ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
21,2 (1991), pp. 211–224

W. WYSOCKI (Warszawa)

GEOMETRICAL ASPECTS OF MEASURES OF DEPENDENCE FOR RANDOM VECTORS

Introduction. The paper consists of two parts: geometrical and probabilistic.

In the first part we introduce the notion of an angle between linear subspaces and of its measure. Let $(\mathcal{X},\mathcal{Y})$ be a pair of closed subspaces of a real separable Hilbert space $(H,\langle\cdot,\cdot\rangle)$ and let P_1', P_2' be the orthogonal projections of $\mathcal{X}$ on $\mathcal{Y}$ and of $\mathcal{Y}$ on $\mathcal{X}$, respectively. The angle is defined for pairs $(\mathcal{X},\mathcal{Y})$ such that $P_2'P_1'$ and $P_1'P_2'$ are compact. The set of all pairs $(\mathcal{X},\mathcal{Y})$ with this property will be denoted by $\mathcal{K}(H)$; by a^2 we will denote the largest eigenvalue of $P_2'P_1'$ among eigenvalues smaller than 1, which is equal to the respective eigenvalue of $P_1'P_2'$. Then, for any $(\mathcal{X},\mathcal{Y})\in\mathcal{K}(H)$ we define the angle between $\mathcal{X}$ and $\mathcal{Y}$ as the angle between unit eigenvectors of $P_2'P_1'$ and $P_1'P_2'$ which correspond to a^2, and $\arccos a$ is said to be the measure of this angle.

If H is finite-dimensional then $\mathcal{K}(H)$ consists of all pairs of subspaces of H.

In the probabilistic part of the paper the cosine of the angle between subspaces $\mathcal{X}$ and $\mathcal{Y}$ of a real separable Hilbert space $(L^2(\Omega,\mathcal{A},P;\mathbf{R}^k),\langle\cdot,\cdot\rangle_\Lambda)$ is considered. This space consists of all k-dimensional centered random vectors with finite covariance matrix, and the scalar product is given by

$$\langle X,Y\rangle_\Lambda = E(X^T\Lambda^{-1}Y)=\mathrm{tr}(\Lambda^{-1}\Sigma_{12}),$$

where Λ is a positive definite symmetric $k\times k$ matrix, and Σ_{12} is the covari-

---

1991 *Mathematics Subject Classification*: 62H20, 62H99.

Key words and phrases: Hilbert space, compact operator, angle between linear subspaces and its measure, Sampson's measure of closed dependence, Jupp & Mardia's measure of dependence, Sampson's correlation ratio, maximal canonical correlation, coefficient of maximal correlation, conditional expectation.

ance matrix of X and Y. Thus, we define for $U \in \mathcal{X}$, $V \in \mathcal{Y}$, $(\mathcal{X}, \mathcal{Y}) \in \mathcal{K}(H)$,

$$\varrho_\Lambda(U,V) = \frac{\langle U,V\rangle}{\|U\|_\Lambda \|V\|_\Lambda},$$

which may be interpreted as the cosine of the angle between U and V. Maximal values of this expression on suitably chosen sets of pairs of random vectors will serve as measures of dependence between X and Y. Let $\mathcal{L}_X$ be the subspace spanned by X, let $\mathcal{H}_X = \{AX \mid A \in \mathcal{M}_k\}$ where $\mathcal{M}_k$ is the set of all $k \times k$ matrices, and let $\mathcal{L}^2_X$ be the set of all square integrable random vectors of the form $f(X)$ where f is a Borel measurable vector function; then for $i = 1, \ldots, 5$

$$\varrho^{(i)}_\Lambda(X,Y) = \sup\{\varrho_\Lambda(U,V) : U \in \mathcal{X}_i,\ V \in \mathcal{Y}_i\},$$

where $(\mathcal{X}_i, \mathcal{Y}_i)$ is respectively equal to $(\mathcal{H}_X, \mathcal{L}_Y)$, $(\mathcal{L}^2_X, \mathcal{L}_Y)$, $(\mathcal{H}_X, \mathcal{H}_Y)$, $(\mathcal{L}^2_X, \mathcal{H}_Y)$, $(\mathcal{L}^2_X, \mathcal{L}^2_Y)$. The supremum is attained if at least one of the subspaces $\mathcal{X}_i$ and $\mathcal{Y}_i$ is of finite dimension, i.e. for $i \le 4$. Then there exist (X_i, Y_i) such that $\varrho^{(i)}_\Lambda(X,Y) = \varrho^{(i)}_\Lambda(X_i, Y_i)$. Moreover, for $i \le 4$, $(\varrho^{(i)}_\Lambda(X,Y))^2$ is the maximal eigenvalue of $P'_2 P'_1$ and $P'_1 P'_2$.

Since the measures $\varrho^{(3)}_\Lambda(X,Y)$ are identical for any Λ, the subscript Λ may be omitted. The same concerns $\varrho^{(4)}_\Lambda$, and also $\varrho^{(5)}_\Lambda$ for any (X,Y) such that $\varrho^{(5)}_\Lambda(X,Y)$ is defined.

The family $(\varrho^{(1)}_\Lambda, \Lambda \in \mathcal{M}_k)$ is strictly related to the measures of dependence proposed by Höschel [1] and by Jupp and Mardia [3]. Höschel's measure is defined in a complicated manner and it is computationally inconvenient, while Jupp and Mardia gave no justification for their proposal. We show that for any (X,Y) Höschel's measure is equal to $\varrho^{(1)}_\Lambda$ for Λ equal to the covariance matrix of Y and that Höschel's measure is equal to the square of that proposed by Jupp and Mardia divided by the squared rank of the covariance matrix of Y.

The measures $\varrho^{(2)}_\Lambda$, $\Lambda \in \mathcal{M}_k$, introduced by Sampson [7], generalize the correlation ratio between a random variable and a random vector. The measure $\varrho^{(3)}$, known as maximal canonical correlation, was proposed by Johnson and Vehrly [2]. The measure $\varrho^{(4)}$ is closely related to another measure defined by Sampson [7]: Sampson's measure is of the same form but $\mathcal{H}_Y$ is restricted to BY such that $\operatorname{cov}(BY) = \Lambda$.

Sampson remarked in his paper that the procedure of maximizing $\varrho_\Lambda(U,V)$ on $\mathcal{L}^2_X \times \mathcal{H}_Y$ can be treated as a half way between maximal canonical correlation $\varrho^{(3)}(X,Y)$ and maximal correlation $\varrho^{(5)}(X,Y)$.

Maximization of $\varrho_\Lambda(f(X), g(Y))$ on $\mathcal{L}^2_X \times \mathcal{L}^2_Y$ which is performed for $i = 5$ requires the compactness of $P'_2 P'_1$ and $P'_1 P'_2$. It is shown that the operators are compact if $P_{(X,Y)}$ is absolutely continuous with respect to

$P_X \otimes P_Y$ and the square of the density of $P_{(X,Y)}$ is integrable. Obviously, P_X, P_Y, $P_{(X,Y)}$ are the distributions generated by the random vectors X, Y and (X, Y), respectively. For (X, Y) satisfying these assumptions maximal correlation exists (Theorem 2) and is identical for any matrix Λ. Clearly, it generalizes the maximal correlation considered by Rényi [5] for a pair of random variables.

Generally, $\varrho_\Lambda^{(i)}(X, Y)$ is the cosine of the angle between the respective subspaces for any $i \le 5$ and for any (X, Y) such that $\varrho_\Lambda^{(i)}(X, Y) < 1$.

1. Angle between linear subspaces. Let $(H, \langle\cdot,\cdot\rangle)$ be a separable real Hilbert space, and let $\mathcal{X}$, $\mathcal{Y}$ be its closed subspaces.

Let $\mathcal{Z}$ denote $\mathcal{X} \cap \mathcal{Y}$ and let $\mathcal{X}_0$, $\mathcal{Y}_0$ be the orthogonal complements of $\mathcal{Z}$ in $\mathcal{X}$ and $\mathcal{Y}$, respectively. We assume that $\mathcal{X} \neq \mathcal{Z}$ and $\mathcal{Y} \neq \mathcal{Z}$. Continuity of the scalar product implies closedness of $\mathcal{X}_0$ and $\mathcal{Y}_0$.

Let $(x_i, i \in I)$ and $(y_j, j \in J)$ be orthonormal bases of $\mathcal{X}_0$ and $\mathcal{Y}_0$, where I and J are some sets of indices, and let

$$B_{\mathcal{X}_0} = \{x \in \mathcal{X}_0 : \|x\| = 1\}, \quad B_{\mathcal{Y}_0} = \{y \in \mathcal{Y}_0 : \|y\| = 1\},$$
$$s = s(\mathcal{X}, \mathcal{Y}) = \sup\{\langle x, y\rangle : x \in B_{\mathcal{X}_0},\ y \in B_{\mathcal{Y}_0}\},$$
$$p_0 = \operatorname{card} I, \quad q_0 = \operatorname{card} J.$$

Furthermore, let $P_1 : \mathcal{X}_0 \to \mathcal{Y}_0$, $P_2 : \mathcal{Y}_0 \to \mathcal{X}_0$ be the linear transformations defined by

$$P_1 x = \sum_{j \in J} \langle x, y_j\rangle y_j, \quad x \in \mathcal{X}_0, \tag{1.1}$$

$$P_2 y = \sum_{i \in I} \langle y, x_i\rangle x_i, \quad y \in \mathcal{Y}_0, \tag{1.2}$$

Thus, P_1 (P_2) is the orthogonal projection of $\mathcal{X}_0$ on $\mathcal{Y}_0$ (of $\mathcal{Y}_0$ on $\mathcal{X}_0$). By the continuity of P_1, P_2 the operators T_1, T_2, defined as

$$T_1 := P_2 \circ P_1, \tag{1.3}$$
$$T_2 := P_1 \circ P_2, \tag{1.4}$$

are also continuous.

THEOREM 1. *If T_1, T_2 are compact then there exist vectors $x_0 \in B_{\mathcal{X}_0}$ and $y_0 \in B_{\mathcal{Y}_0}$ such that*

$$s(\mathcal{X}, \mathcal{Y}) = \langle x_0, y_0\rangle, \tag{1.5}$$
$$T_1 x_0 = s^2 x_0, \tag{1.6}$$
$$T_2 y_0 = s^2 y_0. \tag{1.7}$$

Moreover, s^2 is the largest eigenvalue of T_1 and T_2.

Proof. First, we will show that P_1 and P_2 are adjoint to each other, i.e.

$$\langle P_1 u, z\rangle = \langle u, P_2 z\rangle \tag{1.8}$$

for any $u \in \mathcal{X}_0$, $z \in \mathcal{Y}_0$. The left-hand side of (1.8) can be rewritten as $\overline{u}^T \Sigma_{12} \overline{z}$, where

$$\Sigma_{12} = [\langle x_i, y_j\rangle,\ i = 1,\ldots,p_0,\ j = 1,\ldots,q_0],$$

and the vectors $\overline{u}$, $\overline{z}$ are the counterparts of u, z in the coordinate spaces $\mathbf{R}^{p_0}$ and $\mathbf{R}^{q_0}$ which obvoiusly can be identified ([1]) with $\mathcal{X}_0$ and $\mathcal{Y}_0$.

The right-hand side of (1.8) has the same form, and therefore $P_1^* = P_2$, where P_1^* is adjoint to P_1.

The operators T_1 and T_2 are self-adjoint, since

$$T_1^* = (P_2 \circ P_1)^* = P_1^* \circ P_2^* = P_2 \circ P_1 = T_1$$

and similarly $T_2^* = T_2$. They are also nonnegative definite: for any $x \in \mathcal{X}_0$

$$\langle T_1 x, x\rangle = \langle (P_2 P_1)x, x\rangle = \langle P_1 x, P_1 x\rangle = \|P_1 x\|^2 \geq 0$$

and the same holds for T_2.

Now we will prove that $\|P_1\| = \|P_2\| = s$. Note first that the scalar product $\langle\cdot,\cdot\rangle$, restricted to $\mathcal{X}_0 \times \mathcal{Y}_0$, can be expressed as

$$\langle x, y\rangle = \langle x, P_2 y\rangle, \tag{1.9}$$

$$\langle x, y\rangle = \langle P_1 x, y\rangle. \tag{1.10}$$

Applying the Schwarz inequality to (1.9) we get $\langle x, y\rangle \leq \|x\|\|y\|\|P_2\|$. So by comparing the suprema of both sides of this inequality over $B_{\mathcal{X}_0} \times B_{\mathcal{Y}_0}$, we have $s \leq \|P_2\|$. But by (1.9), $s \geq \langle x, P_2 y\rangle$ for any $\|x\| = 1$, $\|y\| = 1$. Setting $x = P_2 y/\|P_2 y\|$ we get $s \geq \|P_2 y\|$ for any $\|y\| = 1$. Thus $s \geq \|P_2\|$ and, consequently, $s = \|P_2\|$. The proof of $s = \|P_1\|$ is analogous.

The next step is to show that $\|T_1\| = s^2$, where

$$\|T_1\| = \sup\{\langle T_1 x, x\rangle : x \in B_{\mathcal{X}_0}\}.$$

Indeed, for any $x \in \mathcal{X}_0$, $\langle T_1 x, x\rangle = \|P_1\|^2$; comparing the suprema of both sides of this equality we get $\|T_1\| = s^2$. The equality $\|T_2\| = s^2$ is proved in a similar way.

Suppose now that the supremum

$$s = \sup\{\langle x, y\rangle : x \in B_{\mathcal{X}_0},\ y \in B_{\mathcal{Y}_0}\}$$

is attained on some vectors x_0 and y_0.

([1]) If p_0 and q_0 are not finite then the elements of $\mathbf{R}^{p_0}$ and $\mathbf{R}^{q_0}$ are square summable sequences of real numbers.

By the Schwarz inequality applied to (1.9) and (1.10), $P_1x_0 = sy_0$ and $P_2y_0 = sx_0$, which implies $T_1x_0 = s^2x_0$, $T_2y_0 = s^2y_0$, i.e. x_0 and y_0 are eigenvectors of T_1 and T_2, respectively.

Remark 1. Both T_1 and T_2 are compact if either p_0 or q_0 is finite. This follows from the compactness of the superposition of two continuous operators of which at least one is compact (and a finite-dimensional operator is compact).

If p_0 and q_0 are finite then x_0, y_0 and $s(\mathcal{X},\mathcal{Y})$ can be given explicitly. Let $(x_i, i\in I)$ and $(y_j, j\in J)$ be any bases of $\mathcal{X}_0$ and $\mathcal{Y}_0$ respectively. Let

(1.11) $$\Sigma_{11} := [\langle x_i, x_j\rangle], \qquad i,j = 1,\ldots,p_0,$$

(1.12) $$\Sigma_{22} := [\langle y_i, y_j\rangle], \qquad i,j = 1,\ldots,q_0,$$

(1.13) $$\Sigma_{12} := [\langle x_i, y_j\rangle], \qquad i = 1,\ldots,p_0, \quad j = 1,\ldots,q_0,$$

(1.14) $$\Sigma := \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T,$$

(1.15) $$\widetilde{\Sigma} := \Sigma_{22}^{-1}\Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}.$$

The matrices of P_1 and P_2 are equal to $\Sigma_{22}^{-1}\Sigma_{12}^T$ and $\Sigma_{11}^{-1}\Sigma_{12}$, and the matrices of T_1 and T_2 are given by (1.14) and (1.15), respectively, while $s(\mathcal{X},\mathcal{Y})$ is the square root of the largest eigenvalue of (1.14) and thus of (1.15). If $\overline{x}_0$ and $\overline{y}_0$ are vectors in the respective coordinate spaces corresponding to the vectors x_0 and y_0, then to find $\overline{x}_0$ and $\overline{y}_0$ it is necessary to solve the system of equations:

$$(\Sigma - s^2I_{p_0})\overline{x}_0 = \overline{0}, \quad \overline{x}_0^T\Sigma_{11}\overline{x}_0 = 1,$$

$$(\widetilde{\Sigma} - s^2I_{q_0})\overline{y}_0 = \overline{0}, \quad \overline{y}_0^T\Sigma_{22}\overline{y}_0 = 1,$$

where I_k is the $k\times k$ identity matrix.

Let $\mathcal{K}(H)$ be the set of all pairs of closed subspaces $\mathcal{X}$ and $\mathcal{Y}$ such that the corresponding operators T_1 and T_2 are compact.

DEFINITION 1. The angle between the unit eigenvectors x_0 and y_0 of T_1, T_2 corresponding to the common largest eigenvalue $s(\mathcal{X},\mathcal{Y})$ (cf. (1.6) and (1.7)) will be called the *angle between the subspaces* $\mathcal{X}$ and $\mathcal{Y}$ and denoted by $\angle(\mathcal{X},\mathcal{Y})$. Moreover, we shall treat $\arccos(s(\mathcal{X},\mathcal{Y}))$ as the measure of this angle, denoted by $m(\angle(\mathcal{X},\mathcal{Y}))$.

This definition will be generalized so that it can be applied in the case when $\mathcal{Z} = \mathcal{X}\cap\mathcal{Y}$ is equal to $\mathcal{X}$ or $\mathcal{Y}$. Then $\angle(\mathcal{X},\mathcal{Y})$ is defined as the angle between any two unit vectors $z_1, z_2 \in \mathcal{Z}$, and we put $m(\angle(\mathcal{X},\mathcal{Y})) = 0$.

We have

$$\cos(m(\angle(\mathcal{X},\mathcal{Y}))) = \langle x_0, y_0\rangle.$$

Thus, $0 \le m(\angle(\mathcal{X},\mathcal{Y})) \le \pi/2$.

Orthogonality of $\mathcal{X}$ and $\mathcal{Y}$ implies $m(\angle(\mathcal{X},\mathcal{Y})) = \pi/2$, but not conversely. It can easily be seen that $\mathcal{X}$ and $\mathcal{Y}$ are orthogonal iff $m(\angle(\mathcal{X},\mathcal{Y})) = \pi/2$ and

$\mathcal{X} \cap \mathcal{Y} = \{0\}$.

Remark 2. A simpler procedure of deriving the cosine of the angle between subspaces is the following. We choose orthonormal bases $(x'_i, i \in I')$ and $(y'_j, j \in J')$ of $\mathcal{X}$ and $\mathcal{Y}$, respectively, where $\operatorname{card} I' = p$, $\operatorname{card} J' = q$ and I', J' are some sets of indices.

Analogously to (1.1)–(1.4), we define $P'_1 : \mathcal{X} \to \mathcal{Y}$, $P'_2 : \mathcal{Y} \to \mathcal{X}$, T'_1 and T'_2; if T'_1 and T'_2 are both compact then we deal with the eigenvalue a^2 which is maximal among the eigenvalues smaller than 1. Then $\cos(m(\angle(\mathcal{X}, \mathcal{Y}))) = a$.

If p and q are finite then we choose any bases of $\mathcal{X}$ and $\mathcal{Y}$ and form matrices Σ'_{11}, Σ'_{22}, Σ'_{12}, Σ', $\widetilde{\Sigma}'$ analogous to the matrices given by (1.11)–(1.15). Then $\cos(m(\angle(\mathcal{X}, \mathcal{Y}))) = a$ where a^2 is maximal among the eigenvalues of Σ' or $\widetilde{\Sigma}'$ smaller than 1.

The above definitions of the angle between subspaces and of its measures generalize the following elementary cases:

(i) $p = q = 1$,
(ii) $p = 1$, $q = 2$ or $p = 2$, $q = 1$,
(iii) $p = q = 2$.

In all these cases $\mathcal{X}$ and $\mathcal{Y}$ are subspaces of a three-dimensional Hilbert space.

Wysocki [8] defined similar notions of angle between linear subspaces and its volume measure.

Generally, the angle introduced in Definition 1 is different from that proposed in Wysocki [8] but the two definitions are identical if H is a three-dimensional Hilbert space. The volume measure is given by

$$\arcsin\Big(\Big(\prod_{\sigma^2 \neq 1} (1 - \sigma^2)\Big)^{1/2}\Big),$$

where the product is over all eigenvalues of Σ' or $\widetilde{\Sigma}'$ which are different from 1.

2. Maximal correlation. Let $L^2 := L(\Omega, \mathcal{A}, P; \mathbf{R}^k)$ be the vector space of all k-dimensional centered random vectors $X = (X_1, \ldots, X_k)^T$ defined on $(\Omega, \mathcal{A}, P)$ with finite covariance matrix. The scalar product $\langle\cdot,\cdot\rangle_\Lambda$ is given by

$$\langle X, Y\rangle_\Lambda := E(X^T \Lambda^{-1} Y) = \operatorname{tr}(\Lambda^{-1}\Sigma_{12}) \tag{2.1}$$

where Σ_{12} is the covariance matrix between X and Y, and Λ is any given positive definite symmetric $k \times k$ matrix. The norm derived from (2.1), which will be denoted by $\|\cdot\|_\Lambda$, is complete; thus $(L^2(\Omega, \mathcal{A}, P; \mathbf{R}^k), \langle\cdot,\cdot\rangle_\Lambda)$ is a Hilbert space. It can be shown that it is separable.

The dimension of $\mathcal{H}_X$ is not greater than k^2, and it is equal to k^2 iff X is a nondegenerate random vector (cf. Introduction for suitable definitions).

Let I_{ij} be a $k \times k$ matrix with (i,j) entry equal to 1 and all remaining entries zero. The sets

$$\{I_{ij}X : i,j = 1,\ldots,k\}\,, \tag{2.2}$$

$$\{I_{ij}Y : i,j = 1,\ldots,k\} \tag{2.3}$$

are bases of $\mathcal{H}_X$ and $\mathcal{H}_Y$, respectively. Thus for any $k\times k$ matrix $A = (a_{ij})$, $AX = \sum_{i,j} a_{ij}I_{ij}X$. Analogously $BY = \sum_{i,j} b_{ij}I_{ij}Y$.

Clearly, the spaces $\mathcal{H}_X, \mathcal{H}_Y, \mathcal{L}^2_X; \mathcal{L}^2_Y$ are closed in L^2 and the dimensions of $\mathcal{L}^2_X, \mathcal{L}^2_Y$ are at most countable.

LEMMA 2. (a) *For any* $X \in L^2$, $E(\cdot \mid X) : L^2 \to \mathcal{L}^2_X$ *is an orthogonal projector.*

(b) *For any* $X, Y \in L^2$, X *nondegenerate, a transformation* $P_2' : \mathcal{H}_Y \to \mathcal{H}_X$ *is an orthogonal projector iff*

$$P_2'(BY) = B\Sigma_{11}^T \Sigma_{11}^{-1} X\,. \tag{2.4}$$

(c) *If the bases of* $\mathcal{H}_X$ *and* $\mathcal{H}_Y$ *are given by* (2.2), (2.3), *respectively, then the matrix of the transformation* (2.4) *is*

$$I_k \times (\Sigma_{11}^{-1}\Sigma_{12})\,,$$

where $\times$ *stands for Kronecker matrix multiplication.*

Proof. (a) results from the theorem on orthogonal projection and from the following equality stated in Sampson [7]:

$$\inf\{\|Y - f(X)\|^2_\Lambda : f(X) \in \mathcal{L}^2_X\} = \|Y\|^2_\Lambda - \|E(Y \mid X)\|^2_\Lambda\,.$$

The infimum is attained for f_0 given by $f_0(x) = E(Y \mid X = x)$.

(b) Sufficiency of (2.4) is obvious.

Now let $P_2'(BY) = A_0X$ for some $k\times k$ matrix A_0. For any $k\times k$ matrix A, we have

$$\begin{aligned}\langle BY - A_0X, AX\rangle &= \operatorname{tr}(\Lambda^{-1}\operatorname{cov}(BY - A_0X, AX))\\ &= \operatorname{tr}(\Lambda^{-1}(B\Sigma_{12}^T - A_0\Sigma_{11})A^T)\\ &= \operatorname{tr}((B\Sigma_{12}^T - A_0\Sigma_{11})A^T\Lambda^{-1}) = 0\,.\end{aligned}$$

The only solution of the above is

$$A_0 = B\Sigma_{12}^T\Sigma_{11}^{-1}\,.$$

(c) To derive the matrix of the projector (2.4) in the bases (2.2) and (2.3) we use Remark 2 and (1.11), (1.13). The matrix of P_2' is equal to $\check{\Sigma}_{11}^{-1}\check{\Sigma}_{12}$, where

$$\begin{aligned}\check{\Sigma}_{12} &= [\langle I_{ij}X, I_{i'j'}Y\rangle_\Lambda]\,,\\ \check{\Sigma}_{11} &= [\langle I_{ij}X, I_{i'j'}X\rangle_\Lambda]\,, \quad i,i',j,j' = 1,\ldots,k\,.\end{aligned}$$

Since $\langle I_{ij}X, I_{i'j'}Y\rangle_\Lambda = \mathrm{tr}(\Lambda^{-1}I_{ij}\Sigma_{12}I^T_{i'j'}) = \overline{\lambda}_{i'i}\,\mathrm{cov}(X_j, Y_{j'})$, where $\Lambda^{-1} = [\overline{\lambda}_{ij}]$, we have $\check{\Sigma}_{12} = \Lambda^{-1} \times \Sigma_{12}$. Similarly $\check{\Sigma}_{11} = \Lambda^{-1} \times \Sigma_{11}$, and finally $\check{\Sigma}_{11}^{-1}\check{\Sigma}_{12} = I_k \times (\Sigma_{11}^{-1}\Sigma_{12})$, where I_k is the $k \times k$ identity matrix. Notice that P_2' restricted to the subspace $\mathcal{L}_Y$ is given by

$$P_2'(bY) = b\Sigma_{12}^T\Sigma_{11}^{-1}X .$$

Let $X, Y \in L^2$ be nondegenerate and let Σ_{11}, Σ_{22} and Σ_{12} denote the covariance matrices of X, Y and between X and Y.

Let

$$\varrho(X,Y) := \frac{\langle X,Y\rangle_\Lambda}{\|X\|_\Lambda\|Y\|_\Lambda} = \frac{\mathrm{tr}(\Lambda^{-1}\Sigma_{12})}{(\mathrm{tr}(\Lambda^{-1}\Sigma_{11}))^{1/2}(\mathrm{tr}(\Lambda^{-1}\Sigma_{22}))^{1/2}} .$$

In view of the preceding considerations $\varrho_\Lambda(X,Y)$ can be interpreted as the cosine of the angle between X and Y in $(L^2, \langle\cdot,\cdot\rangle_\Lambda)$. According to this interpretation, we treat $\langle X,Y\rangle_\Lambda$ as the covariance of X and Y, and $\|X\|_\Lambda$, $\|Y\|_\Lambda$ as their respective dispersions. Sampson [7] introduced ϱ_Λ as a measure of dependence for a pair of random vectors.

Now, we will maximize ϱ_Λ on suitably chosen sets of random vectors in order to obtain different measures of dependence. We will consider the following measures of dependence:

$$\varrho_\Lambda^{(i)}(X,Y) = \sup\{\varrho_\Lambda(U,V) : U \in \mathcal{X}_i, V \in \mathcal{Y}_i\}, \quad i = 1,\ldots,5,$$

where

$$\mathcal{X}_1 = \mathcal{H}_X, \quad \mathcal{Y}_1 = \mathcal{L}_Y, \tag{2.5}$$

$$\mathcal{X}_2 = \mathcal{L}_X^2, \quad \mathcal{Y}_2 = \mathcal{L}_Y, \tag{2.6}$$

$$\mathcal{X}_3 = \mathcal{H}_X, \quad \mathcal{Y}_3 = \mathcal{H}_Y,$$

$$\mathcal{X}_4 = \mathcal{L}_X^2, \quad \mathcal{Y}_4 = \mathcal{H}_Y,$$

$$\mathcal{X}_5 = \mathcal{L}_X^2, \quad \mathcal{Y}_5 = \mathcal{L}_Y^2. \tag{2.7}$$

Note that in view of Remark 1 (Sec. 1) the assumptions of Theorem 1 are fulfilled for the first four cases. The fifth case requires a separate treatment and will be dealt with in Theorem 2.

In the sequel we consecutively discuss $\varrho_\Lambda^{(i)}$ for $i = 1,\ldots,5$.

1) *Maximization of* ϱ_Λ *on* $\mathcal{H}_X \times \mathcal{L}_Y$. Let $X, Y \in L^2$ and let P_1', P_2' and T_1', T_2' be the orthogonal projections and the operators appearing in Remark 2 which correspond to the spaces (2.5). Thus, P_1' and P_2' are of the form

$$P_1'(AX) = (\mathrm{tr}(\Lambda^{-1}A\Sigma_{12})/\,\mathrm{tr}(\Lambda^{-1}\Sigma_{22}))Y$$

for any $k \times k$ matrix A, and

$$P_2'(bY) = b\Sigma_{12}^T\Sigma_{11}^{-1}X ,$$

where b is any real number, and $\Sigma_{12}^T\Sigma_{11}^{-1}$ is the matrix of P_2'. This is due to the representation of Y which involves linear regression of Y on X, namely

$$Y = \Sigma_{12}^T\Sigma_{11}^{-1}X + U$$

where U is a random vector uncorrelated with X.

Consequently,

$$\langle Y - \Sigma_{12}^T\Sigma_{11}^{-1}X, AX\rangle = 0$$

for any $k \times k$ matrix A. Turning to T_1' and T_2' we have

$$T_1'(AX) = \frac{\operatorname{tr}(\Lambda^{-1}A\Sigma_{12})}{\operatorname{tr}(\Lambda^{-1}\Sigma_{22})}\Sigma_{12}^T\Sigma_{11}^{-1}X,$$

$$T_2'(bY) = b\frac{\operatorname{tr}(\Lambda^{-1}\Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})}{\operatorname{tr}(\Lambda^{-1}\Sigma_{22})}Y.$$

There exist X_0 and Y_0 such that

$$\begin{aligned}(2.8)\qquad \varrho_\Lambda^{(1)}(X,Y) &= \left(\frac{\operatorname{tr}(\Lambda^{-1}\Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})}{\operatorname{tr}(\Lambda^{-1}\Sigma_{22})}\right)^{1/2} \\ &= \sup\{\varrho_\Lambda(AX, bY) : AX \in \mathcal{H}_X,\ bY \in \mathcal{L}_Y\} \\ &= \varrho_\Lambda(X_0, Y_0).\end{aligned}$$

The quantity $(\varrho_\Lambda^{(1)}(X,Y))^2$ is the unique eigenvalue of T_2' and the maximal eigenvalue of T_1'. For any given Λ, $\varrho_\Lambda^{(1)}(X,Y)$ is a measure of dependence between X and Y which generalizes multiple correlation [4]. If $\varrho_\Lambda^{(1)}(X,Y)$ is smaller than 1 then it is the cosine of the angle between $\mathcal{H}_X$ and $\mathcal{L}_Y$.

Let us now discuss the relations between $(\varrho_\Lambda^{(1)}(X,Y),\ \Lambda \in \mathcal{M}_k)$ and the measures proposed by Höschel [1] and by Jupp and Mardia [3]. It is convenient to concentrate first on nondegenerate X and Y. Höschel formulated a set of axioms to be fulfilled by a measure of linear stochastic dependence; then he constructed a suitable measure and proved its uniqueness. For nondegenerate X and Y Höschel's measure is given by

$$(2.9)\qquad \varrho^{(\mathrm{H})}(X,Y) = \frac{\operatorname{tr}(\Sigma_{22}^{-1/2}\Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1/2})}{\operatorname{rank}\Sigma_{22}}.$$

The matrix appearing in the numerator of the right-hand side of (2.9) is equivalent to $\Sigma_{22}^{-1}\Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}$; hence

$$\varrho^{(\mathrm{H})}(X,Y) = \frac{\operatorname{tr}(\Sigma_{22}^{-1}\Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12})}{\operatorname{rank}\Sigma_{22}}.$$

It follows that

$$\varrho^{(\mathrm{H})}(X,Y) = (\varrho_{\mathrm{cov}(Y)}^{(1)}(X,Y))^2.$$

Jupp and Mardia defined their measure as follows:

$$\varrho^{(\text{J-M})}(X,Y) = (\operatorname{tr}(\Sigma_{22}^{-1}\Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}))^{1/2}; \tag{2.10}$$

however, they did not provide any justification for this proposal. Comparing (2.10) with the definition of $\varrho_\Lambda^{(1)}$ we see that

$$\varrho^{(\text{J-M})}(X,Y) = k\varrho^{(1)}_{\operatorname{cov}(Y)}(X,Y).$$

It follows that

$$\varrho^{(\text{H})}(X,Y) = (\varrho^{(\text{J-M})}(X,Y))^2/k^2.$$

If X or Y are degenerate, we obtain similar results replacing (if necessary) the inverse matrices in (2.8), (2.9) and (2.10) by Moore–Penrose inverses.

2) *Maximization of* $\varrho_\Lambda^{(2)}$ *on* $\mathcal{L}_X^2 \times \mathcal{L}_Y$. Now P_1', P_2' and T_1', T_2' will be the orthogonal projections and the operators appearing in Remark 2 which correspond to the spaces (2.6). Thus

$$\begin{aligned}
P_1'(f(X)) &= \frac{\operatorname{tr}(\Lambda^{-1}\operatorname{cov}(f(X),Y))}{\operatorname{tr}(\Lambda^{-1}\Sigma_{22})}Y,\\
P_2'(bY) &= bE(Y \mid X),\\
T_1'(f(X)) &= \frac{\operatorname{tr}(\Lambda^{-1}\operatorname{cov}(f(X),Y))}{\operatorname{tr}(\Lambda^{-1}\Sigma_{22})}E(Y \mid X),\\
T_2'(bY) &= b\frac{\operatorname{tr}(\Lambda^{-1}\operatorname{cov}(E(Y\mid X),Y))}{\operatorname{tr}(\Lambda^{-1}\Sigma_{22})}Y\\
&= b\frac{\operatorname{tr}(\Lambda^{-1}\operatorname{cov}(E(Y\mid X)))}{\operatorname{tr}(\Lambda^{-1}\Sigma_{22})}Y = b\varrho_\Lambda^{(2)}(X,Y)Y,
\end{aligned}$$

where

$$\varrho_\Lambda^{(2)}(X,Y) = \left(\frac{\operatorname{tr}(\Lambda^{-1}\operatorname{cov}(E(Y\mid X)))}{\operatorname{tr}(\Lambda^{-1}\Sigma_{22})}\right)^{1/2}.$$

It follows from the form of T_2' that $(\varrho_\Lambda^{(2)}(X,Y))^2$ is its unique eigenvalue. It is also the maximal eigenvalue of T_1'. Let X_0 and Y_0 be eigenvectors of T_1' and T_2' which correspond to this eigenvalue. Since

$$\begin{aligned}
&\operatorname{cov}(E(Y\mid X),Y) = \operatorname{cov}(E(Y\mid X),Y^T) = E(E(Y\mid X)Y^T)\\
&\quad = E(E((Y\mid X)Y^T)\mid X) = E(E(Y\mid X)(E(Y\mid X))^T) = \operatorname{cov}(E(Y\mid X))
\end{aligned}$$

we have

$$\varrho_\Lambda^{(2)}(X,Y) = \sup\{\varrho_\Lambda(f(X),Y) : f(X)\in\mathcal{L}_X^2,\ bY\in\mathcal{L}_Y\} = \varrho_\Lambda(X_0,Y_0).$$

The measures $\varrho_\Lambda^{(2)}(X,Y)$ were introduced by Sampson [7]. They generalize the correlation ratio between a random variable X and a random vector

Y. If $\varrho_A^{(2)}(X,Y)$ is smaller than 1 then it is the cosine of the angle between $\mathcal{L}_X^2$ and $\mathcal{L}_Y^2$.

3) *Maximization of* ϱ_A *on* $\mathcal{H}_X \times \mathcal{H}_Y$. The orthogonal projectors P_1', P_2', and the operators T_1', T_2' (defined in Remark 2) for $\mathcal{H} = \mathcal{H}_X$, $\mathcal{Y} = \mathcal{H}_Y$ are of the following form:

$$P'(AX) = A\Sigma_{12}\Sigma_{22}^{-1}Y\,, \quad P_2'(BY) = B\Sigma_{12}^T\Sigma_{11}^{-1}X\,,$$
$$T_1'(AX) = A\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T\Sigma_{11}^{-1}X\,, \quad T_2'(BY) = B\Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}Y\,.$$

By Lemma 2(c), the matrices of T_1' and T_2' in the bases of (2.6), (2.7) are

$$I_k \times (\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T\Sigma_{11}^{-1})\,, \tag{2.11}$$
$$I_k \times (\Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1})\,, \tag{2.12}$$

respectively.

Notice that (2.11) is a block matrix with k matrices equal to $\Sigma_{12}\Sigma_{22}^{-1} \times\Sigma_{12}^T\Sigma_{11}^{-1}$ on the diagonal and with zero off-diagonal blocks. Therefore the eigenvalues of (2.11) are equal to those of $\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T\Sigma_{11}^{-1}$. Analogously the eigenvalues of (2.12) are those of $\Sigma_{12}^T\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}$.

Let a^2 denote the maximal eigenvalue of $\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T\Sigma_{11}^{-1}$. We have

$$\varrho^{(3)}(X,Y) = a = \sup\{\varrho_A(AX,BY) : AX \in \mathcal{H}_X, BY \in \mathcal{H}_Y\} = \varrho_A(X_0,Y_0)\,.$$

The random vectors X_0, Y_0 are eigenvectors of T_1', T_2', respectively, for the eigenvalue a^2.

Johnson and Verhly [2] introduced $\varrho^{(3)}(X,Y)$ and called it the *maximal canonical correlation.*

If $\varrho^{(3)}(X,Y)$ is smaller than 1 then it is the cosine of the angle between $\mathcal{H}_X$ and $\mathcal{H}_Y$.

For degenerate X and Y we use the Moore–Penrose inverse matrices in (2.11) and (2.12) if necessary.

4) *Maximization of* ϱ_A *on* $\mathcal{L}_X^2 \times \mathcal{H}_Y$. For $\mathcal{X} = \mathcal{L}_X^2$, $\mathcal{Y} = \mathcal{H}_Y$, the operators T_1', T_2' are of the following form:

$$T_1'(f(X)) = \operatorname{cov}(f(X),Y)\Sigma_{22}^{-1}E(Y \mid X)\,,$$
$$T_2'(BY) = B\operatorname{cov}(E(Y \mid X))\Sigma_{22}^{-1}Y\,.$$

The matrix of the operator T_2' in the base (1.5) is

$$I_k \times (\operatorname{cov}(E(Y \mid X))\Sigma_{22}^{-1})\,. \tag{2.13}$$

The maximal eigenvalue of the matrix (2.26) is equal to that of $\operatorname{cov}(E(Y \mid X))\Sigma_{22}^{-1}$, it is also the maximal eigenvalue of the operators T_1' and T_2'. Let a^2 denote this maximal eigenvalue and let X_0, Y_0 denote the corresponding eigenvectors of the operators T_1' and T_2', respectively.

Hence we have

$$\varrho^{(4)}(X,Y)=a=\sup\{\varrho_\Lambda(f(X),BY):f(X)\in L_X^2, BY\in\mathcal{H}_Y\}=\varrho_\Lambda(X_0,Y_0).$$

If $\varrho^{(4)}(X,Y)$ is less than 1 then it is the cosine of the angle between $\mathcal{L}_X^2$ and $\mathcal{H}_Y$.

In Sampson [7], ϱ_Λ has been maximized on the set $\mathcal{L}_X^2\times\widetilde{\mathcal{H}}_Y$, where $\widetilde{\mathcal{H}}_Y$ is the set of random vectors BY such that $\mathrm{cov}(BY)=\Lambda$. By this procedure he has got

$$\begin{aligned}\widetilde{\varrho}^{(4)}(X,Y)&=\sup\{\varrho_\Lambda(f(X),BY):f(X)\in\mathcal{L}_X^2,\ BY\in\widetilde{\mathcal{H}}_Y\}\\&=\left(\frac{1}{k}\,\mathrm{tr}(\mathrm{cov}(Y\mid X)\Sigma_{22}^{-1})\right)^{1/2}=\varrho_\Lambda(X_0',Y_0')\end{aligned}$$

for suitably chosen random vectors X_0', Y_0'.

5) *Maximization of ϱ_Λ on $\mathcal{L}_X^2\times\mathcal{L}_Y^2$.* Dealing now with the most general case, we will give a simple condition under which maximal correlation exists.

THEOREM 2. *Suppose X and Y have finite respective covariance matrices. If $P_{(X,Y)}$ is absolutely continuous with respect to $P_X\otimes P_Y$, and if the squared density p of $P_{(X,Y)}$ with respect to $P_X\otimes P_Y$ is integrable then there exist random vectors $f_0(X)$ and $g_0(Y)$ such that*

$$\begin{aligned}(2.14)\quad \varrho^{(5)}(X,Y)=a&=\sup\{\varrho_\Lambda(f(X),g(Y)):f(X)\in\mathcal{L}_X^2,\ g(Y)\in\mathcal{L}_Y^2\}\\&=\varrho_\Lambda(f_0(X),g_0(Y)),\end{aligned}$$

where a^2 is the maximal eigenvalue of the operators T_1', T_2' which correspond to the spaces $\mathcal{L}_X^2$, $\mathcal{L}_Y^2$.

The proof is an obvious modification of that given by Rényi [5] in the univariate case. Let P_1', P_2' and T_1', T_2' be the orthogonal projections and operators appearing in Remark 2, related to the spaces (2.7). Hence we have

$$\begin{aligned}P_1'(f(X))&=E(f(X)\mid Y), & P_2'(g(Y))&=E(g(Y)\mid X),\\T_1'(f(X))&=E(E(f(X)\mid Y)\mid X), & T_2'(g(Y))&=E(E(g(Y)\mid X)\mid Y).\end{aligned}$$

In view of Theorem 1 it is sufficient to show that T_1' and T_2' are compact. We will prove it for T_1' since the proof for T_2' is analogous.

Let us introduce the following Hilbert space:

$$\widetilde{\mathcal{L}}_X^2=\{f:f(X)\in\mathcal{L}_X^2\},$$

with the scalar product

$$\int_{\mathbf{R}^k}(f_1(x))^T\Lambda^{-1}f_2(x)\,dP_X(x).$$

It is obvious that $\mathcal{L}_X^2$ and $\widetilde{\mathcal{L}}_X^2$ are isomorphic. Thus, T_1' can be used in $\widetilde{\mathcal{L}}_X^2$ and it has the following form ([6]):

$$T_1'(f(X)) = \int_{\mathbf{R}^k} f(u)\Big(\int_{\mathbf{R}^k} p(u,v)p(x,v)\,dP_Y(v)\Big)\,dP_X(u)\,.$$

It follows that T_1' is an integral operator with kernel

$$\int_{\mathbf{R}^k} p(u,v)p(x,v)\,dP_Y(v)\,.$$

We will show that the squared kernel is integrable, which, as is known, implies the compactness of the integral operator. By the Schwarz inequality,

$$\Big(\int_{\mathbf{R}^k} p(u,v)p(x,v)\,dP_Y(v)\Big)^2 \le \Big(\int_{\mathbf{R}^k} p^2(u,v)\,dP_Y(v)\Big)\Big(\int_{\mathbf{R}^k} p^2(x,v)\,dP_Y(v)\Big)\,.$$

Consequently,

$$\int_{\mathbf{R}^k}\int_{\mathbf{R}^k}\Big(\int_{\mathbf{R}^k} p(u,v)p(x,v)\,dP_Y(v)\Big)^2\,dP_X(u)\,dP_X(x)$$
$$\le \Big(\int_{\mathbf{R}^k}\int_{\mathbf{R}^k} p^2(u,v)\,dP_Y(v)\,dP_X(u)\Big)^2 < \infty\,.$$

It follows by Theorem 1 that there exist functions f_0 and g_0 such that

$$T_1'(f_0) = a^2 f_0\,, \quad T_2'(g_0) = a^2 g_0\,,$$

where a^2 is the maximal eigenvalue of T_1' and T_2'. Moreover, f_0 and g_0 satisfy (2.14).

If $\varrho^{(5)}(X,Y)$ is smaller than 1 then it is equal to the cosine of the angle between $\mathcal{L}_X^2$ and $\mathcal{L}_Y^2$.

References

[1] H. P. Höschel, *Ein axiomatischer Zugang zu Abhängigkeitsmaßen für Zufallsvektoren und die Spurkorrelation*, Math. Operationsforsch. Statist. 7 (1976), 789–802.

[2] R. A. Johnson and T. Verhly, *Measures and models for angular correlation and angular-linear correlation*, J. Roy. Statist. Soc. B 39 (1977), 222–229.

[3] P. E. Jupp and K. V. Mardia, *A general correlation coefficient for directional data and related regression problems*, Biometrika 67 (1980), 163–173.

[4] C. R. Rao, *Linear Statistical Inference and its Applications*, Wiley, New York 1973.

[5] A. Rényi, *On measures of dependence*, Acta Math. Acad. Sci. Hungar. 10 (1959), 441–451.

[6] F. Riesz et B. Nagy, *Leçons d'analyse fonctionnelle*, Akadémiai Kiadó, Budapest 1952.

[7] A. R. Sampson, *Positive dependence properties of elliptically symmetric distributions*. J. Multivariate Anal. 13 (1983), 375–381.
[8] W. Wysocki, *A geometric approach to measuring dependence between random vectors*, Zastos. Mat. 20 (1990), 191–202.

WLODZIMIERZ WYSOCKI
INSTITUTE OF COMPUTER SCIENCE
POLISH ACADEMY OF SCIENCES
P.O. BOX 22
00-901 WARSZAWA, POLAND

Received on 24.2.1988;
revised version on 19.2.1990