Journal of Multivariate Analysis 198 (2023) 105227

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Exact first moments of the RV coefficient by invariant orthogonal integration

François Bavaud

University of Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 16 October 2022
Received in revised form 19 July 2023
Accepted 19 July 2023
Available online 23 July 2023

AMS 2020 subject classifications:
primary 62H15
secondary 62H20
62G10

Keywords:
Invariant orthogonal integration
RV coefficient
Spectral moments
Weighted multidimensional scaling
Weingarten calculus

ABSTRACT

The RV coefficient measures the similarity between two multivariate configurations, and its significance testing has attracted various proposals in the last decades. We present a new approach, the invariant orthogonal integration, permitting to obtain the exact first four moments of the RV coefficient under the null hypothesis.

Our proposal can be applied to any multivariate setting endowed with Euclidean distances between the observations. It also covers the weighted setting of observations of unequal importance, where the exchangeability assumption, justifying the usual permutation tests, breaks down.

The proposed RV moments express as simple functions of the kernel eigenvalues occurring in the weighted multidimensional scaling of the two configurations (spectral effective dimensionality, spectral skewness and spectral excess kurtosis). The expressions for the third and fourth moments seem original, and explain the marked asymmetry and kurtosis of the RV coefficient. They permit to test the significance of the RV coefficient by Cornish–Fisher cumulant expansion, beyond the normal approximation, as illustrated on a small dataset.

The first three moments can be obtained by elementary means, but computing the fourth moment requires a more sophisticated apparatus, the Weingarten calculus for orthogonal groups.

1. Introduction

The RV coefficient, defined in (6) and (7), is a well-known measure of similarity between two datasets, each consisting of multivariate profiles measured on the same n observations or objects. This contribution proposes a new approach, the invariant orthogonal integration, permitting to obtain the exact first four moments of the RV coefficient under the null hypothesis of absence of relation between the two datasets. The main results, Theorem 1 and Corollary 1, are exposed in Section 3.1. The approach is fully nonparametric, and allows the handling of weighted objects, typically made of aggregates such as regions, documents or species, which abound in multivariate analysis.

In the present distance-based data-analytic approach, data sets are constituted by weighted configurations specified by the object weights together with their pair dissimilarities, assumed to be squared Euclidean. Factorial coordinates, reproducing the dissimilarities, and permitting a maximum compression of the configuration inertia, obtain by weighted multidimensional scaling. The latter, seldom exposed in the literature (see however [5,15] and references therein) and hence briefly recalled in Section 2.1, is a direct generalization of Torgerson–Gower classical scaling. The central step is provided by the spectral decomposition of the matrix of weighted centered scalar products or kernel. It permits to

E-mail address: fbavaud@unil.ch.

https://doi.org/10.1016/j.jmva.2023.105227

decompose the spectral eigenspace into a trivial one-dimensional part, determined by the object weights, common to both configurations, and a non-trivial part of dimension $n-1$, orthogonal to the square root of the weights. The weighted RV coefficient obtains as the normalized scalar product between the kernels of the two configurations (Section 2.2), and turns out to be equivalent to its original definition expressed by cross-covariances [17,32].

After recalling the above preliminaries, somewhat lengthy but necessary, the heart of this contribution can be uncovered: invariant orthogonal integration consists in computing the expected null moments of the RV coefficient by averaging, along the invariant Haar orthogonal measure in the non-trivial eigenspace, the orientations of the eigenvectors of the first configuration with respect to the second one (Section 3.2). It constitutes a distinct alternative, with different outcomes, to the traditional permutation approach, whose exchangeability assumption breaks down for weighted objects: typically, the profile dispersion is expected to be larger for lighter objects [6] and the n object scores cannot follow the same distribution. The present approach also yields a novel significance test for the RV coefficient (Eq. (18)), taking into account skewness and kurtosis corrections to the usual normal approximation.

Computing the moments of the RV coefficient requires to evaluate the orthogonal coefficients (26) constituted by Haar expectations of orthogonal monomials. Low-order moments can be computed, with increasing difficulty, by elementary means (Section 3.3), but the fourth-order moment requires a more systematic approach (Section 3.6), provided by the Weingarten calculus developed by workers in random matrix theory and free probability. Both procedures yield the same results for low-order moments (Section 3.7), which is both expected and reassuring.

The first RV moment (12) coincides with all known proposals. The second centered RV moment (13) is simpler than its permutation analog. Both expressions emphasize the effective (spectral) dimensionality of a configuration. The third centered RV moment (14) is particularly enlightening: the RV skewness is simply proportional to the product of the spectral skewness of both configurations, thus elucidating the often noticed positive skewness of the RV coefficient. The expression for the fourth centered RV moment (15) reveals an explicit relation, yet difficult to interpret, between the RV excess kurtosis and the spectral excess kurtosis of both configurations.

A small case study (Section 4) illustrates the theory, and demonstrates the broad applicability of the approach, apt to compare any two multivariate configurations (weighted or not, of numerical or categorical origin), provided that object weights coincide for the two configurations, and that object dissimilarities are squared Euclidean for each configuration.

2. Euclidean configurations in a weighted setting: a concise remainder

2.1. Weighted multidimensional scaling and standard kernels

Consider n objects endowed with positive weights $f_i > 0$ with $\sum_{i=1}^n f_i = 1$, as well with pairwise dissimilarities $\mathbf{D} = (D_{ij})$ between pairs of objects. The $n\times n$ matrix $\mathbf{D}$ is assumed to be squared Euclidean, that is of the form $D_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|^2$ for $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^r$, with $r \leq n-1$. The pair $(\mathbf{f}, \mathbf{D})$ constitutes a weighted configuration, with $f_i = 1/n$ for unweighted configurations.

Weighted multidimensional scaling (where "weighted" here refers to objects and not to features), aims at determining object coordinates $\mathbf{X} = (x_{i\alpha}) \in \mathbb{R}^{n\times r}$ reproducing the dissimilarities $\mathbf{D}$ while expressing a maximum amount of dispersion or inertia Δ (5) in low dimensions. It is performed by the following weighted generalization of the well-known Torgerson–Gower scaling procedure [see, e.g., 8]: first, define the diagonal matrix $\boldsymbol{\Pi} = \mathrm{diag}(\mathbf{f})$, as well as the weighted centering matrix $\mathbf{H} = \mathbf{I}_n - \mathbf{1}_n\mathbf{f}^\top$, transforming $\mathbf{x} \in \mathbb{R}^n$ as $\mathbf{H}\mathbf{x} = \mathbf{x} - \bar{x}\mathbf{1}_n = \mathbf{x}_c$ where $\bar{x} = \mathbf{f}^\top\mathbf{x}$. The centering matrix is a projection ($\mathbf{H}^2 = \mathbf{H}$) whose null space is the set of constant vectors. Also, $\mathbf{H}^\top \neq \mathbf{H}$, unless $\mathbf{f}$ is uniform.

Second, compute the symmetric matrix $\mathbf{B}$ of scalar products by double centering: $\mathbf{B} = -\frac{1}{2}\mathbf{H}\,\mathbf{D}\,\mathbf{H}^\top$. Third, define the $n\times n$ kernel $\mathbf{K}$ as the symmetric matrix of weighted scalar products:

$$\mathbf{K} = \sqrt{\boldsymbol{\Pi}}\,\mathbf{B}\sqrt{\boldsymbol{\Pi}}\ , \quad K_{ij} = \sqrt{f_i f_j}B_{ij}\ . \tag{1}$$

Fourth, perform the kernel spectral decomposition with $\hat{\mathbf{U}} \in \mathbb{R}^{n\times n}$ orthogonal and $\hat{\boldsymbol{\Lambda}}$ diagonal

$$\mathbf{K} = \hat{\mathbf{U}}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{U}}^\top\ , \quad \hat{\mathbf{U}}\hat{\mathbf{U}}^\top = \hat{\mathbf{U}}^\top\hat{\mathbf{U}} = \mathbf{I}_n\ , \quad \hat{\boldsymbol{\Lambda}} = \mathrm{diag}(\boldsymbol{\lambda})\ . \tag{2}$$

By construction, $\mathbf{K}$ possesses one trivial eigenvalue $\lambda_0 = 0$ associated to the eigenvector $\sqrt{\mathbf{f}}$ (since $\mathbf{H}^\top\sqrt{\boldsymbol{\Pi}}\sqrt{\mathbf{f}} = \mathbf{H}^\top\mathbf{f} = \mathbf{0}_n$) and $n-1$ real non-negative eigenvalues decreasingly ordered as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n-1} \geq 0$, among which $r = \mathrm{rank}(\mathbf{K})$ are strictly positive.

From now on the trivial eigenspace will be discarded: set $\hat{\mathbf{U}} = (\sqrt{\mathbf{f}}\,|\,\mathbf{U})$, where $\mathbf{U} = (\mathbf{u}_1|\ldots|\mathbf{u}_{n-1}) \in \mathbb{R}^{n\times(n-1)}$ contains the $n-1$ non-trivial eigenvectors of $\mathbf{K}$, and set $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1,\ldots,\lambda_{n-1})$. Direct substitution from (2) yields

$$\mathbf{K} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top\ , \quad \mathbf{U}\mathbf{U}^\top = \mathbf{I}_n - \sqrt{\mathbf{f}}\sqrt{\mathbf{f}}^\top\ , \quad \mathbf{U}^\top\mathbf{U} = \mathbf{I}_{n-1}\ , \quad \mathbf{U}^\top\sqrt{\mathbf{f}} = \mathbf{0}_n\ . \tag{3}$$

Finally, the searched for coordinates obtain as

$$\mathbf{X} = \boldsymbol{\Pi}^{-\frac{1}{2}}\mathbf{U}\boldsymbol{\Lambda}^{\frac{1}{2}}\ , \quad x_{i\alpha} = \frac{1}{\sqrt{f_i}}\,u_{i\alpha}\sqrt{\lambda_\alpha}\ . \tag{4}$$

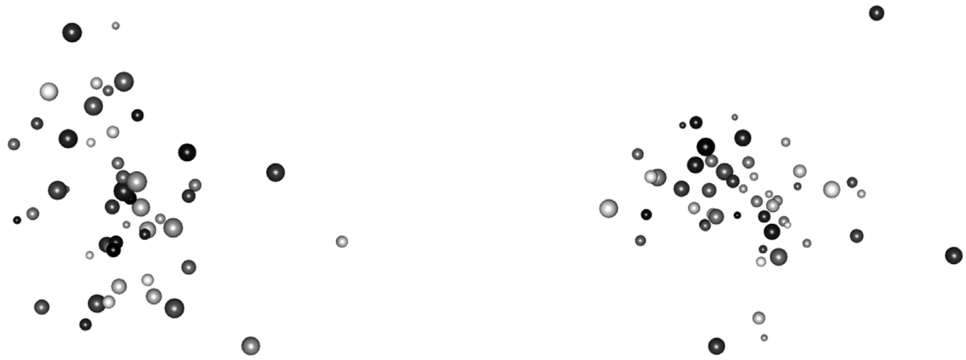

Fig. 1. Two weighted Euclidean configurations $(\mathbf{f}, \mathbf{D}_X)$ (left) and $(\mathbf{f}, \mathbf{D}_Y)$ (right), made of n weighted multivariate observations embedded in $\mathbb{R}^{n-1}$. Their similarity attains the maximum value $\mathrm{RV}_{XY} = 1$ iff $\mathbf{D}_Y = c\mathbf{D}_X$ for some $c > 0$, i.e. iff the two configurations are isometric up to a dilatation.

One verifies easily that

$$D_{ij} = \sum_{\alpha=1}^{n-1}(x_{i\alpha} - x_{j\alpha})^2 \,, \qquad \Delta = \frac{1}{2}\sum_{i,j=1}^{n} f_i f_j D_{ij} = \mathrm{Tr}(\mathbf{K}) = \sum_{\alpha=1}^{n-1}\lambda_\alpha \ . \tag{5}$$

The kernels considered here are symmetric, positive semi-definite and obey in addition $\mathbf{K}\sqrt{\mathbf{f}} = \mathbf{0}_n$. We call them standard kernels. They can be related to the weighted version of centered kernels of Machine Learning [see, e.g., 13]. To each weighted configuration $(\mathbf{f}, \mathbf{D})$ corresponds a unique standard kernel $\mathbf{K}$, and conversely.

The matrix $\mathbf{K}_0 = \mathbf{I}_n - \sqrt{\mathbf{f}}\sqrt{\mathbf{f}}^\top$ appearing in (3) constitutes a standard kernel, referred to as the neutral kernel in view of property $\mathbf{K}_0\mathbf{K} = \mathbf{K}\mathbf{K}_0 = \mathbf{K}$ for any standard kernel $\mathbf{K}$. The corresponding dissimilarities are the weighted discrete distances

$$D_{ij}^0 = \begin{cases} \frac{1}{f_i} + \frac{1}{f_j} \,, & \text{for } i \neq j \,, \\ 0 \,, & \text{otherwise.} \end{cases}$$

A note on unweighted multidimensional scaling: the latter implements the classical Torgerson–Gower scaling procedure, which does not take into account the objects weights, when non-uniform. That is, spectral decomposition is directly performed on the matrix of scalar products $\mathbf{B} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$ (instead on $\mathbf{K}$ in (1)), and factor $1/\sqrt{f_i}$ is absent in the expression for coordinates (4). The resulting dissimilarities $\mathbf{D}$ are exactly the same as those obtained by weighted multidimensional scaling, but the weighted inertia Δ in (5) is no more optimally expressed by the first factors.

2.2. The RV coefficient

Consider two weighted configurations $(\mathbf{f}, \mathbf{D}_X)$ and $(\mathbf{f}, \mathbf{D}_Y)$ endowed with the same weights $\mathbf{f}$, or equivalently two standard kernels $\mathbf{K}_X$ and $\mathbf{K}_Y$ (Fig. 1). Their similarity can be measured by the weighted RV coefficient defined as

$$\mathrm{RV} = \mathrm{RV}_{XY} = \frac{\mathrm{Tr}(\mathbf{K}_X\ \mathbf{K}_Y)}{\sqrt{\mathrm{Tr}(\mathbf{K}_X^2)\,\mathrm{Tr}(\mathbf{K}_Y^2)}} \tag{6}$$

which constitutes the cosine similarity between the vectorized matrices $\mathbf{K}_X$ and $\mathbf{K}_Y$. As a consequence, $\mathrm{RV}_{XY} \geq 0$ (since $\mathbf{K}_X$ and $\mathbf{K}_Y$ are positive semi-definite), $\mathrm{RV}_{XY} \leq 1$ (by the Cauchy–Schwarz inequality) and $\mathrm{RV}_{XX} = 1$.

Quantity (6) is a straightforward weighted generalization of the RV coefficient introduced in [17,32] (where R did refer to "correlation" and V to "vector"): consider multivariate features $\mathbf{X} \in \mathbb{R}^{n\times p}$ and $\mathbf{Y} \in \mathbb{R}^{n\times q}$, directly entering into the definition of $\mathbf{D}_X$ and $\mathbf{D}_Y$ as coordinates, or equivalently as $\mathbf{K}_X = \sqrt{\boldsymbol{\Pi}}\mathbf{X}_c\mathbf{X}_c^\top\sqrt{\boldsymbol{\Pi}}$ and $\mathbf{K}_Y = \sqrt{\boldsymbol{\Pi}}\mathbf{Y}_c\mathbf{Y}_c^\top\sqrt{\boldsymbol{\Pi}}$, where $\mathbf{X}_c = \mathbf{H}\mathbf{X}$ and $\mathbf{Y}_c = \mathbf{H}\mathbf{Y}$ are the centered scores.

The weighted covariances are $\boldsymbol{\Sigma}_{XX} = \mathbf{X}_c^\top\boldsymbol{\Pi}\mathbf{X}_c$ and $\boldsymbol{\Sigma}_{YY} = \mathbf{Y}_c^\top\boldsymbol{\Pi}\mathbf{Y}_c$. The cross-covariances are $\boldsymbol{\Sigma}_{XY} = \mathbf{X}_c^\top\boldsymbol{\Pi}\mathbf{Y}_c$ and $\boldsymbol{\Sigma}_{YX} = \mathbf{Y}_c^\top\boldsymbol{\Pi}\mathbf{X}_c = \boldsymbol{\Sigma}_{XY}^\top$. The original RV coefficient is defined in the feature space as

$$\mathrm{RV}_{XY} = \frac{\mathrm{Tr}(\boldsymbol{\Sigma}_{XY}\ \boldsymbol{\Sigma}_{YX})}{\sqrt{\mathrm{Tr}(\boldsymbol{\Sigma}_{XX}^2)\,\mathrm{Tr}(\boldsymbol{\Sigma}_{YY}^2)}} \ . \tag{7}$$

Proving the equivalence of (6) and (7) is straightforward.

3. Computing the moments of the RV coefficient by invariant orthogonal integration

3.1. Main result and significance testing

Define the CV coefficient by the quantity $\mathtt{CV} = \mathrm{Tr}(\mathbf{K}_X\mathbf{K}_Y)$.

Theorem 1 (*Main Result*). *Under invariant orthogonal integration (Section 3.2), the expectation of the CV coefficient and its centered moments of order 2, 3 and 4 are*

$$\mathbb{E}(\mathtt{CV}) = (n-1)\,\overline{\boldsymbol{\lambda}}\,\overline{\boldsymbol{\mu}}\,, \qquad \mathbb{E}(\mathtt{CV}_c^2) = \frac{2(n-1)^2}{(n-2)(n+1)}\,\overline{\boldsymbol{\lambda}_c^2}\,\overline{\boldsymbol{\mu}_c^2}\,, \qquad \mathbb{E}(\mathtt{CV}_c^3) = \frac{8(n-1)^3}{(n-3)(n-2)(n+1)(n+3)}\,\overline{\boldsymbol{\lambda}_c^3}\,\overline{\boldsymbol{\mu}_c^3} \tag{8}$$

$$\mathbb{E}(\mathtt{CV}_c^4) = \frac{12\,(n-1)^3}{(n-4)(n-3)(n-2)n(n+1)(n+3)(n+5)}\Big\{4\,(n^2-n+2)\,\overline{\boldsymbol{\lambda}_c^4}\,\overline{\boldsymbol{\mu}_c^4} + (n^4+n^3-15n^2-13n + 98)\,\overline{\boldsymbol{\lambda}_c^2}^2\,\overline{\boldsymbol{\mu}_c^2}^2 - 4\,(2n^2-n-7)\,(\overline{\boldsymbol{\lambda}_c^4}\,\overline{\boldsymbol{\mu}_c^2}^2 + \overline{\boldsymbol{\lambda}_c^2}^2\,\overline{\boldsymbol{\mu}_c^4})\Big\}\,. \tag{9}$$

where $\mathtt{CV}_c = \mathtt{CV} - \mathbb{E}(\mathtt{CV})$. In the above, spectral moments and centered spectral moments of order q read

$$\overline{\boldsymbol{\lambda}^q} = \frac{1}{n-1}\sum_{\alpha=1}^{n-1}\lambda_\alpha^q = \frac{1}{n-1}\mathrm{Tr}(\mathbf{K}_X^q) = \mathrm{tr}(\mathbf{K}_X^q)\,, \qquad \overline{\boldsymbol{\lambda}_c^q} = \frac{1}{n-1}\sum_{\alpha=1}^{n-1}(\lambda_\alpha^c)^q\,, \qquad \lambda_\alpha^c = \lambda_\alpha - \overline{\boldsymbol{\lambda}} \tag{10}$$

where $\mathrm{tr}(\mathbf{A}) = \mathrm{Tr}(\mathbf{A})/(n-1)$ denotes the normalized trace. Centered spectral moments can be transformed into normalized traces, and conversely. For instance, $\overline{\boldsymbol{\lambda}_c^3} = \mathrm{tr}(\mathbf{K}_X^3) - 3\,\mathrm{tr}(\mathbf{K}_X^2)\,\mathrm{tr}(\mathbf{K}_X) + 2\,\mathrm{tr}^3(\mathbf{K}_X)$ (see also (43)). The identity

$$\mathtt{RV} = \frac{\mathtt{CV}}{\sqrt{\mathrm{Tr}(\mathbf{K}_X^2)\,\mathrm{Tr}(\mathbf{K}_Y^2)}} = \frac{\mathtt{CV}}{(n-1)\sqrt{\overline{\boldsymbol{\lambda}^2}\;\overline{\boldsymbol{\mu}^2}}} \tag{11}$$

directly implies:

Corollary 1 (*First Cumulants of the RV Coefficient*). *Under invariant orthogonal integration, the first cumulants of the RV coefficient, that is its expectation, variance, skewness and excess kurtosis are, in order,*

$$\mathbb{E}(\mathtt{RV}) = \frac{1}{n-1}\,\frac{Tr(\mathbf{K}_X)\,Tr(\mathbf{K}_Y)}{\sqrt{Tr(\mathbf{K}_X^2)\,Tr(\mathbf{K}_Y^2)}} = \frac{\overline{\boldsymbol{\lambda}}\,\overline{\boldsymbol{\mu}}}{\sqrt{\overline{\boldsymbol{\lambda}^2}\;\overline{\boldsymbol{\mu}^2}}} = \frac{\sqrt{\nu(\boldsymbol{\lambda})\,\nu(\boldsymbol{\mu})}}{n-1}\,, \tag{12}$$

$$\mathbb{V}\mathrm{ar}(\mathtt{RV}) = \mathbb{E}(\mathtt{RV}_c^2) = \frac{2\,(n-1-\nu(\boldsymbol{\lambda}))\,(n-1-\nu(\boldsymbol{\mu}))}{(n-2)(n-1)^2(n+1)}\,, \tag{13}$$

$$\mathbb{A}(\mathtt{RV}) = \frac{\mathbb{E}(\mathtt{RV}_c^3)}{\mathbb{E}^{\frac{3}{2}}(\mathtt{RV}_c^2)} = \frac{\sqrt{8(n-2)(n+1)}}{(n-3)(n+3)}\,a(\boldsymbol{\lambda})\,a(\boldsymbol{\mu})\,, \tag{14}$$

$$\mathbb{\Gamma}(\mathtt{RV}) = \frac{\mathbb{E}(\mathtt{RV}_c^4)}{\mathbb{E}^2(\mathtt{RV}_c^2)} - 3 = \frac{3\,(n-2)(n+1)}{(n-4)(n-3)(n-1)n(n+3)(n+5)}\Big\{4(n^2-n+2)\,\gamma(\boldsymbol{\lambda})\,\gamma(\boldsymbol{\mu}) + (4n^2-8n + 52)(\gamma(\boldsymbol{\lambda})+\gamma(\boldsymbol{\mu})) - \frac{4(5n^3-57n^2+27n+169)}{(n-2)(n+1)}\Big\}\,, \tag{15}$$

where $\mathtt{RV}_c = \mathtt{RV} - \mathbb{E}(\mathtt{RV})$. Here, the quantities

$$a(\boldsymbol{\lambda}) = \frac{\overline{\boldsymbol{\lambda}_c^3}}{(\overline{\boldsymbol{\lambda}_c^2})^{\frac{3}{2}}}\,, \qquad \gamma(\boldsymbol{\lambda}) = \frac{\overline{\boldsymbol{\lambda}_c^4}}{(\overline{\boldsymbol{\lambda}_c^2})^2} - 3 \tag{16}$$

denote the spectral skewness, respectively the excess spectral kurtosis. The quantity

$$\nu(\boldsymbol{\lambda}) = \frac{\mathrm{Tr}^2(\mathbf{K}_X)}{\mathrm{Tr}(\mathbf{K}_X^2)} = \frac{(\sum_{\alpha\ge 1}\lambda_\alpha)^2}{\sum_{\alpha\ge 1}\lambda_\alpha^2} = (n-1)\frac{\overline{\boldsymbol{\lambda}}^2}{\overline{\boldsymbol{\lambda}^2}} \tag{17}$$

has appeared at times as an adjusted degrees of freedom in multivariate tests of the general linear model [see, e.g., 1,18,33,35]. It provides a measure of sphericity or effective dimensionality of configuration $(\mathbf{f}, \mathbf{D}_X)$. Interestingly enough, the first two cumulants (12) and (13) depend, in addition to n, only on the effective dimensionality of the two configurations. The latter also measures the similarity of the configuration with the neutral configuration in view of identity $\nu(\boldsymbol{\lambda}) = (n-1)\,\mathtt{RV}^2(\mathbf{K}_X, \mathbf{K}_0)$.

The minimum $\nu(\boldsymbol{\lambda}) = 1$ is attained for univariate configurations. The maximum $\nu(\boldsymbol{\lambda}) = n-1$ is attained for uniform dilatations of the discrete distances $\mathbf{D}_X^0$ (Section 2.1), in which case $\mathbb{V}\mathrm{ar}(\mathtt{RV}) = 0\mathbb{V}$ since RV is then concentrated on $\sqrt{\nu(\boldsymbol{\mu})/(n-1)}$.

The second-order Cornish–Fisher cumulant expansion permits to approximatively redress the normal quantiles by taking into account the skewness and the "taildeness" of a non-normal distribution [see, e.g., 2,25]. The observed RV is statistically significant at level α if (one-tailed test)

$$\underbrace{\mathtt{RV}_s = \frac{\mathtt{RV} - \mathbb{E}(\mathtt{RV})}{\sqrt{\mathbb{V}\mathrm{ar}(\mathtt{RV})}}}_{z-score} > \underbrace{u_{1-\alpha}}_{standard\ normal\ quantile}$$
$$+ \underbrace{\frac{\mathbb{A}(\mathtt{RV})}{6}(u_{1-\alpha}^2 - 1) + \frac{\mathbb{\Gamma}(\mathtt{RV})}{24}(u_{1-\alpha}^3 - 3u_{1-\alpha}) - \frac{\mathbb{A}^2(\mathtt{RV})}{36}(2u_{1-\alpha}^3 - 5u_{1-\alpha})}_{correction\ to\ the\ normal\ distribution} . \tag{18}$$

3.2. Invariant orthogonal integration

The rest of the paper is devoted to presenting invariant orthogonal integration [see, e.g., 9,10,30,36, and references therein] and proving Theorem 1. Invariant orthogonal integration is a major theme of random matrix theory, itself developed ever since the end of the *XIX*th Century [see 16, for an historical perspective].

Consider two standard kernels together with their spectral decomposition (3) $\mathbf{K}_X = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ and $\mathbf{K}_Y = \mathbf{V}\mathbf{M}\mathbf{V}^\top$ with $\mathbf{M} = \mathrm{diag}(\boldsymbol{\mu})$, where $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{n\times(n-1)}$. Define the matrix $\mathbf{W} = (w_{\alpha\beta}) \in \mathbb{R}^{(n-1)\times(n-1)}$ as

$$\mathbf{W} = \mathbf{U}^\top\mathbf{V} . \tag{19}$$

The numerator CV of RV in (11) reads

$$\mathtt{CV}(\mathbf{W}) = \mathrm{Tr}(\mathbf{K}_X\mathbf{K}_Y) = \mathrm{Tr}(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top\mathbf{V}\mathbf{M}\mathbf{V}^\top) = \sum_{\alpha=1}^{n-1}\sum_{\beta=1}^{n-1}\lambda_\alpha\mu_\beta P_{\alpha\beta} = \sum_{\alpha=1}^{n-1}\sum_{\beta=1}^{n-1}\lambda_\alpha\mu_\beta w_{\alpha\beta}^2 , \tag{20}$$

where

$$P_{\alpha\beta} = \sum_{i,j=1}^{n} u_{i\alpha}u_{j\alpha}v_{i\beta}v_{j\beta} = (\sum_{i=1}^{n} u_{i\alpha}v_{i\beta})^2 = w_{\alpha\beta}^2 = \cos^2\angle(u_\alpha, v_\beta) \tag{21}$$

is a measure of alignment between the eigenvectors u_α of $\mathbf{K}_X$ and v_β of $\mathbf{K}_Y$.

Identities $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_{n-1}$, $\mathbf{U}^\top\sqrt{\mathbf{f}} = \mathbf{0}_n$ and $\mathbf{V}\mathbf{V}^\top = \mathbf{I}_n - \sqrt{\mathbf{f}}\sqrt{\mathbf{f}}^\top$ in (3) imply the identity $\mathbf{W}\mathbf{W}^\top = \mathbf{U}^\top\mathbf{V}\mathbf{V}^\top\mathbf{U} = \mathbf{I}_{n-1}$. Similarly, $\mathbf{W}^\top\mathbf{W} = \mathbf{I}_{n-1}$, that is $\mathbf{W} \in \mathbb{R}^{(n-1)\times(n-1)}$ in (19) constitutes an orthogonal matrix acting in the non-trivial eigenspace. In particular (where "•" denotes a sum over all values of the replaced index)

$$P_{\bullet\beta} = \sum_{\alpha=1}^{n-1} P_{\alpha\beta} = \sum_{\alpha=1}^{n-1} w_{\alpha\beta}^2 = 1 \tag{22}$$

and, similarly, $P_{\alpha\bullet} = 1$: the matrix $\mathbf{P} = (P_{\alpha\beta}) \in \mathbb{R}^{(n-1)\times(n-1)}$ is non-negative, and doubly stochastic: it expresses as a mixture of permutations of $\mathcal{S}_{n-1}$ (Birkhoff–von Neumann theorem [7]). In particular, one gets the crude estimate

$$\sum_{\alpha\ge1}\lambda_\alpha\mu_{n-\alpha} \le \mathtt{CV} \le \sum_{\alpha\ge1}\lambda_\alpha\mu_\alpha .$$

The null hypothesis H_0 states that the two configurations $(\mathbf{f}, \mathbf{D}_X)$ and $(\mathbf{f}, \mathbf{D}_Y)$ are unrelated: under H_0, any angle $\angle(u_\alpha, v_\beta)$ in (21) is equally likely. Hence, the eigenvectors of the first configuration will be rotated and/or reflected by replacing $\mathbf{U} = (u_{i\alpha})$ in (21) by $\mathbf{U}\mathbf{T}$, where $\mathbf{T} = (t_{\alpha a}) \in \mathbb{O}_{n-1}$, the orthogonal group of dimension $n-1$. This transformation acts in the non-trivial eigenspace only, leaving the weights $\mathbf{f}$ unchanged.

The transformed coordinates (4) $\mathbf{X}(\mathbf{T}) = \mathbf{\Pi}^{-\frac{1}{2}}\mathbf{U}\mathbf{T}\mathbf{\Lambda}^{\frac{1}{2}}$ differ from the rotated-reflected coordinates $\mathbf{X}\mathbf{T}$ (unless $\mathbf{\Lambda}$ is constant). Similarly, $\mathbf{K}_X(\mathbf{T}) = \mathbf{U}\mathbf{T}\mathbf{\Lambda}\mathbf{T}^\top\mathbf{U}^\top$ differs from $\mathbf{K}_X$, but $\mathrm{Tr}(\mathbf{K}_X^q(\mathbf{T})) = \mathrm{Tr}(\mathbf{K}_X^q)$ for $q \in \{1, 2, \ldots\}$ In particular, the inertia Δ in (5) and the denominators in (11) are left unchanged by the transformation.

Also, the orthogonal matrix (19) transforms as $\widetilde{\mathbf{W}} = \mathbf{T}^\top\mathbf{U}^\top\mathbf{V} = \mathbf{T}^\top\mathbf{W} \in \mathbb{O}_{n-1}$, and the CV coefficient becomes

$$\mathtt{CV}(\mathbf{T}^\top\mathbf{W}) = \mathtt{CV}(\widetilde{\mathbf{W}}) = \mathrm{Tr}(\mathbf{\Lambda}\mathbf{T}^\top\mathbf{U}^\top\mathbf{V}\mathbf{M}\mathbf{V}^\top\mathbf{U}\mathbf{T}) = \mathrm{Tr}(\mathbf{\Lambda}\widetilde{\mathbf{W}}\mathbf{M}\widetilde{\mathbf{W}}^\top) = \sum_{\alpha=1}^{n-1}\sum_{\beta=1}^{n-1}\lambda_\alpha\mu_\beta\tilde{w}_{\alpha\beta}^2 . \tag{23}$$

The idea of invariant orthogonal integration is to compute the expectation of the moments

$$\mathbb{E}(\mathtt{CV}^q) := \int_{\mathbb{O}_{n-1}} \mathtt{CV}^q(\mathbf{T}^\top\mathbf{W})\, d\mu(\mathbf{T}) = \int_{\mathbb{O}_{n-1}} \mathtt{CV}^q(\mathbf{W})\, d\mu(\mathbf{W}) \qquad q \in \{1, 2, \ldots\} \tag{24}$$

by averaging over all possible orthogonal transformations $\mathbf{T} = (t_{a\alpha}) \in \mathbb{O}_{n-1}$ distributed by the invariant Haar measure $d\mu(\mathbf{T})$ normalized to $\int_{\mathbb{O}_{n-1}} d\mu(\mathbf{T}) = 1$. The last identity results from the Haar invariance property $d\mu(\mathbf{T}) = d\mu(\mathbf{T}^\top) = d\mu(\mathbf{T}^\top\mathbf{W})$. Note that the moment generating function reads

$$\mathbb{E}(\exp(t\,\mathtt{CV})) = \int_{\mathbb{O}_{n-1}} \exp(t\,\mathrm{Tr}(\boldsymbol{\Lambda}\mathbf{WMW}^\top))\,d\mu(\mathbf{W})\ . \tag{25}$$

Define $[n] = \{1, 2, \ldots, n\}$. Computing (24) involves the following integrals, we shall refer to as orthogonal coefficients, defined in whole generality as

$$\mathcal{I}_{\mathbf{a}}^{\mathbf{b}} = \mathbb{E}(w_{a_1b_1}\, w_{a_2b_2} \ldots w_{a_{2q}b_{2q}}) = \int_{\mathbb{O}_{n-1}} d\mu(\mathbf{W})\, w_{a_1b_1}\, w_{a_2b_2} \ldots w_{a_{2q}b_{2q}}\ , \tag{26}$$

where the multi-indices $\mathbf{a} = (a_1a_2\ldots a_{2q})$ and $\mathbf{b} = (b_1b_2\ldots b_{2q})$ are elements of $[n-1]^{2q}$ that is, $\mathbf{a}$ and $\mathbf{b}$ are words of length $2q$ on the alphabet $[n-1]$.

To ease the notations, define $A_q = [n-1]^q$ and, for $\boldsymbol{\alpha} = (\alpha_1\ldots\alpha_q) \in A_q$, define $\boldsymbol{\alpha\alpha} = (\alpha_1\alpha_1\alpha_2\alpha_2\ldots\alpha_q\alpha_q) \in A_{2q}$. Identities (23), (24) and (26) yield

$$\mathbb{E}(\mathtt{CV}^q) = \sum_{\alpha_1\ldots\alpha_q=1}^{n-1} \underbrace{\lambda_{\alpha_1}\cdots\lambda_{\alpha_q}}_{\lambda_{\boldsymbol{\alpha}}} \sum_{\beta_1\ldots\beta_q=1}^{n-1} \underbrace{\mu_{\beta_1}\cdots\mu_{\beta_q}}_{\mu_{\boldsymbol{\beta}}} \underbrace{\mathbb{E}(w_{\alpha_1\beta_1}\, w_{\alpha_1\beta_1}\, w_{\alpha_2\beta_2}\, w_{\alpha_2\beta_2} \ldots w_{\alpha_q\beta_q}\, w_{\alpha_q\beta_q})}_{\mathcal{I}_{\boldsymbol{\alpha\alpha}}^{\boldsymbol{\beta\beta}}} = \sum_{\boldsymbol{\alpha}\in A_q} \lambda_{\boldsymbol{\alpha}} \sum_{\boldsymbol{\beta}\in A_q} \mu_{\boldsymbol{\beta}}\, \mathcal{I}_{\boldsymbol{\alpha\alpha}}^{\boldsymbol{\beta\beta}}\ . \tag{27}$$

Determining the orthogonal coefficients $\mathcal{I}_{\boldsymbol{\alpha\alpha}}^{\boldsymbol{\beta\beta}}$ will yield exact expressions for $\mathbb{E}(\mathtt{CV}^q)$ in terms of spectral moments of $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$, or equivalently in terms of traces of integer powers of $\mathbf{K}_X$ and $\mathbf{K}_Y$, as demonstrated in the next sections for $q \in \{1, 2, 3, 4\}$.

3.3. Computing low-order orthogonal coefficients

Evaluating the orthogonal coefficients (26) is a major topic in random matrix theory and free probability, and its systematic handling is presented in Section 3.6. Yet, as observed by some authors [see, e.g., 3,9,36], well-inspired invariance considerations (Lemmas 1 and 2) suffice in determining more directly the values of the orthogonal coefficients of low order.

Since $d\mu(-\mathbf{W}) = d\mu(\mathbf{W})$, coefficients (26) are zero unless each index in $\mathbf{a}$ and in $\mathbf{b}$ occurs an even number of times, with a total of $2q$ occurrences, where q defines the order of the orthogonal coefficient. Also, applying the same permutation on the two multi-indexes, or exchanging the multi-indexes leaves the coefficients unchanged. Furthermore, the particular value taken by an index is irrelevant: only matters its multiplicity. For instance, in general:

$$\mathcal{I}_{\alpha\beta\gamma\delta}^{abcd} = \mathcal{I}_{abcd}^{\alpha\beta\gamma\delta} = \mathcal{I}_{\delta\alpha\gamma\beta}^{dacb} \neq \mathcal{I}_{\delta\alpha\gamma\beta}^{adcb}\ ; \qquad \mathcal{I}_{\alpha\alpha\gamma\gamma}^{abcc} = \mathcal{I}_{\alpha\alpha\gamma\gamma}^{accb} = 0\ ,\ a \neq b\ .$$

Also, for $\alpha \neq \gamma$ and $a \neq b$,

$$\mathcal{I}_{\alpha\alpha\gamma\gamma}^{aabb} = \mathcal{I}_{\alpha\alpha\gamma\gamma}^{bbaa} = \mathcal{I}_{1122}^{2211} = \mathcal{I}_{1122}^{1122}\ .$$

Lemma 1 (*Proved in the* Appendix)**.** *Let $\alpha \neq \gamma$ and let $\boldsymbol{\varepsilon}$ be a multi-index not containing α, γ. For any indices a, b, c, d and multi-index $\boldsymbol{e}$ of the same size as $\boldsymbol{\varepsilon}$*

$$\mathcal{I}_{\alpha\alpha\alpha\alpha\boldsymbol{\varepsilon}}^{abcd\boldsymbol{e}} = \mathcal{I}_{\alpha\alpha\gamma\gamma\boldsymbol{\varepsilon}}^{abcd\boldsymbol{e}} + \mathcal{I}_{\alpha\gamma\alpha\gamma\boldsymbol{\varepsilon}}^{abcd\boldsymbol{e}} + \mathcal{I}_{\alpha\gamma\gamma\alpha\boldsymbol{\varepsilon}}^{abcd\boldsymbol{e}}\ .$$

In particular,

$$\mathcal{I}_{\alpha\alpha\alpha\alpha\boldsymbol{\varepsilon}}^{aacc\boldsymbol{e}} = \mathcal{I}_{\alpha\alpha\gamma\gamma\boldsymbol{\varepsilon}}^{aacc\boldsymbol{e}} + 2\mathcal{I}_{\alpha\gamma\alpha\gamma\boldsymbol{\varepsilon}}^{aacc\boldsymbol{e}}\ , \qquad \mathcal{I}_{\alpha\alpha\alpha\alpha\boldsymbol{\varepsilon}}^{aaaa\boldsymbol{e}} = 3\mathcal{I}_{\alpha\alpha\gamma\gamma\boldsymbol{\varepsilon}}^{aaaa\boldsymbol{e}}\ . \tag{28}$$

Lemma 2. *For any multi-index $\boldsymbol{e}$ not containing β, and for any unrestricted multi-index $\boldsymbol{\varepsilon}$ of the same size,*

$$\sum_{\beta=1}^{n-1} \mathcal{I}_{\alpha\gamma\boldsymbol{\varepsilon}}^{\beta\beta\boldsymbol{e}} = \delta_{\alpha\gamma}\, \mathcal{I}_{\boldsymbol{\varepsilon}}^{\boldsymbol{e}}\ . \tag{29}$$

Proof. Eq. (29) follows directly from $\boldsymbol{WW}^\top = \boldsymbol{I}_{n-1}$, that is $\sum_{\beta=1}^{n-1} w_{\alpha\beta}w_{\gamma\beta} = \delta_{\alpha\gamma}$. □

3.4. The first and second moments

The computation of the first moment, identical to the proposal of Kazi-Aoual et al. [24] who used averaging on all permutations between the n (unweighted) objects, is straightforward: $\mathcal{I}^{\beta\delta}_{\alpha\alpha} = \delta_{\beta\delta}\, \mathcal{I}^{\beta\beta}_{\alpha\alpha}$, where $\mathcal{I}^{\beta\beta}_{\alpha\alpha}$ is independent of α and β. By (29), (20) and (27)

$$\mathcal{I}^{\beta\beta}_{\alpha\alpha} = \frac{1}{n-1}\ , \qquad \mathbb{E}(P_{\alpha\beta}) = \frac{1}{n-1}\ , \qquad \mathbb{E}(\mathrm{CV}) = \frac{1}{n-1}\sum_{\alpha,\beta=1}^{n-1} \lambda_\alpha \mu_\beta\ . \tag{30}$$

The computation of the second moment involves four orthogonal coefficients, namely (all super- and sub-indices in (31) are distinct)

$$E := \mathcal{I}^{\beta\beta\beta\beta}_{\alpha\alpha\alpha\alpha}\ , \quad F := \mathcal{I}^{\beta\beta\zeta\zeta}_{\alpha\alpha\alpha\alpha} = \mathcal{I}^{\alpha\alpha\gamma\gamma}_{\beta\beta\beta\beta}\ , \quad G := \mathcal{I}^{\beta\beta\zeta\zeta}_{\alpha\alpha\gamma\gamma}\ , \quad H := \mathcal{I}^{\beta\beta\zeta\zeta}_{\alpha\gamma\alpha\gamma}\ . \tag{31}$$

They satisfy

$$F \overset{(28)}{=} G + 2H\ , \qquad E \overset{(28)}{=} 3F\ , \qquad (n-2)G + F \overset{(29),(30)}{=} \frac{1}{n-1}\ , \qquad (n-2)H + F \overset{(29),(30)}{=} 0$$

with solution

$$E = 3(n-2)\kappa\ , \quad F = (n-2)\kappa\ , \quad G = n\kappa\ , \quad H = -\kappa\ , \quad \kappa = \frac{1}{(n-2)(n-1)(n+1)}\ . \tag{32}$$

Hence (the expression is also valid for four possibly coinciding sub-indices, since $E = 3F$),

$$\mathcal{I}^{\beta\delta\zeta\theta}_{\alpha\alpha\alpha\alpha} = (\delta_{\beta\delta}\delta_{\zeta\theta} + \delta_{\beta\zeta}\delta_{\delta\theta} + \delta_{\beta\theta}\delta_{\delta\zeta})F \tag{33}$$

and, for $\alpha \neq \gamma$,

$$\mathcal{I}^{\beta\delta\zeta\theta}_{\alpha\alpha\gamma\gamma} = \delta_{\beta\delta}\delta_{\zeta\theta}\, G + (\delta_{\beta\zeta}\delta_{\delta\theta} + \delta_{\beta\theta}\delta_{\delta\zeta})H\ . \tag{34}$$

As a result, performing $\delta_{\alpha\gamma} \times$ (33) $+ (1-\delta_{\alpha\gamma}) \times$ (34) yields the general formula, where the super- and the sub-indices may be distinct or not

$$\mathcal{I}^{\beta\delta\zeta\theta}_{\alpha\alpha\gamma\gamma} = \kappa\, [n\delta_{\beta\delta}\delta_{\zeta\theta} - (\delta_{\delta\zeta}\delta_{\delta\theta} + \delta_{\beta\theta}\delta_{\delta\zeta}) - 2\delta_{\alpha\gamma}\delta_{\beta\delta}\delta_{\zeta\theta} + (n-1)\delta_{\alpha\gamma}(\delta_{\beta\zeta}\delta_{\delta\theta} + \delta_{\beta\theta}\delta_{\delta\zeta})]$$

which finally implies from (27)

$$\mathbb{E}(P_{\alpha\beta}P_{\gamma\delta}) = \kappa\, [n - 2\delta_{\alpha\gamma} - 2\delta_{\beta\delta} + 2(n-1)\delta_{\alpha\gamma}\delta_{\beta\delta}]\ . \tag{35}$$

Hence, by (27) and (10)

$$\mathbb{E}(\mathrm{CV}^2) = \kappa(n-1)^3\, [n(n-1)\overline{\boldsymbol{\lambda}}^2\, \overline{\boldsymbol{\mu}}^2 - 2\overline{\boldsymbol{\lambda}}^2\, \overline{\boldsymbol{\mu}^2} - 2\overline{\boldsymbol{\lambda}^2}\, \overline{\boldsymbol{\mu}}^2 + 2\overline{\boldsymbol{\lambda}^2}\, \overline{\boldsymbol{\mu}^2}]\ . \tag{36}$$

Subtracting $\mathbb{E}^2(\mathrm{CV})$ obtained in (30), substituting the value of κ in (32) and rearranging terms yields the second identity in (8).

Expressions for the second centered moments in (8) and (13) are simpler than the corresponding quantities obtained, in the unweighted setting, by averaging over all permutations of the n objects: the latter contain additional correction terms, as derived in Kazi-Aoual et al. [24]. See also Heo and Ruben Gabriel [21], Josse et al. [23] and Abdi [1].

3.5. The third moment

The third moment reads

$$\mathbb{E}(\mathrm{CV}^3) = \sum_{\alpha,\beta,\gamma,\delta,\varepsilon,\zeta=1}^{n-1} \lambda_\alpha\, \lambda_\gamma\, \lambda_\varepsilon\, \mu_\beta\, \mu_\delta\, \mu_\zeta\, \mathbb{E}(P_{\alpha\beta}P_{\gamma\delta}P_{\varepsilon\zeta}) = \sum_{\alpha,\beta,\gamma,\delta,\varepsilon,\zeta=1}^{n-1} \lambda_\alpha\, \lambda_\gamma\, \lambda_\varepsilon\, \mu_\beta\, \mu_\delta\, \mu_\zeta\, \mathcal{I}^{\beta\beta\delta\delta\zeta\zeta}_{\alpha\alpha\gamma\gamma\varepsilon\varepsilon} \tag{37}$$

and involves eleven third-order orthogonal coefficients, namely (all super- and sub-indices in (38) are distinct)

$$\begin{aligned} &L := \mathcal{I}^{\beta\beta\beta\beta\beta\beta}_{\alpha\alpha\alpha\alpha\alpha\alpha}\ , \quad M := \mathcal{I}^{\beta\beta\beta\beta\beta\beta}_{\alpha\alpha\alpha\alpha\gamma\gamma}\ , \quad N := \mathcal{I}^{\beta\beta\beta\beta\beta\beta}_{\alpha\alpha\gamma\gamma\varepsilon\varepsilon}\ , \quad P := \mathcal{I}^{\beta\beta\beta\beta\delta\delta}_{\alpha\alpha\alpha\alpha\gamma\gamma}\ , \quad Q := \mathcal{I}^{\beta\beta\beta\beta\delta\delta}_{\alpha\alpha\gamma\gamma\alpha\alpha}\ , \quad R := \mathcal{I}^{\beta\beta\beta\beta\delta\delta}_{\alpha\alpha\alpha\gamma\alpha\gamma} \\ &S := \mathcal{I}^{\beta\beta\beta\beta\delta\delta}_{\alpha\alpha\gamma\gamma\varepsilon\varepsilon}\ , \quad T := \mathcal{I}^{\beta\beta\beta\beta\delta\delta}_{\alpha\gamma\varepsilon\varepsilon\alpha\gamma} \quad U := \mathcal{I}^{\beta\beta\delta\delta\zeta\zeta}_{\alpha\alpha\gamma\gamma\varepsilon\varepsilon}\ , \quad V := \mathcal{I}^{\beta\beta\delta\delta\zeta\zeta}_{\alpha\gamma\alpha\gamma\varepsilon\varepsilon}\ , \quad W := \mathcal{I}^{\beta\beta\delta\delta\zeta\zeta}_{\alpha\varepsilon\gamma\varepsilon\alpha\gamma}\ . \end{aligned} \tag{38}$$

Handcrafted computations are a bit awkward, yet feasible, with the result

Lemma 3 (*Proved in the* Appendix)**.**

$$\begin{aligned} \frac{1}{\hat{\kappa}}\, \mathbb{E}(P_{\alpha\beta}P_{\gamma\delta}P_{\varepsilon\zeta}) = {} & (n^2+n-4) - 2(n+1)(\sigma+\tau) + 16(\varphi+\psi) + 8\sigma\tau \\ & - 8(n-1)(\sigma\psi + \tau\varphi) + 8(n-1)^2\varphi\psi + 2(n-3)(n+3)\omega \end{aligned} \tag{39}$$

where

$$\hat{\kappa} = \frac{\kappa}{(n-3)(n+3)} = \frac{1}{(n-3)(n-2)(n-1)(n+1)(n+3)} \tag{40}$$

and

$$\begin{aligned} &\sigma = \delta_{\alpha\gamma} + \delta_{\alpha\varepsilon} + \delta_{\gamma\varepsilon}\,, \quad \tau = \delta_{\beta\delta} + \delta_{\beta\zeta} + \delta_{\delta\zeta}\,, \quad \omega = \delta_{\alpha\gamma}\delta_{\beta\delta} + \delta_{\alpha\varepsilon}\delta_{\beta\zeta} + \delta_{\gamma\varepsilon}\delta_{\delta\zeta}\,, \\ &\varphi = \delta_{\alpha\gamma}\delta_{\alpha\varepsilon}\delta_{\gamma\varepsilon} = \delta_{\alpha\gamma}\delta_{\alpha\varepsilon} = \delta_{\alpha\gamma}\delta_{\gamma\varepsilon} = \delta_{\alpha\varepsilon}\delta_{\gamma\varepsilon}\,, \quad \psi = \delta_{\beta\delta}\delta_{\beta\zeta}\delta_{\delta\zeta} = \delta_{\beta\delta}\delta_{\beta\zeta} = \delta_{\beta\delta}\delta_{\delta\zeta} = \delta_{\beta\zeta}\delta_{\delta\zeta}\;. \end{aligned} \tag{41}$$

One can check with (35) that

$$\sum_{\varepsilon=1}^{n-1} \mathbb{E}(P_{\alpha\beta}P_{\gamma\delta}P_{\varepsilon\zeta}) = \sum_{\zeta=1}^{n-1} \mathbb{E}(P_{\alpha\beta}P_{\gamma\delta}P_{\varepsilon\zeta}) = \mathbb{E}(P_{\alpha\beta}P_{\gamma\delta})$$

as it must be. Inserting (39) in (37) and using (10) yields

$$\begin{aligned} \frac{\mathbb{E}(\mathtt{CV}^3)}{(n-1)^4\hat{\kappa}} &= (n^2+n-4)(n-1)^2\bar{\boldsymbol{\lambda}}^3\bar{\boldsymbol{\mu}}^3 - 6(n+1)(n-1)(\bar{\boldsymbol{\lambda}}\overline{\boldsymbol{\lambda}^2}\bar{\boldsymbol{\mu}}^3 + \bar{\boldsymbol{\lambda}}^3\bar{\boldsymbol{\mu}}\overline{\boldsymbol{\mu}^2}) \\ &+ 16(\overline{\boldsymbol{\lambda}^3}\bar{\boldsymbol{\mu}}^3 + \bar{\boldsymbol{\lambda}}^3\overline{\boldsymbol{\mu}^3}) + 6(n^2+3)\bar{\boldsymbol{\lambda}}\overline{\boldsymbol{\lambda}^2}\bar{\boldsymbol{\mu}}\overline{\boldsymbol{\mu}^2} - 24(\bar{\boldsymbol{\lambda}}\overline{\boldsymbol{\lambda}^2}\,\overline{\boldsymbol{\mu}^3} + \overline{\boldsymbol{\lambda}^3}\bar{\boldsymbol{\mu}}\overline{\boldsymbol{\mu}^2}) + 8\overline{\boldsymbol{\lambda}^3}\,\overline{\boldsymbol{\mu}^3}\;. \end{aligned} \tag{42}$$

The centered third moment

$$\mathbb{E}(\mathtt{CV}_c^3) = \mathbb{E}((\mathtt{CV}-\mathbb{E}(\mathtt{CV}))^3) = \mathbb{E}(\mathtt{CV}^3) - 3\,\mathbb{E}(\mathtt{CV}^2)\,\mathbb{E}(\mathtt{CV}) + 2\,\mathbb{E}^3(\mathtt{CV}) \tag{43}$$

finally reads, by (30), (36) and (40)

$$\begin{aligned} \frac{\mathbb{E}(\mathtt{CV}_c^3)}{8(n-1)^4\hat{\kappa}} &= 4\bar{\boldsymbol{\lambda}}^3\bar{\boldsymbol{\mu}}^3 - 6(\bar{\boldsymbol{\lambda}}\overline{\boldsymbol{\lambda}^2}\bar{\boldsymbol{\mu}}^3 + \bar{\boldsymbol{\lambda}}^3\bar{\boldsymbol{\mu}}\overline{\boldsymbol{\mu}^2}) + 2(\overline{\boldsymbol{\lambda}^3}\bar{\boldsymbol{\mu}}^3 + \bar{\boldsymbol{\lambda}}^3\overline{\boldsymbol{\mu}^3}) + 9\,\bar{\boldsymbol{\lambda}}\overline{\boldsymbol{\lambda}^2}\bar{\boldsymbol{\mu}}\overline{\boldsymbol{\mu}^2} \\ &-3(\bar{\boldsymbol{\lambda}}\overline{\boldsymbol{\lambda}^2}\,\overline{\boldsymbol{\mu}^3} + \overline{\boldsymbol{\lambda}^3}\bar{\boldsymbol{\mu}}\overline{\boldsymbol{\mu}^2}) + \overline{\boldsymbol{\lambda}^3}\,\overline{\boldsymbol{\mu}^3} = (\overline{\boldsymbol{\lambda}^3} - 3\bar{\boldsymbol{\lambda}}\overline{\boldsymbol{\lambda}^2} + 2\bar{\boldsymbol{\lambda}}^3)(\overline{\boldsymbol{\mu}^3} - 3\bar{\boldsymbol{\mu}}\overline{\boldsymbol{\mu}^2} + 2\bar{\boldsymbol{\mu}}^3) = \overline{\boldsymbol{\lambda}_c^3}\,\overline{\boldsymbol{\mu}_c^3}, \end{aligned}$$

thus proving the third identity in (8). This exact expression for the third moment seems original, and is considerably simpler than the corresponding expression derived by averaging on the $n!$ object permutations [24]. It depends directly on n, but only indirectly on $\mathbf{f}$ through the eigenvalue spectra. Expression (14) for the RV skewness (see Section 3.7 for an alternative derivation) is particularly transparent, and elucidates the cause of the marked positive asymmetry of the RV coefficient, often reported in the literature [see, e.g., 21,23,29,37]: plainly, $a(\boldsymbol{\lambda}) > 0$ and $a(\boldsymbol{\mu}) > 0$ for typical scree plots (see, e.g., Fig. 3).

3.6. The fourth moment

Computing $\mathbb{E}(\mathtt{RV}^4)$, or equivalently $\mathbb{E}(\mathtt{CV}^4)$ is clearly untractable with the former pedestrian approach, and a more structured strategy is needed. The latter is provided by the so-called Weingarten calculus [see 10,12,27,30,31], elaborated as a systematic machinery to evaluate Haar integrals of the form (26) over the unitary, symplectic or (in the present case) orthogonal compact groups. See [11] for a pedagogical and historical account of this vast topic.

Consider $\mathcal{P}_{2q}$, the set of all partitions of $\{1, 2, \ldots, 2q\}$ whose all blocks are of length two, also called pairings [31]. There are $(2q-1)!! = (2q-1)(2q-3)\cdots 5\cdot 3$ distinct pairings. For instance, for $q = 4$, the two partitions

$$\sigma = (13|25|46|78)\,, \qquad \tau = (15|26|34|78)$$

constitute such pairings. Their join $\sigma \vee \tau$ (i.e., the finest partition coarser than both σ and τ) is $\sigma \vee \tau = (123456|78)$.

In general, the join $\sigma \vee \tau$ of two pairings $\sigma, \tau \in \mathcal{P}_{2q}$ is a partition made of $N(\sigma \vee \tau)$ blocks of even sizes $2l_1, 2l_2, 2l_3, \ldots$, with $l_1 \geq l_2 \geq l_3 \geq \cdots$ and $\sum_{c=1}^{N(\sigma\vee\tau)} l_c = q$. The multi-index $\ell = (l_1, l_2, l_3 \ldots)$ constitutes an integer partition of q (noted $\ell \vdash q$), and defines the type $\ell(\sigma \vee \tau)$ of $\sigma \vee \tau$.

For $q = 4$, five integer partitions or types are possible, namely

$$\ell = (1,1,1,1) \equiv (1^4)\,, \qquad \ell = (2,1,1)\,, \qquad \ell = (2,2)\,, \qquad \ell = (3,1)\,, \qquad \ell = (4)\;.$$

Any pairing $\sigma \in \mathcal{P}_{2q}$ also defines a particular permutation between $2q$ indices, exchanging the indices belonging to the same block of two [31]. The orthogonal coefficients (26) turn out to express [see 4,11, and references therein]

$$\mathcal{I}_{\mathbf{a}}^{\mathbf{b}} = \sum_{\sigma\in\mathcal{P}_{2q}} \sum_{\tau\in\mathcal{P}_{2q}} \delta_\sigma(\mathbf{a})\,\delta_\tau(\mathbf{b})\,\mathrm{Wg}(\ell(\sigma\vee\tau)) \tag{44}$$

where the multi-Kronecker symbol $\delta_\sigma(\mathbf{a})$ is equal to 1 if the indices $a_{\sigma(2r-1)}$ and $a_{\sigma(2r)}$ (permutation notation) belonging to the same rth block of pairing σ coincide for all blocks $r \in \{1, \ldots .q\}$. Otherwise, $\delta_\sigma(\mathbf{a}) = 0$. Explicitly,

$$\delta_\sigma(\mathbf{a}) = \prod_{r=1}^{q} \delta_{a_{\sigma(2r-1)},a_{\sigma(2r)}}\,, \qquad \delta_\tau(\mathbf{b}) = \prod_{r=1}^{q} \delta_{b_{\tau(2r-1)},b_{\tau(2r)}} \tag{45}$$

which restricts the sum in (44) over the pairings σ compatible with $\mathbf{a}$ in the above sense (and pairings τ compatible with $\mathbf{b}$), implying in particular that all indices in $\mathbf{a}$ and $\mathbf{b}$ must occur an even number of times.

The quantities $\mathrm{Wg}(\ell(\sigma\vee\tau))$ appearing in (44) are the orthogonal Weingarten functions, and depend upon the dimension $d = n-1$ as well. They have been computed up to order $q = 6$ [10]. For $q = 4$:

$$\mathrm{Wg}(1^4) = \phi\,(n-3)(n+2)(n^2+4n-4) \qquad \mathrm{Wg}(2,1,1) = \phi\,(-n^3-3n^2+6n+4) \qquad \mathrm{Wg}(2,2) = \phi\,(n^2+3n+14)$$
$$\mathrm{Wg}(3,1) = \phi\,(n-1)(2n+6) \qquad \mathrm{Wg}(4) = -\phi\,(5n+1) \tag{46}$$

where

$$\phi = \frac{1}{(n-4)(n-3)(n-2)(n-1)n(n+1)(n+3)(n+5)}\ . \tag{47}$$

Substituting (44) in (27) yields

$$\mathbb{E}(\mathrm{CV}^q) = \sum_{\sigma\in\mathcal{P}_{2q}}\sum_{\tau\in\mathcal{P}_{2q}} \mathrm{Wg}(\ell(\sigma\vee\tau)) \sum_{\boldsymbol{\alpha}\in A_q} \delta_\sigma(\boldsymbol{\alpha}\boldsymbol{\alpha})\,\lambda_{\boldsymbol{\alpha}} \sum_{\boldsymbol{\beta}\in A_q} \delta_\tau(\boldsymbol{\beta}\boldsymbol{\beta})\,\mu_{\boldsymbol{\beta}} = \sum_{\sigma\in\mathcal{P}_{2q}}\sum_{\tau\in\mathcal{P}_{2q}} \mathrm{Wg}(\ell(\sigma\vee\tau))\,\mathrm{Tr}_\sigma(\mathbf{K}_X)\,\mathrm{Tr}_\tau(\mathbf{K}_Y)$$
$$= \sum_{\sigma\in\mathcal{P}_{2q}}\sum_{\tau\in\mathcal{P}_{2q}} (n-1)^{N(\sigma\vee\sigma_0)+N(\tau\vee\sigma_0)}\ \mathrm{Wg}(\ell(\sigma\vee\tau)) \prod_{c=1}^{N(\sigma\vee\sigma_0)} \overline{\boldsymbol{\lambda}^{l_c}} \prod_{\tilde{c}=1}^{N(\tau\vee\sigma_0)} \overline{\boldsymbol{\mu}^{l_{\tilde{c}}}}\ , \tag{48}$$

where the following lemma and definition have been used:

Lemma 4. *Consider the reference pairing $\sigma_0 = (12|34|\dots|2q-1,q) \in \mathcal{P}_{2q}$, and consider the type $\ell(\sigma\vee\sigma_0)$, also called coset-type of σ [see, e.g., 11,27]. Define*

$$Tr_\sigma(\mathbf{K}) = \prod_{c=1}^{N(\sigma\vee\sigma_0)} Tr(\mathbf{K}^{l_c})\,, \qquad tr_\sigma(\mathbf{K}) = \prod_{c=1}^{N(\sigma\vee\sigma_0)} tr(\mathbf{K}^{l_c})\,. \tag{49}$$

Then $\sum_{\boldsymbol{\alpha}\in A_q} \delta_\sigma(\boldsymbol{\alpha}\boldsymbol{\alpha})\,\lambda_{\boldsymbol{\alpha}} = Tr_\sigma(\mathbf{K}_X)$, which also reads

$$Tr_\sigma(\mathbf{K}_X) = (n-1)^{N(\sigma\vee\sigma_0)}\, tr_\sigma(\mathbf{K}_X) = (n-1)^{N(\sigma\vee\sigma_0)} \prod_{c=1}^{N(\sigma\vee\sigma_0)} \overline{\boldsymbol{\lambda}^{l_c}}\ . \tag{50}$$

Proof of Lemma 4. Consider $\mathbf{a} = \boldsymbol{\alpha}\boldsymbol{\alpha} \in A_{2q}$. By construction, $a_{(2r-1)} = a_{2r}$, that is $\delta_{\sigma_0}(\mathbf{a}) = 1$. On the other hand, the term $\delta_\sigma(\boldsymbol{\alpha}\boldsymbol{\alpha})$ imposes $\omega_{\sigma(2r-1)} = \omega_{\sigma(2r)}$. Hence all indices of $\mathbf{a}$ in the blocks of $\sigma\vee\sigma_0$ (of sizes $2l_1, 2l_2, 2l_3\dots, 2l_{N(\sigma\vee\sigma_0)}$) are identical, that is the sum on $\boldsymbol{\alpha} \in A_q$ involves $N(\sigma\vee\sigma_0)$ unconstrained indices respectively repeated exactly $(l_1, l_2, l_3\dots, l_{N(\sigma\vee\sigma_0)})$ times. □

Transforming expression (48) into an effective formula requires to determine, among the $((2q-1)!!)^2$ pairings (σ,τ) entering into the sum, how many are jointly of type $\ell(\sigma\vee\tau)$, $\ell(\sigma\vee\sigma_0)$ and $\ell(\tau\vee\sigma_0)$.

For $q = 4$, Table 1 gives the distribution of joint counts of the $105^2 = 11\,025$ pairings (σ,τ), among the $5^3 = 125$ possible trivariate types. Those counts have, for lack of foreseeable analytical approach, been mechanically computed with the help of the `R` package `igraph` [14], by the functions `union()` (determining the join of two pairings coded as binary graphs) and `components()` (determining the join type).

Working with centered quantities notably simplifies the computations:

Lemma 5.

$$\mathrm{CV}_c = \mathrm{CV} - \mathbb{E}(\mathrm{CV}) = \sum_{\alpha=1}^{n-1}\sum_{\beta=1}^{n-1} \lambda^c_\alpha \mu^c_\beta P_{\alpha\beta}\ .$$

Proof of Lemma 5. By the first identity in (8), (20) and $P_{\alpha\bullet} = P_{\bullet\beta} = 1$,

$$\sum_{\alpha=1}^{n-1}\sum_{\beta=1}^{n-1} \lambda^c_\alpha \mu^c_\beta P_{\alpha\beta} = \sum_{\alpha\beta} (\lambda_\alpha - \overline{\boldsymbol{\lambda}})(\mu_\beta - \overline{\boldsymbol{\mu}}) P_{\alpha\beta} = \sum_{\alpha\beta} \lambda_\alpha\mu_\beta P_{\alpha\beta} - (n-1)\overline{\boldsymbol{\lambda}}\,\overline{\boldsymbol{\mu}}$$
$$-\,(n-1)\overline{\boldsymbol{\lambda}}\,\overline{\boldsymbol{\mu}} + (n-1)\overline{\boldsymbol{\lambda}}\,\overline{\boldsymbol{\mu}} = \sum_{\alpha\beta} \lambda_\alpha\mu_\beta P_{\alpha\beta} - (n-1)\overline{\boldsymbol{\lambda}}\,\overline{\boldsymbol{\mu}} = \mathrm{CV} - \mathbb{E}(\mathrm{CV}) = \mathrm{CV}_c\ . \quad \square$$

Table 1
Trivariate type counts for $q = 4$: each table refers to the type of $\sigma \vee \tau$. Rows refer to the type of $\sigma \vee \sigma_0$, and columns to the type of $\tau \vee \sigma_0$.

	$\sigma \vee \tau = (1^4)$					$\sigma \vee \tau = (2, 1, 1)$					$\sigma \vee \tau = (2, 2)$				
	(1^4)	(2, 1, 1)	(2, 2)	(3, 1)	(4)	(1^4)	(2, 1, 1)	(2, 2)	(3, 1)	(4)	(1^4)	(2, 1, 1)	(2, 2)	(3, 1)	(4)
(1^4)	1					12							12		
(2, 1, 1)		12				12	12	24	96			24	24		96
(2, 2)			$\boxed{12}$				24	$\boxed{24}$		$\boxed{96}$	12	24	$\boxed{60}$		$\boxed{48}$
(3, 1)				32			96		96	192				192	192
(4)					$\boxed{48}$			$\boxed{96}$	192	$\boxed{288}$		96	$\boxed{48}$	192	$\boxed{240}$

	$\sigma \vee \tau = (3, 1)$					$\sigma \vee \tau = (4)$				
	(1^4)	(2, 1, 1)	(2, 2)	(3, 1)	(4)	(1^4)	(2, 1, 1)	(2, 2)	(3, 1)	(4)
(1^4)				32						48
(2, 1, 1)		96		96	192			96	192	288
(2, 2)				192	$\boxed{192}$		96	$\boxed{48}$	192	$\boxed{240}$
(3, 1)	32	96	192	320	384		192	192	384	768
(4)		192	$\boxed{192}$	384	$\boxed{768}$	48	288	$\boxed{240}$	768	$\boxed{960}$

Consequently, (48) entails

$$\mathbb{E}(\mathrm{CV}_c^q) = \sum_{\sigma \in \mathcal{P}_{2q}} \sum_{\tau \in \mathcal{P}_{2q}} (n-1)^{N(\sigma\vee\sigma_0)+N(\tau\vee\sigma_0)} \; \mathrm{Wg}(\ell(\sigma \vee \tau)) \prod_{c=1}^{N(\sigma\vee\sigma_0)} \overline{\boldsymbol{\lambda}_c^{l_c}} \prod_{\tilde{c}=1}^{N(\tau\vee\sigma_0)} \overline{\boldsymbol{\mu}_c^{l_{\tilde{c}}}} \tag{51}$$

in which, for $q = 4$, the contributions of σ and τ coset types (1^4), (2, 1, 1) and (3, 1), associated to $\overline{\boldsymbol{\lambda}_c} = 0$ or $\overline{\boldsymbol{\mu}_c} = 0$, are zeroed: only (2, 2) and (4) survive, with contributions indicated by the boxed counts in Table 1. Explicitly, the coefficient of $\overline{\boldsymbol{\lambda}_c^2}\,\overline{\boldsymbol{\lambda}_c^2}\,\overline{\boldsymbol{\mu}_c^2}\,\overline{\boldsymbol{\mu}_c^2}$ in (51) is

$$(n-1)^4[12\,\mathrm{Wg}(1^4) + 24\,\mathrm{Wg}(2,1,1) + 60\,\mathrm{Wg}(2,2) + 48\,\mathrm{Wg}(4)] = 12\,\phi\,(n-1)^4\,(n^4 + n^3 - 15n^2 - 13n + 98) \ ,$$

the coefficient of $\overline{\boldsymbol{\lambda}_c^4}\,\overline{\boldsymbol{\mu}_c^4}$ is

$$(n-1)^2[48\,\mathrm{Wg}(1^4) + 288\,\mathrm{Wg}(2,1,1) + 240\,\mathrm{Wg}(2,2) + 768\,\mathrm{Wg}(3,1) + 960\,\mathrm{Wg}(4)] = 48\,\phi\,(n-1)^4\,(n^2 - n + 2) \ ,$$

and the coefficient of $\overline{\boldsymbol{\lambda}_c^4}\,\overline{\boldsymbol{\mu}_c^2}\,\overline{\boldsymbol{\mu}_c^2}$ and $\overline{\boldsymbol{\lambda}_c^2}\,\overline{\boldsymbol{\lambda}_c^2}\,\overline{\boldsymbol{\mu}_c^4}$ is

$$(n-1)^3[96\,\mathrm{Wg}(2,1,1) + 48\,\mathrm{Wg}(2,2) + 192\,\mathrm{Wg}(3,1) + 240\,\mathrm{Wg}(4)] = 48\,\phi\,(n-1)^4\,(2n^2 - n - 7) \ .$$

The final expressions in the above follow from (46) and, together with (47), prove (9). They have been further checked with the software *Mathematica*. Expression (9) for the fourth moment is relatively simple, but it lacks elegance and direct interpretation.

3.7. The third moment, revisited

Let us apply the steps of the previous Section for $q = 3$ to verify the coincidence of the Weingarten and pedestrian approaches. Three types occur for $q = 3$, namely $\ell = (1, 1, 1)$, $\ell = (2, 1)$ and $\ell = (3)$. The contribution of coset-types (1, 1, 1) and (2, 1) for σ or τ is zero by consequence of centration. Hence, only the coset-types (3) contribute to (51), which is therefore simply proportional to $\overline{\boldsymbol{\lambda}_c^3}\,\overline{\boldsymbol{\mu}_c^3}$. The conciseness of the last identity in (8) and (14) is thus elucidated. The proportionality coefficient is determined by the boxed components of Table 2 as

$$(n-1)^2[8\,\mathrm{Wg}(1^3) + 24\,\mathrm{Wg}(2,1) + 32\,\mathrm{Wg}(3)] = \frac{8(n-1)^3}{(n-3)(n-2)(n+1)(n+3)} \tag{52}$$

which is exactly the third identity in (8), obtained much more indirectly in Section 3.5. The values of the Weingarten coefficients in (52) were obtained from Collins and Śniady [12]. They read with the present notations ($d = n - 1$ and (40)) as

$$\mathrm{Wg}(1^3) = \hat{\kappa}\,(n^2 + n - 4)\,, \quad \mathrm{Wg}(2,1) = -\hat{\kappa}\,(n+1)\,, \quad \mathrm{Wg}(3) = 2\hat{\kappa} \ .$$

Those coefficients coincide, in order, with the values U, V and W defined in (38) and determined in (53) in the Appendix, as they must in view of (44). Hence, for $q = 3$, the pedestrian approach of Sections 3.3, 3.4 and 3.5 exactly matches

Table 2
Trivariate type counts for $q = 3$: each table refers to the type of $\sigma \vee \tau$. Rows refer to the type of $\sigma \vee \sigma_0$, and columns to the type of $\tau \vee \sigma_0$.

	$\sigma \vee \tau = (1^3)$			$\sigma \vee \tau = (2, 1)$			$\sigma \vee \tau = (3)$		
	(1^3)	$(2, 1)$	(3)	(1^3)	$(2, 1)$	(3)	(1^3)	$(2, 1)$	(3)
(1^3)	1				6				8
$(2, 1)$		6		6	6	24		24	24
(3)			8		24	24	8	24	32

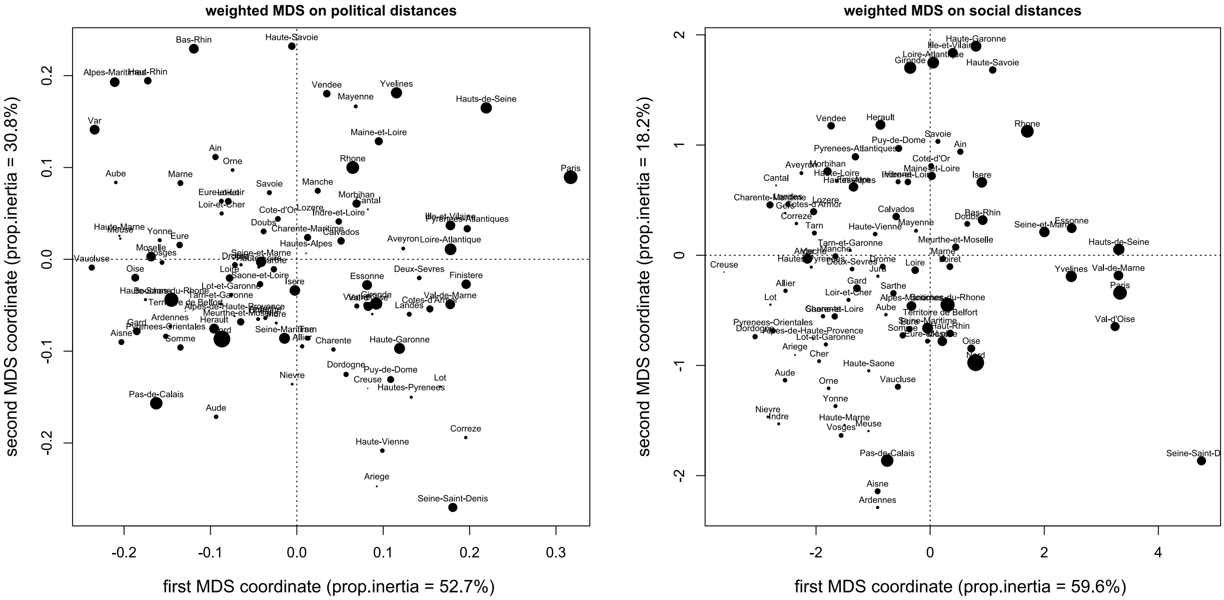

Fig. 2. Left: weighted MDS of the political configuration $(\mathbf{f}, \mathbf{D}_X)$, whose first factor expresses mainly a political right-left gradient. Right: weighted MDS of the social configuration $(\mathbf{f}, \mathbf{D}_Y)$, whose first factor opposes ageing departments (left) to departments with a strong foreign population and natural balance (right).

the systematic Weingarten approach of Sections 3.6 and 3.7: a circumstance both expected and relieving, apt to boost confidence in the results obtained so far.

4. Illustration: comparing political and social configurations of french departments

Consider the $n = 94$ continental French departments, together with the $n \times p$ contingency table $\mathbf{N} = (n_{ik})$ counting the number of votes n_{ik} obtained in department i for candidate k, among the $p = 10$ candidates of the 2012 presidential primary election [22]. The first weighted configuration $(\mathbf{f}, \mathbf{D}_X)$ of the n objects is defined by $f_i = n_{i\bullet}/n_{\bullet\bullet}$ (proportion of voters in the ith department), with a squared Euclidean political distance $\mathbf{D}_X$ defined by the chi-square dissimilarity [see, e.g., 19,26]:

$$D_{ij}^X = \sum_{k=1}^{p} \frac{n_{\bullet\bullet}}{n_{\bullet k}} \left(\frac{n_{ik}}{n_{i\bullet}} - \frac{n_{jk}}{n_{j\bullet}} \right)^2 \quad .$$

Weighted MDS on the corresponding kernel $\mathbf{K}_X$ turns out to be equivalent to the Simple Correspondence Analysis of $\mathbf{N}$. Fig. 2 left depicts the resulting MDS coordinates in dimensions $\alpha \in \{1, 2\}$.

The above configuration will be compared to a rudimentary social configuration $(\mathbf{f}, \mathbf{D}_Y)$, where $\mathbf{f}$ still represents the department share of voters, and $\mathbf{D}_Y$ is the squared Euclidean dissimilarity constructed from the $n \times q$ departmental profiles $\mathbf{Y} = (y_{ik})$ made of the $q = 5$ standardized proportions of "natural demographic balance", "migratory demographic balance", "population over 65", "foreign population" and "unactive young people" around 2015 [22], namely

$$D_{ij}^Y = \sum_{k=1}^{q} (y_{ik} - y_{jk})^2 \quad .$$

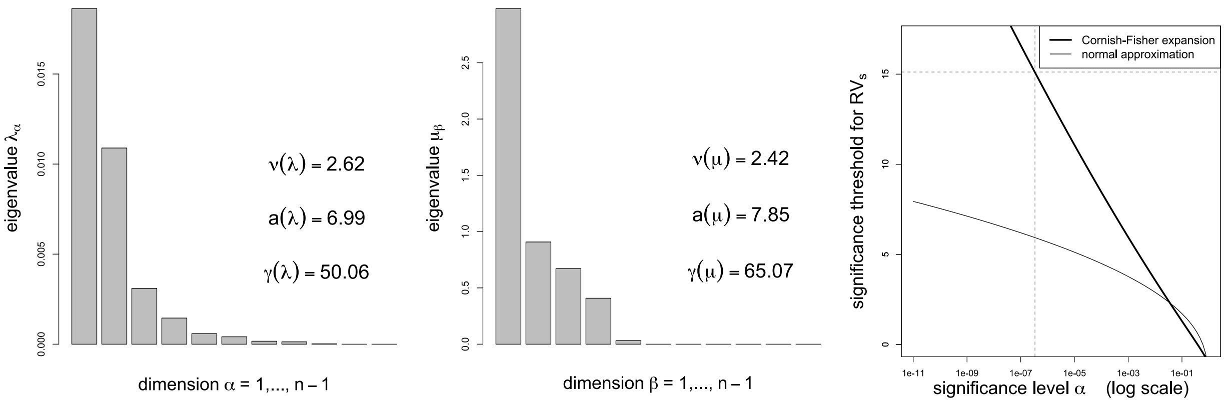

Fig. 3. Scree plots of the political kernel $\mathbf{K}_X$ (left) and of the social kernel $\mathbf{K}_Y$ (middle), together with the values for effective dimensionality, spectral asymmetry and spectral excess kurtosis. Right: significance threshold (r.h.s. of (18), thick line) and normal approximation $u_{1-\alpha}$ (thin line) for the standardized RV (l.h.s. of (18)).

Weighted MDS on the corresponding kernel $\mathbf{K}_Y$ turns out to be equivalent to the Principal Component Analysis of the (weighted) $q \times q$ correlation matrix $\mathbf{R}$. Fig. 2 right depicts the resulting MDS coordinates in dimensions $\alpha \in \{1, 2\}$.

Fig. 3 depicts the scree plots of the eigenvalues of $\mathbf{K}_X$ (left) and $\mathbf{K}_Y$ (middle), together with the associated values for effective dimensionality ν, spectral asymmetry a and spectral excess kurtosis γ.

The value (6) of the coefficient turns out to be $\mathtt{RV} = 0.2496$. Expected moments obtain from the spectral moments by means of formulas (12) to (15), with the result

$$\mathbb{E}(\mathtt{RV}) = 0.02707\,, \quad \mathbb{V}\mathrm{ar}(\mathtt{RV}) = 0.0002166\,, \quad \mathbb{A}(\mathtt{RV}) = 1.644\,, \quad \mathbb{\Gamma}(\mathtt{RV}) = 4.497$$

yielding a standardized value (z-score) as high as $\mathtt{RV}_s = 15.12$. Fig. 3 (right) depicts the significance threshold (r.h.s. of (18)) as a function of α (thick curve), as well as its normal approximation $u_{1-\alpha}$ (thin curve). As expected, the similarity between the political and social configurations is extremely significant, with a resulting p-value (i.e., the α for which both sides of (18) coincide) as low as $p = 3.3 \cdot 10^{-7}$ (Cornish–Fisher expansion), yet much larger than the p-value $p = 6.0 \cdot 10^{-52}$ obtained under normal approximation.

5. Discussion and conclusion

The weighted RV coefficient measures the similarity between two weighted Euclidean configurations, and this contribution proposes exact expressions for the first four moments of the RV. Considering weighted objects extends the traditional uniform framework. It also provides precious guidance for separating the trivial and non-trivial eigenspaces resulting form the spectral decomposition of the standard kernels occurring in the weighted multidimensional scaling of both configurations.

Our approach, invariant orthogonal integration, is nonparametric, and consists in averaging the relative orientation of the kernel eigenvectors of both configurations by performing Haar integration on orthogonal matrices $\mathbf{W} \in \mathbb{O}_{n-1}$ acting in the non-trivial eigenspace only. The resulting expressions are simpler and easier to interpret than their traditional counterparts obtained by averaging on permutation matrices $\mathbf{S}$ between n objects. In view of $\mathbf{S}\mathbf{S}^\top = \mathbf{I}_n$, permutations also do constitute orthogonal transformations, but in $\mathbb{O}_n$, and their undiscriminate use is furthermore questionable in the weighted setting. Comparing the present approach to parametric approaches, typically postulating a multivariate normal distribution for the object features, is left open for future investigations.

Also, our approach is object-oriented, as in traditional Data Analysis and Machine Learning, rather than variable-oriented as in Mathematical Statistics. Its use requires to dispose of squared Euclidean dissimilarities between objects, possibly weighted, and some of its numerous applications (including spatial autocorrelation and network clustering) will be illustrated in forthcoming publications. This contribution underlines in particular the key role played by the standard kernel, central to weighted multidimensional scaling, and whose spectrum governs the values of the RV moments. Correlatively, it appears that the humble scree plot should deserve more consideration, beyond its limited role in selecting the number of factors: mentioning and interpreting its effective dimensionality, spectral skewness and spectral excess kurtosis could arguably become more systematic in practice.

Computing the fourth RV moment did require to recourse to the Weingarten calculus, whose apparatus, arguably demanding for the neophyte, turned out decisive for the pursuit of our objective. One may reasonably hope that future developments along that line will enrich the present results, replacing in particular the mechanical computation of Tables 1 and 2 by true mathematical arguments. However, determining the analytical, exact null distribution of RV, may reveal itself out of reach: as a matter of fact, the moment generating function (25) is an orthogonal analog of the celebrated

Harish-Chandra trace integral for the unitary group, whose analytical expression has been determined ever since the fifties [20] (see also, e.g., Tao [34] and McSwiggen [28]). Yet, discovering a corresponding expression for the orthogonal case, precisely, has not been achieved so far.

Acknowledgments

The author warmly thanks Guillaume Guex as well as an anonymous reviewer for their remarks and suggestions, having contributed to improve the clarity of this paper.

Appendix

Proof of Lemma 1. Let $\mathbf{W}=(w_{\alpha a})\in\mathbb{O}_{n-1}$. Let $\alpha\neq\gamma$ and consider the matrix $\widetilde{\boldsymbol{W}}$ with components

$$\tilde{w}_{\alpha a}=\cos\xi\, w_{\alpha a}-\sin\xi\, w_{\gamma a}\,,\quad \tilde{w}_{\gamma a}=\sin\xi\, w_{\alpha a}+\cos\xi\, w_{\gamma a}\,,\quad \tilde{w}_{\beta a}=w_{\beta a}\,,\ \beta\neq\alpha,\gamma\,,$$

for any a, where ξ is an arbitrary, fixed angle. Then $\widetilde{\boldsymbol{W}}$ is an orthogonal matrix, as likely as $\boldsymbol{W}$, that is $d\mu(\widetilde{\boldsymbol{W}})=d\mu(\boldsymbol{W})$. To ease the notations, take $\boldsymbol{\varepsilon}$ and $\boldsymbol{e}$ in Lemmas 1 and 2, playing no active role in what follows, as empty. Then, by (26)

$$\begin{aligned}\mathcal{I}^{abcd}_{\alpha\alpha\gamma\gamma}&=\mathbb{E}([\cos\xi\, w_{\alpha a}-\sin\xi\, w_{\gamma a}][\cos\xi\, w_{\alpha b}-\sin\xi\, w_{\gamma b}][\sin\xi\, w_{\alpha c}+\cos\xi\, w_{\gamma c}][\sin\xi\, w_{\alpha d}+\cos\xi\, w_{\gamma d}])\\&=2\cos^2\xi\sin^2\xi\,\mathcal{I}^{abcd}_{\alpha\alpha\alpha\alpha}+(\cos^4\xi+\sin^4\xi)\mathcal{I}^{abcd}_{\alpha\alpha\gamma\gamma}-2\cos^2\xi\sin^2\xi\,\mathcal{I}^{abcd}_{\alpha\gamma\alpha\gamma}-2\cos^2\xi\sin^2\xi\,\mathcal{I}^{abcd}_{\alpha\gamma\gamma\alpha}\ .\end{aligned}$$

Multiplying the l.h.s. by $\cos^4\xi+\sin^4\xi+2\cos^2\xi\sin^2\xi=1$ and simplifying yields

$$\mathcal{I}^{abcd}_{\alpha\alpha\alpha\alpha}=\mathcal{I}^{abcd}_{\alpha\alpha\gamma\gamma}+\mathcal{I}^{abcd}_{\alpha\gamma\alpha\gamma}+\mathcal{I}^{abcd}_{\alpha\gamma\gamma\alpha}\ .\quad\square$$

Proof of Lemma 3. Lemmas 1 and 2 entail the following relations between orthogonal coefficients (38)

$$\begin{aligned}&S\overset{(28)}{=}U+2V\,,\quad T\overset{(28)}{=}V+2W\,,\quad M\overset{(28)}{=}3N\,,\quad P\overset{(28)}{=}3S\,,\quad R\overset{(28)}{=}3T\,,\quad E\overset{(29)}{=}(n-2)M+L\,,\\&E\overset{(29)}{=}(n-2)P+M\,,\quad F\overset{(29)}{=}(n-3)N+2M\,,\quad F\overset{(29)}{=}(n-3)S+2Q\,,\quad F\overset{(29)}{=}(n-2)Q+M\,,\\&G\overset{(29)}{=}(n-3)U+2S\,,\quad H\overset{(29)}{=}(n-3)V+2T\,,\quad 0\overset{(29)}{=}(n-2)R+M\,,\quad 0\overset{(29)}{=}(n-2)T+N\,,\\&0\overset{(29)}{=}(n-3)W+2T\end{aligned}$$

with solution (recall that E,F,G,H in (32) are already known)

$$\begin{aligned}&L=\frac{15(n-2)}{n+3}\kappa\,,\qquad M=\frac{3(n-2)}{n+3}\kappa\,,\qquad N=\frac{n-2}{n+3}\kappa\,,\qquad P=\frac{3(n+2)}{n+3}\kappa\,,\\&Q=\frac{n}{n+3}\kappa\,,\qquad R=\frac{-3}{n+3}\kappa\,,\qquad S=\frac{n+2}{n+3}\kappa\,,\qquad T=\frac{-1}{n+3}\kappa\,,\\&U=\frac{n^2+n-4}{(n-3)(n+3)}\kappa\,,\qquad V=\frac{-(n+1)}{(n-3)(n+3)}\kappa\,,\qquad W=\frac{2}{(n-3)(n+3)}\kappa\ .\end{aligned}\tag{53}$$

Consider first $\alpha=\gamma=\varepsilon$, and assume the sub-indices of the orthogonal coefficients to be matched into three distinct pairs. There are $5\times3=15$ such pairings, namely

$$\begin{aligned}\mathcal{I}^{abcdef}_{\alpha\alpha\alpha\alpha\alpha\alpha}=&N\{\delta_{ab}\delta_{cd}\delta_{ef}+\delta_{ac}\delta_{bd}\delta_{ef}+\delta_{ad}\delta_{bc}\delta_{ef}+\delta_{ae}\delta_{cd}\delta_{bf}+\delta_{af}\delta_{cd}\delta_{be}+\delta_{ab}\delta_{ce}\delta_{df}+\delta_{ab}\delta_{cf}\delta_{de}\\&+\ \delta_{ac}\delta_{be}\delta_{df}+\delta_{ac}\delta_{bf}\delta_{de}+\delta_{ad}\delta_{ce}\delta_{bf}+\delta_{ad}\delta_{cf}\delta_{be}+\delta_{ae}\delta_{bc}\delta_{df}+\delta_{ae}\delta_{bd}\delta_{cf}+\delta_{af}\delta_{bc}\delta_{de}+\delta_{af}\delta_{bd}\delta_{ce}\}\ .\end{aligned}\tag{54}$$

In (54), the first term preserves the three pairs in the reference partition $(ab|cd|ef)$, the next six terms preserve one pair only, and the eight remaining terms mix all pairs. It turns out that (54) also holds for coinciding pairs in view of $M=3N$ and $L=5M$. By (26)

$$\mathbb{E}(P_{\alpha\beta}P_{\alpha\delta}P_{\alpha\zeta})=(1+2\delta_{\beta\delta}+2\delta_{\beta\zeta}+2\delta_{\delta\zeta}+8\delta_{\beta\delta}\delta_{\beta\zeta}\delta_{\delta\zeta})N\ .\tag{55}$$

Consider now $\alpha=\gamma\neq\varepsilon$. Distinguishing between cases preserving or not the pair (ef) yields

$$\begin{aligned}\mathcal{I}^{abcdef}_{\alpha\alpha\alpha\alpha\varepsilon\varepsilon}=&S\{\delta_{ab}\delta_{cd}\delta_{ef}+\delta_{ac}\delta_{bd}\delta_{ef}+\delta_{ad}\delta_{bc}\delta_{ef}\}+T\{\delta_{ae}\delta_{cd}\delta_{bf}+\delta_{af}\delta_{cd}\delta_{be}+\delta_{ab}\delta_{ce}\delta_{df}+\delta_{ab}\delta_{cf}\delta_{de}\\&+\delta_{ac}\delta_{be}\delta_{df}+\delta_{ac}\delta_{bf}\delta_{de}+\delta_{ad}\delta_{ce}\delta_{bf}+\delta_{ad}\delta_{cf}\delta_{be}+\delta_{ae}\delta_{bc}\delta_{df}+\delta_{ae}\delta_{bd}\delta_{cf}+\delta_{af}\delta_{bc}\delta_{de}+\delta_{af}\delta_{bd}\delta_{ce}\}\end{aligned}$$

which also holds for three preserved pairs since $3S+12T=M$. By (26)

$$\alpha=\gamma\neq\varepsilon\,,\qquad \mathbb{E}(P_{\alpha\beta}P_{\gamma\delta}P_{\varepsilon\zeta})=(1+2\delta_{\beta\delta})S+(2\delta_{\beta\zeta}+2\delta_{\delta\zeta}+8\delta_{\beta\delta}\delta_{\beta\zeta}\delta_{\delta\zeta})T\tag{56}$$

$$\alpha=\varepsilon\neq\gamma\,,\qquad \mathbb{E}(P_{\alpha\beta}P_{\gamma\delta}P_{\varepsilon\zeta})=(1+2\delta_{\beta\zeta})S+(2\delta_{\beta\delta}+2\delta_{\delta\zeta}+8\delta_{\beta\delta}\delta_{\beta\zeta}\delta_{\delta\zeta})T\tag{57}$$

$$\gamma=\varepsilon\neq\alpha\,,\qquad \mathbb{E}(P_{\alpha\beta}P_{\gamma\delta}P_{\varepsilon\zeta})=(1+2\delta_{\delta\zeta})S+(2\delta_{\beta\delta}+2\delta_{\beta\zeta}+8\delta_{\beta\delta}\delta_{\beta\zeta}\delta_{\delta\zeta})T\ .\tag{58}$$

In the remaining case $\alpha \neq \gamma \neq \varepsilon$, the same reasoning yield

$$\mathcal{I}^{abcdef}_{\alpha\alpha\gamma\gamma\varepsilon\varepsilon} = U\{\delta_{ab}\delta_{cd}\delta_{ef}\} + V\{\delta_{ac}\delta_{bd}\delta_{ef}+\delta_{ad}\delta_{bc}\delta_{ef}+\delta_{ae}\delta_{cd}\delta_{bf}+\delta_{af}\delta_{cd}\delta_{be}+\delta_{ab}\delta_{ce}\delta_{df}+\delta_{ab}\delta_{cf}\delta_{de}\}$$
$$+\ W\{\delta_{ac}\delta_{be}\delta_{df}+\delta_{ac}\delta_{bf}\delta_{de}+\delta_{ad}\delta_{ce}\delta_{bf}+\delta_{ad}\delta_{cf}\delta_{be}+\delta_{ae}\delta_{bc}\delta_{df}+\delta_{ae}\delta_{bd}\delta_{cf}+\delta_{af}\delta_{bc}\delta_{de}+\delta_{af}\delta_{bd}\delta_{ce}\}$$

also valid for three preserved pairs since $U + 6V + 8W = N$, and finally for $\alpha \neq \gamma \neq \varepsilon$

$$\mathbb{E}(P_{\alpha\beta}P_{\gamma\delta}P_{\varepsilon\zeta}) = U + (2\delta_{\beta\delta} + 2\delta_{\beta\zeta} + 2\delta_{\delta\zeta})V + 8\delta_{\beta\delta}\delta_{\beta\zeta}\delta_{\delta\zeta}W \ . \tag{59}$$

To ease notations, use definitions (41), multiply both sides of (55) by φ, of (56) by $\delta_{\alpha\gamma}(1-\delta_{\alpha\varepsilon})(1-\delta_{\gamma\varepsilon}) = \delta_{\alpha\gamma} - \varphi$, of (57) by $\delta_{\alpha\varepsilon} - \varphi$, of (58) by $\delta_{\gamma\varepsilon} - \varphi$, of (59) by $(1-\delta_{\alpha\gamma})(1-\delta_{\alpha\varepsilon})(1-\delta_{\gamma\varepsilon}) = 1 - \sigma + 2\varphi$, and add the whole to obtain the unrestricted expression (39). □

References

[1] H. Abdi, Congruence: Congruence coefficient, RV coefficient, and Mantel coefficient, Encyclopedia Res. Des. 3 (2010) 222–229.
[2] C.-O. Amédée-Manesme, F. Barthélémy, D. Maillard, Computation of the corrected Cornish-Fisher expansion using the response surface methodology: application to VaR and CVaR, Ann. Oper. Res. 281 (1) (2019) 423–453.
[3] S. Aubert, C. Lam, Invariant integration over the unitary group, J. Math. Phys. 44 (12) (2003) 6112–6131.
[4] T. Banica, The orthogonal Weingarten formula in compact form, Lett. Math. Phys. 91 (2) (2010) 105–118.
[5] F. Bavaud, On the Schoenberg transformations in data analysis: Theory and illustrations, J. Classification 28 (2011) 297–314.
[6] F. Bavaud, Testing spatial autocorrelation in weighted networks: the modes permutation test, J. Geogr. Syst. 15 (3) (2013) 233–247.
[7] G. Birkhoff, Three observations on linear algebra, Univ. Nac. Tacuman Rev. Ser. A 5 (1946) 147–151.
[8] I. Borg, P.J. Groenen, Modern Multidimensional Scaling: Theory and Applications, Springer Science & Business Media, 2005.
[9] D. Braun, Invariant integration over the orthogonal group, J. Phys. A: Math. Gen. 39 (47) (2006) 14581.
[10] B. Collins, S. Matsumoto, On some properties of orthogonal Weingarten functions, J. Math. Phys. 50 (11) (2009) 113516.
[11] B. Collins, S. Matsumoto, J. Novak, The Weingarten calculus, Not. Am. Math. Soc. 69 (5) (2022).
[12] B. Collins, P. Śniady, Integration with respect to the Haar measure on unitary, orthogonal and symplectic group, Comm. Math. Phys. 264 (3) (2006) 773–795.
[13] C. Cortes, M. Mohri, A. Rostamizadeh, Algorithms for learning kernels based on centered alignment, J. Mach. Learn. Res. 13 (2012) 795–828.
[14] G. Csardi, T. Nepusz, et al., The igraph software package for complex network research, InterJ. Complex Syst. 1695 (5) (2006) 1–9.
[15] C. Cuadras, J. Fortiana, Weighted continuous metric scaling, in: A.K. Gupta, V. Girko (Eds.), Multidimensional Statistical Analysis and Theory of Random Matrices, in: Proceedings of the Sixth Eugene Lukacs Symposium, De Gruyter, 1996, pp. 27–40.
[16] P. Diaconis, P.J. Forrester, Hurwitz and the origins of random matrix theory in mathematics, Random Matrices: Theory Appl. 6 (01) (2017) 1730001.
[17] Y. Escoufier, Le traitement des variables vectorielles, Biometrics 29 (4) (1973) 751–760.
[18] S. Geisser, S.W. Greenhouse, An extension of Box's results on the use of the F distribution in multivariate analysis, Ann. Math. Stat. 29 (3) (1958) 885–891.
[19] M. Greenacre, Correspondence Analysis in Practice, Chapman and Hall, 2017.
[20] Harish-Chandra, Differential operators on a semisimple Lie algebra, Amer. J. Math. (1957) 87–120.
[21] M. Heo, K. Ruben Gabriel, A permutation test of association between configurations by means of the RV coefficient, Comm. Statist. Simulation Comput. 27 (3) (1998) 843–856.
[22] INSEE, Institut national de la statistique et des études économiques (National institute of statistics and economic studies), 2022, https://www.insee.fr/en/accueil, Accessed: 2022.
[23] J. Josse, J. Pagès, F. Husson, Testing the significance of the RV coefficient, Comput. Statist. Data Anal. 53 (1) (2008) 82–91.
[24] F. Kazi-Aoual, S. Hitier, R. Sabatier, J.-D. Lebreton, Refined approximations to permutation tests for multivariate inference, Comput. Stat. Data Anal. 20 (6) (1995) 643–656.
[25] M. Kendall, A. Stuart, The Advanced Theory of Statistics. Vol. 1: Distribution Theory, Griffin, London, 1977.
[26] L. Lebart, A. Morineau, K.M. Warwick, Multivariate Descriptive Statistical Analysis; Correspondence Analysis and Related Techniques for Large Matrices, Wiley, 1984.
[27] S. Matsumoto, General moments of the inverse real Wishart distribution and orthogonal Weingarten functions, J. Theoret. Probab. 25 (3) (2012) 798–822.
[28] C. McSwiggen, The Harish-Chandra integral: An introduction with examples, L'Enseignement Math. 67 (3) (2021) 229–299.
[29] P.W. Mielke, Meteorological applications of permutation techniques based on distance functions, Handbook of Statist. 4 (1984) 813–830.
[30] J.A. Mingo, M. Popa, Real second order freeness and Haar orthogonal matrices, J. Math. Phys. 54 (5) (2013) 051701.
[31] J.A. Mingo, R. Speicher, Free Probability and Random Matrices, in: Fields Institute Monographs, Vol. 35, Springer, 2017.
[32] P. Robert, Y. Escoufier, A unifying tool for linear multivariate statistical methods: the RV-coefficient, J. R. Stat. Soc. Ser. C. Appl. Stat. 25 (3) (1976) 257–265.
[33] P. Schlich, Defining and validating assessor compromises about product distances and attribute correlations, in: Data Handling in Science and Technology, Vol. 16, Elsevier, 1996, pp. 259–306.
[34] T. Tao, The Harish-Chandra-Itzykson-Zuber integral formula, What's new (blog) (2013) https://terrytao.wordpress.com/2013/02/08/the-harish-chandra-itzykson-zuber-integral-formula.
[35] K.J. Worsley, K.J. Friston, Analysis of fMRI time-series revisited—again, Neuroimage 2 (3) (1995) 173–181.
[36] Y. Yamamoto, S. Kudo, Probabilistic analysis of an estimator for the Frobenius norm of a matrix product, JSIAM Lett. 9 (2017) 9–12.
[37] H. Zhang, J. Tian, J. Li, J. Zhao, RV-coefficient and its significance test in mapping brain functional connectivity, in: Medical Imaging 2009: Biomedical Applications in Molecular, Structural, and Functional Imaging, Vol. 7262, SPIE, 2009, pp. 627–635.