NORTH-HOLLAND

A Family of Matrices, the Discretized Brownian Bridge, and Distance-Based Regression*

J. Fortiana and C. M. Cuadras
Department d'Estadística
Universitat de Barcelona
08028-Barcelona, Spain

Submitted by George P. H. Styan

ABSTRACT

The investigation of a distance-based regression model, using a one-dimensional set of equally spaced points as regressor values and $\sqrt{|x-y|}$ as a distance function, leads to the study of a family of matrices which is closely related to a discrete analog of the Brownian-bridge stochastic process. We describe its eigenstructure and several properties, recovering in particular well-known results on tridiagonal Toeplitz matrices and related topics. © 1997 Elsevier Science Inc.

1. INTRODUCTION

The distance-based regression model (Cuadras, 1989; Cuadras and Arenas, 1990; Cuadras et al., 1996) is an extension of the ordinary linear model which can be applied to qualitative or, in general, to mixed continuous and discrete explanatory variables, provided that a distance δ can be defined on the set of values of these variables.

A brief description of the method is as follows: Assume we are given n cases or individuals, on which the values $y_1, \ldots, y_n$ of a continuous response variable y have been observed, corresponding to the values $\mathbf{w}_1, \ldots, \mathbf{w}_n$ of a

* Work supported in part by grants CGYCIT PB93-0784 and 1995SGR-00085. E-mail: C. M. Cuadras, carlesm@porthos.bio.ub.es; J. Fortiana, fortiana@cerber.mat.ub.es.
The authors thank M. A. Stephens for his helpful comments.

LINEAR ALGEBRA AND ITS APPLICATIONS 264:173–188 (1997)

655 Avenue of the Americas, New York, NY 10010

0024-3795/97/$17.00
PII S0024-3795(97)00051-7

set of explanatory variables. We compute the $n \times n$ distance matrix $\mathbf{\Delta} = (\delta(\mathbf{w}_i, \mathbf{w}_j))$, and from it, we obtain the matrix $\mathbf{X}$ of principal coordinates (see below). Then we perform an ordinary least-squares regression of y, taking the columns of $\mathbf{X}$ as predictors.

Principal-coordinate analysis, also called *classic metric scaling*, is a solution to the following problem: given an $n \times n$ symmetric matrix $\mathbf{\Delta}$ whose entries are the distances between the elements of a set $\mathscr{U}$, we want to obtain n vectors $\{\mathbf{x}_i,\ 1 \leqslant i \leqslant n\}$ in some $\mathbb{R}^k$, such that the Euclidean distance $\|\mathbf{x}_i - \mathbf{x}_j\|$ equals the (i, j) entry in $\mathbf{\Delta}$, $1 \leqslant i, j \leqslant n$. Such a set of vectors is called a *Euclidean configuration* for $\mathscr{U}$. See, e.g., Mardia et al. (1979) for a detailed account of the technique. The $n \times p$ matrix $\mathbf{X}$ referred to in the previous paragraph is built by stacking together these (row) vectors. The columns of $\mathbf{X}$ form an orthogonal set in $\mathbb{R}^n$, which we take as the linear predictors in the distance-based regression model

We refer to Cuadras and Arenas (1990) and Cuadras et al. (1996) for a more thorough discussion of the model and its properties. Here we remark only that if the values $\mathbf{w}$ of the explanatory variables actually belong to some Euclidean space and if we choose the Euclidean distance for δ, we recover the ordinary least-squares solution. Even if $\mathbf{w} \in \mathbb{R}^p$, distance functions other than the Euclidean can be used, and in fact

$$\delta(\mathbf{w}_i, \mathbf{w}_j) = \sqrt{|w_{i1} - w_{j1}| + \cdots + |w_{ip} - w_{jp}|},$$

the square root of the L^1 distance, has excellent properties, and with its use the distance-based model can often replace advantageously a nonlinear model.

To examine theoretically the reasons for this behavior, we consider a one-dimensional model, with a set $\mathscr{U} = \{0, 1, \ldots, n\}$ of $n + 1$ equally spaced points as predictor values, and compute the principal coordinates $\mathbf{X}$ for these points with the distance $\delta_{ij} = \sqrt{|i - j|}$. In principle the first step is to write the matrix $\mathbf{\Delta}$ of distances, but the following alternative is more convenient: We observe that the Euclidean distance between the ith and jth rows of $(n + 1) \times n$ matrix

$$\mathbf{U} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & 1 & 1 & \ddots & 0 \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix} \tag{1}$$

is $\sqrt{|i-j|}$; hence from the property of duality between principal components and principal coordinates (Mardia et al., 1979, p. 404) we have $\mathbf{X} = \mathbf{HUV}$, where $\mathbf{H} = \mathbf{I} - \mathbf{1}_{n+1}\mathbf{1}'_{n+1}/(n+1)$ is the $(n+1) \times (n+1)$ centering matrix and $\mathbf{V}$ is the matrix of orthonormalized eigenvectors of the covariance matrix of $\mathbf{U}$.

The covariance s_{ij} between columns i and j of $\mathbf{U}$ equals $(n+1)^{-2}$ times

$$c_{ij} = (n+1)\min\{i,j\} - ij, \qquad 1 \leqslant i,j \leqslant n. \tag{2}$$

This covariance,

$$s_{ij} = \min\left\{\frac{i}{n+1}, \frac{j}{n+1}\right\} - \frac{i}{n+1}\frac{j}{n+1}, \qquad 1 \leqslant i,j \leqslant n,$$

can be regarded as the discretization of the covariance function of the Brownian bridge

$$K(s,t) = \min\{s,t\} - st, \qquad 0 \leqslant s,t \leqslant 1, \tag{3}$$

on a partition of $[0,1]$ in $n+1$ subintervals of equal length. As such, it appears in Anderson and Stephens (1993, 1997) and in tests of goodness of fit; see Durbin and Knott (1972). Properties of the Brownian bridge are discussed by Anderson and Stephens (1997). In our case the underlying continuous structure is a Bernoulli rather than a Gaussian process (Cuadras and Fortiana, 1993, 1995).

This process is presented and compared with the Brownian bridge in Section 2. The family of matrices containing the covariance of the discretized Brownian bridge is defined in Section 3, and some historical references are given in Section 4. The eigenvectors and eigenvalues are obtained in Section 5, and closed formulae for three particular matrices are given in Section 6.

2. CONTINUOUS AND DISCRETE PROCESSES AND SOME REPRESENTATIONS

A continuous version of the matrix $\mathbf{U}$ is the function

$$u : [0,1] \times [0,1] \to \mathbb{R},$$

$$(x,s) \mapsto u(x,s) = \begin{cases} 0 & \text{if} \quad x \leqslant s, \\ 1 & \text{if} \quad x > s. \end{cases}$$

By considering a uniform probability on [0, 1], i.e., the Lebesgue measure, we see that for each $s \in [0, 1]$, the indicator function $U_s = u(\cdot, s)$ of the interval $[s, 1]$ is a Bernoulli random variable with parameter $p = 1 - s$. Hence

$$\mathbb{U} = \{U_s, 0 \leqslant s \leqslant 1\}, \tag{4}$$

is a Bernoulli stochastic process. Its covariance function is readily computed: Since $U_s U_t = U_{\max(s,t)}$ for $s, t \in [0, 1]$, we have $\mathrm{cov}(U_s, U_t) = 1 - \max\{s, t\} - (1 - s)(1 - t) = K(s, t)$, the kernel defined in (3).

The Karhunen-Loève (principal components) expansion of the Brownian bridge $\mathbb{B} = \{B_s,\ s \in [0, 1]\}$, obtained from the eigenvalues and eigenfunctions of (3), is

$$B_s = \sum_{j=1}^{\infty} \frac{\sqrt{2}}{j\pi} (\sin j\pi s) X_j, \qquad s \in [0, 1], \tag{5}$$

where $\{X_j\}_{j \in \mathbb{N}}$ is a countable set of i.i.d. $N(0, 1)$ variables (see, e.g., Anderson and Darling, 1952). This representation means that both sides of (5) have the same probabilistic distribution. The representation of the Bernoulli process (4) is

$$U_s = \sum_{j=1}^{\infty} \frac{\sqrt{2}}{j\pi} (\sin j\pi s) Y_j, \qquad s \in [0, 1], \tag{6}$$

where $\{Y_j = \sqrt{2}(1 - \cos j\pi U)\}_{j \in \mathbb{N}}$ is a countable set of i.d. and *uncorrelated* variables, each with mean $\sqrt{2}$ and variance 1.

Both the Brownian bridge and the Bernoulli process (4) can be related to goodness-of-fit tests. From Parseval's identity we have

$$W^2 = \int_0^1 B_s^2 \, ds = \sum_{j=1}^{\infty} \frac{X_j^2}{j^2 \pi^2},$$

where W^2 is the limit of the Cramér–von Mises statistic W_n^2, which also has a similar but finite decomposition (see Anderson and Stephens, 1997). Similarly,

$$U = \int_0^1 U_s^2 \, ds = \sum_{j=1}^{\infty} \frac{Y_j^2}{j^2 \pi^2},$$

where U is a uniform [0, 1] random variable, namely the identity function on [0, 1]. This follows by direct integration of $U_s^2 = U_s$:

$$U(x) = \int_0^1 U_s(x)\, ds = \int_0^x 1\, ds = x.$$

As a measure of goodness of fit for the hypothesis that a sample follows a uniform (0, 1) distribution, Cuadras and Fortiana (1993) proposed the *maximum Hoeffding correlation*, ρ_n^+, between the sample empirical distribution F_n and the uniform (0, 1) distribution. In general, given two univariate distributions F and G, the maximum Hoeffding correlation is defined as the maximum of the set of correlations for bivariate distributions having F and G as marginals (see Dall-Aglio et al., 1991, for details). When $F = F_n$ and G is the uniform distribution,

$$\rho_n^+ = \frac{6}{S_Z n^2} \sum_{i=1}^{n} (2i - n - 1) Z_{(i)},$$

where $Z_{(1)} \leqslant \cdots \leqslant Z_{(n)}$ is the ordered sample and S_Z is the empirical standard deviation. Note that any test for a completely specified continuous distribution can be reduced to a test for a uniform (0, 1) distribution. Similarly to W_n^2, the correlation ρ_n^+ admits the decomposition (see Cuadras and Fortiana, 1993)

$$\rho_n^+ = \frac{4\sqrt{6}}{\pi^2} \sum_{j=0}^{\infty} \frac{\hat{\beta}_{2j+1}}{(2j+1)^2},$$

where

$$\hat{\beta}_j = \frac{\sqrt{2}}{j\pi S_Z} \sum_{i=1}^{n} \left(\sin \frac{(i-1)j\pi}{n} - \sin \frac{ij\pi}{n} \right) Z_{(i)}, \qquad j \geqslant 1,$$

are the sample counterparts of the correlation coefficients

$$\beta_j = \operatorname{corr}(U, Y_j) = \begin{cases} 4\sqrt{6}\,/j^2\pi^2 & \text{if } j \text{ is odd,} \\ 0 & \text{if } j \text{ is even.} \end{cases}$$

A sequence of goodness-of-fit tests can be constructed by comparing the coefficients β_j and their estimates $\hat{\beta}_j$. Alternatively, instead of using them to test the hypothesis, these coefficients are used to ascertain the distribution of the data in a geometrical way (Cuadras and Fortiana, 1994).

3. A FAMILY OF TRIDIAGONAL MATRICES

Consider the $n \times n$ matrix $\mathbf{C} = \mathbf{C}[n] = (c_{ij})$, $1 \leqslant i, j \leqslant n$, where c_{ij} is given in (2). The entries in $\mathbf{C}$ satisfy $c_{ij} = c_{n+1-j, n+1-i}$. Matrices with this property are called *centrosymmetric*. Equivalently, in terms of the permutation matrix

$$\mathbf{W} = \begin{pmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ \vdots & ⋰ & \vdots & \vdots \\ 1 & \cdots & 0 & 0 \end{pmatrix}, \tag{7}$$

the equality $\mathbf{WCW} = \mathbf{C}$ holds. A Toeplitz matrix is centrosymmetric, and nonsingular centrosymmetric matrices form a multiplicative group; see Cord and Sylvester (1962) and Good (1970). In general, the inverse of a Toeplitz matrix is a centrosymmetric matrix. A centrosymmetric matrix is not necessarily symmetric, but $\mathbf{C}$, as defined above, has both properties.

We now introduce two matrices $\mathbf{B}$ and $\tilde{\mathbf{B}}$ related to $\mathbf{C}$. Let $\mathbf{B} = \mathbf{B}[n] = (\min\{i, j\})_{1 \leqslant i, j \leqslant n}$, $\mathbf{b} = (1, 2, \ldots, n)'$, and $\tilde{\mathbf{B}} = \tilde{\mathbf{B}}[n] = (2n + 1)\mathbf{B} - 2\mathbf{bb}'$.

From the definitions we see that

$$\mathbf{C} = (n + 1)\mathbf{B} - \mathbf{bb}' = \frac{\mathbf{B} + \tilde{\mathbf{B}}}{2}.$$

The three matrices $\mathbf{B}[n]$, $\tilde{\mathbf{B}}[n]$, and $\mathbf{C}[n]$ can be described in terms of the one-parameter family of tridiagonal $n \times n$ matrices

$$\mathbf{F}[a, n] = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & \cdots & 0 & -1 & a \end{pmatrix} \tag{8}$$

where $a \in \mathbb{R}$.

A direct computation shows the equalities $\mathbf{B} = \mathbf{T}^2$, $\mathbf{T} = \mathbf{R}^{-1}$, and $\mathbf{R}^2 = \mathbf{F}[1, n]$, where the entries in $\mathbf{T} = (t_{ij})$ and $\mathbf{R} = (r_{ij})$ are defined by

$$t_{ij} = \begin{cases} 1 & \text{if } \ i + j \geqslant n + 1, \\ 0 & \text{otherwise}, \end{cases} \qquad r_{ij} = \begin{cases} 1 & \text{if } \ i + j = n + 1, \\ -1 & \text{if } \ i + j = n, \\ 0, & \text{otherwise}, \end{cases}$$

$$1 \leqslant i, j \leqslant n.$$

In particular, $\mathbf{F}[1, n] = \mathbf{B}[n]^{-1}$. From this equality and

$$\mathbf{F}[a, n] = \mathbf{B}^{-1} + (a - 1)\mathbf{u}_0\mathbf{u}_0',$$

where $\mathbf{u}_0 = (0, \dots, 0, 1)'$, we can obtain the inverse $\mathbf{F}[a, n]^{-1}$:

$$[n(a - 1) + 1]\mathbf{F}[a, n]^{-1} = [n(a - 1) + 1]\mathbf{B}[n] - (a - 1)\mathbf{bb}'.$$

Actually, $n(a - 1) + 1 = \det \mathbf{F}[a, n]$, as is easily checked. Using special cases of a, we have, in particular,

$$\mathbf{B}[n] = \mathbf{F}[1, n]^{-1}, \qquad \mathbf{C}[n] = (n + 1)\mathbf{F}[2, n]^{-1},$$

$$\tilde{\mathbf{B}}[n] = (2n + 1)\mathbf{F}[3, n]^{-1}.$$

It is worth noting that $\mathbf{F}[2, n]$ is a Toeplitz matrix, but its inverse is $(n + 1)^{-1}$ times the non-Toeplitz, but centrosymmetric matrix $\mathbf{C}[n]$.

4. HISTORICAL REMARKS

Besides distance-based regression and the discretized Brownian bridge, the matrices $\mathbf{F}[a, n]$, their inverses, and other related matrices have been studied in many contexts. Some examples are the following.

Numerical Linear Algebra

A particular instance (for $n = 12$) of

$$\mathbf{A} = (n + 1)\mathbf{11}' - \mathbf{B} = \max\{n + 1 - i, n + 1 - j\},$$

where **1** is the column vector of 1's, is used by Frank (1958) to validate algorithms for eigenvalue computation. Frank notices that $\mathbf{A}^{-1}$ is a tridiagonal matrix and gives a closed formula for its eigenvalues:

$$\lambda_q = \frac{1}{2}\left(1 - \cos\frac{2q-1}{2n+1}\pi\right)^{-1}, \qquad q = 1, \dots, n. \tag{9}$$

A related tridiagonal matrix is described by Longley (1981), who cites an unpublished note of M. Newman as its origin. This matrix, $\mathbf{D}_t'\mathbf{D}_t$, where $\mathbf{D}_t = \mathbf{I} - t\mathbf{N}$, t is a real parameter, and $\mathbf{N}$ is the nilpotent $n \times n$ matrix with ones in its subdiagonal and zeros elsewhere, is a remarkable example of a poorly conditioned matrix on which many numerical inversion algorithms fail for moderate values of n and t.

Theory of Inequalities

Fan, Taussky, and Todd (1955) find discrete versions of inequalities relating the integral of a function and that of its derivatives, e.g., if $x(0) = 0$,

$$\int_0^{\pi/2}[x(t)]^2\,dt < \int_0^{\pi/2}[x'(t)]^2\,dt, \text{ unless } x(t) = a\sin t.$$

Their proofs are based on finding the minimum value, i.e., the least eigenvalue, of quadratic forms build on $\mathbf{F}[a, n]$ and related matrices.

Oscillations of Discrete Mechanical Systems

Fan et al. use results of Rutherford (1947, 1951) concerning eigenvalues of several matrices, which arise in investigations of the motion of mechanical systems consisting of n equal material particles constrained to move in a straight line and linked together by elastic springs. In this context, eigenvectors give the normal modes of vibration, and the corresponding eigenvalues have the meaning of normal periods.

Partial Differential Equations

$\mathbf{F}[2, n]$ is a Toeplitz matrix and the most popular of this family of matrices, since it is the second-difference matrix, which appears in the discretization of the second-derivative operator (see Shintani, 1968).

Serial Correlation

The matrices obtained with this family are similar to some matrices of quadratic forms, used for computing serial correlations, especially in the circular model. See Anderson (1971), Durbin and Watson (1950, 1951), and von Neumann (1941). See also Anderson and Stephens (1997).

Goodness-of-Fit Statistics

$\mathbf{C}[n]$ and $\mathbf{F}[2, n]$ appear in the context of uniform goodness-of-fit statistics built from an ordered uniform (0, 1) sample $\mathbf{Z} = (Z_{(1)} \leqslant \cdots \leqslant Z_{(n)})$, like the ones considered by Anderson and Stephens (1997) and in Section 2 above, and others, e.g.

$$M_n^2 = \frac{n+1}{n} \sum_{j=1}^{n} \left(Z_{(j)} - \frac{j}{n+1} \right)^2,$$

an analog to W_n^2, studied by Durbin and Knott (1972). This ubiquity is due to the well-known fact that the matrix of variances and covariances of $\mathbf{Z}$ equals $(n+1)^{-2}(n+2)^{-1}\mathbf{C}[n]$.

5. EIGENVECTORS

The eigenvalues and eigenvectors of $\mathbf{F}[a, n]$ can be expressed in terms of Chebyshev polynomials of the second kind. These are defined by

$$U_0(\xi) = 1, \qquad U_1(\xi) = 2\xi, \qquad U_{k+2}(\xi) = 2\xi U_{k+1}(\xi) - U_k(\xi) \qquad (k \geqslant 0). \quad (10)$$

The polynomial $U_p(\xi)$ has the trigonometric representation

$$U_p(\xi) = \frac{\sin(p+1)\theta}{\sin\theta}, \qquad \text{where} \quad \xi = \cos\theta,$$

and direct computation gives the identities, for $p > 0$,

$$(1 - \xi^2)U_p'(\xi) = (p+1)U_{p-1}(\xi) - p\xi U_p(\xi),$$

$$(1 - \xi^2)U_p'(\xi) = (p+2)\xi U_p(\xi) - (p+1)U_{p+1}(\xi)$$

and

$$U_p^2(\xi) - \xi U_p(\xi)U_{p-1}(\xi) = \frac{1 + U_{2p}(\xi)}{2},$$

$$U_{p-1}^2(\xi) - \xi U_p(\xi)U_{p-1}(\xi) = \frac{1 - U_{2p}(\xi)}{2}.$$

THEOREM 1. *Let* μ *be an eigenvalue of* $\mathbf{F}[a, n]$ *associated with the eigenvector* $\mathbf{v} = (v_1, \dots, v_n)'$. *Then*

(a) $\xi = 1 - \mu/2$ *is a zero of the polynomial*

$$Q[a, n](\xi) = U_n(\xi) + (a - 2)U_{n-1}(\xi). \tag{11}$$

(b) *If* $\mathbf{v}$ *is normalized to unit length, its entries (up to a sign) are given by*

$$v_p = \frac{2 \sin p\theta}{\sqrt{2n + 1 - U_{2n}(\xi)}}, \qquad p = 1, \dots, n, \tag{12}$$

where θ *is defined by* $\xi = \cos\theta$.

Proof. Let $P[a, n](\xi)$ be the result of evaluating the characteristic polynomial of $\mathbf{F}[a, n]$ at $\mu = 2(1 - \xi)$, i.e.,

$$P[a, n](\xi) = \det\{\mathbf{F}[a, n] - 2(1 - \xi)\mathbf{I}_n\}. \tag{13}$$

Taking $a = 2$, a direct computation shows the recurrence $P[2, n](\xi) = 2\xi P[2, n-1](\xi) - P[2, n-2](\xi)$. Thus $P[2, n](\xi) = U_n(\xi)$. Expanding the determinant (13) on its last row, whose nth entry is $2\xi + (a - 2)$, we obtain

$$P[a, n](\xi) = \{2\xi + (a - 2)\}P[2, n - 1](\xi) - P[2, n - 2](\xi); \tag{14}$$

hence $P[a, n]$ coincides with $Q[a, n]$, as defined in (11).

Let $\mathbf{v} = (v_1, v_2, \dots, v_n)'$ be an eigenvector of $\mathbf{F}[a, n]$ with eigenvalue μ. Rearranging the equation $\mathbf{F}[a, n]\mathbf{v} = \mu\mathbf{v}$, we have

$$\begin{aligned} v_2 &= 2\xi v_1, \\ v_3 &= 2\xi v_2 - v_1, \\ &\vdots \\ v_n &= 2\xi v_{n-1} - v_{n-2}, \\ (2-a)v_n &= 2\xi v_n - v_{n-1}. \end{aligned} \tag{15}$$

Use of the three-point recurrence formula in (10) gives

$$v_p = U_{p-1}(\xi)v_1, \qquad p = 1, \dots, n-1. \tag{16}$$

We obtain v_1 from the normalization

$$1 = \sum_{p=1}^{n} v_p^2 = v_1^2 \sum_{p=0}^{n-1} U_p^2(\xi),$$

and using the Christoffel-Darboux formula

$$\sum_{p=0}^{n-1} U_p^2(\xi) = \tfrac{1}{2}[U_n'(\xi)U_{n-1}(\xi) - U_n(\xi)U_{n-1}'(\xi)],$$

together with the identities given above.

Finally,

$$v_1 = \frac{\pm 2\sin\theta}{\sqrt{2n+1-U_{2n}(\xi)}}.$$

and (12) follows from (16). The eigenvalue associated with $\mathbf{v}$ is

$$\mu = 2(1-\cos\theta), \tag{17}$$

where θ is such that

$$U_n(\cos\theta) = (2-a)U_{n-1}(\cos\theta). \quad \blacksquare \tag{18}$$

The eigenvalue λ of $(\det \mathbf{F}[a,n])\mathbf{F}[a,n]^{-1}$ corresponding to the eigenvalue μ of $\mathbf{F}[a,n]$ is

$$\lambda = \frac{(a-1)n+1}{\mu}. \tag{19}$$

6. EIGENSTRUCTURE OF B, C, AND $\tilde{\mathbf{B}}$

For $a = 1, 2, 3$, the roots of $Q[a,n](\xi) = U_n(\xi) + (a-2)U_{n-1}(\xi)$ can be expressed in closed form, and hence explicit expressions for eigenvalues and eigenvectors can be found.

THEOREM 2. *Suppose* $\mathbf{T}$ *is any one of the matrices* $\mathbf{B}$, $\mathbf{C}$, *and* $\tilde{\mathbf{B}}$. *The eigenvalues and eigenvectors of* $\mathbf{T}$ *can be ordered so that the matrix* $\mathbf{V} = (v_{pq})$ *of orthonormalized eigenvectors is symmetric. With these orderings, the entries* v_{pq} *of* $\mathbf{V}$ *are given by*

	λ_q	v_{pq}
$\mathbf{B}$	$\frac{1}{4}\sec^2 \frac{q}{2n+1}\pi$	$(-1)^{p+q+1} \frac{2}{\sqrt{2n+1}} \sin \frac{2pq}{2n+1}\pi$
$\mathbf{C}$	$\frac{n+1}{2}\left(1 - \cos\frac{q}{n+1}\pi\right)^{-1}$	$\frac{2}{\sqrt{2n+2}} \sin \frac{pq}{n+1}\pi$
$\tilde{\mathbf{B}}$	$\frac{2n+1}{2}\left(1 - \cos\frac{2q}{2n+1}\pi\right)^{-1}$	$\frac{2}{\sqrt{2n+1}} \sin \frac{2pq}{2n+1}\pi$

Proof. *Matrix* $\mathbf{B}$: The substitution $a = 1$ in (18) gives the equation $\sin(n+1)\theta - \sin n\theta = 0$, which, from the identity $\sin x - \sin y = 2\sin[(x-y)/2]\cos[(x+y)/2]$, is equivalent to

$$\sin\frac{\theta}{2}\cos\frac{(2n+1)\theta}{2} = 0.$$

The solutions to this equation are given by

$$\theta_q = \frac{2q-1}{2n+1}\pi, \qquad q = 1, \ldots, n,$$

and satisfy $U_{2n}(\theta_q) = 0$. Substitution gives the entries in the eigenvector $\mathbf{v}_q = (v_{1q}, \dots, v_{nq})'$

$$v_{pq} = \frac{2}{\sqrt{2n+1}} \sin \frac{p(2q-1)}{2n+1}\pi, \qquad p = 1, \dots, n, \tag{20}$$

and the eigenvalue is

$$\mu_q = 2\left(1 - \cos \frac{2q-1}{2n+1}\pi\right).$$

From (19), the corresponding eigenvalue of $\mathbf{B}$ is

$$\lambda_q = \frac{1}{2}\left(1 - \cos \frac{2q-1}{2n+1}\pi\right)^{-1}. \tag{21}$$

in concordance with the result (9).

The sequence $\{\lambda_q\}$ in (21) is decreasing. If it is reordered in increasing order, we obtain the first part of Theorem 2.

Let $q' = n + 1 - q$. Then

$$\begin{aligned} \sin \frac{p(2q-1)}{2n+1}\pi &= \sin\left(p\pi - \frac{2pq'}{2n+1}\pi\right) \\ &= -\cos p\pi \sin \frac{2pq'}{2n+1}\pi \\ &= (-1)^{p+1} \sin \frac{2pq'}{2n+1}\pi. \end{aligned}$$

Multiplying each vector $\mathbf{v}_q$ by the constant factor $(-1)^q$, we obtain the statement. An analogous computation gives the eigenvalues λ_q.

Matrix $\mathbf{C}$: For $a = 2$, we have $Q[2, n](\xi) = U_n(\xi)$, and the characteristic roots are the zeros of the polynomial $U_n(\xi)$, that is,

$$\theta_q = \frac{q}{n+1}\pi, \qquad q = 1, \dots, n.$$

As $U_{2n}(\xi_q) = -1$ for these values, the entries in $\mathbf{v}_q = (v_{1q}, \dots, v_{nq})'$ are

$$v_{pq} = \frac{2}{\sqrt{2n+2}} \sin \frac{pq}{n+1}\pi, \qquad p = 1, \dots, n, \tag{22}$$

and the eigenvalue is

$$\mu_q = 2\left(1 - \cos \frac{q}{n+1}\pi\right).$$

The corresponding eigenvalue of $\mathbf{C}$ is obtained from (19).

Matrix $\tilde{\mathbf{B}}$: For $a = 3$, (18) is equivalent to $\sin(n+1)\theta + \sin\theta = 0$. From the identity $\sin x + \sin y = 2\sin[(x+y)/2]\cos[(x-y)/2]$ follows the equation

$$\sin \frac{(2n+1)\theta}{2} \cos \frac{\theta}{2} = 0.$$

The solutions are

$$\theta_q = \frac{2q}{2n+1}\pi, \qquad q = 1, \dots, n.$$

For these values, $U_{2n}(\xi) = 0$; hence the entries in $\mathbf{v}_q = (v_{1q}, \dots, v_{nq})'$ are

$$v_{pq} = \frac{2}{\sqrt{2n+1}} \sin \frac{2pq}{2n+1}\pi, \qquad p = 1, \dots, n. \tag{23}$$

The eigenvalue is

$$\mu_q = 2\left(1 - \cos \frac{2q}{2n+1}\pi\right),$$

and the corresponding eigenvalues of $\tilde{\mathbf{B}}$ is

$$\lambda_q = \frac{2n+1}{2}\left(1 - \cos \frac{2q}{2n+1}\pi\right)^{-1}. \qquad \blacksquare$$

7. CONCLUSIONS

Some of the above eigenvalues and eigenvectors were also obtained by Anderson and Stephens (1997). Use of the family (8) gives a unified approach. As we have seen, this family contains, as particular cases, three matrices related to several statistical problems: goodness-of-fit tests, distance-based regression, the Brownian bridge, and continuous scaling. Finally, it is worthwhile studying the eigenstructure of **B**. Cuadras (1990) has suggested that the entries of the eigenvectors can be obtained, up to a sign change, by permutation of the entries of the first eigenvector. An algorithm has been constructed to show that this property is true for any n such that $2n + 1$ is a prime number, and has been used to verify the result for primes up to a very large number. The present proof is, however, quite complicated, and work is in progress on an easier proof.

REFERENCES

Anderson, T. W. 1971. *The Statistical Analysis of Time Series*, Wiley, New York.

Anderson, T. W. and Darling, D. A. 1952. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes, *Ann. Math. Statist.* 23:193–212.

Anderson, T. W. and Stephens, M. A. 1993. The modified Cramér–von Mises goodness-of-fit criterion for time series, *Sankhyā Ser. A* 55:357–369.

Anderson, T. W. and Stephens, M. A. 1997. The continuous and discrete Brownian bridges: Representations and applications, *Linear Algebra Appl.*, this issue.

Cord, M. S. and Sylvester, R. J. 1962. The property of cross-symmetry. *J. Soc. Indust. Appl. Math.* 10:632–637.

Cuadras, C. M. 1989. Distance analysis in discrimination and classification using both continuous and categorical variables, in *Statistical Data Analysis and Inference*, (Y. Doge, Ed.), Elsevier Science (North-Holland), pp. 459–473.

Cuadras, C. M. 1990. An eigenvector pattern arising in nonlinear regression, *Qüestiió* 14:89–95.

Cuadras, C. M. and Arenas, C. 1990. A distance based regression model for prediction with mixed data, *Comm. Statist. A Theory Methods* 19:2261–2279.

Cuadras, C. M. and Fortiana, J. 1993. Continuous metric scaling and prediction, in *Multivariate Analysis, Future Directions* 2, (C. M. Cuadras and C. R. Rao, Eds.), Elsevier Science (North-Holland), Amsterdam, pp. 47–66.

Cuadras, C. M. and Fortiana, J. 1994. Ascertaining the underlying distribution of a data set, in *Selected Topics on Stochastic Modelling* (R. Gutiérrez and M. J. Valderrama, Eds.), World Scientific, Singapore, pp. 223–230.

Cuadras, C. M. and Fortiana, J. 1995. A continuous metric scaling solution for a random variable, *J. Multivariate Anal.* 52:1–14.

Cuadras, C. M., Arenas, C., and Fortiana, J. 1996. Some computational aspects of a distance-based model for prediction, *Comm. Statist. B Simulation Comput.* 25(3):1–18.

Dall'Aglio, G., Kotz, S., and Salinetti, G. (Eds.). 1991. *Advances in Probability Distributions with Given Marginals*, Kluwer Academic.

Durbin, J. and Knott, M. 1972. Components of Cramér–Von Mises statistics. I, *J. Roy. Statist. Soc. Ser. B* 34:290–307.

Durbin, J. and Watson, G. S. 1950. Testing for serial correlation in least squares regression, I, *Biometrika* 37:409–428; reprinted in Kotz and Johnson (1992), pp. 237–259.

Durbin, J. and Watson, G. S. 1951. Testing for serial correlation in least squares regression, II, *Biometrika* 38:159–178; reprinted in Kotz and Johnson (1992), pp. 260–266.

Fan, K., Taussky, O., and Todd, J. 1955. Discrete analogs of inequalities of Wirtinger, *Monatsh. Math.* 59:73–90.

Frank, W. L. 1958. Computing eigenvalues of complex matrices by determinant evaluation and by methods of Danilevski and Wielandt, *J. Soc. Indust. Appl. Math.* 6:378–392.

Good, I. J. 1970. The inverse of a centrosymmetric matrix, *Technometrics* 12:925–928.

Kotz, S. and N. L. Johnson (Eds.). 1992. *Breakthroughs in Statistics. Vol. II: Methodology and Distribution*, Springer-Verlag, New York.

Longley, J. W. 1981. Least squares computations and the condition of the matrix, *Comm. Statist. B Simulation Comput.* 10:593–615.

Mardia, K. V., Kent, J. T., and Bibby, J. M. 1979. *Multivariate Analysis*, Academic, London.

Rutherford, D. E. 1947. Some continuant determinants arising in physics and chemistry—I, *Proc. Roy. Soc. Edinburgh Sect. A* LXII:229–236.

Rutherford, D. E. 1951. Some continuant determinants arising in Physics and Chemistry—II, *Proc. Roy. Soc Edinburgh Sect. A* LXIII:232–241.

Shintani, H. 1968. Direct solution of partial difference equations for a rectangle, *J. Sci. Hiroshima Univ. Ser. A-I* 32:17–53.

von Neumann, J. 1941. Distribution of the ratio of the mean square successive difference to the variance, *Ann. Math. Statist.* 12:367–395.

Received 12 October 1996; final manuscript accepted 7 March 1997