Measures of dependence between random vectors and tests of independence.

Julie Josse\textsuperscript{a}, Susan Holmes\textsuperscript{b}

\textsuperscript{a} Applied Mathematics department, Agrocampus Ouest, Rennes, France.  
\textsuperscript{b} Statistics department, Stanford University, California.

Abstract

The simple correlation coefficient between two variables has been generalized to measures of association between two matrices several times. Coefficients such as the RV coefficient, the distance covariance (dCov) coefficient or the Hilbert Schmidt Information Criterion (HSIC) have all been adopted by different communities. Scientists also use tests to measure whether two random variables are linked and then interpret the coefficients in context; many branches of science currently need multiway measures of association.

The aim of this paper is to provide a small state of the art on the topic of measures of dependence between random vectors and tests of independence and to show links between different approaches. We document some of the interesting rediscoveries and lack of interconnection between bodies of literature.

This review starts with a short history of randomization tests using distance matrices and some motivating examples. We then provide definition of the coefficients and associated tests. Finally we review a few of the recent modifications that have been proposed that provide improved properties and enhance ease of interpretation, as well as some prospective directions for future research.

Keywords: measures of relationship between matrices, RV coefficient, dCov coefficient, kernel methods, tests of independence, permutation tests

1 Introduction

Studying and assessing the relationship between two matrices can be traced back at least to Knox (1964). Following this work and work by David and Barton (1962); Barton and David (1962) on testing clumping in spatial data, David and Barton (1966) used the comparison of two distances, one distance measuring the differences in time between disease occurrences, the other measuring the spatial distance between events enabling the detection of epidemics in diseases such as leukemia. If the two distances concurred, it indicated a high chance of an occurrence of the epidemic. This was done by vectorizing the matrices of distances, computing the correlation coefficient between the two vectors and using a permutation test to detect significant values. Mantel (1967b) wrote a review paper on the subject and made suggestions on improving the test. His name is now associated to this method of randomized testing between two distances which is widely used in ecology and biometry (Sokal and Sneath, 1963).

Today, applied statisticians study relationships across two matrices of data or two sets of variables measured on the same samples in many different contexts. For instance, in sensory analysis, the same products (such as wines, yogurts or fruit) can be described by both sensory descriptor variables (such as bitterness, sweetness or texture) and physical-chemical measurements (such as pH, NaCl or sugars). Scientists often need ways of comparing the sensory profile with the physical-chemical’s one (Génard et al., 1994; Pagès and Husson, 2005).

In biology, the emergence of many new technologies has generated heterogeneous data collected on the same samples. For instance, information can be available at both the genome level (with CGH data) and at the transcriptome level (with expression data) for the same sample of tumors and one may want to study the similarity between the two groups of variables (de Tayrac et al., 2009; Witten et al., 2009).
Studying the relationship between two such sets can enable the replacement of an expensive set of measurements by inexpensive ones, often called instrumental variables for instance in Rao (1964).

As suggested by the original work cited above on detection of diseases using distances in space and time, when dealing with two sets of measurements, one may want to first identify if there is a significant relationship between the two sets by using a test. Then, the coefficient of association can be interpreted even if as pointed out by Reimherr and Nicolae (2013), the task is not easy.

Several different coefficients have been published as measures of association between groups of variables. Recently, Szekely et al. (2007) introduced the distance covariance (dCov) coefficient which has the property of being equal to zero if and only if the random vectors are independent. In the machine learning community, kernel based measures of independence have been developed. Sejdinovic et al. (2013) made the link between one of them, the Hilbert-Schmidt Independence Criterion (HSIC) presented in Gretton et al. (2005) and the dCov coefficient. This literature on the dCov coefficient and on the kernel based coefficients, however, seems to have overlooked the literature on the RV coefficient despite many common features. The RV coefficient presented in Escofier (1973) can be seen as an early instance of a natural generalization of the notion of correlation to groups of variables. This coefficient has been a standard measurement in sensory analyses for many years (Schlich 1996; Risvik et al. 1997; Noble and Ebeler 2002; Giacalone et al. 2013; Cadena et al. 2013). It has also been successfully applied in many fields such as morphology (Klingenberg 2009; Fruciano et al. 2013; Santana and Lofgren 2013; Foth et al. 2013), neuroscience where Shinkareva et al. (2008) and Abdi (2010) used it to compute the level of association between stimuli and brain images captured using fMRI and transcriptomics where, for instance, Culhane et al. (2003) used it to assess the similarity of expression data coming from different technologies.

We present a short survey of measures of dependence between random vectors and association tests. Our review is far from exhaustive, as we only focus on three main classes of coefficients. First, we consider linear relationships by defining the RV coefficient in section 2. We note its properties and present asymptotic tests as well as permutation tests to assess its significance. Then, we present two modified versions of the RV coefficient recently proposed that aim at correcting the bias of the RV coefficient when the covariance between the two sets is null. Such bias was shown to be exacerbated in high dimensions. Section 3 focuses on the detection of non-linear relationships using the dCov coefficient. For each type of coefficient, we cover the same topics (asymptotic tests, permutation tests, modified coefficients). A small simulation study comparing the performance of these coefficients is shown in Section 3. The RV coefficient and the dCov coefficient rely on Euclidean distances. We also include coefficients based on other distances in Section 4, presenting kernel based coefficients such as the HSIC coefficient. As an aside it seemed interesting to look at the citation record of papers covering the subject, showing that different disciplines have adopted different types of coefficients with strong within discipline preferences.

2 The RV coefficient

2.1 Definition

Let us consider two random vectors $X$ in $\mathbb{R}^p$ and $Y$ in $\mathbb{R}^q$. The aim is to study and test the association between these two vectors. The matrices $X_{n \times p}$ and $Y_{n \times q}$ represent $n$ independent realizations of the random vectors and are assumed to be centered.

The rationale of the RV coefficient is to consider that two sets of variables are correlated if the relative position of the samples in one set is similar to the relative position of the samples in the other set. The matrices representing the relative positions of the samples are the cross-product matrices: $W_X = XX'$ and $W_Y = YY'$. They are of size $n \times n$ and so can be compared directly. To measure their proximity, the inner product between matrices is computed:

$$< W_X, W_Y > = tr(XX'YY') = \sum_{l=1}^{p} \sum_{m=1}^{q} \text{cov}^2(X_{.l}, Y_{.m})$$ (1)
Since the two matrices may have different norms, a generalized correlation coefficient is computed by normalizing by the matrix norms. The RV coefficient of Escoufier (1973) is defined as a correlation coefficient between the two cross-product matrices:

$$RV(X, Y) = \frac{\langle W_X, W_Y \rangle}{\|W_X\| \|W_Y\|} = \frac{tr(XX'YY')}{\sqrt{tr(XX')^2tr(YY')^2}}$$

(2)

This represents the cosine of the angle between the two vectors representing the cross-product matrices. It may be convenient to write the RV coefficient in a different way to better understand its properties, for instance using the covariance matrices: $RV(X, Y) = \frac{tr(S_{XY}S_{YX})}{\sqrt{tr(S_{XX})tr(S_{YY})}}$, with $S_{XY} = X'Y'$ being the empirical covariance matrix between $X$ and $Y$. It is also possible to express the coefficient using distance matrices. More precisely, let $\Delta_{n \times n}$ be the matrix where element $d_{ij}$ represents the Euclidean distance between the samples $i$ and $j$, $d_i.$ and $d_j.$ being the mean of the row $i$ and the mean of column $j$ and $d_.$ being the global mean of the distance matrix. Using the formulae relating the inner product and the euclidean distance between two samples (Schoenberg, 1935; Gower, 1966), $W_{ij} = -\frac{1}{2}(d_{ij}^2 - d_i.^2 - d_j.^2 + d_.^2)$, the RV coefficient (2) can be written as:

$$RV(X, Y) = \frac{\langle C\Delta X^2 C, C\Delta Y^2 C \rangle}{\|C\Delta X^2 C\| \|C\Delta Y^2 C\|}$$

(3)

with $C = I_n - \frac{1n^T_n}{n}$, $I_n$ the identity matrix of order $n$ and $1_n$ a vector of ones of size $n$. The numerator of (3) is the inner product between the double centered (by rows and by columns) squared Euclidean distance matrices. This last expression (3) will be important for the sequel of the paper since it allows an easiest comparison with other coefficients.

The properties of the RV coefficient are:

- it is consistent: it converges to its population counterpart $\rho V$ when $n \to \infty$
- for $p = q = 1$, $RV = r^2$ the square of the simple correlation coefficient
- $0 \leq RV(X, Y) \leq 1$
- $RV(X, Y) = 0$ if and only if $X'Y = 0$: all the variables of one group are orthogonal to all the variables in the other group
- $RV(X, aBX + c) = 1$, with $B$ an orthogonal matrix, $a$ a constant and $c$ a constant vector

Remarks:

1. If the column-variables of both matrices $X$ and $Y$ are standardized (thus have unit variance), the numerator of the RV coefficient (1) is equal to the sum of the squared correlations between the variables of the first group and the variables of the second group. The “pre-processing” step is as usual a crucial step in the analysis which implies a different value for the coefficient.

2. It is interesting to note that the principal component analysis (PCA) problem i.e. finding the matrix of rank $k$ that approximates $X$ the best in the least squares sense, can also be rephrased as the problem of finding the rank $k$ matrix which is the most correlated to $X$ as measured by the RV matrix correlation (Robert and Escoufier, 1976).
2.2 Tests

As with the simple correlation coefficient, a high value of the RV coefficient does not necessarily imply a significant relationship between the two sets of measurement. Indeed, as we will show in Section 2.2.2, the value of the RV coefficient depends on the sample size as well as on the covariance structure of each matrix, hence the need for a valid inferential procedure for testing the significance of the association. One usually sets up the hypothesis test by taking

\[
\begin{align*}
   & H_0 & \rho_V = 0, \text{there is no significant association between the two sets} \\
   & H_1 & \rho_V > 0, \text{there is a significant association between the two sets}
\end{align*}
\]

The test assesses if there is any linear relationship between the two sets. The fact that \( \rho_V = 0 \) (which corresponds to the population covariance matrix \( \Sigma_{XY} = 0 \)) does not necessarily imply independence between \( X \) and \( Y \) (except when they are multivariate normal), only the absence of a linear relationship between them.

2.2.1 Asymptotic tests

Under the null hypothesis, the asymptotic distribution of the \( nRV \) is available when the joint distribution of the random variables is multivariate normal \( \text{\cite{Robert1985}} \) or when it belongs to the class of elliptical distributions \( \text{\cite{Cleroux1989}} \). More precisely, \( nRV \) converges to:

\[
\frac{1 + k}{tr(\Sigma_{XX}^2)tr(\Sigma_{YY}^2)} \sum_{l=1}^{p} \sum_{m=1}^{q} \lambda_l \gamma_m Z_{lm}^2,
\]

where:
- \( k \) is the kurtosis parameter of the elliptical distribution,
- \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \) are the eigenvalues of the covariance matrix \( \Sigma_{XX} \),
- \( \gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_q \) are the eigenvalues of the covariance matrix \( \Sigma_{YY} \),
- \( Z_{lm} \) are i.i.d. \( \mathcal{N}(0,1) \) random variables.

To eliminate the need for any distributional hypotheses, \( \text{\cite{Cleroux1995}} \) suggested a test based on the rank. However, \( \text{\cite{Josse2008}} \) showed that these tests are accurate (providing accurate type I errors) only for large sample size \( (n > 300) \). An alternative is to use permutation tests.

2.2.2 Permutation tests

As we saw in the introduction, the idea of using permutation tests to ascertain a link between two sets of variables are the earliest instances of multi-table association testing.

The simple permutation rows of one matrix and computation of the statistic such as the RV coefficient gives the null distribution of no association. There are \( n! \) possible permutations to consider and the p-value will be the proportion of the values that are greater or equal to the observed coefficient.

Note that care must be taken in the implementation as this is not equivalent to a complete permutation test of the vectorized cross-product matrices for which the exhaustive distribution is much larger \( (n(n-1)/2)! \).

Computing the exact permutation distribution is computationally costly when \( n > 15 \). Consequently, the permutation distribution is usually approximated by Monte Carlo, although a moment matching approach is also possible. The latter consists of approximating the distribution by a continuous distribution without doing any permutations and using the analytical moment of the exact permutation distribution under the null. \( \text{\cite{Kazi-Aoual1995}} \) defined the first moments of the quantity \( 1 \) under the null which yields the moments of the RV coefficient. The expectation is:

\[
\mathbb{E}_{H_0}(RV) = \frac{\sqrt{\beta_x \times \beta_y}}{n - 1} \quad \beta_x = \frac{(tr(XX))^2}{tr((XX)^2)} = \frac{(\sum \lambda_i)^2}{\sum \lambda_i^2}
\]
Equation (5) provides insights into the expected behavior of the RV coefficient. Indeed, the coefficient $\beta_x$ can be seen as a measure of the complexity of the matrix. It varies between 1 when all the variables are perfectly correlated and $p$ when all the variables are orthogonal. Thus, equation (5) shows that under the null, the RV coefficient takes high values when the sample size is small (as with the simple correlation coefficient) and when the data matrices $X$ and $Y$ are very multi-dimensional. The expression of the variance and the skewness are detailed in Josse et al. (2008). With the first three moments, Josse et al. (2008) compared different moment based methods such as the Edgeworth expansions or the Pearson family and pointed out the quality of the Pearson type III approximation for permutation distributions. The RV coefficient is implemented in the R (R Core Team, 2013) packages ade4 (Dray, 2007, function RV.rtest) and FactoMineR where the function coeffRV also provides the Pearson type III approximation to assess its significance (Husson et al., 2013).

2.3 Modified coefficients

In practice, the statistical significance of the association test is not sufficient and one often wants to interpret the value of the coefficient as a measure of the intensity of the relationship. However, as shown using equation (5), a value of the RV coefficient is not informative in itself since it depends among other things on the sample size. For this reason modified versions of the coefficient have been suggested. Smilde et al. (2009) noted that even under the null, the values of the RV coefficient can be very high. Previous results on the expectation of the coefficient (Kazi-Aoual et al., 1995; Josse et al., 2008) went in this direction (see (5)) but Smilde et al. (2009)’s work is not based on these works. They estimated the expected values of the RV coefficient under the null by considering the matrices $X$ and $Y$ as random matrices (with elements drawn from a standard normal distribution) and using results from random matrix theory. They show that the problem can be traced back to the diagonal elements of the matrices $XX'$ and $YY'$. Thus, Smilde et al. (2009) defined a new coefficient, the modified RV, by removing these terms:

$$
RV_{mod}(X, Y) = \frac{\text{tr}((XX' - \text{diag}(XX'))(YY' - \text{diag}(YY')))}{\sqrt{\text{tr}(X'X - \text{diag}(XX'))^2\text{tr}(Y'Y - \text{diag}(XX'))^2}}
$$

This new coefficient can take negative values. They showed in a simulation study that their coefficient has the expected behavior, meaning that even in high dimensional setting ($n = 20$ and $p = q = 100$), the values of the $RV_{mod}$ are around 0 under the null. In addition, for a fixed value of $n$, they simulated two matrices uncorrelated to each other and slowly increased the correlation between the two groups. They showed that the $RV_{mod}$ varies between 0 and 1 whereas the RV varies between 0.85 to 0.99. Thus, they argued that the modified coefficient allows an easier interpretation.

Mayer et al. (2011) extended this work by highlighting the fact that the $RV_{mod}$ (6) is still biased under the null and referred to the results of Kazi-Aoual et al. (1995) and Josse et al. (2008) on the expectation. The rationale of their approach is to replace the simple correlation coefficient $r^2$ in the expression of the RV coefficient (which can be seen in equation (1) when the variables are standardized) by an adjusted coefficient. More precisely, they defined the adjusted RV as:

$$
RV_{adj} = \frac{\sum_{l=1}^{p} \sum_{m=1}^{q} r_{adj}^2(X_{.l}, Y_{.m})}{\sqrt{\sum_{l,l'=1}^{p} r_{adj}^2(X_{.l}, X_{.l'}) \sum_{m,m'=1}^{q} r_{adj}^2(Y_{.m}, Y_{.m'})}}
$$

with $r_{adj}^2 = 1 - \frac{n-1}{n-2}(1 - r^2)$

They suggested testing its significance using permutation tests. The p-values are the same as those obtained using the RV coefficient since the denominator is invariant under permutation and the numerator
is a monotone function. In their simulation study, they focused on the comparison between \(RV_{adj}\) and \(RV_{mod}\) by computing the mean square error (MSE) between the sample coefficients and the population coefficient (\(\rho V\)) and showed the smallest values with their new coefficient. We stress this approach here, as very few papers studying these coefficients refer to a population coefficient.

Both [Smilde et al. (2009)] and [Mayer et al. (2011)] use their coefficients on real data from biology (such as samples described by groups of genes) and emphasized the relevant interpretation from a biological perspective. In addition, [Mayer et al. (2011)] applied a Multidimensional Scaling method on the matrix gathering the adjusted RV coefficients between the groups (more than two groups of genes were studied) to visualize the proximity between the groups. Such an analysis is extremely similar to the earlier STATIS method ([Escoufier 1987]) that uses RV coefficients to compute a compromise eigenstructure on which to project each table. We can also emphasize the steps of the analysis: study the significance of the relationships, the intensity of the links and describe the relationships.

As an aside, note that the idea of removing the diagonal terms of the inner product matrix (as in equation (6)) can also be found in the framework of multiple correspondence analysis (MCA) ([Greenacre and Blasius 2006]). Greenacre ([1988, 1994]) defined joint correspondence analysis to focus only on the non-diagonal part of the Burt matrix to get percentages of variability that are less pessimistic than in MCA.

3 The dCov coefficient

3.1 Definition

[Szekely et al. (2007)] defined a measure of dependence between random vectors: the distance covariance (dCov) coefficient. They show that for all random variables with finite first moments, the dCov coefficient generalizes the idea of correlation in two ways. First, this coefficient can be applied when \(X\) and \(Y\) are of any dimensions, they construe their coefficient as a generalization of the simple case where \(p = q = 1\) without referring to the earlier RV literature. Second, the dCov coefficient is equal to zero, if and only if there is independence between the random vectors. Indeed, a correlation coefficient measures linear relationships and can be equal to 0 even when the variables are related. This can be seen as a major shortcoming of the correlation coefficient.

The dCov coefficient is defined as a weighted \(L^2\) distance between the joint and the product of the marginal characteristic functions of the random vectors. The choice of the weights is crucial and ensures the independence property. The dCov coefficient can also be written in terms of the expectations of Euclidean distances which is easier to interpret:

\[
V^2 = E(|X - X'||Y - Y'|) + E(|X - X'|E(|Y - Y''|)) - 2E(|X - X'||Y - Y'|) \tag{7}
\]

\[
cov(|X - X'|, |Y - Y'|) - 2cov(|X - X'||Y - Y'|) \tag{8}
\]

with \(X', X''\) and \(Y', Y''\) being independent copies of the random variables \(X\) and \(Y\) and \(|X - X'|\) being the Euclidean distance (we stick here with their notations). Expression (8) shows that the covariance of the distances can be equal to 0 with a value for distance covariance not equal to 0 (no independence). Expression (7) implies a straightforward empirical estimate \(V^2_n(X, Y)\) also known as \(\text{dCov}^2_n(X, Y)\):

\[
\text{dCov}^2_n(X, Y) = \frac{1}{n^2} \sum_{i,j=1}^{n} d_{ij}^X d_{ij}^Y + d_{i.}^X d_{.j}^Y - 2 \frac{1}{n} \sum_{i=1}^{n} d_{i.}^X d_{i.}^Y
\]

\[
= \frac{1}{n^2} \sum_{i,j=1}^{n} (d_{ij}^X - d_{i.}^X - d_{.j}^X + d_{..}^X)(d_{ij}^Y - d_{i.}^Y - d_{.j}^Y + d_{..}^Y)
\]
Once \( \text{dCov}_n^2 \) is defined, the correlation coefficient can be defined:

\[
\text{dCor}_n^2(X, Y) = \frac{\langle C \Delta X C, C \Delta Y C \rangle}{\| C \Delta X C \| \| C \Delta Y C \|}
\]

(9)

The only difference between this and the RV coefficient (3) is that Euclidean distances \( \Delta_X \) and \( \Delta_Y \) are used in (9) and not squared Euclidean distances. This difference implies that the dCor coefficient is able to detect non-linear relationships whereas the RV coefficient is restricted to linear ones. Indeed, when the distances are squared, many terms cancel whereas when the distances are not squared no cancellation occurs and allowing more complex associations to be detected.

The properties of the coefficient are:

- it is consistent: it converges to its population counterpart \( R \) when \( n \to \infty \)
- \( p = q = 1 \) with Gaussian distribution: \( \text{dCor}_n \leq |r| \)
- \( 0 \leq \text{dCor}_n(X, Y) \leq 1 \)
- \( R(X, Y) = 0 \) if and only if \( X \) and \( Y \) are independent

Note the similarities to some of the properties of the RV coefficient presented in Section 2.1. Now, as in Section 2 derivations of asymptotic and permutation tests and a definition of a modified coefficient are provided.

3.2 Asymptotic tests

Asymptotic tests are derived to assess if there is any relationship between the two sets. An appealing property of the distance correlation coefficient is that the associated test assesses independence between the random vectors. Szekely et al. (2007) showed that under the null hypothesis of independence, \( n \text{dCor}^2 \) converges in distribution to a quadratic form:

\[
Q = \sum_{j=1}^{\infty} \lambda_j Z_j^2,
\]

where \( Z_j \) are independent standard Gaussian variables and \( \lambda_j \) depend on the distribution of \((X, Y)\). Under the null, the expectation of \( Q \) is equal to 1 and it tends to infinity otherwise. Thus, the null hypothesis is rejected for large values of \( n \text{dCor}^2(X, Y) \). One main feature of this test is that it is consistent against all dependent alternatives whereas some alternatives are ignored in the test based on the RV coefficient (4).

3.3 Permutation tests

Permutation tests are used to assess the significance of the distance correlation coefficient in practice. The coefficient and the test are implemented in the R package energy (Rizzo and Szekely, 2013) in the function \texttt{dcov.test}. Monte Carlo is used to generate a random subset of permutation. Methods based on moment approximations could be considered for this coefficient (see Section 4.1).

3.4 Modified coefficients

As in Smilde et al. (2009), Szekely and Rizzo (2013b) remarked that the \( \text{dCor}_n \) coefficient can takes high values even under independence especially in high-dimensional settings. In addition, they showed that \( \text{dCor}_n \) tends to 1 when \( p \) and \( q \) tend to infinity. Thus, they defined a corrected coefficient \( \text{dCor}^*(X, Y) \) to make the interpretation easier. The rationale is to remove the bias under the null (Szekely and Rizzo, 2013a). The \( \text{dCor}^* \) coefficient can take negative values. Its distribution under the null in the modern setting where \( p \) and \( q \) tend to infinity has been derived and can be used to perform a test. This coefficient and the test are implemented in the function \texttt{dcor.ttest}. 

7
3.5 Generalization

Szekely et al. (2007) showed that the theory still holds when the Euclidean distance $d_{ij}$ is replaced by $d_{ij}^\alpha$ with $0 \leq \alpha < 2$. This means that a whole set of coefficients can be derived and that the tests will still be consistent against all alternatives. Lyons (2013) generalized this result to metric spaces of strong negative type.

3.6 Simulations

To assess the performance of the dCor coefficient, Szekely et al. (2007) used the following simulations. First, matrices $X_{n \times 5}$ and $Y_{n \times 5}$ were generated from a multivariate Gaussian distribution with a within-matrix covariance structure equal to the identity matrix and the covariances between all the variables of $X$ and $Y$ equals to 0.1. We generated 1000 draws and computed the RV test (using the Pearson approximation) as well as the dCov test (using 500 permutations) for each draw. Figure 1, on the left, shows the power of the tests for different sample sizes $n$ demonstrating the similar behavior of the RV (black curve) and dCov (dark blue curve) tests. We also added the tests using different exponents $\alpha = (0.1, 0.5, 1.5)$ and we can remark that it leads to very different performances in term of power.

Then, another data structure was simulated by generating the matrix $Y$ such that $Y_{ml} = log(X_{ml}^2)$ for $m,l = 1, \ldots, 5$ and the same procedure was applied. Results are displayed in Figure 1 on the right. The "usual" dCov test (using $\alpha = 1$), as well as all the other tests are more powerful compared to the RV test in this non-linear setting which is expected.

![Figure 1: Power of the RV test and dCov test. Right: linear case - left: non-linear case. The dCov test is performed using different exponents $\alpha (0.1, 1, 1.5, 2)$ on the Euclidean distances.](image)

4 Moving away from Euclidean distances...

The RV coefficient and the dCov coefficient rely on Euclidean distances. In this section we focus on coefficients based on other distances or dissimilarities.
4.1 The Generalized RV

Minas et al. (2013) highlighted the fact that the data are not always attribute data (with samples described by variables) but often come directly from the form of distances or dissimilarity matrices. The typical situation occurs when the data are coming from graphs such as social networks. They noted that the RV coefficient is only defined for Euclidean distances whereas other distances can be better fitted depending on the nature of the data. They referred for instance to the identity by state distance or the Sokal and Sneath’s distance which are well suited for specific biological data such known as SNP data. To overcome this drawback of the RV coefficient, they defined the generalized RV (GRV) coefficient as follows:

\[
\text{GRV}(X, Y) = \frac{\langle \mathbf{C}\Delta^2_X \mathbf{C}, \mathbf{C}\Delta^2_Y \mathbf{C} \rangle}{\| \mathbf{C}\Delta^2_X \mathbf{C} \| \| \mathbf{C}\Delta^2_Y \mathbf{C} \|}
\]

(10)

where \(\Delta_X\) and \(\Delta_Y\) being arbitrary dissimilarity matrices. The properties of their coefficient depend on the properties of the matrices \(\mathbf{C}\Delta^2_X \mathbf{C}\) and \(\mathbf{C}\Delta^2_Y \mathbf{C}\). If both are positive semi-definite, then GRV varies between 0 and 1; if both have positive or negative eigenvalues then the GRV can take negative values but the value 1 can still be reached; if one is semi-definite positive and the other one not, the value 1 cannot be reached.

To assess the significance of the GRV coefficient, they derived the first three moments of the coefficient based on Kazi-Aoual et al. (1995)'s results and use the Pearson type III approximation of the permutation distribution. They performed simulations, using different choices of distances for each matrix \(X\) and \(Y\). They considered as a strength this flexibility to use different distances for the same object, since different choices highlight different aspects of the data. However, different choices of distance matrices can result in vastly different power for hypothesis testing. Thus, they suggested strategies to aggregate the results which relies on the principle that when many tests reject the null hypothesis, there is definitely an association between the two sets. This work raised the issue of selecting an appropriate distance before assessing the relationship.

Minas et al. (2013) do not refer to the literature on the dCov coefficient as well as the literature on the kernel coefficient (described in Section 4.2).

4.2 Kernel measures

In the machine learning community, similarity measures between kernels are also available. Kernels are similarity matrices and consequently can be computed from attribute data but can also be obtained for other types of data such as graphs, trees, etc. The most famous measure of similarity is defined as the maximum mean discrepancy (MMD) between the joint distribution of two random variables and the product of their marginal distributions. This criterion introduced by Gretton et al. (2005) is called the Hilbert Schmidt independent Criterion (HSIC) and can be written as:

\[
\text{HSIC} = \text{tr}(K_XCK_Y) (11)
\]

with \(K_X\) being a \(n \times n\) matrix defined as \((K_X)_{i,j} = k_X(x_i, x_j)\) (resp. \((K_Y)_{i,j} = k_Y(y_i, y_j)\)). Note that this measure can be seen as a direct extension of the numerator of the RV coefficient since the numerator is the inner product between simple cross-product (kernel) matrices. The asymptotic distribution under the null has the same form as the ones derived for the other coefficients (Sections 2.2.1 and 3.2). 

\[
n\text{HSIC} \sim \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \lambda_i \eta_j Z^2_{ij}
\]

where \(\lambda_i\) and \(\eta_j\) are the eigenvalues of the operators. This provides a test which is consistent against all the alternatives when a characteristic kernel is used. The empirical version of HSIC is also biased. Song et al. (2012) showed that the bias comes from the self interaction terms and defined an unbiased estimator by removing these terms.

Sejdinovic et al. (2013) established the link between the distance covariance coefficient and the HSIC one. In addition, they generalized Lyons (2013)'s work to semi-metric space (where the triangular equality
is not met) of negative type. It means that many kernels can be used (leading to different coefficients) and that the resulting tests are consistent. In the same vein as the work of Minas et al. (2013), they highlighted that this flexibility allows the discovery of different features and kinds of association. Sejdinovic et al. (2013) also performed a simulation study generating random variables that are dependent but uncorrelated. They used Gaussian kernels as well as distance based kernels which are equivalent to using the dCov coefficient with different power on the Euclidean distances. As in Section 3.6, they showed that the different exponents \( \alpha \) led to a wide performance range in term of power.

Others measures of similarity were suggested in the framework of kernels with different starting points. Cristianini et al. (2001) suggested the kernel target alignment coefficient: 

\[
K_{\text{TMA}} = \frac{\text{tr}(K_X K_Y)}{\|K_X\| \|K_Y\|}
\]

where \( K_Y = yy' \) with \( y \) a vector (and not a matrix) of labels (with 0 and 1 for instance). The aim is to assess the similarity between features (in \( K_X \)) and a response \( y \).

Purdom (2006) defined a RV coefficient for the kernels as a direct extension of the RV coefficient for similarities matrices other than \( W_X \) and \( W_Y \). This corresponds to the correlation version of the HSIC (11) which represents the covariance. All these coefficients are mainly implemented in the software MATLAB (MATLAB, 2012).

Recently, Lopez-Paz et al. (2013) suggested studying non-linear dependencies between random coefficients by computing the largest canonical correlation between random non-linear projections of their respective empirical copula-transformations. Their method showed better results in term of power than the ones obtained by the dCov coefficient and other kernel based coefficient on the simulations detailed in Section 3.6.

5 Comments on other coefficients

Many other coefficients of association are available in the literature (Abdi, 2007, 2010) such as the congruence coefficient or the \( L_g \) coefficient (Pagès, 2014). Lazraq and Robert (1988) compared 9 measures of relationship. Many of these coefficients have been completely forgotten, in part because many of the relevant references were written in French. The coefficients that are still used are the ones which are implemented in softwares. We describe in the following sections two of these coefficients, the Procrustes coefficient and we also give details on the Mantel coefficient since as mentioned in the introduction it can be seen as a pioneer on the topic.

5.1 The Procrustes coefficient

The Procrustes coefficient (Gower, 1971) also known as the Lingoes and Schönemann (RLS) coefficient (Lingoes and Schönemann, 1974) is defined as follows:

\[
\text{RLS}(X,Y) = \frac{\text{tr}(XX'YY')^{1/2}}{\sqrt{\text{tr}(X'X)\text{tr}(Y'Y)}}
\]

Its properties are close to the ones of the RV and dCov coefficient. When \( p = q = 1 \), RLS is equal to \(|r|\). It varies between 0 and 1, being equal to 0 when \( X'Y = 0 \) and to 1 when one matrix is equivalent to the other up to an orthogonal transformation. Lazraq et al. (1992) showed that \( \sqrt{pq}\text{RLS}^2 \leq \text{RV} \leq \frac{1}{\sqrt{pq}}\text{RLS}^2 \).

To assess the significance of the RLS coefficient, a permutation test (Jackson, 1995; Peres-Neto and Jackson, 2001) is used. The coefficient and the test are implemented in the R package ade4 (Dray, 2007) in the function \texttt{procuste.randtest} and in the package vegan (Oksanen et al., 2013) in the function \texttt{protest}. Based on some simulations and real datasets, the tests based on the RV and on the Procrustes coefficients are known to give roughly similar results (Dray et al., 2003) in terms of power. The use of Procrustes is spread in morphometrics (Rohlf and Slice, 1990) since the rationale of Procrustes analysis is to find the optimal translation, rotation and dilatation to superimpose configurations of points. Ecologists also refer to this coefficient to assess the relationship between tables (Jackson, 1995).
5.2 The Mantel coefficient

As we noted in the introduction, the Mantel (Mantel, 1967a; Legendre and Fortin, 2010) test uses a distance based coefficient mainly popular in ecology. Given arbitrary dissimilarities matrices, it is defined as:

\[
r_m(X, Y) = \frac{\sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} (d_{X}^{i,j} - \bar{d}_{X})(d_{Y}^{i,j} - \bar{d}_{Y})}{\sqrt{\sum_{i,j,j \neq i}^{n} (d_{X}^{i,j} - \bar{d}_{X})^2 \sum_{i,j,j \neq i}^{n} (d_{Y}^{i,j} - \bar{d}_{Y})^2}}
\]

This is the correlation coefficient between the upper diagonal terms of the distance matrices. Its significance is assessed via permutation tests. The coefficient and the tests are implemented in several R packages: ade4 (Dray, 2007), vegan (Oksanen et al., 2013), ecodist (Goslee and Urban, 2007).

The Mantel coefficient is probably one of the most popular one and consequently many studies suggesting new coefficients compared their performance to the Mantel’s one. For instance, using simulations, Minas et al. (2013) showed that the Mantel test is less powerful than the test based on the GRV coefficient (10). In the same way, Omelka and Hudecová (2013) underlined the superiority of the dCov test over the Mantel test. However, despite its widespread use, some of the properties of the Mantel test are unclear and recently its use is questioned by many studies (Omelka and Hudecová, 2013). Legendre and Fortin (2010) showed that the Mantel coefficient is not equal to 0 when the covariance between the two sets of variables is null and thus can’t be used to detect linear relationships. Non-linear relationships can be discovered but no result is available to determine when.

Nevertheless, the extensive use in ecology and spatial statistics has led to a large number of extensions of the Mantel coefficient. Smouse et al. (1986) proposed a generalization of this ‘matrix-correlation’ so that the tests can account for a third type of variable, i.e. allowing for partial correlations. Recently, the lack of power and high type error rate for this test has been noted, calling into doubt the validity of its use Guillot and Rousset (2013). Note that Szekely and Rizzo (2013c) also considered this extension and suggested a partial correlation coefficient based on the dCov coefficient.

6 Bibliometric study

Of the 370 papers citing the original RV paper (Robert and Escoufier, 1976), almost half are methodological papers which do not have a particular field of application, of the others 40% come from ecology, almost 30% come from food science and sensory analyses, whereas 20% originate from neuroscience, other well represented disciplinary areas are chemometrics, shape analyses and genomics.

If we look at the list of the 7,000 papers citing Mantel (1967b), ranked according to citations, more than half of the books and references are in the ecological and genetic disciplines, with other areas that use spatial statistics intensively are well represented.

Although recent, a hundred of papers cite Szekely et al. (2007) and most of them are theoretical ones.

We can remark that it is a growing field with many applied or theoretical papers published this year.

7 Conclusion

Currently many situations enable researchers to collect several different types of data for each sample. These heterogeneous sources of information can lead to improved power in the statistical analysis and a better explanatory resolution.

Many coefficients of association have been proposed in the literature, several of which are functions of general dissimilarity (or similarity) matrices, very convenient for comparing heterogeneous data. Unfortunately, recent papers have not made many connections to previous work on the topic, often because the research is being done in disjoint disciplines without a common vocabulary. We have seen that researchers
using the RV coefficient, heavily biased towards ecologists and food scientists are often also familiar with the Mantel coefficient, for instance, but are not aware of the more recent dCov coefficient. On the other hand, more recent statistical and machine learning articles mention the dCov coefficient and are often familiar with the kernel literature but not with the other coefficients. In this review, we have emphasized the similarities of the coefficients and their contiguous studies, such as asymptotic tests, permutation tests and extensions of the coefficients.

Another common feature between some of these coefficients is that they can be seen as a “unifying tool” encompassing many methods: a whole set of methods can be derived by maximizing the association coefficient under specific constraints. The first work on this topic is Robert and Escoufier (1976), who showed for instance that the PCA of \( \mathbf{X} \) can be presented as maximizing \( \text{RV}(\mathbf{X}, \mathbf{Y} = \mathbf{X}\mathbf{A}) \) with \( \mathbf{A} \) being an \( n \times k \) matrix under the constraints that \( \mathbf{Y}'\mathbf{Y} \) is diagonal. Discriminant analysis, canonical analysis as well as multivariate regression can also be derived in the same way. In the kernel literature, Kernel PCA, Kernel Canonical Correlation Analysis can also be defined as maximizing the “RV for kernels” between different kernels under constraints (Purdom, 2006).

We can underline the fact that users of the RV coefficient have had 30 years of experience in developing a large array of methods for dealing with multiway tables and heterogeneous multi-table data (Kroonenberg, 2008; Acar and Yener, 2009; Escoufier, 1987; Lavit et al., 1994; Dray, 2007; Lé et al., 2008; Pages, 2014). The success with which these methods have allowed ecologists and food scientists to explore and visualize their complex multi-table data suggests that adapting them to incorporate nonlinear coefficients such as dCov or HSIC could be a worthwhile enterprise. For instance, one can considered the analogous of STATIS for kernels and get as a compromise kernel a linear combination of kernels with optimal weights. It is also worth mentioning that even if at first the RV coefficient is not dedicated to study non-linear relationships, it can be used (and it is used) for such purposes using transformation of the variables such as a transformation into categorical variables.

We have not given details on using the coefficients in the context of binary, categorical or mixed data. There are interesting approaches to treating binary data and a simple correlation coefficient would overweight the commonality of zeros over the more important co-occurrences, thus a Jaccard index approach seems preferable for these types of data. One paper has used the Jaccard as an inspiration in comparing graphs using a Jaccard kernel (Gärtnert et al., 2006). Researchers who use multiple correspondence analyses (Greenacre and Blasius, 2006) have developed special weighting metrics for the contingency table and indicator matrix of dummy variables that replace correlations and variances with chi-square based statistics. With these particular row and column weights, the RV coefficient between two groups of categorical variables is related to the sum of the \( \Phi^2 \) between all the variables and the RV between one group of continuous and one group of categorical variables to the sum of the squared correlation ratio \( \eta^2 \) between the variables (Escoufier, 2006; Holmes, 2008; Pages, 2014).

Finally, even if there has been recent theoretical progress in evaluation of the different coefficients, important choices still need to be made by the user. Results depend on the particular preprocessing choice (such as scaling), distance or kernel choices. This flexibility can be viewed as a strength, since many types of dependencies can be discovered. On the other hand, of course, it underscores the subjectivity of the analysis and the importance of educated decisions by the analyst.

Acknowledgements

Julie Josse acknowledges support from the Agreenskills grant for an academic visit to Stanford in 2013. Susan Holmes acknowledges support from the NIH grant R01 GM086884.

References

H. Abdi. RV coefficient and congruence coefficient. In (Eds.) N.J. Salkind, editor, Encyclopedia of Measurement and Statistics, pages 849–853. Thousand Oaks (CA): Sage, 2007.
H. Abdi. Congruence: Congruence coefficient, RV coefficient, and mantel coefficient. In N. J. Salkind, D. M. Dougherty, and B. Frey (Eds.), editors, Encyclopedia of Research Design, pages 222–229. Thousand Oaks (CA): Sage, 2010.

E. Acar and B. Yener. Unsupervised multiway data analysis: A literature survey. Knowledge and Data Engineering, IEEE Transactions on, 21(1):6–20, 2009.

D. E. Barton and F. N. David. Randomization bases for multivariate tests. I. the bivariate case. randomness of n points in a plane. Bulletin of the international statistical institute, page i39, 1962.

R. S. Cadena, A. G. Cruz, R. R. Netto, W. F. Castro, J-d-A. F. Faria, and H. M. A. Bolini. Sensory profile and physicochemical characteristics of mango nectar sweetened with high intensity sweeteners throughout storage time. Food Research International, 2013.

R. Cl´eroux and G. R. Ducharme. Vector correlation for elliptical distribution. Communications in Statistics Theory and Methods, 18:1441–1454, 1989.

R. Cl´eroux, A. Lazraq, and Y. Lepage. Vector correlation based on ranks and a non parametric test of no association between vectors. Communications in Statistics Theory and Methods, 24:713–733, 1995.

N. Cristianini, J. Shawe-Taylor, A. Elisseeff, and J. Kandola. On kernel-target alignment. NIPS, 2001.

A. Culhane, G. Perrière, and D. Higgins. Cross-platform comparison and visualisation of gene expression data using co-inertia analysis. BMC bioinformatics, 4(1):59, 2003.

F. N. David and D. E. Barton. Combinatorial chance. Griffin London, 1962.

F. N. David and D. E. Barton. Two space-time interaction tests for epidemicity. British Journal of Preventive & Social Medicine, 20(1):44–48, 1966.

M. de Tayrac, S. Le, M. Aubry, J. Mosser, and F. Husson. Simultaneous analysis of distinct omics data sets with integration of biological knowledge: Multiple factor analysis approach. BMC Genomics, 10 (1):32, 2009.

S. Dray. The ade4 package: implementing the duality diagram for ecologists. Journal of Statistical Software, 22 (4):1–20, 2007.

S. Dray, D. Chessel, and J. Thioulouse. Procrustean co-inertia analysis for the linking of multivariate datasets. Ecoscience, 10:110–119, 2003.

Y. Escoufier. Le traitement des variables vectorielles. Biometrics, 29:751–760, 1973.

Y. Escoufier. Three-mode data analysis: the STATIS method. In Method for multidimensional analysis, pages 153–170. Lecture notes from the European Course in Statistic, 1987.

Y. Escoufier. Operator related to a data matrix: a survey. In Compstat 2006-Proceedings in Computational Statistics, pages 285–297. Springer, 2006.

C. Foth, P. Bona, and J. B. Desojo. Intraspecific variation in the skull morphology of the black caiman melanosuchus niger (alligatoridae, cainainae). Acta Zoologica, 2013.

C. Fruciano, P. Franchini, and A. Meyer. Resampling-based approaches to study variation in morphological modularity. PLoS ONE, 8:e69376, 2013.

T. Gärtner, G. Q. V. Le, and A.J. Smola. A short tour of kernel methods for graphs. 2006.
M. Génard, M. Souty, S. Holmes, M. Reich, and L. Breuils. Correlations among quality parameters of peach fruit. *Journal of the Science of Food and Agriculture*, 66(2):241–245, 1994.

D. Giacalone, L. M. Ribeiro, and M. B. Frøst. Consumer-based product profiling: Application of partial napping® for sensory characterization of specialty beers by novices and experts. *Journal of Food Products Marketing*, 19(3):201–218, 2013.

S. C. Goslee and D. L. Urban. The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, 22:1–19, 2007.

J. C. Gower. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53:325–338, 1966.

J. C. Gower. Statistical methods of comparing different multivariate analyses of the same data. In F. R. Hodson, D. G. Kendall, and P. Tautu, editors, *Mathematics in the archaeological and historical sciences*, pages 138–149. Edinburgh University Press, 1971.

M. J. Greenacre. Correspondence analysis of multivariate categorical data by weighted least-squares. *Biometrika*, 75:457–477, 1988.

M. J. Greenacre. Multiple and joint correspondence analysis. In J. Blasius and M. J. Greenacre, editors, *Correspondence Analysis in the social science*, pages 141–161. London: Academic Press, 1994.

M. J. Greenacre and J. Blasius. *Multiple Correspondence Analysis and Related Methods*. Chapman & Hall/CRC, 2006.

A. Gretton, R. Herbrich, A. Smola, O. Bousquet, and B. Schoelkopf. Kernel methods for measuring independence. *Journal of Machine Learning Research*, 6:2075–2129, 2005.

G. Guillot and F. Rousset. Dismantling the Mantel tests. *Methods in Ecology and Evolution*, 2013.

S. Holmes. Multivariate data analysis: the French way. *Probability and Statistics: Essays in Honor of David A. Freedman. Institute of Mathematical Statistics, Beachwood, Ohio*, pages 219–233, 2008.

F. Husson, J. Josse, S. Le, and J. Mazet. *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R*, 2013. URL [http://CRAN.R-project.org/package=FactoMineR](http://CRAN.R-project.org/package=FactoMineR). R package version 1.24.

D. A. Jackson. Protest: a procustean randomization test of community environment concordance. *Ecology*, 2:297–303, 1995.

J. Josse, J. Pagès, and F. Husson. Testing the significance of the RV coefficient. *Computational Statistics and Data Analysis*, 53:82–91, 2008.

F. Kazi-Aoual, S. Hitier, R. Sabatier, and J. D. Lebreton. Refined approximations to permutation tests for multivariate inference. *Computational Statistics and Data Analysis*, 20:643–656, 1995.

C. P. Klingenberg. Morphometric integration and modularity in configurations of landmarks: tools for evaluating a priori hypotheses. *Evolution & Development*, 11:405–421, 2009.

E. G. Knox. The detection of space-time interactions. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 13(1):25–30, 1964.

P. M. Kroonenberg. *Applied Multiway Data Analysis*. Wiley series in probability and statistics, 2008.

C. Lavit, Y. Escoufier, R. Sabatier, and P. Traissac. The ACT (STATIS method). *Computational Statistics & Data Analysis*, 18(1):97–119, 1994.
A. Lazraq and C. Robert. Etude comparative de différentes mesures de liaison entre deux vecteurs aléatoires et tests d’indépendance. Statistique et analyse de données, 1:15–38, 1988.

A. Lazraq, Cléroux R., and H. A. L. Kiers. Mesures de liaison vectorielle et généralisation de l’analyse canonique. Statistique et analyse de données, 40 (1):23–35, 1992.

S. Lê, J. Josse, and F. Husson. Factominer: An R package for multivariate analysis. Journal of Statistical Software, 25(1):1–18, 3 2008.

P. Legendre and M. Fortin. Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. Molecular Ecology Resources, 10:831–844, 2010.

J. C. Lingoes and P. H Schönenmann. Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. Psychometrika, 39: 423–427, 1974.

D. Lopez-Paz, P. Hennig, and B. Schölkopf. The Randomized Dependence Coefficient. ArXiv e-prints, 2013.

R. Lyons. Distance covariance in metric spaces. Annals of Probability, 41 (5):3284–3305, 2013.

N. Mantel. The detection of disease clustering and a generalized regression approach. Cancer Research, 27:209–220, 1967a.

N. Mantel. The detection of disease clustering and a generalized regression approach. Cancer research, 27(2 Part 1):209–220, 1967b.

MATLAB. MATLAB and Statistics Toolbox Release. The MathWorks, Inc., Natick, Massachusetts, United States, 2012. URL [http://www.mathworks.com/products/matlab/](http://www.mathworks.com/products/matlab/).

C-D. Mayer, T. Lorent, and G. W. Horgan. Exploratory analysis of multiples omics datasets using the adjusted RV coefficient. Statistical applications in genetics and molecular biology, 10, 2011.

C. Minas, E. Curry, and G. Montana. A distance-based test of association between paired heterogeneous genomic data. Bioinformatics, 29 (22):2555–2563, 2013.

A. C. Noble and S. E. Ebeler. Use of multivariate statistics in understanding wine flavor. Food Reviews International, 18(1):1–20, 2002.

J. Oksanen, F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O’Hara, G. L. Simpson, P. Solymos, M. H. Stevens, and H. Wagner. vegan: Community Ecology Package, 2013. URL [http://CRAN.R-project.org/package=vegan](http://CRAN.R-project.org/package=vegan) R package version 2.0-9.

M. Omelka and S. Hudecová. A comparison of the mantel test with a generalised distance covariance test. Environmetrics, 2013. URL [http://dx.doi.org/10.1002/env.2238](http://dx.doi.org/10.1002/env.2238).

J. Pagès. Multiple Factor Analysis with R. Springer, 2014.

J. Pagès and F. Husson. Multiple factor analysis with confidence ellipses: A methodology to study the relationships between sensory and instrumental data. Journal of Chemometrics, 19:138–144, 2005.

P. R. Peres-Neto and D. A. Jackson. How well do multivariate data sets match? the advantages of a procrustean superimposition approach over the mantel test. Oecologia, 129:169–178, 2001.

E. Purdom. Multivariate kernel methods in the analysis of graphical structures. PhD thesis, University of Stanford, 2006.
C. R. Rao. The use and interpretation of principal component analysis in applied research. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 329–358, 1964.

M. Reimherr and D. L. Nicolae. On quantifying dependence: A framework for developing interpretable measures. *Statistical Science*, 28 (1):116–139, 2013.

E. Risvik, J. A. McEwan, and M. Rødbotten. Evaluation of sensory profiling and projective mapping data. *Food quality and preference*, 8(1):63–71, 1997.

M. L. Rizzo and G. J. Szekely. *energy: E-statistics (energy statistics)*, 2013. URL [http://CRAN.R-project.org/package=energy](http://CRAN.R-project.org/package=energy). R package version 1.6.0.

P. Robert and Y. Escoufier. A unifying tool for linear multivariate statistical methods: The RV-coefficient. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 3:257–265, 1976.

P. Robert, R. Cléroux, and N. Ranger. Some results on vector correlation. *Computational Statistics and Data Analysis*, 3:25–32, 1985.

F. J Rohlf and D. Slice. Extensions of the procrustes method for the optimal superimposition of landmarks. *Systematic Biology*, 39(1):40–59, 1990.

S. E. Santana and S. E. Lofgren. Does nasal echolocation influence the modularity of the mammal skull? *Journal of evolutionary biology*, 26(11):2520–2526, 2013.

P. Schlich. Defining and validating assessor compromises about product distances and attribute correlations. *Data handling in science and technology*, 16:259–306, 1996.

I. J. Schoenberg. Remarks to maurice fréchet’s article “sur la définition axiomatique d’une classe d’espace distancié vectoriellement applicable sur l’espace de hilbert. *Annals of Mathematics*, 36 (2):724–732, 1935.

D. Sejdinovic, B. Sriperumbudur, A. Gretton, and K. Fukumizu. Equivalence of distance-based and rkhs-based statistics in hypothesis testing. Submitted to *Annals of Statistics*, 2013.

S. V. Shinkareva, R. A. Mason, V. L. Malave, W. Wang, T. M. Mitchell, and M. A. Just. Using fmri brain activation to identify cognitive states associated with perception of tools and dwellings. *PLoS One*, 3(1):e1394, 2008.

A. K. Smilde, H. A. L. Kiers, S. Bijlsma, C. M. Rubingh, and M. J van Erk. Matrix correlations for high-dimensional data: the modified RV-coefficient. *Bioinformatics*, 25:401–405, 2009.

P. E. Smouse, J. C. Long, and R. R. Sokal. Multiple regression and correlation extensions of the mantel test of matrix correspondence. *Systematic zoology*, 35(4):627–632, 1986.

R. R. Sokal and P. Sneath. Principles of numerical taxonomy. *Principles of numerical taxonomy*, 1963.

L. Song, A. Smola, A. Gretton, J. Bedo, and K. Borgwardt. Feature selection via dependence maximization. *Journal of Machine Learning Research*, 13:1393–1434, 2012.

G. J. Szekely and M. L. Rizzo. Energy statistics: A class of statistics based on distances. *Journal of statistical planning and inference*, 143:1249–1272, 2013a.

G. J. Szekely and M. L. Rizzo. The distance correlation t-test of independence in high dimension. *Journal of Multivariate Analysis*, 117:193–213, 2013b.
G. J. Szekely and M. L. Rizzo. Partial distance correlation with methods for dissimilarities.
*arXiv:1310.2926*, 2013c.

G. J. Szekely, M. L. Rizzo, and N. K. Bakirov. Measuring and testing dependence by correlation of
distances. *The Annals of Statistics*, 35 (6):2769–2794, 2007.

D. M. Witten, R. Tibshirani, and T. Hastie. A penalized matrix decomposition, with applications to
sparse principal components and canonical correlation analysis. *Biostatistics*, 10:515–534, 2009.