

Nonlinear Principal and Canonical Directions from Continuous Extensions of Multidimensional Scaling

Carles M. Cuadras

Department of Statistics, University of Barcelona, Barcelona, Spain
Email: ccuadras@ub.edu

Received December 26, 2013; revised January 26, 2014; accepted February 4, 2014

Copyright © 2014 Carles M. Cuadras. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. In accordance of the Creative Commons Attribution License all Copyrights © 2014 are reserved for SCIRP and the owner of the intellectual property Carles M. Cuadras. All Copyright © 2014 are guarded by law and by SCIRP as a guardian.

ABSTRACT

A continuous random variable is expanded as a sum of a sequence of uncorrelated random variables. These variables are principal dimensions in continuous scaling on a distance function, as an extension of classic scaling on a distance matrix. For a particular distance, these dimensions are principal components. Then some properties are studied and an inequality is obtained. Diagonal expansions are considered from the same continuous scaling point of view, by means of the chi-square distance. The geometric dimension of a bivariate distribution is defined and illustrated with copulas. It is shown that the dimension can have the power of continuum.

KEYWORDS

Statistical Distances; Orthogonal Expansions; Principal Directions of Random Variables; Diagonal Expansions; Copulas; Uncountable Dimensionality

1. Introduction

Let X be a random variable on a probability space (Ω, \mathcal{A}, P) with range $I = [a, b] \subset \overline{\mathbb{R}}$, absolutely continuous cdf F and density f w.r.t. the Lebesgue measure. Our main purpose is to expand (a function of) X as

$$X = \mu + \sum_{n \geq 1} b_n X_n, \quad (1)$$

where (X_n) is a sequence of uncorrelated random variables, which can be seen as a decomposition of the so-called geometric variability $V_\delta(X)$, defined below, a dispersion measure of X in relation with a suitable distance function $\delta(x, x'), x, x' \in I$. Here orthogonality is synonymous with a lack of correlation.

Some goodness-of-fit statistics, which can be expressed as integrals of the empirical processes of a sample, have expansions of this kind [1-3]. Expansion (1) is obtained following a similar procedure, except that we have a sequence of uncorrelated rather than independent random variables. Finite orthogonal expansions appear in analysis of variance and in factor analysis. Orthogonal expansions and series also appear in the theory of stochastic processes, in martingales in the wide sense ([4], Chap. 4; [5], Chap. 10), in non-parametric statistics [6], in goodness-of-fit tests [7,8], in testing independence [9] and in characterizing distributions [10].

The existence of an orthogonal expansion and some classical expansions is presented in Section 2. A continuous extension of matrix formulations in multidimensional scaling (MDS), which provides a wide class of expansions, is presented in Section 3. Some interesting expansions are obtained in Section 4 for a particular distance as well as additional results, such as an optimal property of the first dimension. Section 5 contains an inequality concerning the variance of a function. Section 6 is devoted to diagonal expansions from the continuous scaling point of view. Sections 7 and 8 are devoted to canonical correlation analysis, including a continuous generalization. This paper extends the previous results on continuous scaling [11-14], and other related topics

[15,16].

2. Existence and Classical Expansions

There are many ways of obtaining expansion (1). Our aim is to obtain some explicit expansions with good properties from a multivariate analysis point of view. However, before doing this, let us prove that a such expansion exists and present some classical expansions.

Theorem 1. *Let X be an absolutely continuous r.v. with density f and support I and h a measurable function such that $E[h(X)] = 0$ and $\text{var}[h(X)] < \infty$. Then there exists a sequence (V_n) of uncorrelated r.v.'s such that*

$$h(X) = \sum_{n \geq 1} V_n, \quad (2)$$

with $\text{var}[h(X)] = \sum_{n \geq 1} \text{var}(V_n)$, where the series $\sum_{n \geq 1} V_n$ converges in the mean square as well as almost surely.

Proof. Consider the Lebesgue spaces $L^2(I)$ and $L^2(f)$ of measurable functions u on I such that $\|u\|^2 = \int_I |x|^2 dx < \infty$, and $\|u\|_f^2 = \int_I |x|^2 f(x) dx < \infty$, respectively. Obviously, $L^2(I)$ and $L^2(f)$ are separable Hilbert spaces with quadratic-norms $\|\cdot\|$ and $\|\cdot\|_f$, respectively. Moreover T_f given by $T_f u = f^{-\frac{1}{2}} u$, ($u \in L^2(I)$), is a linear isometry of $L^2(I)$ onto $L^2(f)$ i.e., $\|T_f u\|_f = \|u\|$.

Let (u_n) be an orthonormal basis for $L^2(I)$. Accordingly, (u_n^*) , given by

$$u_n^* = T_f u_n = f^{-\frac{1}{2}} u_n \quad n \geq 1,$$

is an orthonormal basis for $L^2(f)$. The assumption $\text{var}[h(X)] < \infty$, together with $E[h(X)] = 0$, is equivalent to $h \in L^2(f)$. In fact, $\|h\|_f^2 = \int_I h(x)^2 f(x) dx = \text{var}[h(X)] < \infty$. Hence $h = \sum_{n \geq 1} b_n u_n^*$, where b_n are the Fourier coefficients and $\sum_{n \geq 1} b_n^2 = \|h\|_f^2$. Letting $v_n = b_n u_n^*$, we deduce that $h = \sum_{n \geq 1} v_n$, where the series converges in $L^2(f)$ and $\langle v_m, v_n \rangle_f = b_m^2 \delta_{mn}$, $m, n \geq 1$, where δ_{mn} is Kronecker's delta. Replacing x by X , and defining $V_n = v_n(X)$, we obtain $h(X) = \sum_{n \geq 1} V_n$.

This series converges almost surely, since the series $\sum_{n \geq 1} v_n$ converges in $L^2(f)$. We may suppose $E(V_n) = 0$. Next, for $m, n \geq 1$, we have $\text{cov}(V_m, V_n) = E(V_m V_n) = \langle v_m, v_n \rangle_f = b_m^2 \delta_{mn}$, as asserted. Finally, note that $\text{var}[h(X)] = \|h\|_f^2 = \sum_{n \geq 1} \text{var}(V_n)$. In particular, the series in (2) converges also in the mean square. \square

Some classical expansions for X or a function of X are next given.

2.1. Legendre Expansions

Let F be the cdf of X . An all-purpose expansion can be obtained from the shifted Legendre polynomials (L_n) on $[0,1]$, where $L_n(x) = \sqrt{2n+1} P_n(2x-1)$. The first three are

$$L_1(x) = \sqrt{3}(2x-1), \quad L_2(x) = \sqrt{5}(6x^2 - 6x + 1), \quad L_3(x) = \sqrt{7}(20x^3 - 30x^2 + 12x - 1).$$

Note that $E[L_n(F(X))] = 0$, $\text{cov}(L_m(F(X)), L_n(F(X))) = \delta_{mn}$. Thus we may consider the orthogonal expansion

$$X = \mu + \sum_{n \geq 1} b_n L_n[F(X)],$$

where $\mu = E(X)$ and $b_n = \text{cov}(X, L_n(X))$ are the Fourier coefficients. This expansion is quite useful due to the simplicity of these polynomials, is optimal for the logistic distribution, but may be improved for other distributions, as it is shown below.

2.2. Univariate Expansions

A class of orthogonal expansions arises from

$$h(x) = g(x) \left[1 + \sum_{n \geq 1} a_n A_n(x) \right], \tag{3}$$

where both h, g are probability density functions. Then $(A_n(X))$ is a complete orthonormal set w.r.t. g .

2.3. Diagonal Expansions

Lancaster [17] studied the orthogonal expansions

$$h(x, y) = f(x)g(y) \left[1 + \sum_{n \geq 1} \rho_n u_n(x)v_n(y) \right], \tag{4}$$

where h is a bivariate probability density with marginal densities f, g . Then $(u_n), (v_n)$ are complete orthonormal sets w.r.t f and g , respectively. Moreover $E[u_n(X)] = E[v_n(Y)] = 0$ and $E[u_n(X)v_m(Y)] = \rho_n \delta_{mn}$, where ρ_n is the n -th canonical correlation between X and Y .

Expansion (4) can be viewed as a particular extension of Theorem 3, proved below, when the distance is the so-called chi-square distance. This is proved in [18]. See Section 6.

3. Continuous Scaling Expansions

In this section we propose a distance-based approach for obtaining orthogonal expansions for a r.v. X , which contains the Karhunen-Loève expansion of a Bernoulli process related to X as a particular case. We will prove that we can obtain suitable expansions using continuous multidimensional scaling on a Euclidean distance.

Let $\delta: I \times I \rightarrow \mathbb{R}^+$ be a dissimilarity function, i.e., $\delta(x, y) = \delta(y, x) \geq 0$ and $\delta(x, x) = 0$ for all $x, y \in I$.

Definition 1. We say that δ is a Euclidean distance function if there exists an embedding $x \rightarrow \phi(x) \in E$, where E is a real separable Hilbert space with quadratic norm $\|\cdot\|$, such that $\delta(x, x') = \|\phi(x) - \phi(x')\|$.

We may always view the Hilbert space E as a closed linear subspace of ℓ^2 . In this case, we may identify $\phi(x)$ with $\mathbf{Q}(x)$, where for $x \in I$, $\mathbf{Q}(x) = (Q_1(x), Q_2(x), \dots)$ is a vector in ℓ^2 . Accordingly, for $x, x' \in I$

$$\delta^2(x, x') = \sum_{n \geq 1} [Q_n(x) - Q_n(x')]^2 = \|\mathbf{Q}(x) - \mathbf{Q}(x')\|^2. \tag{5}$$

Definition 2. $\mathbf{Q}(x) = (Q_1(x), Q_2(x), \dots)$ is called a Euclidean configuration to represent (I, δ) . The geometric variability of X w.r.t. δ is defined by

$$V_\delta(X) = \frac{1}{2} \int_I \int_I \delta^2(x, x') f(x) f(x') dx dx'.$$

The proximity function of x to X is defined by

$$D_\delta^2(x, X) = E[\delta^2(x, X)] - V_\delta(X). \tag{6}$$

The double-centered inner product related to δ is the symmetric function

$$G(x, x') = -\frac{1}{2} [\delta^2(x, x') - D_\delta^2(x, X) - D_\delta^2(x', X)]. \tag{7}$$

These definitions can easily be extended to random vectors. For example, if \mathbf{X} is $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, \mathbf{x} is an observation of \mathbf{X} and δ is the Euclidean distance in \mathbb{R}^p , then $V_\delta(\mathbf{X}) = \text{tr}(\boldsymbol{\Sigma})$ and $D_\delta^2(\mathbf{x}, \mathbf{X}) = (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$ is the Mahalanobis distance from \mathbf{x} to $\boldsymbol{\mu}$.

The function G is symmetric, semi-definite positive and satisfies

$$\delta^2(x, x') = G(x, x) + G(x', x') - 2G(x, x'). \tag{8}$$

It can be proved [19], that there is an embedding $x \rightarrow (Q_1(x), Q_2(x), \dots)$ of I into ℓ^2 such that

$$V_\delta(X) = E[\|\mathbf{Q}(X)\|^2] - \|E[\mathbf{Q}(X)]\|^2,$$

$$D_\delta^2(x_0, X) = \|\mathbf{Q}(x_0) - E[\mathbf{Q}(X)]\|^2.$$

G is the continuous counterpart of the centered inner product matrix computed from a $n \times n$ distance matrix

and used in metric multidimensional scaling [20,21]. The Euclidean embedding or method of finding the Euclidean coordinates from the Euclidean distances was first given by Schoenberg [22]. The concepts of geometric variability and proximity function, were originated in [23] and are used in discriminant analysis [19], and in constructing probability densities from distances [24]. In fact, $V_\delta(X)$ is a variant of Rao’s quadratic entropy [25]. See also [14,26].

In order to obtain expansions, our interest is on $\delta(X, X'), Q(X)$ and $G(X, X')$, i.e., we substitute x by the r.v. X .

For convenience and analogy with the finite classic scaling, let us use a generalized product matrix notation (i.e., a “multivariate analysis notation”), following [18]. We write $G(X, X')$ as $G = QQ'$, where $Q = (Q_1(X), Q_2(X), \dots)$ is a Euclidean centered configuration to represent (I, δ) , according to (5), (8), i.e., we substitute x by X in $Q(x) = (Q_1(x), Q_2(x), \dots)$ and suppose that each $Q_n(X)$ has mean 0. The covariance matrix of Q is $\Sigma = (\sigma_{ij})$, where for $i, j \geq 1$, $\sigma_{ij} = E(Q_i Q_j) = \int_I Q_i(x) Q_j(x) f(x) dx$, and may be expressed as

$$\Sigma = Q' f^{1/2} * f^{1/2} Q,$$

where $f^{1/2}$ stands for the continuous diagonal matrix $\text{diag}(f^{1/2}(x))$ and the row \times column multiplication, denoted by $*$, is evaluated at $x \in I$ and then follows an integration w.r.t. x .

In the theorems below, $f^{\frac{1}{2}} \otimes f^{\frac{1}{2}} G$ and $u_n \otimes u_n$ stands for $f^{\frac{1}{2}}(x) f^{\frac{1}{2}}(x') G(x, x')$ and $u_n(x) u_n(x')$.

Theorem 2. Suppose that for a Euclidean distance δ the geometric variability $V_\delta(X)$ is finite. Define $tr_X(G) = E[G(X, X)]$. Then:

1. $V_\delta(X) = tr_X(G)$.
2. $f^{\frac{1}{2}} \otimes f^{\frac{1}{2}} G$ is a Hilbert-Schmidt kernel, i.e.,

$$\int_I \int_I G(x, x')^2 f(x) f(x') dx dx' < \infty.$$

Proof. Let X, X' be i.i.d. and write $v = V_\delta(X)$. From (7) $E[G(X, X')] = -\frac{1}{2}(2v - v - v) = 0$, and hence using (8),

$$\begin{aligned} v &= \frac{1}{2} \int_I \int_I \delta^2(x, x') f(x) dx f(x') dx' \\ &= \frac{1}{2} [2E(G(X, X)) - 2E(G(X, X'))] \\ &= tr_X(G). \end{aligned}$$

This proves 1). Next, G is p.s.d., so for x_1, \dots, x_n the $n \times n$ matrix with entries $G(x_i, x_j)$ is p.s.d. In particular, for $n = 2$ with $x_1 = x, x_2 = x'$, the determinant

$$\begin{vmatrix} G(x, x) & G(x, x') \\ G(x, x') & G(x', x') \end{vmatrix}$$

is non-negative. Thus

$$f(x)G(x, x)G(x', x')f(x') \geq f(x)G(x, x')^2 f(x').$$

Integrating this inequality over I^2 gives

$$tr_X(G)^2 \geq \int_I \int_I G(x, x')^2 f(x) f(x') dx dx',$$

and 2) is also proved. \square

Theorem 3. Suppose that $V_\delta(X)$ for a Euclidean distance δ is finite. Then the kernel $f^{\frac{1}{2}} \otimes f^{\frac{1}{2}} G$ admits the expansion

$$f^{\frac{1}{2}} \otimes f^{\frac{1}{2}} G = \sum_{n \geq 1} \lambda_n u_n \otimes u_n, \tag{9}$$

where (u_n) is a complete orthonormal set of eigenfunctions in $L^2(I)$. Let

$$c_n = \sqrt{\lambda_n} f^{-\frac{1}{2}} u_n, \tag{10}$$

and $C_n = c_n(X), n \geq 1$. Then $\lambda_n = \text{var}(C_n)$ and (C_n) is a sequence of centered and uncorrelated r.v.'s, which are principal components of \mathbf{Q} , where $\mathbf{G} = \mathbf{Q}\mathbf{Q}'$ and

$$\text{tr}_X(\mathbf{G}) = V_\delta(X) = \sum_{n \geq 1} \lambda_n.$$

Proof. The eigendecomposition (9) exists because $f^{1/2} \otimes f^{1/2} \mathbf{G}$ is a Hilbert-Schmidt kernel and $\text{tr}_X(\mathbf{G}) < \infty$, by Theorem 2. Next, (9) and (10) can be written as

$$f^{\frac{1}{2}} \mathbf{G} f^{\frac{1}{2}} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}', \tag{11}$$

i.e., $f^{\frac{1}{2}} \mathbf{G} f^{\frac{1}{2}} * \mathbf{U} = \mathbf{U} \mathbf{\Lambda}$, and $\mathbf{C} = f^{-\frac{1}{2}} \mathbf{U} \mathbf{\Lambda}^{\frac{1}{2}}$, $\mathbf{G} = \mathbf{C} \mathbf{C}'$, where $\mathbf{U} = (u_i)$, $\mathbf{C} = (c_i)$, $f^{1/2} = \text{diag}(f^{1/2})$, $\mathbf{\Lambda} = \text{diag}(\lambda_i)$. Thus, for all $x, x' \in I$,

$$G(x, x') = \sum_{n \geq 1} c_n(x) c_n(x').$$

Next, for $x \in I$ we have $E[G(x, X)] = E[G(X, x)] = 0$. In particular, since

$$G(X, x) = \sum_{n \geq 1} c_n(X) c_n(x), \quad (x \in I),$$

we have $E[c_n(X)] = 0$. Moreover, from (10) we also have

$$\text{cov}(c_m(X), c_n(X)) = \lambda_n \delta_{mn},$$

where δ_{mn} is Kronecker's delta, showing that the variables $c_n(X)$ are centered and uncorrelated.

Recall the product matrix notation. The principal components of \mathbf{Q} such that $\mathbf{Q}\mathbf{Q}' = \mathbf{G}$ are $\mathbf{Q}\mathbf{\Phi}$ where

$$\mathbf{\Sigma} = \mathbf{Q} f^{1/2} * f^{1/2} \mathbf{Q} = \mathbf{\Phi} \mathbf{\Lambda} \mathbf{\Phi}' \tag{12}$$

is the spectral decomposition of $\mathbf{\Sigma}$. Premultiplying (12) by $f^{\frac{1}{2}} \mathbf{Q}$ and postmultiplying by $\mathbf{\Phi}'$ we obtain

$$f^{\frac{1}{2}} \mathbf{G} f^{\frac{1}{2}} * f^{\frac{1}{2}} \mathbf{Q} \mathbf{\Phi} = f^{\frac{1}{2}} \mathbf{Q} \mathbf{\Phi} \mathbf{\Lambda}$$
 and therefore

$$f^{\frac{1}{2}} \mathbf{G} f^{\frac{1}{2}} * f^{\frac{1}{2}} \mathbf{Q} \mathbf{\Phi} \mathbf{\Lambda}^{-1/2} = f^{\frac{1}{2}} \mathbf{Q} \mathbf{\Phi} \mathbf{\Lambda}^{-1/2} \mathbf{\Lambda}.$$

Thus the columns of $f^{\frac{1}{2}} \mathbf{Q} \mathbf{\Phi} \mathbf{\Lambda}^{-1/2}$ are eigenfunctions of $f^{\frac{1}{2}} \mathbf{G} f^{\frac{1}{2}}$. This shows that $f^{\frac{1}{2}} \mathbf{Q} \mathbf{\Phi} \mathbf{\Lambda}^{-1/2} = \mathbf{U}$, see (11), and $\mathbf{C} = \mathbf{Q} \mathbf{\Phi} = (c_1(X), c_2(X), \dots)$ contains the principal components of \mathbf{Q} . The rows in \mathbf{C} may be accordingly called the *principal coordinates* of distance δ . This name can be justified as follows.

Let us write the coordinates $C(x) = (c_1(x), c_2(x), \dots)$ and suppose that $P(x) = (p_1(x), p_2(x), \dots)$ is another Euclidean configurations giving the same distance δ . The linear transformation $C(x) \rightarrow P(x)$ is orthogonal and $p_m(x) = C(x) \mathbf{a}_m$ with $\|\mathbf{a}_m\|^2 = \sum_{n \geq 1} a_{nm}^2 = 1$. Then the r.v.'s $p_1(X), p_2(X), \dots$ are uncorrelated and

$$\text{var}[p_m(X)] = \sum_{n \geq 1} a_{nm}^2 \lambda_n \leq \|\mathbf{a}_m\|^2 \lambda_1 = \text{var}[c_1(X)].$$

Thus $c_1(x)$ is the first principal coordinate in the sense that $\text{var}[c_1(X)]$ is maximum. The second coordinate $c_2(x)$ is orthogonal to $c_1(x)$ and has maximum variance, and so on with the others coordinates. \square

The following expansions hold as an immediate consequence of this theorem:

$$G(X, X') = \sum_{n \geq 1} C_n C'_n, \tag{13}$$

$$G(X, \alpha) = \sum_{n \geq 1} c_n(\alpha) C_n, \tag{14}$$

$$\delta^2(X, X') = \sum_{n \geq 1} (C_n - C'_n)^2, \tag{15}$$

where $\alpha \in I$ and $(C_n), (C'_n)$, with $C_n = c_n(X), C'_n = c_n(X'), n \geq 1$, are sequences of centered and uncorrelated random variables, which are principal components of Q . We next obtain some concrete expansions.

4. A Particular Expansion

If X is a continuous r.v. with finite mean and variance, $\mu = E(X)$, $\sigma^2 = \text{var}(X)$, say, and δ is the ordinary Euclidean distance $|x - x'|$, then it is easy to prove that $V_\delta(X) = \sigma^2$, $D_\delta^2(x, X) = (x - \mu)^2$, $G(x, x') = (x - \mu)(x' - \mu)$. Then from (14) and taking $x' = \alpha$, we obtain $G(X, \alpha) = (X - \mu)(\alpha - \mu)$, which provides the trivial expansion $X = X$. A much more interesting expansion can be obtained by taking the square root of $|x - x'|$.

4.1. The Square Root Distance

Let us consider the distance function

$$\delta(x, x') = \sqrt{|x - x'|}. \tag{16}$$

The double-centered inner product $G(x, x')$ is next given.

Definition 3. Let $h_n(x)$ be defined as $h_n(x) = c_n(x) - c_n(a)$, where c_n is defined in (10).

We immediately have the following result.

Proposition 1. The function h_n satisfies:

1. $E[h_n(X)] = \mu_n = -c_n(a)$.
2. $h_n(a) = 0$.
3. $h_n = c_n + \mu_n$.

Proposition 2. Assuming X, X' i.i.d., if $\tilde{\mu} = E_{XX'}[\min\{X, X'\}]$, then

$$\min\{x, x'\} = \tilde{\mu} + G(x, x') - G(x, a) - G(a, x'). \tag{17}$$

Proof. From $|x - x'| = x + x' - 2\min\{x, x'\}$ and combining (7) and (6), we obtain $G(x, x') = \min\{x, x'\} - \varphi(x) - \varphi(x') + \tilde{\mu}$, where $\varphi(x) = E_X \min\{X, x\}$, which satisfies $\varphi(a) = a$, $\varphi(b) = E(X) = \mu_0$. Hence $\varphi(x) = \tilde{\mu} - G(x, a)$ and (17) holds. \square

Proposition 3. The following expansion holds

$$\min\{x, x'\} = x_0 + \sum_{n \geq 1} [h_n(x)h_n(x') - h_n(x_0)h_n(b)]. \tag{18}$$

Proof. Using (17) and expanding $\min\{x, x'\} - \min\{y, y'\}$ and setting $y = b, y' = x_0$, we get (18). \square
 Replacing x by X we have the expansion

$$\min\{X, X'\} = x_0 + \sum_{n \geq 1} [X_n X'_n - h_n(x_0)h_n(b)], \tag{19}$$

and, as a consequence [12]:

$$|X - X'| = \sum_{n \geq 1} (X_n - X'_n)^2,$$

where X' is distributed as X and $X'_n = h_n(X')$. This expansion also follows from (15).

If $\mu_0 = E(X)$, $\mu_n = E(X_n)$, from (19) we can also obtain the expansions

$$\begin{aligned} X &= x_0 + \sum_{n \geq 1} h_n(b)[X_n - h_n(x_0)], \\ X &= \mu_0 + \sum_{n \geq 1} h_n(b)(X_n - \mu_n), \end{aligned} \tag{20}$$

as well as

$$\begin{aligned} X &= x_0 + \sum_{n \geq 1} [X_n^2 - h_n(x_0)h_n(b)], \\ X &= \mu_0 + \sum_{n \geq 1} [X_n^2 - \mu_n h_n(b)], \\ X &= \tilde{\mu} + \sum_{n \geq 1} (X_n + \mu_n)(X_n - \mu_n), \end{aligned} \tag{21}$$

where the convergence is in the mean square sense [27].

4.2. Principal Components

Related to the r.v. X with cdf F , let us define the stochastic process $\mathbf{X} = \{X_t, t \in I\}$, where X_t is the indicator of $[X > t]$ and follows the Bernoulli $B(1, p)$ distribution with $p = 1 - F(t)$. For the distance (16), the relation between X , the Bernoulli process \mathbf{X} and δ is

$$\int_a^b (x_t - x'_t)^2 dt = \int_x^{x'} (0-1)^2 dt = |x - x'|, \tag{22}$$

where $X(\omega) = x$ and $X_t(\omega) = x_t$ is a realization of \mathbf{X} . Thus \mathbf{X} is a continuous configuration for δ . Note that, if a is finite, then

$$X = a + \int_I X_t dt.$$

The covariance kernel $K(s, t)$ of \mathbf{X} is given by $K(s, t) = \min\{F(s), F(t)\} - F(s)F(t)$. Let us consider the integral operator \mathcal{K} with kernel K

$$(\mathcal{K}\varphi)(t) = \int_I K(s, t)\varphi(s) ds,$$

where φ is an integrable function on $I = [a, b]$. Let $(\psi_n(t)), n \geq 1$, be the countable orthonormal set of eigenfunctions of \mathcal{K} , i.e., $\mathcal{K}\psi_n = \lambda_n\psi_n$. We may suppose that (ψ_n) constitutes a basis of $L_2(I)$ and the eigenvalues (λ_n) are arranged in descending order. As a consequence of Mercer's theorem, the covariance kernel $K(s, t) = \min\{F(s), F(t)\} - F(s)F(t)$ can be expanded as

$$K(s, t) = \sum_{n \geq 1} \lambda_n \psi_n(s)\psi_n(t). \tag{23}$$

Theorem 4. The functions $h_n = c_n - c_n(a)$ (see definition 3), satisfy:

1. $h_n(x) = \int_a^x \psi_n(s) ds$.
2. $(h_n(X))$ is a countable set of uncorrelated r.v.'s such that $\text{var}[h_n(X)] = \lambda_n, n \geq 1$.
3. $(h_1(x), h_2(x), \dots)$ are principal coordinates for the distance $\sqrt{|x - x'|}$, i.e.,

$$|x - x'| = \sum_{n \geq 1} [h_n(x) - h_n(x')]^2.$$

Proof. To prove 1), let us use the multiplication “*” and write $K(s, t)$ as $\mathbf{K} = \bar{\mathbf{X}}\mathbf{f}^{1/2} * \mathbf{f}^{1/2}\bar{\mathbf{X}} = \Phi\Lambda\Phi'$, where

$$\bar{\mathbf{X}} = \{\bar{X}_t = X_t - [1 - F(t)], t \in [a, b]\},$$

i.e., $K(s, t) = \int_x^t (x_s(x) - \mu_s)f(x)[x_t(x) - \mu_t]dx$, with $\mu_t = 1 - F(t)$, where $x_s(x) = 1$ if $x > s$ and 0 otherwise. Then $\bar{\mathbf{X}}$ is a centered continuous configuration for δ and clearly $\mathbf{G} = \bar{\mathbf{X}} * \bar{\mathbf{X}}'$. Arguing as in Theorem 3, the centered principal coordinates are $\mathbf{C} = \bar{\mathbf{X}} * \Phi$, i.e.,

$$\begin{aligned} c_n(x) &= \int_I x_t \psi_n(t) dt - \int_I [1 - F(t)] \psi_n(t) dt \\ &= \int_a^x \psi_n(t) dt - \mu_n = h_n(x) - \mu_n. \end{aligned}$$

2) is a consequence of Theorem 3. An alternative proof follows by taking $\alpha = h_m, \beta = h_n, L = K$, in the formula for the covariance [28]:

$$\text{cov}(\alpha(X), \beta(X)) = \int_I \int_I M(x, y) d\alpha(x) d\beta(y). \tag{24}$$

where $M(x, y) = \min\{F(x), F(y)\} - F(x)F(y)$. See Theorem 6.

To prove 3), let $X(\omega) = x < X(\omega') = x'$ and $x_t = X_t(\omega), x'_t = X_t(\omega')$. Then $x_t = \sum_{n \geq 1} h_n(x)\psi_n(t)$ and $|x - x'|$, see (22), is

$$\begin{aligned} \int_I (x_t - x'_t)^2 dt &= \int_I \sum_{n \geq 1} [h_n(x) - h_n(x')]^2 \psi_n^2(t) dt \\ &= \sum_{n \geq 1} [h_n(x) - h_n(x')]^2. \end{aligned}$$

This proves that $(h_1(x), h_2(x), \dots)$ are principal coordinates for the distance $\delta(x, x') = \sqrt{|x - x'|}$. \square

The above results have been obtained via continuous scaling. For this particular distance, we get the same results by using the Karhunen-Loève (also called Kac-Siebert) expansion of \mathbf{X} , namely,

$$X_t = \sum_{n \geq 1} \psi_n(t) X_n, \tag{25}$$

where $X_n = \int_I X_t \psi_n(t) dt$, i.e., $X_n = h_n(X)$. Thus each X_n is a principal component of \mathbf{X} and the sequence $X_n = h_n(X), n \geq 1$, constitutes a countable set of uncorrelated random variables.

4.3. The Differential Equation

Let $g_n = h_n - \mu_n$, where $\mu_n = E(X_n)$. It can be proved [12] that the means μ_n , eigenvalues λ_n and functions g_n , satisfy the second order differential equation

$$\lambda g''(x) + g(x) f(x) = 0, \quad g(a) = -\mu, g'(a) = 0. \tag{26}$$

The solution of this equation is well-known when X is $(0,1)$ uniform.

Examples of eigenfunctions ψ_n , principal components $h_n(X)$ and the corresponding variances λ_n are [12,27,29]:

1. X is $(0,1)$ uniform: $\psi_n(x) = \sqrt{2} \sin(n\pi x)$, $h_n(X) = [\sqrt{2}/(n\pi)](1 - \cos n\pi X)$, $\lambda_n = 1/(n\pi)^2$.
2. X is exponential with unit mean. If $c_n = \xi_n J_0(\xi_n)$,

$$\begin{aligned} \psi_n(x) &= \exp(-x/2) J_1(\xi_n \exp(-x/2)) / J_0(\xi_n), \\ h_n(X) &= [2J_0(\xi_n \exp(-X/2)) - 2J_0(\xi_n)] / c_n, \end{aligned}$$

$\lambda_n = 4/\xi_n^2$, where ξ_n is the n -th positive root of J_1 and J_0, J_1 are the Bessel functions of the first order.

3. X is standard logistic $F(x) = 1/[1 + \exp(-x)]$, $-\infty < x < \infty$. If $c_n = n(n+1)$, $d_n = 2n+1$,

$$\begin{aligned} \psi_n(x) &= \frac{1}{2} \sqrt{\frac{c_n}{d_n}} \left[\frac{L_{n-1}(F(x))}{\sqrt{d_n-2}} - \frac{L_{n+1}(F(x))}{\sqrt{d_n+2}} \right], \\ h_n(X) &= c_n^{-1/2} [L_n(F(X)) + (-1)^{n+1} \sqrt{d_n}], \end{aligned} \tag{27}$$

$\lambda_n = 1/[n(n+1)]$, where (L_n) are the shifted Legendre polynomials on $[0,1]$.

4. X is Pareto with $F(x) = 1 - x^{-3}, x > 1$:

$$\begin{aligned} \psi_n(x) &= c_n [\sin(\xi_n/x) - (1/x) \cos(\xi_n/x)], \\ h_n(X) &= c_n [X \sin(\xi_n/X) - \sin(\xi_n)], \end{aligned}$$

$\lambda_n = 3/\xi_n^2$, where

$$c_n = 2\xi_n^{-1/2} [2\xi_n - \sin(2\xi_n)]^{-1/2}$$

and $\xi_n = \tan(\xi_n)$.

Note that the change of variable $y = x/\alpha$ transforms g in g_α and (26) in

$$(\lambda/\alpha) g_\alpha''(y) / \sqrt{\alpha} + g_\alpha(y) / \sqrt{\alpha} \cdot \alpha f(\alpha y) = 0.$$

Hence $h_n(x) \rightarrow h_n(\alpha y) / \sqrt{\alpha}$, $\psi_n(x) \rightarrow \sqrt{\alpha} \psi_n(\alpha y)$ and $\lambda_n \rightarrow \lambda_n / \alpha$, providing solutions for the variable X/α . For instance, we immediately can obtain the principal dimensions of the Pareto distribution with cdf $1 - (x/\alpha)^{-3}$, $x > \alpha$.

4.4. A Comparison

The results obtained in the previous sections can be compared and summarized in **Table 1**, where X is a random variable with absolutely continuous cdf F , density f and support $I = [a,b]$. The continuous scaling expansion is found w.r.t. the distance (16). Note that we reach the same orthogonal expansions (we only include two), but this continuous scaling approach is more general, since by changing the distance we may find other

Table 1. Principal components and principal directions of a random variable.

Principal components expansion	Continuous scaling expansion
Bernoulli process $X_t = \mathbf{1}_{[X>s]}$	Euclidean distance $\delta(x, x') = \sqrt{ x-x' } = \ \mathcal{Q}(x) - \mathcal{Q}(x')\ $
Mercer's theorem $K(s, t) = \text{cov}(X_s, X_t) = \sum_{n \geq 1} \lambda_n \psi_n(t) \psi_n(s)$	Eigen decomposition of G $G(x, x') = \mathcal{Q}\mathcal{Q}' = \sum_{n \geq 1} \lambda_n f(x)^{\frac{1}{2}} u_n(x) f(x')^{\frac{1}{2}} u_n(x')$
Orthogonal sequence $h_n(x) = \int_a^x \psi_n(t) dt$	Orthogonal sequence (centered) $c_n(x) = \sqrt{\lambda_n} f(x)^{\frac{1}{2}} u_n(x)$
Principal components $X_n = \int_t X_t \psi_n(t) dt = h_n(X)$	Principal dimensions $X_n = h_n(X) = c_n(X) - c_n(a)$
Karhunen-Loève expansion $X_t = \sum_{n \geq 1} \psi_n(t) X_n$	Continuous scaling expansion $G(x, x') = \sum_{n \geq 1} c_n(x) c_n(x')$
Trace of K $\text{tr}(K) = \int_t K(s, s) ds$	Geometric variability $V_\delta(X) = \frac{1}{2} \int_t \delta(x, x')^2 dF(x) dF(x')$
Decomposition of $\text{tr}(K)$ $\text{tr}(K) = \sum_{n \geq 1} \lambda_n = \sum_{n \geq 1} \text{var}(X_n)$	Decomposition of $V_\delta(X)$ $V_\delta(X) = \sum_{n \geq 1} \lambda_n = \sum_{n \geq 1} \text{var}(X_n)$
Orthogonal expansions	
$X = x_0 + \sum_{n \geq 1} h_n(b) [h_n(X) - h_n(x_0)]$	
$ X - X' = \sum_{n \geq 1} [h_n(X) - h_n(X')]^2$	

principal directions and expansions. This distance-based approach may be an alternative to the problem of finding nonlinear principal dimensions [30].

4.5. Some Properties of the Eigenfunctions

In this section we study some properties of the eigenfunctions ψ_n and their integrals h_n .

Proposition 4. *The first eigenfunction ψ_1 is strictly positive and satisfies*

$$\psi_1(x) > \psi_n(a) = \psi_n(b) = 0, \quad x \in (a, b), n \geq 2.$$

Proof. K is positive, so ψ_1 is also positive (Perron-Frobenius theorem). On the other hand $K(t, t) = F(t)[1 - F(t)] = \sum_{n \geq 1} \lambda_n \psi_n(t)^2$, which satisfies $K(a, a) = K(b, b) = 0$. \square

Clearly, if ψ_1 is positive, h_1 is increasing and positive. Moreover, any h_n satisfies the following bound.

Proposition 5. *If $\sigma^2 = \text{var}(X)$ is finite then $h_n(b)$ is also finite and*

$$h_n(b) < \frac{\sigma}{\sqrt{\lambda_n}}, \quad n \geq 1. \tag{28}$$

Proof. $\psi_n = h'_n$ is an eigenfunction and from (24)

$$\begin{aligned} & \int_t \left(\int_t K(x, y) \psi_n(x) dx \right) dy \\ &= \lambda_n \int_t \psi_n(y) dy = \lambda_n h_n(b) = \text{cov}(h_n(X), X). \end{aligned}$$

Hence $\lambda_n^2 h_n(b)^2 < \lambda_n \sigma^2$. \square

Proposition 6. *The principal components $(h_n(X))$ of X_t constitutes a complete orthogonal system of $L^2(F)$.*

Proof. The orthogonality has been proved as a consequence of (24). Let $\phi \in L^2(F)$ be a continuous function

such that $\text{cov}(\phi(X), h_n(X)) = 0$, $n \geq 1$. Suppose that ϕ' exists. As (ψ_n) is a complete system $\phi' = \sum_{n \geq 1} c_n \psi_n$ and integrating, we have $\phi = c_0 + \sum_{n \geq 1} c_n h_n$. But $\text{cov}(\phi(X), h_n(X)) = c_n \lambda_n = 0$, $n \geq 1$, which shows that ϕ must be constant. \square

4.6. The First Principal Component

In this section we prove two interesting properties of h_1 and the first principal component $h_1(X)$, see [10].

Proposition 7. *The increasing function h_1 characterizes the distribution of X .*

Proof. Write $z = h_1$. Then z satisfies the differential equation (see (26))

$$\lambda z'' + (z - \mu)z' = 0, \quad z(a) = z'(a) = 0, \quad (29)$$

where $\mu = E[z(X)]$, $\lambda = \text{var}[z(X)]$. When the function z is given, μ and λ may be obtained by solving the equations

$$\int_I \frac{-\lambda z''}{(z - \mu)} dx = 1, \quad \int_I \frac{-\lambda z''}{(z - \mu)} z dx = \mu,$$

and the density of X is $f = -\lambda z'' / (z - \mu)$. \square

Proposition 8. *For a fixed t let $\rho^2(X_t, g(X))$ denote the squared correlation between X_t and a function $g(X)$. The average of $\rho^2(X_t, \phi(X))$ weighted by $K(t, t) = F(t)[1 - F(t)]$ is maximum for $g \equiv h_1$, i.e.,*

$$\sup_g \int_I \rho^2(X_t, g(X)) K(t, t) dt = \int_I \rho^2(X_t, f_1(X)) K(t, t) dt.$$

Proof. Let us write (see Proposition 6) $g = \sum_{n \geq 1} a_n h_n$. Then $\text{var}(g(X)) = \sum_{n \geq 1} a_n^2 \lambda_n$ and we can suppose $\sum_{n \geq 1} a_n^2 = 1$. From (25)

$$\int_I \text{cov}(X_t, g(X))^2 dt = \sum_{n \geq 1} a_n^2 \lambda_n^2.$$

As $\text{var}(X_t) = K(t, t)$, we have

$$\begin{aligned} & \sup_g \int_I \rho^2(X_t, g(X)) K(t, t) dt \\ &= \left(\sum_{n \geq 1} a_n^2 \lambda_n^2 \right) / \left(\sum_{n \geq 1} a_n^2 \lambda_n \right) \leq \lambda_1 \left(\sum_{n \geq 1} a_n^2 \lambda_n \right) / \left(\sum_{n \geq 1} a_n^2 \lambda_n \right) = \lambda_1. \end{aligned}$$

Thus the supreme is attained at $g \equiv h_1$. \square

5. An Inequality

The following inequality holds for X with the normal $N(0, 1)$ distribution [31,32]:

$$\{E[g'(X)]\}^2 \leq \text{var}[g(X)] \leq E\{[g'(X)]^2\},$$

where g is an absolutely continuous function and $g(X)$ has finite variance. This inequality has been extended to other distributions by Klaassen [33]. Let us prove a related inequality concerning the function of a random variable and its derivative.

If $h_1(X)$ is the first principal dimension, then $\psi_1 = h_1' \geq 0$ and $h_1(b) = \int_I \psi_1(x) dx < \infty$. We can define the probability density f_1 with support $I = [a, b]$ given by

$$f_1(y) = \frac{\psi_1(y)}{h_1(b)}.$$

Theorem 5. *Let Y be a r.v. with pdf f_1 . If g is an absolutely continuous function and $g(X)$ has finite variance, the following inequality holds*

$$h_1(b)^2 \lambda_1 \{E[g'(Y)]\}^2 \leq \text{var}[g(X)] \leq \lambda_1 \int_I [g'(x)]^2 dx, \quad (30)$$

with equality if g is h_1 .

Proof. From Proposition 6, we can write $g' = \sum_{n \geq 1} a_n \psi_n$, where $a_1 = \int_I \psi_1(x) g'(x) dx = h_1(b) E(g'(Y))$. Then $g = \sum_{n \geq 1} a_n h_n$ and $\text{var}[h_n(X)] = \lambda_n$, so

$$\text{var}[g(X)] = \sum_{n \geq 1} a_n^2 \lambda_n.$$

From Parseval's identity $\gamma = \int_I [g'(x)]^2 dx = \sum_{n \geq 1} a_n^2$. Thus we have

$$a_1^2 \lambda_1 \leq \sum_{n \geq 1} a_n^2 \lambda_n \leq \gamma \lambda_1,$$

proving (30). Moreover, if $g = h_1$ we have $h_1(b)^2 [E(g'(Y))]^2 = 1$ and $\int_I [g'(x)]^2 dx = 1$. \square

Inequality (30) is equivalent to

$$\lambda_1 \left\{ \int_I g'(x) \psi_1(x) dx \right\}^2 \leq \text{var}[g(X)] \leq \lambda_1 \int_I [g'(x)]^2 dx. \tag{31}$$

Some examples are next given.

5.1. Uniform Distribution

Suppose X uniform $(0,1)$. Then $f_1(y) = \pi \sin(\pi y)/2$, $0 < y < 1$ and $\lambda_1 = 1/\pi^2$. We obtain

$$\frac{2}{\pi^2} \left\{ \int_0^1 g'(x) \sin(\pi x) dx \right\} \leq \text{var}[g(X)] \leq \frac{1}{\pi^2} \int_0^1 [g'(x)]^2 dx.$$

5.2. Exponential Distribution

Suppose X exponential with mean 1. Then $\lambda_1 = 4/\xi_1^2$ and

$$f_1(y) = \frac{\xi_1 \exp(-y/2) J_1(\xi_1 \exp(-y/2))}{2[1 - J_0(\xi_1)]},$$

where $\xi_1 = 3.83171$ satisfies $J_1(\xi_1) = 0$. Inequality (31) is

$$c_1 \left\{ \int_0^\infty g'(x) e^{-x/2} J_1(\xi_1 e^{-x/2}) dx \right\} \leq \text{var}[g(X)] \leq \frac{4}{\xi_1^2} \int_0^\infty [g'(x)]^2 dx,$$

where $c_1 = 4/[\xi_1^2 J_0(\xi_1)]$.

5.3. Pareto Distribution

Suppose X with density $f(x) = 1 - x^{-3}$ for $x > 1$. Then $\lambda_1 = 3/\xi_1^2$ and

$$f_1(y) = \frac{\sin(\xi_1/y) - (1/y) \cos(\xi_1/y)}{\xi_1 - \sin(\xi_1)},$$

where $\xi_1 = 4.49341$ satisfies $\xi_1 = \tan(\xi_1)$. Inequality (31) is

$$c_1 \left\{ \int_1^\infty g'(x) \left[\sin\left(\frac{\xi_1}{x}\right) - \frac{1}{x} \cos\left(\frac{\xi_1}{x}\right) \right] dx \right\} \leq \text{var}[g(X)] \leq \frac{3}{\xi_1^2} \int_1^\infty [g'(x)]^2 dx,$$

where $c_1 = 3/\{4\xi_1^3 [2\xi_1 - \sin(2\xi_1)]\}$.

5.4. Logistic Distribution

Suppose that X follows the standard logistic distribution. The cdf is

$$F(x) = [1 + \exp(-x)]^{-1}, \quad -\infty < x < +\infty,$$

and the density is $f = F(1 - F)$.

This distribution has especial interest, as the two first principal components are $h_1 = \sqrt{6}F, h_2 = \sqrt{30}f$, i.e., proportional to the cdf and the density, respectively. Note that h_1 can be obtained directly, as if we write

$h_1 = cF$, then $\mu = c/2, \lambda = c^2/12$ and (29) gives

$$\frac{c^2}{12}(1-2F) + \left(F - \frac{1}{2}\right) = 0,$$

so $c = \sqrt{6}$. Similarly, we can obtain h_2 . Besides

$$\sup_{\phi} \int_I \rho^2(X_t, \phi(X)) f(t) dt = \int_I \rho^2(X_t, F(X)) f(t) dt,$$

i.e., the expectation of $\rho^2(X_t, \phi(X))$ w.r.t. t with density $f(t)$ is maximum for $\phi = F$.

Now $\lambda_1 = 1/2$ and $h_1(b) = \sqrt{6}$. The density $f_1 = \psi_1/h_1(b)$ is just f , therefore inequality (30) for the logistic distribution reduces to

$$3[E(g'(X))]^2 \leq \text{var}[g(X)] \leq \frac{1}{2} \int_{-\infty}^{+\infty} [g'(x)]^2 dx.$$

In general, if Z is logistic with variance σ^2 then $Z = \alpha X$ with $\alpha = (\sqrt{3}/\pi)\sigma$. Noting that the functions $h'_n, n \geq 2$, are orthogonal to f and using (24), we obtain

$$\begin{aligned} & \text{cov}(F(X), g(\alpha X)) \\ &= \int_I (\sqrt{6}f(x))^2 dx \int_I \frac{\alpha}{2} f(y) g'(\alpha y) dy = \frac{\alpha}{2} E(g'(\alpha X)). \end{aligned}$$

As $\text{var}[F(X)] = 1/12$, the Cauchy-Schwarz inequality proves that

$$\text{var}[g(Z)] \geq \left(\frac{3^2}{\pi^2}\right) \text{var}(Z) [E(g'(Z))]^2.$$

6. Diagonal Expansions

Correspondence analysis is a variant of multidimensional scaling, used for representing the rows and columns of a contingency table, as points in a space of low-dimension separated by the chi-square distance. See [15,34]. This method employs a singular value decomposition (SVD) of a transformed matrix. A continuous scaling expansion, viewed as a generalization of correspondence analysis, can be obtained from (3) and (4).

6.1. Univariate Case

Let h, f be two densities with the same support I . Define the squared distance

$$\delta^2(x, x') = [h(x)/f(x) - h(x')/f(x')]^2.$$

The double-centered inner product is given by

$$G(x, x') = [h(x)/f(x) - 1][h(x')/f(x') - 1]$$

and the geometric variability is

$$\text{tr}_x(G) = \int_I [h(x)/f(x) - 1]^2 f(x) dx,$$

which is a Pearson measure of divergence between h and f . If (u_n) is an orthonormal basis for $L^2(I)$, we may consider the expansion

$$[h(x)/f(x) - 1] f(x)^{1/2} = \sum_{n \geq 1} \alpha_n u_n(x)$$

and defining $A_n = \alpha_n f^{-1/2} u_n$, then

$$h(x) = f(x) \left[1 + \sum_{n \geq 1} A_n(x) \right]$$

and

$$\delta^2(x, x') = \sum_{n \geq 1} [A_n(x) - A_n(x')]^2.$$

However, $(A_1(x), A_2(x), \dots)$ are not the principal coordinates related to the above distance. In fact, the continuous scaling dimension is 1 for this distance and it can be found in a straightforward way.

6.2. Bivariate Case

Let us write (4) as

$$h - f \otimes g = \sum_{n \geq 1} \rho_n f a_n \otimes g b_n, \tag{32}$$

where $a_n = f^{-1/2} u_n, b_n = g^{-1/2} v_n, n \geq 1$, are the canonical functions and (ρ_n) is the sequence of canonical correlations. Note that $E_x [a_n(X)] = E_y [b_n(Y)] = 0$ and $E_x [a_n(X)^2] = E_y [b_n(Y)^2] = 1$.

Suppose that h is absolutely continuous w.r.t. $f \times g$ and let us consider the Radom-Nikodym derivative

$$r(x, y) = \frac{h(x, y)}{f(x)g(y)}.$$

The so-called chi-square distance between $x, x' \in [a, b]$ is given by

$$\begin{aligned} \delta^2(x, x') &= \int_c^d [f(y/x) - f(y/x')]^2 g(y)^{-1} dy \\ &= E_y \{ [r(x, Y) - r(x', Y)]^2 \}. \end{aligned}$$

Let $z = \int_a^b \int_c^d [h(x, y)/(f(x)g(y))]^2 dx dy$, the Pearson contingency coefficient is defined by

$$\phi^2 = \int_a^b \int_c^d \frac{[h(x, y) - f(x)g(y)]^2}{f(x)g(y)} dx dy = z - 1.$$

The geometric variability of the chi-square distance is $V_\delta = \phi^2$. In fact

$$2V_\delta = \int_a^b \int_a^b \int_c^d [f(y/x) - f(y/x')]^2 g(y)^{-1} dx dx' dy = z + z - 2.$$

The proximity function is

$$\begin{aligned} D_\delta^2(x) &= E_x [\delta^2(x, X)] - V_\delta = E_y [r^2(x, Y)] - 1 \\ &= E_y \{ [r(x, Y) - 1][r(x, Y) - 1] \}, \end{aligned}$$

and the double-centered inner product is

$$\begin{aligned} G(x, x') &= E_y \{ r(x, Y)r(x', Y) \} - 1 \\ &= E_y \{ [r(x, Y) - 1][r(x', Y) - 1] \}. \end{aligned}$$

We can express (32) as

$$f^{1/2} \otimes g^{1/2} (r - 1) = \sum_{n \geq 1} \rho_n u_n \otimes v_n.$$

This SVD exists provided that $\phi^2 < \infty$. Multiplying $f^{1/2} \otimes g^{1/2} (r - 1)$ by himself and integrating w.r.t. y we readily obtain

$$f^{1/2} \otimes f^{1/2} G = \sum_{n \geq 1} \rho_n^2 u_n \otimes u_n.$$

Comparing with (9) we have the principal coordinates $(A_1(x), A_2(x), \dots)$, where $A_n = \rho_n f^{-1/2} u_n$, which satisfy

$$\delta^2(x, x') = \sum_{n \geq 1} [A_n(x) - A_n(x')]^2.$$

Then

$$\phi^2 = \sum_{n \geq 1} \rho_n^2.$$

See [18] for further details.

Finally, the following expansion in terms of cdf's holds [35]:

$$H(x, y) - F(x)G(y) = \sum_{n \geq 1} \rho_n \int_a^b M_1(x, s) da_n(s) \int_c^d M_2(t, y) db_n(t), \quad (33)$$

where $M_1(x, y) = \min\{F(x), F(y)\} - F(x)F(y)$, $M_2(x, y) = \min\{G(x), G(y)\} - G(x)G(y)$. Using a matrix notation, this diagonal expansion can be written as

$$H - F \otimes G = D_\rho M_1 * (dA) \otimes M_2 * (dB),$$

where D_ρ stands for the diagonal matrix with the canonical correlations, and $dA = (da_1, da_2, \dots)'$, $dB = (db_1, db_2, \dots)'$.

7. The Covariance between Two Functions

Here we generalize the well-known Hoeffding's formula

$$\text{cov}(X, Y) = \int_a^b \int_c^d [H(x, y) - F(x)G(y)] dx dy,$$

which provides the covariance in terms of the bivariate and univariate cdf's. The proof of the generalization below uses Fubini's theorem and integration by parts, being different from the proof given in [28].

Let us suppose that the supports of X, Y are the intervals $[a, b], [c, d] \subset \mathbb{R}$, respectively, although the results may also hold for other subsets of \mathbb{R} . We then have

$$F(a) = G(c) = 0, F(b) = G(d) = 1.$$

Theorem 6. *If $\alpha(x), \beta(y)$ are two functions defined on $[a, b], [c, d]$, respectively, such that:*

1. *Both functions are of bounded variation.*
2. *$\alpha(a)F(a) = \beta(c)G(c) = 0$.*
3. *The expectations $E(|\alpha(X)\beta(Y)|), E(|\alpha(X)|), E(|\beta(Y)|)$ are finite.*

Then:

$$\text{cov}(\alpha(X), \beta(Y)) = \int_a^b \int_c^d [H(x, y) - F(x)G(y)] d\alpha(x) d\beta(y). \quad (34)$$

Proof. The covariance exists and is

$$I_K = \int_S \alpha(x)\beta(y) dH(x, y) - \int_a^b \alpha(x) dF(x) \int_c^d \beta(y) dG(y),$$

where $S = [a, b] \times [c, d]$. Integration by parts gives

$$A = \int_a^b \alpha(x) dF(x) = \alpha(b) - \int_a^b F(x) d\alpha(x) = \alpha(b) - I_F,$$

and similarly $B = \beta(d) - I_G$. By Fubini's theorem for transition probabilities

$$E[\phi(X, Y)] = \int_a^b \left(\int_c^d \phi(x, y) dF_x(y) \right) dF(x),$$

where $F_x(y)$ is the cdf of Y given $X = x$, we can write

$$C = \int_S \alpha(x)\beta(y) dH(x, y) = \int_a^b \alpha(x) dF(x) \int_c^d \beta(y) dF_x(y).$$

We first integrate with respect to y . Setting $u = \beta(y), v = \int_c^y dF_x(t)$, to find $\int_c^d u dv$, integration by parts gives

$$\int_c^d \beta(y) dF_x(y) = \beta(d) \int_c^d dF_x(t) - \int_c^y \int_c^y dF_x(t) d\beta(y).$$

Since $dF(x)dF_x(t) = dH(x, t)$, setting $u = \alpha(x), v = \int_a^x \int_c^d dH(x, t) = F(x)$, we find

$$\begin{aligned} C_1 &= \beta(d) \int_a^b \alpha(x) \int_c^d dF(x) dF_x(t) \\ &= \beta(d) \left[\alpha(b) - \int_a^b F(x) d\alpha(x) \right], \end{aligned}$$

and setting $u = \alpha(x), v = \int_c^d \int_a^x \int_c^y dH(s,t) d\beta(y) = \int_c^d H(x,y) d\beta(y)$, again integration by parts gives

$$\begin{aligned} C_2 &= \int_a^b \alpha(x) dF(x) \int_c^d \int_c^y dF_x(t) d\beta(y) \\ &= \alpha(x) \int_c^d H(x,y) d\beta(y) \Big|_a^b - I_H, \\ &= \alpha(b) \int_c^d G(y) d\beta(y) - I_H, \end{aligned}$$

where $I_H = \int_S H(x,y) d\alpha(x) d\beta(y)$. Therefore

$$C = C_1 - C_2 = \alpha(b)\beta(d) - \beta(d) \int_a^b F(x) d\alpha(x) - \alpha(b) \int_c^d G(y) d\beta(y) + I_H.$$

A last simplification shows that $I_K = C - A \times B = I_H - I_F \times I_G$. \square

8. Canonical Analysis

Given two sets of variables, the purpose of canonical correlation analysis is to find sets of linear combinations with maximal correlation. In Section 6.2 we studied, from a multidimensional scaling perspective, the nonlinear canonical functions of X and Y , with joint pdf h . Here we find the canonical correlations and functions for several copulas.

Let (U, V) be a bivariate random vector with cdf $C(u, v)$, where U and V are $(0,1)$ uniform. Then C is a cdf called copula [36]. Let us suppose C symmetric. Then $\text{cov}(\phi(U), \phi(V)) = \text{cov}(\phi(U), \phi(V))$. Therefore, in finding the canonical functions, we can suppose $\phi = \varphi$.

Let us consider the symmetric kernels

$$K(u, v) = C(u, v) - uv, \quad L(u, v) = \min\{u, v\} - uv.$$

We seek the canonical functions $\phi(U), \phi(V)$ and the corresponding canonical correlations, i.e.,

$$\rho = \sup_{\phi} \text{cor}(\phi(U), \phi(V)). \tag{35}$$

Definition 4. A generalized eigenfunction of K w.r.t. L with eigenvalue λ is a function ϕ such that

$$\int_0^1 K(u, v) d\phi(v) = \lambda \int_0^1 L(u, v) d\phi(v). \tag{36}$$

From the theory of integral equations, we have

$$\begin{aligned} \lambda &= \sup_{\phi} \int_0^1 K(u, v) d\phi(u) d\phi(v) \\ &\text{constrained to } \int_0^1 L(u, v) d\phi(u) d\phi(v) = 1. \end{aligned} \tag{37}$$

Definition 5. On the set of functions in $L^2(I)$, with $I = [0, 1]$, we define the inner products

$$\begin{aligned} (\phi, K\phi) &= \int_{I^2} K(u, v) d\phi(u) d\phi(v), \\ (\phi, L\phi) &= \int_{I^2} L(u, v) d\phi(u) d\phi(v). \end{aligned}$$

Clearly, see Theorem 6, if ϕ is eigenfunction of K w.r.t. L with eigenvalue λ , we have the covariance $\text{cov}(\phi(U), \phi(V)) = (\phi, K\phi)$ and the variance $\text{var}[\phi(U)] = (\phi, L\phi)$. Therefore the correlation between $\phi(U), \phi(V)$ is:

$$\text{cor}(\phi(U), \phi(V)) = \frac{(\phi, K\phi)}{(\phi, L\phi)} = \lambda.$$

Thus the canonical correlations of (U, V) are the eigenvalues of K w.r.t. L .

Justified by the embedding in a Euclidean or Hilbert space via the chi-square distance, the geometric dimension of a copula C is defined as the cardinal of the set (ρ_n) of canonical correlations. The dimension can be finite, infinite countable (cardinality of \aleph_0) or uncountable (cardinality of the continuum \mathcal{C}).

We next illustrate these results with some copulas. Since $\min\{u, v\}$ is a copula, the so-called Fréchet-

Hoeffding upper bound, we firstly suggest a procedure for constructing copulas and performing canonical analysis. This procedure is based on the expansion (23) for the logistic distribution.

For this distribution, if $c_n = n(n+1)$, $d_n = 2n+1$, we have [27]:

$$\min\{F(x), F(y)\} - F(x)F(y) = \sum_{n \geq 1} c_n^{-1} \psi_n(F(x)) \psi_n(F(y)),$$

where $\psi_n(x)$ is given in (27). With the change $u = F(x)$, $v = F(y)$, we find:

$$\min\{u, v\} = uv + \sum_{n \geq 1} c_n^{-1} \psi_n^*(u) \psi_n^*(v), \quad (38)$$

with $0 \leq u, v \leq 1$ and

$$\begin{aligned} \psi_n^*(u) &= \frac{1}{2} \sqrt{\frac{d_n}{c_n}} [nP_{n-1}(2u-1) - n(2u-1)P_n(2u-1)] \\ &= \frac{1}{2} \sqrt{\frac{c_n}{d_n}} [P_{n-1}(2u-1) - P_{n+1}(2u-1)] \\ &= \frac{1}{2} \sqrt{\frac{c_n}{d_n}} \left[\frac{1}{\sqrt{d_n-2}} L_{n-1}(u) - \frac{1}{\sqrt{d_n+2}} L_{n+1}(u) \right], \end{aligned}$$

where P_n are Legendre polynomials and L_n are shifted Legendre polynomials on $[0, 1]$. Thus $\psi_1^*(u) = \sqrt{6}u(1-u)$, $\psi_2^*(u) = \sqrt{30}u(1-u)(2u-1)$, $\psi_3^*(u) = \sqrt{84}u(-5u^3+10u^2-6u+1)$, etc.

8.1. FGM Copula

If we consider only ψ_1^* in (38), we obtain the Farlie-Gumbel-Morgenstern copula:

$$FGM_\theta(u, v) = uv[1 + \theta(1-u)(1-v)], \quad -1 \leq \theta \leq 1.$$

Then $\int_0^1 K(u, v) d\phi(v) = (\theta/6)u(1-u)$ and $\int_0^1 L(u, v) d\phi(v) = (1/2)u(1-u)$ if $\phi(v) = v$. Thus v is an eigenfunction with eigenvalue $\theta/3$. Then u and v are the canonical functions with canonical correlation $|\theta|/3$. The geometric dimension is one.

8.2. Extended FGM Copula

By taking more terms in expansion (38), we may consider the following copula

$$\begin{aligned} C_3(u, v) &= uv + \theta_1 u(1-u)v(1-v) + \theta_2(2u-1)u(1-u)(2v-1)v(1-v) \\ &\quad + \theta_3 u(5u^3 - 10u^2 + 6u - 1)v(5v^3 - 10v^2 + 6v - 1), \end{aligned}$$

where $-1 \leq \theta_i \leq 1$. The canonical correlations are $\rho_1 = |\theta_1/3|$, $\rho_2 = |\theta_2/5|$, $\rho_3 = |\theta_3/7|$. The canonical functions are $\psi_1^{*'}(u) = \sqrt{6}(1-2u)$, $\psi_2^{*'}(u) = \sqrt{30}(6u^2-6u+1)$, $\psi_3^{*'}(u) = \sqrt{84}(20u^3-30u^2+6u-1)$, respectively. This copula has dimension 3.

Clearly C_3 reduces to the FGM copula if $\theta_2 = \theta_3 = 0$. When the dimension is 2, i.e., $\theta_3 = 0$, we have a copula with cubic sections [37]. A generalization is given in [38].

8.3. Cuadras-Augé Copula

The Cuadras-Augé family of copulas [39] is defined as the weighted geometric mean of $\min\{u, v\}$ and uv :

$$CA_\theta(u, v) = (uv)^{1-\theta} (\min\{u, v\})^\theta, \quad 0 \leq \theta \leq 1.$$

For this copula, the canonical correlations constitute a continuous set. If $\phi_\gamma(x) = 1$ for $x = \gamma$ and 0 otherwise, it can be proved [40] that the set $(\phi_\gamma, \lambda_\gamma)$ of eigenpairs of $K_\theta(u, v) = CA_\theta(u, v) - uv$ w.r.t. $L = K_1$, is given by

$$\phi_\gamma, \lambda_\gamma = \theta\gamma^{1-\theta}, \quad 0 \leq \gamma \leq 1.$$

Thus, the set of canonical functions and correlations for the Cuadras-Augé copula is the uncountable set $(\phi_\gamma, \theta\gamma^{1-\theta})$, $0 \leq \gamma \leq 1$, with dimension of the power of the continuum. In particular, the maximum correlation is the parameter θ with canonical function the Heaviside distribution \mathcal{H}_1 . The maximum correlation θ was obtained by Cuadras [28].

REFERENCES

- [1] T. W. Anderson and M. A. Stephens, "The Continuous and Discrete Brownian Bridges: Representation and Applications," *Linear Algebra and Its Applications*, Vol. 264, 1996, pp. 145-171. [http://dx.doi.org/10.1016/S0024-3795\(97\)00015-3](http://dx.doi.org/10.1016/S0024-3795(97)00015-3)
- [2] J. Durbin and M. Knott, "Components of Cramér-von Mises Statistics. I," *Journal of the Royal Statistical Society: Series B*, Vol. 34, 1972, pp. 290-307.
- [3] J. Fortiana and C. M. Cuadras, "A Family of Matrices, the Discretized Brownian Bridges and Distance-Based Regression," *Linear Algebra and Its Applications*, Vol. 264, 1997, pp. 173-188. [http://dx.doi.org/10.1016/S0024-3795\(97\)00051-7](http://dx.doi.org/10.1016/S0024-3795(97)00051-7)
- [4] J. L. Doob, "Stochastic Processes," Wiley, New York, 1953.
- [5] M. Loève, "Probability Theory," 3rd Edition, Van Nostrand, Princeton, 1963.
- [6] G. K. Eagleson, "Orthogonal Expansions and U-Statistics," *Australian Journal of Statistics*, Vol. 21, No. 3, 1979, pp. 221-237. <http://dx.doi.org/10.1111/j.1467-842X.1979.tb01141.x>
- [7] C. M. Cuadras and D. Cuadras, "Orthogonal Expansions and Distinction between Logistic and Normal," In: C. Huber-Carol, N. Balakrishnan, M. S. Nikulin and M. Mesbah, Eds., *Goodness-Of-Fit Tests and Model Validity*, Birkhäuser, Boston, 2002, pp. 327-339. http://dx.doi.org/10.1007/978-1-4612-0103-8_24
- [8] J. Fortiana and A. Grané, "Goodness-Of-Fit Tests Based on Maximum Correlations and Their Orthogonal Decompositions," *Journal of the Royal Statistical Society: Series B*, Vol. 65, No. 1, 2003, pp. 115-126. <http://dx.doi.org/10.1111/1467-9868.00375>
- [9] C. M. Cuadras, "Diagonal Distributions via Orthogonal Expansions and Tests of Independence," In: C. M. Cuadras, J. Fortiana and J. A. Rodríguez-Lallena, Eds., *Distributions with Given Marginals and Statistical Modelling*, Kluwer Academic Press, Dordrecht, 2002, pp. 35-42. http://dx.doi.org/10.1007/978-94-017-0061-0_5
- [10] C. M. Cuadras, "First Principal Component Characterization of a Continuous Random Variable," In: N. Balakrishnan, I. G. Bairamov and O. L. Gebizlioglu, Eds., *Advances on Models, Characterizations and Applications*, Chapman and Hall/CRC, London, 2005, pp. 189-199. <http://dx.doi.org/10.1201/9781420028690.ch12>
- [11] C. M. Cuadras and J. Fortiana, "Continuous Metric Scaling and Prediction," In: C. M. Cuadras and C. R. Rao, Eds., *Multivariate Analysis, Future Directions 2*, Elsevier Science Publishers B. V. (North-Holland), Amsterdam, 1993, pp. 47-66.
- [12] C. M. Cuadras and J. Fortiana, "A Continuous Metric Scaling Solution for a Random Variable," *Journal of Multivariate Analysis*, Vol. 52, No. 1, 1995, pp. 1-14. <http://www.sciencedirect.com/science/article/pii/S0047259X85710019>
<http://dx.doi.org/10.1006/jmva.1995.1001>
- [13] C. M. Cuadras and J. Fortiana, "Weighted Continuous Metric Scaling," In: A. K. Gupta and V. L. Girko, Eds., *Multidimensional Statistical Analysis and Theory of Random Matrices*, VSP, Zeist, 1996, pp. 27-40.
- [14] C. M. Cuadras and J. Fortiana, "The Importance of Geometry in Multivariate Analysis and Some Applications," In: C. R. Rao and G. Székely, Eds., *Statistics for the 21st Century*, Marcel Dekker, New York, 2000, pp. 93-108.
- [15] C. M. Cuadras and D. Cuadras, "A Parametric Approach to Correspondence Analysis," *Linear Algebra and Its Applications*, Vol. 417, No. 1, 2006, pp. 64-74. <http://www.sciencedirect.com/science/article/pii/S0024379505005203>
<http://dx.doi.org/10.1016/j.laa.2005.10.029>
- [16] C. M. Cuadras and D. Cuadras, "Eigenanalysis on a Bivariate Covariance Kernel," *Journal of Multivariate Analysis*, Vol. 99, No. 10, 2008, pp. 2497-2507. <http://www.sciencedirect.com/science/article/pii/S0047259X08000754>
<http://dx.doi.org/10.1016/j.jmva.2008.02.039>
- [17] H. O. Lancaster, "The Chi-Squared Distribution," Wiley, New York, 1969.
- [18] C. M. Cuadras, J. Fortiana and M. J. Greenacre, "Continuous Extensions of Matrix Formulations in Correspondence Analysis, with Applications to the FGM Family of Distributions," In: R. D. H. Heijmans, D. S. G. Pollock and A. Satorra, Eds., *Innovations in Multivariate Statistical Analysis*, Kluwer Academic Publisher, Dordrecht, 2000, pp. 101-116. http://dx.doi.org/10.1007/978-1-4615-4603-0_7
- [19] C. M. Cuadras, J. Fortiana and F. Oliva, "The Proximity of an Individual to a Population with Applications in Discriminant Analysis," *Journal of Classification*, Vol. 14, No. 1, 1997, pp. 117-136. <http://dx.doi.org/10.1007/s003579900006>
- [20] K. V. Mardia, J. T. Kent and J. M. Bibby, "Multivariate Analysis," Academic Press, London, 1979.
- [21] T. F. Cox and M. A. Cox, "Multidimensional Scaling," Chapman and Hall, London, 1994.
- [22] I. J. Schoenberg, "Remarks to Maurice Fréchet's Article 'Sur la définition axiomatique d'une classe d'espaces vectoriels distanciés applicables vectoriellement sur l'espace de Hilbert'," *Annals of Mathematics*, Vol. 36, No. 3, 1935, pp. 724-732.

<http://dx.doi.org/10.2307/1968654>

- [23] C. M. Cuadras, "Distance Analysis in Discrimination and Classification Using Both Continuous and Categorical Variables," In: Y. Dodge, Ed., *Statistical Data Analysis and Inference*, Elsevier Science Publishers B. V. (North-Holland), Amsterdam, 1989, pp. 459-473.
- [24] C. M. Cuadras, E. A. Atkinson and J. Fortiana, "Probability Densities from Distances and Discriminant Analysis," *Statistics and Probability Letters*, Vol. 33, No. 4, 1997, pp. 405-411. [http://dx.doi.org/10.1016/S0167-7152\(96\)00154-X](http://dx.doi.org/10.1016/S0167-7152(96)00154-X)
- [25] C. R. Rao, "Diversity: Its Measurement, Decomposition, Apportionment and Analysis," *Sankhyā: The Indian Journal of Statistics, Series A*, Vol. 44, No. 1, 1982, pp. 1-21.
- [26] Z. Liu and C. R. Rao, "Asymptotic Distribution of Statistics Based on Quadratic Entropy and Bootstrapping," *Journal of Statistical Planning and Inference*, Vol. 43, No. 1-2, 1995, pp. 1-18. [http://dx.doi.org/10.1016/0378-3758\(94\)00005-G](http://dx.doi.org/10.1016/0378-3758(94)00005-G)
- [27] C. M. Cuadras and Y. Lahlou, "Some Orthogonal Expansions for the Logistic Distribution," *Communications in Statistics—Theory and Methods*, Vol. 29, No. 12, 2000, pp. 2643-2663. <http://dx.doi.org/10.1080/03610920008832629>
- [28] C. M. Cuadras, "On the Covariance between Functions," *Journal of Multivariate Analysis*, Vol. 81, No. 1, 2002, pp. 19-27. <http://www.sciencedirect.com/science/article/pii/S0047259X01920007>
<http://dx.doi.org/10.1006/jmva.2001.2000>
- [29] C. M. Cuadras and Y. Lahlou, "Principal Components of the Pareto Distribution," In: C. M. Cuadras, J. Fortiana and J. A. Rodriguez-Lallena, Eds., *Distributions with Given Marginals and Statistical Modelling*, Kluwer Academic Press, Dordrecht, 2002, pp. 43-50. http://dx.doi.org/10.1007/978-94-017-0061-0_6
- [30] E. Salinelli, "Nonlinear Principal Components, II: Characterization of Normal Distributions," *Journal of Multivariate Analysis*, Vol. 100, No. 4, 2009, pp. 652-660. <http://dx.doi.org/10.1016/j.jmva.2008.07.001>
- [31] H. Chernoff, "A Note on an Inequality Involving the Normal Distribution," *Annals of Probability*, Vol. 9, No. 3, 1981, pp. 533-535. <http://dx.doi.org/10.1214/aop/1176994428>
- [32] T. Cacoullos, "On Upper and Lower Bounds for the Variance of a Function of a Random Variable," *Annals of Probability*, Vol. 10, No. 3, 1982, pp. 799-809. <http://dx.doi.org/10.1214/aop/1176993788>
- [33] C. A. J. Klaassen, "On an Inequality of Chernoff," *Annals of Probability*, Vol. 13, No. 3, 1985, pp. 966-974. <http://dx.doi.org/10.1214/aop/1176992917>
- [34] M. J. Greenacre, "Theory and Applications of Correspondence Analysis," Academic Press, London, 1984.
- [35] C. M. Cuadras, "Correspondence Analysis and Diagonal Expansions in Terms of Distribution Functions," *Journal of Statistical Planning and Inference*, Vol. 103, No. 1-2, 2002, pp. 137-150. [http://dx.doi.org/10.1016/S0378-3758\(01\)00216-6](http://dx.doi.org/10.1016/S0378-3758(01)00216-6)
- [36] R. B. Nelsen, "An Introduction to Copulas," 2nd Edition, Springer, New York, 2006.
- [37] R. B. Nelsen, J. J. Quesada-Molina and J. A. Rodriguez-Lallena, "Bivariate Copulas with Cubic Sections," *Journal of Nonparametric Statistics*, Vol. 7, No. 3, 1997, pp. 205-220. <http://dx.doi.org/10.1080/10485259708832700>
- [38] C. M. Cuadras and W. Daz, "Another Generalization of the Bivariate FGM Distribution with Two-Dimensional Extensions," *Acta et Commentationes Universitatis Tartuensis de Mathematica*, Vol. 16, No. 1, 2012, pp. 3-12. <http://math.ut.ee/acta/16-1/CuadrasDiaz.pdf>
- [39] C. M. Cuadras and J. Augé, "A Continuous General Multivariate Distribution and Its Properties," *Communications in Statistics—Theory and Methods*, Vol. 10, No. 4, 1981, pp. 339-353. <http://dx.doi.org/10.1080/03610928108828042>
- [40] C. M. Cuadras, "Continuous Canonical Correlation Analysis," *Research Letters in the Information and Mathematical Sciences*, Vol. 8, 2005, pp. 97-103. <http://muir.massey.ac.nz/handle/10179/4454>