Statistics and Computing **14**: 99–108, 2004

On the discretization of nonparametric isotropic covariogram estimators

DAVID J. GORSICH* and MARC G. GENTON†

**National Automotive Center, U.S. Army RDECOM, AMSRD-TR-N (MS 157), Warren, MI 48397-5000, USA*
gorsichd@tacom.army.mil
†*Department of Statistics, 209-E, North Carolina State University, Raleigh, NC 27695-8203,USA*
genton@stat.ncsu.edu

Received June 2002 and accepted September 2003

In this article, we describe the discretization of nonparametric covariogram estimators for isotropic stationary stochastic processes. The use of nonparametric estimators is important to avoid the difficulties in selecting a parametric model. The key property the isotropic covariogram must satisfy is to be positive definite and thus have the form characterized by Yaglom's representation of Bochner's theorem. We present an optimal discretization of the latter in the sense that the resulting nonparametric covariogram estimators are guaranteed to be smooth and positive definite in the continuum. This provides an answer to an issue raised by Hall, Fisher and Hoffmann (1994). Furthermore, from a practical viewpoint, our result is important because a nonlinear constrained algorithm can sometimes be avoided and the solution can be found by least squares. Some numerical results are presented for illustration.

Keywords: Bochner's theorem, Fourier-Bessel series, nonnegative least squares, positive definiteness, spatial prediction.

1. Introduction

Before statisticians perform optimal linear spatial prediction, or kriging, two steps are taken to estimate the covariogram in the continuum. The first step is to estimate the covariogram values at a fixed number of lags or distances, using a realization of the spatial stochastic process. If the process is isotropic, the covariogram is only a function of distance. It can be estimated with various estimators (see Yaglom 1987, Cressie 1993). In the second step, one of several valid (Cressie 1993) parametric models is typically fitted to these covariogram estimates. This is done either because the estimated points do not guarantee the positive definiteness of the covariogram, or values of the covariogram in the continuum are not known. We assume the covariogram estimator being used provides positive definite samples from a covariogram valid in $\mathbb{R}^d$. When parametric model fitting is used instead of a nonparametric covariogram estimator, it is often not clear which model should be used even though the choice of the parametric model is crucial (Gorsich and Genton 2000). For this reason, a nonparametric estimator for the covariogram is important and the choice of a parametric model is no longer needed.

Nonparametric covariogram estimators for isotropic stationary stochastic processes were first discussed by Shapiro and Botha (1991), followed by Lele (1995), Cherry, Banfield and Quimby (1996), Cherry (1997), and Ecker and Gelfand (1997). These methods are based on the isotropic spectral representation of positive definite functions derived by Yaglom (1957) from Bochner's (1955) theorem. Nevertheless, in order to preserve the positive definiteness property required for the covariogram, statisticians commonly resort to fitting a valid parametric model. Such an approach is often inadequate because the choice of the model is generally subjective and tests of goodness of fit are difficult for dependent data, see e.g. Armstrong and Diamond (1984), Christakos (1984), Stein (1999), and Gorsich and Genton (2000). Hall, Fisher and Hoffmann (1994) have proposed nonparametric covariogram estimators constructed with kernel methods. In particular, they criticized the approach of Shapiro and Botha (1991) because it provides a positive definite estimator only on a discrete set of points and not in the continuum. In this article, we resolve this problem by introducing an optimal discretization of Yaglom's (1957) formula, in the sense that the resulting nonparametric covariogram estimators are guaranteed

0960-3174

to be positive definite in the continuum. Our method does not always require numerical nonlinear optimization under constraints, and does not suffer from spurious oscillations in the continuum as in the approaches first described by Shapiro and Botha (1991).

The article is organized as follows. In Section 2, we start by reviewing how the series expansion of a covariogram in a Bessel basis is derived from Yaglom's representation of Bochner's theorem. We explain the positivity requirement on the coefficients of the series, and review the current techniques for determining them. In Section 3, we present an optimal choice of the nodes of the series that makes the Bessel functions an orthogonal basis. Using this result, Section 4 gives explicit formulas for the coefficients of the series by exploiting weighted orthogonality of the basis. Finally, Section 5 presents some numerical results.

2. Nonparametric covariogram estimation

Consider a spatial stochastic process $\{Z(\mathbf{x}) \mid \mathbf{x} \in D\}$, where the spatial domain D is a fixed subset of $\mathbb{R}^d$, $d \geq 1$. Assume that this process is second-order stationary with covariogram $c(\mathbf{h}) = \text{Cov}(Z(\mathbf{x}+\mathbf{h}), Z(\mathbf{x}))$, $\mathbf{h} \in \mathbb{R}^d$. We further assume that the covariogram is isotropic, i.e. it depends only on the distance $h = \|\mathbf{h}\| \in \mathbb{R}_+$ and not on the direction of the lag vector $\mathbf{h}$. Anisotropies can be explored with the usual techniques. For instance, geometric anisotropies can be corrected by a linear transformation $A\mathbf{h}$ of the lag vector $\mathbf{h}$, where A is a $d \times d$ matrix, see e.g. Cressie (1993, p. 64). More general anisotropies can be corrected with a spatial deformation, see Sampson and Guttorp (1992).

Nonparametric estimation of isotropic covariograms involves the use of a series derived from Yaglom's (1957) representation of Bochner's (1955) theorem:

$$c(h) = \sum_{j=1}^{\infty} p_j \Omega_d(t_j h), \tag{1}$$

where the p_j form a set of positive coefficients, the scalars t_j are nodes, and Ω_d is defined by:

$$\Omega_d(x) = (2/x)^{(d-2)/2} \Gamma(d/2) J_{(d-2)/2}(x), \tag{2}$$

Γ is the gamma function, and $J_{(d-2)/2}$ is a Bessel function of the first kind of order $(d-2)/2$. The functions in (2) form a basis for isotropic covariograms that are valid in $\mathbb{R}^d$, where they automatically satisfy the positive definiteness property. Previous approaches use Bayesian, nonlinear, or ad hoc methods to estimate the parameters of the series (1). In this paper, we determine the optimal choice of the nodes t_j for the series and we establish linear relationships between sampled covariogram values and the coefficients p_j of the expansion (1). This significantly improves the fit of the continuous covariogram to the covariogram sample values as well as avoids a nonlinear algorithm and the spurious oscillations that result in the other methods. Note that the nonparametric estimation of variograms ($\gamma(h)$—e.g. Cressie 1993) follows in a straightforward manner from the relation $\gamma(h) = c(0) - c(h)$.

When the spatial problem is in $\mathbb{R}^1$, and the nodes t_j are chosen as zeros of $\cos(x)$, the series (1) becomes a discrete cosine transform which has been studied rigorously, see e.g. Strang (1999). For $d = 2$, the series (1) becomes a Bessel expansion of the first kind of order zero and for $d = 3$ the expansion is made in a sinc basis. Specifically, if the nodes t_j are chosen as zeros of Bessel functions, the expansion of $2^{(2-d)/2} h^{(d-2)/2} c(h)/\Gamma(d/2)$ becomes similar to a Fourier-Bessel (or Hankel) expansion (Tolstov 1962), whereas if the nodes are zeros of the derivative of Bessel functions, the expansion is similar to a Dini expansion (Bowman 1958) of $2^{(2-d)/2} h^{(d-2)/2} c(h)/\Gamma(d/2)$:

$$\begin{aligned} c^*(h) &= 2^{(2-d)/2} h^{(d-2)/2} c(h)/\Gamma(d/2) \\ &= \sum_{j=1}^{\infty} \frac{p_j}{t_j^{(d-2)/2}} J_{(d-2)/2}(t_j h). \end{aligned} \tag{3}$$

By using the expansion (3), the nonparametric estimator of the covariogram now has excellent convergence properties. This allows Bochner's theorem to be used to prove that under certain conditions on the covariogram, the coefficients p_j in the series must be positive. If this is the case, then nonnegative least squares becomes least squares.

It is important to investigate the best way to discretize the series (1). Normally, only a finite number of basis functions can be taken so the series is truncated to n:

$$\sum_{j=1}^{n} p_j \Omega_d(t_j h). \tag{4}$$

In this paper we answer the natural questions: How should n and the nodes t_j be chosen? In which dimension d should the expansion be made? How should the coefficients p_j be solved for? Are they positive given some nodes t_j? The answer to the latter question is yes when the covariogram has certain properties and the nodes are zeros of the functions $\Omega_d(x)$. Then, because the coefficients p_j are positive, there is no need for a constrained nonlinear algorithm, and convergence is guaranteed. Before we describe a methodology for choosing these values, we discuss Yaglom's representation of Bochner's theorem.

The key idea behind a nonparametric estimator for the covariogram is Bochner's (1955) theorem, which gives the spectral representation of positive definite functions. In particular, a covariance function $c(\mathbf{h})$ is positive definite if and only if it has the form:

$$c(\mathbf{h}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \cos(\mathbf{u}^T \mathbf{h}) F(d\mathbf{u}), \tag{5}$$

where $F(d\mathbf{u})$ is a positive bounded symmetric measure. If $c(\mathbf{h})$ is isotropic, Bochner's theorem can be written as (Yaglom 1957):

$$c(h) = \int_{0}^{\infty} \Omega_d(th) F(dt), \tag{6}$$

where Ω_d form a basis for functions in $\mathbb{R}^d$, $\Omega_d(x) = (2/x)^{(d-2)/2}\Gamma(d/2)J_{(d-2)/2}(x)$. Here $F(t)$ is any nondecreasing bounded function, $\Gamma(d/2)$ is the gamma function, and J_v is the Bessel function of the first kind of order v. Some familiar examples of Ω_d are $\Omega_1(x) = \cos(x)$, $\Omega_2(x) = J_0(x)$, and $\Omega_3(x) = \sin(x)/x$.

To solve for $c(h)$, we choose a vector $\mathbf{t}$, which represents the locations of the jump points, or nodes, in a discretization of $F(t)$. Let $c(h_1), \ldots, c(h_l)$ be equally spaced sampled values from a valid covariogram c, or be covariogram estimates. To find $c(h)$ from $c(h_i)$, let $F(t) = \sum_{j=1}^{\infty} p_j \Delta(t - t_j)$ where Δ is the step function:

$$\Delta(t - t_j) = \begin{cases} 1 & \text{if } t \geq t_j, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

All that is known about $F(t)$ is that it is nondecreasing and bounded, so the coefficients p_j are only required to be nonnegative and $\sum_{j=1}^{\infty} p_j < \infty$. It is the coefficients p_j that we are interested in solving for. Given $\mathbf{t} = (t_1, \ldots, t_\infty)^T$, $t_j \geq 0$ and p_j for $j = 1, \ldots, \infty$, the nonparametric estimator has the form:

$$c(h) = \sum_{j=1}^{\infty} p_j \Omega_d(t_j h). \tag{8}$$

In moving from Yaglom's integral (6) to this series, it is assumed that:

$$\int_0^{\infty} h^{d-1}|c(h)|dh < \infty, \tag{9}$$

so the derivative of $F(t)$ exists, i.e. $dF(t) = f(t)dt$ (Yaglom 1987). For most applications, it is generally true that the covariogram goes to zero as $h \to \infty$ (Yaglom 1987).

The usual way to determine the nonnegative coefficients p_j from the l points $c(h_i)$, $i = 1, \ldots, l$, is to minimize, under the constraint $\mathbf{p} \geq \mathbf{0}$, the objective function:

$$S[\mathbf{p}] = \sum_{i=1}^{l} w_i \left(c(h_i) - \sum_{j=1}^{n} p_j \Omega_d(t_j h_i) \right)^2, \tag{10}$$

or equivalently in matrix notation:

$$S[\mathbf{p}] = (\mathbf{c} - M\mathbf{p})^T W (\mathbf{c} - M\mathbf{p}), \tag{11}$$

where $\mathbf{c} = (c(h_1), \ldots, c(h_l))^T$, $\mathbf{p} = (p_1, \ldots, p_n)^T$, $M_{ij} = \Omega_d(t_j h_i)$, and $W = \text{diag}(w_1, \ldots, w_l)$ is a weighting matrix that approximates the variance structure of the covariogram estimates $\mathbf{c}$, see Genton (1999) where extensions to the whole covariance structure are described. Generally only a finite number of p_j can be estimated, n of them, and to avoid arbitrary oscillations n should be kept less than or equal to l, the number of covariogram values. Using fewer than l coefficients gives a smoother estimate of the covariance function. These are nonnegative least squares problems, or can be formulated as quadratic programming problems (Björck 1996). There are many problems with this approach. For instance, the nonlinear algorithm does not necessarily converge (Gorsich 2000). In practice, it is observed that, if the optimization algorithm does converge, it converges only to the covariogram values. For a reasonable fit to the sample points, the algorithm typically needs a large number of nodes t_j. These nodes are related to the frequencies of the covariogram in the continuum. Because there are typically more nodes used than there are covariogram values, the optimization algorithm finds a solution that is not smooth in the continuum, even though it fits the actual values well. This causes spurious oscillations in the estimated continuous covariogram, oscillations that have nothing to do with the statistics of the problem. When inference based on the derivative of the estimated covariogram is of interest (Gorsich and Genton 2000), these oscillations are amplified and the problem becomes even more dramatic. These problems are completely related to improper node selection, and their resolution is discussed in the next section.

3. Node selection and orthogonality

The choice of the vector $\mathbf{t}$ of nodes significantly affects the solution of the minimization problem (11). If $\mathbf{t}$ is chosen adequately, the approximation error can be greatly reduced. In fact, if the components of $\mathbf{c}$ are samples from a valid covariogram, it will be shown with the correct choice of $\mathbf{t}$, that a nonlinear algorithm is no longer needed to find the vector of coefficients $\mathbf{p}$. It is important to notice that the series in equation (3) can be made orthogonal. Then the matrix M is nonsingular, and numerical error is reduced (Björck 1996). If the coefficients $\mathbf{p}$ are positive for a given node selection $\mathbf{t}$, then nonnegative least squares becomes ordinary least squares, and covariogram values can be fitted with only a small number of nodes. It is well known in physics, and for the solution of partial differential equations (Bowman 1958), that if the nodes t_j are chosen as the roots of some Bessel functions, the series (3) becomes very similar to a Fourier-Bessel series. Another type of orthogonal series is found when the nodes t_j are chosen as roots of a multiple of a Bessel function plus its derivative. This is known as a Fourier-Bessel series of the second type, or a Dini expansion (Bowman 1958, Tolstov 1962). Actually, there are many ways to orthogonalize the basis functions.

Using the zeros of the Bessel functions, the Fourier-Bessel series of order v is written as:

$$c^*(h) = \sum_{j=1}^{\infty} \tilde{p}_j J_v(t_j h), \tag{12}$$

where the t_j are the roots of $J_v(t) = 0$, $0 \leq h \leq 1$, and $\tilde{p}_j = p_j / t_j^v$. Here, $v = (d-2)/2$ and d is again a positive integer corresponding to the dimension of the spatial domain. The Dini expansion is very similar but uses the roots of the equation $tJ_v'(t) + uJ_v(t) = 0$, where u is a constant. These choices of $\mathbf{t}$ make the basis orthogonal (Bowman 1958). Because the series is orthogonal, multiplying by $hJ_v(t_k h)$ and integrating from

0 to 1 gives:

$$\tilde{p}_k = \frac{2}{J_{v+1}^2(t_k)} \int_0^1 hc^*(h) J_v(t_k h)\, dh, \tag{13}$$

for the Fourier-Bessel expansion of the covariogram. For the Dini expansion of the covariogram with $u > -v$ we have:

$$\tilde{p}_k = \frac{2}{J_v^2(t_k) - J_{v-1}(t_k) J_{v+1}(t_k)} \int_0^1 hc^*(h) J_v(t_k h)\, dh. \tag{14}$$

Zeros of Bessel functions of order v and order $v+1$ are interlaced, so the denominator of (13) and (14) is never zero. For large values of $x = t_j h$, $t_k = (k+(d-3)/4)\pi$ are good approximations of the zeros of the Bessel function. In fact for large x, $x \gg (\frac{1}{2}v^2 - \frac{1}{8})$, we have (Arfken 1985):

$$J_v(x) \approx \sqrt{\frac{2}{\pi x}} \cos\left(x - \left(v + \frac{1}{2}\right)\frac{\pi}{2}\right), \tag{15}$$

so these zeros arrive at the same asymptotic spacing as those of a cosine. The accuracy of the zeros can easily be improved using Newton's method and the first few are well known. If $u < -v$ then $tJ_v'(t) + uJ_v(t) = 0$ has two imaginary roots. The vector **t** cannot take on imaginary numbers since it is associated with a distribution, so in general $u \geq -v$.

In practice, h will not be between 0 and 1, but we will derive everything from now on from 0 to 1. The lag domain can always be rescaled. When $0 \leq h < h_{\max}$, the coefficients in equation (13) take the form:

$$\tilde{p}_k = \frac{2}{h_{\max}^2 J_{v+1}^2(t_k)} \int_0^{h_{\max}} hc^*(h) J_v\left(\frac{t_k h}{h_{\max}}\right) dh. \tag{16}$$

The other equations are just as easily transformed.

4. The discrete case and solving for p

There are two problems with equations (13), (14), and (16). First, the integral will have to be approximated since we only know $c(h)$ at a finite set of lags. Second, there is no guarantee that all coefficients p_j will be positive. Rewriting equation (4) in matrix form gives the following expansion for the covariogram:

$$\mathbf{c} = b_v D_v J_v \tilde{D}_v \mathbf{p}, \tag{17}$$

where $(\tilde{D}_v)_{ij} = \frac{\delta(i-j)}{t_j^v}$, $(D_v)_{ij} = \frac{\delta(i-j)}{h_i^v}$, $b_v = 2^v\Gamma(v+1)$ and δ is the Kronecker index. Thus, the matrix M for the covariogram, $M_{ij} = \Omega_v(t_j h_i)$, is equal to $D_v J_v \tilde{D}_v$. Assume the nodes t_j satisfy $J_v(t_j) = 0$ and the spatial lags h_i be equally spaced with $0 < h_i \leq 1$, $i = 1, \ldots, l$, and $v \geq -1/2$. The coefficients of the expansion in equation (17) are given by the solution of (11) without constraint on **p**:

$$\mathbf{p} = (M^T W^{-1} M)^{-1} M^T W^{-1} \mathbf{c}. \tag{18}$$

This is a simple linear equation for the coefficients **p**. More importantly, if there is flexibility in choosing the lags h_i as $\frac{t_i}{t_l}$, then a nearly orthogonal matrix J_v can be found which allows the coefficients **p** to be found without inverting a matrix, see Genton and Gorsich (2002).

The next step is to determine if and when the coefficients **p** given by (18) will be positive. We know that such coefficients exist if $c(h)$ is positive definite in $\mathbb{R}^d$ by Bochner's (1955) theorem. The question is whether the coefficients will be positive for a particular discretization **t**. This question depends mainly on whether the covariogram itself is valid in $\mathbb{R}^d$ and what expansion is used. In general, if the covariogram is valid in $\mathbb{R}^r$, then any expansion in Ω_d is valid provided $d \leq r$. Even though all covariograms may be valid in $\mathbb{R}^1$, this does not mean the discrete cosine transform will always give positive coefficients for any covariogram. A simulation shows this is not true (although the real coefficients of the discrete Fourier transform are always positive for positive definite functions). The most restrictive condition on a covariogram is asking that it be valid in all dimensions. A function $c(h)$ is positive definite in $\mathbb{R}^\infty$ if and only if it is related to a completely monotone function, $g(h)$, by $c(h^{1/2}) = g(h)$, see Schoenberg (1938). The covariogram $c(h)$ must fall off more quickly as d increases, as might be expected by examining the basis functions (2) for the covariogram. Among all covariograms, those in $\mathbb{R}^\infty$ have the greatest restrictions placed on them. In general, valid covariograms in $\mathbb{R}^d$ form a nested family of subspaces. When $d \to \infty$ the basis given by (2) goes to $\exp(-(t_j h)^2)$. Schoenberg (1938) proved that if β_d is the class of positive definite functions of the form given by Bochner, then the classes for all d have the property $\beta_\infty \subset \cdots \subset \beta_d \subset \cdots \subset \beta_2 \subset \beta_1$, so that as d is increased, the space of available functions is reduced. Only functions with the basis $\exp(-(t_j h)^2)$ are contained in all the classes. The following theorem determines when the coefficients of an Ω_d series expansion will be positive. It implies that all expansions in Ω_r with $r \leq d$ will also have positive coefficients. The theorem relies on Bochner's (1955) theorem as well as the double integral representation of the Fourier-Bessel series (Bowman 1958).

Theorem 1. *Let $c(h)$ be a valid covariogram in $\mathbb{R}^d, d \geq 1$, with a finite range a, i.e. $c(h) = \phi(h)$ for $0 \leq h < a$ and $c(h) = 0$ for $a \leq h \leq \infty$, so that $\int_0^\infty h^{d-1}|c(h)|dh < \infty$. Let the nodes t_j be chosen such that $J_{(d-2)/2}(t_j) = 0$. Then the coefficients $p_j \geq 0$ for all j in the series* (3), *and $\sum_{j=1}^\infty p_j < \infty$.*

The proof is given in the appendix. Because the integral $\int_0^\infty h^{d-1}|c(h)|\, dh$ is bounded, the range of the integral for p_j, $h_{\max}$, can be allowed to go to infinity. The theorem then says that the coefficients p_j will still be positive, if the integral for $f(\hat{t})$ (see appendix) is evaluated to infinity. In the discrete case, this cannot be done, and the integral is only approximated up to some lag h_l. In the next section, this will be seen as the cause of a Gibbs effect (Gray and Pinsky 1993) in the coefficients p_j caused by the truncation of $c(h)$. This oscillation in the coefficients causes some of them to drop below zero.

The following example illustrates the previous theorem. Consider a Bessel function of order zero with a node k other than a root, i.e. $c(h) = J_0(kh)$ and $J_0(k) \neq 0$. This is a valid covariogram in $\mathbb{R}^2$ by equation (6). The Fourier-Bessel expansion of this function is:

$$J_0(kh) = \sum_{j=0}^{\infty} 2J_0(k)\frac{t_j}{\left(t_j^2 - k^2\right)J_1(t_j)}J_0(t_j h). \qquad (19)$$

The function does not satisfy the boundary condition, i.e. does not have a finite range a, nor does it satisfy $\int_0^\infty h^1|c(h)|dh < \infty$. The sum of the absolute value of the coefficients does not converge, and the coefficients are not positive. In the next section, numerical computations will be carried out to illustrate these results on different valid covariograms, as well as covariograms that are not valid. From this theorem, we can see the possibility of also determining whether a given covariogram is valid and meets the fall off condition by testing the coefficients. The validity of the solution (18) is very sensitive to the two conditions, $\int_0^\infty h^{d-1}|c(h)|dh < \infty$ and setting the nodes to correspond to the boundary condition of the covariogram. In our case, the covariogram is always assumed to be 0 beyond some maximum lag so the nodes must be zeros of the Bessel functions. This means of course that the above integral is always finite if the covariogram is bounded.

If an expansion of $c(h)$ in Ω_d is valid, will a similar expansion in Ω_r, $r < d$ be valid? For example, will the coefficients in an expansion in Ω_{d-2} be positive if the coefficients in Ω_d are? The next theorem relates the coefficients p_j between two different expansions of $c(h)$.

Theorem 2. *Let* $c(h)$ *be a valid covariogram in* $\mathbb{R}^{d+2}$ *with expansions* $c(h) = \sum_{j=1}^{\infty} \tilde{p}_j \Omega_{d+2}(\tilde{t}_j h)$ *and* $c(h) = \sum_{j=1}^{\infty} p_j \Omega_d(t_j h)$. *If the expansion in* Ω_{d+2} *is valid* ($\tilde{p}_j \geq 0$, $\sum_{j=1}^{\infty} \tilde{p}_j < \infty$) *and* $\tilde{t}_1 < \tilde{t}_2 < \cdots < \tilde{t}_k$, *such that* t_k *interlaces* $\tilde{t}_{k-1}$ *and* $\tilde{t}_k$ (*as in two orthogonal expansions*) *for all* k, *then* $p_j \geq 0$ *for all* j *and* p_j *are related only to* $\tilde{p}$*'s and* $\tilde{t}$*'s by*:

$$p_{k+1} = \frac{1}{2(v+1)}\sum_{j=k+1}^{\infty} \tilde{p}_j \left(\frac{\tilde{t}_{k+1}^{2(v+1)} - \tilde{t}_k^{2(v+1)}}{\tilde{t}_j^{2(v+1)}}\right), \qquad (20)$$

where $v = (d-2)/2$.

The proof is given in the appendix. Note that the conditions for the theorem are met for all orthogonal node selections (including Dini expansions) because the zeros of Bessel functions of different orders always interlace (Watson 1958). It is also interesting to look at the implication of the theorem regarding solving problems with $d \geq 3$. The coefficients of the Ω_d basis are a weighted average of those in the Ω_{d+2} basis. Therefore, if some coefficients are negative in the Ω_{d+2} basis, it may be possible to smooth them and have a valid expansion in Ω_d. Also, if the coefficients are positive in Ω_{d+2}, they are certainly positive in Ω_d. We will take another look at this in the next section on numerical results. A similar theorem for going from a lower dimensional expansion to a higher one is easily derived taking the coefficients p_k and finding $\tilde{p}_j$. Instead of taking a weighted average of the coefficients, a weighted finite difference is used.

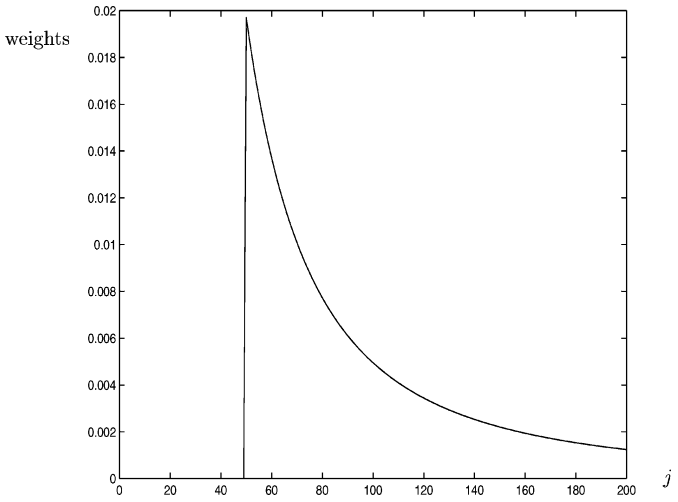

Fig. 1. *Weights applied to the coefficients of an* $\mathbb{R}^4$ *expansion to give the 50th coefficient of an* $\mathbb{R}^2$ *expansion. The weights smooth and reduce the coefficients of the higher dimensional space to find* p_{50}. *The sum of the weights is approximately .7522*

Figure 1 depicts the weights for the $\tilde{p}_j$ to find p_{50}, where $\tilde{t}_j$ are the zeros of J_1. So we are converting the coefficients of the $\mathbb{R}^4$ expansion to coefficients of the $\mathbb{R}^2$ expansion. These weights clearly smooth the coefficients of the $\mathbb{R}^4$ expansion as well as reduce their size (the sum of the coefficients in Fig. 1 is approximately .7522). This is exactly the opposite of what happens to $c(h)$ as we go down in dimension. As can be seen by Theorem 2, going from an $\mathbb{R}^{d+2}$ to an $\mathbb{R}^d$ expansion, the covariogram becomes less smooth while the coefficients become smoother.

If the expansion is not valid because the coefficients are not all positive, it may be possible to expand using a smaller d to achieve a valid expansion. On the other hand, if a smoother fit of $c(h)$ is desired, d may be increased. If it is known that the covariogram is valid in dimension d and satisfies equation (9), then it is recommended that an expansion in d is used. This guarantees the correct form and smoothness of the covariogram. Unfortunately, covariance estimates may be positive definite in $\mathbb{R}^1$, but do not have a positive definite representation in $\mathbb{R}^d$, primarily because of the variability of the estimates. When this is the case, an expansion in $r < d$ may be required.

5. Numerical results

In this section, we illustrate the results obtained previously by first computing the coefficients **p** given by equation (18) on equally spaced sampled values from various covariograms, and by checking when the coefficients are all positive. Next we discuss a typical example of nonparametric covariogram estimation in the continuum using both our method, and the method first proposed by Shapiro and Botha (1991).

In order to illustrate Theorem 1, we choose the nodes **t** as zeros of $J_{(d-2)/2}$, and use equally spaced sampled values from covariograms that are known to be valid in $\mathbb{R}^d$. These covariograms may or may not have the correct boundary condition or the property $\int_0^\infty h^{d-1}|c(h)|dh < \infty$. We compute the coefficients **p** by means of the matrix equation (18) to determine whether the smallest coefficient p_j is positive or not. For simplicity, the number of nodes and lags are both taken to be equal to l so that only square matrices are involved. Because there are l nodes for l lags, there is no error with the fitted covariogram values, if ordinary least squares and nonnegative least squares coincide. Then there is no truncation error and the coefficients fall off quickly. No aliasing issues are present with any of the covariograms tested when the nodes are properly chosen. This means the series can be used to find the covariogram values at any lag h continuously.

Table 1 lists the results using $l = 100$, and $0 \le h \le 1$. For covariograms like the first one, the spherical model, and the following four, the value of the covariogram is 0 when $h \ge 1$. The others decay as h increases. For the discrete problem, only values in $0 \le h \le 1$ are fitted. The algebra was done with MATLAB and a small set of very fast MATLAB scripts. For each covariogram, the smallest coefficient p_j is reported along with the dimension the covariogram is valid in, the value of the covariogram at $h = 1$, and the basis Ω_d in which the covariogram was expanded. For simplicity, the matrix W was taken to be the identity matrix.

The following comments can be made about the results presented in Table 1. The first five covariograms behave very nicely. The boundary condition is satisfied, and when the expansion is in the correct basis, the smallest coefficient is positive. For the covariogram $(1 - h)^2$, expanded with the basis Ω_4, the smallest coefficient falls below zero because the covariogram is not valid in that dimension. Moreover, this happens at the same time that the infinite sum of the coefficients no longer is finite. The Gaussian covariograms in Table 1, $\exp(-\alpha h^2)$, $\alpha > 0$, fail to have all positive coefficients, but only few of them are negative and they become very small when the value of the Gaussian covariogram is closer to being zero at $h = 1$. This failure is strictly due to the p_j being estimated from a truncated, sampled covariogram. The true covariogram is not zero for $h \ge 1$ and is only sampled from 0 to 1.

Table 1. *Smallest value of p_j using equation (18) for various covariograms*

Covariogram	Valid in dimension	Expanded in dimension	Smallest p_j	Value at $h = 1$
$1 - \frac{3}{2}h + \frac{1}{2}h^3$	3	2	2.57E-6	0
$(1-h)^2$	3	2	3.26E-6	0
$(1-h)^2$	3	3	5.89E-6	0
$(1-h)^2$	3	4	−9.56E-3	0
$(1-h)^3$	5	5	4.74E-5	0
e^{-10h^2}	∞	2	−1.44E-5	5.5E-5
e^{-20h^2}	∞	2	−4.45E-10	3.0E-9
e^{-60h^2}	∞	2	−3.19E-16	8.76E-27
$\Omega_2(th)$	2	2	0	0
$\Omega_6(th)$	6	2	−1.00E-5	0
$\Omega_8(th)$	8	2	−4.85E-5	0
$hK_1(th)$	∞	2	9.80E-7	8.40E-42
$hK_1(th)$	∞	3	1.47E-6	3.84E-42
$h^2K_2(th)$	∞	3	1.29E-8	3.85E-42
$\frac{1}{(1+h^2)^2}$	∞	2	−1.78E-1	.25
$\frac{1}{(1+h^2)^4}$	∞	2	−3.68E-2	6.25E-2
$\frac{1}{(1+20h^2)^4}$	∞	2	−1.06E-6	5.14E-6
$\frac{1}{(1+50h^2)^6}$	∞	2	−1.44E-12	5.68E-11

The next basis functions, $\Omega_d(th)$, are valid covariograms provided d is large enough as discussed in Theorem 2. The value of t in each case is chosen to be the fifteenth zero of the function so that the function is zero at $h = 1$. When $d = 2$, the vector **c** is a column of the matrix M and the basis is perfect. Thus, the smallest p_j is zero. This is of course unusual, given that the integral in the continuous case is unbounded. For $d = 6$ we find that the last few coefficients are negative coefficients. The problem here is that $\Omega_6(th)$ is not zero beyond $h = 1$. The coefficients are all positive when the integral in equation (13) is evaluated to infinity. For the continuous case, we obtain positive coefficients, but not from equation (18). This error is due to the discretization of the integral in equation (13). Figure 2 shows the coefficients **p** for $l = 200$. There are only a few negative coefficients at the right on the hump, and these are completely due to the discretization. The rest of the coefficients are near zero. The oscillation of the coefficients is a Gibbs phenomenon, a well known issue with truncated Fourier series. Here the oscillation is in the coefficients p_j instead of the estimate of $c(h)$. Gray and Pinsky (1993) discuss the same Gibbs phenomenon for Fourier-Bessel series. Zeroing out the negative coefficients does little to the fit of the basis function since they are typically small and there is only a small number of them.

The covariogram $c(h)$ is only sampled at discrete values, and only to $h = 1$. For $h \ge 1$ the function is assumed to be zero,

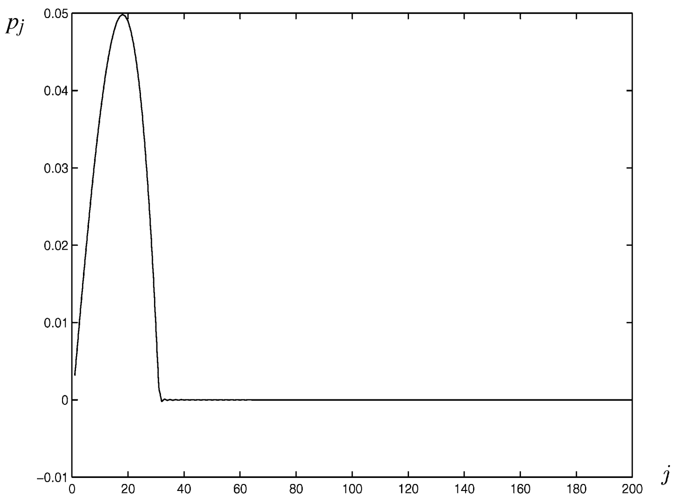

Fig. 2. *The 200 coefficients* **p** *of an expansion in $J_0(t_j h_i)$ of the basis function $\Omega_6(th_i)$. The coefficients are not all positive due to the Gibbs effect since the integral in equation (13) is only approximated*

causing an oscillation in the coefficients. The infinite continuous integral is in fact positive for all t. The key problem is the discretization and truncation (see Theorem 1) of $c(h)$, not the truncation of the series! Clearly, nonnegative least squares is not really needed here. The basis function $\Omega_8(th)$ is first expanded with $d = 1$ with the same issues as before. Then the coefficients are smoothed using the result from Theorem 2. The smallest coefficient goes from -2.1E-3 to -4.85E-5. The procedure can be repeated, dropping the smallest negative value by powers of 100.

Stein (1999) mentions that the covariograms $h^v K_v(th)$, known as the Matérn class, have considerable practical value for applications. For these covariograms, equation (18) yields positive coefficients $\mathbf{p}$. The last set of functions are valid in $\mathbb{R}^\infty$ but are not zero at $h = 1$. They perform very similarly to the covariogram basis functions above giving nearly all positive coefficients except a few, which are small in magnitude. The negative values come closer to zero as the minimum point of the covariogram decreases.

To summarize, Table 1 confirms the positivity of the coefficients provided the covariogram sampled values belong to a positive definite $\mathbb{R}^d$ function. Even when the covariogram samples are estimated, it is still beneficial to choose the nodes as before. If the variability is small, generally the solution may still be close to having positive coefficients, and a nonnegative least squares algorithm may not be needed. If the spatial dimension d is high enough, the coefficients may be smoothed (Theorem 2), and the estimated points may be fitted in a lower dimension. Because covariogram estimates that are positive definite in $\mathbb{R}^1$ can always be found, the nonparametric covariogram estimator (4) with $d = 1$ can always be reached through smoothing. In order to demonstrate the issues raised by Hall, Fisher and Hoffmann (1994) regarding the Shapiro-Botha approach, we plot the effects of different choices of nodes on the nonparametric covariogram estimator (4). A spherical covariogram, $c(h) = 1 - \frac{3}{2}h + \frac{1}{2}h^3$ for $0 \le h \le 1$ and $c(h) = 0$ for $h > 1$, is sampled at 30 equally spaced lags from 0 to 1. This covariogram has a finite range $a = 1$, and is valid up to $\mathbb{R}^3$. In Fig. 3, three nonparametric covariogram estimators (4) are computed using $d = 3$ and nodes chosen randomly (+), uniformly (*), and as zeros of the Bessel functions (dots) as discussed in Theorem 1. All nodes are in the same range, between the smallest and largest of the first 30 zeros of the Bessel function. For the estimator with the nodes t_j chosen as zeros, the fit with the sampled values of the spherical covariogram is perfect and lies exactly on them. Here equation (18) gives 30 positive coefficients p_j and nonnegative least squares is not needed. For the other choices of nodes, equation (18) gives many negative p_j, and a constrained nonlinear algorithm must be used in order to obtain valid nonparametric covariogram estimators. But here, no guarantee of convergence exists and nonnegative least squares is unable to converge to the proper solution, as can be seen in Fig. 3. The solution with random nodes has oscillations not present in the true covariogram, and the fit with a uniform set of nodes gives a covariogram that drops off much more sharply than the spherical covariogram does.

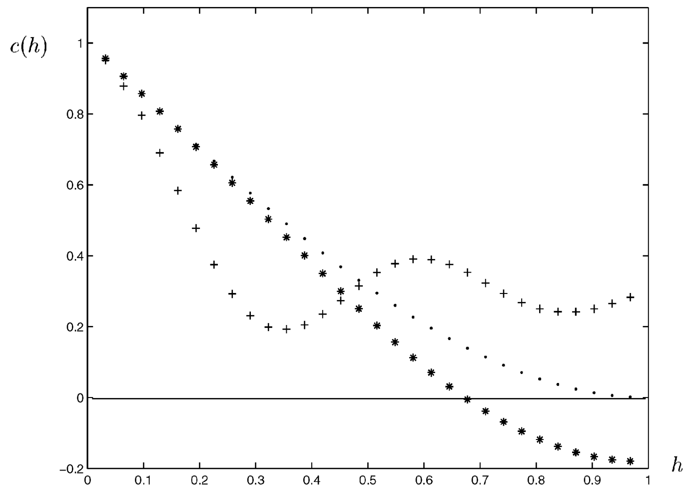

Fig. 3. *Shown are three different nonparametric covariogram estimators (4) with $d = 3$. The difference between the estimators is the choice of nodes. The estimators are computed by solving for p_j either with (18) or (10). The covariogram values used to find the estimators come from a spherical covariogram. The estimator with nodes selected as in Theorem 1 fits the covariogram values exactly and are represented as dots. The estimator with random nodes is plotted with pluses (+) and is unable to represent the true covariogram. The estimator with equally spaced nodes, represented here with stars (*), is also unable to fit the true covariogram values*

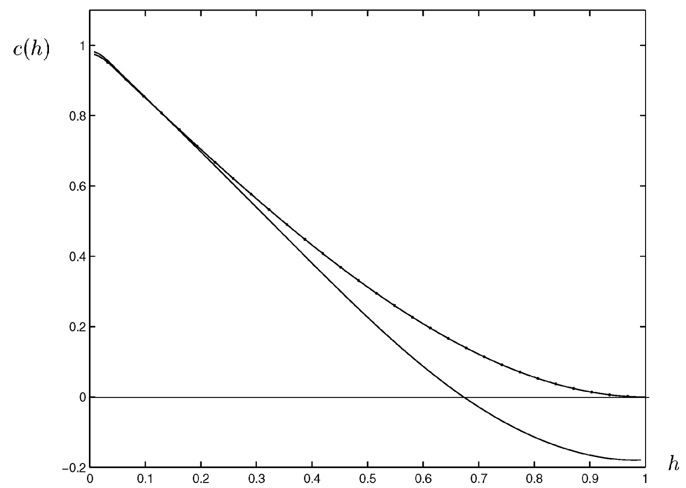

Fig. 4. *The nonparametric covariogram estimators (4) with nodes chosen uniformly (solid line) and as zeros of Bessel functions in the same range (solid line through dots). Both estimators are smooth in the continuum, but the first one is unable to represent the true spherical covariogram. The covariogram sampled values are represented as dots*

The random node selection is generally poor, and we now only compare the two others in the continuum. In Fig. 4, the nonparametric covariogram estimators with nodes chosen equally spaced and as zeros are plotted as a function of h in the continuum. Both are smooth because only 30 nodes are being used, but only the one with nodes chosen as zeros actually fits the

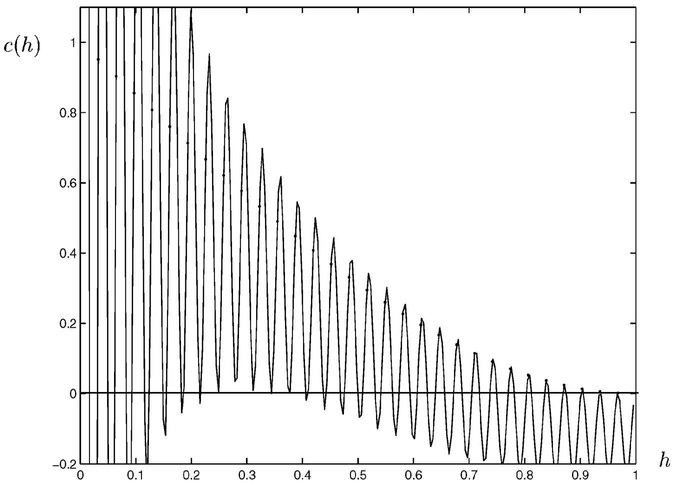

Fig. 5. *The spurious oscillations (mentioned by Hall, Fisher and Hoffmann 1994) of the nonparametric covariogram estimator (4) in the continuum. Now the estimator can represent the true covariogram only on a discrete set of values, shown here as dots, by using 90 nodes instead of 30. The result is an unsmooth covariogram which does not fit the true spherical covariogram in the continuum*

covariogram values. To force the nonparametric covariogram estimator with equally spaced nodes (but not zeros) to fit the covariogram values, Shapiro and Botha (1991), Cherry, Banfield and Quimby (1996), and Cherry (1997), increase the number of nodes used. In order to fit most of the spherical covariogram values, 90 nodes are required, and nonnegative least squares must again be used because (18) gives negative coefficients. Nonnegative least squares does find a solution, but now aliasing causes the solution to have spurious oscillations in the continuum as shown in Fig. 5. As pointed out by Hall, Fisher and Hoffmann (1994), the problem here is in the continuum, where the wild oscillations have nothing to do with the true spherical covariogram. Our method fits the true covariogram values very accurately, without nonnegative least squares and without spurious oscillations.

6. Conclusions

In this article, we have demonstrated the benefits of choosing the nodes of a discretization so that the Bessel functions for the basis of isotropic covariograms in $\mathbb{R}^d$ are orthogonal. With this selection of nodes, the problems discussed by Hall, Fisher and Hoffmann (1994) in the other nonparametric covariogram estimators, first proposed by Shapiro and Botha (1991), are resolved. Specifically, the nonparametric estimator in the continuum is positive definite and no longer has spurious oscillations. We proved that a nonlinear algorithm is not needed because these formulas will give positive coefficients, provided the covariogram is valid and has the correct boundary condition. Then we gave relationships for expansions in different dimensions d and the effects on the coefficients of the nonparametric estimator. Finally we described a number of numerical computations to illustrate the results and offered some general techniques for nonparametric estimation of isotropic covariograms. These nonparametric estimators provide a mechanism to determine any value of $c(h)$ in the continuum given only a finite number of function values, a crucial requirement for kriging techniques. A practical application of nonparametric covariogram estimation to photometric data arising from Astronomy can be found in Genton and Gorsich (2002).

Appendix

Proof of Theorem 1: Because $c(h) = \int_0^\infty \Omega_d(th)\,dF(t)$ and $\int_0^\infty h^{d-1}|c(h)|dh < \infty$, we have:

$$\int_0^\infty \left| \int_0^\infty \frac{J_{(d-2)/2}(th)}{t^{(d-2)/2}} f(t)\,dt \right| h^{d/2} dh < \infty. \tag{21}$$

This implies (Watson 1958, Tolstov 1962) that the series (3) converges uniformly to $c^*(h)$. Thus, $\sum_{j=1}^\infty p_j \Omega_d(t_j h)$ converges uniformly to $c(h)$. Now using the fact that the series is orthogonal, i.e.:

$$\int_0^{h_{\max}} h J_{(d-2)/2}(t_j h/h_{\max}) J_{(d-2)/2}(t_k h/h_{\max})\,dh = \delta(j-k) h_{\max}^2 J_{d/2}^2(t_j)/2, \tag{22}$$

where $a \le h_{\max} < \infty$, we can solve for each p_j using equation (13) and arrive at:

$$c^*(h) = \sum_{j=1}^\infty \left(\frac{2}{h_{\max}^2 J_{d/2}^2(t_j)} \int_0^{h_{\max}} \tilde{h}^{d/2} c(\tilde{h}) J_{(d-2)/2}(t_j \tilde{h}/h_{\max})\,d\tilde{h} \right) \times J_{(d-2)/2}(t_j h/h_{\max}). \tag{23}$$

This is equivalent to:

$$c^*(h) = \sum_{j=1}^\infty \left(\frac{2}{h_{\max}^2 J_{d/2}^2(t_j)} \int_0^a \tilde{h}^{d/2} \phi(\tilde{h}) J_{(d-2)/2}(t_j \tilde{h}/h_{\max}) \times J_{(d-2)/2}(t_j h/h_{\max})\,d\tilde{h} \right). \tag{24}$$

The term $J_{d/2}^2(t_j)$ can be approximated as $\frac{2}{t_j \pi} \sin^2(t_j - d\pi/2 + \pi/4)$ where the error goes to zero as t_j gets large. The zeros t_j can also be approximated as $(j + (d-3)/4)\pi$ with the error going to zero as t_j gets large (Watson 1958, Tolstov 1962). Moreover, every term in the expansion above of $b_{(2-d)/2} h^{(d-2)/2} \phi(h)$, $b_{(d-2)/2} = 2^{(d-2)/2}\Gamma(d/2)$, tends to zero as $h_{\max}$ gets large, so the error in the first terms can be neglected for large $h_{\max}$. Using this approximation for both t_j and $J_{d/2}^2(t_j)$ we have:

$$J_{d/2}^2(t_j) \approx \frac{2}{t_j \pi}. \tag{25}$$

Now $b_{(2-d)/2}h^{(d-2)/2}\phi(h)$ can be written as:

$$\sum_{j=1}^{\infty}\frac{\pi t_j}{h_{\max}^2}\int_0^a \tilde{h}^{d/2}\phi(\tilde{h})J_{(d-2)/2}(t_j\tilde{h}/h_{\max})J_{(d-2)/2}(t_j h/h_{\max})\,d\tilde{h}. \tag{26}$$

Setting $\hat{t}_j = t_j/h_{\max}$ and $\delta\hat{t} = \hat{t}_{j+1} - \hat{t}_j$, we have $\delta\hat{t} = \pi/h_{\max}$. Now

$$b_{(2-d)/2}h^{(d-2)/2}\phi(h) = \sum_{j=1}^{\infty}\hat{t}_j\delta\hat{t}\int_0^a \tilde{h}^{d/2}\phi(\tilde{h})J_{(d-2)/2}(t_j\tilde{h}/h_{\max}) \times J_{(d-2)/2}(t_j h/h_{\max})\,d\tilde{h}. \tag{27}$$

Therefore, as $h_{\max}\to\infty$, the sum can be rewritten as an integral:

$$b_{(d-2)/2}h^{(d-2)/2}\phi(h) = \int_0^{\infty} J_{(d-2)/2}(\hat{t}h)\hat{t}d\hat{t}\int_0^a \tilde{h}^{d/2}\phi(\tilde{h})J_{(d-2)/2}(\hat{t}\tilde{h})d\tilde{h}. \tag{28}$$

Setting $f(\hat{t}) = \int_0^a \tilde{h}^{d/2}\phi(\tilde{h})J_{(d-2)/2}(\hat{t}\tilde{h})d\tilde{h}$, we have:

$$b_{(2-d)/2}h^{(d-2)/2}\phi(h) = \int_0^{\infty} J_{(d-2)/2}(\hat{t}h)f(\hat{t})\hat{t}\,d\hat{t}. \tag{29}$$

Now we have Yaglom's representation (1957) of Bochner's theorem:

$$\phi(h) = b_{(d-2)/2}\int_0^{\infty}\frac{J_{(d-2)/2}(\hat{t}h)}{(\hat{t}h)^{(d-2)/2}}f(\hat{t})\hat{t}^{d/2}\,d\hat{t}, \tag{30}$$

or simply:

$$c(h) = b_{(d-2)/2}\int_0^{\infty}\frac{J_{(d-2)/2}(\hat{t}h)}{(\hat{t}h)^{(d-2)/2}}\,dF(t), \tag{31}$$

because $dF(\hat{t}) = \hat{t}^{d/2}f(\hat{t})\,dt$. The condition that $\int_0^{\infty} h^{d-1}|c(h)|dh < \infty$ is equivalent to the condition $\int_0^{\infty}\hat{t}^{d-1}f(\hat{t})d(\hat{t}) < \infty$ (Stein 1999). Because $c(h)$ is valid, $f(\hat{t}) \geq 0$ for all $\hat{t}$. This means $f(\hat{t}_j) \geq 0$ for all j. Moreover since $F(\hat{t})$ is bounded, $\sum_{j=1}^{\infty} p_j < \infty$, thus proving the theorem. □

Proof of Theorem 2: We know from Theorem 1 that if $c(h)$ is valid in $d+2$ and $\int_0^{\infty} h^{d+1}|c(h)|dh < \infty$, then assuming appropriate boundary condition, we can represent $h^{v+1}c(h)$ using an orthogonal series, where again $v = (d-2)/2$. We assume this series is known for $v+1$ and that the coefficients are positive and have a finite sum. Then the series representation in Ω_d is valid if the nodes between $v+1$ and v are interlaced. This is always true in a Fourier-Bessel or Dini expansion (Watson 1958, Tolstov 1962, Tranter 1969), so this is a natural request. To prove the relation above, that surprisingly does not depend on t_j, start by writing:

$$\sum_{j=1}^{\infty}\tilde{p}_j\Omega_{d+2}(\tilde{t}_j h) = \sum_{j=1}^{\infty}p_j\Omega_d(t_j h). \tag{32}$$

Now multiplying both sides by $h^v J_{v+1}(\tilde{t}_l h)$ for some l and integrating with respect to h gives:

$$\sum_{j=1}^{\infty}\frac{d\tilde{p}_j}{\tilde{t}_j^{n+1}}\int_0^{\infty}\frac{J_{v+1}(\tilde{t}_l h)J_{v+1}(\tilde{t}_j h)}{h}\,dh = \sum_{j=1}^{\infty}\frac{p_j}{t_j^n}\int_0^{\infty}J_{v+1}(\tilde{t}_l h)J_v(t_j h)\,dh. \tag{33}$$

The values of these integrals are known (Wheelon 1968). Evaluating the integrals and then multiplying each side by $\tilde{t}_l^{v+1}$ gives:

$$\frac{1}{2(v+1)}\left(\sum_{\tilde{t}_l\leq\tilde{t}_j}\tilde{p}_j\left(\frac{\tilde{t}_l}{\tilde{t}_j}\right)^{2(v+1)} + \sum_{\tilde{t}_j<\tilde{t}_l}\tilde{p}_j\right) = \sum_{t_j<\tilde{t}_l}p_j, \tag{34}$$

where now the t_j have canceled, and the interlacing of the nodes was used. It is assumed $\tilde{t}_l > t_l$, which is the case for the Fourier-Bessel and Dini expansions. Now for $l = 1$ we have:

$$p_1 = \frac{1}{2(v+1)}\sum_{j=1}^{\infty}\tilde{p}_j\left(\frac{\tilde{t}_1}{\tilde{t}_j}\right)^{2(v+1)}, \tag{35}$$

so that $p_1 \geq 0$ since $\tilde{p}_j \geq 0$ and $\tilde{t}_j \geq 0$ for all j. By subtracting the equations for $l+1$ and l we then arrive at:

$$p_{l+1} = \frac{1}{2(v+1)}\sum_{j=l+1}^{\infty}\tilde{p}_j\left(\frac{\tilde{t}_{l+1}^{2(v+1)} - \tilde{t}_l^{2(v+1)}}{\tilde{t}_j^{2(v+1)}}\right), \tag{36}$$

so again this implies $p_l \geq 0$ as well since $\tilde{t}_{l+1} > \tilde{t}_l$. Therefore all p_l are positive, finishing the proof. □

Acknowledgments

We would like to thank the Editor, the Associate Editor, and two anonymous referees for valuable comments that have improved the paper.

References

Arfken G. 1985. Mathematical Methods for Physicists. Academic Press, New York.

Armstrong A.G. and Diamond P. 1984. Testing variograms for positive-definiteness. Math. Geol. 16: 407–421.

Bochner S. 1955. Harmonic Analysis and the Theory of Probability. University of California Press, Los Angeles.

Bowman F. 1958. Introduction to Bessel Functions. Dover, New York.

Björck A. 1996. Numerical Methods for Least Squares Problems. SIAM, Philadelphia.

Cherry S., Banfield J. and Quimby W.F. 1996. An evaluation of a non-parametric method of estimating semi-variograms of isotropic spatial processes. Journal of Applied Statistics 23: 435–449.

Cherry S. 1997. Non-parametric estimation of the sill in geostatistics. Environmetrics 8: 13–27.

Christakos G. 1984. On the problem of permissible covariance and variogram models. Water Resources Research 20: 251–265.

Cressie N. 1993. Statistics for Spatial Data. John Wiley & Sons, New York.

Ecker M.D. and Gelfand A.E. 1997. Bayesian variogram modeling for an isotropic spatial process. Journal of Agricultural, Biological and Environmental Statistics 2: 347–368.

Genton M.G. 1999. The correlation structure of the sample Autocovariance function for a particular class of time series with elliptically contoured distribution. Statistics and Probability Letters 41: 131–137.

Genton M.G. and Gorsich D.J. 2002. Nonparametric variogram and covariogram estimation with Fourier-Bessel matrices. Computational Statistics and Data Analysis. Special issue on Matrix Computations and Statistics 41: 47–57.

Gorsich D.J. and Genton M.G. 2000. Variogram model selection via nonparametric derivative estimation. Math. Geol. 32: 249–270.

Gorsich D.J. 2000. Nonparametric estimation of dependencies for spatial interpolation. Ph.D. Thesis, MIT.

Gray A. and Pinsky M. 1993. Gibbs phenomenon for Fourier-Bessel series. Expo. Math. 11: 1058–1062.

Hall P., Fisher N. and Hoffmann B. 1994. On the Nonparametric estimation of covariance functions. The Annals of Statistics 22: 2115–2134.

Lele S. 1995. Inner product matrices, kriging, and Nonparametric estimation of variogram. Math. Geol. 27: 673–692.

Sampson P.D. and Guttorp P. 1992. Nonparametric estimation of nonstationary spatial covariance structure. Journal of the American Statistical Association 87: 108–119.

Schoenberg I.J. 1938. Metric spaces and completely monotone functions. Annals of Mathematics 39: 811–841.

Shapiro A. and Botha J.D. 1991. Variogram fitting with a general class of conditionally nonnegative definite functions. Computational Statistics and Data Analysis 11: 87–96.

Stein M.L. 1999. Interpolation of Spatial Data: Some Theory for Kriging. Springer Series in Statistics, New York.

Strang G. 1999. The discrete cosine transform. SIAM Review 41: 135–147.

Tolstov G.P. 1962. Fourier Series. Prentice-Hall, New Jersey.

Tranter C.J. 1969. Bessel Functions with Some Physical Applications. Hart Publishing Co., New York.

Watson G.N. 1958. Treatise on the Theory of Bessel Functions. The University Press, Cambridge.

Wheelon A.D. 1968. Tables of Summable Series and Integrals Involving Bessel Functions. Holden-Day, San Francisco.

Yaglom A.M. 1957. Some classes of random fields in n-dimensional space, related to stationary random processes, Theory of Probability and its Applications 2: 273–320.

Yaglom A.M. 1987. Correlation Theory of Stationary and Related Random Functions, Vol. I & II. Springer Series in Statistics, New York.