Probab. Theory Relat. Fields 117, 133–144 (2000)
Digital Object Identifier (DOI) 10.1007/s004400000048

F. Abramovich · T. Sapatinas · B.W. Silverman

Stochastic expansions in an overcomplete wavelet dictionary

Received: 21 September 1998 / Revised version: 20 August 1999 /
Published online: 30 March 2000

Abstract. We consider random functions defined in terms of members of an overcomplete wavelet dictionary. The function is modelled as a sum of wavelet components at arbitrary positions and scales where the locations of the wavelet components and the magnitudes of their coefficients are chosen with respect to a marked Poisson process model. The relationships between the parameters of the model and the parameters of those Besov spaces within which realizations will fall are investigated. The models allow functions with specified regularity properties to be generated. They can potentially be used as priors in a Bayesian approach to curve estimation, extending current standard wavelet methods to be free from the dyadic positions and scales of the basis functions.

1. Introduction

1.1. Background

Wavelets have recently been of great interest in various statistical areas such as nonparametric regression, density estimation, inverse problems, change point problems, and time series analysis. Surveys of wavelet applications in these and other related statistical areas can be found, for example, in Ogden (1997), Härdle, Kerkyacharian, Picard & Tsybakov (1998), Antoniadis (1999), Silverman (1999), Vidakovic (1999) and Abramovich, Bailey & Sapatinas (2000). An interesting development, motivated by Bayesian approaches to curve estimation, is the modelling of a function as an orthonormal wavelet expansion with random coefficients. Abramovich, Sapatinas & Silverman (1998) considered such models in detail, and studied the Besov regularity properties of the functions produced by the models. They consider the application of the models in a Bayesian context, and also give

F. Abramovich: Department of Statistics and Operations Research, Raymond & Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Ramat Aviv 69978, Israel

B.W. Silverman: School of Mathematics, University of Bristol, Bristol BS8 1TW, United Kingdom

T. Sapatinas (*author for correspondence*): Institute of Mathematics and Statistics, University of Kent at Canterbury, Canterbury, Kent CT2 7NF, United Kingdom. e-mail: `T.Sapatinas@ukc.ac.uk`

Mathematics Subject Classification (1991): 60G55, 62G07

Key words and phrases: Besov spaces – Continuous wavelet transform – Overcomplete wavelet dictionaries – Poisson processes

references to related work by other authors; the results obtained have generally been very encouraging.

1.2. Abandoning dyadic constraints

Orthonormal wavelet bases have the disadvantage that the positions and the scales of the basis functions are subject to dyadic constraints. In order to avoid these constraints, this paper considers random functions defined by expansions in a continuous wavelet dictionary, where functions are built up from wavelet components that may have arbitrary positions and scales. The models provide a constructive method of simulating functions with varying degrees of regularity and spatial homogeneity, and our results give the explicit regularity properties of the functions thus produced, in terms of Besov spaces.

Some users might wish to be able to simulate or construct functions with specific Besov parameters in mind. Others might wish to use the models to gain intuition about the meaning of the Besov parameters, by generating functions that lie just inside and just outside particular Besov spaces. The models lay open the possibility of building a Bayesian curve estimation approach with the advantages of standard wavelet methods, in that inhomogeneous functions can be modelled under the prior, but without the artificial dyadic constraints on the positions and scales of the basis functions. The improvement to standard wavelet thresholding methods obtained by moving from the discrete (decimated) wavelet transform to the non-decimated wavelet transform (see, for example, Coifman & Donoho, 1995; Nason & Silverman, 1995; Johnstone & Silverman, 1997) suggests that a Bayesian approach freed from dyadic positions and scales may result in yet better wavelet shrinkage estimators. The algorithmic details, probably involving modern Bayesian computational methods, have yet to be worked out in detail, and this is an interesting topic for further research.

1.3. Models in continuous wavelet dictionaries

For simplicity of exposition we work with functions periodic on $[0, 1]$. Suppose that ϕ and ψ are the compact-support scaling function and mother wavelet respectively that correspond to an r-regular multiresolution analysis, for some integer $r > 0$ (see, for example, Daubechies, 1992). Take $a_0 = 2^{j_0}$, for some integer j_0, such that a_0 is at least twice the length of the support of ψ. For indices $\lambda = (a, b)$ with $a > a_0$ and $0 < b < 1$ we define $\psi_\lambda(t) = a^{1/2}\psi(a(t-b))$ wrapping ψ_λ periodically if necessary.

We model our function as the sum of a coarse-scale function f_0 and a fine-scale function f. The function f_0 is given by

$$f_0(t) = \sum_{i=1}^{M} \eta_i \phi_{\lambda_i}(t) \tag{1}$$

for some finite set of indices (a_i, b_i), $i = 1, 2, \ldots, M$, with $a_i \le a_0$, and some real numbers η_i. Here ϕ_λ has an analogous definition to ψ_λ. The function f is generated by a stochastic mechanism and is given by

$$f(t) = \sum_{\lambda \in S} \omega_\lambda \psi_\lambda(t) \quad . \tag{2}$$

The locations of the wavelet components and the magnitudes of their coefficients are chosen with respect to a marked Poisson process model. Specifically, the set Λ of indices $\lambda = (a, b)$ is sampled from a Poisson process S on $[a_0, \infty) \times [0, 1]$ with intensity $\mu(\lambda)$. Conditional on S, the wavelet coefficients ω_λ are assumed to be independent normal random variables

$$\omega_\lambda \mid S \sim N(0, \tau^2(\lambda)) \quad . \tag{3}$$

It is assumed that the variance $\tau^2(\lambda)$ and the intensity $\mu(\lambda)$ depend on the scale a only, and are of the form

$$\tau_a^2 \propto a^{-\delta} \quad \text{and} \quad \mu_a \propto a^{-\zeta}, \quad a \geq 1 \ , \tag{4}$$

where $\delta, \zeta \geq 0$, with $\delta + \zeta > 0$.

The stochastic wavelet expansions we consider allow intuitive notions about the functions genuinely to be modelled. The parameter ζ controls the relative rarity of fine-scale wavelet components in the function, while the parameter δ controls the size of the contribution of these components when they appear. For example, if ζ is small and δ is large, there will be a considerable number of fine-scale components but these will each have fairly low contribution, so one might expect the functions to be reasonably smooth and homogeneous. On the other hand, if ζ is large and δ is small, there will be occasional large fine-scale effects in the functions.

In the remainder of the paper, we investigate the regularity properties of the random functions generated by the proposed model. We reveal relations between the parameters δ and ζ of the model and the parameters of those Besov spaces within which realizations from the model will fall.

2. Regularity properties of the random functions

An important tool in our argument will be the equivalence between the Besov norm of the function f on $[0, 1]$ and the corresponding sequence norm of its orthonormal wavelet coefficients. For details of Besov spaces see, for example, Meyer (1992, Chapter 6), Härdle, Kerkyacharian, Picard & Tsybakov (1998, Chapter 9).

For $j \geq j_0$, define w_j to be the vector of orthonormal wavelet coefficients $w_{jk} = \langle f, \psi_{jk} \rangle,\ 0 \leq k \leq 2^j - 1$. Define also the vector u_{j_0} to have elements $u_{j_0 k} = \langle f, \phi_{j_0 k} \rangle,\ 0 \leq k \leq 2^{j_0} - 1$. Let $s' = s + 1/2 - 1/p$ and define the norm of the array w by

$$||w||_{b^s_{p,q}} = \left\{ \sum_{j=j_0}^{\infty} 2^{js'q} ||w_j||_p^q \right\}^{1/q}, \quad 1 \leq q < \infty \ ,$$

$$||w||_{b^s_{p,\infty}} = \sup_{j \geq j_0} \left\{ 2^{js'} ||w_j||_p \right\} \quad .$$

Then, for $0 < s < r$, $1 \le p, q \le \infty$, the Besov norm $||f||_{B^s_{p,q}}$ on [0, 1] is equivalent to the sequence space norm $||u_{j_0}||_p + ||w||_{b^s_{p,q}}$ (see, for example, Donoho & Johnstone, 1998, Theorem 2).

Because f_0 given by (1) is a finite linear combination of functions ϕ_λ, it will belong to the same Besov spaces as the scaling functions, including all those for which $0 < s < r$. For these parameter values, we consider in detail the necessary and sufficient conditions for f given by (2) to fall (with probability one) in any particular Besov space.

Theorem 1. *Let ϕ and ψ be the compact-support scaling function and mother wavelet respectively that correspond to an r-regular multiresolution analysis. Consider a function f as defined in (2), with the conditional variances $\tau_a^2 \propto a^{-\delta}$ and the intensity of the Poisson process $\mu_a \propto a^{-\zeta}$. Assume that $\delta \ge 0$, $0 \le \zeta \le 1$, and that $\delta + \zeta > 0$. Assume also that ϕ (and hence ψ) are sufficiently regular that $2(r + \rho) > 1 + \delta$, where $\rho \in (0, 1)$ is the exponent of Hölder continuity of the r-th derivative of ϕ and ψ. Then, for $0 < s < r$, $1 \le p, q \le \infty$*

$$f \in B^s_{p,q} \quad \textit{almost surely}$$

if and only if

$$\begin{cases} s + 1/2 - \zeta/p - \delta/2 < 0 & \textit{if } 1 \le p < \infty \\ s + 1/2 - \delta/2 < 0 & \textit{if } p = \infty \ . \end{cases}$$

Proof. Define $\gamma = \zeta/p + \delta/2 - s - 1/2$ if $1 \le p < \infty$ and $\gamma = \delta/2 - s - 1/2$ if $p = \infty$.

2.1. Sufficiency: Case $1 \le p < \infty$

Consider the orthonormal wavelet coefficients $w_{jk} = \langle f, \psi_{jk} \rangle$ and set $\lambda_{jk} = (2^j, 2^{-j}k)$. For resolution and spatial indices j and k with $j \ge j_0$ and $k = 0, 1, \ldots, 2^j - 1$ respectively, we then have

$$w_{jk} = \sum_{\lambda \in S} K(\lambda, \lambda_{jk}) \omega_\lambda \ , \tag{5}$$

where $K(\lambda, \lambda') = \langle \psi_\lambda, \psi_{\lambda'} \rangle$.

We now explore some properties of the reproducing kernel K that we use in subsequent calculations. Firstly, since $\int \psi_\lambda^2 = 1$ for all λ, we always have $K^2(\lambda, \lambda') \le 1$. Now define $K_0(u, v) = \langle \psi, \psi_{uv} \rangle$. Let $\lambda = (a, b)$ and $\lambda' = (a', b')$. Simple calculus shows that

$$K(\lambda, \lambda') = K_0(a/a', a'(b - b')) \ . \tag{6}$$

In the particular case where $\lambda' = \lambda_{jk}$, we have $K(\lambda, \lambda_{jk}) = K_0(2^{-j}a, 2^j b - k)$. Let $[L_\psi, U_\psi]$ be the support of the mother wavelet ψ. Then $K_0(u, v) = \langle \psi, \psi_{uv} \rangle \ne 0$ only if

$$L_\psi - U_\psi/u \le v \le U_\psi - L_\psi/u \ . \tag{7}$$

In what follows we use C to denote a generic positive constant, not necessarily the same each time it is used. We have, from Daubechies (1992, p. 48),

$$|K_0(u,v)| \le Cu^{-(r+\rho+1/2)}, \quad \text{uniformly in } u \ge 1 \ . \tag{8}$$

For $u < 1$, apply the symmetry of K to show that $K_0(u,v) = K_0(1/u, -uv)$ and hence

$$|K_0(u,v)| \le Cu^{(r+\rho+1/2)}, \quad \text{uniformly in } u \le 1 \ . \tag{9}$$

We now study the moments of the orthonormal wavelet coefficients w_{jk}. It follows from (5) that, conditionally on S, the distribution of w_{jk} is normal with mean zero and variance

$$\sigma^2_{jk}(S) = \sum_{\lambda \in S} K^2(\lambda, \lambda_{jk}) a^{-\delta} \ , \tag{10}$$

where, as usual, $\lambda = (a, b)$. The unconditional distribution of w_{jk} will have finite variance if the expectation of $\sigma^2_{jk}(S)$ over S is finite. If the sum in (10) is infinite for a particular S, then, conditionally on S, the sum defining w_{jk} cannot converge, because it will not converge in distribution. More generally, for $p > 0$, the pth absolute moment of w_{jk} will be given by

$$E|w_{jk}|^p = \nu_p E_S \left\{ \sum_{\lambda \in S} K^2(\lambda, \lambda_{jk}) a^{-\delta} \right\}^{p/2} \ , \tag{11}$$

where ν_p is the pth absolute moment of the standard normal distribution.

Let $\mathscr{T}^0_{jk}$ be the set $[a_0, \infty) \times (2^{-j}k - 1/2, 2^{-j}k + 1/2)$. By the definition of a_0 and j_0, we may restrict attention in the sum (10) to $S \cap \mathscr{T}^0_{jk}$ since the support of any ψ_λ with $a > a_0$ will be of length at most $\frac{1}{2}$, and so the terms excluded by restricting the sum will all be zero. Now let S'_j be a Poisson process on the half-plane $\{(u, v) : u > 0, -\infty < v < \infty\}$ of intensity $2^{-j\zeta} u^{-\zeta}$. Define the set $\mathscr{T}_j = [2^{-j}, \infty) \times [-2^{j-1}, 2^{j-1}]$. Consider the transformation of $\lambda = (a, b)$ given by $(u, v) = (2^{-j}a, 2^j b - k)$. Applied to the process $S \cap \mathscr{T}^0_{jk}$ this gives a process with the same distribution as $S'_j \cap \mathscr{T}_j$. In addition, for each λ, we have from (6) that $K(\lambda, \lambda_{jk}) = K_0(u, v)$. It follows that

$$\begin{aligned} E|w_{jk}|^p &= \nu_p 2^{-j\delta p/2} E \left\{ \sum_{(u,v) \in S'_j \cap \mathscr{T}_j} K_0^2(u,v) u^{-\delta} \right\}^{p/2} \\ &\le \nu_p 2^{-j\delta p/2} E \left\{ \sum_{(u,v) \in S'_j} K_0^2(u,v) u^{-\delta} \right\}^{p/2} \ . \end{aligned} \tag{12}$$

To obtain a bound on the expectation in (12), define the random sum

$$Z_j = \sum_{(u,v) \in S'_j} K_0^2(u,v) u^{-\delta} \ . \tag{13}$$

The bounds (8) and (9) imply that $K_0^2(u,v)u^{-\delta}$ is bounded by $Cu^{2r+2\rho+1-\delta}$ for $0 < u \le 1$, and by $Cu^{-2r-2\rho-1-\delta}$ for $u \ge 1$; hence it is uniformly bounded for all u and v.

We now apply Corollary 1 in the Appendix to investigate the behaviour of $EZ_j^{p/2}$. To verify the finiteness of the first integral in the corollary, it will be sufficient to have finiteness of the integral

$$\int_0^\infty \int_{-\infty}^\infty K_0^2(u,v)u^{-\delta-\zeta}\,dv\,du \quad . \tag{14}$$

The bounds (7) on the support of the integrand, and those stated above on its order of magnitude, allow (14) to be dominated by

$$C\int_0^1 u^{2r+2\rho+1-\delta-\zeta}(1+1/u)du + C\int_1^\infty u^{-2r-2\rho-1-\delta-\zeta}(1+1/u)du \quad .$$

The assumptions of the theorem about the regularity of the wavelets imply that the first integral is finite since

$$2r+2\rho+1-\delta-\zeta > 2-\zeta > 1-\zeta \ge 0 \quad ,$$

while the second integral is clearly finite since $2r+2\rho+\delta+\zeta > 0$.

We now verify the finiteness of the second integral in the corollary. By similar arguments to those just used,

$$\begin{aligned}&\int_0^\infty \int_{-\infty}^\infty \{K_0^2(u,v)u^{-\delta}\}^{p/2}u^{-\zeta}\,dv\,du \\ &\quad \le C\int_0^1 u^{(2r+2\rho+1-\delta)p/2-\zeta-1}du + C\int_1^\infty u^{-(2r+2\rho+1+\delta)p/2-\zeta}du \quad .\end{aligned}$$

The assumptions of the theorem about the regularity of the wavelets imply that these integrals are both finite: the first integral is finite since

$$(2r+2\rho+1-\delta)p/2-\zeta > p-\zeta \ge 1-\zeta \ge 0$$

and the second is finite since

$$(2r+2\rho+1+\delta)p/2+\zeta > (1+\delta)p+\zeta > p \ge 1 \quad .$$

It now follows from Corollary 1 in the Appendix that, for each fixed p,

$$EZ_j^{p/2} = C2^{-j\zeta} + o(2^{-j\zeta}) \quad \text{as } j \to \infty \quad . \tag{15}$$

Now define w_j to be the vector with elements w_{jk} for $k = 0, \dots, 2^j - 1$. Substituting (15) into (12) gives

$$E\|w_j\|_p^p \le C2^{-j(\delta p/2+\zeta-1)} \quad \text{for all } j \quad . \tag{16}$$

By Jensen's inequality and (16), we have

$$E||w||_{b^s_{p,1}} = \sum_{j=j_0}^{\infty} 2^{js'} E||w_j||_p \le \sum_{j=j_0}^{\infty} 2^{js'} (E||w_j||_p^p)^{1/p} \le C \sum_{j=j_0}^{\infty} 2^{-j\gamma} \ ,$$

which is finite for $\gamma > 0$, and hence $||w||_{b^s_{p,1}}$ is finite almost surely for $\gamma > 0$.

To complete the proof we use similar methods to show that the norm $||u_{j_0}||_p$ is finite almost surely. For fixed k, let $\mathscr{T}'$ be the range of indices $\lambda = (a, b)$ with $a > a_0 = 2^{j_0}$ for which the support of ψ_λ overlaps that of ϕ_{j_0k} for some fixed k. Then

$$u_{j_0k} = \sum_{\lambda \in S \cap \mathscr{T}'} W(\lambda, \lambda_{j_0k}) \omega_\lambda \ , \tag{17}$$

where $W(\lambda, \lambda_{j_0k}) = \langle \psi_\lambda, \phi_{j_0k} \rangle = W_0(2^{-j_0}a, 2^{j_0}(b - 2^{-j_0}k))$ and $W_0(u, v) = \langle \phi, \psi_{uv} \rangle$. Note that for $\lambda \in S \cap \mathscr{T}'$, $u \ge 1$. One can easily verify that $W_0(u, v) = \langle \psi, \psi_{uv} \rangle \ne 0$ only if

$$L_\phi - U_\psi / u \le v \le U_\phi - L_\psi / u \ , \tag{18}$$

where $[L_\phi, U_\phi]$ and $[L_\psi, U_\psi]$ are the supports of the scaling function ϕ and the mother wavelet ψ respectively. From Daubechies (1992, p. 48) we have again

$$|W_0(u, v)| \le C u^{-(r+\rho+1/2)}, \quad \text{uniformly in } u \ge 1 \ . \tag{19}$$

Exploiting (18), (19) and the fact that $u \ge 1$, the same techniques used for wavelet coefficients ω_{jk} show that $E||u_{j_0}||_p^p$ is finite, and hence $||u_{j_0}||_p$ is finite almost surely.

By the equivalence of norms, we conclude that $f \in B^s_{p,1}$ almost surely, and therefore, by the embedding properties of Besov spaces (see, for example, Härdle, Kerkyacharian, Picard & Tsybakov, 1998, Corollary 9.2, p. 124), $f \in B^s_{p,q}$ almost surely for all $1 \le q \le \infty$, completing the proof for this case.

2.2. *Sufficiency: Case $p = \infty$*

For any positive θ and c, Markov's inequality implies that

$$P(|w_{jk}| > c) \le 2e^{-\theta c} E(e^{\theta w_{jk}}) \ . \tag{20}$$

To evaluate the expectation, we use the standard expression for the moment generating function of a normal distribution and Campbell's theorem (see, for example, Kingman, 1993, p. 28) applied to the random sum (10) to obtain

$$\begin{aligned}
\log E(e^{\theta w_{jk}}) &= \log E\{E(e^{\theta w_{jk}} | S)\} = \log E \ \exp\{\theta^2 \sigma^2(S)/2\} \\
&= \int_{\mathscr{T}^0_{jk}} [\exp\{\theta^2 K^2(\lambda, \lambda_{jk}) a^{-\delta}/2\} - 1] a^{-\zeta} d\lambda \\
&= 2^{-\zeta j} \int_{\mathscr{T}_j} [\exp\{\theta^2 K_0^2(u, v) 2^{-1-\delta j} u^{-\delta}\} - 1] u^{-\zeta} du\, dv
\end{aligned}$$

by the usual change of variable. Now let $M = \sup\{K_0^2(u,v)u^{-\delta}\}$, which was shown earlier to be finite. Extending the integral to the whole of the half-plane $u > 0$, we have, by the convexity of the exponential function,

$$\log E(e^{\theta w_{jk}}) \le C2^{-\zeta j} \exp(2^{-1-\delta j}\theta^2 M) \int_0^\infty \int_{-\infty}^\infty K_0^2(u,v)u^{-\delta-\zeta}\,dv\,du$$
$$\le C2^{-\zeta j} \exp(M\theta^2 2^{-1-\delta j}) \ .$$

Suppose $2^{j\delta/2}c > 1$. To obtain a bound for $P(|w_{jk}| > c)$, choose $\theta^2 = M^{-1}2^{1+\delta j} \log(2^{j\delta/2}c)$, and substitute into (20) to obtain, for positive constants C_1 and C_2,

$$\log P(|w_{jk}| > c) \le \log 2 + C_1 2^{(\delta/2-\zeta)j}c - C_2 2^{\delta j/2}c\sqrt{\log\{2^{j\delta/2}c\}} \ .$$

Both $s' = s + 1/2$ and γ are positive by the hypotheses of the theorem. Choose ϵ such that $0 < \epsilon < \gamma$ and set $c = 2^{-(s'+\epsilon)j}$. Then $2^{j\delta/2}c = 2^{(\gamma-\epsilon)j} > 1$. We now have

$$\log P(2^{(s'+\epsilon)j}|w_{jk}| > 1) \le \log 2 + C_1 2^{(\gamma-\epsilon-\zeta)j} - C_2 2^{(\gamma-\epsilon)j}\sqrt{\log 2^{(\gamma-\epsilon)j}}$$
$$\le -2^{(\gamma-\epsilon)j} \ ,$$

for sufficiently large j. Since w_j is of length 2^j, it follows that, for sufficiently large j,

$$P(2^{s'j}||w_j||_\infty > 2^{-\epsilon j}) < 2^j \exp(-2^{(\gamma-\epsilon)j}) \ .$$

This very rapidly decreasing bound on the tail probabilities implies that, with probability one, the sequence $2^{s'j}||w_j||_\infty$ is bounded by a multiple of $2^{-\epsilon j}$, and hence $||w||_{b^s_{\infty,1}}$ is finite almost surely. The same arguments used in the case of finite p for scaling coefficients show that $||u_{j_0}||_\infty$ is finite almost surely.

By the equivalence of norms, we conclude that $f \in B^s_{\infty,1}$ almost surely and, therefore, by the embedding properties of Besov spaces (see, for example, Härdle, Kerkyacharian, Picard & Tsybakov, 1998, Corollary 9.2, p. 124), $f \in B^s_{\infty,q}$ almost surely for all $1 \le q \le \infty$, completing the proof for this case, and hence we have the sufficiency.

2.3. *Necessity*

Noting that the function $K_0(u,v)$ is continuous and that $K_0(1,0) = 1$, choose c_0 with $0 < c_0 < 1$ such that $K_0(u,v) > 1/2$ for all (u,v) with $1 \le u \le 1 + c_0$ and $0 \le v \le c_0$. For $j \ge j_0$ and $k = 0, 1, \ldots, 2^j - 1$, define the nonoverlapping rectangles $\mathscr{I}_{jk}$ in the range of indices Λ as $\mathscr{I}_{jk} = [2^j, 2^j(1+c_0)] \times [2^{-j}k, 2^{-j}(k+c_0)]$.

Using (4), the expected number of wavelet components λ falling within $\mathscr{I}_{jk}$ is then $\int_{\mathscr{I}_{jk}} a^{-\zeta}\,db\,da = c_1 2^{-\zeta j}$ for some $c_1 > 0$, and hence the probability that there is one or more wavelet components in $\mathscr{I}_{jk}$ is at least $c_2 2^{-\zeta j}$ for some $c_2 > 0$. Now define

$$w'_{jk} = \sum_{\lambda \in S \cap \mathscr{I}_{jk}} K(\lambda, \lambda_{jk})\omega_\lambda$$

and observe that the w'_{jk} are independent because the $\mathscr{I}_{jk}$ are disjoint. It follows from (6) that $K(\lambda, \lambda_{jk}) > 1/2$ for λ in $\mathscr{I}_{jk}$. Note also that from (3) and (4), $\mathrm{Var}(\omega_\lambda \mid \lambda \in S \cap \mathscr{I}_{jk}) \geq 4c_3 2^{-j\delta}$ for some c_3, so that $\mathrm{Var}(K(\lambda, \lambda_{jk})\omega_\lambda \mid \lambda \in S \cap \mathscr{I}_{jk}) \geq c_3 2^{-j\delta}$.

For $j \geq j_0$ and $k = 0, 1, \ldots, 2^j - 1$, now define independent random variables w^0_{jk} to have the mixture distribution

$$w^0_{jk} \sim \pi_j N(0, \tau_j^2) + (1 - \pi_j)\delta(0) \ ,$$

where $\pi_j = c_2 2^{-\zeta j}$ and $\tau_j^2 = c_3 2^{-j\delta}$. For the orthonormal wavelet coefficients $w_{jk} = \langle f, \psi_{jk} \rangle$, it is obvious that $|w_{jk}|$ is stochastically larger than $|w'_{jk}|$ which in turn is stochastically larger than $|w^0_{jk}|$. Hence, stochastically $||w||_{b^s_{p,q}} \geq ||w'||_{b^s_{p,q}} \geq ||w^0||_{b^s_{p,q}}$ for any $0 < s < r$, $1 \leq p, q \leq \infty$.

The methods of Abramovich, Sapatinas & Silverman (1998) for independent orthonormal wavelet coefficients w^0_{jk} (see their Theorem 1) now show that if $||w^0||_{b^s_{p,q}}$ is finite almost surely, then $\gamma > 0$. This completes the proof of necessity, and hence that of the theorem as a whole.

3. Concluding remarks

Theorem 1 places an upper bound restriction on the value of ζ. In the case $\zeta > 1$, the intensity $\mu_a \propto a^{-\zeta}$ is integrable over the range of λ for which ψ_λ has support intersecting $[0, 1]$. Therefore, the number of relevant terms in the stochastic expansion of f is finite almost surely. With probability one, f will belong to the same Besov spaces as the mother wavelet ψ, namely those for which $0 < s < r$, $1 \leq p \leq \infty$, $1 \leq q \leq \infty$.

The key conclusion of Theorem 1 we have proved is that, under suitable conditions, the function f falls in $B^s_{p,q}$ if $\delta + (2/p)\zeta$ exceeds $2s + 1$. Since the fine-scale content of model functions depends both on the intensity of fine-scale components and on their size, it is not surprising that the smoothness as measured by the parameter s should depend on both parameters. The parameter p can be seen as discouraging inhomogeneity, in that the larger the value of p the more emphasis is placed on the parameter δ. For large δ, no matter how many fine-scale components there are, they each make a relatively low contribution. On the other hand, if p is small, then there is a trade-off where large weights on fine-scale components (small δ) can be tolerated if the corresponding components are relatively rare (large ζ).

The constraints placed on s in the statistical literature, for example in the optimality results of Donoho & Johnstone (1998), are often stronger than those we have assumed. Typical conditions are $\max(0, 1/p - 1/2) < s < r$ or $1/p < s < r$. These constraints ensure that the Besov spaces are function spaces rather than spaces of more general distributions (see, for example, Meyer, 1992, Chapter 6).

Acknowledgements. We are delighted to thank Boris Tsirelson for fruitful discussions. The financial support of the Engineering and Physical Sciences Research Council, the Israel Academy of Science, and the Royal Society are gratefully acknowledged. During periods of the work TS was a Research Associate in the School of Mathematics, University of Bristol, supported by EPSRC grant GR/K70236, and BWS was a Fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford, partly supported by NSF grant SBR-9601236. Helpful comments of Peter Hall and three referees are gratefully acknowledged.

Appendix: Moments of sums of thinned Poisson processes

In this appendix we prove a lemma and corollary used in the proof of Theorem 1. They are of interest in their own right.

Lemma 1. *Let μ be a measure on $\mathbb{R}$ and let S_ε be a Poisson Process on $\mathbb{R}$ with intensity measure $\varepsilon\mu$, where $\varepsilon > 0$. Assume that*

$$\int_{-\infty}^{\infty} \min(1, |x|)\ \mu(dx) < \infty \ \ \textit{and}\ \ c_l = \int_{-\infty}^{\infty} |x|^l\ \mu(dx) < \infty \quad \textit{for some } l > 0\ . \tag{21}$$

Define $Y_\varepsilon = \sum_{X\in S_\varepsilon} X$. Then

$$E|Y_\varepsilon|^l = \varepsilon\, c_l + o(\varepsilon) \quad \text{as} \quad \varepsilon \to 0\ . \tag{22}$$

Proof. Applying Campbell's Theorem (Kingman, 1993, p. 28), condition (21) shows that the sum defining Y_ε is absolutely convergent with probability one. For any $\delta > 0$, define $B_\delta = \mathbb{R}\backslash[-\delta, \delta]$. It follows from (21) that $\mu(B_\delta) < \infty$; define $F(\delta) = \mu(B_\delta)$. Now choose $\delta < 1$ to depend on ε in such a way that $\delta \to 0$ and $\varepsilon F(\delta) \to 0$ as $\varepsilon \to 0$. The dependence of δ on ε will not be expressed explicitly. Now define

$$Y_\varepsilon^{(1)} = \sum_{X\in S_\varepsilon \bigcap B_\delta} X \quad \text{and} \quad Y_\varepsilon^{(2)} = \sum_{X\in S_\varepsilon \bigcap[-\delta,\delta]} X\ .$$

Consider, first, the asymptotic behaviour of $Y_\varepsilon^{(1)}$. The number of X in $S_\varepsilon \bigcap B_\delta$ is a Poisson $(\varepsilon F(\delta))$ random variable and so

$$E|Y_\varepsilon^{(1)}|^l = \sum_{j=1}^{\infty} \exp(-\varepsilon F(\delta)) \frac{\varepsilon^j F(\delta)^j}{j!} E|\sum_{i=1}^{j} X_i|^l\ , \tag{23}$$

where $X_1, X_2, \ldots$ are independent and identically distributed random variables on B_δ with distribution $\mu/F(\delta)$. Let $c_l^{(1)} = \int_{B_\delta} |x|^l\ \mu(dx)$. For $j \geq 2$, we consider a bound for the expectation in (23). For $l \leq 1$, we immediately see that

$$E(|\sum_{i=1}^{j} X_i|^l \leq E\sum_{i=1}^{j} |X_i|^l = jE|X_i|^l = \frac{j c_l^{(1)}}{F(\delta)}\ . \tag{24}$$

For $l > 1$, using Jensen's inequality, we have

$$E|\sum_{i=1}^{j} X_i|^l \leq j^{l-1} E\sum_{i=1}^{j} |X_i|^l = \frac{j^l c_l^{(1)}}{F(\delta)}\ . \tag{25}$$

Hence, in either (24) or (25), we have from (23), separating the terms for $j = 1$ and $j > 1$,

$$E|Y_\varepsilon^{(1)}|^l = \varepsilon\, c_l^{(1)} \exp(-\varepsilon F(\delta)) + R_\varepsilon^{(1)} \;\; , \tag{26}$$

where

$$\begin{aligned} R_\varepsilon^{(1)} &\le \sum_{j=1}^{\infty} \exp(-\varepsilon F(\delta)) \frac{\varepsilon^j F(\delta)^j}{j!} \frac{j^{(l\wedge 1)} c_l^{(1)}}{F(\delta)} \\ &= \varepsilon^2 c_l^{(1)} F(\delta) \sum_{k=0}^{\infty} \exp(-\varepsilon F(\delta)) \frac{\varepsilon^k F(\delta)^k}{k!} \frac{(k+2)^{(l-1)_+}}{(k+1)} \;\; . \end{aligned} \tag{27}$$

As $\varepsilon \to 0$, the sum in (27) is a Poisson expectation that converges to $2^{(l-1)_+}$, since $\varepsilon F(\delta) \to 0$. It follows that $R_\varepsilon^{(1)} = o(\varepsilon)$ and hence, from (26), that, as $\varepsilon \to 0$:

$$\varepsilon^{-1} E|Y_\varepsilon^{(1)}|^l \to c_l \;\; , \tag{28}$$

using the facts that $c_l^{(1)} \to c_l$ and $\varepsilon F(\delta) \to 0$.

Now consider the asymptotic behaviour of $Y_\varepsilon^{(2)}$. For $l \le 1$, by Campbell's theorem applied to the Poisson process $S_\varepsilon \bigcap [-\delta, \delta]$, we have

$$E|Y_\varepsilon^{(2)}|^l \le E \left(\sum_{X \in S_\varepsilon \bigcap [-\delta,\delta]} |X| \right)^l \le E \sum_{X \in S_\varepsilon \bigcap [-\delta,\delta]} |X|^l = \varepsilon \int_{-\delta}^{\delta} |x|^l \,\mu(dx) \;\; . \tag{29}$$

Therefore, we have from (29), as $\delta \to 0$, that $E|Y_\varepsilon^{(2)}|^l = o(\varepsilon)$. Since $|Y_\varepsilon|^l \le |Y_\varepsilon^{(1)}|^l + |Y_\varepsilon^{(2)}|^l$, it follows from (28) that $E|Y_\varepsilon|^l = \varepsilon c_l + o(\varepsilon)$ as $\varepsilon \to 0$.

Finally, consider the case $l > 1$. Define $Z_\varepsilon = \sum_{X \in S_\varepsilon \bigcap [-\delta,\delta]} |X|$. For $-\delta \le x \le \delta$, we have $0 \le e^{|x|} - 1 \le \delta^{-1}(e^\delta - 1)|x|$, and so, using (21), $I_\delta = \int_{-\delta}^{\delta} (e^{|x|} - 1)\;\mu(dx) < \infty$. It then follows (using equation (3.17) of Kingman, 1993), that $E \exp(Z_\varepsilon) = \exp(\varepsilon I_\delta)$. Since $l > 1$, there exists a constant k_l such that $z^l \le k_l(e^z - 1)$ for all $z \ge 0$. We then have

$$\begin{aligned} E Z_\varepsilon^l &\le k_l(E \exp(Z_\varepsilon) - 1) \\ &= k_l(\exp(\varepsilon I_\delta) - 1) \\ &= o(\varepsilon) \text{ as } \varepsilon \to 0 \;\; , \end{aligned} \tag{30}$$

since $I_\delta \to 0$ as $\delta \to 0$. Therefore, from (30), it follows at once that $E|Y_\varepsilon^{(2)}|^l = o(\varepsilon)$ as $\varepsilon \to 0$.

Using Minkowski's inequality applied to the norm $||X||_l = (E|X|^l)^{1/l}$ we have:

$$(E|Y_\varepsilon^{(1)}|^l)^{1/l} - (E|Y_\varepsilon^{(2)}|^l)^{1/l} \le (E|Y_\varepsilon|^l)^{1/l} \le (E|Y_\varepsilon^{(1)}|^l)^{1/l} + (E|Y_\varepsilon^{(2)}|^l)^{1/l}$$

and hence, since $\varepsilon^{-1} E|Y_\varepsilon^{(2)}|^l \to 0$, the limiting values of $\varepsilon^{-1} E|Y_\varepsilon^{(1)}|^l$ and $\varepsilon^{-1} E|Y_\varepsilon|^l$ are the same. It follows from (28) that $\varepsilon^{-1} E|Y_\varepsilon|^l \to c_l$, which gives (22), completing the proof of the lemma.

Corollary 1. *Let μ be a measure on a set Ω, and let g be a real-valued function on Ω. Let S_ε be a Poisson Process on Ω with intensity measure $\varepsilon\mu$, where $\varepsilon > 0$. Assume that the induced measure $\mu(g^{-1}(A))$ is non-atomic for every measurable set $A \subseteq \mathbb{R}$, and assume that*

$$\int_\Omega \min(1, |g(x)|)\ \mu(dx) < \infty \quad \textit{and}$$

$$c_l = \int_\Omega |g(x)|^l\ \mu(dx) < \infty \quad \textit{for some } l > 0\ .$$

Define $Y_\varepsilon = \sum_{X \in S_\varepsilon} g(X)$. Then

$$E|Y_\varepsilon|^l = \varepsilon\, c_l + o(\varepsilon) \quad \textit{as} \quad \varepsilon \to 0\ .$$

Proof. Define a measure on $\mathbb{R}$ by $\mu_g(A) = \mu(g^{-1}(A))$ and let $Z = g(X)$. Then, appealing to the Mapping theorem for Poisson processes (see, for example, Kingman, 1993, p. 18), Z is a Poisson process on $\mathbb{R}$ with intensity measure $\varepsilon\mu_g$ and $Y_\varepsilon = \sum_{X \in S_\varepsilon} Z$. The proof of the corollary is completed by applying Lemma 1 to Z and μ_g.

References

1. Abramovich, F., Bailey, T.C., Sapatinas, T.: Wavelet analysis and its statistical applications. The Statistician, **49**, 1–29 (2000)
2. Abramovich, F., Sapatinas, T., Silverman, B.W.: Wavelet thresholding via a Bayesian approach. J. R. Statist. Soc. B, **60**, 725–749 (1998)
3. Antoniadis, A.: Wavelets in statistics: a review (with discussion). J. It. Statist. Soc., **6**, 000–000 (1999)
4. Coifman, R.R., Donoho, D.L.: Translation-invariant de-noising. In: Antoniadis, A., Oppenheim, G. (Eds.), Wavelets and Statistics, Lecture Notes in Statistics **103**, pp. 125–150, New York: Springer-Verlag, 1995
5. Daubechies, I.: Ten Lectures on Wavelets. Philadelphia: SIAM, 1992
6. Donoho, D.L., Johnstone, I.M.: Minimax estimation via wavelet shrinkage. Ann. Statist., **26**, 879–921 (1998)
7. Johnstone, I.M., Silverman, B.W.: Wavelet threshold estimators for data with correlated noise. J. R. Statist. Soc. B, **59**, 319–351 (1997)
8. Härdle, W., Kerkyacharian, G., Picard, D., Tsybakov, A.: Wavelets, Approximation, and Statistical Applications. Lecture Notes in Statistics, **129**, New York: Springer-Verlag, 1998
9. Kingman, J.F.C.: Poisson Processes. Oxford: Oxford University Press, 1993
10. Meyer, Y.: Wavelets and Operators. Cambridge: Cambridge University Press, 1992
11. Nason, G.P., Silverman, B.W.: The stationary wavelet transform and some statistical applications. In: Antoniadis, A., Oppenheim, G. (Eds.), Wavelets and Statistics, Lecture Notes in Statistics **103**, pp. 281–300, New York: Springer-Verlag, 1995
12. Ogden, T.R.: Essential Wavelets for Statistical Applications and Data Analysis. Boston: Birkhäuser, 1997
13. Silverman, B.W.: Wavelets in statistics: beyond the standard assumptions. Phil. Trans. Roy. Soc. Lond. A, **357**, 2459–2474 (1999)
14. Vidakovic, B.: Statistical Modeling by Wavelets. New York: John Wiley & Sons, 1999