

# CREATING AND CLASSIFYING MEASURES OF LINEAR ASSOCIATION BY OPTIMIZATION TECHNIQUES

WIEBE R. PESTMAN

## Abstract

The idea of measures of linear association, such as Pearson's correlation coefficient, can be put in a general framework by axiomization. Groups of linear transformations on  $\mathbb{R}^n$  can be exploited to create new and classify existing measures according to their invariance properties. Thus invariance under the Euclidean transformation group leads to the class of so-called geometric measures. Similarly, a measure is called algebraic if it is invariant under scalings. Pearson's coefficient is an example of an algebraic measure; it is not geometric. It is proved that, generally, a measure of linear association cannot possibly be both geometric and algebraic. A procedure is developed to convert a geometric measure into an algebraic and vice versa. Thus a kind of a duality between algebraic and geometric measures arises. In this duality measures can be reflexive or not.

## 1. Introduction

When talking about measures of linear association, the first thing that comes up in mind is usually Pearson's correlation coefficient (see for example [10]), hereafter denoted as  $\rho_P$ . This measure (or better its square) can be useful to evaluate the fraction of explained variance in an ordinary linear regression. Besides this, however, Pearson's coefficient is often used as a kind of a geometrical quantity that would measure the degree to which the points in a scatterplot cluster around a straight line. This is really a risky usage, for the coefficient changes its values when applying simple Euclidean transformations. For example, the value of  $\rho_P$  changes when applying a rotation, say, around the origin, on the set of points in a scatter plot (see [5], [7], [8], [11], [12]). In mathematics this is usually referred to by saying that  $\rho_P$  is not invariant under (the group of) rotations. In the following the invariance properties of Pearson's coefficient will be discussed in some detail. Besides Pearson's coefficient, however, some other measures of linear association will be discussed. It will appear to be comfortable to capture one thing and another in a small framework. This framework will allow for a deliberative view on the position of Pearson's coefficient relative to various other measures of linear association.

## 2. Setting the notations

Characters like  $\mathbf{x}$ ,  $\mathbf{y}$ ,  $\mathbf{z}$  will stand for vectors in  $\mathbb{R}^n$ , hence one has for example

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

The vector with all components equal to 1 will be denoted as  $\mathbf{1}$ . The mean of the components  $x_1, x_2, \dots, x_n$  of  $\mathbf{x}$  will be denoted by  $\bar{x}$ . The covariance of two vectors  $\mathbf{x}$  and  $\mathbf{y}$  will be denoted by  $\text{cov}(\mathbf{x}, \mathbf{y})$ , hence one has

$$\text{cov}(\mathbf{x}, \mathbf{y}) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

The variance of a vector  $\mathbf{x}$  is by definition the quantity  $\text{cov}(\mathbf{x}, \mathbf{x})$  and will be denoted by  $\text{var}(\mathbf{x})$ .

Given any set of vectors  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$  in  $\mathbb{R}^n$ , a linear combination

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \dots + a_m\mathbf{x}_m$$

will be called *spherical* if the coefficient vector  $\mathbf{a} = (a_1, a_2, \dots, a_m)$  is of unit length in  $\mathbb{R}^m$ , that is, if

$$\|\mathbf{a}\|^2 = a_1^2 + a_2^2 + \dots + a_m^2 = 1.$$

As is wellknown (see for example [10]), maximization and minimization of the expression

$$(1) \quad \text{var}(a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \dots + a_m\mathbf{x}_m)$$

over all coefficient vectors  $\mathbf{a}$  of unit length can be carried out through diagonalization of the covariance matrix  $\mathbf{C}$  of the vectors  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$ , that is, the matrix

$$\mathbf{C} = \begin{pmatrix} \text{cov}(\mathbf{x}_1, \mathbf{x}_1) & \text{cov}(\mathbf{x}_1, \mathbf{x}_2) & \dots & \text{cov}(\mathbf{x}_1, \mathbf{x}_m) \\ \text{cov}(\mathbf{x}_2, \mathbf{x}_1) & \text{cov}(\mathbf{x}_2, \mathbf{x}_2) & \dots & \text{cov}(\mathbf{x}_2, \mathbf{x}_m) \\ \vdots & \vdots & \dots & \vdots \\ \text{cov}(\mathbf{x}_m, \mathbf{x}_1) & \text{cov}(\mathbf{x}_m, \mathbf{x}_2) & \dots & \text{cov}(\mathbf{x}_m, \mathbf{x}_m) \end{pmatrix}.$$

Thus the maximum and minimum values of (1) are equal to the largest and smallest eigenvalue of  $\mathbf{C}$ . The coefficient vectors that bring this about are

eigenvectors (of unit length) of  $\mathbf{C}$ . The largest and the smallest eigenvalue of  $\mathbf{C}$  will be denoted by  $\lambda_{\max}$  and  $\lambda_{\min}$  respectively.

The character  $\mathbf{d}$ , unless otherwise stated, will stand for an  $n \times m$  matrix. The rows of such a matrix can be looked upon as being  $n$  points in  $\mathbb{R}^m$  and these points can be subject of (geometrical) transformations. Row number  $i$  of matrix  $\mathbf{d}$  will be denoted as  $\mathbf{d}_{i\bullet}$  and column number  $j$  as  $\mathbf{d}_{\bullet j}$ .

A matrix  $\mathbf{d}$  will be called a *datamatrix* if there is among the columns of  $\mathbf{d}$  at least one that is not constant. An  $n \times m$  datamatrix  $\mathbf{d}$  will be called *degenerate* if the vectors  $\{\mathbf{1}, \mathbf{d}_{\bullet 1}, \mathbf{d}_{\bullet 2}, \dots, \mathbf{d}_{\bullet m}\}$  form a linear dependent system in  $\mathbb{R}^n$  in the sense defined in linear algebra. A datamatrix is degenerate if and only if the  $n$  associated points are situated in an affine subspace of  $\mathbb{R}^m$  of dimension strictly less than  $m$ . A datamatrix is called *centered* if the column means are all zero.

Sometimes the last column in a datamatrix will play a particular role. If so then the datamatrix will be announced as an  $n \times (m + 1)$  matrix. In such cases the matrix consisting of the first  $m$  columns of  $\mathbf{d}$  will be denoted by  $\mathbf{x}$  and the last column of  $\mathbf{d}$  by  $\mathbf{y}$ . Hence, when writing out an  $n \times (m + 1)$  datamatrix  $\mathbf{d}$  under these conventions one would have:

$$\mathbf{d} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} & y_1 \\ x_{21} & x_{22} & \dots & x_{2m} & y_2 \\ x_{31} & x_{32} & \dots & x_{3m} & y_3 \\ \vdots & \vdots & & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} & y_n \end{pmatrix}.$$

In these notations an  $n \times (m + 1)$  datamatrix  $\mathbf{d}$  defines a set of  $n$  points

$$(\mathbf{x}_{1\bullet}, y_1), (\mathbf{x}_{2\bullet}, y_2), \dots, (\mathbf{x}_{n\bullet}, y_n)$$

in  $\mathbb{R}^{m+1}$ .

To define the Pearson correlation coefficient of a non-degenerate  $n \times (m + 1)$  datamatrix  $\mathbf{d}$ , suppose that in a linear regression the column  $\mathbf{y}$  is taken as response variable and the  $m$  columns of  $\mathbf{x}$  as explanatory variables. Pearson's correlation coefficient is then by definition the square root of the fraction of explained variance in this regression. This coefficient will be denoted by  $\rho_P(\mathbf{d})$ . Denoting the residual sum of squares in the regression by SSE one could also define  $\rho_P(\mathbf{d})$  for non-degenerate datamatrices as (see [10]):

$$\rho_P(\mathbf{d}) = \sqrt{1 - \frac{\text{SSE}}{n \text{ var}(\mathbf{y})}}$$

For degenerate datamatrices Pearson's coefficient will be set to the value 1 by definition. Note that  $\rho_P(\mathbf{d})$  is always a number in between 0 and 1. The value 1 is only attained in cases where the  $n$  points happen to lie in an affine subspace of dimension less than  $m$ . Also note that in the setting above Pearson's coefficient can never be negative. In cases where  $\mathbf{d}$  is a centered datamatrix Pearson's coefficient can be regarded (see [10]) as the cosine of the angle between  $\mathbf{y}$  and the linear space spanned by the columns of  $\mathbf{x}$ .

### 3. Transformation of datasets

Given some  $n \times m$  matrix  $\mathbf{a}$  of type

$$\mathbf{a} = \begin{pmatrix} a_1 & a_2 & \dots & a_m \\ a_1 & a_2 & \dots & a_m \\ a_1 & a_2 & \dots & a_m \\ \vdots & \vdots & & \vdots \\ a_1 & a_2 & \dots & a_m \end{pmatrix}$$

translations of  $n \times m$  matrices can be presented as a transformation of the form

$$\mathbf{d} \mapsto \mathbf{d} + \mathbf{a}.$$

In terms of matrix products a scaling can be captured as a transformation of type

$$\mathbf{d} \mapsto \mathbf{d}\mathbf{s}$$

where  $\mathbf{s}$  is a  $m \times m$  diagonal matrix with strictly positive entries on the diagonal. The set of all such matrices will be denoted by  $\mathbf{D}$ .

Similarly, orthogonal transformations can be captured as:

$$\mathbf{d} \mapsto \mathbf{d}\mathbf{q}$$

where  $\mathbf{q}$  is a  $m \times m$  orthogonal matrix. The set of all such matrices will be denoted by  $\mathbf{O}$ . Orthogonal transformations (especially those with determinant equal to 1) are often regarded as generalized rotations. Given a datamatrix  $\mathbf{d}$ , an orthogonal matrix that is special to  $\mathbf{q}$  is presented by a matrix  $\mathbf{q}$  that consists of normalized eigenvectors of the covariance matrix of  $\mathbf{d}$ . For such a particular  $\mathbf{q}$  the columns in the datamatrix  $\mathbf{d}\mathbf{q}$  are mutually uncorrelated. Hence, provided  $\mathbf{d}$  is non-degenerate, for such  $\mathbf{q}$  one always has a zero value for  $\rho_P(\mathbf{d}\mathbf{q})$ . It follows from this that Pearson's coefficient changes as a rule when applying orthogonal transformations on a datamatrix.

General linear transformations can be captured in terms of matrix products as:

$$\mathbf{d} \mapsto \mathbf{d}\mathbf{g}$$

where  $\mathbf{g}$  is an arbitrary  $m \times m$  invertible matrix.

Note that all these classes of transformations form groups in the sense of Galois. Actually, with the exception of translations, they all form matrix multiplication groups.

#### 4. General measures of linear association

In this section a small framework will be set up. This will turn out to be comfortable in subsequent sections. General measures of linear association will be defined as being maps  $\mathbf{d} \mapsto \gamma(\mathbf{d})$  that satisfy a suitable set of properties (see also [9]).

DEFINITION 4.1. A matrix function  $\gamma$  is said to be *continuous* in a datamatrix  $\mathbf{d}$  if for every sequence  $\mathbf{d}_1, \mathbf{d}_2, \dots$  of matrices (of the same size) converging to  $\mathbf{d}$  the sequence  $\gamma(\mathbf{d}_1), \gamma(\mathbf{d}_2), \dots$  converges to  $\gamma(\mathbf{d})$ .

DEFINITION 4.2. An  $m$ -dimensional *measure* of linear association is understood to be a function  $\mathbf{d} \mapsto \gamma(\mathbf{d})$ , defined on the set of all  $m$ -column datamatrices, that has the next six properties:

- a) For every  $\mathbf{d}$  one has  $0 \leq \gamma(\mathbf{d}) \leq 1$ .
- b) One has  $\gamma(\mathbf{d}) = 1$  if and only if  $\mathbf{d}$  is degenerate.
- c)  $\gamma$  is invariant under row permutations.
- d)  $\gamma$  is invariant under translations.
- e)  $\gamma$  is invariant under scalar multiplication, that is, one has  $\gamma(\mathbf{d}) = \gamma(s\mathbf{d})$  for every non-zero scalar  $s$ .
- f) If a datamatrix does not contain columns that are constant then  $\gamma$  is continuous in it.

A measure will be called *symmetric* if it is invariant under column permutations and it will be called *continuous* if it is continuous in every datamatrix.

In the 2-dimensional case Pearson's correlation coefficient is an example of a measure of linear association in the sense of the definition above. In higher dimensions, however, Pearson's coefficient fails to meet the necessary continuity requirements. Nevertheless Pearson's coefficient will prove to be useful in the construction of measures of linear association.

DEFINITION 4.3. A measure of linear association  $\gamma$  is said to be *invariant* under a matrix multiplication group  $\mathbf{G}$  if for all possible datamatrices  $\mathbf{d}$  one has

$$\gamma(\mathbf{d}) = \gamma(\mathbf{d}\mathbf{g}) \quad \text{for every } \mathbf{g} \in \mathbf{G}.$$

It is difficult to imagine that a measure that is neither invariant under the diagonal group  $\mathbf{D}$  nor under the orthogonal group  $\mathbf{O}$  can serve useful purposes.

So, starting from this way of thinking, the measures of linear association will be divided into two classes: a class where the measures are invariant under  $\mathbf{D}$  and a class where they are invariant under  $\mathbf{O}$ . Because (by definition) all measures are automatically invariant under translations,  $\mathbf{O}$ -invariant measures are actually invariant under the Euclidean transformation group. In modern mathematical terms they therefore present geometric characteristics (see [14]).

DEFINITION 4.4. If a measure of linear association is  $\mathbf{O}$ -invariant, then it will be called a *geometric* measure. If it is  $\mathbf{D}$ -invariant it will be called an *algebraic* measure.

In this terminology, the 2-dimensional Pearson's correlation coefficient is algebraic, but not geometric. A question that could arise is whether it is possible to construct a measure of linear association that is both algebraic and geometric. The answer to this question turns out to be negative. In proving this it will appear to be comfortable to have a function  $\zeta$  defined as:

DEFINITION 4.5. For every  $n \times m$  matrix the number  $\zeta(\mathbf{d})$  is defined as

$$\zeta(\mathbf{d}) = \text{var}(\mathbf{d}_{\bullet 1}) \text{var}(\mathbf{d}_{\bullet 2}) \dots \text{var}(\mathbf{d}_{\bullet m}).$$

Note that  $\zeta(\mathbf{d}) = 0$  if and only if  $\mathbf{d}$  has one or more constant columns. The following property of  $\zeta$  will be useful:

LEMMA 4.6. *If the columns of an  $n \times m$  datamatrix  $\mathbf{d}$  are not of equal variance, then there exists an orthogonal matrix  $\mathbf{q}$  such that*

$$\zeta(\mathbf{d}) < \zeta(\mathbf{d}\mathbf{q}).$$

PROOF. This lemma will first be proved for cases where  $\mathbf{d}$  is a 2-column datamatrix. Define for such matrices the function  $\mathbf{d} \mapsto f(\mathbf{d})$  as

$$f(\mathbf{d}) = \text{var}(\mathbf{d}_{\bullet 1}) - \text{var}(\mathbf{d}_{\bullet 2}).$$

Moreover define for every  $\varphi \in [0, \pi]$  the orthogonal matrix  $\mathbf{q}(\varphi)$  as

$$\mathbf{q}(\varphi) = \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix}.$$

Now the function  $\varphi \mapsto f(\mathbf{d}\mathbf{q}(\varphi))$  is continuous and assumes values of opposite sign in 0 and  $\pi/2$ . It follows that there must be a zero somewhere in between 0 and  $\pi/2$ . Let  $\varphi$  be any such zero and let  $\mathbf{q}$  be the corresponding matrix  $\mathbf{q}(\varphi)$ . Then  $\mathbf{d}\mathbf{q}$  has columns of equal variance. Set

$$v_1 = \text{var}(\mathbf{d}_{\bullet 1}) \quad \text{and} \quad v_2 = \text{var}(\mathbf{d}_{\bullet 2}).$$

Then the variances of the columns in  $\mathbf{d}\mathbf{q}$  must be equal to  $(v_1 + v_2)/2$ . If  $v_1 \neq v_2$  then one has

$$\zeta(\mathbf{d}) = v_1 v_2 < \left(\frac{v_1 + v_2}{2}\right) \left(\frac{v_1 + v_2}{2}\right) = \zeta(\mathbf{d}\mathbf{q})$$

which proves the lemma in cases where  $\mathbf{d}$  is a 2-column matrix.

Next, let  $\mathbf{d}$  be an arbitrary  $n \times m$  matrix the columns of which are not of equal variance. Without loss of generality one may assume that the first two columns have different variances. Let  $\mathbf{d}_0$  be the  $n \times 2$  matrix consisting of these two columns. By the first part of the proof there exists a  $2 \times 2$  orthogonal matrix  $\mathbf{q}_0$  such that

$$\zeta(\mathbf{d}_0) < \zeta(\mathbf{d}_0\mathbf{q}_0).$$

Define  $\mathbf{q}$  to be the  $m \times m$  identity matrix in which the left upper  $2 \times 2$  block has been replaced by  $\mathbf{q}_0$ . For this matrix  $\mathbf{q}$  one has

$$\begin{aligned} \zeta(\mathbf{d}) &= \zeta(\mathbf{d}_0) \operatorname{var}(\mathbf{d}_{\bullet 3}) \dots \operatorname{var}(\mathbf{d}_{\bullet m}) \\ &< \zeta(\mathbf{d}_0\mathbf{q}_0) \operatorname{var}(\mathbf{d}_{\bullet 3}) \dots \operatorname{var}(\mathbf{d}_{\bullet m}) = \zeta(\mathbf{d}\mathbf{q}) \end{aligned}$$

which proves the lemma for the general  $\mathbf{d}$ .

Leaning on this lemma one can prove:

**THEOREM 4.7.** *A geometric measure of linear association is always continuous. An algebraic measure of linear association is necessarily discontinuous in certain datamatrices that contain one or more constant columns.*

**PROOF.** To prove the continuity of geometric measures, let  $\gamma$  be any such measure. The only point of concern are the datamatrices that contain one or more constant columns. If  $\mathbf{d}$  is such a matrix, however, then, by the previous lemma, there exists an orthogonal matrix  $\mathbf{q}$  such that  $\mathbf{d}\mathbf{q}$  does not contain constant columns. Hence  $\gamma$  is continuous in  $\mathbf{d}\mathbf{q}$ . Because  $\gamma$  is  $\mathbf{O}$ -invariant this implies that there is also continuity in  $\mathbf{d}$ . To prove discontinuity of algebraic measures, suppose that  $\gamma$  is any such measure. Let  $\mathbf{d}_0$  be a datamatrix that has one or more constant columns. Assume that  $\mathbf{d}_0$  is such that the non-constant columns in it form a non-degenerate datamatrix. Because of the translation invariance of  $\gamma$  one may assume, without loosing generality, that the constant columns consist of zeros only. Now consider datamatrices that can be derived from  $\mathbf{d}_0$  by modifying its zero columns. Among these matrices there are non-degenerate ones. Choose any such non-degenerate  $\mathbf{d}$ . Then there exists a sequence  $\mathbf{s}_1, \mathbf{s}_2, \dots$  of matrices in the diagonal group  $\mathbf{D}$  such that

$$\lim_{n \rightarrow \infty} \mathbf{d}\mathbf{s}_n = \mathbf{d}_0.$$

However, in virtue of the D-invariance of  $\gamma$ , one has

$$\lim_{n \rightarrow \infty} \gamma(\mathbf{d}\mathbf{s}_n) = \gamma(\mathbf{d}) < 1.$$

This shows that the limit on the left side cannot possibly be equal to  $\gamma(\mathbf{d}_0)$ , which has the value 1. It follows that  $\gamma$  is discontinuous in  $\mathbf{d}_0$ .

**COROLLARY 4.8.** *A measure of linear association cannot possibly be both algebraic and geometric.*

If, for example, as to a set of points in  $\mathbb{R}^2$  there is a wish to quantify their degree of clustering around a line then it seems natural to use, rather than Pearson's coefficient, a geometric measure. To have an example of a geometric measure in mind, one could think of the measure defined by:

**DEFINITION 4.9.** Given an arbitrary  $n \times (m + 1)$  matrix  $\mathbf{d}$  its *Euclidean correlation coefficient* is given by

$$\rho_E(\mathbf{d}) = \max_{\mathbf{q} \in \mathcal{O}} \rho_P(\mathbf{d}\mathbf{q})$$

where the maximum is taken over all orthogonal matrices  $\mathbf{q}$ .

More colloquially, the Euclidean correlation coefficient is the maximum value of Pearson's correlation coefficient that can be obtained by transforming the cloud of points by means of orthogonal matrices. As will be clarified later on, the map  $\mathbf{d} \mapsto \rho_E(\mathbf{d})$  is a measure of linear association indeed. Besides this, exploiting the fact that the orthogonal matrices form a group, it is easy to see that the Euclidean correlation coefficient is geometric. As a consequence of Corollary 4.8, the Euclidean correlation coefficient cannot possibly be algebraic. For  $n \times 2$  datamatrices the maximization above has been studied in [7]. From the results there one can derive the following explicit expression for  $\rho_E(\mathbf{d})$  in the case of  $n \times 2$  datamatrices:

$$\rho_E(\mathbf{d}) = \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}$$

where  $\lambda_{\max}$  and  $\lambda_{\min}$  are the largest and smallest eigenvalue of the covariance matrix of  $\mathbf{d}$ . In a subsequent section it will be proved that the above is generally valid for arbitrary  $n \times m$  datamatrices. From this result it follows that  $\rho_E$  meets the necessary continuity requirements and that it is a proper measure of linear association by consequence.



## 5. Creating invariance through optimization

The maximizing construction, used in the previous section to define the Euclidean correlation coefficient, can easily be captured in a more general setting:

DEFINITION 5.1. For an arbitrary matrix multiplication group  $\mathbf{G}$  and an arbitrary measure of linear association  $\gamma$  the  $\mathbf{G}$ -maximization and  $\mathbf{G}$ -minimization of  $\gamma$  are understood to be the maps

$$\mathbf{d} \mapsto \sup_{\mathbf{g} \in \mathbf{G}} \gamma(\mathbf{d}\mathbf{g}) \quad \text{and} \quad \mathbf{d} \mapsto \inf_{\mathbf{g} \in \mathbf{G}} \gamma(\mathbf{d}\mathbf{g})$$

These maps will be denoted by  $\sup_{\mathbf{G}}(\gamma)$  and  $\inf_{\mathbf{G}}(\gamma)$  respectively.

It should be noted that the maps  $\sup_{\mathbf{G}}(\gamma)$  and  $\inf_{\mathbf{G}}(\gamma)$  are always  $\mathbf{G}$ -invariant, but that they do not always present measures of linear association. An example of this emerges when minimizing Pearson's coefficient over the (compact) orthogonal group  $\mathbf{O}$  to get  $\inf_{\mathbf{O}}(\rho_P)$ . This map can only assume the values 0 and 1 and can therefore impossibly meet the necessary continuity conditions. This negative phenomenon is more generally described by the next theorem:

THEOREM 5.2. *A  $\mathbf{D}$ -maximization of a continuous measure or an  $\mathbf{O}$ -minimization of an algebraic measure is never a measure of linear association.*

PROOF. To prove the first statement, suppose that  $\gamma$  is a continuous measure of linear association. Choose any non-degenerate datamatrix  $\mathbf{d}$ . Then there exists a sequence  $\mathbf{s}_1, \mathbf{s}_2, \dots$  of matrices in the diagonal group  $\mathbf{D}$  such that the sequence

$$\mathbf{d}\mathbf{s}_1, \mathbf{d}\mathbf{s}_2, \mathbf{d}\mathbf{s}_3, \dots$$

converges to a degenerate datamatrix. Continuity of  $\gamma$  implies that the sequence

$$\gamma(\mathbf{d}\mathbf{s}_1), \gamma(\mathbf{d}\mathbf{s}_2), \gamma(\mathbf{d}\mathbf{s}_3), \dots$$

converges to 1. For this reason the  $\mathbf{D}$ -maximization of  $\gamma$  systematically assumes the value 1 for non-degenerate datamatrices. It is therefore not a measure of linear association.

To prove the second statement, suppose that  $\gamma$  is an algebraic measure of linear association. Again, fix any non-degenerate datamatrix  $\mathbf{d}$ . Then one has

$$\gamma(\mathbf{d}\mathbf{s}) = \gamma(\mathbf{d}) < 1 \quad \text{for all } \mathbf{s} \in \mathbf{D}.$$

It follows from this that

$$\inf_{\mathbf{O}}(\gamma)(\mathbf{d}\mathbf{s}) \leq \gamma(\mathbf{d}\mathbf{s}) \leq \gamma(\mathbf{d}) < 1 \quad \text{for all } \mathbf{s} \in \mathbf{D}.$$

Now, if  $\inf_{\mathcal{O}}(\gamma)$  were a measure of linear association, then it would be a geometric and thus a continuous one. Using the arguments of the first part of the proof, the inequality above could not be true for all  $\mathbf{s} \in \mathcal{D}$ . This proves that the  $\mathcal{O}$ -minimization of  $\gamma$  cannot possibly be a measure of linear association.

It will turn out, in contrast to the theorem above, that  $\mathcal{D}$ -minimization of geometric measures or  $\mathcal{O}$ -maximization of algebraic measures results in proper measures of linear association. In proving this the next topological lemma will be useful:

LEMMA 5.3. *Let  $X$  and  $Y$  be compact metric spaces and let  $f : X \times Y \rightarrow \mathbb{R}$  be a continuous function. Then the maps*

$$x \mapsto \min_{y \in Y} f(x, y) \quad \text{and} \quad x \mapsto \max_{y \in Y} f(x, y)$$

*are continuous on  $X$ .*

PROOF. This can be proved by using standard arguments in mathematical analysis. See for example [2], [6], [13] for the necessary tools in this.

THEOREM 5.4. *Let  $\mathbf{G}$  be an arbitrary compact multiplicative matrix group and  $\gamma$  an arbitrary measure of linear association. Then the  $\mathbf{G}$ -maximization of  $\gamma$  is a  $\mathbf{G}$ -invariant measure of linear association. If  $\gamma$  is continuous then the  $\mathbf{G}$ -minimization of  $\gamma$  is a continuous  $\mathbf{G}$ -invariant measure of linear association.*

PROOF. As to  $\sup_{\mathbf{G}}(\gamma)$ , the only point deserving attention is its continuity in datamatrices that do not contain constant columns. Let  $\gamma$  be an  $m$ -dimensional measure of linear association and  $\mathbf{G}$  a compact multiplicative group of  $m \times m$  matrices. To prove continuity of  $\sup_{\mathbf{G}}(\gamma)$  in non-degenerate datamatrices, let  $\mathbf{d}$  be any such matrix. The non-degenerate datamatrices can be looked upon as being an open set in  $\mathbb{R}^{n \times m}$ . It follows that  $\mathbf{d}$  has a compact neighborhood  $X$  that exclusively contains non-degenerate datamatrices. Now the map

$$(\mathbf{x}, \mathbf{g}) \mapsto \gamma(\mathbf{x}\mathbf{g})$$

is continuous on the Cartesian product  $X \times \mathbf{G}$ . Applying the foregoing lemma it follows that  $\sup_{\mathbf{G}}(\gamma)$  is continuous on  $X$  and thus in  $\mathbf{d}$ . This proves that  $\sup_{\mathbf{G}}(\gamma)$  is continuous in all non-degenerate  $\mathbf{d}$ . Next, let  $\mathbf{d}$  be a degenerate datamatrix that contains no constant columns. To prove that  $\sup_{\mathbf{G}}(\gamma)$  is continuous in  $\mathbf{d}$ , let  $\mathbf{d}_1, \mathbf{d}_2, \dots$  be any sequence that converges to  $\mathbf{d}$ . Then one evidently has

$$\gamma(\mathbf{d}_n) \leq \sup_{\mathbf{G}}(\gamma)(\mathbf{d}_n) \leq 1.$$

The sequence  $\gamma(\mathbf{d}_1), \gamma(\mathbf{d}_2), \dots$  converging to 1 it follows from this that

$$\lim_{n \rightarrow \infty} \sup_{\mathbf{G}}(\gamma)(\mathbf{d}_n) = 1 = \sup_{\mathbf{G}}(\gamma)(\mathbf{d}).$$

This shows that  $\sup_{\mathbf{G}}(\gamma)$  is continuous in  $\mathbf{d}$ . It can be proved in very much the same way that  $\inf_{\mathbf{G}}(\gamma)$  is a measure of linear association in cases where  $\gamma$  is continuous.

The theorem above can be successfully applied by taking for  $\mathbf{G}$  the orthogonal group  $\mathbf{O}$ , which is a compact group indeed. Thus the theorem provides a tool to create geometric measures.

In Theorem 5.4 compactness of  $\mathbf{G}$  is essential. To illustrate this, when maximizing a geometric measure  $\gamma$  with respect to the non-compact diagonal group  $\mathbf{D}$  the resulting map  $\sup_{\mathbf{D}}(\gamma)$  is (in virtue of Theorem 5.2) not a measure of linear association. In contrast to this one has:

**THEOREM 5.5.** *Let  $\gamma$  be a continuous measure of linear association. Then the  $\mathbf{D}$ -minimization of  $\gamma$  is an algebraic measure of linear association.*

**PROOF.** As in the proof of the previous theorem, the only point deserving attention is the continuity of  $\inf_{\mathbf{D}}(\gamma)$ . To start the proof of this, suppose that  $\gamma$  is an  $m$ -dimensional continuous measure of linear association and that  $\mathbf{d}$  is a matrix with  $\zeta(\mathbf{d}) > 0$ . It has to be proved then that  $\gamma$  is continuous in  $\mathbf{d}$ . Lemma 5.3 will be used to bring this about. To define the space  $X$  in this lemma, note that the set of all  $n \times m$  matrices  $\mathbf{x}$  with  $\zeta(\mathbf{x}) > 0$  can be considered an open subset of  $\mathbb{R}^{n \times m}$ . It follows that  $\mathbf{d}$  has a compact neighborhood  $X$  that consists exclusively of datamatrices  $\mathbf{x}$  with  $\zeta(\mathbf{x}) > 0$ . In preparing the definition of  $Y$ , let  $\mathbf{D}_1$  be the subset of  $\mathbf{D}$  defined by

$$\mathbf{D}_1 = \{\mathbf{g} \in \mathbf{D} \mid \text{trace}(\mathbf{g}) = 1\}.$$

Then, because  $\gamma$  is invariant under scalar multiplication, one has

$$(2). \quad \inf_{\mathbf{D}}(\gamma)(\mathbf{x}) = \inf_{\mathbf{g} \in \mathbf{D}} \gamma(\mathbf{x}\mathbf{g}) = \inf_{\mathbf{g} \in \mathbf{D}_1} \gamma(\mathbf{x}\mathbf{g})$$

Define  $Y$  to be the closure of  $\mathbf{D}_1$ . Then  $Y$  consists of all  $m \times m$  diagonal matrices of trace 1 with with entries  $\geq 0$ . Contrary to  $\mathbf{D}_1$  the set  $Y$  is a compact metric space. For every  $\mathbf{x} \in X$  and every  $\mathbf{g} \in Y$  the matrix  $\mathbf{x}\mathbf{g}$  is a datamatrix and therefore  $\gamma$  is continuous in it. It follows that the map

$$(\mathbf{x}, \mathbf{g}) \mapsto \gamma(\mathbf{x}\mathbf{g})$$

is continuous on the Cartesian product  $X \times Y$ . Therefore, by Lemma 5.3, the map

$$\mathbf{x} \mapsto \min_{\mathbf{g} \in Y} \gamma(\mathbf{x}\mathbf{g})$$

is continuous on  $X$  and thus in  $\mathbf{d}$ . However, by (2) and continuity of  $\gamma$ , the above is just the map  $\inf_{\mathbf{D}}(\gamma)$ . It follows that this map is continuous in  $\mathbf{d}$ , which was to be proved.

REMARK 5.6. Geometric measures being continuous in every datamatrix, the result above can be applied to them.

The theorems in this section will be useful in creating examples of invariant measures of linear association.

## 6. Shuttling between the algebraic and geometric class

Given an algebraic measure of linear association, it can be turned into a geometric one by means of Theorem 5.4. Given a geometric measure, it can be turned into an algebraic one by means of Theorem 5.5. This makes shuttling between the algebraic and geometric class possible. As to this process one evidently has

$$\inf_{\mathbf{D}}\{\sup_{\mathbf{O}}(\gamma)\} \geq \gamma$$

for algebraic measures and

$$\sup_{\mathbf{O}}\{\inf_{\mathbf{D}}(\gamma)\} \leq \gamma$$

for geometric measures.

DEFINITION 6.1. An algebraic or geometric measure of linear association will be called *reflexive* if there is equality in the corresponding inequality above.

In this terminology one has:

THEOREM 6.2. *An algebraic measure of linear association is reflexive if and only if it is a  $\mathbf{D}$ -minimization of a geometric measure. A geometric measure is reflexive if and only if it is an  $\mathbf{O}$ -maximization of an algebraic measure.*

PROOF. To prove the statement about algebraic measures, suppose that  $\gamma$  is an algebraic measure of the form  $\inf_{\mathbf{D}}(\delta)$ , where  $\delta$  is a geometric measure. Then one evidently has

$$\gamma \leq \delta.$$

It follows from this that

$$\sup_{\mathbf{O}}(\gamma) \leq \sup_{\mathbf{O}}(\delta) = \delta.$$

It thus appears that

$$\inf_{\mathbf{D}}\{\sup_{\mathbf{O}}(\gamma)\} \leq \inf_{\mathbf{D}}(\delta) = \gamma.$$

The opposite inequality being structurally true, it follows that there is actually equality in the above and thus that  $\gamma$  is reflexive. If, conversely,  $\gamma$  is reflexive, then it is evidently a measure of the form  $\inf_{\mathcal{D}}(\delta)$ . Just take for  $\delta$  the measure  $\sup_{\mathcal{O}}(\gamma)$ . This proves the statement about algebraic measures. For geometric measures one can use similar arguments.

The next sections will provide examples of reflexive and non-reflexive measures.

### 7. Maximizing the Hadamard coefficient

In this section a useful algebraic measure of linear association will be defined. The measure will be  $\mathcal{O}$ -maximized to make it geometric.

Let  $\mathbf{d}$  be an arbitrary  $n \times m$  datamatrix and let  $\mathbf{C} = \mathbf{C}(\mathbf{d})$  be its covariance matrix. By Hadamard's inequality (see for example [15], p. 119) one then has

$$\det(\mathbf{C}) \leq C_{11}C_{22} \dots C_{mm}.$$

Leaning on this inequality a useful measure of linear association can be defined:

DEFINITION 7.1. For every  $n \times m$  datamatrix  $\mathbf{d}$  the *Hadamard correlation coefficient*  $\rho_H(\mathbf{d})$  is understood to be:

$$\rho_H(\mathbf{d}) = \sqrt{1 - \frac{\det(\mathbf{C})}{C_{11}C_{22} \dots C_{mm}}}.$$

The Hadamard coefficient is set to 1 if the denominator in the above vanishes.

Note that the Hadamard coefficient is a symmetric algebraic measure. The symmetry contrasts to Pearson's coefficient, which is non-symmetric in cases where  $m > 2$ . For  $m = 2$  the Hadamard coefficient is the same as Pearson's coefficient.

The remainder of this section is devoted to the maximization of the Hadamard coefficient over the orthogonal group. As before, the character  $\mathbf{q}$  will systematically stand for an orthogonal matrix.

Note that the Hadamard coefficient can be expressed in terms of the function  $\zeta$ , which was defined in a previous section, as

$$(3) \quad \rho_H(\mathbf{d}) = \sqrt{1 - \frac{\det[\mathbf{C}(\mathbf{d})]}{\zeta(\mathbf{d})}}.$$

In maximizing the expression  $\rho_H(\mathbf{d}\mathbf{q})$  one should keep in mind that

$$\det[\mathbf{C}(\mathbf{d}\mathbf{q})] = \det[\mathbf{C}(\mathbf{d})].$$

Hence maximization of the expression  $\rho_H(\mathbf{d}\mathbf{q})$  amounts to the same as maximizing  $\zeta(\mathbf{d}\mathbf{q})$ .

**THEOREM 7.2.** *Let  $\mathbf{d}$  be an arbitrary  $n \times m$  matrix and let  $\lambda_1, \lambda_2, \dots, \lambda_m$  be the eigenvalues of the covariance matrix of  $\mathbf{d}$ . Then one has*

$$\max_{\mathbf{q} \in \mathcal{O}} \rho_H(\mathbf{d}\mathbf{q}) = \sqrt{1 - m^m \frac{\lambda_1 \lambda_2 \dots \lambda_m}{(\lambda_1 + \lambda_2 + \dots + \lambda_m)^m}}.$$

*In the above an orthogonal matrix  $\mathbf{q}$  is maximizing if and only if the columns in  $\mathbf{d}\mathbf{q}$  are of equal variance.*

**PROOF.** Given a fixed  $n \times m$  matrix  $\mathbf{d}$ , the map

$$\mathbf{q} \mapsto \zeta(\mathbf{d}\mathbf{q})$$

is a continuous function of  $\mathbf{q}$ . The orthogonal group being a compact set, it follows (see [13]) that there is a  $\mathbf{q}$  that maximizes the expression  $\zeta(\mathbf{d}\mathbf{q})$ . Using Lemma 4.6 and the fact that the orthogonal matrices form a group, such a maximizing  $\mathbf{q}$  is necessarily such that  $\mathbf{d}\mathbf{q}$  has columns of equal variance. The sum of the variances of the columns in the matrix  $\mathbf{d}\mathbf{q}$  is equal to the trace of the covariance matrix  $\mathbf{C}(\mathbf{d}\mathbf{q})$ . The trace of  $\mathbf{C}(\mathbf{d}\mathbf{q})$ , however, is the same as the trace of  $\mathbf{C}(\mathbf{d})$ . It follows that for a maximizing  $\mathbf{q}$  the variances of the columns in  $\mathbf{d}\mathbf{q}$  are all equal to

$$\frac{\lambda_1 + \lambda_2 + \dots + \lambda_m}{m}.$$

It thus appears that

$$\max_{\mathbf{q} \in \mathcal{O}} \zeta(\mathbf{d}\mathbf{q}) = \left( \frac{\lambda_1 + \lambda_2 + \dots + \lambda_m}{m} \right)^m.$$

The theorem follows when plugging the above into (3).

**COROLLARY 7.3.** *In the case of a 2-column datamatrix  $\mathbf{d}$  the Euclidean correlation coefficient  $\rho_E(\mathbf{d})$  can be expressed as*

$$\rho_E(\mathbf{d}) = \max_{\mathbf{q} \in \mathcal{O}} \rho_P(\mathbf{d}\mathbf{q}) = \max_{\mathbf{q} \in \mathcal{O}} \rho_H(\mathbf{d}\mathbf{q}) = \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}.$$

This corollary has been proved in a somewhat different way in [7]. In the next section there will be reliance on this specific result in order to generalize it.

REMARK 7.4. In the notations of the previous section the map

$$\mathbf{d} \mapsto \max_{\mathbf{q} \in \mathcal{O}} \rho_H(\mathbf{d}\mathbf{q})$$

is just the measure  $\sup_{\mathcal{O}}(\rho_H)$ . By the foregoing theorem one has

$$\sup_{\mathcal{O}}(\rho_H)(\mathbf{d}) = \rho_H(\mathbf{d})$$

if and only if the columns of the datamatrix  $\mathbf{d}$  are of equal variance. This can be brought about by multiplying  $\mathbf{d}$  by a suitable positive diagonal matrix. It follows that the D-minimization of  $\sup_{\mathcal{O}}(\rho_H)$  is equal to  $\rho_H$ . Otherwise stated, the Hadamard coefficient represents an example of a reflexive symmetric algebraic measure.

### 8. Maximizing Pearson's coefficient

This section focuses on the maximization of Pearson's correlation coefficient in order to obtain a general explicit expression for the Euclidean coefficient. Surprisingly this expression turns out to be generally the same as in the case for 2-column datamatrices in that it only depends on the ratio of  $\lambda_{\min}$  and  $\lambda_{\max}$ .

THEOREM 8.1. *For an arbitrary  $n \times m$  datamatrix  $\mathbf{d}$  one has*

$$\rho_E(\mathbf{d}) = \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}$$

where  $\lambda_{\min}$  and  $\lambda_{\max}$  are the smallest and largest eigenvalue of the covariance matrix of  $\mathbf{d}$ .

PROOF. First of all note that, because of translation invariance, it suffices to prove this theorem for centered datamatrices. The particular case of 2-column datamatrices was already described in the corollary to Theorem 7.2. The proof of the general case will lean on this. Note that for degenerate datamatrices the result is trivial. For the case of a non-degenerate datamatrix  $\mathbf{d}$  the first step in the proof is to show that for such matrices one has

$$(4) \quad \rho_P(\mathbf{d}) \leq \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}.$$

To see this, choose an arbitrary  $n \times (m + 1)$  centered non-degenerate datamatrix  $\mathbf{d}$  and run a linear regression in which  $\mathbf{y}$  is taken as the response variable and the columns of  $\mathbf{x}$  as explanatory variables. Let, in a self-explanatory notation, the vector  $\mathbf{b} = (b_0, b_1, \dots, b_m)$  present the regression coefficients that emanate in this way. Because  $\mathbf{d}$  is centered one has  $b_0 = 0$ . Now set

$$\hat{\mathbf{y}} = b_1 \mathbf{x}_{\bullet 1} + b_2 \mathbf{x}_{\bullet 2} + \dots + b_m \mathbf{x}_{\bullet m}.$$

The vector  $\hat{\mathbf{y}}$  is (see [10]) the orthogonal projection of the vector  $\mathbf{y}$  on the linear space spanned by the columns of  $\mathbf{x}$  and  $\rho_P(\mathbf{d})$  is the cosine of the angle between  $\mathbf{y}$  and this linear space. Thus one has

$$\rho_P(\mathbf{d}) = \rho_P([\hat{\mathbf{y}}, \mathbf{y}]) = \rho_P([b_1\mathbf{x}_{\bullet 1} + b_2\mathbf{x}_{\bullet 2} + \cdots + b_m\mathbf{x}_{\bullet m}, \mathbf{y}]).$$

In the above, because Pearson's coefficient is invariant under scaling transformations, the coefficients  $b_1, b_2, \dots, b_m$  can be modified in such a way that

$$b_1^2 + b_2^2 + \cdots + b_m^2 = 1.$$

From now on it will be assumed that the  $b_1, b_2, \dots, b_m$  satisfy the above. Denote the largest and the smallest eigenvalue of the covariance matrix of the  $n \times 2$  datamatrix

$$[b_1\mathbf{x}_{\bullet 1} + b_2\mathbf{x}_{\bullet 2} + \cdots + b_m\mathbf{x}_{\bullet m}, \mathbf{y}]$$

by  $\tilde{\lambda}_{\max}$  and  $\tilde{\lambda}_{\min}$ . The two eigenvalues  $\tilde{\lambda}_{\max}$  and  $\tilde{\lambda}_{\min}$  present the maximum and minimum value of the expression

$$\text{var}(c_1(b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + \cdots + b_m\mathbf{x}_m) + c_2\mathbf{y})$$

when optimizing over all coefficient vectors  $(c_1, c_2)$  with  $c_1^2 + c_2^2 = 1$ . However, for every such pair of numbers  $c_1$  and  $c_2$  one has

$$(c_1b_1)^2 + (c_1b_2)^2 + \cdots + (c_1b_m)^2 + c_2^2 = 1.$$

It follows that one necessarily has

$$\tilde{\lambda}_{\max} \leq \lambda_{\max} \quad \text{and} \quad \tilde{\lambda}_{\max} \geq \lambda_{\min}.$$

Now, glueing the pieces together and applying the corollary to Theorem 7.2, one arrives at

$$\begin{aligned} \rho_P(\mathbf{d}) = \rho_P([b_1\mathbf{x}_{\bullet 1} + b_2\mathbf{x}_{\bullet 2} + \cdots + b_m\mathbf{x}_{\bullet m}, \mathbf{y}]) &\leq \frac{\tilde{\lambda}_{\max} - \tilde{\lambda}_{\min}}{\tilde{\lambda}_{\max} + \tilde{\lambda}_{\min}} \\ &\leq \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}. \end{aligned}$$

This proves (4), which was the first step in the proof of this theorem. To start the second step, note that for an arbitrary orthogonal matrix  $\mathbf{q}$  the eigenvalues of the covariance matrices of  $\mathbf{d}$  and  $\mathbf{d}\mathbf{q}$  are the same. Hence (4) actually implies that for all  $\mathbf{d}$  and all  $\mathbf{q}$  one has

$$(5) \quad \rho_P(\mathbf{d}\mathbf{q}) \leq \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}.$$



Taking the maximum over all  $\mathbf{q}$  it follows that

$$(6) \quad \rho_E(\mathbf{d}) \leq \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}.$$

This completes the second step in the proof of this theorem. The third and last step consists of showing that for every  $\mathbf{d}$  there is an orthogonal matrix  $\mathbf{q}$  that comes with equality in (5). To prove this, choose an arbitrary datamatrix  $\mathbf{d}$ . Let  $\mathbf{q}_0$  be an orthogonal matrix such that the columns of

$$\mathbf{d}' = \mathbf{d}\mathbf{q}_0$$

are uncorrelated. Without loss of generality one may replace  $\mathbf{d}$  by  $\mathbf{d}'$ : if there is a  $\mathbf{q}'$  that makes  $\rho_P(\mathbf{d}'\mathbf{q}')$  equal to the right side of (5), then  $\mathbf{q}'\mathbf{q}$  will do the same job for  $\mathbf{d}$ . Hence there is no problem in assuming the columns of  $\mathbf{d}$  to be uncorrelated. Similar arguments show that permutation of the columns is allowed. For this reason one may also assume that the two last columns in  $\mathbf{d}$  have the lowest and highest variance of all columns respectively. Because the columns of  $\mathbf{d}$  are uncorrelated, the variances of the last two columns are necessarily equal to  $\lambda_{\min}$  and  $\lambda_{\max}$ . Let  $\mathbf{d}_0$  be the  $n \times 2$  matrix consisting of the last two columns of  $\mathbf{d}$ . Then, by Theorem 7.2 and its corollary, there exists a  $2 \times 2$  orthogonal matrix  $\mathbf{q}_0$  that makes the variances of the columns in  $\mathbf{d}_0\mathbf{q}_0$  equal. For this matrix one has

$$\rho_P(\mathbf{d}_0\mathbf{q}_0) = \rho_H(\mathbf{d}_0\mathbf{q}_0) = \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}.$$

Let  $\mathbf{d}''$  be the matrix  $\mathbf{d}$  where the last two columns have been replaced by the two columns in  $\mathbf{d}_0\mathbf{q}_0$ . Then in  $\mathbf{d}''$  the last two columns are uncorrelated to the remaining columns. For this reason one has

$$\rho_P(\mathbf{d}'') = \rho_P(\mathbf{d}_0\mathbf{q}_0) = \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}.$$

Now define  $\mathbf{q}$  to be the identity matrix in which the lower right  $2 \times 2$  block has been replaced by  $\mathbf{q}_0$ . Then  $\mathbf{q}$  is orthogonal and one evidently has  $\mathbf{d}'' = \mathbf{d}\mathbf{q}$ . In virtue of the above this is a  $\mathbf{q}$  that brings equality in (5), thus completing the proof of the theorem.

**REMARK 8.2.** The eigenvalues  $\lambda_{\min}$  and  $\lambda_{\max}$  depend in a continuous way on the underlying datamatrix  $\mathbf{d}$ . For this reason one may conclude that, in spite of the fact that Pearson's coefficient generally does not meet the necessary continuity requirements to be a measure, the Euclidean correlation coefficient is a

genuine geometric measure of linear association. By exploiting the inequality

$$\inf_{\mathbf{D}}(\rho_E) \geq \rho_P$$

it is easily derived that  $\rho_E$  is reflexive. Note that, for dimensions higher than two, the algebraic measure  $\inf_{\mathbf{D}}(\rho_E)$  cannot possibly be equal to the non-measure  $\rho_P$ .

### 9. A non-reflexive geometric measure

To construct an example of a non-reflexive geometric measure, define for every 2-column datamatrix  $\mathbf{d}$  the number  $\gamma(\mathbf{d})$  by

$$\gamma(\mathbf{d}) = \inf_{\mathbf{O}}(\rho_E) \left( \mathbf{d} \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \right) = \min_{\mathbf{q} \in \mathbf{O}} \rho_E \left( \mathbf{d}\mathbf{q} \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \right).$$

By Theorem 5.4 the map  $\mathbf{d} \mapsto \gamma(\mathbf{d})$  is a geometric measure of linear association. It will be shown in this section that  $\gamma$  is non-reflexive. To start the reasoning, define for every real number  $\theta$  the datamatrix  $\mathbf{e}(\theta)$  by

$$(7) \quad \mathbf{e}(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.$$

It will turn out that for this type of datamatrices one systematically has

$$(8) \quad \inf_{\mathbf{D}}(\gamma)(\mathbf{e}(\theta)) = 0.$$

To see this, let  $\mathbf{C}(\theta)$  be the covariance matrix of  $\mathbf{e}(\theta)$  and let for arbitrary  $\mathbf{s} \in \mathbf{D}$  the numbers  $\lambda_{\max}(\theta, \mathbf{s})$  and  $\lambda_{\min}(\theta, \mathbf{s})$  be the largest and smallest eigenvalue of the matrix

$$(9) \quad \mathbf{s}\mathbf{C}(\theta)\mathbf{s}.$$

Then one may write

$$\frac{\lambda_{\max}(\theta, \mathbf{s})}{\lambda_{\min}(\theta, \mathbf{s})} = \frac{1 + \rho_E(\mathbf{e}(\theta)\mathbf{s})}{1 - \rho_E(\mathbf{e}(\theta)\mathbf{s})}.$$

From the fact that  $\rho_E$  is the  $\mathbf{O}$ -maximization of Pearson's measure  $\rho_P$  and the fact that Pearson's measure is reflexive in the 2-dimensional case it follows that

$$(10) \quad \min_{\mathbf{s}} \frac{\lambda_{\max}(\theta, \mathbf{s})}{\lambda_{\min}(\theta, \mathbf{s})} = \frac{1 + \rho_P(\mathbf{e}(\theta))}{1 - \rho_P(\mathbf{e}(\theta))}.$$

Besides this one has

$$\sup_{\mathbf{s}} \frac{\lambda_{\max}(\theta, \mathbf{s})}{\lambda_{\min}(\theta, \mathbf{s})} = +\infty.$$

Denoting the right side of (10) by  $m(\theta)$ , one may conclude that, given any fixed  $\theta$ , all values in the interval  $[m(\theta), +\infty)$  are taken on by the expression  $\lambda_{\max}(\theta, \mathbf{s})/\lambda_{\min}(\theta, \mathbf{s})$  when  $\mathbf{s}$  runs through  $\mathbf{D}$ . In virtue of (7) and (10), one has

$$\max_{\theta} m(\theta) = \max_{\theta} \min_{\mathbf{s}} \frac{\lambda_{\max}(\theta, \mathbf{s})}{\lambda_{\min}(\theta, \mathbf{s})} = \frac{1 + \rho_E(\mathbf{e}(0))}{1 - \rho_E(\mathbf{e}(0))} = 3.$$

Hence, there is for every  $\theta$  an  $\mathbf{s}$  such that the matrix (9) shows the proportion 1 : 9 in its eigenvalues. Consequently there is for every  $\theta$  an  $\mathbf{s} \in \mathbf{D}$  and an orthogonal matrix  $\mathbf{q}$  such that the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}^{\dagger} \mathbf{q}^{\dagger} \mathbf{s}^{\dagger} \mathbf{C}(\theta) \mathbf{s} \mathbf{q} \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$$

is proportional to the identity matrix. The above, however, is the covariance matrix of the datamatrix

$$\mathbf{e}(\theta) \begin{pmatrix} 1 & 0 \\ 0 & s \end{pmatrix} \mathbf{q} \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}.$$

It follows that the value of  $\rho_E$  for the matrix above must be zero. This proves (8), which, in turn, proves that

$$\sup_{\mathbf{O}} \{\inf_{\mathbf{D}}(\gamma)\}(\mathbf{e}(0)) = 0.$$

Opposite to this, however, one has

$$\gamma(\mathbf{e}(0)) = \frac{1}{2}.$$

It thus appears that the  $\mathbf{O}$ -invariant open set of all  $3 \times 2$  datamatrices  $\mathbf{d}$  such that

$$\sup_{\mathbf{O}} \{\inf_{\mathbf{D}}(\gamma)\}(\mathbf{d}) < \gamma(\mathbf{d})$$

is non-empty. For this reason  $\gamma$  cannot possibly be reflexive. As a peculiar detail, Monte Carlo experiments seem to suggest that the complement of the set above is of positive Lebesgue measure.

### 10. A non-reflexive symmetric algebraic measure

A wellknown way (see [4], [11]) to turn an arbitrary measure of linear association into an algebraic measure is by *standardizing* it. In this section a non-reflexive algebraic measure will be created in this way. To be more precise, given an arbitrary datamatrix  $\mathbf{d}$  that contains no constant columns, one can define its standardization  $\text{std}(\mathbf{d})$  as the datamatrix consisting of the standardized columns of  $\mathbf{d}$ . For an arbitrary measure of linear association  $\gamma$  one could thus define a matrix function  $\text{std}(\gamma)$  by setting

$$\text{std}(\gamma)(\mathbf{d}) = \begin{cases} \gamma(\text{std}(\mathbf{d})) & \text{if } \mathbf{d} \text{ contains no constant columns,} \\ 1 & \text{if } \mathbf{d} \text{ contains constant columns.} \end{cases}$$

In this way the map  $\mathbf{d} \mapsto \text{std}(\gamma)(\mathbf{d})$  presents an algebraic measure of linear association. From now on  $\gamma$  will stand for the non-reflexive geometric measure constructed in the previous section. It will be proved that the measure  $\text{std}(\gamma)$  is not reflexive. Generally, in proving that an algebraic measure  $\delta$  is not reflexive it suffices to show that

$$(11) \quad \sup_{\mathbf{0}}(\delta)(\mathbf{e}) > \delta(\mathbf{e})$$

where  $\mathbf{e}$  is an  $n \times m$  datamatrix in which the upper  $m \times m$  block consists of the identity matrix and the lower block of a zero matrix. By symmetry, namely, one has for such datamatrices that

$$\inf_{\mathbf{D}} \{ \sup_{\mathbf{0}}(\delta) \}(\mathbf{e}) = \sup_{\mathbf{0}}(\delta)(\mathbf{e}).$$

In order to apply this to the case where  $\delta = \text{std}(\gamma)$ , the matrix  $\mathbf{e}$  is set to

$$\mathbf{e} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Note that, in the notation of the previous section,  $\mathbf{e} = \mathbf{e}(0)$ . For this datamatrix one has

$$\text{std}(\gamma)(\mathbf{e}) = \gamma(\mathbf{e}) = \frac{1}{2}.$$

Now, in the notations of the previous section, define the datamatrix  $\mathbf{d}$  to be the matrix  $\mathbf{e}(\frac{\pi}{4})$ . Then the covariance matrix of the standardization of  $\mathbf{d}$  is proportional to the  $2 \times 2$  identity matrix. By exploiting this it is easily derived that

$$\sup_{\mathbf{0}}(\text{std}(\gamma))(\mathbf{e}) \geq \text{std}(\gamma)(\mathbf{d}) = \frac{4}{5}.$$

Hence the inequality (11) is valid and for this reason the measure  $\text{std}(\gamma)$  cannot possibly be reflexive. Note that  $\text{std}(\gamma)$  is symmetric.

## 11. Closing remarks

As to the computational aspects, given a datamatrix, minimization of a geometric measure over  $D$  can be carried out straightforward. Optimization over  $O$  can be carried out by exploiting the fact that the exponential map maps the Lie algebra of  $O$  onto the set of all special orthogonal matrices (see [3]). Thus, by using computational software, one could get hints, for example, whether a measure is reflexive or not in the duality between algebraic and geometric measures. A software library, working under the free computational program Scilab, is obtainable by just mailing the author. An accompanying manual goes with it.

## REFERENCES

1. Bourbaki, N., *Algèbre*, Hermann, Paris 1965.
2. Bourbaki, N., *Topologie générale*, Hermann, Paris 1997.
3. Curtis, M. L., *Matrix Groups*, Springer, Berlin 1998.
4. Coleman, B. J., *A coefficient of linear correlation based on the method of least squares and the line of best fit*, Ann. Math. Stat. 3 (1932), no. 2., 79–85.
5. Gebelein, H., *Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und sein Zusammenhang mit der Ausgleichsrechnung*, Z. Angew. Math. Mech. 21 (1941), 364–379.
6. Lang, S., *Real and Functional Analysis*, Grad. Texts in Math. 142, Springer, Berlin 1993.
7. Loh, W. Y., *Does the correlation coefficient really measure the degree of clustering around a line?*, J. Educ. Stat. 12 (1987), 235–239.
8. Moore, D. S., *Statistics: Concepts and Controversies*, W. H. Freeman, San Francisco 2006.
9. Peña, D., and Rodríguez, J., *Descriptive measures of multivariate scatter and linear dependence*, J. Multivariate Anal. 85 (2003), 361–374.
10. Pestman, W. R., *Mathematical Statistics*, second edition, de Gruyter, Berlin 2009.
11. Pestman, W. R., *Classification of bivariate measures of linear association*, Elem. Math. (to appear 2010).
12. Rényi, A., *On measures of dependence*, Acta Math. Acad. Sci. Hungar. 10 (1959), 441–451.
13. Rudin, W., *Principles of Mathematical Analysis*, 2nd edition, McGraw-Hill, New York 1964.
14. Sénéchal, B., *Groupes et géométries*, Hermann, Paris 1979.
15. Voïévodine, V., *Algèbre linéaire*, Éditions Mir, Moscow 1976.