Journal of Multivariate Analysis **80**, 138–165 (2002)
doi:10.1006/jmva.2000.1976, available online at http://www.idealibrary.com on IDEAL®

Eigenstructures of Spatial Design Matrices

David J. Gorsich, Marc G. Genton, and Gilbert Strang

Massachusetts Institute of Technology
E-mail: gorsich@math.mit.edu, genton@math.mit.edu, gs@math.mit.edu

Received April 7, 2000; published online June 19, 2001

In estimating the variogram of a spatial stochastic process, we use a spatial design matrix. This matrix is the key to Matheron's variogram estimator. We show how the structure of the matrix for any dimension is based on the one-dimensional spatial design matrix, and we compute explicit eigenvalues and eigenvectors for all dimensions. This design matrix involves Kronecker products of second order finite difference matrices, with cosine eigenvectors and eigenvalues. Using the eigenvalues of the spatial design matrix, the statistics of Matheron's variogram estimator are determined. Finally, a small simulation study is performed.

AMS 2000 subject classifications: 62H11; 65F15.

Key words and phrases: discrete cosine transform; eigenvalue; eigenvector; kriging; Kronecker product; Matheron's estimator; variogram; spatial statistics.

1. INTRODUCTION

We study the structure of spatial design matrices that occur in several applications related to spatial statistics. Kriging is an established technique (Cressie [3], Wackernagel [22]) for estimating unknown values of a spatial stochastic process by a weighted average of known values. To determine these weights in an optimal linear spatial prediction, we estimate a variogram from the data. That empirical variogram comes directly from the spatial design matrix.

Consider a spatial stochastic process $\{Z(\mathbf{x}) : \mathbf{x} \in D\}$, where D is a fixed subset of $\mathbb{R}^d$. The process is assumed to be intrinsically stationary and isotropic: for all $\mathbf{x}$ and $\mathbf{x} + \mathbf{h}$ in D,

$$\mathrm{E}[Z(\mathbf{x})] = \mu = \text{constant},$$

$$\mathrm{Var}[Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})] = 2\gamma(\mathbf{h}),$$

$$\gamma(\mathbf{h}) = \gamma(\|\mathbf{h}\|).$$

0047-259X/01 $35.00

The variogram, 2γ, is typically estimated by Matheron's classical technique (Matheron [14]). His unbiased estimator, $2\hat{\gamma}_M$, is computed from values $Z(\mathbf{x}_i)$ at a discrete set of locations $\mathbf{x}_i \in D$ with separation $\mathbf{h}$,

$$2\hat{\gamma}_M(\mathbf{h}) = \frac{1}{N_{\mathbf{h}}} \sum_{N(\mathbf{h})} (Z(\mathbf{x}_i) - Z(\mathbf{x}_j))^2, \tag{1}$$

where $N(\mathbf{h}) = \{(\mathbf{x}_i, \mathbf{x}_j) \in D \mid \mathbf{x}_i - \mathbf{x}_j = \mathbf{h}\}$ and $N_{\mathbf{h}}$ is the cardinality of $N(\mathbf{h})$. Note that if data are irregularly spaced, "tolerance" regions around $\mathbf{h}$ are often used (see, e.g., Cressie [3]). Some distributional properties of the estimator (1) were first discussed by von Neumann *et al.* [21], Shah [16], and Davis and Borgman [4, 5].

To understand the properties and performance of Matheron's estimator, we start by expressing it as a quadratic form,

$$2\hat{\gamma}_M(\mathbf{h}) = \frac{\mathbf{z}^T A^{(d)}(n^d, h)\, \mathbf{z}}{N_{\mathbf{h}}}, \tag{2}$$

where it is assumed that the points lie on a uniform grid in $\mathbb{R}^d$, a hypercube. The spatial design matrix $A^{(d)}(n^d, h)$ is a difference matrix that depends on the distance $h = \|\mathbf{h}\|$ and the dimension d. (In d dimensions, the matrix combines differences in directions that share the same h.) The vector $\mathbf{z} = (Z(\mathbf{x}_1), ..., Z(\mathbf{x}_{n^d}))^T$ gives the data at n^d points in the domain D. The properties of Matheron's variogram estimator can be determined from the eigenstructure of the matrix $A^{(d)}(n^d, h)$.

A second application is spatial tests, or tests for generalized time series. Examples are Durbin–Watson tests (Durbin and Watson [6]) for nonzero lag autocorrelations and randomness. These tests have been generalized by Ali [1], Vinod [20], and Wallis [23]. Ali [2] generalized the Durbin–Watson tests using the matrices $A(n, h)$ for autocorrelations at lag h. This paper will primarily focus on the application to the variogram using Matheron's estimator (2).

In Section 2 we describe the structure of the matrix $A^{(1)}(n, h)$ in one dimension. Its eigenvalues and eigenvectors are found in Section 3. The matrix is permuted into a direct sum of second order finite difference matrices whose eigenvectors form the type 2 discrete cosine transform, DCT2. The general structure of the spatial design matrix for $\mathbb{R}^d$ is determined in Section 4. In Section 5, we show how the eigenstructure changes for grids in $\mathbb{R}^2$ considering only differences in non-diagonal (coordinate) directions. In Section 6, properties of the matrix with diagonal directions are studied. Then in Section 7, our results are combined to determine the properties of Matheron's variogram estimator. Finally, Section 8 discusses a small simulation study for illustration of these properties.

2. THE SPATIAL DESIGN MATRIX IN $\mathbb{R}^1$

In this section we introduce the one-dimensional spatial design matrix $A(n, h) = A^{(1)}(n, h)$. In Section 4 we will define the matrix for any dimension d.

The matrix comes directly from the quadratic form (2); see Genton [8]. For distance $h < n/2$ there will be $2h$ points affected by the boundary. Therefore the first h and last h diagonal entries of $A(n, h)$ will have unit weight. The rest have weight 2 because there are two neighbors for each of those interior points. All row sums are zero (so that constant $Z(\mathbf{x})$ makes no contribution). The matrix then has entries -1 along an upper and a lower diagonal, separated by h columns or rows from the main diagonal:

$$
A(n,h)=\left(\begin{array}{ccc|ccccc|ccc}
1 & & & -1 & & & & & & & \\
 & \ddots & & & \ddots & & & & & & \\
 & & 1 & & & \ddots & & & & & \\
\hline
-1 & & & 2 & & & \ddots & & & & \\
 & \ddots & & & \ddots & & & \ddots & & & \\
 & & \ddots & & & \ddots & & & \ddots & & \\
 & & & \ddots & & & \ddots & & & \ddots & \\
 & & & & \ddots & & & 2 & & & -1 \\
\hline
 & & & & & \ddots & & & 1 & & \\
 & & & & & & \ddots & & & \ddots & \\
 & & & & & & & -1 & & & 1
\end{array}\right)
\begin{matrix} h \\ \\ \\ n-2h \\ \\ \\ h \end{matrix}
$$

When $h = n/2$, the matrix has all ones along the diagonal; the upper and lower diagonals are at distance $n/2$. If $h > n/2$, the matrix has $2h - n$ zeros at the center of the main diagonal. The rest of the diagonal entries are 1. Again the diagonals with -1 are h columns above and below the center diagonal. The upper diagonal starts at column $h + 1$ and ends in column n, and the lower diagonal goes from column 1 to column $n - h$.

$A(n, h)$ is the key spatial design matrix. The matrices $A^{(d)}(n^d, h)$ will all be derived from Kronecker products of these one-dimensional design matrices. In practice, the spatial design matrix is not often used for $h \geqslant n/2$ (Journel and Huijbregts [12]), and our paper will focus on the case when $h < n/2$, as illustrated above.

3. EIGENSTRUCTURES IN $\mathbb{R}^1$

We start by describing the eigenstructure of the spatial design matrix in $\mathbb{R}^1$. This will later be used to prove several properties of Matheron's estimator. The W^TW factorization and the eigenstructure of $A(n, h)$ are particularly simple.

When $h < n/2$ the matrix has the form $A(n, h) = W^T(n, h)\, W(n, h)$ (Ali [2]) where $W(n, h)$ is a first order finite difference matrix. This $(n-h) \times n$ matrix has $W_{ii} = 1$ and $W_{i,\, i+h} = -1$. Then $A(n, h)$ has h zero eigenvalues (and we could add h zero rows to W). As an example with $n = 5$ and $h = 2$,

$$W(5, 2) = \begin{pmatrix} 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \end{pmatrix} \text{ and} \tag{3}$$

$$A(5, 2) = \begin{pmatrix} 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ -1 & 0 & 2 & 0 & -1 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \end{pmatrix}. \tag{4}$$

When $h = 1$, A is the standard tridiagonal second order difference matrix. Its eigenvectors yield the discrete cosine transform (DCT2) and its eigenvalues are $2 - 2\cos(k\pi/n)$, $0 \leqslant k < n$ (Strang [19]). What happens when h increases and the diagonals in A are separated?

The key is a permutation of the rows and columns of A. The permutation turns A into a direct sum of h smaller matrices, each one in the simple tridiagonal form (with separation $h = 1$). That form is completely understood. In our 5×5 example we take rows and columns in the order 1, 3, 5, 2, 4. This ordering jumps over the spaces between diagonals ($h = 2$ in this case) to produce a block tridiagonal matrix:

$$P^T A(5, 2)\, P = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}. \tag{5}$$

The corresponding W has the same 1, 3, 5, 2, 4 permutation applied to its columns:

$$W(5,2)\,P = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}. \tag{6}$$

Then $(WP)^T(WP) = P^T(W^TW)\,P = P^TAP$ as desired.

To describe this permutation for the general matrix $A(n, h)$, we express the order n as $ph+q$ with remainder $0 \leqslant q < h$. The permutation will produce q tridiagonal matrices of order $p+1$ and $h-q$ tridiagonal matrices of order p. (The example had $5 = 2(2)+1$, so there was one matrix of order $p+1=3$ and one of order $p=2$.) The permutation jumps h columns at a time, until the next column number would exceed n, and then restarts:

$$1,\, 1+h,\, ...,\, 1+ph,\, 2,\, 2+h,\, ...,\, 2+ph,\, ...,\, q,\, q+h,\, ...,\, q+ph,$$

$$q+1,\, q+1+h,\, ...,\, q+1+(p-1)\,h,\, ...,\, h,\, 2h,\, ...,\, (p-1)\,h.$$

There are q sets of $p+1$ columns in the first group and $h-q$ sets of p columns in the second group. This permutation of the rows and columns rearranges A into the direct sum of h tridiagonal submatrices:

$$P^TA(n,h)\,P = \left(\bigoplus_{i=1}^{q} A(p+1, 1)\right) \oplus \left(\bigoplus_{i=1}^{h-q} A(p, 1)\right) \tag{7}$$

The matrix WP splits into a similar combination of submatrices $W(p+1, 1)$ and $W(p, 1)$.

The eigenvalues of P^TAP (and of the original A) are $2-2\cos(k\pi/(p+1))$ each with multiplicity q, and $2-2\cos(k\pi/p)$ each with multiplicity $h-q$. The total count is $(p+1)\,q+p(h-q)=n$.

The eigenvectors of the matrix $A(n, h)$ of order n are the columns of the discrete cosine matrices of order $p+1$ or p, completed to length n by zeros. If h divides n then the eigenvector matrix is a direct sum of h type 2 discrete cosine transforms $C(p)$ of size $p=n/h$. These matrices are defined as $(C(p))_{ij} = \cos(\pi i(2j+1)/2p)$ (Strang [19]). If h does not divide n then the eigenvector matrix X of A is permuted by P into

$$P^TXP = \left(\bigoplus_{i=1}^{q} C(p+1)\right) \oplus \left(\bigoplus_{i=1}^{h-q} C(p)\right). \tag{8}$$

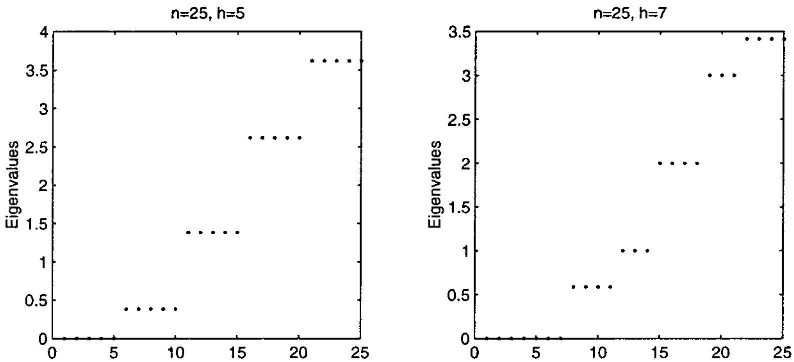


FIG. 1. Eigenvalues of the spatial design matrix in $\mathbb{R}^1$ with $n=25$ and distance $h=5$ on the left and $h=7$ on the right. On the left, 5 divides 25 and there is only one DCT and one set of repeated eigenvalues. On the right, 7 does not divide 25 and there are two sets of eigenvalues.

Figure 1 displays two plots of the eigenvalues for $n=25$ points on a line, with distances $h=5$ and $h=7$. Since $h=5$ divides 25, we have $q=0$ and the set of $p=5$ eigenvalues is repeated with multiplicity 5. For $h=7$, we have $25=3(7)+4$. There are two types of eigenvectors, and two sets of eigenvalues interwoven: $p+1=4$ eigenvalues with multiplicity $q=4$ and $p=3$ eigenvalues with multiplicity $h-q=3$.

For $\mathbb{R}^1$, this describes the exact structure of the matrix. The maximum eigenvalue is 4 for all h and n. This is the maximum of the quadratic form $\mathbf{z}^T A \mathbf{z}$ for unit vectors $\mathbf{z}$. As h increases, $A(n, h)$ has fewer distinct eigenvalues (and h zero eigenvalues). The eigenvalue multiplicities increase and the ability of the estimator to predict the variogram dramatically diminishes. This will be discussed more completely after we examine the eigenstructure for any dimension d.

4. THE SPATIAL DESIGN MATRIX IN $\mathbb{R}^d$

How does the form of the spatial design matrix $A^{(d)}(n^d, h)$ change for higher dimensions? The answer is found using Kronecker products. The spatial data are assumed to be located on a hypercube with each edge holding n points (a total of n^d points). Note it is straightforward to generalize the hypercube to have unequal edges (Genton [8]). On the hypercube, the separation distance (the lag) is given by $h=\sqrt{h_1^2+\cdots+h_d^2}$ where h_k is the distance along the kth axis. The distance vector is $\mathbf{h}=(h_1, \ldots, h_d)$.

We start with the non-diagonal case, when only one component h_l is non-zero. The spatial design matrix along the lth axis is

$$A_l^{(d)} = \left(\bigotimes_{k=1}^{l-1} I(n) \right) \otimes A(n, h_l) \otimes \left(\bigotimes_{k=l+1}^{d} I(n) \right). \tag{9}$$

This defines $A^{(d)}$ only in one direction. When the stochastic process is isotropic the variogram is not dependent on direction, only on distance. Then the matrix $A^{(d)}(n^d, h)$ must take into account all directions, and l ranges from 1 to d. The form of the spatial design matrix for non-diagonal directions (coordinate directions) of a fixed distance h is (Genton [8])

$$A^{(d)}(n^d, h) = \sum_{l=1}^{d} A_l^{(d)}, \tag{10}$$

This form is particularly nice, as will be seen later when its eigenvalues are determined.

In general there are more directions to be accounted for, and the matrix becomes more difficult to handle. For example, in $\mathbb{R}^2$ the distance $h = 5$ can be reached not only by $(5, 0)$ or $(0, 5)$, but also by $(3, 4)$ and $(4, 3)$. This introduces "diagonal" vectors $\mathbf{h}$ with two nonzero components. Fortunately these are also handled with Kronecker products. The simplest case is when all h_k are non-zero. Then the spatial design matrix along the diagonal $\mathbf{h} = (\pm h_1, ..., \pm h_d)$ is

$$\bigotimes_{k=1}^{d} D(n, h_k) + (-1)^{d-1} \bigotimes_{k=1}^{d} O(n, h_k), \tag{11}$$

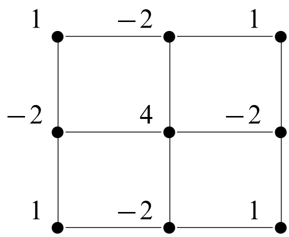
where $D(n, h)$ is the diagonal matrix of $A(n, h)$ and $O(n, h)$ is the off-diagonal matrix: $A(n, h) = D(n, h) + O(n, h)$.

There are no diagonal directions in one dimension (along a line). For larger d the matrix $\bigotimes_{k=1}^{d} D(n, h_k)$ gives the diagonal part, and the matrix $(-1)^{d-1} \bigotimes_{k=1}^{d} O(n, h_k)$ gives the off-diagonal part of $A^{(d)}(n^d, h)$. The off-diagonal part always consists of -1 entries and $(-1)^{d-1}$ keeps the sign correct. The diagonal entries represent the number of neighbors of a point in the directions $(\pm h_1, ..., \pm h_d)$.

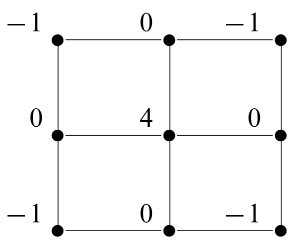
To see that Eq. (11) is the correct form for a given diagonal $\mathbf{h}$, start in $\mathbb{R}^2$. There $A(n, h_1) \otimes A(n, h_2)$ gives a typical finite difference mask (nine points without boundary and four or six points with boundary), taking differences on axis one, then taking differences of the prior in the next orthogonal direction. As an example consider $A(3, 1) \otimes A(3, 1)$. This gives a 9×9 matrix of finite differences:

$$A(3,1)\otimes A(3,1)=\begin{pmatrix} 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & 0\\ -1 & 2 & -1 & 1 & -2 & 1 & 0 & 0 & 0\\ 0 & -1 & 1 & 0 & 1 & -1 & 0 & 0 & 0\\ -1 & 1 & 0 & 2 & -2 & 0 & -1 & 1 & 0\\ 1 & -2 & 1 & -2 & 4 & -2 & 1 & -2 & 1\\ 0 & 1 & -1 & 0 & -2 & 2 & 0 & 1 & -1\\ 0 & 0 & 0 & -1 & 1 & 0 & 1 & -1 & 0\\ 0 & 0 & 0 & 1 & -2 & 1 & -1 & 2 & -1\\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 \end{pmatrix}.$$

The diagonal directions of interest are $\mathbf{h}=(\pm 1, \pm 1)$. Each row involves the use of $2^2=4$ or $2(3)=6$ or $3^2=9$ points. For the distance $\sqrt{2}$, there should only be one, two or four neighbors from each point. Consider the center point on the grid, $\mathbf{x}_5$, corresponding to the fifth row and column of the matrix. This is the only point with four neighbors. The finite difference mask given by row 5 of $A(3,1)\otimes A(3,1)$ is



This does not give differences of points a fixed distance h apart. The problem is that we are not interested in the horizontal and vertical directions (non-diagonal directions), $A(n,h_1)\otimes I(n)$ and $I(n)\otimes A(n,h_2)$. The correct finite difference mask is:



In general we want to remove the non-diagonal directions, given in Eq.(10). A simple subtraction from $A(n,h_1)\otimes A(n,h_2)$ does not work because after the first differences are taken, the Kronecker product multiplies the prior axes' weights to the next orthogonal direction. This is because $A(n, h_1) \otimes A(n, h_2) = (A(n, h_1) \otimes I(n))(I(n) \otimes A(n, h_2))$. Therefore we must also add these weights to $A(n, h_1) \otimes I(n)$ and $I(n) \otimes A(n, h_2)$. The weights are exactly contained in $D(n, h_2)$ and $D(n, h_1)$, giving the following matrices to be subtracted from $A(n, h_1) \otimes A(n, h_2)$:

$$A(n, h_1) \otimes D(n, h_2) + D(n, h_1) \otimes A(n, h_2). \tag{12}$$

By subtracting Eq. (12) from $A(n, h_1) \otimes A(n, h_2)$, the horizontal and vertical directions have been taken out leaving the diagonal directions, except the sign must be corrected so that the diagonal is again positive. This gives the following form for diagonal (non-coordinate) directions in $\mathbb{R}^2$:

$$A(n, h_1) \otimes D(n, h_2) + D(n, h_1) \otimes A(n, h_2) - A(n, h_1) \otimes A(n, h_2).$$

Expanding each $A(n, h)$ as $D(n, h) + O(n, h)$ we arrive at Eq. (11) for $d = 2$. For larger d the procedure is the same. All non-diagonal directions (coordinate directions) in Eq. (10) must be subtracted out with the correct weights, namely the $D(n, h)$ matrices, leaving only the diagonal directions as in Eq. (11).

Isotropy requires us to sum the additional directions to Eq. (11), which only gives one particular set of directions. There will generally be other diagonal directions that give the same distance h and have the property that all h_k are non-zero. To take into account all diagonal directions of a given distance h, the labels of the axes must be permuted. Taking into account all permutations the spatial design matrix for the diagonal directions of distance h is given by

$$\sum_{j=1}^{N} P_j^T \left(\bigotimes_{k=1}^{d} D(n, h_k) + (-1)^{d-1} \bigotimes_{k=1}^{d} O(n, h_k) \right) P_j, \tag{13}$$

where P_j is a permutation operator, and $N = d!/x_1!x_2! \cdots x_m!$ is the number of permutations of Eq. (11). The multinomial gives the number of distinct partitions of the components of $\mathbf{h}$, where m is the number of distinct non-zero components of $\mathbf{h}$, and x_k, $k = 1, \ldots, m$ is the number of times a component is repeated. This means $x_1 + x_2 + \cdots + x_m = d$. If all components have distinct values, then there are $d!/(1! \cdots 1!) = d!$ permutations. This gives all possible diagonal directions in $\mathbb{R}^d$ with distance h and components h_1 through h_d.

The most general (and most complicated) form is when r component directions are zero, $r < d$. Now identity matrices $I(n)$ are mixed into Eq. (13) for the component directions that are zero. Assuming isotropy, the identity matrices must be permuted in the same fashion as before, keeping only distinct permutations. This gives the following form for the most general spatial design matrix:

$$\sum_{j=1}^{MN} P_j^T \left(\left[\bigotimes_{k=1}^{d-r} D(n, h_k) + (-1)^{d-r-1} \bigotimes_{k=1}^{d-r} O(n, h_k) \right] \bigotimes_{i=1}^{r} I(n) \right) P_j. \quad (14)$$

This is because there are $N = (d-r)!/(x_1! \cdots x_m!)$ permutations of the non-zero components of $\mathbf{h}$ and $M = d!/(r!(d-r)!)$ permutations of the zero components with the prior. When $r = d-1$, Eq. (14) simplifies to Eq. (10).

We now give several examples of the spatial design matrix. Let $d = 4$, $n = 3$ and two of the components of $\mathbf{h}$ be zero so $r = 2$. Also assume the directions of interest are of length $\sqrt{2}$. This means there are two repeated components of $\mathbf{h}$, so that $2!/2! = 1$. There is only one unique permutation of $D(n, h_1) \otimes D(n, h_2)$, namely itself, because $D(n, h_1) \otimes D(n, h_2) = D(n, h_2) \otimes D(n, h_1)$. But, there are permutations with the identity matrix, $4!/2!2! = 6$. The permuted diagonal matrices to be summed are

$$\begin{gathered} I(n) \otimes I(n) \otimes D(n, 1) \otimes D(n, 1) \\ I(n) \otimes D(n, 1) \otimes D(n, 1) \otimes I(n) \\ I(n) \otimes D(n, 1) \otimes I(n) \otimes D(n, 1) \\ D(n, 1) \otimes D(n, 1) \otimes I(n) \otimes I(n) \\ D(n, 1) \otimes I(n) \otimes D(n, 1) \otimes I(n) \\ D(n, 1) \otimes I(n) \otimes I(n) \otimes D(n, 1). \end{gathered}$$

The off-diagonal matrices $O(n, h)$ are permuted in exactly the same way and subtracted. The sum gives all possible combinations in the $(\pm 1, \pm 1)$ directions, or $h = \sqrt{2}$ in $\mathbb{R}^4$. The first diagonal entry is 6 since the first diagonal entries of $D(n, h_1)$ and $D(n, h_2)$ are 1. This is exactly the number of neighbors for each corner of the hypercube.

Now assume we are interested in the distance $h = \sqrt{5}$. Only diagonal directions can reach this length, with $r = 2$. One component of $\mathbf{h}$ must be 2 and the other 1. The components of $\mathbf{h}$ are no longer repeated, and now $N = 2!/1!1! = 2$. This means the same permutations above are summed:

$$\begin{gathered} I(n) \otimes I(n) \otimes D(n, 1) \otimes D(n, 2) \\ I(n) \otimes D(n, 1) \otimes D(n, 2) \otimes I(n) \\ I(n) \otimes D(n, 1) \otimes I(n) \otimes D(n, 2) \\ D(n, 1) \otimes D(n, 2) \otimes I(n) \otimes I(n) \\ D(n, 1) \otimes I(n) \otimes D(n, 2) \otimes I(n) \\ D(n, 1) \otimes I(n) \otimes I(n) \otimes D(n, 2) \end{gathered}$$

But $D(n,1)\otimes D(n,2)\neq D(n,2)\otimes D(n,1)$ so there are another six distinct permutations:

$$I(n)\otimes I(n)\otimes D(n,2)\otimes D(n,1)$$

$$I(n)\otimes D(n,2)\otimes D(n,1)\otimes I(n)$$

$$I(n)\otimes D(n,2)\otimes I(n)\otimes D(n,1)$$

$$D(n,2)\otimes D(n,1)\otimes I(n)\otimes I(n)$$

$$D(n,2)\otimes I(n)\otimes D(n,1)\otimes I(n)$$

$$D(n,2)\otimes I(n)\otimes I(n)\otimes D(n,1)$$

Now the first diagonal entry is 12 instead of 6 to reflect that there are 12 neighbors in the diagonal distances of length $\sqrt{5}$.

The last concern as to the form of the spatial design matrix is when we wish to consider distances such as $\|h\|=5$ in dimension $d=4$. There are additional directions that give length 5 that are not accounted for in the permutations. The non-diagonal directions of the same distance need to be included, which means adding Eq. (10) with $d=4$. We assume this is added when the square root of the components is a natural number. There can also be other combinations of the h_k that can give the same distance and this matrix. The value of r may change, but the matrix still has the form given by Eq. (14). In our example with $h_1=3$, $h_2=4$ and $d=4$, the matrices to be summed are those for the non-diagonal directions (Eq. (10)) and those for the directions (3, 4, 0, 0) and (1, 2, 2, 4) (Eq. (14)).

All forms of the spatial design matrix have been described, and the next sections will explore the eigenstructure. We begin in $\mathbb{R}^2$ and look at the eigenstructure for diagonal and non-diagonal cases separately.

5. EIGENSTRUCTURES FOR $\mathbb{R}^d$ (NON-DIAGONAL DIRECTIONS)

We start by looking at the eigenstructure of the spatial design matrix that only considers non-diagonal directions. This matrix was defined in Eq. (10). Surprisingly, the eigenvalues do not change significantly for the $d=2$ case. The eigenvalues in the $\mathbb{R}^2$ case now range from 0 to 8 instead of 0 to 4, and follow a similar pattern having a structured set of multiplicities. In fact these eigenvalues are the sums of two similar finite difference matrices, each having the eigenvalues $2-2\cos(k\pi/p)$. The eigenvectors are products of discrete cosines.

This can be seen by recalling that for non-diagonal directions, $A^{(2)}(n^2, h)$ has the following form:

$$A^{(2)}(n^2, h) = I(n) \otimes A(n, h) + A(n, h) \otimes I(n). \tag{15}$$

Here $I(n) \otimes A(n, h)$ represents the horizontal direction and $A(n, h) \otimes I(n)$ represents the vertical direction. On a 3×3 grid with $h = 1$ we have a 9×9 matrix:

$$A^{(2)}(9, 1) = \begin{pmatrix} 2 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 3 & -1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 4 & -1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & -1 & 3 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & -1 & 3 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 2 \end{pmatrix}.$$

This matrix $A^{(2)}$ is equal to $I(3) \otimes A(3, 1) + A(3, 1) \otimes I(3)$.

The structure of $A^{(2)}(n^2, h)$ can be rewritten as

$$A^{(2)}(n^2, h) = \bigoplus_{i=1}^{n} A(n, h) + A(n^2, nh). \tag{16}$$

This is because the points on the grid are numbered left to right, top to bottom, so that the horizontal differences give n tridiagonal spatial design matrices of size n. In the vertical direction, the differences are at a distance of nh instead of h due to the numbering.

Since we know the eigenstructure of $A(n, h)$, we know the eigenstructure of the horizontal and vertical directions. The eigenvalues of the sum $A^{(2)}(n^2, h)$ are the eigenvalues of $A(n^2, nh)$ plus the same eigenvalues permuted.

PROPOSITION 5.1. *The eigenvalue matrix of $A^{(2)}(n^2, h)$ using only non-diagonal directions is $P^T \Lambda P + \Lambda$ where Λ is the eigenvalue matrix of $A(n^2, nh)$ and P is the permutation matrix that reorders the gridpoints by columns*: $P^T A(n^2, nh)\, P = \bigoplus_{i=1}^{n} A(n, h)$.

Proof. The eigenvalues of $I(n) \otimes A(n,h) + A(n,h) \otimes I(n)$ were noted by Steeb [17]. Because of the numbering of $\mathbf{z}$ we simply have $A(n,h) \otimes I(n) = A(n^2, nh)$. The permutation matrices were described by Graham [11]. ▮

More generally it is easy to show $A(n,h) \otimes (\bigotimes_{i=1}^{k} I(n)) = A(n^{k+1}, n^k h)$. The non-diagonal d-dimensional hypercube case follows directly from the $d=2$ case. From Eq. (10), the matrix $A^{(d)}(n^d, h)$ is a sum of a direct product of $d-1$ identity matrices of size n with a single one-dimensional matrix $A(n,h)$. When $d=2$ this simplifies to Eq. (15). The numbering of the spatial locations is assumed to continue lexicographically from the $d=2$. As i increases, the differences spread apart until we reach $A(n,h) \otimes I(n) \otimes \cdots \otimes I(n)$.

In this case, the exact eigenvalues can be found by using a generalization of Lancaster's theorem (Lancaster [13]) found by Searle and Henderson [15]. Using their results, we give the eigenvalues in the next lemma.

LEMMA 5.1. *Let Λ be the eigenvalue matrix of $A(n,h)$. Then the eigenvalue matrix of $A^{(d)}(n^d, h)$, using only non-diagonal directions of length h is*

$$\sum_{i=0}^{d-1} (P^T)^i \, (\Lambda \otimes I(n) \otimes \cdots \otimes I(n) \, P^i. \tag{17}$$

Each term in this sum represents repeated eigenvalues of $A(n,h)$ in order i with n^{d-1} times the multiplicity of the eigenvalues of Λ. For all n and h the maximum eigenvalue is $4d$.

The additional multiplicity is caused by the Kronecker products of the identity matrix with $A(n,h)$. This equation is the sum of every combination of the repeated eigenvalues for $A(n,h)$ giving $n^{d-1}(q(p+1) + p(h-q)) = n^d$ eigenvalues. Since the eigenvalues of $A(n,h)$ are $2 - 2\cos(k\pi/(p+1))$ of multiplicity q and $2 - 2\cos(k\pi/p)$ of multiplicity $h-q$, the eigenvalues of $A^{(d)}(n^d, h)$ range from 0 to $4d$. It is interesting to note that the eigenvalues of $A(n^d, n^{d-1}h)$ are the same as the eigenvalues of $A(n,h)$ but the multiplicity has a factor n^{d-1}.

6. EIGENSTRUCTURES IN $\mathbb{R}^d$ (WITH DIAGONAL DIRECTIONS)

We now admit diagonal directions for the spatial design matrix. The distance is $h = \sqrt{h_1^2 + \cdots + h_d^2}$ and the matrix $A^{(d)}(n^d, h)$ depends on all components of $\mathbf{h}$.

On a two-dimensional grid, there are many more directions and distances: $h=5$ can be formed by $h_1 = 3$ and $h_2 = 4$ or $h_1 = 5$ and $h_2 = 0$. In this case, the form of $A^{(2)}$ changes although the matrix can still be placed in the form of Kronecker products and additions. In Section 4 we found that for $\mathbf{h} = (h_1, h_2)$, the matrix $A^{(2)}(n^2, h)$ has the form

$$A^{(2)}(n^2, h) = D(n, h_1) \otimes D(n, h_2) - O(n, h_1) \otimes O(n, h_2). \tag{18}$$

Using this form we can find a bound on the eigenvalues, given in Lemma 6.1.

LEMMA 6.1. *Let λ_i be the eigenvalues of $A(n, h_1)$, μ_j the eigenvalues of $A(n, h_2)$, and $\lambda_{max} = \max\ (\lambda_i, \mu_j)$. Then the eigenvalues of $A^{(2)}(n^2, h)$ for a grid in $\mathbb{R}^2$ are bounded by*:

$$\begin{array}{ll} 2\,\lambda_{max} & \textit{if} \quad h_1 = h_2,\ h_1 = 0 \textit{ or } h_2 = 0 \\ 4\,\lambda_{max} & \textit{if} \quad h_1 \neq h_2 \textit{ and } h \textit{ is not an integer} \\ 4\,\lambda_{max} + 8 & \textit{otherwise.} \end{array}$$

Since $\lambda_{max} = 4$, these three upper bounds are 8, 16, *and* 24.

Proof. There are four different cases for the eigenvalues of the spatial design matrix. These cases are defined by the distance $h = \sqrt{h_1^2 + h_2^2}$ on a grid. The non-diagonal case is when h_1 or h_2 is zero. Here we know the exact eigenvalues (Lemma 5.1), and therefore the maximum. The next case is when $h_1 = h_2$ and h is not an integer. This is the diagonal case with only one permutation matrix to (h_2, h_1) in Eq. (13) so the matrix is defined as

$$A^{(2)}(n^2, h) = D(n, h_1) \otimes D(n, h_2) - O(n, h_1) \otimes O(n, h_2).$$

This relation can also be rewritten as

$$A(n, h_1) \otimes D(n, h_2) + D(n, h_1) \otimes A(n, h_2) - A(n, h_1) \otimes A(n, h_2). \tag{19}$$

We know the eigenstructure of $D(n, h)$ and $A(n, h)$, so we know the eigenstructure of each of the three terms. This gives an upper bound (but not the exact eigenvalues) of the sum. Because $D(n, h_2)$ and $A(n, h_1)$ are both positive definite we have

$$A(n, h_1) \otimes D(n, h_2) < 2A(n, h_1) \otimes I(n) \tag{20}$$

where 2 is the largest diagonal entry of $D(n, h_2)$. We also have $D(n, h_1) \otimes A(n, h_2) < 2I(n) \otimes A(n, h_2)$, so that $A^{(2)}(n^2, h)$ is dominated by

$$A(n, h_1) \otimes D(n, h_2) + D(n, h_1) \otimes A(n, h_2) - A(n, h_1) \otimes A(n, h_2).$$

Therefore the maximum eigenvalue of

$$\tilde{A} = 2A(n, h_1) \otimes I(n) + 2I(n) \otimes A(n, h_2) - A(n, h_1) \otimes A(n, h_2)$$

is never less than the maximum eigenvalue of $A^{(2)}(n^2, h)$. The eigenvalues of $\tilde{A}$ are known from Searle and Henderson [15] to be

$$2\lambda_i + 2\lambda_j - \lambda_i \lambda_j, \tag{21}$$

where λ_i is the ith eigenvalue of $A(n, h_1)$ and λ_j is the jth eigenvalue of $A(n, h_2)$. The maximum of Eq. (21) is $2\lambda_{max}$, where λ_{max} is the maximum of both sets of eigenvalues, λ_i and λ_j. This is because both sets always have 0 as an eigenvalue. Because the maximum eigenvalue of either matrix for any n is 4, and 0 is always an eigenvalue, the bound is 8. Eq. (21) has a saddle point in the center of its domain and extreme values of 8. This provides the bound for the diagonal case.

The third case is when $h_1 \neq h_2$ and the distance is non-integer, meaning diagonal directions only. The difference now is that there are two permutations in Equation (13), since $2!/1!1! = 2$, so there are two matrices. Since both are positive definite, the maximum eigenvalue of the sum is less than the sum of the maximum eigenvalues. Therefore the maximum eigenvalue is two times the second case, giving 16.

Finally the last case is when a third matrix needs to be added. It is also positive definite, representing the non-diagonal directions. An example is the distance 5. Then not only do diagonal directions (4, 3) and (3, 4) need to be added. The non-diagonal directions (5, 0) and (0, 5) must be included. Adding those matrices gives an additional 8 to the bound, setting a maximum bound of 24. This proves the Lemma. ∎

Figure 2 displays the tightness of the bound when $d = 2$ and $h_1 = h_2$. Distances are shown on the horizontal axis and the maximum eigenvalue is plotted on the vertical axis. The bound $2\lambda_{max}$ given by Lemma 6.1 is indicated by $+$. The actual eigenvalues are given by dots and are below the bound.

For increasing n, the bounds given in Fig. 2 are in fact tight. As n increases, the maximum eigenvalues go to 8 in two dimensions. This is seen in Fig. 3 for the direction $\sqrt{2}$. Can we find a similar bound for the $\mathbb{R}^d$ case with diagonal directions? The answer is yes. The matrix $A^{(d)}(n^d, h)$ can be written in a Kronecker product form just as was done in the $\mathbb{R}^2$ case (Eq. (21)).

THEOREM 6.1. *Let m be the number of distinct non-zero components in $(h_1, ..., h_{d-r})$ and x_k be the number of times that h_k is repeated, $k = 1, ..., m$. Let r be the number of components that are zero. Then the maximum eigenvalue,*

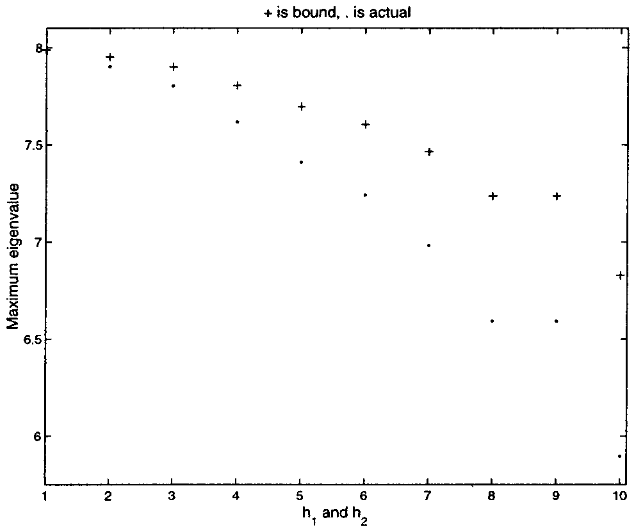


FIG. 2. Maximum eigenvalues of $A^{(2)}(40^2, \sqrt{2k^2})$ for the directions $h_1 = \pm k$ and $h_2 = \pm k$, $k = 1, ..., 10$. The actual maximum eigenvalues are given under each bound, and the bounds $2\lambda_{max}$ are labeled by $+$. The bound becomes worse as h increases.

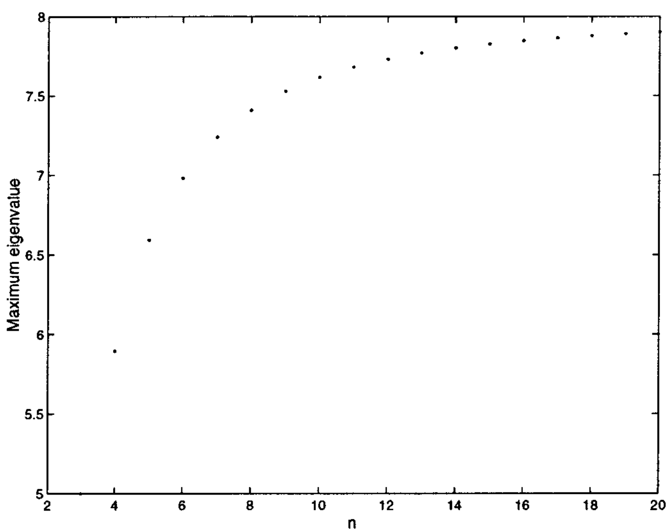


FIG. 3. The maximum eigenvalue of the spatial design matrix in $\mathbb{R}^2$ with the lag $h = \sqrt{2}$ given by directions $h_1 = \pm 1$ and $h_2 = \pm 1$. The maximum eigenvalue is asymptotic to 8 as n increases.

λ_{max}, of the spatial design matrix $A^{(d)}(n^d, h)$ using diagonal directions with non-integer distance h is bounded by

$$2^{d-r+1}\left(\frac{d!}{r!x_1!x_2!\cdots x_m!}\right). \tag{22}$$

If h is an integer and $d-r>1$, λ_{max} is bounded by

$$2^{d-r+1}\left(\frac{d!}{r!x_1!x_2!\cdots x_m!}\right)+4d. \tag{23}$$

Proof. To determine a bound, we approximate the spatial design matrix by another matrix that fits the theorem given by Searle and Henderson [15]. We start with the form given by (13). Assume $r=0$. When $d=1$ this reduces to the one-dimensional case with a bound of 4. Otherwise we have $d!/(x_1!x_2!\cdots x_m!)$ distinct permutations to consider. Consider just one permutation, $\bigotimes_{k=1}^{d} D(n, h_k)+(-1)^{d-1}\bigotimes_{k=1}^{d} O(n, h_k)$. Since these matrices are positive definite with real eigenvalues, we can write

$$\begin{aligned}\bigotimes_{k=1}^{d} D(n, h_k)+(-1)^{d-1}\bigotimes_{k=1}^{d} O(n, h_k)\\ <2^d I(n^d)+(-1)^{d-1}\bigotimes_{k=1}^{d} O(n, h_k).\end{aligned} \tag{24}$$

Writing the off-diagonal matrices as $A-D$ is one approach to reach a bound, but it involves many cross terms. A better approach is to determine the eigenvalues of the off-diagonal matrices, $O(n, h_k)$. These matrices simply have -1 along an upper and lower diagonal, h_k away from the main diagonal. Closely related to the eigenvalues of $A(n, h_k)$, we find the eigenvalues

$$-2\cos\left(\frac{(j+1)\pi}{p_k+1}\right), \qquad j=0, ..., p_k-1, \tag{25}$$

of multiplicity h_k-q_k, and

$$-2\cos\left(\frac{(j+1)\pi}{p_k+2}\right), \qquad j=0, ..., p_k, \tag{26}$$

of multiplicity q_k with $n=p_k h_k+q_k$. Using this result we can apply the generalized version of Lancaster's theorem and find the eigenvalues of the bounding matrix in Eq. (24),

$$2^d+2^d(-1)^{2d-1}\prod_{k=1}^{d}\cos(y_{k_j}), \tag{27}$$

where y_{k_j} is the jth eigenvalue of $O(n, h_k)$. The maximum occurs when the product of cosines is near -1. Then the bound is simply 2^{d+1}. For each permutation the same bound of 2^{d+1} applies. By the Cauchy–Schwartz inequality, the bound is therefore:

$$2^{d+1}\left(\frac{d!}{x_1!x_2!\cdots x_m!}\right). \tag{28}$$

When $r \neq 0$, there are also identity matrices to be permuted. They do not change the value of the eigenvalues, but because there are $d!/(r!(d-r)!)$ new matrices to be added, Cauchy–Schwartz gives the bound as:

$$2^{d+1}\left(\frac{d!}{r!x_1!x_2!\cdots x_m!}\right). \tag{29}$$

This is because $\mathbf{h}$ has $(d-r)!$ non-zero components so there are $(d-r)!$ identity matrices to permute. Similarly using Lemma 5.1 and the Cauchy–Schwartz inequality, we add $4d$ if h is an integer and $d-r>1$. If $d-r=1$ then the problem reduces to the non-diagonal case and is completely explained by Lemma 5.1. Therefore the bound is proved. ▮

7. MATHERON'S VARIOGRAM ESTIMATOR

Matheron's estimator (2) computes an empirical variogram by using the spatial design matrix. The key properties of the estimator can be described as functions of $A^{(d)}(n^d, h)$ and the actual covariance matrix of the data, $\mathrm{Var}(\mathbf{z}) = \Sigma$. The process Z is assumed to have zero mean and a Gaussian distribution. The latter assumption can be relaxed to elliptically contoured distributions (Genton [9]) or skew-normal distributions (Genton *et al.* [10]). The expectation, variance, and covariance of $2\hat{\gamma}_M(h)$ are given by the trace (sum of the eigenvalues) (Cressie [3], Genton [8]):

$$\mathrm{E}(2\hat{\gamma}_M(h)) = \frac{\mathrm{tr}(A^{(d)}(n^d, h)\,\Sigma)}{N_h} \tag{30}$$

$$\mathrm{Var}(2\hat{\gamma}_M(h)) = 2\,\frac{\mathrm{tr}(A^{(d)}(n^d, h)\,\Sigma A^{(d)}(n^d, h)\,\Sigma)}{N_h^2} \tag{31}$$

$$\mathrm{Cov}(2\hat{\gamma}_M(h_a), 2\hat{\gamma}_M(h_b)) = 2\,\frac{\mathrm{tr}(A^{(d)}(n^d, h_a)\,\Sigma A^{(d)}(n^d, h_b)\,\Sigma)}{N_{h_a}N_{h_b}}. \tag{32}$$

If the covariance goes to zero, then $\Sigma \to \sigma^2 I(n^d)$, and the expressions simplify. They are used for fitting a valid parametric variogram model to variogram estimates by generalized least squares. (Furrer and Genton [7], Genton [8, 9]). This simplification is based on knowing all the eigenvalues and is described in the next lemma.

LEMMA 7.1. *Let $\Sigma = \sigma^2 I(n^d)$ and suppose that $A^{(d)}$ in Matheron's estimator for Gaussian data involves only non-diagonal coordinate directions on a hypercube of n^d points in $\mathbb{R}^d$. For $h < n/2$ the number of differences $\|\mathbf{x}_i - \mathbf{x}_j\| = h$ is $N_h = dn^{d-1}(n-h)$. The estimator's properties are given by*

$$\mathrm{E}(2\hat{\gamma}_M(h)) = \frac{2\sigma^2 dn^{d-1}(n-h)}{N_h}$$

$$\mathrm{Var}(2\hat{\gamma}_M(h)) = 2\sigma^4 \frac{n^{d-1}d(6n-8h) + d(d-1)\, n^{d-2}(2n-2h)^2}{N_h^2}$$

$$\mathrm{Cov}(2\hat{\gamma}_M(h_a), 2\hat{\gamma}_M(h_b)) = 2\sigma^4 \left(\frac{2dn^{d-1}(2n - h_a - 2h_b)}{N_{h_a} N_{h_b}} + \frac{4d(d-1)\, n^{d-2}(n-h_a)(n-h_b)}{N_{h_a} N_{h_b}} \right) \text{ for } h_a < h_b$$

Proof. The value of N_h in $\mathbb{R}^1$ is just $n-h$, the number of differences of distance h. In $\mathbb{R}^2$ there are n differences in the horizontal and vertical directions giving $2n(n-h)$ differences. Continuing in this way, there are $N_h = dn^{d-1}(n-h)$ differences of size h in $\mathbb{R}^d$. Of course N_h would not have such a simple form if diagonal directions were included. Given this value of N_h, the expectation is $\mathrm{E}(2\hat{\gamma}_M(h)) = 2\sigma^2$, reflecting that Matheron's estimator is unbiased (Cressie [3]).

To prove that $\mathrm{tr}(A^{(d)}(n^d, h)) = 2\sigma^2 dn^{d-1}(n-h)$ we want to add the eigenvalues λ_i of $A^{(d)}(n^d, h)$. From Lemma 5.1 those eigenvalues are $\sum_{i=0}^{d-1} (P^T)^i \, (\Lambda \otimes (\bigotimes_{k=1}^{d-1} I(n)))\, P^i$. Therefore from the properties of the trace we have

$$\mathrm{tr}\left(\sum_{i=0}^{d-1} (P^T)^i \left(\Lambda \otimes \left(\bigotimes_{k=1}^{d-1} I(n) \right) \right) P^i \right) = dn^{d-1}\, tr(\Lambda), \tag{33}$$

where Λ is the eigenvalue matrix of $A(n, h)$. The trace (sum of the diagonal entries) of $A(n, h)$ is $2n - 2h$. Thus the sum of the eigenvalues of $A^{(d)}(n^d, h)$ is $2dn^{d-1}(n-h)$, and the expectation of Matheron's estimator is $2\sigma^2 dn^{d-1}(n-h)/N_h$.

The variance of Matheron's estimator for $\Sigma = \sigma^2 I(n^d)$ is found the same way, by summing the squared eigenvalues:

$$\operatorname{tr}\left(\left[\sum_{i=1}^{d} \bigotimes_{j=1}^{d} (A(n,h))^{\delta(i-j)}\right]^2\right). \tag{34}$$

Using the properties of the trace and Kronecker product, the sum comes outside. Since the trace of $\bigotimes_{j=1}^{d} A(n,h)^{\delta(i-j)}$ is equal to the trace of $\bigotimes_{j=1}^{d} A(n,h)^{\delta(k-j)}$, we can rewrite the trace in (34) as

$$dn^{d-1} \operatorname{tr}(A^2) + d(d-1)\, n^{d-2} (\operatorname{tr}(A))^2. \tag{35}$$

Recall that the trace of A is $2(n-h)$. Since A is symmetric, the trace of A^2 is the sum of $6(n-2h)$, the diagonal entries unaffected by the boundary and $2(2h)$, the top and bottom diagonal entries, which is simply $6n-8h$. Therefore the trace in (34) is

$$dn^{d-1}(6n-8h) + d(d-1)\, n^{d-2}(2n-2h)^2. \tag{36}$$

The form for the covariance follows the same reasoning, but the trace of $A(n, h_a)\, A(n, h_b)$ is needed. This trace has been found in Genton [8] as

$$\frac{4(2n-h_a-2h_b)}{(n-h_a)(n-h_b)}, \tag{37}$$

where $h_a < h_b$ and $h_a + h_b < n$. Now the entire lemma is proved. ∎

What do these results tell us about the performance of the estimator as d and h vary? Keeping n fixed, we plot the variance for increasing d and h on two axes. The third axis is the logarithm of the variance. Figure 4 indicates that the performance of the estimator dramatically improves as d increases. The largest increases in performance occur as soon as d goes beyond 1.

As h increases, the variance increases linearly. This is true for all d. Experimentally we find that this increase in performance as d increases happens for only some of the Σ tested. Although we have not proven the exact form of the variance for different d and h for general Σ, it appears that a similar result might be possible.

Example. The covariogram $(1-h)^j$ corresponds to $\Sigma_{ik} = (1 - \|\mathbf{x}_i - \mathbf{x}_k\|)^j$. This is a valid covariance function in $\mathbb{R}^{(2j-1)}$ with $0 \leqslant h \leqslant 1$. The variogram is $1-(1-h)^j$. As j increases, the covariogram approaches a delta function. For $d=1$ we plot the variance of the estimator for

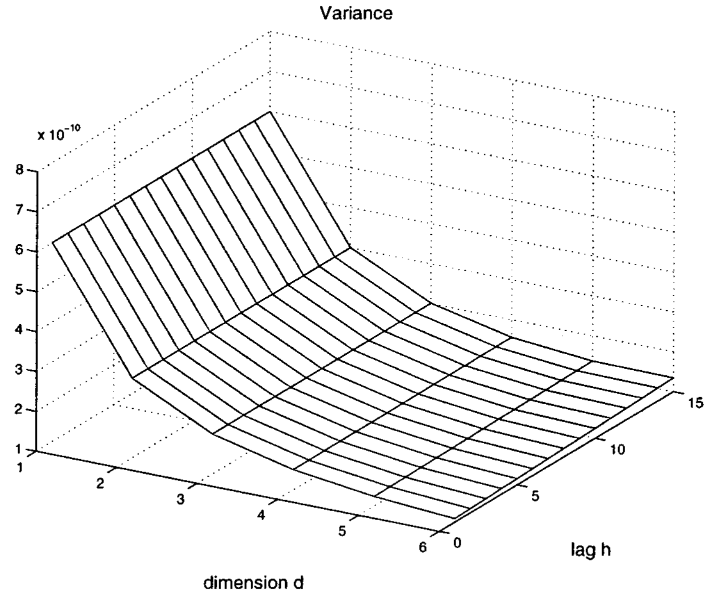


FIG. 4. The variance of Matheron's estimator keeping n fixed and $\Sigma = I$. The first axis is the dimension d going from 1 to 6; the second axis is the lag h going from 1 to 15. The number of spatial points is fixed at 10^{10}, so when $d = 6$ the hypercube has about 46 points along each axis. The variance of Matheron's estimator decreases exponentially when d increases, and increases linearly with h.

$j = 1, ..., 100$ and $h = 1, ..., 20$. There are 20 plots in Fig. 5, with the variance strictly increasing for each h. The variance of the estimator tends to be larger for covariances other than $\sigma^2 I(n^d)$ as h increases. It is interesting that the peak variance is at $j = 11$ when $h = 20$. When $h = 45$ to $h = 49$ the peak variance is caused by the covariance $(1-h)^4$.

For $h \leqslant 3$, the variance increases as the covariance goes to a delta. This means the variance of the estimator could be bounded by Lemma 7.1. But for larger h, a new peak variance occurs for some j. Figure 6 shows the value of j that gives the maximum variance for a given h. As h increases, the peak variance is reached for smaller j.

Lemma 7.1 only applies to the non-diagonal variogram estimator. How do the properties of the estimator change when diagonal directions $\mathbf{h}$ are also used in the spatial design matrix? For the expectation of $2\hat{\gamma}_M(h)$ we want the trace of Eq. (11). Using the properties of the trace, this is equal to

$$\prod_{k=1}^{d} \operatorname{tr}(D(n, h_k)) - \prod_{k=1}^{d} (\operatorname{tr}(O(n, h_k)). \tag{38}$$

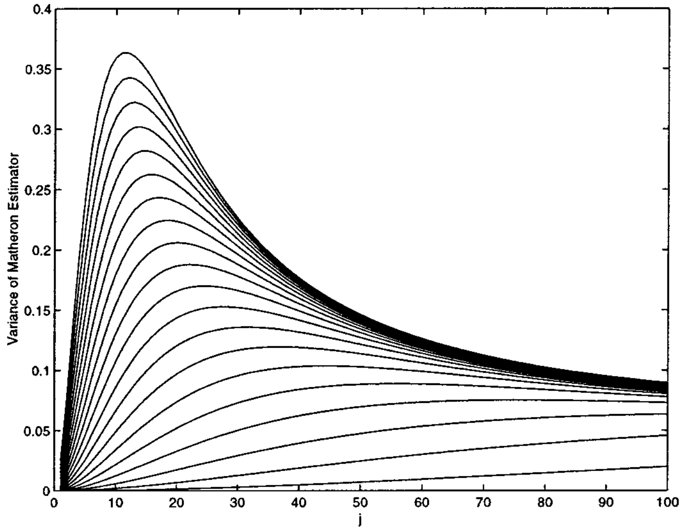


FIG. 5. The variance of Matheron's estimator (31) as the covariance $(1-h)^j$ goes from $1-h$ toward a delta function (j is on the horizontal axis). The last covariance entry is $(1-h)^{100}$. There are 20 plots displaying the variance for $h=1, ..., 20$. The variance increases with h.

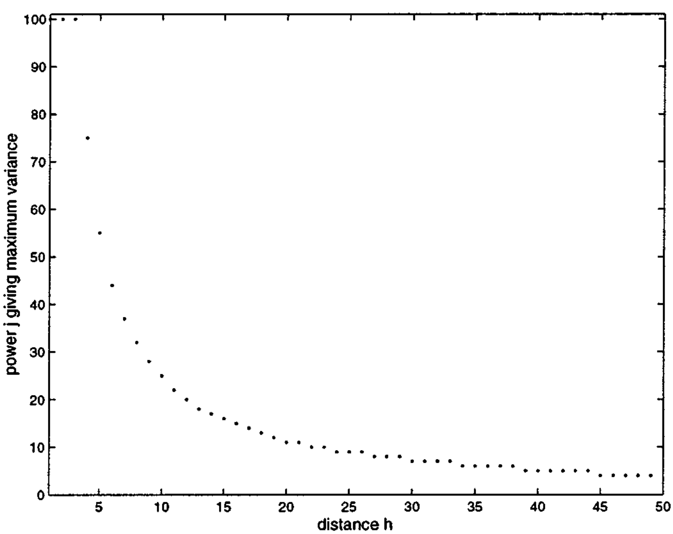


FIG. 6. The covariance of Z, $(1-h)^j$, parameterized by j giving the largest variance of Matheron's estimator (31) for a given distance h. For small h, the largest variance is given by covariances close to the delta functions, but for large h, the estimator's variance peaks at a covariance far from the delta function.

The trace of $O(n, h_k)$ is zero, so

$$\operatorname{tr}(A^{(d)}(n^d, h)) = \prod_{k=1}^{d} (2n - 2h_k). \tag{39}$$

Because the estimator is unbiased, the expectation must be 2. Therefore we know for a diagonal direction that the number of differences is

$$N_h = 2^{d-1} \prod_{k=1}^{d} (n - h_k). \tag{40}$$

These results yield the variance of the estimator when diagonal directions are used.

LEMMA 7.2. *Let $h = \sqrt{h_1^2 + \cdots + h_d^2}$ with $0 < h_i < n/2$ for all components. If $\Sigma = \sigma^2 I(n^d)$ then the variance of the estimator $\mathbf{z}^T A^{(d)}(n^d, h)\, \mathbf{z}/N_{\mathbf{h}}$ using only one diagonal direction is given by*

$$\sigma^4 \frac{\prod_{k=1}^{d} (4n - 6h_k) + \prod_{k=1}^{d} (2h_k - 2n)}{2^{2d-2} \prod_{k=1}^{d} (n - h_k)^2}. \tag{41}$$

Proof. Recall that the matrix $A^{(d)}(n^d, h)$ for a diagonal direction can be represented by the diagonal and off-diagonal matrices of the individual components of $\mathbf{h}$ as in Eq. (11). The variance of the estimator assuming $\Sigma = \sigma^2 I(n^d)$ is determined by the trace:

$$\operatorname{Var}(\mathbf{z}^T A^{(d)}(n^d, h)\, \mathbf{z}/N_{\mathbf{h}}) = \sigma^4 \operatorname{tr}([A^{(d)}(n^d, h)]^2)/N_{\mathbf{h}}^2. \tag{42}$$

Using Eq. (11) the trace of the squared matrix is equal to

$$\prod_{k=1}^{d} (4n - 6h_k) + \prod_{k=1}^{d} \operatorname{tr}(O(n, h_k)^2), \tag{43}$$

because the main diagonal of $O(n, h_k)$ is entirely zero. This means the cross terms $D(n, h_k)\, O(n, h_k)$ also have trace zero. Earlier we found the trace of the diagonal matrices to be $2n - 2h_k$. The trace of $O(n, h_k)^2$ is unlikely to be zero. To find this trace we recall the eigenvalues of $O(n, h_k)$ from Theorem 6.1:

$$-2\cos\left(\frac{(j+1)\pi}{p_k+1}\right) \text{ of multiplicity } h_k - q_k \text{ where } j = 0, ..., p_k - 1 \tag{44}$$

$$-2\cos\left(\frac{(j+1)\pi}{p_k+2}\right) \text{ of multiplicity } q_k \text{ where } j = 0, ..., p_k. \tag{45}$$

The sum squared is therefore

$$2h_k - 2n, \tag{46}$$

the negative of the diagonal $D(n, h_k)$ sum. Therefore the trace of $[A^{(d)}(n^d, h)]^2$ using a diagonal direction $\mathbf{h}$ is given by

$$\prod_{k=1}^{d} (4n - 6h_k) + \prod_{k=1}^{d} (2h_k - 2n). \tag{47}$$

Using this trace, the lemma is now proved. ∎

Another property that can be found using the eigenvalues is the maximum value of the normalized form of Matheron's estimator:

$$\frac{n^d \mathbf{z}^T A^{(d)}(n^d, h)\, \mathbf{z}}{N_h \mathbf{z}^T \mathbf{z}}. \tag{48}$$

The normalization constant is in fact $c(0)$, the covariogram at lag 0. Suppose that the spatial process is second order stationary, so that

$$\gamma(h) = c(0) - c(h). \tag{49}$$

The maximum eigenvalue of $A^{(d)}(n^d, h)$ is the upper bound of the normalized quadratic form. Moreover, there are well established bounds on normalized covariograms that depend on the dimension $c(\|\mathbf{h}\|)$ is valid in (Stein [18], Yaglom [24, 25]). If it is assumed that the covariogram is isotropic, then it can be represented as a $\mathbb{R}^1$ function $c(h)$, representing the radial function in $\mathbb{R}^d$. All covariograms valid in $\mathbb{R}^1$ are bounded by $-1 \leqslant c(h)/c(0) \leqslant 1$. For covariograms valid in $\mathbb{R}^2$, the bound is $-0.403 \leqslant c(h)/c(0)$ (Stein [18]). For covariograms valid in any dimension, they must be positive and convex, $0 \leqslant c(h)/c(0)$. This then sets a bound on $\gamma(h)/c(0)$. For example, for $\mathbb{R}^1$ the bound on $\gamma(h)$ is $0 \leqslant \gamma(h) \leqslant 2$.

What is the bound on Matheron's estimator? It turns out to be set by the maximum eigenvalue of Eq. (48). In $\mathbb{R}^1$ this maximum is

$$4n/(n-h). \tag{50}$$

This equation represents $2\gamma(h)$ so that for large n the bound goes to 2. For smaller n, the estimator loses its ability to match the possible range of actual $\gamma(h)$. For covariograms valid in higher d, the bound on Matheron's estimator does not change, and the estimates can be much larger than the variograms. For variograms valid in all dimensions, the bound on the variogram is $0 \leqslant \gamma(h) \leqslant 1$ but Matheron's estimator could still take values as high as 2.

In general, the bound for Matheron's estimator considering only non-diagonal directions is given for all dimensions by

$$\frac{2\,dn^d}{dn^{d-1}(n-h)} = \frac{2n}{(n-h)}. \tag{51}$$

From the eigenstructure of the spatial design matrix, we have found simple forms for the expectation, variance, covariance and maximum value of Matheron's estimator when $\Sigma = \sigma^2 I$. It is seen that the performance of the estimator decreases as h increases linearly, but exponentially increases as d increases for diagonal Σ. Therefore we would expect the estimator in a simulation to do much worse at approximating the variogram in $\mathbb{R}^1$ than in $\mathbb{R}^2$. In the next section we simulate an isotropic stochastic process for both $d=1$ and $d=2$ and determine the variance of the estimator for a given variogram, using this spatial design matrix. We find that higher dimensions yield an improved estimator for nearly diagonal or diagonal Σ, but in general, higher dimensions prove detrimental to the variance of the estimator.

8. SIMULATION STUDY

To illustrate these results, we perform a simulation estimating variograms from spatial processes in different dimensions. We choose a variogram that is valid in all dimensions tested. We keep n constant while increasing d, and see that the performance of Matheron's estimator seriously degrades. The spacing of the grid points, n in each direction, never changes. To run the simulation in a reasonable time, we test only $d=1$ and $d=2$.

The stochastic process is simulated with the spherical variogram $3h/10 - h^3/250$ for $0 \leqslant h \leqslant 5$, which is valid in both $\mathbb{R}^1$ and $\mathbb{R}^2$. The variogram has a range of 5, a sill of 1 and a nugget of 0. For $h \geqslant 5$ the variogram is taken to be just 1. The number of points is kept fixed at $n = 256$ so in $\mathbb{R}^2$ only a 16×16 grid is used.

Splus is used to perform the simulation using the function `rfsim` of the module S + SpatialStats. This function is based on a Choleski decomposition of the covariance matrix of the process and the results below pertain to this particular technique. For both $d=1$ and $d=2$, 200 simulations were performed using Matheron's estimator. For $d=1$ we find the variance at distance 1 to be 8.65×10^{-4} but for $d=2$ at the same distance we get 1.01×10^{-3}, larger than the one-dimensional case. The theoretical variance of the estimator is 1.53×10^{-3} in $\mathbb{R}^1$ and 1.84×10^{-3} for $\mathbb{R}^2$. It is interesting to notice that the variance was worse in $\mathbb{R}^2$. From our results in Lemma 7.2

and Fig. 6, we would have expected the variance to go down as d increases. But notice that the lemma is only for $\Sigma = \sigma^2 I(n^d)$. In fact, over most simulations as Σ goes away from $\sigma^2 I(n^d)$, the variance decreases in $\mathbb{R}^1$ and increases for $d > 1$. As an example, Fig. 7 shows the $d = 1$ and $d = 2$ cases for the spherical variogram with the range varying from 0.5 to 6 in increments of 0.5. The line with dots is the variance for $d = 1$ and the line with plus marks is the $d = 2$ graph. The $d = 2$ graph is smaller when the variance-covariance of $\mathbf{z}$ is essentially white noise, but decreases slower than the $d = 1$ case beyond a certain range (in this case three). Note that the number of data points and the spacing between the points is identical in both cases.

Since the grid is from 1 to 16 in increments of 1, the first two variances for $d = 1$ (and $d = 2$) are the same. A spherical variogram, with a range of 0.5 or 1, is essentially the same as a white noise process. In general, Matheron's estimator seems to perform worse as d increases and when Σ is not a diagonal matrix. In the diagonal case, the estimator does better in higher dimensions because N_h increases dramatically. There are many more combinations to take to reduce the variance of the estimator. When Σ diverges from a diagonal, the number of combinations does not change, but they are no longer independent, and information is lost.

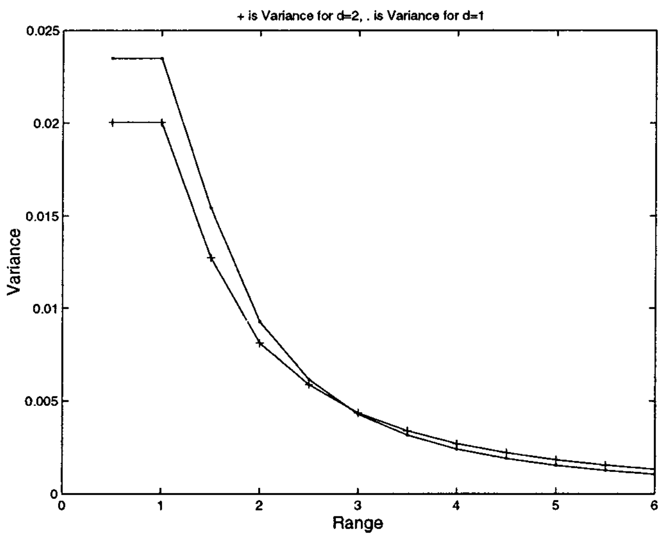


FIG. 7. Plots of the variance of Matheron's estimator for $d = 1$ (the dotted line) and $d = 2$ (the "+" line) as the range of the underlying variogram of the data increases. The variogram of the data is the spherical variogram with a nugget of 0 and a sill of 1. The estimator does better at estimating white noise in dimension two, but loses its performance when the range is three or larger.

9. CONCLUSIONS

This paper describes the simple and striking eigenstructure of the spatial design matrix that is used in Matheron's estimator of the variogram. This eigenstructure is related to finite differences and discrete cosine transforms. Moreover, this structure can be extended to arbitrary dimensions with the use of Kronecker products. Using the eigenstructure, we prove the estimator's properties for $\Sigma = \sigma^2 I(n^d)$. When Σ diverges from I, the effect of increasing dimension goes from positive to negative. This again is a property of the eigenvalues when $A^{(d)}(n^d, h)$ is multiplied by Σ. A next step would be to determine a general form for the eigenvalues when Σ is not diagonal.

REFERENCES

1. M. Ali, An approximation to the null distribution and power of the Durbin–Watson statistic, *Biometrika* **71** (1984), 253–261.
2. M. Ali, Durbin–Watson and generalized Durbin–Watson tests for autocorrelations and randomness, *J. Amer. Statist. Assoc.* **5**, No. 2 (1987), 195–203.
3. N. Cressie, "Statistics for Spatial Data," Wiley, New York, 1993.
4. B. M. Davis and L. E. Borgman, Some exact sampling distributions for variogram estimators, *Math. Geol.* **11** (1979), 643–653.
5. B. M. Davis and L. E. Borgman, A note on the asymptotic distribution of the sample variogram, *Math. Geol.* **14** (1982), 189–193.
6. J. Durbin and G. S. Watson, Testing for serial correlation in least squares regression, *Biometrika* **37** (1950), 409–428.
7. R. Furrer and M. G. Genton, Robust spatial data analysis of Lake Geneva sediments with S + SpatialStats, *Systems Res. Inform. Systems* **8** (1999), 257–272.
8. M. G. Genton, Variogram fitting by generalized least squares using an explicit formula for the covariance structure, *Math. Geology* **30** (1998), 323–345.
9. M. G. Genton, The correlation structure of Matheron's classical variogram estimator under elliptically contoured distributions, *Math. Geol.* **32** (2000), 127–137.
10. M. G. Genton, L. He, and X. Liu, Moments of skew-normal random vectors and their quadratic forms, *Statist. and Probab. Lett.* **51**, No. 4 (2001), 319–325.
11. A. Graham, "Kronecker Products and Matrix Calculus with Applications," Ellis Horwood, Chichester, 1981.
12. A. G. Journel and C. J. Huijbregts, "Miming Geostatistics," Academic Press, London, 1978.
13. P. M. Lancaster, "Theory of Matrices," Academic Press, New York, 1969.
14. G. Matheron, "Traité de Géostatistique Appliquée, Tome I," Mémoires du Bureau de Recherches Géologiques et Minières, No. 14, Editions Technip, Paris, 1962.
15. S. R. Searle and H. V. Henderson, Dispersion matrices for variance components models, *J. Amer. Statist. Assoc.* **74**, No. 366 (1979), 465–470.
16. B. K. Shah, Distribution of half-mean square differences, *Biometrika* **57** (1970), 193–198.
17. H. Steeb, "Matrix Calculus and Kronecker Product with Applications and C++ Programs," World Scientific, Singapore, 1997.
18. M. L. Stein, "Interpolation of Spatial Data," Springer-Verlag, Berlin, 1999.

19. G. Strang, The discrete cosine transform, *SIAM Rev*. **41** (1999), 135–147.
20. H. D. Vinod, Generalization of the Durbin–Watson statistic for higher order autoregressive processes, *Comm. Statist*. **2** (1973), 115–144.
21. J. von Neumann, R. H. Kent, H. R. Bellinson, and B. I. Hart, The mean square successive differences, *Ann. of Math. Stat*. **12** (1941), 153–162.
22. H. Wackernagel, "Multivariate Geostatistics," 2nd ed., Springer-Verlag, Berlin/New York, 1998.
23. K. F. Wallis, Testing for fourth-order autocorrelation in quarterly regression equations, *Econometrica* **40** (1972), 617–636.
24. A. M. Yaglom, "Correlations Theory of Stationary and Related Random Functions. I. Basic Results," Springer Series in Statistics, Springer-Verlag, Berlin/New York, 1987.
25. A. M. Yaglom, "Correlation Theory of Stationary and Related Random Functions. II. Supplementary Notes and References," Springer Series in Statistics, Springer-Verlag, Berlin/New York, 1987.