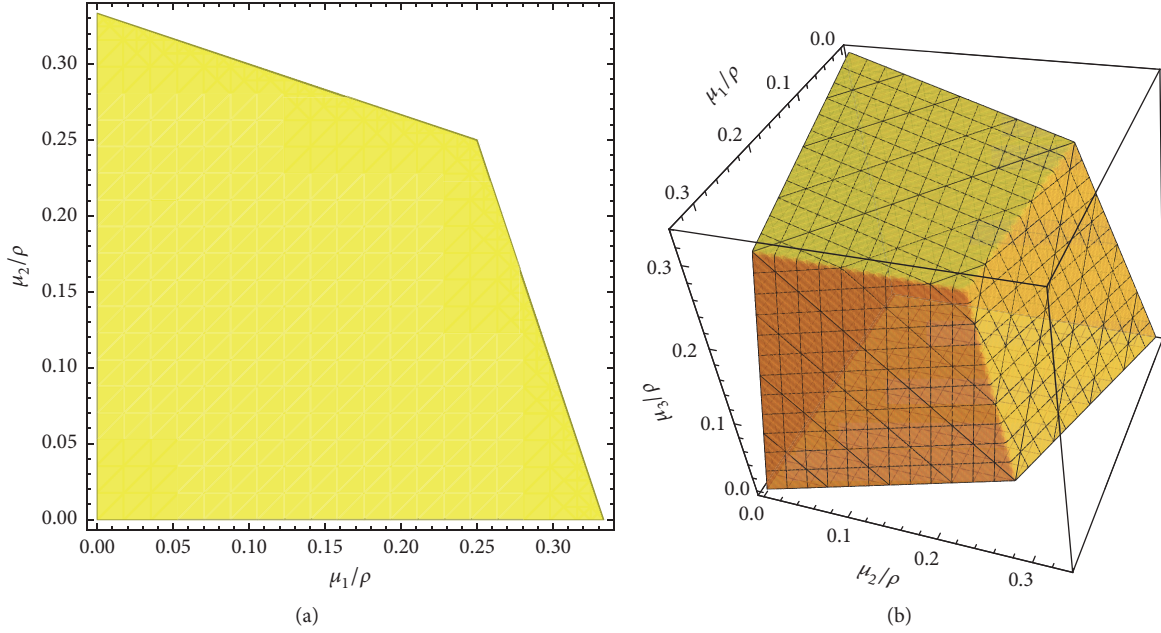\tag{36}
$$

FIGURE 2: (a) Numerical reconstruction of $\mathscr{D}(\tau_\rho^{-1})$ in $\nu = 2$ dimensions. (b) Numerical reconstruction of $\mathscr{D}(\tau_\rho^{-1})$ in $\nu = 3$ dimensions.

A graphical representation of (36) in $\nu = 2$ and $\nu = 3$ dimensions is depicted in Figure 2. The reader should note that until Section 7 we shall always assume that μ comes from the application of τ_ρ to some λ; thus $\mu \in \mathscr{D}(\tau_\rho^{-1})$ by construction.

Now, we observe that (19) can be written in the equivalent form

$$\lambda = T\left(\lambda; \mu; \rho\right), \tag{37}$$

$$T : \mathbb{R}_+^\nu \times \mathbb{R}_+^\nu \times \mathbb{R}_+ \longrightarrow \mathbb{R}_+^\nu;$$

$$T_k\left(\lambda; \mu; \rho\right) = \mu_k \frac{\alpha}{\alpha_k}\left(\rho; \lambda\right), \quad k = 1, \ldots, \nu. \tag{38}$$

Since ρ and μ are (nonindependent) input parameters for the covariance reconstruction problem (and in order to keep the notation light), in the sequel we shall leave the dependence of T upon ρ and μ implicitly understood; that is, we shall write (37) as $\lambda = T(\lambda)$. Hence, we see that the full eigenvalue spectrum λ is a fixed point of the operator T. This suggests obtaining it as the limit of a sequence

$$\lambda^{(0)} = \mu,$$

$$\lambda^{(n+1)} = T\left(\lambda^{(n)}\right), \tag{39}$$

$$n = 0, 1, \ldots,$$

$$\lambda = \lim_{n \to \infty} \lambda^{(n)}, \tag{40}$$

provided that this can be shown to converge. Note that since $\alpha_k < \alpha$, it follows that $T_k(\lambda^{(n)}) > \mu_k \; \forall n$, so the sequence is bounded from below by μ. In particular, this holds for $n = 0$.

Therefore, the sequence moves to the right direction at least at the beginning. A formal proof of convergence, based on the monotonicity properties stated by Proposition 2, is given in the next section.

4. Convergence of the Fixed Point Equation

We split our argument into three propositions, describing different properties of the sequence $\lambda^{(n)}$. They assert, respectively, that (i) the sequence is componentwise monotically increasing; (ii) the sequence is componentwise bounded from above by any fixed point of T; and (iii) if T has a fixed point, this must be unique. Statements (i) and (ii) are sufficient to guarantee the convergence of the sequence to a finite limit (the unconstrained spectrum is a fixed point of T). In addition, the limit is easily recognized to be a fixed point of T. Hence, statement (iii) guarantees that the sequence converges to the unconstrained eigenvalue spectrum. We remark that all the monotonicities discussed in Proposition 2 are strict; that is, the ratios α_k/α have no stationary points at finite ρ and λ, which is crucial for the proof.

Proposition 6 (increasing monotonicity). *Given ν, ρ, and $\mu \in \mathscr{D}(\tau_\rho^{-1})$, the sequence $\lambda^{(0)} = \mu$, $\lambda^{(n+1)} = T(\lambda^{(n)})$, $n = 0, 1, \ldots$, is monotically increasing; namely, $\lambda_k^{(n+1)} > \lambda_k^{(n)} \; \forall k = 1, \ldots, \nu$.*

Proof. The proof is by induction. We first notice that

$$\lambda_k^{(1)} = T_k\left(\lambda^{(0)}\right) = T_k\left(\mu\right) = \mu_k \frac{\alpha}{\alpha_k}\left(\rho; \mu\right) > \mu_k = \lambda_k^{(0)},$$

$$k = 1, \ldots, \nu; \tag{41}$$

the inequality follows from $\alpha_k(\rho; \mu) < \alpha(\rho; \mu)$. Suppose now that the property of increasing monotonicity has been checked off up to the nth element of the sequence. Then,

$$\lambda_k^{(n+1)} = \mu_k \frac{\alpha}{\alpha_k} \left(\rho; \lambda^{(n)}\right) > \mu_k \frac{\alpha}{\alpha_k} \left(\rho; \lambda^{(n-1)}\right) = \lambda_k^{(n)}; \quad (42)$$

the inequality follows this time from the inductive hypothesis and from properties (p_2) and (p_3) of Proposition 2. $\square$

Proposition 7 (boundedness). *Given v, ρ, and $\mu \in \mathscr{D}(\tau_\rho^{-1})$, the sequence $\lambda^{(0)} = \mu$, $\lambda^{(n+1)} = T(\lambda^{(n)})$, $n = 0, 1, \ldots$, is bounded from above; namely, $\lambda_k^{(n)} < \lambda_k^* \ \forall k = 1, \ldots, v$, λ^* being a fixed point of T.*

Proof. We proceed again by induction. We first notice that

$$\lambda_k^{(0)} = \mu_k < \mu_k \frac{\alpha}{\alpha_k} \left(\rho; \lambda^*\right) = \lambda_k^*, \quad k = 1, \ldots, v; \quad (43)$$

the inequality follows as previously from $\alpha_k(\rho; \lambda^*) < \alpha(\rho; \lambda^*)$. Suppose now that the property of boundedness has been checked off up to the nth element of the sequence. Then,

$$\lambda_k^{(n+1)} = \mu_k \frac{\alpha}{\alpha_k} \left(\rho; \lambda^{(n)}\right) = \lambda_k^* \frac{\alpha_k}{\alpha} \left(\rho; \lambda^*\right) \frac{\alpha}{\alpha_k} \left(\rho; \lambda^{(n)}\right)$$

$$< \lambda_k^* \frac{\alpha_k}{\alpha} \left(\rho; \lambda^{(n)}\right) \frac{\alpha}{\alpha_k} \left(\rho; \lambda^{(n)}\right) = \lambda_k^*; \quad (44)$$

the inequality follows for the last time from the inductive hypothesis and from properties (p_2) and (p_3) of Proposition 2. $\square$

According to Propositions 6 and 7, the sequence converges. Now, let $\tilde{\lambda} = \lim_{n\to\infty} \lambda^{(n)}$ be the limit of the sequence. Effortlessly, we prove that $\tilde{\lambda}$ is a fixed point of T. Indeed,

$$\tilde{\lambda}_k = \lim_{n\to\infty} \lambda_k^{(n)} = \lim_{n\to\infty} \mu_k \frac{\alpha}{\alpha_k} \left(\rho; \lambda^{(n-1)}\right)$$

$$= \mu_k \frac{\alpha}{\alpha_k} \left(\rho; \lim_{n\to\infty} \lambda^{(n-1)}\right) = T_k \left(\tilde{\lambda}\right). \quad (45)$$

Note that passing the limit over n under the integral sign is certainly allowed for Gaussian integrals.

Proposition 8 (uniqueness of the fixed point). *Let $\lambda' = T(\lambda')$ and $\lambda'' = T(\lambda'')$ be two fixed points of T, corresponding to the same choice of v, ρ, and $\mu \in \mathscr{D}(\tau_\rho^{-1})$. Then, it must be that $\lambda' = \lambda''$.*

Proof. According to the hypothesis, λ' and λ'' fulfill the equations

$$\lambda_k' = \mu_k \frac{\alpha}{\alpha_k} \left(\rho; \lambda'\right) \Longrightarrow$$

$$\mu_k = \lambda_k' \frac{\alpha_k}{\alpha} \left(\rho; \lambda'\right),$$

$$\lambda_k'' = \mu_k \frac{\alpha}{\alpha_k} \left(\rho; \lambda''\right) \Longrightarrow \quad (46)$$

$$\mu_k = \lambda_k'' \frac{\alpha_k}{\alpha} \left(\rho; \lambda''\right).$$

Hence,

$$0 = \mu_k - \mu_k = \lambda_k' \frac{\alpha_k}{\alpha} \left(\rho; \lambda'\right) - \lambda_k'' \frac{\alpha_k}{\alpha} \left(\rho; \lambda''\right)$$

$$= \sum_{\ell=1}^{v} \left[\int_0^1 dt\, J_{k\ell} \left(\rho; \lambda'' + t \left(\lambda' - \lambda''\right)\right)\right] \left(\lambda_\ell' - \lambda_\ell''\right), \quad (47)$$

where J denotes the Jacobian matrix of τ_ρ and is given by

$$J_{k\ell} \left(\rho; \lambda\right) = \partial_k \left(\lambda_\ell \frac{\alpha}{\alpha} \left(\rho; \lambda\right)\right) = \frac{1}{2} \frac{\lambda_\ell}{\lambda_k} \left(\frac{\alpha_{k\ell}}{\alpha} - \frac{\alpha_k \alpha_\ell}{\alpha^2}\right)$$

$$= \left[\Lambda^{-1} \Omega \left(\rho; \lambda\right) \Lambda\right]_{k\ell}, \quad (48)$$

having set $\Omega_{k\ell} \equiv (1/2)(\alpha_{k\ell}/\alpha - \alpha_k \alpha_\ell/\alpha^2)$. It will be noted that $\Omega = \{\Omega_{k\ell}\}_{k,\ell=1}^{v}$ is essentially the covariance matrix of the square components of X under spherical truncation (we have come across its matrix elements in (21)–(23)). As such, Ω is symmetric positive definite. Indeed,

$$\Omega_{k\ell} = \frac{1}{2\lambda_k \lambda_\ell} \text{cov} \left(X_k^2, X_\ell^2 \mid X \in \mathscr{B}_v(\rho)\right) = \frac{1}{2\lambda_k \lambda_\ell}$$

$$\cdot \mathbb{E}\left[\left(X_k^2 - \mathbb{E}\left[X_k^2 \mid X \in \mathscr{B}_v(\rho)\right]\right) \quad (49)$$

$$\cdot \left(X_\ell^2 - \mathbb{E}\left[X_\ell^2 \mid X \in \mathscr{B}_v(\rho)\right]\right) \mid X \in \mathscr{B}_v(\rho)\right].$$

On setting $Z_k = (X_k^2 - \mathbb{E}[X_k^2 \mid X \in \mathscr{B}_v(\rho)])/\sqrt{2}\lambda_k$, we can represent Ω as $\Omega = \mathbb{E}[ZZ^T \mid X \in \mathscr{B}_v(\rho)]$. If $x \in \mathbb{R}^v$ is not the null vector, then $x^T \Omega x = \mathbb{E}[x^T ZZ^T x \mid X \in \mathscr{B}_v(\rho)] = \mathbb{E}[(x^T Z)^2 \mid X \in \mathscr{B}_v(\rho)] > 0$. Moreover, the eigenvalues of Ω fulfill the secular equation

$$0 = \det \left(\Omega - \phi \mathbb{I}_v\right) = \det \left[\Lambda^{-1} \left(\Omega - \phi \mathbb{I}_v\right) \Lambda\right]$$

$$= \det \left(\Lambda^{-1} \Omega \Lambda - \phi \mathbb{I}_v\right) = \det \left(J - \phi \mathbb{I}_v\right), \quad (50)$$

whence it follows that J is positive definite as well (though it is not symmetric). Since the sum of positive definite matrices is positive definite, we conclude that $\int_0^1 dt\, J(\rho; \lambda'' + t(\lambda' - \lambda''))$ is positive definite too. As such, it is nonsingular. Therefore, from (47), we conclude that $\lambda' = \lambda''$. $\square$

5. Numerical Computation of Gaussian Integrals over $\mathscr{B}_v(\rho)$

Let us now see how to compute $\alpha_{k\ell m\cdots}$ with controlled precision. Most of the relevant work has been originally done by Ruben in [2], where the case of α is discussed. We extend Ruben's technique to Gaussian integrals containing powers of the integration variable. Specifically, it is shown in [2] that $\alpha(\rho; \lambda)$ can be represented as a series of chi-square cumulative distribution functions:

$$\alpha \left(\rho; \lambda\right) = \sum_{m=0}^{\infty} c_m \left(s; \lambda\right) F_{v+2m} \left(\frac{\rho}{s}\right). \quad (51)$$

The scale factor s has the same physical dimension as ρ and λ. It is introduced in order to factorize the dependence of α upon ρ and λ at each order of the expansion. The series on the *right-hand side* of (51) converges uniformly on every finite interval of ρ. The coefficients c_m are given by

$$c_m(s;\lambda) = \frac{1}{m!} \frac{s^{\nu/2+m}}{|\Lambda|^{1/2}} \frac{\Gamma(\nu/2+m)}{\Gamma(\nu/2)} \mathbb{M}\left[(-Q)^m\right], \tag{52}$$

$$m = 0, 1, \ldots,$$

having defined $Q(x) \equiv x^{\mathrm{T}}[\Lambda^{-1} - s^{-1}\mathbb{I}_\nu]x$ for $x \in \mathbb{R}^\nu$ and $\mathbb{M}$ as the uniform average operator on the $(\nu-1)$-sphere $\partial \mathcal{B}_\nu(1) \equiv \{u \in \mathbb{R}^\nu : u^{\mathrm{T}}u = 1\}$; namely,

$$\mathbb{M}[\phi] \equiv \frac{\Gamma(\nu/2)}{2\pi^{\nu/2}} \int_{\partial \mathcal{B}_\nu(1)} \mathrm{d}u\, \phi(u), \tag{53}$$

$$\forall \phi \in \mathscr{C}^0\left(\partial \mathcal{B}_\nu(1)\right) \text{ a.e.}$$

Unfortunately, (52) is not particularly convenient for numerical computations, since $\mathbb{M}[(-Q)^m]$ is only given in integral form. However, it is also shown in [2] that the coefficients c_m can be extracted from the Taylor expansion (at $z_0 = 0$) of the generating function

$$\psi(z) = \prod_{k=1}^{\nu} \left(\frac{s}{\lambda_k}\right)^{1/2} \left[1 - \left(1 - \frac{s}{\lambda_k}\right)z\right]^{-1/2}, \tag{54}$$

that is $\psi(z) = \sum_{m=0}^{\infty} c_m(s;\lambda) z^m$.

This series converges uniformly for $|z| < \min_i|1 - s/\lambda_i|^{-1}$. On evaluating the derivatives of $\psi(z)$, it is then shown that c_m's fulfill the recursion:

$$c_0 = \prod_{m=1}^{\nu} \sqrt{\frac{s}{\lambda_m}};$$

$$c_n = \frac{1}{2n} \sum_{r=0}^{n-1} g_{n-r} c_r, \tag{55}$$

$$n = 1, 2, \ldots;$$

$$g_n \equiv \sum_{m=1}^{\nu} \left(1 - \frac{s}{\lambda_m}\right)^n.$$

Finally, the systematic error produced on considering only the lowest k terms of the chi-square series of (51) is estimated by

$$\mathscr{R}_n(\rho;\lambda) \equiv \left|\sum_{m=n}^{\infty} c_m(s;\lambda) F_{\nu+2m}\left(\frac{\rho}{s}\right)\right| \le c_0(s;\lambda)$$

$$\cdot \frac{\Gamma\left(\dfrac{\nu}{2}+n\right)}{\Gamma\left(\dfrac{\nu}{2}\right)} \frac{\eta^n}{n!} (1-\eta)^{-\left(\frac{\nu}{2}+n\right)} F_{\nu+2n}\left[\frac{(1-\eta)\rho}{s}\right] \tag{56}$$

$$\equiv \mathfrak{R}_n,$$

with $\eta = \max_i|1 - s/\lambda_i|$.

Now, as mentioned, it is possible to extend the above expansion to all Gaussian integrals $\alpha_{k\ell m\cdots}$. Here, we are interested only in α_k and α_{jk}, since these are needed in order to implement the fixed point iteration and to compute the Jacobian matrix of τ_ρ. The extension is provided by the following.

Theorem 9 (Ruben's expansions). *The integrals α_k and α_{jk} admit the series representations:*

$$\alpha_k(\rho;\lambda) = \sum_{m=0}^{\infty} c_{k;m}(s;\lambda) F_{\nu+2(m+1)}\left(\frac{\rho}{s}\right), \tag{57}$$

$$\alpha_{jk}(\rho;\lambda) = \sum_{m=0}^{\infty} c_{jk;m}(s;\lambda) F_{\nu+2(m+2)}\left(\frac{\rho}{s}\right), \tag{58}$$

with s being an arbitrary positive constant. The series coefficients are given, respectively, by

$$c_{k;m}(s;\lambda) = \frac{s}{\lambda_k} \frac{\nu+2m}{m!} \frac{s^{\nu/2+m}}{|\Lambda|^{1/2}} \frac{\Gamma(\nu/2+m)}{\Gamma(\nu/2)}$$

$$\cdot \mathbb{M}\left[(-Q)^m u_k^2\right], \tag{59}$$

$$c_{jk;m}(s;\lambda) = \left(1 + 2\delta_{jk}\right) \frac{s}{\lambda_j} \frac{s}{\lambda_k} \frac{(\nu+2m+2)(\nu+2m)}{m!}$$

$$\cdot \frac{s^{\nu/2+m}}{|\Lambda|^{1/2}} \frac{\Gamma(\nu/2+m)}{\Gamma(\nu/2)} \mathbb{M}\left[(-Q)^m u_j^2 u_k^2\right], \tag{60}$$

with δ_{jk} denoting the Kronecker symbol. The series on the right-hand side of (57)-(58) converge uniformly on every finite interval of ρ. The functions

$$\psi_k(z) = \left(\frac{s}{\lambda_k}\right)^{3/2} \left[1 - \left(1 - \frac{s}{\lambda_k}\right)z\right]^{-3/2} \prod_{i \ne k} \left(\frac{s}{\lambda_i}\right)^{1/2}$$

$$\cdot \left[1 - \left(1 - \frac{s}{\lambda_i}\right)z\right]^{-1/2}, \tag{61}$$

$$\psi_{kk}(z) = 3\left(\frac{s}{\lambda_k}\right)^{5/2}\left[1-\left(1-\frac{s}{\lambda_k}\right)z\right]^{-5/2}$$

$$\cdot\prod_{i\neq k}\left(\frac{s}{\lambda_i}\right)^{1/2}\left[1-\left(1-\frac{s}{\lambda_i}\right)z\right]^{-1/2},$$

(62)

$$\psi_{jk}(z) = \left(\frac{s}{\lambda_j}\frac{s}{\lambda_k}\right)^{3/2}$$

$$\cdot\left\{\left[1-\left(1-\frac{s}{\lambda_j}\right)z\right]\left[1-\left(1-\frac{s}{\lambda_k}\right)z\right]\right\}^{-3/2}$$

(63)

$$\times\prod_{i\neq j,k}\left(\frac{s}{\lambda_i}\right)^{1/2}\left[1-\left(1-\frac{s}{\lambda_i}\right)z\right]^{-1/2}\quad(j\neq k)$$

are generating functions, respectively, for the coefficients $c_{k;m}$, $c_{kk;m}$ and $c_{jk;m}$ $(j\neq k)$; that is, they fulfill

$$\psi_k(z) = \sum_{m=0}^{\infty}c_{k;m}(s;\lambda)z^m,$$

$$\psi_{jk}(z) = \sum_{m=0}^{\infty}c_{jk;m}(s;\lambda)z^m,$$

(64)

for $|z| < \min_i|1-s/\lambda_i|^{-1}$. Finally, the coefficients $c_{k;m}$, $c_{kk;m}$, and $c_{jk;m}$ $(j\neq k)$ can be obtained iteratively from the recursions

$$c_{k;0} = \left(\frac{s}{\lambda_k}\right)c_0;$$

$$c_{k;m} = \frac{1}{2m}\sum_{r=0}^{m-1}g_{k;m-r}c_{k;r};$$

$$g_{k;m} \equiv \sum_{i=1}^{v}e_{k;i}\left(1-\frac{s}{\lambda_i}\right)^m,$$

(65)

$$m\geq 1;$$

$$c_{jk;0} = \left(1+2\delta_{jk}\right)\left(\frac{s}{\lambda_j}\right)\left(\frac{s}{\lambda_k}\right)c_0;$$

$$c_{jk;m} = \frac{1}{2m}\sum_{r=0}^{m-1}g_{jk;m-r}c_{jk;r};$$

$$g_{jk;m} \equiv \sum_{i=1}^{v}e_{jk;i}\left(1-\frac{s}{\lambda_i}\right)^m,$$

(66)

$$m\geq 1,$$

where the auxiliary coefficients $e_{k;i}$ and $e_{jk;i}$ are defined by

$$e_{k;i} = \begin{cases}3 & \text{if } i=k\\ 1 & \text{otherwise,}\end{cases}$$

(67)

$$e_{kk;i} = \begin{cases}5 & \text{if } i=k\\ 1 & \text{otherwise,}\end{cases}$$

(68)

$$e_{jk;i} = \begin{cases}3 & \text{if } i=j \text{ or } k\\ 1 & \text{otherwise}\end{cases}\quad(j\neq k).$$

It is not difficult to further generalize this theorem, so as to provide a chi-square expansion for any Gaussian integral $\alpha_{k\ell m\cdots}$. The proof follows closely the original one given by Ruben. We reproduce it in the appendix for α_k, just to highlight the differences arising when the Gaussian integral contains powers of the integration variable.

Analogously to (56), it is possible to estimate the systematic error produced when considering only the lowest k terms of the chi-square series of α_k and α_{jk}. Specifically, we find that

$$\mathcal{R}_{k;n} \equiv \left|\sum_{m=n}^{\infty}c_{k;m}(s;\lambda)F_{v+2(m+1)}\left(\frac{\rho}{s}\right)\right| \leq c_{k;0}$$

$$\cdot\frac{\eta^n}{n!}(1-\eta)^{-(v/2+n+1)}\frac{\Gamma(v/2+n+1)}{\Gamma(v/2)}$$

$$\cdot F_{v+2(n+1)}\left[\frac{(1-\eta)\rho}{s}\right] \equiv \mathfrak{R}_{k;n},$$

(69)

$$\mathcal{R}_{jk;n} \equiv \left|\sum_{m=n}^{\infty}c_{jk;m}(s;\lambda)F_{v+2(m+2)}\left(\frac{\rho}{s}\right)\right| \leq c_{jk;0}$$

$$\cdot\frac{\eta^n}{n!}(1-\eta)^{-(v/2+n+2)}\frac{\Gamma(v/2+n+2)}{\Gamma(v/2)}$$

$$\cdot F_{v+2(n+2)}\left[\frac{(1-\eta)\rho}{s}\right] \equiv \mathfrak{R}_{jk;n}.$$

In order to evaluate all Ruben series with controlled uncertainty, we first set (see once more [2] for an exhaustive discussion on how to choose s) $s = 2\lambda_1\lambda_v/(\lambda_1+\lambda_v)$; then we choose a unique threshold ε representing the maximum tolerable systematic error; for example, $\varepsilon_{dp} = 1.0\times10^{-14}$ (roughly corresponding to double floating-point precision), for all α, α_k, and α_{jk}, and finally for each α_X we compute the integer

$$k_{th} \equiv \min_{n\geq 1}\{n : \mathfrak{R}_{X;n} < \varepsilon\},$$

(70)

providing the minimum number of chi-square terms, for which the upper bound $\mathfrak{R}_{X;n}$ to the residual sum $\mathcal{R}_{X;n}$ lies below ε. Of course, this procedure overshoots the minimum number of terms really required for the $\mathcal{R}$'s to lie below ε, since we actually operate on the $\mathfrak{R}$'s instead of the $\mathcal{R}$'s. Nevertheless, the computational overhead is acceptable, as it will be shown in the next section. For the sake of

completeness, it must be said that typically the values of k_{th} for α, α_k, and α_{jk} with the same ϵ (and ρ, λ) are not much different from each other.

To conclude, we notice that k_{th} depends nontrivially upon λ. By contrast, since $F_\nu(x)$ is monotically increasing in x, we clearly see that k_{th} is monotically increasing in ρ. Now, should one evaluate α and the like for a given λ at several values of ρ, say $\rho_1 \leq \rho_2 \leq \cdots \leq \rho_{max}$, it is advisable to save computing resources and work out Ruben coefficients just once, up to the order k_{th} corresponding to ρ_{max}, since $k_{th}(\rho_1) \leq \cdots \leq k_{th}(\rho_{max})$. We made use of this trick throughout our numerical experiences, as reported in the sequel.

6. Numerical Analysis of the Reconstruction Process

The fixed point (39) represents the simplest iterative scheme that can be used in order to reconstruct the solution $\lambda = \tau_\rho^{-1} \cdot \mu$. In the literature of numerical methods, this scheme is known as a nonlinear Gauss–Jacobi (GJ) iteration (see, e.g., [20]). Accordingly, we shall rewrite it as $\lambda_{GJ,k}^{(n+1)} = T_k(\lambda_{GJ}^{(n)})$. As we have seen, the sequence $\lambda_{GJ}^{(n)}$ converges with no exception as $n \to \infty$, provided $\mu \in \mathcal{D}(\tau_\rho^{-1})$. Given $\epsilon_T > 0$, the number of steps n_{it} needed for an approximate convergence with relative precision ϵ_T, that is,

$$n_{it} \equiv \min_{n \geq 1} \left\{ n : \frac{\left\| \lambda_{GJ}^{(n)} - \lambda_{GJ}^{(n-1)} \right\|_\infty}{\left\| \lambda_{GJ}^{(n-1)} \right\|_\infty} < \epsilon_T \right\}, \qquad (71)$$

depends not only upon ϵ_T, but also on ρ and μ (note that the stopping rule is well conditioned, since $\|\lambda^{(n)}\|_\infty > 0$ $\forall n$ and also $\lim_{n \to \infty} \|\lambda^{(n)}\|_\infty > 0$). In order to characterize statistically the convergence rate of the reconstruction process, we must integrate out the fluctuations of n_{it} due to changes of μ; that is, we must average n_{it} by letting μ fluctuate across its own probability space. In this way, we obtain the quantity $\bar{n}_{it} \equiv \mathbb{E}_\mu[n_{it} \mid \epsilon_T, \rho]$, which better synthesizes the cost of the reconstruction for given ϵ_T and ρ. It should be evident that carrying out this idea analytically is hard, for on the one hand n_{it} depends upon μ nonlinearly and on the other hand μ has a complicated distribution, as we briefly explain below.

6.1. Choice of the Eigenvalue Ensemble. Since λ is the eigenvalue spectrum of a full covariance matrix, it is reasonable to assume its distribution to be a Wishart $\mathcal{W}_\nu(p, \Sigma_0)$ for some scale matrix Σ_0 and for some number of degrees of freedom $p \geq \nu$. In the sequel, we shall make the ideal assumption $\Sigma_0 = p^{-1} \cdot \mathbb{I}_\nu$, so that the probability measure of λ is (see, e.g., [21])

$$dw_\nu(p; \lambda) = p^{(p+\nu^2-1)/2}$$

$$\cdot \frac{\pi^{\nu^2/2} \prod_{k=1}^\nu \lambda_k^{(p-\nu-1)/2} \exp\left((-p/2)\sum_{k=1}^\nu \lambda_k\right) \prod_{k<j} \left(\lambda_j - \lambda_k\right)}{2^{\nu p/2} \Gamma_\nu(p/2)\, \Gamma_\nu(\nu/2)} \quad (72)$$

$$\cdot d^\nu \lambda.$$

Under this assumption, the probability measure of μ is obtained by performing the change of variable $\lambda = \tau_\rho^{-1} \cdot \mu$ in (72). Unfortunately, we have no analytic representation of τ_ρ^{-1}. Thus, we have neither an expression for the distribution of μ. However, μ can be extracted numerically as follows:

(i) generate randomly $\Sigma \sim \mathcal{W}_\nu(p, p^{-1} \cdot \mathbb{I}_\nu)$ by means of the Bartlett decomposition [22];

(ii) take the ordered eigenvalue spectrum λ of Σ;

(iii) obtain μ by applying the truncation operator τ_ρ to λ.

Note that since $\mathcal{W}_\nu(p, p^{-1} \cdot \mathbb{I}_\nu)$ is only defined for $p \geq \nu$, we need to rescale p as ν increases. The simplest choice is to keep the ratio p/ν fixed. The larger this ratio, the closer Σ fluctuates around $\mathbb{I}_\nu$ (recall that if $\Sigma \sim \mathcal{W}_\nu(p, p^{-1} \cdot \mathbb{I}_\nu)$, then $\mathbb{E}[\Sigma_{ij}] = \delta_{ij}$ and $\text{var}(\Sigma_{ij}) = p^{-1}[1+\delta_{ij}]$). In view of this, large values of p/ν are to be avoided, since they reduce the probability of testing the fixed point iteration on eigenvalue spectra characterized by large condition numbers $n_{cond} \equiv \lambda_\nu/\lambda_1$. For this reason, we have set $p = 2\nu$ in our numerical study.

Having specified an ensemble of matrices from which the eigenvalue spectra are extracted, we are now ready to perform numerical simulations. To begin with, we report in Figure 3 the marginal probability density function of the ordered eigenvalues $\{\lambda_k\}_{k=1}^\nu$ and their truncated counterparts $\{\mu_k\}_{k=1}^\nu$ for the Wishart ensemble $\mathcal{W}_{10}(20, 20^{-1} \cdot \mathbb{I}_{10})$ at $\rho = 1$, as obtained numerically from a rather large sample of matrices ($\simeq 10^6$ units). It will be noted that (i) the effect of the truncation is severe on the largest eigenvalues, as a consequence of the analytic bounds of Corollary 2.1 and Proposition 2.2; (ii) while the skewness of the lowest truncated eigenvalues is negative, it becomes positive for the largest ones. This is due to a change of relative effectiveness of (29) (i) with respect to (29) (ii).

6.2. Choice of the Simulation Parameters. In order to explore the dependence of $\bar{n}_{it}$ upon ρ, we need to choose one or more simulation points for the latter. Ideally, it is possible to identify three different regimes in our problem: $\rho \lesssim \lambda_1$ (strong truncation regime), $\lambda_1 \lesssim \rho \lesssim \lambda_\nu$ (crossover), and $\rho \lesssim \lambda_\nu$ (weak truncation regime). We cover all of them with the following set of points:

$$\rho \in \{\text{Mo}(\lambda_1), \ldots, \text{Mo}(\lambda_\nu)\}$$
$$\cup \left\{ \frac{1}{2}\text{Mo}(\lambda_1), 2\text{Mo}(\lambda_\nu) \right\}, \qquad (73)$$

where $\text{Mo}(\cdot)$ stands for the mode. In principle, it is possible to determine $\text{Mo}(\lambda_k)$ with high accuracy by using analytic representations of the marginal probability densities of the ordered eigenvalues [24]. In practice, the latter become computationally demanding at increasingly large values of ν: for instance, the determination of the probability density of λ_2 requires $(\nu!)^2$ sums, which is unfeasible even at $\nu \sim 10$. Moreover, to our aims, it is sufficient to choose approximate values, provided that these lie not far from the exact ones. Accordingly, we have determined the eigenvalue modes
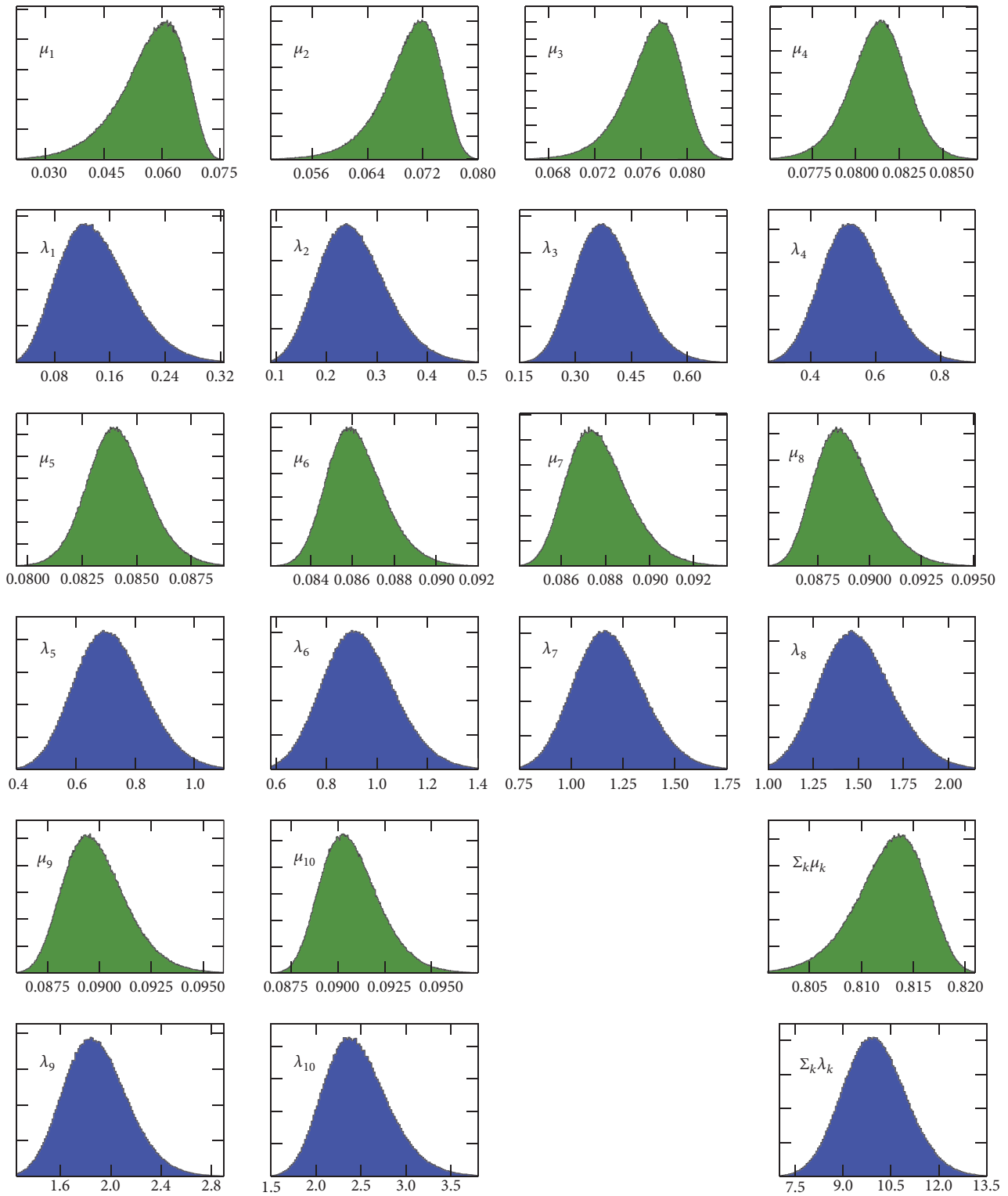
Figure 3: Monte Carlo simulation of the probability density function of the ordered eigenvalues λ_k (even rows) and their truncated counterparts μ_k at $\rho = 1$ (odd rows) for the Wishart ensemble $\mathscr{W}_{10}(20, 20^{-1} \cdot \mathbb{I}_{10})$. The last two plots (bottom right) display the distribution of the sum of eigenvalues.

TABLE 1: Numerical estimates of the mode of the ordered eigenvalues $\{\lambda_1, \ldots, \lambda_\nu\}$ of $\Sigma \sim \mathcal{W}_\nu(2\nu, (2\nu)^{-1} \cdot \mathbb{I}_\nu)$ with $\nu = 3, \ldots, 10$. The estimates have been obtained from Grenander's mode estimator [23].

	$\nu = 3$	$\nu = 4$	$\nu = 5$	$\nu = 6$	$\nu = 7$	$\nu = 8$	$\nu = 9$	$\nu = 10$
$\widehat{\mathrm{Mo}}(\lambda_1)$	0.1568	0.1487	0.1435	0.1383	0.1344	0.1310	0.1269	0.1258
$\widehat{\mathrm{Mo}}(\lambda_2)$	0.6724	0.4921	0.4017	0.3424	0.3039	0.2745	0.2554	0.2399
$\widehat{\mathrm{Mo}}(\lambda_3)$	1.6671	1.0112	0.7528	0.6071	0.5138	0.4543	0.4048	0.3693
$\widehat{\mathrm{Mo}}(\lambda_4)$	—	1.8507	1.2401	0.9621	0.7854	0.6684	0.5858	0.5288
$\widehat{\mathrm{Mo}}(\lambda_5)$	—	—	2.0150	1.4434	1.1269	0.9263	0.7956	0.7032
$\widehat{\mathrm{Mo}}(\lambda_6)$	—	—	—	2.1356	1.5789	1.2559	1.0527	0.9111
$\widehat{\mathrm{Mo}}(\lambda_7)$	—	—	—	—	2.2190	1.6764	1.3673	1.1603
$\widehat{\mathrm{Mo}}(\lambda_8)$	—	—	—	—	—	2.2763	1.7687	1.4624
$\widehat{\mathrm{Mo}}(\lambda_9)$	—	—	—	—	—	—	2.3210	1.8473
$\widehat{\mathrm{Mo}}(\lambda_{10})$	—	—	—	—	—	—	—	2.3775

numerically from samples of $N \simeq 10^6$ Wishart matrices. Our estimates are reported in Table 1 for $\nu = 3, \ldots, 10$. They have been obtained from Grenander's estimator [23]:

$$\mathrm{Mo}\left(\lambda_k\right)_{rs} = \frac{1}{2} \frac{\sum_{i=1}^{N-r}\left(\lambda_k^{(i)} + \lambda_k^{(i+r)}\right)\left(\lambda_k^{(i)} - \lambda_k^{(i+r)}\right)^{-s}}{\sum_{i=1}^{N-r}\left(\lambda_k^{(i)} - \lambda_k^{(i+r)}\right)^{-s}}, \quad (74)$$

with properly chosen parameters r, s.

We are now in the position to investigate numerically how many terms in Ruben's expansions must be considered as ε is set to $\varepsilon_{\mathrm{dp}} = 1.0 \times 10^{-14}$, for our choice of the eigenvalue ensemble $\lambda \sim \mathcal{W}_\nu(2\nu, (2\nu)^{-1} \cdot \mathbb{I}_\nu)$ and with ρ set as in Table 1. As an example, we report in Figure 4 the discrete distributions of k_{th} for the basic Gaussian integral α at $\nu = 10$, the largest dimension that we have simulated. As expected, we observe an increase of k_{th} with ρ. Nevertheless, we see that the number of Ruben's components to be taken into account for a double precision result keeps altogether modest even in the weak truncation regime, which proves the practical usefulness of the chi-square expansions.

6.3. Fixed Point Iteration at Work.

The GJ iteration is too slow to be of practical interest. For instance, at $\nu = 10$, $\rho \simeq \mathrm{Mo}(\lambda_1)$ and $\epsilon_{\mathrm{T}} = 1.0 \times 10^{-7}$ (corresponding to a reconstruction of λ with single floating-point precision), it is rather easy to extract realizations of μ which require $n_{\mathrm{it}} \simeq 15,000$ to converge. An improvement of the GJ scheme is achieved via overrelaxation (GJOR); that is,

$$\lambda_{\mathrm{GJOR},k}^{(0)} = \mu_k,$$

$$\lambda_{\mathrm{GJOR},k}^{(n+1)} = \lambda_{\mathrm{GJOR},k}^{(n)} + \omega\left[T_k\left(\lambda_{\mathrm{GJOR}}^{(n)}\right) - \lambda_{\mathrm{GJOR},k}^{(n)}\right], \quad (75)$$

$$k = 1, \ldots, \nu.$$

Evidently, at $\omega = 1$, the GJOR scheme coincides with the standard GJ one. The optimal value ω_{opt} of the relaxation factor ω is not obvious even in the linear Jacobi scheme, where ω_{opt} depends upon the properties of the coefficient matrix of the system. For instance, if the latter is symmetric positive definite, it is demonstrated that the best choice is provided by $\omega_{\mathrm{opt}} \equiv 2(1 + \sqrt{1 - \sigma^2})^{-1}$, σ being the spectral radius of the

Jacobi iteration matrix [25]. In our numerical tests with the GJOR scheme, we found empirically that the optimal value of ω at $\rho \lesssim \lambda_\nu$ is close to the linear prediction, provided that σ is replaced by $\|J\|_\infty$, J being defined as in Section 3 (note that $\|J\|_\infty < 1$).

By contrast, the iteration diverges after few steps with increasing probability as $\rho/\lambda_\nu \rightarrow \infty$ if ω is kept fixed at $\omega = \omega_{\mathrm{opt}}$; in order to restore the convergence, ω must be lowered towards $\omega = 1$ as such limit is taken.

To give an idea of the convergence rate of the GJOR scheme, we show in Figure 5(a) a joint box-plot of the distributions of n_{it} at $\nu = 10$ and $\epsilon_{\mathrm{T}} = 1.0 \times 10^{-7}$. From the plot we observe that the distribution of n_{it} shifts rightwards as ρ decreases: clearly, the reconstruction is faster if ρ is in the weak truncation regime (where μ is closer to λ), whereas it takes more iterations in the strong truncation regime. The dependence of $\bar{n}_{\mathrm{it}}$ upon ρ, systematically displayed in Figure 6, is compatible with a scaling law

$$\log \bar{n}_{\mathrm{it}}\left(\rho, \nu, \epsilon_{\mathrm{T}}\right) = a\left(\nu, \epsilon_{\mathrm{T}}\right) - b\left(\nu, \epsilon_{\mathrm{T}}\right)\log \rho, \quad (76)$$

apart from small corrections occurring at large ρ. Equation (76) tells us that $\bar{n}_{\mathrm{it}}$ increases polynomially in $1/\rho$ at fixed ν. In order to estimate the parameters a and b in the strong truncation regime (where the algorithm becomes challenging), we performed jackknife fits to (76) of data points with $\rho \lesssim 1$. Results are collected in Figure 5(b), showing that b is roughly constant, while a increases almost linearly in ν. Thus, while the cost of the eigenvalue reconstruction is only polynomial in $1/\rho$ at fixed ν, it is exponential in ν at fixed ρ. The scaling law of the GJOR scheme is therefore better represented by $\bar{n}_{\mathrm{it}} = C e^{\kappa \nu}/\rho^b$, with C being a normalization constant independent of ρ and ν and κ representing approximately the slope of a as a function of ν. Although the GJOR scheme improves the GJ one, the iteration reveals to be still inefficient in a parameter subspace, which is critical for the applications.

6.4. Boosting the GJOR Scheme.

A further improvement can be obtained by letting ω depend on the eigenvalue index in the GJOR scheme. Let us discuss how to work out such an adjustment. On commenting on Figure 3, we have already noticed that the largest eigenvalues are affected by the truncation to a larger extent than the smallest ones. Therefore,
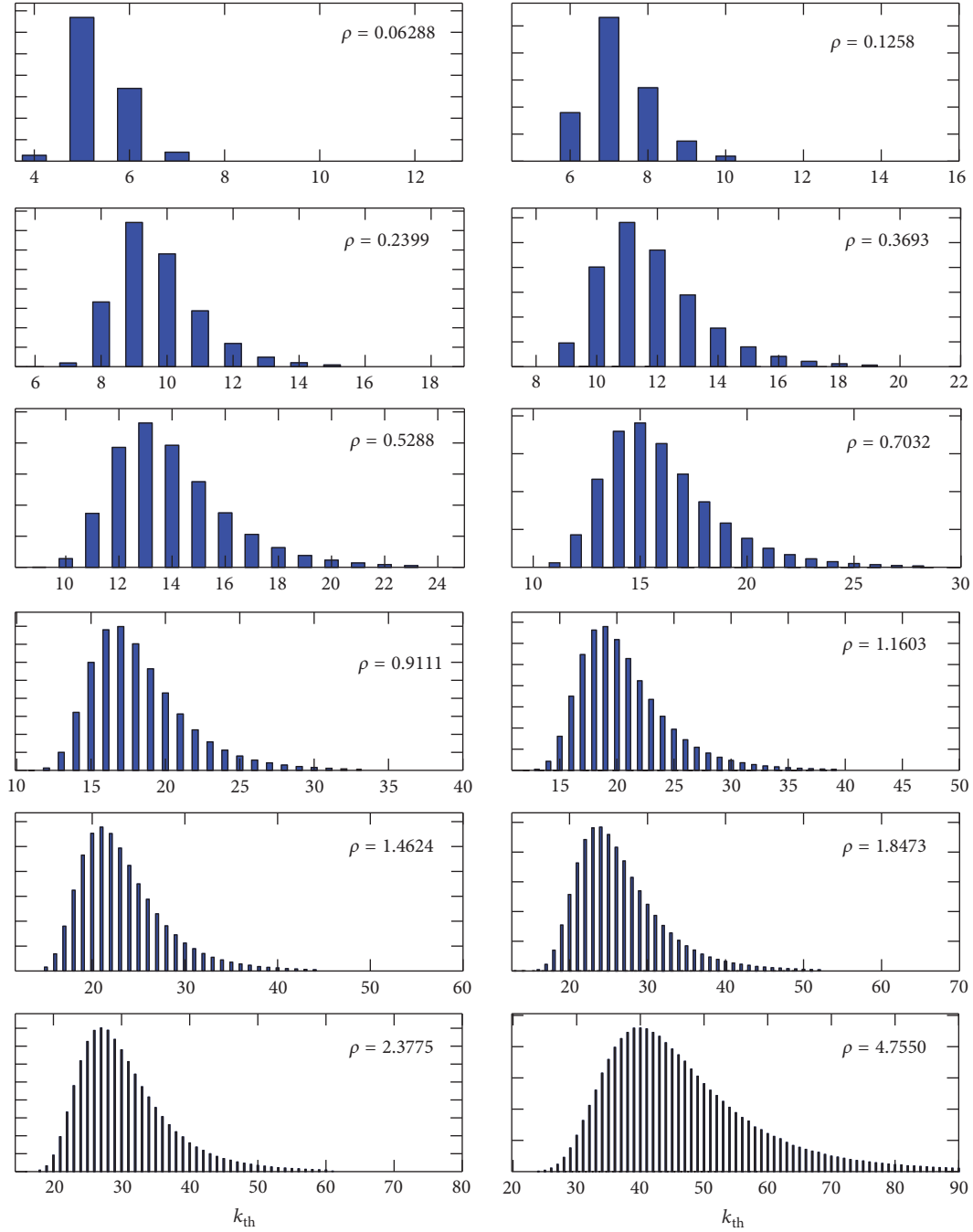
FIGURE 4: Monte Carlo simulation of the probability mass function of the parameter k_{th} for the Gaussian probability content α. The histograms refer to the eigenvalue ensemble λ of $\Sigma \sim \mathscr{W}_{10}(20, 20^{-1} \cdot \mathbb{I}_{10})$, with ρ chosen as in Table 1 and $\varepsilon = 1.0 \times 10^{-14}$.

they must perform a longer run through the fixed point iteration, in order to converge to the untruncated values. This is a possible qualitative explanation for the slowing down of the algorithm as $\rho \to 0$. In view of it, we expect to observe some acceleration of the convergence rate, if ω is replaced, for instance, by

$$\omega \longrightarrow \omega_k \equiv (1 + \beta \cdot k)\,\omega_{\mathrm{opt}}, \quad \beta \geq 0, \ k = 1, \ldots, \nu. \quad (77)$$

The choice $\beta = 0$ corresponds obviously to the standard GJOR scheme. Any other choice yields $\omega_k > \omega_{\mathrm{opt}}$. Therefore, the new scheme is also expected to display a higher rate of failures than the GJOR one at $\rho \gg \lambda_\nu$, for the reason explained in Section 5.3. The componentwise overrelaxation proposed in (77) is only meant to enhance the convergence speed in the strong truncation regime and in the crossover, where the improvement is actually needed.
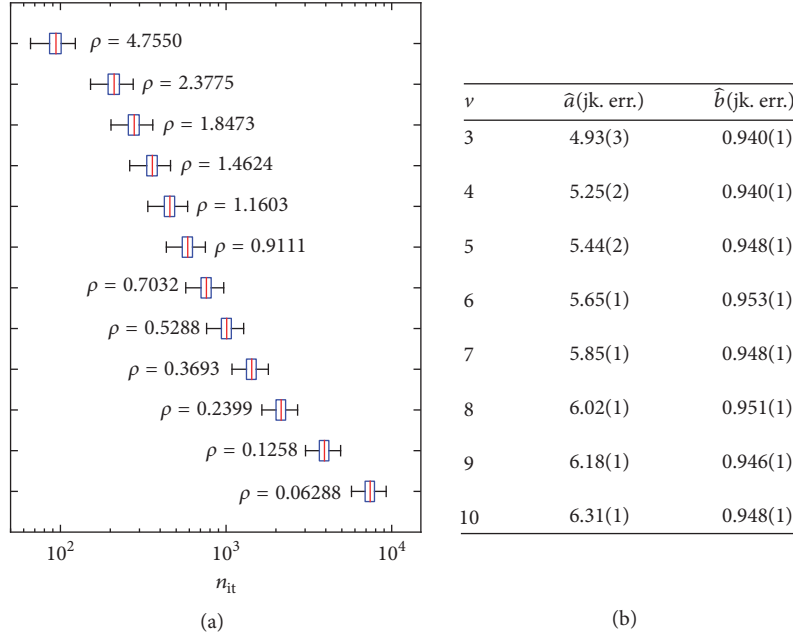
v	$\hat{a}$(jk. err.)	$\hat{b}$(jk. err.)
3	4.93(3)	0.940(1)
4	5.25(2)	0.940(1)
5	5.44(2)	0.948(1)
6	5.65(1)	0.953(1)
7	5.85(1)	0.948(1)
8	6.02(1)	0.951(1)
9	6.18(1)	0.946(1)
10	6.31(1)	0.948(1)

(a)   (b)

FIGURE 5: (a) Box-plot of n_{it} in the GJOR scheme at $v = 10$, with $\epsilon_{\text{T}} = 1.0 \times 10^{-7}$ and ρ chosen as in Table 1. The distributions have been reconstructed from a sample of $N \simeq 10^3$ eigenvalue spectra extracted from $\mathscr{W}_{10}(20, 20^{-1} \cdot \mathbb{I}_{10})$. The whiskers extend to the most extreme data point within $(3/2)(75\%-25\%)$ data range. (b) Numerical estimates of the scaling parameters a and b of the GJOR scheme, as obtained from jackknife fits to (76) of data points with $\rho \lesssim 1$ and $\epsilon_{\text{T}} = 1.0 \times 10^{-7}$. We quote in parentheses the jackknife error.

In order to confirm this picture, we have explored systematically the effect of β on $\overline{n}_{\text{it}}$ by simulating the reconstruction process at $v = 3, \ldots, 10$, with β varying from 0 to 2 in steps of 1/5. First of all, we have observed that the rate of failures at large ρ is fairly reduced if the first $30 \div 50$ iterations are run with $\omega_k = \omega_{\text{opt}}$, and only afterwards β is switched on. Having minimized the failures, we have checked that for each value of β, the scaling law assumed in (76) is effectively fulfilled. Then, we have computed jackknife estimates of the scaling parameters a and b. These are plotted in Figure 7 as functions of v. Each trajectory (represented by a dashed curve) corresponds to a given value of β. Those with darker markers refer to smaller values of β and the other way round. From the plots we notice that

(i) all the trajectories with $\beta > 0$ lie below the one with $\beta = 0$;

(ii) the trajectories of a display a clear increasing trend with v; yet their slope lessens as β increases. By contrast, the trajectories of b develop a mild increasing trend with v as β increases, though this is not strictly monotonic;

(iii) the trajectories of both a and b seem to converge to a limit trajectory as β increases; we observe a saturation phenomenon, which thwarts the benefit of increasing β beyond a certain threshold close to $\beta_{\text{max}} \simeq 2$.

We add that pushing β beyond β_{max} is counterproductive, as the rate of failures becomes increasingly relevant in the crossover and eventually also in the strong truncation regime.

By contrast, if $\beta \lesssim \beta_{\text{max}}$ the rate of failures keeps very low for essentially all simulated values of ρ.

Our numerical results signal a strong reduction of the slowing down of the convergence rate. Indeed, (i) means qualitatively that C and b are reduced as β increases. (ii) means that κ is reduced as β increases (this is the most important effect, as κ is mainly responsible for the exponential slowing down with v). The appearance of a slope in the trajectories of b as β increases indicates that a mild exponential slowing down is also developed at denominator of the scaling law $\overline{n}_{\text{it}} = Ce^{\kappa v}/\rho^b$, but the value of b is anyway smaller than at $\beta = 0$. Finally, (iii) means that choosing $\beta > \beta_{\text{max}}$ has a minor impact on the performance of the algorithm. In Figure 8, we report a plot of the parameter κ (obtained from least-squares fits of data to a linear model $a = a_0 + \kappa \cdot v$) as a function of β. We see that $\kappa(\beta = 0)/\kappa(\beta = 2) \simeq 4$. This quantifies the maximum exponential speedup of the convergence rate, which can be achieved by our proposal. When β is close to β_{max}, $\overline{n}_{\text{it}}$ amounts to few hundreds at $v = 10$ and $\rho \simeq \lambda_1/2$.

7. On the Ill-Posedness of the Reconstruction in Sample Space

So far we have discussed the covariance reconstruction under the assumption that $\mu = \tau_\rho \cdot \lambda$ represents the exact truncated counterpart of some $\lambda \in \mathbb{R}^v$ and we have looked at the algorithmic properties of the iteration schemes which operatively define τ_ρ^{-1}. Such analysis is essential in order to characterize τ_ρ^{-1} mathematically; yet it is not sufficient in real situations, specifically when μ is perturbed by statistical noise.
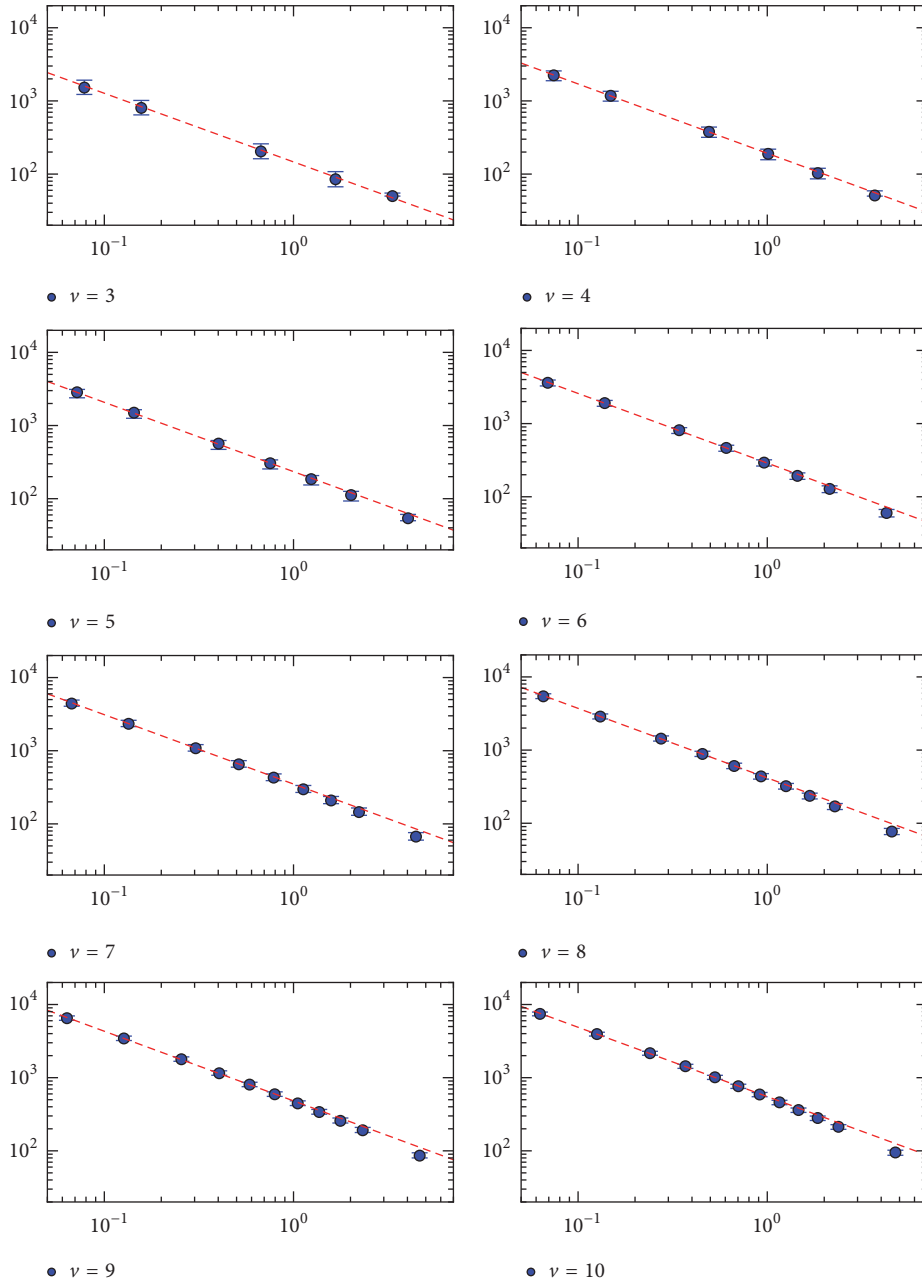
FIGURE 6: Log-log plots of $\overline{n}_{it}$ versus ρ in the GJOR scheme at $\epsilon_T = 1.0 \times 10^{-7}$. The parameter ρ has been chosen as in Table 1. The (red) dashed line in each plot represents our best jackknife linear fit to (76) of data points with $\rho \lesssim 1$.

In this section, we examine the difficulties arising when performing the covariance reconstruction in sample space. We first recall that, according to Hadamard [26], a mathematical problem is well posed provided that the following conditions are fulfilled:

(H$_1$) there exists always a solution to the problem;

(H$_2$) the solution is unique;

(H$_3$) the solution depends smoothly on the input data.

Inverse problems are often characterized by violation of one or more of them; see, for instance, [13]. In such cases,

the standard practice consists in regularizing the inverse operator, that is, in replacing it by a stable approximation. With regard to our problem, the reader will recognize that (H$_1$) is violated (and the problem becomes ill-posed) as soon as the space of the input data is allowed to be a superset of $\mathscr{D}(\tau_\rho^{-1})$: once clarifying how μ is concretely estimated in the applications (Sections 7.1 and 7.2), we propose a perturbative regularization of τ_ρ^{-1}, which improves effectively the fulfillment of (H$_1$) (Section 7.3). By contrast, Proposition 8 guarantees that whenever a solution exists, it is also unique; thus (H$_2$) is never of concern. Finally, the fulfillment of (H$_3$)
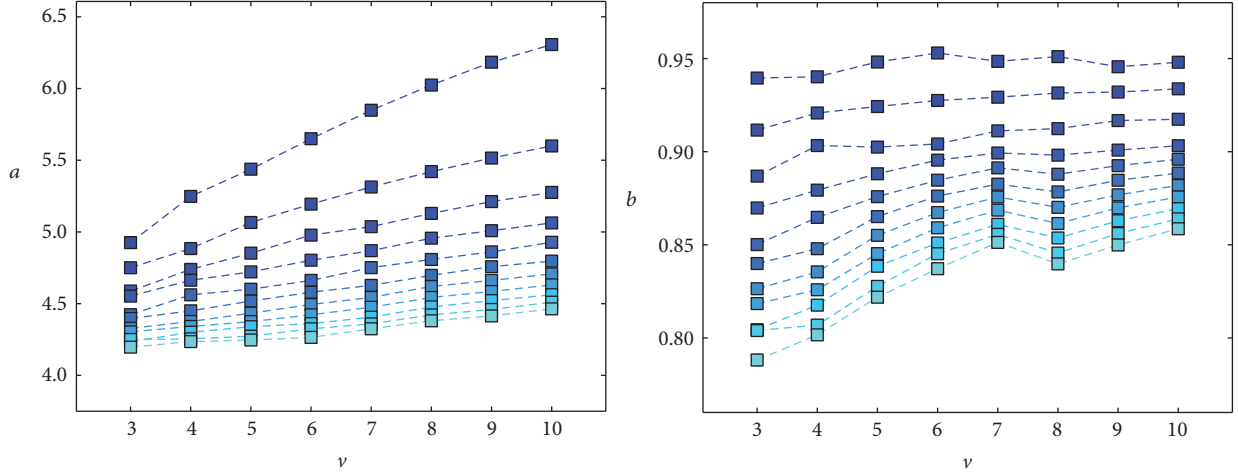
FIGURE 7: Scaling parameters a and b of the modified GJOR scheme as functions of v at $\epsilon_T = 1.0 \times 10^{-7}$, with β varying in the range $0 \div 2$ in steps of $1/5$. Each trajectory (represented by a dashed curve) refers to a different value of β. Those with darker markers correspond to smaller values of β and the other way round.
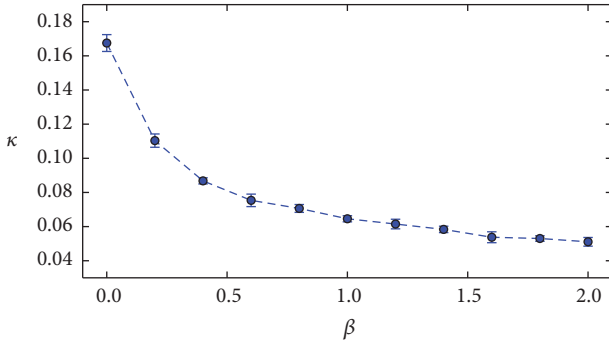

FIGURE 8: The parameter κ as a function of β. Estimates of κ are obtained from least-squares fits of data to a linear model $a = a_0 + \kappa \cdot v$.

depends on how the statistical noise on μ is nonlinearly inflated by the action of τ_ρ^{-1}. For the sake of conciseness, in the present paper, we just sketch the main ideas underlying the perturbative regularization of τ_ρ^{-1}, whereas a technical implementation of it and a discussion of (H$_3$) are deferred to a separate paper [27].

7.1. Definition of the Sample Truncated Covariance Matrix.
The examples of Section 2 assume that (i) spherical truncations are operated on a representative sample $\mathscr{P}_N = \{x^{(k)}\}_{k=1}^N$ of $X \sim \mathscr{N}_v(0, \Sigma)$ with finite size N, (ii) ρ is known exactly, and (iii) the input budget for the covariance reconstruction is given by the subset

$$\mathcal{Q}_M = \left\{ x \in \mathscr{P}_N : \|x\|^2 < \rho \right\}, \quad \text{with } |\mathcal{Q}_M| = M \leq N. \quad (78)$$

As usual in the analysis of stochastic variables in sample space, we assume that the observations $x^{(k)}$ are realizations

of i.i.d. stochastic variables $X^{(k)} \sim \mathscr{N}_v(0, \Sigma)$, $k = 1, \ldots, N$. Thus, M is itself a stochastic variable in sample space, where it reads

$$M = \sum_{k=1}^N \mathbb{I}_{\mathscr{B}_v(\rho)}\left(X^{(k)}\right) \equiv \sum_{k=1}^N \mathbb{I}_k, \quad (79)$$

with $\mathbb{I}_{\mathscr{B}_v(\rho)}(\cdot)$ denoting the characteristic function of $\mathscr{B}_v(\rho)$ and $\mathbb{I}_k \equiv \mathbb{I}_{\mathscr{B}_v(\rho)}(X^{(k)})$ being just a shortcut for its extended counterpart. It is easily recognized that $M \sim B(N, \alpha)$ is a binomial variate. If we indeed denote by $\mathfrak{E}$ the sample expectation operator (*i.e.*, the integral with respect to the product measure of the joint variables $\{X^{(k)}\}_{k=1}^N$), then a standard calculation yields

$$\mathfrak{E}[M] = \mathfrak{E}\left[\sum_{k=1}^N \mathbb{I}_k\right] = \sum_{k=1}^N \mathfrak{E}[\mathbb{I}_k] = \sum_{k=1}^N \alpha = \alpha N,$$

$$\mathfrak{var}[M] = \mathfrak{E}\left[M^2\right] - \mathfrak{E}[M]^2$$

$$= \sum_{k=1}^N \mathfrak{E}\left[\mathbb{I}_k^2\right] + \sum_{k,s:k\neq s}^{1\cdots N} \mathfrak{E}[\mathbb{I}_k \mathbb{I}_s] - \alpha^2 N^2$$

$$= \sum_{k=1}^N \mathfrak{E}[\mathbb{I}_k] + \sum_{k,s:k\neq s}^{1\cdots N} \mathfrak{E}[\mathbb{I}_k]\, \mathfrak{E}[\mathbb{I}_s] - \alpha^2 N^2$$

$$= \alpha N + \alpha^2 N (N-1) - \alpha^2 N^2$$

$$= \alpha (1-\alpha) N.$$

(80)

Hence, we see that the relative dispersion of M is $O(N^{-1/2})$. Now, the simplest way to measure Σ and $\mathfrak{S}_{\mathcal{B}}$, respectively, from the sets $\mathscr{P}_N$ and $\mathcal{Q}_M$ is via the classical estimators

$$\widehat{\Sigma}_{ij} = \frac{1}{N-1} \sum_{x \in \mathscr{P}_N} (x - \overline{x})_i \, (x - \overline{x})_j \,,$$

$$\overline{x}_i = \frac{1}{N} \sum_{x \in \mathscr{P}_N} x_i, \tag{81}$$

$$\left(\widehat{\mathfrak{S}}_{\mathcal{B}} \right)_{ij} = \frac{1}{M-1} \sum_{x \in \mathcal{Q}_M} (x - \widetilde{x})_i \, (x - \widetilde{x})_j \,,$$

$$\widetilde{x}_i = \frac{1}{M} \sum_{x \in \mathcal{Q}_M} x_i. \tag{82}$$

We define the sample estimates $\widehat{\lambda}$ and $\widehat{\mu}$, respectively, of λ and μ as the eigenvalue spectra of $\widehat{\Sigma}$ and $\widehat{\mathfrak{S}}_{\mathcal{B}}$. By symmetry arguments we see that $\widetilde{x}_i$ is unbiased. Indeed, it holds

$$\mathfrak{E} \left[\widetilde{x}_i \right] = \sum_{k=1}^{N} \mathfrak{E} \left[X_i^{(k)} \frac{\mathbb{I}_k}{\sum_{s=1}^{N} \mathbb{I}_s} \right]. \tag{83}$$

The *right-hand side* of (83) makes only sense if we conventionally define the integrand to be zero in the integration subdomain $\{X^{(k)} \notin \mathcal{B}_\nu(\rho), \ \forall k\}$ or equivalently if we interpret $\mathfrak{E}[\widetilde{x}_i]$ as the conditional one $\mathfrak{E}[\widetilde{x}_i \mid M > 0]$ (the event $M > 0$ occurs a.s. only as $N \to \infty$). Since the sample measure is even under $X^{(k)} \to -X^{(k)}$ while the integrand is odd, we immediately conclude that $\mathfrak{bias}[\widetilde{x}_i] = 0$.

7.2. Bias of the Sample Truncated Covariance Matrix. The situation gets somewhat less trivial with $\widehat{\mathfrak{S}}_{\mathcal{B}}$: the normalization factor $(M-1)^{-1}$, which has been chosen in analogy with (81), is not sufficient to remove completely the bias of $\widehat{\mathfrak{S}}_{\mathcal{B}}$ at finite N, though we aim at showing here that the residual bias is exponentially small and asymptotically vanishing. In order to see this, we observe that

$$\mathfrak{E} \left[\left(\widehat{\mathfrak{S}}_{\mathcal{B}} \right)_{ij} \right]$$

$$= \sum_{k=1}^{N} \mathfrak{E} \left[\left(X_i^{(k)} - \widetilde{X}_i \right) \left(X_j^{(k)} - \widetilde{X}_j \right) \frac{\mathbb{I}_k}{\sum_{s=1}^{N} \mathbb{I}_s - 1} \right]$$

$$= \sum_{\ell,r=1}^{\nu} R_{i\ell} R_{jr}$$

$$\cdot \sum_{k=1}^{N} \mathfrak{E}_{\text{diag}} \left[\left(X_\ell^{(k)} - \widetilde{X}_\ell \right) \left(X_r^{(k)} - \widetilde{X}_r \right) \frac{\mathbb{I}_k}{\sum_{s=1}^{N} \mathbb{I}_s - 1} \right], \tag{84}$$

with $\mathfrak{E}_{\text{diag}}$ denoting the sample expectation corresponding to a multinormal measure with diagonal covariance matrix $\Lambda = \text{diag}(\lambda) = R^{\text{T}} \Sigma R$, conditioned to $M > 1$. Having diagonalized the product measure, we observe that the integrand on the *right-hand side* is odd for $\ell \neq r$ and even for $\ell = r$ under the joint change of variables $X_\ell^{(k)} \to -X_\ell^{(k)}$ for $k = 1, \ldots, N$,

similarly to what we did in Section 2. As a consequence, it holds

$$\mathfrak{E} \left[\left(\widehat{\mathfrak{S}}_{\mathcal{B}} \right)_{ij} \right]$$

$$= \sum_{\ell=1}^{\nu} R_{i\ell} R_{j\ell} \left\{ \sum_{k=1}^{N} \mathfrak{E}_{\text{diag}} \left[\left(X_\ell^{(k)} - \widetilde{X}_\ell \right)^2 \frac{\mathbb{I}_k}{\sum_{s=1}^{N} \mathbb{I}_s - 1} \right] \right\}, \tag{85}$$

whence we infer that the matrix $\mathfrak{E}[\widehat{\mathfrak{S}}_{\mathcal{B}}]$ is diagonalized by the same matrix R as Σ. From (85) we also conclude that

$$\mathfrak{bias} \left[\widehat{\mathfrak{S}}_{\mathcal{B}} \right] = R \, \text{diag}(w) \, R^{\text{T}},$$

$$w_i \equiv \sum_{k=1}^{N} \mathfrak{E}_{\text{diag}} \left[\left(X_i^{(k)} - \widetilde{X}_i \right)^2 \frac{\mathbb{I}_k}{\sum_{s=1}^{N} \mathbb{I}_s - 1} \right] \tag{86}$$

$$- \mu_i,$$

$i = 1, \ldots, \nu$. It should be observed that in general $w_i \neq \mathfrak{bias}[\widehat{\mu}_i]$ since the computation of $\widehat{\mu}_i$ requires the diagonalization of $\widehat{\mathfrak{S}}_{\mathcal{B}}$, which is in general performed by a diagonalizing matrix $\widehat{R} \neq R$. Nevertheless, if w vanishes then $\mathfrak{bias}[\widehat{\mathfrak{S}}_{\mathcal{B}}]$ vanishes too. Now, we observe that w_i splits into three contributions:

$$w_{i1} = \sum_{k=1}^{N} \mathfrak{E}_{\text{diag}} \left[\left(X_i^{(k)} \right)^2 \frac{\mathbb{I}_k}{\sum_{s=1}^{N} \mathbb{I}_s - 1} \right] - \mu_i,$$

$$w_{i2} = -2 \sum_{k=1}^{N} \mathfrak{E}_{\text{diag}} \left[X_i^{(k)} \widetilde{X}_i \frac{\mathbb{I}_k}{\sum_{s=1}^{N} \mathbb{I}_s - 1} \right], \tag{87}$$

$$w_{i3} = \sum_{k=1}^{N} \mathfrak{E}_{\text{diag}} \left[\left(\widetilde{X}_i \right)^2 \frac{\mathbb{I}_k}{\sum_{s=1}^{N} \mathbb{I}_s - 1} \right],$$

which can be exactly calculated and expressed in terms of μ_i, α, and N. For instance,

$$w_{i1} = N \mathfrak{E}_{\text{diag}} \left[\frac{1}{M-1} \left(X_i^{(1)} \right)^2 \mathbb{I}_1 \mid M > 1 \right] - \mu_i$$

$$= N \sum_{m=2}^{N} \frac{1}{m-1} \mathfrak{E}_{\text{diag}} \left[\left(X_i^{(1)} \right)^2 \mathbb{I}_1 \mid M = m \right] - \mu_i$$

$$= N \sum_{m=2}^{N} \frac{1}{m-1} \alpha \mu_i \binom{N-1}{m-1} \alpha^{m-1} (1-\alpha)^{N-m} - \mu_i \tag{88}$$

$$= \mu_i \sum_{m=2}^{N} \frac{m}{m-1} \binom{N}{m} \alpha^m (1-\alpha)^{N-m} - \mu_i.$$

Analogously, we have

$$w_{i2} = -2 \mu_i \sum_{m=2}^{N} \frac{1}{m-1} \binom{N}{m} \alpha^m (1-\alpha)^{N-m},$$

$$w_{i3} = \mu_i \sum_{m=2}^{N} \frac{1}{m-1} \binom{N}{m} \alpha^m (1-\alpha)^{N-m}. \tag{89}$$

Hence, it follows that

$$w_i = -\mu_i \left[1 + \alpha \left(N - 1\right)\right] \left(1 - \alpha\right)^{N-1}. \tag{90}$$

Since $\alpha > 0$, we see that $\lim_{N\to\infty} w_i = 0$. Thus, we conclude that $\widehat{\mathfrak{S}}_{\mathscr{B}}$ is asymptotically unbiased.

A discussion of the variance of the sample truncated covariance matrix is beyond the scope of the present paper. We just observe that, apart from the above calculation, studying the sample properties of the truncated spectrum is made hard by the fact that eigenvalues and eigenvectors of a diagonalizable matrix are intimately related from their very definition; thus such study would require a careful consideration of the distribution of the sample diagonalizing matrix $\widehat{R} \neq R$ of $\widehat{\mathfrak{S}}_B$.

7.3. Perturbative Regularization of τ_ρ^{-1}. When μ is critically close to the internal boundary of $\mathscr{D}(\tau_\rho^{-1})$, a sample estimate $\widehat{\mu}$ may fall outside of it due to statistical fluctuations. In that case the iterative procedure described in the previous sections diverges. On the quantitative side, the ill-posedness of the reconstruction problem is measured by the failure probability

$$p_{\text{fail}}\left(\rho, \Sigma, N\right) = \mathbb{P}\left[\widehat{\mu} \notin \mathscr{D}\left(\tau_\rho^{-1}\right) \mid X^{(k)} \sim \mathcal{N}_\nu\left(0; \Sigma\right), \; k \right. \\ \left. = 1, \ldots, N\right], \tag{91}$$

which is a highly nontrivial function of ρ, Σ, and N. An illustrative example of it is reported in Figure 9(a), which refers to a specific case with $\nu = 4$ and $\Sigma = \text{diag}(0.1, 0.3, 0.8, 2.2)$. The plot suggests that the iterative procedure becomes severely ill-posed in the regime of strong truncation.

In order to regularize the problem, we propose to go back to (19) and consider it from a different perspective. Specifically, we move from the observation that a simplified framework occurs in the special circumstance when the eigenvalue spectra are fully degenerate, which is essentially equivalent to the setup of [1]. If $\mu_1 = \cdots = \mu_\nu \equiv \widetilde{\mu}$, by symmetry arguments it follows that $\lambda_1 = \cdots = \lambda_\nu \equiv \widetilde{\lambda}$ and the other way round. Equation (19) reduces in this limit to

$$\widetilde{\mu} = \widetilde{\lambda} \frac{F_{\nu+2}}{F_\nu}\left(\frac{\rho}{\widetilde{\lambda}}\right) \equiv \mathcal{T}_\rho\left(\widetilde{\lambda}\right). \tag{92}$$

It can be easily checked that the function $\mathcal{T}_\rho(\widetilde{\lambda})$ is monotonically increasing in $\widetilde{\lambda}$. In addition, we have

$$\begin{aligned} &\text{(i)} \quad \lim_{\widetilde{\lambda}\to 0} \mathcal{T}_\rho\left(\widetilde{\lambda}\right) = 0, \\ &\text{(ii)} \quad \lim_{\widetilde{\lambda}\to\infty} \mathcal{T}_\rho\left(\widetilde{\lambda}\right) = \frac{\rho}{\nu + 2}; \end{aligned} \tag{93}$$

thus (92) can be surely (numerically) inverted provided that $0 < \widetilde{\mu} < \rho/(\nu+2)$. We can regard (92) as an approximation to the original problem (19). When μ is not degenerate, we must define $\widetilde{\mu}$ in terms of the components of μ. One possibility is to average them, that is, to choose

$$\widetilde{\mu} = \frac{1}{\nu}\sum_{i=1}^{\nu} \mu_i. \tag{94}$$

Subject to this, we expect $\widetilde{\lambda}$ to lie somewhere between λ_1 and λ_ν. Equation (92) can be thought of as the lowest order approximation of a perturbative expansion of (19) around the point $\lambda_T = \{\widetilde{\lambda}, \ldots, \widetilde{\lambda}\}$. If the condition number of Σ is not extremely large, such an expansion is expected to quickly converge, so that a few perturbative corrections to λ_T should be sufficient to guarantee a good level of approximation.

As mentioned above, a technical implementation of the perturbative approach and a thorough discussion of its properties are deferred to a separate paper [27]. Here, we limit ourselves to observing that the definition domain of perturbation theory is ultimately set by its lowest order approximation, since corrections to (92) are all algebraically built in terms of it, with no additional constraints. Following (94), the domain of $\mathcal{T}_\rho^{-1}$ comes to be defined as

$$\mathscr{D}\left(\mathcal{T}_\rho^{-1}\right) = \left\{\mu \in \mathbb{R}_+^\nu : \sum_{i=1}^{\nu}\mu_i \leq \frac{\rho\nu}{\nu + 2}\right\}, \tag{95}$$

and it is clear that $\mathscr{D}(\tau_\rho^{-1}) \subset \mathscr{D}(\mathcal{T}_\rho^{-1})$ (it is sufficient to sum term by term all the inequalities contributing to (36)). In Figure 10, we show the set difference $\mathscr{D}(\mathcal{T}_\rho^{-1}) \setminus \mathscr{D}(\tau_\rho^{-1})$ in $\nu = 2$ and $\nu = 3$ dimensions. When $\mu \in \mathscr{D}(\tau_\rho^{-1})$ but its estimate $\widehat{\mu} \notin \mathscr{D}(\tau_\rho^{-1})$, it may well occur $\widehat{\mu} \in \mathscr{D}(\mathcal{T}_\rho^{-1})$; that is, the set difference acts as an absorbing shield of the statistical noise. Therefore, if we define the failure probability of the perturbative reconstruction as

$$q_{\text{fail}}\left(\rho, \Sigma, N\right) = \mathbb{P}\left[\widehat{\mu} \notin \mathscr{D}\left(\mathcal{T}_\rho^{-1}\right) \mid X^{(k)} \right. \\ \left. \sim \mathcal{N}_\nu\left(0; \Sigma\right), \; k = 1, \ldots, N\right], \tag{96}$$

we expect the inequality $q_{\text{fail}}(\rho, \Sigma, N) \ll p_{\text{fail}}(\rho, \Sigma, N)$ to generously hold. An example is given in Figure 9(b): we see that q_{fail} becomes lower than p_{fail} by orders of magnitude as soon as ρ and N are not exceedingly small. In this sense, the operator $\mathcal{T}_\rho^{-1}$ can be regarded as the lowest order approximation of a regularizing operator for τ_ρ^{-1}.

8. Conclusions

In this paper we have studied how to reconstruct the covariance matrix Σ of a normal multivariate $X \sim \mathcal{N}_\nu(0, \Sigma)$ from the matrix $\mathfrak{S}_{\mathscr{B}}$ of the spherically truncated second moments, describing the covariances among the components of X when the probability density is cut off outside a centered Euclidean ball. We have shown that Σ and $\mathfrak{S}_{\mathscr{B}}$ share the same eigenvectors. Therefore, the problem amounts to relating the eigenvalues of Σ to those of $\mathfrak{S}_{\mathscr{B}}$. Such relation entails the inversion of a system of nonlinear integral equations, which admits unfortunately no closed-form solution. Having found a necessary condition for the invertibility of the system, we have shown that the eigenvalue reconstruction can be achieved numerically via a converging fixed point iteration. In order to prove the convergence, we rely ultimately upon some probability inequalities, known in the literature as *square correlation inequalities*, which have been recently proved in [17].
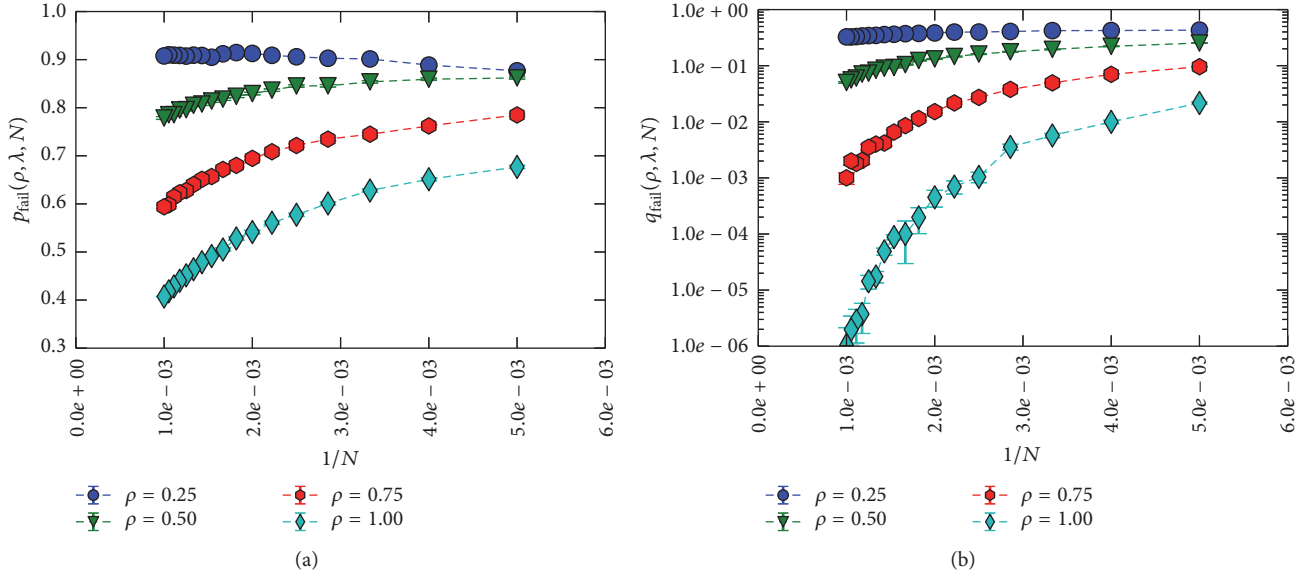
FIGURE 9: (a) Numerical reconstruction of the failure probability of the iterative procedure as $\nu = 4$ and $\Sigma = \text{diag}(0.1, 0.3, 0.8, 2.2)$, for several values of ρ and for $N = 200, 250, \dots, 1000$. (b) Failure probability of the perturbative regularization with the same parameters.
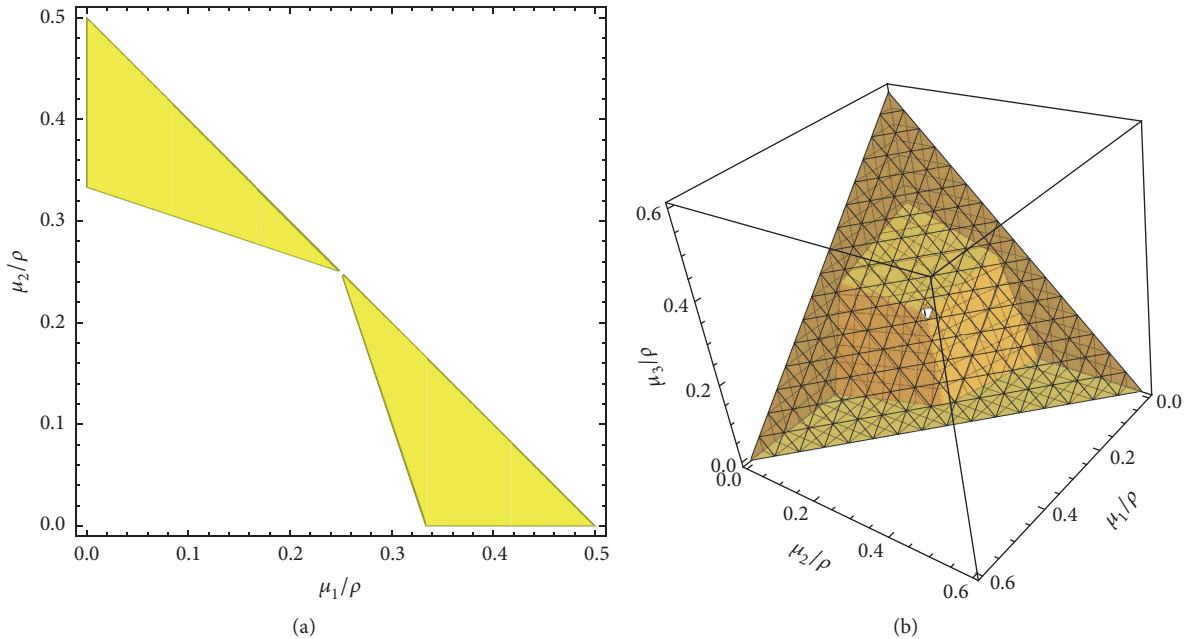

FIGURE 10: (a) Set difference $\mathscr{D}(\mathscr{T}_\rho^{-1}) \setminus \mathscr{D}(\tau_\rho^{-1})$ in $\nu = 2$ dimensions. (b) Set difference $\mathscr{D}(\mathscr{T}_\rho^{-1}) \setminus \mathscr{D}(\tau_\rho^{-1})$ in $\nu = 3$ dimensions.

In order to explore the convergence rate of the fixed point iteration, we have implemented some variations of the nonlinear Gauss–Jacobi scheme. Specifically, we have found that overrelaxing the basic iteration enhances the convergence rate by a moderate factor. However, the overrelaxed algorithm still slows down exponentially in the number of eigenvalues and polynomially in the truncation radius of the Euclidean ball. We have shown that a significant reduction

of the slowing down can be achieved in the regime of strong truncation by adapting the relaxation parameter to the eigenvalue that is naturally associated with, so as to boost the higher components of the spectrum.

We have also discussed how the iterative procedure works when the eigenvalue reconstruction is performed on sample estimates of the truncated covariance spectrum. Specifically, we have shown that the statistical fluctuations make the

problem ill-posed. We have sketched a possible way out based on perturbation theory, which is thoroughly discussed in a separate paper [27].

A concrete implementation of the proposed approach requires the computation of a set of multivariate Gaussian integrals over the Euclidean ball. For this, we have extended to the case of interest a technique, originally proposed by Ruben for representing the probability content of quadratic forms of normal variables as a series of chi-square distributions. In the paper, we have shown the practical feasibility of the series expansion for the integrals involved in our computations.

Appendix

Proof of Theorem 9. As already mentioned in Section 4, the proof follows in the tracks of the original one of [2]. We detail the relevant steps for α_k, while for α_{jk} we only explain why it is necessary to distinguish between equal or different indices and the consequences for either case.

In order to prove (57), we first express α_k in spherical coordinates; that is, we perform the change of variable $x = ru$, being $r = \|x\|$ and $u \in \partial \mathcal{B}_\nu(1)$ (recall that $d^\nu x = r^{\nu-1} dr\, du$, with du embodying the angular part of the spherical Jacobian and the differentials of $\nu - 1$ angles); then we insert a factor of $1 = \exp(r^2/2s) \exp(-r^2/2s)$ under the integral sign. Hence, α_k reads

$$
\begin{aligned}
\alpha_k(\rho;\lambda) = {} & \frac{1}{(2\pi)^{\nu/2} |\Lambda|^{1/2}} \int_0^{\sqrt{\rho}} dr\, r^{\nu-1} \frac{r^2}{\lambda_k} \exp\left(-\frac{r^2}{2s}\right) \\
& \cdot \int_{\partial \mathcal{B}_\nu(1)} du\, u_k^2 \exp\left(-\frac{Q(u)\, r^2}{2}\right).
\end{aligned} \tag{A.1}
$$

The next step consists in expanding the inner exponential in Taylor series (in his original proof, Ruben considers a more general setup, with the center of the Euclidean ball shifted by a vector $b \in \mathbb{R}^\nu$ from the center of the distribution. In that case, the Gaussian exponential looks different and must be expanded in series of Hermite polynomials. Here, we work in a simplified setup, where the Hermite expansion reduces to Taylor's); namely,

$$
\exp\left(-\frac{Q(u)\, r^2}{2}\right) = \sum_{m=0}^{\infty} \frac{1}{m!} \frac{r^{2m}}{2^m} (-Q)^m. \tag{A.2}
$$

This series converges uniformly in u. We review the estimate just for the sake of completeness:

$$
\left| \sum_{m=0}^{\infty} \frac{1}{m!} \frac{r^{2m}}{2^m} (-Q)^m \right| \leq \sum_{m=0}^{\infty} \frac{1}{m!} \frac{r^{2m}}{2^m} q_0^m = \exp\left(\frac{r^2 q_0}{2}\right), \tag{A.3}
$$

where $q_0 = \max_i |1/s - 1/\lambda_i|$. It follows that we can integrate the series term by term. With the help of the uniform average operator introduced in (53), α_k is recast to

$$
\begin{aligned}
& \alpha_k(\rho;\lambda) \\
&= \sum_{m=0}^{\infty} \frac{1}{m!} \frac{1}{\lambda_k} \frac{1}{|\Lambda|^{1/2}} \mathbb{M}\left[(-Q)^m u_k^2\right] \frac{1}{2^{\nu/2+m-1}\Gamma(\nu/2)} \\
& \quad \cdot \int_0^{\sqrt{\rho}} dr\, r^{\nu+2m+1} \exp\left(-\frac{r^2}{2s}\right).
\end{aligned} \tag{A.4}
$$

The presence of an additional factor of u_k^2 in the angular average is harmless, since $|u_k^2| < 1$. We finally notice that the radial integral can be expressed in terms of a cumulative chi-square distribution function on replacing $r \to \sqrt{rs}$; namely,

$$
\begin{aligned}
& \int_0^{\sqrt{\rho}} dr\, r^{\nu+2m+1} \exp\left(-\frac{r^2}{2s}\right) \\
&= 2^{\nu/2+m} s^{\nu/2+m+1} \Gamma\left(\frac{\nu}{2} + m + 1\right) F_{\nu+2(m+1)}\left(\frac{\rho}{s}\right).
\end{aligned} \tag{A.5}
$$

Inserting (A.5) into (A.4) results in Ruben's representation of α_k. This completes the first part of the proof.

As a next step, we wish to demonstrate that the function ψ_k of (61) is the generating function of the coefficients $c_{k;m}$. To this aim, we first recall the identities

$$
\begin{aligned}
a^{-1/2} &= (2\pi)^{-1/2} \int_{-\infty}^{\infty} dx \exp\left(-\frac{a}{2}x^2\right), \\
a^{-3/2} &= (2\pi)^{-1/2} \int_{-\infty}^{\infty} dx\, x^2 \exp\left(-\frac{a}{2}x^2\right),
\end{aligned} \tag{A.6}
$$

valid for $a > 0$. On setting $a_i = [1 - (1 - s/\lambda_i)z]$, we see that ψ_k can be represented in the integral form

$$
\begin{aligned}
\psi_k(z) = {} & \left(\frac{s}{\lambda_k}\right)^{3/2} (2\pi)^{-\nu/2} \int_{-\infty}^{\infty} dx_k\, x_k^2 \\
& \cdot \exp\left(-\frac{1}{2}\left[1 - \left(1 - \frac{s}{\lambda_k}\right)z\right]x_k^2\right) \times \prod_{i \neq k}\left(\frac{s}{\lambda_i}\right)^{1/2} \\
& \cdot \int_{-\infty}^{\infty} dx_i \exp\left(-\frac{1}{2}\left[1 - \left(1 - \frac{s}{\lambda_k}\right)z\right]x_i^2\right) = \frac{s}{\lambda_k} \\
& \cdot \frac{s^{\nu/2}}{(2\pi)^{\nu/2} |\Lambda|^{1/2}} \int_{\mathbb{R}^\nu} d^\nu x\, x_k^2 \\
& \cdot \exp\left(-\frac{1}{2}zsQ(x) - \frac{x^{\mathrm{T}} \cdot x}{2}\right),
\end{aligned} \tag{A.7}
$$

provided $|z| < \min_i |1 - s/\lambda_i|^{-1}$. As previously done, we introduce spherical coordinates $x = ru$ and expand $\exp\{-(1/2)zsQ(x)\} = \exp\{-(1/2)zsr^2Q(u)\}$ in Taylor series.

By the same argument as above, the series converges uniformly in u (the factor of zs does not depend on u), thus allowing term-by-term integration. Accordingly, we have

$$\psi_k(z) = \frac{s}{\lambda_k} \frac{s^{v/2}}{(2\pi)^{v/2} |\Lambda|^{1/2}} \sum_{m=0}^{\infty} z^m \frac{s^m}{2^m m!}$$

$$\cdot \int_0^{\infty} dr\, r^{v+2(m+1)-1} e^{-r^2/2} \qquad (A.8)$$

$$\cdot \int_{\partial \mathscr{B}_v(1)} du\, [-Q(u)]^m u_k^2.$$

We see that the *right-hand side* of (A.8) looks similar to (A.4), the only relevant differences being the presence of the factor of z^m under the sum sign and the upper limit of the radial integral. With some algebra, we arrive at

$$\psi_k(z) = \sum_{m=0}^{\infty} z^m \left\{ \frac{2}{m!} \frac{s}{\lambda_k} \frac{s^{v/2+m}}{|\Lambda|^{1/2}} \frac{\Gamma(v/2+m+1)}{\Gamma(v/2)} \right.$$

$$\left. \cdot \mathbb{M}\left[(-Q)^m u_k^2\right] \right\}. \qquad (A.9)$$

The series coefficients are recognized to be precisely those of (59).

In the last part of the proof, we derive the recursion fulfilled by the coefficients $c_{k;m}$. To this aim, the mth derivative of ψ_k has to be evaluated at $z = 0$ and then identified with $m! c_{k;m}$. The key observation is that differentiating ψ_k reproduces ψ_k itself; that is to say,

$$\psi'_k(z) = \Psi_k(z)\, \psi_k(z), \qquad (A.10)$$

with

$$\Psi_k(z) = \frac{1}{2} \sum_{i=1}^{v} e_{k;i} \left(1 - \frac{s}{\lambda_i}\right) \left[1 - \left(1 - \frac{s}{\lambda_i}\right) z\right]^{-1}, \qquad (A.11)$$

and the auxiliary coefficient $e_{k;i}$ being defined as in (67). Equation (A.10) lies at the origin of the recursion. Indeed, from (A.10), it follows that that ψ''_k is a function of ψ'_k and ψ_k; namely, $\psi''_k = \Psi'_k \psi_k + \Psi_k \psi'_k$. Proceeding analogously yields the general formula

$$\psi_k^{(m)}(z) = \sum_{r=0}^{m-1} \binom{m-1}{r} \Psi_k^{(m-r-1)}(z)\, \psi_k^{(r)}(z). \qquad (A.12)$$

At $z = 0$, this reads

$$m! c_{k;m} = \sum_{r=0}^{m-1} \frac{(m-1)!}{(m-r-1)! r!} \Psi_k^{(m-r-1)}(0)\, r! c_{k;r}. \qquad (A.13)$$

The last step consists in proving that

$$\Psi_k^{(m)}(0) = \frac{1}{2} m!\, g_{k;m+1}, \qquad (A.14)$$

with $g_{k;m}$ defined as in (65). This can be done precisely as explained in [2].

Having reiterated Ruben's proof explicitly in a specific case, it is now easy to see how the theorem is extended to any other Gaussian integral. First of all, from (A.1), we infer that each additional subscript in $\alpha_{k\ell m\cdots}$ enhances the power of the radial coordinate under the integral sign by 2 units. This entails a shift in the number of degrees of freedom of the chi-square distributions in Ruben's expansion, amounting to twice the number of subscripts. For instance, since α_{jk} has two subscripts, its Ruben's expansion starts by F_{v+4}, independently of whether $j = k$ or $j \neq k$. In second place, we observe that in order to correctly identify the generating functions of Ruben's coefficients for a higher-order integral $\alpha_{k\ell m\cdots}$, we need to take into account the multiplicities of the indices k, ℓ, m,.... As an example, consider the case of ψ_{jk} ($j \neq k$) and ψ_{kk}. By going once more through the argument presented in (A.7), we see that (A.6) are sufficient to show that (63) is the generating function of α_{jk}. By contrast, in order to repeat the proof for the case of ψ_{kk}, we need an additional integral identity; namely,

$$a^{-5/2} = \frac{1}{3} (2\pi)^{-1/2} \int_{-\infty}^{+\infty} dx\, x^4 \exp\left(-\frac{a}{2} x^2\right), \qquad (A.15)$$

valid once more for $a > 0$. Hence, we infer that ψ_{kk} must depend upon λ_k via a factor of $[1 - (1 - s/\lambda_k)z]^{-5/2}$, whereas ψ_{jk} ($j \neq k$) must depend on λ_j and λ_k via factors of, respectively, $[1 - (1 - s/\lambda_j)z]^{-3/2}$ and $[1 - (1 - s/\lambda_k)z]^{-3/2}$. The different exponents are ultimately responsible for the specific values taken by the auxiliary coefficients $e_{kk;i}$ and $e_{jk;i}$ of (68).

To conclude, we observe that the estimates of the residuals $\mathscr{R}_{k;m}$ and $\mathscr{R}_{jk;m}$, presented in Section 4 without an explicit proof, do not require any further technical insight than already provided by [2] plus our considerations. We leave them to the reader, since they can be obtained once more in the tracks of the original derivation of $\mathscr{R}_m$. $\qquad \square$

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

References

[1] G. M. Tallis, "Elliptical and radial truncation in normal populations," *The Annals of Mathematical Statistics*, vol. 34, no. 3, pp. 940–944, 1963.

[2] H. Ruben, "Probability content of regions under spherical normal distributions. IV. The distribution of homogeneous and non-homogeneous quadratic functions of normal variables," *The Annals of Mathematical Statistics*, vol. 33, pp. 542–570, 1962.

[3] J. Aitchison, *The Statistical Analysis of Compositional Data*, Monographs on Statistics and Applied Probability, Chapman and Hall, London, UK, 1986.

[4] V. Pawlowsky-Glahn and A. Buccianti, *Compositional Data Analysis: Theory and Applications*, John Wiley & Sons, 2011.

[5] J. J. Egozcue, V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barceló-Vidal, "Isometric logratio transformations for compositional data analysis," *Mathematical Geology*, vol. 35, no. 3, pp. 279–300, 2003.

[6] J. Aitchison and S. M. Shen, "Logistic-normal distributions: some properties and uses," *Biometrika*, vol. 67, no. 2, pp. 261–272, 1980.

[7] V. Pawlowsky-Glahn and J. J. Egozcue, "Geometric approach to statistical analysis on the simplex," *Stochastic Environmental Research and Risk Assessment*, vol. 15, no. 5, pp. 384–398, 2001.

[8] P. Filzmoser and K. Hron, "Outlier detection for compositional data using robust methods," *Mathematical Geosciences*, vol. 40, no. 3, pp. 233–248, 2008.

[9] A. C. Atkinson, "Fast very robust methods for the detection of multiple outliers," *Journal of the American Statistical Association*, vol. 89, no. 428, pp. 1329–1339, 1994.

[10] M. Riani, A. C. Atkinson, and A. Cerioli, "Finding an unknown number of multivariate outliers," *Journal of the Royal Statistical Society B*, vol. 71, part 2, pp. 447–466, 2009.

[11] W. C. Guenther, "An easy method for obtaining percentage points of order statistics," *Technometrics*, vol. 19, no. 3, pp. 319–321, 1977.

[12] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems. Mathematics and Its Applications*, vol. 375, Springer, Berlin, Germany, 1996.

[13] L. Cavalier, "Inverse problems in statistics," in *Inverse Problems and High-Dimensional Estimation: Stats in the Château Summer School, August 31–September 4, 2009*, P. Alquier, E. Gautier, and G. Stoltz, Eds., vol. 203 of *Lecture Notes in Statistics*, pp. 3–96, Springer, Berlin, Germany, 2011.

[14] F. Palombi and S. Toti, "A note on the variance of the square components of a normal multivariate within a Euclidean ball," *Journal of Multivariate Analysis*, vol. 122, pp. 355–376, 2013.

[15] M. Anttila, K. Ball, and I. Perissinaki, "The central limit problem for convex bodies," *Transactions of the American Mathematical Society*, vol. 355, no. 12, pp. 4723–4735, 2003.

[16] J. O. Wojtaszczyk, "The square negative correlation property for generalized Orlicz balls," in *Proceedings of the Geometric Aspects of Functional Analysis, Israel Seminar*, pp. 305–313, 2004-2005.

[17] R. Mukerjee and S. H. Ong, "Variance and covariance inequalities for truncated joint normal distribution via monotone likelihood ratio and log-concavity," *Journal of Multivariate Analysis*, vol. 139, pp. 1–6, 2015.

[18] F. G. Friedlander and M. S. Joshi, *Introduction to the Theory of Distributions*, Cambridge University Press, Cambridge, UK, 2nd edition, 1998.

[19] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover Publications, New York, 1964., New York, NY, USA, 1964.

[20] K. L. Judd, *Numerical Methods in Economics*, The MIT Press, Cambridge, Mass, USA, 1998.

[21] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, New York, NY, USA, 2nd edition, 1984.

[22] M. S. Bartlett, "XX.—on the theory of statistical regression," *Proceedings of the Royal Society of Edinburgh*, vol. 53, pp. 260–283, 1934.

[23] U. Grenander, "Some direct estimates of the mode," *The Annals of Mathematical Statistics*, vol. 36, no. 1, pp. 131–138, 1965.

[24] A. Zanella, M. Chiani, and M. Z. Win, "On the marginal distribution of the eigenvalues of wishart matrices," *IEEE Transactions on Communications*, vol. 57, no. 4, pp. 1050–1060, 2009.

[25] D. M. Young, *Iterative Solution of Large Linear Systems*, Computer Science and Applied Mathematics, Academic Press, 1971.

[26] J. Hadamard, "Sur les problèmes aux dérivés partielles et leur signification physique," *Princeton University Bulletin*, vol. 13, pp. 49–52, 1902.

[27] F. Palombi and S. Toti, "A perturbative approach to the reconstruction of the eigenvalue spectrum of a normal covariance matrix from a spherically truncated counterpart," https://arxiv.org/abs/1207.1256.

[28] G. Ponti, F. Palombi, D. Abate et al., "The role of medium size facilities in the HPC ecosystem: the case of the new CRESCO4 cluster integrated in the ENEAGRID infrastructure," in *Proceedings of the International Conference on High Performance Computing and Simulation (HPCS '14)*, 6903807, pp. 1030–1033, July 2014.