IMA Journal of Numerical Analysis (2012) **32**, 294–319
doi:10.1093/imanum/drr018
Advance Access publication on June 22, 2011

Representation of Gaussian fields in series with independent coefficients

CLAUDE JEFFREY GITTELSON
Department of Mathematics, Seminar for Applied Mathematics, Eidgenössische Technische Hochschule Zürich,
Rämistrasse 101, CH-8092 Zurich, Switzerland
claude.gittelson@sam.math.ethz.ch

[Received on 9 June 2010; revised on 16 November 2010]

The numerical discretization of problems with stochastic data or stochastic parameters generally involves the introduction of coordinates that describe the stochastic behaviour, such as coefficients in a series expansion or values at discrete points. The series expansion of a Gaussian field with respect to any orthonormal basis of its Cameron–Martin space has independent standard normal coefficients. A standard choice for numerical simulations is the Karhunen–Loève series, which is based on eigenfunctions of the covariance operator. We suggest an alternative, the hierarchic discrete spectral expansion, which can be constructed directly from the covariance kernel. The resulting basis functions are often well localized, and the convergence of the series expansion seems to be comparable to that of the Karhunen–Loève series. We provide explicit formulas for particular cases and general numerical methods for computing exact representations of such bases. Finally, we relate our approach to numerical discretizations based on replacing a random field by its values on a finite set.

Keywords: Gaussian random fields; simulation; Karhunen–Loève series; Cameron–Martin space; reproducing kernel Hilbert space; covariance kernel; representation of random fields.

1. Introduction

The numerical discretization of problems with stochastic data or stochastic parameters requires that the random inputs are approximated by finite quantities. This is generally done in one of two ways. Either the random data are expanded in a series, which can be truncated for numerical computations (see, e.g. Babuška *et al.*, 2004; Frauenfelder *et al.*, 2005; Todor & Schwab, 2007; Nobile *et al.*, 2008; Bieri *et al.*, 2009; Wan & Karniadakis, 2009), or it is replaced by a finite-dimensional random variable, describing, for example, the value of a random field on a discrete set of points, or a projection of the field onto a finite-dimensional function space (see, e.g. Dietrich & Newsam, 1997; Matthies & Keese, 2005; Graham *et al.*, 2010).

A standard approach, falling strictly into the first category, is to expand a random field into its Karhunen–Loève series (see, e.g. Schwab & Todor, 2006). For Gaussian fields, the coefficients in this series are independent standard normal random variables.

The independence of these coefficients is crucial to many numerical methods. For example, in Monte Carlo simulation coefficient sequences can be generated by independent draws of pseudorandom numbers. The construction of polynomial chaos bases as tensor products of orthonormal bases with respect to the distributions of the coefficients also requires that these are independent. Similarly, in collocation and quasi-Monte Carlo methods constructions of collocation points make use of the product structure of the joint distribution of the coefficients.

Nevertheless, the Karhunen–Loève series is often ill suited for numerical computations, as it requires eigenfunctions of the covariance operator. These are usually not known exactly and

are expensive to approximate numerically. Furthermore, the eigenfunctions generally have global supports.

We suggest an alternative to the Karhunen–Loève series for general continuous Gaussian fields on bounded domains, which we call the *hierarchic discrete spectral expansion*. Assuming that the covariance kernel is given, the basis functions in our series expansion can be constructed exactly. As these form an orthonormal basis of the Cameron–Martin space, independence of the coefficients in the series expansion is preserved.

Our method does not assume any particular structure of the covariance kernel or of the underlying domain. Alternative methods based on the fast Fourier transform and wavelet expansions exist for stationary Gaussian fields (Ripley, 1987; Elliott & Majda, 1994; Elliott *et al.*, 1997; Lang & Potthoff, 2009) sometimes with further restrictions on the domain and the covariance.

In Section 2 we provide a brief overview of the theory of Gaussian measures on Banach spaces and lay the theoretical foundation for our basis in Proposition 2.9. We refer to Vakhania *et al.* (1987, Chapter III) and Bogachev (1998, Chapters 2, 3) for further details.

We consider the case of an exponential covariance in one dimension in Section 3. For this setting, we derive explicit formulas for a sequence of basis functions. In Section 3.3 we apply our approach to the Wiener measure, giving an alternative derivation of the well-known piecewise linear construction of Brownian motion.

The numerical construction of a hierarchic sequence of basis functions is discussed in Section 4. We consider in particular two special cases of our general algorithm. Gram–Schmidt orthogonalization in the Cameron–Martin space is closely related to the Cholesky decomposition of the covariance matrix. Hierarchic spectral decomposition of the covariance matrix can also be used to exactly construct an orthonormal basis of the Cameron–Martin space, providing an alternative interpretation of the naive approximation of the Karhunen–Loève series given by eigenvectors and eigenvalues of the covariance matrix.

Finally, in Section 5 we provide numerical examples of hierarchic discrete spectral bases computed by the two aforementioned special cases of our algorithm for several covariance kernels. We study in particular the decay of these basis functions.

2. The Cameron–Martin space of a Gaussian distribution

2.1 *Gaussian measures on Banach spaces*

Let X be a real separable Banach space with Borel σ-algebra $\mathscr{B}(X)$ and let γ be a Gaussian measure on $(X, \mathscr{B}(X))$, i.e. for all $\varphi \in X^*$, the image measure $\varphi(\gamma)$ on $\mathbb{R}$ is Gaussian. By Bogachev (1998, Theorem 3.2.3) there is a unique element $a_\gamma \in X$, called the *mean* of γ, such that

$$\varphi(a_\gamma) = \int_X \varphi(h)\,\mathrm{d}\gamma(h) \quad \forall\, \varphi \in X^*\,. \tag{2.1}$$

The *covariance operator* R_γ is given formally by

$$\langle R_\gamma \varphi, \psi\rangle = \int_X \varphi(h - a_\gamma)\,\psi(h - a_\gamma)\,\mathrm{d}\gamma(h) \quad \forall\, \varphi, \psi \in X^*. \tag{2.2}$$

Again by Bogachev (1998, Theorem 3.2.3) (2.2) defines a unique linear operator $R_\gamma\colon X^* \to X$. We define

$$\sigma(\varphi) := \sqrt{\langle R_\gamma \varphi, \varphi\rangle}, \quad \varphi \in X^*\,. \tag{2.3}$$

LEMMA 2.1 The assignment $\|h\| := \sigma(\varphi)$ if $h = R_\gamma \varphi$ for $\varphi \in X^*$ defines a norm on range$(R_\gamma) \subset X$.

Proof. Let $\varphi, \psi \in X^*$ with $R_\gamma \varphi = R_\gamma \psi$. Using the symmetry of $\langle R_\gamma \cdot, \cdot \rangle$,

$$\langle R_\gamma \varphi, \varphi \rangle = \langle R_\gamma \psi, \varphi \rangle = \langle R_\gamma \varphi, \psi \rangle = \langle R_\gamma \psi, \psi \rangle \ .$$

Therefore, $\|R_\gamma \varphi\|$ is well defined on range(R_γ). It is a norm since it coincides with $\|\varphi(\cdot - a_\gamma)\|_{L^2(\gamma)}$. □

We define the *Cameron–Martin space* $H(\gamma)$ as the completion of range(R_γ) with respect to the norm from Lemma 2.1. By Bogachev (1998, Theorem 3.2.7) $H(\gamma)$ is a separable Hilbert space, and due to Bogachev (1998, Proposition 2.4.6) $H(\gamma)$ embeds continuously into X. Furthermore, by Bogachev (1998, Lemma 3.2.2) $H(\gamma)$ is independent of the space X on which γ is considered.

LEMMA 2.2 For all $\varphi \in X^*$ and all $f \in H(\gamma)$,

$$(f, R_\gamma \varphi)_{H(\gamma)} = \varphi(f). \tag{2.4}$$

Proof. Let $f = R_\gamma \psi$ for $\psi \in X^*$. Then using the parallelogram identity,

$$(f, R_\gamma \varphi)_{H(\gamma)} = (R_\gamma \psi, R_\gamma \varphi)_{H(\gamma)} = \langle R_\gamma \psi, \varphi \rangle = \varphi(f) \ .$$

This extends to all $f \in H(\gamma)$ by density. □

LEMMA 2.3 Let $\varphi \in X^*$ and let $(\varphi_n)_{n\in\mathbb{N}} \subset X^*$ be bounded with $\varphi_n(h) \to \varphi(h)$ for all $h \in X$. Then $R_\gamma \varphi_n \to R_\gamma \varphi$ in $H(\gamma)$.

Proof. We assume without loss of generality that $\varphi = 0$. By definition

$$\sigma(\varphi_n)^2 = \langle R_\gamma \varphi_n, \varphi_n \rangle = \int_X \varphi_n(h - a_\gamma)^2 \, \mathrm{d}\gamma(h) \leqslant \|\varphi_n\|_{X^*}^2 \int_X \|h - a_\gamma\|_X^2 \, \mathrm{d}\gamma(h) \ .$$

The last integral is finite as a consequence of Fernique's theorem (Fernique, 1970); see also Bogachev (1998, Theorem 2.8.5). Therefore, by dominated convergence, $\sigma(\varphi_n) \to \sigma(\varphi) = 0$ and thus $R_\gamma \varphi_n \to 0$ in $H(\gamma)$. □

The conditions of Lemma 2.3 are satisfied, e.g. if $\varphi_n \rightharpoonup \varphi$, since weakly convergent sequences are bounded. Furthermore, the statement of Lemma 2.3 implies in particular that $R_\gamma \varphi_n \to R_\gamma \varphi$ in X.

2.2 *Orthonormal bases of the Cameron–Martin space*

The Gaussian measure γ on X is uniquely characterized by its mean a_γ and its covariance operator R_γ. The covariance operator, in turn, is determined by the Cameron–Martin space $H(\gamma)$. It can be expressed in terms of an orthonormal basis of $H(\gamma)$.

PROPOSITION 2.4 Let $(e_m)_{m\in\Theta}$ be an orthonormal basis of $H(\gamma)$. Then,

$$R_\gamma \varphi = \sum_{m\in\Theta} \varphi(e_m) e_m \quad \forall \varphi \in X^* \tag{2.5}$$

with unconditional convergence in $H(\gamma)$ and in X. Furthermore,

$$\langle R_\gamma \varphi, \psi \rangle = \sum_{m\in\Theta} \varphi(e_m)\psi(e_m) \quad \forall \varphi, \psi \in X^* \tag{2.6}$$

with unconditional convergence in $\mathbb{R}$.

Proof. Let $\varphi \in X^*$. By Lemma 2.2,

$$(R_\gamma \varphi, e_m)_{H(\gamma)} = \varphi(e_m) \quad \forall m \in \Theta \ .$$

Expanding $R_\gamma \varphi$ in the orthonormal basis $(e_m)_{m\in\Theta}$ of $H(\gamma)$ we have

$$R_\gamma \varphi = \sum_{m\in\Theta} (R_\gamma \varphi, e_m)_{H(\gamma)} e_m = \sum_{m\in\Theta} \varphi(e_m) e_m$$

with convergence in $H(\gamma)$, and thus also in X since $H(\gamma)$ embeds continuously into X. The convergence is unconditional since all the above is independent of any ordering of the basis $(e_m)_{m\in\Theta}$. In particular, for any $\psi \in X^*$,

$$\langle R_\gamma \varphi, \psi \rangle = \psi(R_\gamma \varphi) = \psi\left(\sum_{m\in\Theta} \varphi(e_m) e_m\right) = \sum_{m\in\Theta} \varphi(e_m)\psi(e_m) \ ,$$

again with unconditional convergence. □

An orthonormal basis $(e_m)_{m\in\Theta}$ of $H(\gamma)$ is useful not only for computing the covariance R_γ but also for sampling the distribution γ on X.

We define the product Gaussian measure π on $(\mathbb{R}^\Theta, \mathscr{B}(\mathbb{R}^\Theta))$,

$$\pi := \bigotimes_{m\in\Theta} \pi_m \ , \tag{2.7}$$

where each π_m is a standard Gaussian measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. In principle, π can be sampled numerically by independent standard Gaussian draws for each index $m \in \Theta$. If, in some sense, these values are of decreasing importance, then the sequence can be truncated to finitely many random values. Using an orthonormal basis $(e_m)_{m\in\Theta}$ of $H(\gamma)$ we parameterize γ by π, allowing γ to be sampled by mapping a sample of π from $\mathbb{R}^\Theta$ to X. Also, series expansions of the form (2.8) are a prerequisite for many stochastic Galerkin and collocation methods.

THEOREM 2.5 Let $(e_m)_{m\in\Theta}$ be an orthonormal basis of $H(\gamma)$. Then the series in

$$T\colon \mathbb{R}^\Theta \to X, \quad \xi = (\xi_m)_{m\in\Theta} \mapsto a_\gamma + \sum_{m\in\Theta} \xi_m e_m \tag{2.8}$$

converges unconditionally in X for π-a.e. $\xi = (\xi_m)_{m\in\Theta}$ in $\mathbb{R}^\Theta$, and the distribution of T is $T(\pi) = \gamma$.

For a proof of Theorem 2.5 we refer to Bogachev (1998, Theorem 3.5.1) (see also Jain & Kallianpur, 1970; Kuelbs, 1971; Page, 1972; Vakhania *et al.*, 1987, Section V.5.5). Due to the product structure of π, $\xi = (\xi_m)_{m\in\Theta}$ are independent standard Gaussian random variables.

REMARK 2.6 Theorem 2.5 applies directly to R_γ-Wiener processes. For any orthonormal basis $(e_m)_{m\in\Theta}$ of $H(\gamma)$ let $(\beta_t^m)_{m\in\Theta}$ be independent scalar Wiener processes. Then an R_γ-Wiener process W_t in X is given by

$$W_t = \sum_{m\in\Theta} \beta_t^m e_m \ , \tag{2.9}$$

with unconditional convergence in X a.s. for all $t \geqslant 0$. Convergence of this series follows by applying Theorem 2.5 for every $t > 0$. Equation (2.9) defines an $H(\gamma)$-Wiener process since for all $t > s$,

$$W_t - W_s = \sum_{m \in \Theta} (\beta_t^m - \beta_s^m) e_m = \sqrt{t-s} \sum_{m \in \Theta} \xi_m e_m \tag{2.10}$$

with independent standard normal $(\xi_m)_{m \in \Theta}$, and the last sum represents an X-valued centred Gaussian random variable with covariance R_γ. In particular, if $t_n \to t$, then (2.10) with $s = t_n$ implies that W_t defined by (2.9) has continuous paths.

REMARK 2.7 The series representation (2.8) also allows conditional simulation of the distribution γ. For a finite set $\Theta_n \subset \Theta$ the distribution of the series

$$a_\gamma + \sum_{m \in \Theta_n} y_m e_m + \sum_{m \in \Theta \setminus \Theta_n} \xi_m e_m \tag{2.11}$$

serves as a conditional probability of γ, conditioned on $\xi_m = y_m$ for $m \in \Theta_n$, where $(\xi_m)_{m \in \Theta \setminus \Theta_n}$ are independent standard normal random variables (see Bogachev, 1998, Corollary 3.5.2). In particular, if γ describes a prior model for a random element of X and y_m are measurements of ξ_m for $m \in \Theta_n$, then (2.11) is the resulting posterior model. This is the foundation for the interpolation techniques known in geostatistics as kriging (see, e.g. Stein, 1999; Diggle & Ribeiro, 2007).

2.3 *Continuous Gaussian fields*

We consider the case $X = C(D)$ for a compact set $D \subset \mathbb{R}^d$. Then γ describes a Gaussian field on D that is as above continuous.

For all $x \in D$ the Dirac functional $\delta_x(f) := f(x)$ is in X^*. We define

$$k_x := R_\gamma \delta_x, \quad x \in D. \tag{2.12}$$

Then the *covariance kernel* of γ is

$$k(x, y) := k_x(y) = \langle R_\gamma \delta_x, \delta_y \rangle, \quad x, y \in D. \tag{2.13}$$

The function $k(\cdot, \cdot)$ is symmetric since

$$k(x, y) = \langle R_\gamma \delta_x, \delta_y \rangle = \langle R_\gamma \delta_y, \delta_x \rangle = k(y, x) \quad \forall x, y \in D. \tag{2.14}$$

By the Riesz representation theorem for $X = C(D)$, X^* can be identified with the space of signed measures μ on the Borel σ-algebra of D with the total variation norm. Accordingly, we will use the same symbol, e.g. μ, for the signed measure on $(D, \mathscr{B}(D))$ and for the element of X^* given by integration against this measure.

PROPOSITION 2.8 Let μ and ν be signed measures on $(D, \mathscr{B}(D))$. Then

$$(R_\gamma \mu)(x) = \int_D k_x \, \mathrm{d}\mu \quad \forall x \in D \tag{2.15}$$

and

$$\langle R_\gamma \mu, \nu \rangle = \int_D \int_D k(x, y) \, \mathrm{d}\mu(y) \, \mathrm{d}\nu(x). \tag{2.16}$$

Proof. For all $x \in D$, by symmetry of $\langle R_\gamma \cdot, \cdot \rangle$,

$$(R_\gamma \mu)(x) = \langle R_\gamma \mu, \delta_x \rangle = \langle R_\gamma \delta_x, \mu \rangle = \int_D k_x \, \mathrm{d}\mu.$$

Consequently,

$$\langle R_\gamma \mu, \nu \rangle = \int_D (R_\gamma \mu)(x) \, \mathrm{d}\nu(x) = \int_D \int_D k(x, y) \, \mathrm{d}\mu(y) \, \mathrm{d}\nu(x). \qquad \square$$

In particular, if $\mathrm{d}\mu(x) = f(x)\,\mathrm{d}x$ and $\mathrm{d}\nu(x) = g(x)\,\mathrm{d}x$ for functions f and g in $L^1(D)$ we recover the well-known formula describing the covariance operator R_γ as integration against $k(\cdot, \cdot)$,

$$\langle R_\gamma f, g \rangle = \int_D \int_D k(x, y) f(y) g(x) \, \mathrm{d}y \, \mathrm{d}x. \tag{2.17}$$

Here f and g are interpreted as the elements of X^* given by integration against f and g, respectively.

PROPOSITION 2.9 Let $(p_i)_{i \in \Lambda}$ be a dense subset of D. Then the linear span of $(k_{p_i})_{i \in \Lambda}$ is dense in $H(\gamma)$.

Proof. Let $f \in H(\gamma)$ be orthogonal to k_{p_i} for all $i \in \Lambda$. Lemma 2.2 implies

$$f(p_i) = \delta_{p_i}(f) = (f, R_\gamma \delta_{p_i})_{H(\gamma)} = (f, k_{p_i})_{H(\gamma)} = 0 \quad \forall\, i \in \Lambda.$$

Since $(p_i)_{i \in \Lambda}$ is dense in D by assumption and f is continuous due to $H(\gamma) \subset X = C(D)$ it follows that $f = 0$. $\square$

REMARK 2.10 Proposition 2.9 suggests a construction for an orthonormal basis of $H(\gamma)$. Given a dense sequence $(p_i)_{i \in \mathbb{N}}$ in D we can apply Gram–Schmidt orthogonalization to $(k_{p_i})_{i \in \mathbb{N}}$. This is explored in Section 4.2. Note that by Lemma 2.2, $f \in H(\gamma)$ is orthogonal to k_x for an $x \in D$ if and only if $f(x) = 0$ since

$$(f, k_x)_{H(\gamma)} = (f, R_\gamma \delta_x)_{H(\gamma)} = \delta_x(f) = f(x).$$

Therefore, constructing an orthonormal basis of $H(\gamma)$ can reduce to finding functions in the span of $(k_{p_i})_{i=1}^n$ with certain zeros.

For any sequence $p = (p_i)_{i \in \Lambda}$ in D let K denote the covariance matrix of the functionals $(\delta_{p_i})_{i \in \Lambda}$, i.e.

$$K = [\langle R_\gamma \delta_{p_i}, \delta_{p_j} \rangle]_{i,j \in \Lambda} = [k(p_i, p_j)]_{i,j \in \Lambda}. \tag{2.18}$$

For a finitely supported vector $a = (a_i)_{i \in \Lambda} \in \mathbb{R}^\Lambda$ we define the functional

$$a\delta_{\boldsymbol{p}} := \sum_{i \in \Lambda} a_i \delta_{p_i} \in C(D)^* \tag{2.19}$$

and the function

$$ak_{\boldsymbol{p}} := R_\gamma a\delta_{\boldsymbol{p}} = \sum_{i \in \Lambda} a_i k_{p_i} \in H(\gamma) \subset C(D). \tag{2.20}$$

Note that no convergence issues emerge since by assumption, $a_i = 0$ for all but finitely many $i \in \Lambda$. Then, using Lemma 2.2, for all finitely supported a and b in $\mathbb{R}^\Lambda$,

$$(ak_{\boldsymbol{p}}, bk_{\boldsymbol{p}})_{H(\gamma)} = \langle R_\gamma a\delta_{\boldsymbol{p}}, b\delta_{\boldsymbol{p}}\rangle = a^{\mathrm{T}} K b\ . \tag{2.21}$$

Therefore, orthogonality of the functions $ak_{\boldsymbol{p}}$ and $bk_{\boldsymbol{p}}$ in $H(\gamma)$, which is equivalent to the functionals $a\delta_{\boldsymbol{p}}$ and $b\delta_{\boldsymbol{p}}$ being uncorrelated, is characterized by orthogonality of the vectors a and b with respect to K.

COROLLARY 2.11 Let $p = (p_i)_{i\in\Lambda}$ be a sequence in D, and let $(a^m)_{m\in\Theta} \subset \mathbb{R}^\Lambda$ such that a^m is finitely supported for all $m \in \Theta$ and

$$(a^m)^{\mathrm{T}} K a^{m'} = \delta_{mm'} \quad \forall m, m' \in \Theta\ . \tag{2.22}$$

Then $(a^m k_{\boldsymbol{p}})_{m\in\Theta}$ is an orthonormal system in $H(\gamma)$. If p is dense in D and for all $i \in \Lambda$, k_{p_i} is in the span of $(a^m k_{\boldsymbol{p}})_{m\in\Theta}$ in $H(\gamma)$, then $(a^m k_{\boldsymbol{p}})_{m\in\Theta}$ is an orthonormal basis of $H(\gamma)$.

Proof. Orthonormality of $(a^m k_{\boldsymbol{p}})_{m\in\Theta}$ follows from (2.22) due to (2.21). Density is a consequence of Proposition 2.9. □

REMARK 2.12 The assumption that coefficient vectors are finitely supported can be weakened to a more general summability condition. In the interest of a concise presentation, and since all numerically representable coefficient vectors are finitely supported, we consider only this setting.

3. Examples

3.1 *Exponential covariance in one dimension*

Let $I \subset \mathbb{R}$ be a compact interval. We consider a Gaussian measure γ on $C(I)$ with an exponential covariance kernel

$$k(x, y) = \sigma^2 e^{-|x-y|/\lambda}, \quad x, y \in I, \tag{3.1}$$

for constants $\sigma > 0$ and $\lambda > 0$.

We define a hierarchic sequence of grids on I. For all $\ell \in \mathbb{N}$, let $\Delta_\ell \subset I$ be a finite set with $\Delta_\ell \cap \Delta_{\ell'} = \varnothing$ if $\ell \neq \ell'$. We define the unions

$$\Lambda_\ell := \bigsqcup_{j=1}^{\ell} \Delta_j \quad \text{and} \quad p := \Lambda := \bigsqcup_{\ell=1}^{\infty} \Delta_\ell = \bigcup_{\ell=1}^{\infty} \Lambda_\ell. \tag{3.2}$$

Then $n_\ell := \#\Lambda_\ell < \infty$. For all $\ell \in \mathbb{N}$, we enumerate the elements of Λ_ℓ in increasing order, $\Lambda_\ell = \{p_1^\ell, \dots, p_{n_\ell}^\ell\}$ with $p_i^\ell < p_{i+1}^\ell$. We assume that elements of Δ_ℓ are never adjacent in this ordering.

ASSUMPTION 3.1 For all $\ell \in \mathbb{N}$, if $p_i^\ell, p_j^\ell \in \Delta_\ell$ and $i \neq j$, then $|i - j| \geqslant 2$.

Assumption 3.1 implies in particular

$$n_1 \leqslant 1 \quad \text{and} \quad n_{\ell+1} \leqslant 2n_\ell + 1 \quad \forall \ell \in \mathbb{N}\ . \tag{3.3}$$

It is always satisfied if $\#\Delta_\ell = 1$ for all $\ell \in \mathbb{N}$.

To each $p \in \Lambda$ we associate a function $a^p k_{\boldsymbol{p}}$ in $H(\gamma)$ with $a^p \in \mathbb{R}^\Lambda$ finitely supported. Let $\ell \in \mathbb{N}$ and $p \in \Lambda_\ell$ such that $p = p_i^\ell$ in the above ordering. We define the coefficient vector a^p by allowing a_q^p to be different from zero only if q is adjacent to p in Λ_ℓ, and requiring that $a^p k_{\boldsymbol{p}}$ is zero at all such adjacent nodes q. Finally, we normalize the coefficients such that $a^p k_{\boldsymbol{p}}$ has unit norm in $H(\gamma)$.

If i is different from 1 and n_ℓ, i.e. if p is an interior point of Λ_ℓ, this results in a 3×3 linear system with the following solution. For $h_+ := (p_{i+1}^\ell - p_i^\ell)/\lambda$ and $h_- := (p_i^\ell - p_{i-1}^\ell)/\lambda$ let

$$\tilde{a}_0 := \frac{e^{2h_-+2h_+} - 1}{(e^{2h_-} - 1)(e^{2h_+} - 1)}, \quad \tilde{a}_{-1} := \frac{-e^{h_-}}{e^{2h_-} - 1}, \quad \tilde{a}_1 := \frac{-e^{h_+}}{e^{2h_+} - 1} \tag{3.4}$$

and define $a^p = (a_q^p)_{q \in \Lambda}$ by

$$a^p_{p_j^\ell} := \frac{\tilde{a}_{j-i}}{\sigma\sqrt{\tilde{a}_0}}, \quad j \in \{i-1, i, i+1\}, \tag{3.5}$$

and $a_q^p := 0$ for all other $q \in \Lambda$.

If p is a boundary point of Λ_ℓ and $n_\ell \geqslant 2$ the above conditions lead to a 2×2 linear system. If $i = 1$, then $h := (p_2^\ell - p_1^\ell)/\lambda$ and

$$\tilde{a}_0 := 1, \quad \tilde{a}_1 := -e^{-h} \quad \text{and} \quad a^p_{p_j^\ell} := \frac{\tilde{a}_{j-1}}{\sigma\sqrt{1 - e^{-2h}}}. \tag{3.6}$$

Similarly, if $i = n_\ell$, we set $h := (p_{n_\ell}^\ell - p_{n_\ell - 1}^\ell)/\lambda$ and

$$\tilde{a}_0 := 1\,, \quad \tilde{a}_{-1} := -e^{-h} \quad \text{and} \quad a^p_{p_j^\ell} := \frac{\tilde{a}_{j-n_\ell}}{\sigma\sqrt{1 - e^{-2h}}}\,. \tag{3.7}$$

In both cases, $a_q^p := 0$ for all other $q \in \Lambda$. Finally, if p is the only element of Λ_ℓ, then $\tilde{a}_0 := 1$, $a_p^p := 1/\sigma$, and $a_q^p := 0$ for all $q \neq p$. Some basis functions are plotted in Fig. 1.

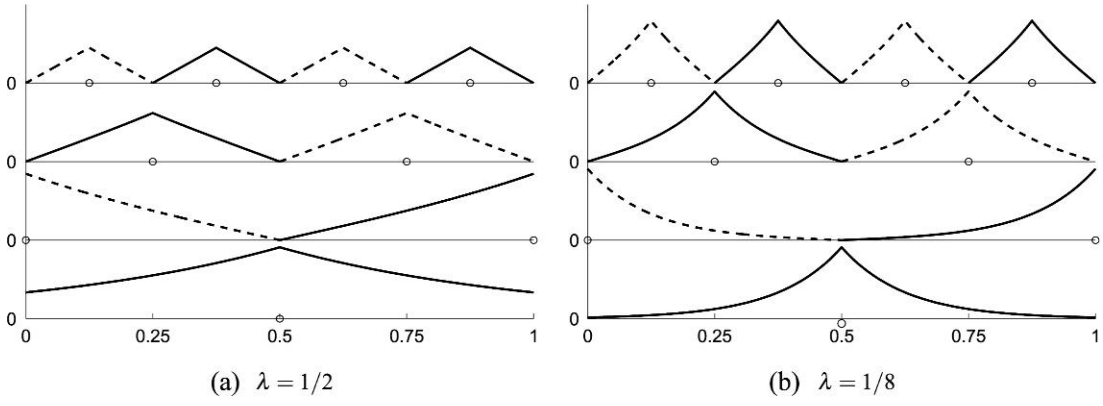


FIG. 1. The first few functions $a^p k_{\boldsymbol{p}}$ on $I = [0, 1]$ constructed on dyadic grids (circles).

PROPOSITION 3.2 Let $\ell \in \mathbb{N}$ with $n_\ell \geqslant 2$. Then for all $p = p_i^\ell \in \Lambda_\ell$ the support of $a^p k_{\boldsymbol{p}}$ in I is

$$\operatorname{supp} a^p k_{\boldsymbol{p}} = \big[p_{i-1}^\ell, p_{i+1}^\ell\big], \tag{3.8}$$

where we set $p_0^\ell := \min(I)$ and $p_{n_\ell+1}^\ell := \max(I)$.

Proof. We first consider an interior point $p = p_i^\ell$ of Λ_ℓ. Any $x \geqslant p_{i+1}^\ell$ is of the form $x = p_i^\ell + t\lambda$ for some $t \geqslant h_+ = (p_{i+1}^\ell - p_i^\ell)/\lambda$. Then using the notation from (3.4), for some $c > 0$,

$$\begin{aligned} c(a^p k_{\boldsymbol{p}})(x) &= \tilde{a}_{-1}\, e^{-(t+h_-)} + \tilde{a}_0\, e^{-t} + \tilde{a}_1\, e^{-(t-h_+)} = \frac{-e^{-t}}{e^{2h_-}-1} + \frac{e^{-t}\left(e^{2h_-+2h_+}-1\right)}{(e^{2h_-}-1)(e^{2h_+}-1)} + \frac{-e^{-t}\,e^{2h_+}}{e^{2h_+}-1} \\ &= \frac{e^{-t}\big(-e^{2h_+}+1+e^{2h_-+2h_+}-1-e^{2h_-+2h_+}+e^{2h_+}\big)}{(e^{2h_-}-1)(e^{2h_+}-1)} = 0\,. \end{aligned}$$

This implies $a^p k_{\boldsymbol{p}}(x) = 0$ for all $x \geqslant p_{i+1}^\ell$. By symmetry, also $a^p k_{\boldsymbol{p}}(x) = 0$ for all $x \leqslant p_{i-1}^\ell$. Similar, but shorter, computations lead to the same result for boundary points. □

LEMMA 3.3 For all $p \in \Lambda$ and all $x \in I$,

$$0 \leqslant (a^p k_{\boldsymbol{p}})(x) \leqslant (a^p k_{\boldsymbol{p}})(p) = \frac{1}{a_p^p}. \tag{3.9}$$

In particular, the maximal value of $a^p k_{\boldsymbol{p}}$ has the following form. If $p \in \Lambda_\ell$ is an interior node of Λ_ℓ, then

$$(a^p k_{\boldsymbol{p}})(p) = \sigma\sqrt{\frac{(e^{2h_-}-1)(e^{2h_+}-1)}{e^{2h_-+2h_+}-1}} \tag{3.10}$$

with h_- and h_+ as above. If p is a boundary node of Λ_ℓ with $n_\ell \geqslant 2$, then $(a^p k_{\boldsymbol{p}})(p) = \sigma\sqrt{1-e^{-2h}}$ with h as above, and if $n_\ell = 1$, then $(a^p k_{\boldsymbol{p}})(p) = \sigma$.

Proof. We first compute $(a^p k_{\boldsymbol{p}})(p)$. Let $p \in \Lambda_\ell$ be an interior node of Λ_ℓ. Then

$$(a^p k_{\boldsymbol{p}})(p) = \frac{1}{\sigma\sqrt{\tilde{a}_0}}\big(\tilde{a}_{-1}\sigma^2\, e^{-h_-} + \tilde{a}_0\sigma^2 + \tilde{a}_1\sigma^2\, e^{-h_+}\big) = \sigma\sqrt{\frac{(e^{2h_-}-1)(e^{2h_+}-1)}{e^{2h_-+2h_+}-1}}\,.$$

If $p \in \Lambda_\ell$ is a boundary node of Λ_ℓ with $n_\ell \geqslant 2$,

$$(a^p k_{\boldsymbol{p}})(p) = \frac{1}{\sigma\sqrt{1-e^{-2h}}}\,\sigma^2 - \frac{e^{-h}}{\sigma\sqrt{1-e^{-2h}}}\,\sigma^2\, e^{-h} = \sigma\sqrt{1-e^{-2h}}\,.$$

Finally, if $n_\ell = 1$, then $a^p k_{\boldsymbol{p}} = \sigma^{-1} k_p$, and therefore $(a^p k_{\boldsymbol{p}})(p) = \sigma$.

To prove (3.9) we show that the derivative of $a^p k_{\boldsymbol{p}}$ is monotonic between elements of Λ_ℓ. It suffices to show this on the interval (p_i^ℓ, p_{i+1}^ℓ) if $p = p_i^\ell$. For interior nodes this is a consequence of

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(\tilde{a}_{-1}\, e^{-(t+h_-)} + \tilde{a}_0\, e^{-t} + \tilde{a}_1\, e^{-(h_+-t)}\big) = -\tilde{a}_{-1}\, e^{-t-h_-} - \tilde{a}_0\, e^{-t} + \tilde{a}_1\, e^{t-h_+} = \frac{-e^{-t}\left(e^{2h_+}+e^{2t}\right)}{e^{2h_+}-1} \leqslant 0$$

for all $0 < t < h_+$. Similar estimates hold for boundary nodes. □

THEOREM 3.4 The functions $(a^p k_{\boldsymbol{p}})_{p\in\Lambda}$ form an orthonormal system in $H(\gamma)$. If Λ is dense in I, then $(a^p k_{\boldsymbol{p}})_{p\in\Lambda}$ is an orthonormal basis of $H(\gamma)$.

Proof. Let $p = p_i^\ell \in \Delta_\ell$ and $q \in \Lambda_\ell \setminus \{p\}$. By Lemma 2.2 and Proposition 3.2, $(a^p k_{\boldsymbol{p}}, k_q)_{H(\gamma)} = (a^p k_{\boldsymbol{p}})(q) = 0$ since q is not in the interior of the support of $a^p k_{\boldsymbol{p}}$.

Let $p' = p_j^\ell$ with $j \neq i$. If $|i-j| \geqslant 2$, then $a^{p'} k_{\boldsymbol{p}}$ is a linear combination of k_q for $q \in \Lambda_\ell \setminus \{p\}$, and therefore $(a^p k_{\boldsymbol{p}}, a^{p'} k_{\boldsymbol{p}})_{H(\gamma)} = 0$. By Assumption 3.1, if $|i-j| = 1$, then $p' \in \Lambda_\ell \setminus \Delta_\ell$, i.e. $p' \in \Delta_{\ell'}$ for some $\ell' \leqslant \ell - 1$. Consequently, $a^{p'} k_{\boldsymbol{p}}$ is a linear combination of k_q for $q \in \Lambda_{\ell'}$, and thus again $(a^p k_{\boldsymbol{p}}, a^{p'} k_{\boldsymbol{p}})_{H(\gamma)} = 0$. Therefore, the functions $(a^p k_{\boldsymbol{p}})_{p\in\Lambda}$ are mutually orthogonal in $H(\gamma)$.

Using Lemma 3.3 and the orthogonality of $a^p k_{\boldsymbol{p}}$ to k_q for all $q \in \Lambda_\ell \setminus \{p\}$,

$$\|a^p k_{\boldsymbol{p}}\|^2_{H(\gamma)} = (a^p k_{\boldsymbol{p}}, a_p^p k_p)_{H(\gamma)} = a_p^p \delta_p(a^p k_{\boldsymbol{p}}) = 1 \,.$$

This shows that the functions $(a^p k_{\boldsymbol{p}})_{p\in\Lambda}$ are normalized in $H(\gamma)$.

Proposition 2.9 implies that if Λ is dense in I, then the linear span of $(k_p)_{p\in\Lambda}$ is dense in $H(\gamma)$. For every ℓ, $(a^p k_{\boldsymbol{p}})_{p\in\Lambda_\ell}$ are n_ℓ linearly independent elements of the span of $(k_q)_{q\in\Lambda_\ell}$. Consequently, for all $q \in \Lambda_\ell$, k_q is a linear combination of $(a^p k_{\boldsymbol{p}})_{p\in\Lambda_\ell}$. □

3.2 *Decay of basis elements*

PROPOSITION 3.5 For all $p \in \Lambda$, if $p \in \Delta_\ell$ is an interior node of Λ_ℓ, then

$$\|a^p k_{\boldsymbol{p}}\|_{C(I)} \leqslant \sigma \min\left(e^{h_- + h_+} \sqrt{\frac{2h_- h_+}{h_- + h_+}} \,,\; 1 \right) \tag{3.11}$$

with h_- and h_+ as above. If p is a boundary node and $n_\ell \geqslant 2$, then $\|a^p k_{\boldsymbol{p}}\|_{C(I)} \leqslant \sigma\sqrt{\min(2h, 1)}$ with h as above, and if $n_\ell = 1$, then $\|a^p k_{\boldsymbol{p}}\|_{C(I)} = \sigma$.

Proof. We use the elementary estimates $1 - e^{-t} \leqslant t \leqslant e^t - 1 \leqslant t\, e^t$ for $t \geqslant 0$. By Lemma 3.3, the maximum of $a^p k_{\boldsymbol{p}}$ is attained at p and is equal to $1/a_p^p$. The first part of (3.11) follows from (3.10) and

$$\frac{\big(e^{2h_-} - 1\big)\big(e^{2h_+} - 1\big)}{e^{2h_- + 2h_+} - 1} \leqslant \frac{\big(2h_-\, e^{2h_-}\big)\big(2h_+\, e^{2h_+}\big)}{2h_- + 2h_+} = e^{2h_- + 2h_+} \frac{2h_- h_+}{h_- + h_+} \,.$$

For boundary nodes and in the case $n_\ell = 1$, the claim is a consequence of Lemma 3.3.

It remains to be shown that (3.10) is bounded by σ. To this end we compute the derivative

$$\frac{\mathrm{d}}{\mathrm{d}h_-} \frac{\big(e^{2h_-} - 1\big)\big(e^{2h_+} - 1\big)}{e^{2h_- + 2h_+} - 1} = \frac{2\, e^{2h_-} \big(e^{2h_+} - 1\big)^2}{(e^{2h_- + 2h_+} - 1)^2} \geqslant 0 \,,$$

and similarly with h_+ and h_- switched. Let $h := h_- + h_+$ and $\vartheta := h_-/h$. Then by monotonicity, for any fixed $\vartheta \in (0, 1)$,

$$\sup_{h\to\infty} \frac{\big(e^{2\vartheta h} - 1\big)\big(e^{2(1-\vartheta)h} - 1\big)}{e^{2h} - 1} = \lim_{h\to\infty} 2\vartheta\big(1 - e^{-2(1-\vartheta)h}\big) + 2(1-\vartheta)\big(1 - e^{-2\vartheta h}\big) = 1,$$

which concludes the proof. □

PROPOSITION 3.6 For all $p \in \Lambda$, if $p \in \Lambda_\ell$ is an interior node of Λ_ℓ, then

$$\|a^p k_{\boldsymbol{p}}\|_{L^2(I)} \leqslant \sigma \min\left(e^{h_- + h_+}\sqrt{2h_- h_+}, \sqrt{h_- + h_+}\right) \tag{3.12}$$

with h_- and h_+ as above. If p is a boundary node and $n_\ell \geqslant 2$, then $\|a^p k_{\boldsymbol{p}}\|_{L^2(I)} \leqslant \sigma \sqrt{\min(2h, 1)(h + h_\partial)}$ with h as above and $h_\partial := (p - \min(I))/\lambda$ if $p = p_1^\ell$ and $h_\partial := (\max(I) - p)/\lambda$ if $p = p_{n_\ell}^\ell$. If $n_\ell = 1$, then $\|a^p k_{\boldsymbol{p}}\|_{L^2(I)} \leqslant \sigma\sqrt{|I|}$.

Proof. The claim follows from Propositions 3.2 and 3.5 using the estimate

$$\|a^p k_{\boldsymbol{p}}\|_{L^2(I)} \leqslant \sqrt{|\operatorname{supp} a^p k_{\boldsymbol{p}}|}\, \|a^p k_{\boldsymbol{p}}\|_{C(I)}. \qquad \square$$

For small correlation lengths λ, i.e. large values of h, h_-, h_+ and h_∂, the estimate in Proposition 3.6 is quite crude but sufficient for our purposes.

EXAMPLE 3.7 Let $I := [0, 1]$, $\Lambda_0 := \{1/2\}$ and $\Delta_\ell := \{i2^{-\ell}\,;\ i = 0, 1, \ldots, 2^\ell\}$ for all $\ell \geqslant 1$, as in Fig. 1. Then for all $\ell \in \mathbb{N}$ and all $p \in \Delta_\ell$, $h = h_- = h_+ = 2^{-\ell}$, and $h_\partial = 0$. In particular, Propositions 3.5 and 3.6 imply

$$\|a^p k_{\boldsymbol{p}}\|_{C(I)} \leqslant C\sigma\,\frac{1}{\sqrt{n_\ell}} \quad \text{and} \quad \|a^p k_{\boldsymbol{p}}\|_{L^2(I)} \leqslant C\sigma\,\frac{1}{n_\ell} \quad \forall p \in \Delta_\ell. \tag{3.13}$$

Thus, the convergence rate in $L^2(I)$ of (2.8) and (2.9) coincides with that of the Karhunen–Loève series. At any given point $x \in [0, 1]$, since only at most two basis functions per level are nonzero, the convergence of the series is exponential if all other basis functions are disregarded.

3.3 *The Wiener measure*

The same approach as in Section 3.1 can be used to construct an orthonormal basis of the Cameron–Martin space of the Wiener measure. Let $I = [0, 1]$ and

$$k(x, y) := \min(x, y), \quad x, y \in I. \tag{3.14}$$

Furthermore, let $\Lambda_\ell = \{p_1^\ell, \ldots, p_{n_\ell}^\ell\}$, $\Lambda_\ell \uparrow \Lambda =: p$, be a hierarchic sequence of grids satisfying Assumption 3.1, and not containing 0.

Proceeding as in the case of an exponential covariance kernel we construct for each $p \in \Lambda$ a function $a^p k_{\boldsymbol{p}}$ that is a linear combination of k_q for at most three $q \in \Lambda$. Since k_q is piecewise linear for all $q \in \Lambda$, the orthogonality conditions on $a^p k_{\boldsymbol{p}}$ imply that $a^p k_{\boldsymbol{p}}$ is a multiple of a hat function.

For each $p \in \Lambda$ we define a hat function ζ_p as follows. For $p \in \Delta_\ell$, ζ_p is the piecewise linear function on I subordinate to the nodes $\Lambda_\ell \cup \{0, 1\}$ such that $\zeta_p(q) = \delta_{pq}$ for all $q \in \Lambda_\ell$, $\zeta_p(0) = 0$ and $\zeta_p(1) = \zeta_p(p_{n_\ell}^\ell)$.

PROPOSITION 3.8 For all $p \in \Lambda$, $a^p k_{\boldsymbol{p}} = \zeta_p / a_p^p$.

Proof. We consider the case $p = p_i^\ell \in \Delta_\ell$ with $2 \leqslant i \leqslant n_\ell - 1$. The case of boundary points is similar. Then, since $a^p k_{\boldsymbol{p}}$ is piecewise linear, the condition $0 = (a^p k_{\boldsymbol{p}}, k_q)_{H(\gamma)} = (a^p k_{\boldsymbol{p}})(q)$ for $q = p_{i\pm1}^\ell$ implies that $a^p k_{\boldsymbol{p}}$ is a multiple of ζ_p. The third defining condition,

$$1 = \|a^p k_{\boldsymbol{p}}\|_{H(\gamma)}^2 = (a^p k_{\boldsymbol{p}}, a_p^p k_p)_{H(\gamma)} = a_p^p (a^p k_{\boldsymbol{p}})(p)$$

leads to the claim. $\square$

The values of a_p^p can be determined by explicitly solving the linear system defining a^p. Let $p = p_i^\ell \in \Lambda_\ell$. If $i = n_\ell = 1$, then $a_p^p = 1/\sqrt{p}$. If $n_\ell \geqslant 2$ and $i = 1$, for $h := p_2^\ell - p_1^\ell$, $a_p^p = \sqrt{(p+h)/ph}$, whereas if $i = n_\ell$, for $h := p_{n_\ell}^\ell - p_{n_\ell-1}^\ell$, $a_p^p = 1/\sqrt{h}$. For interior points, i.e. if $2 \leqslant i \leqslant n_\ell - 1$, $a_p^p = \sqrt{(h_- + h_+)/h_- h_+}$ for $h_- := p_i^\ell - p_{i-1}^\ell$ and $h_+ := p_{i+1}^\ell - p_i^\ell$.

This method provides an alternative derivation of the well-known piecewise linear basis for Brownian motion due to Lévy and Ciesielski, which is more commonly derived by integrating $L^2([0,1])$-orthonormal Haar wavelets (see, e.g. Karatzas & Shreve, 1991).

The statements of Proposition 3.2 and Lemma 3.3 for the Wiener measure are a direct consequence of Proposition 3.8. In particular, Theorem 3.4 applies, as the proof requires only these two statements and is not specific to exponential covariances. Bounds similar to those given in Section 3.2 also hold in the case of the Wiener measure and follow from Proposition 3.8 and the above values of a_p^p.

4. Numerical construction of hierarchic bases

4.1 *A general algorithm*

Let $(e_m)_{m\in\Theta_0}$ be an orthonormal system in $H(\gamma)$ for a finite, possible empty, set Θ_0. We wish to determine a finite sequence $(e_m)_{m\in\Theta_1}$ in $H(\gamma)$ such that $(e_m)_{m\in\Theta_0\cup\Theta_1}$ is an orthonormal system in $H(\gamma)$.

Let $p := (p_i)_{i\in\Lambda}$ be a sequence in D. We assume that, for a finite set $\Lambda_0 \subset \Lambda$, e_m is in the span of $(k_{p_i})_{i\in\Lambda_0}$ for all $m \in \Theta_0$ and allow e_m to be in the span of $(k_{p_i})_{i\in\Lambda_0\cup\Lambda_1}$ for a second finite set $\Lambda_1 \subset \Lambda$. The coefficients of $(e_m)_{m\in\Theta_0}$ with respect to $(k_{p_i})_{i\in\Lambda_0}$ can be stored as the columns of a matrix,

$$A_0 := [a^m]_{m\in\Theta_0}, \quad e_m = \sum_{i\in\Lambda_0} a_i^m k_{p_i}, \quad m \in \Theta_0. \tag{4.1}$$

We also consider the covariance matrix of $(e_m)_{m\in\Theta_0}$ and $(k_{p_i})_{i\in\Lambda_0}$,

$$F_0 := [(k_{p_i}, e_m)_{H(\gamma)}]_{m\in\Theta_0, i\in\Lambda_0} = [e_m(p_i)]_{m\in\Theta_0, i\in\Lambda_0}, \tag{4.2}$$

where m indexes the rows and i the columns of F_0. Although this matrix is not required in order to augment the orthonormal system $(e_m)_{m\in\Theta_0}$ it is computed as a byproduct of our algorithm.

AUGMENTBASIS$[k, p, \Lambda_0, \Lambda_1, A_0, F_0] \mapsto [A, F]$

$K_0 \longleftarrow [k(p_i, p_j)]_{i\in\Lambda_0, j\in\Lambda_1}$
$K_1 \longleftarrow [k(p_i, p_j)]_{i,j\in\Lambda_1}$
$C \longleftarrow A_0^{\mathrm{T}} K_0$
$[V, D] \longleftarrow \mathtt{eig}_\epsilon(K_1 - C^{\mathrm{T}}C)$
$B \longleftarrow V D^{-1/2}$

$$A \longleftarrow \begin{pmatrix} A_0 & -A_0 C B \\ 0 & B \end{pmatrix}$$

$$F \longleftarrow \begin{pmatrix} F_0 & C \\ 0 & D^{1/2} V^{\mathrm{T}} \end{pmatrix}$$

In AUGMENTBASIS, $\mathtt{eig}_\epsilon$ with $\epsilon \geqslant 0$ is a function that computes some nonzero eigenpairs of a matrix. We assume that all eigenvalues not computed by $\mathtt{eig}_\epsilon$ have magnitude less than or equal to ϵ.

REMARK 4.1 If $\Lambda_0 \cap \Lambda_1 \neq \varnothing$, then the matrices A and F constructed by the algorithm AUGMENTBASIS do not have the block structure indicated in the formulation of the algorithm. Rather, the rows of A

corresponding to the same index $i \in \Lambda_0 \cap \Lambda_1$ are assumed to be summed. In F the new value of a column corresponding to $i \in \Lambda_0 \cap \Lambda_1$ overwrites the old value.

PROPOSITION 4.2 Let A_0 and F_0 be as in (4.1) and (4.2) for an orthonormal system $(e_m)_{m\in\Theta_0}$ in $H(\gamma)$. Let the set Θ index the columns of A generated by AUGMENTBASIS, and $\Theta_1 := \Theta \setminus \Theta_0$. For all $m \in \Theta$ let

$$e_m := \sum_{i\in\Lambda_0\cup\Lambda_1} a_i^m k_{p_i}, \tag{4.3}$$

where $a^m := (a_i^m)$ is the mth column of A. Then the sequence $(e_m)_{m\in\Theta}$ is an orthonormal system in $H(\gamma)$. For all $m \in \Theta_0$, (4.3) coincides with (4.1). Furthermore, for all $m \in \Theta$, the mth row of F is the vector $(e_m(p_i))_{i\in\Lambda_0\cup\Lambda_1}$. In particular, if $m \in \Theta_1$ and $i \in \Lambda_0 \setminus \Lambda_1$, then $e_m(p_i) = 0$.

Proof. We can assume without loss of generality that $\Lambda_0 \cap \Lambda_1 = \varnothing$ since $p_i = p_j$ is possible for $i \neq j$. Identifying such indices leads to the compression described in Remark 4.1.

By definition of A and (4.1) the columns of

$$E := \begin{pmatrix} I & -CB \\ 0 & B \end{pmatrix}$$

represent $(e_m)_{m\in\Theta}$ with respect to $(e_m)_{m\in\Theta_0} \cup (k_{p_i})_{i\in\Lambda_1}$. Furthermore, the covariance matrix of the sequence $(e_m)_{m\in\Theta_0} \cup (k_{p_i})_{i\in\Lambda_1}$ is

$$L := \begin{pmatrix} I & A_0^{\mathrm{T}}K_0 \\ (A_0^{\mathrm{T}}K_0)^{\mathrm{T}} & K_1 \end{pmatrix} = \begin{pmatrix} I & C \\ C^{\mathrm{T}} & K_1 \end{pmatrix}.$$

Thus, the covariance matrix of $(e_m)_{m\in\Theta}$ is

$$E^{\mathrm{T}}LE = \begin{pmatrix} I & 0 \\ -B^{\mathrm{T}}C^{\mathrm{T}} & B^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} I & C \\ C^{\mathrm{T}} & K_1 \end{pmatrix} \begin{pmatrix} I & -CB \\ 0 & B \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & B^{\mathrm{T}}(K_1 - C^{\mathrm{T}}C)B \end{pmatrix}.$$

By definition of B,

$$B^{\mathrm{T}}(K_1 - C^{\mathrm{T}}C)B = I.$$

Therefore, $(e_m)_{m\in\Theta}$ is an orthonormal system in $H(\gamma)$; see Corollary 2.11.

Finally, we have

$$FA = \begin{pmatrix} F_0A_0 & -F_0A_0CB + CB \\ 0 & D^{1/2}V^{\mathrm{T}}B \end{pmatrix} = I$$

since $F_0A_0 = I$ by orthonormality of $(e_m)_{m\in\Theta_0}$ in $H(\gamma)$. Therefore, FA is the matrix representation of

$$h = \sum_{m\in\Theta} (h, e_m)_{H(\gamma)} e_m \quad \forall h \in \operatorname{span}(e_m)_{m\in\Theta}$$

in the basis $(k_{p_i})_{i\in\Lambda_0\cup\Lambda_1}$. This implies that the mth row of F is the vector $((k_{p_i}, e_m)_{H(\gamma)})_{i\in\Lambda_0\cup\Lambda_1}$, and the claim follows using Lemma 2.2. □

REMARK 4.3 Let $(p_i)_{i\in\Lambda}$ be a finite sequence in D such that $(k_{p_i})_{i\in\Lambda}$ are linearly independent, and let

$$K := [\langle R_\gamma \delta_{p_i}, \delta_{p_j}\rangle]_{i,j\in\Lambda} = [k(p_i,p_j)]_{i,j\in\Lambda} = [(k_{p_i},k_{p_j})_{H(\gamma)}]_{i,j\in\Lambda} \tag{4.4}$$

be the covariance matrix of $(\delta_{p_i})_{i\in\Lambda}$. Then K is symmetric positive definite and thus allows a factorization of the form

$$K = F^{\mathrm{T}}F \tag{4.5}$$

with a regular matrix F. Defining $A := F^{-1}$ we have

$$A^{\mathrm{T}}KA = A^{\mathrm{T}}F^{\mathrm{T}}FA = I,$$

and by Corollary 2.11, the columns of A are the coefficients with respect to $(k_{p_i})_{i\in\Lambda}$ of an orthonormal system in $H(\gamma)$. Since A is square this orthonormal system is an orthonormal basis of the linear span of $(k_{p_i})_{i\in\Lambda}$ in $H(\gamma)$. Generalizing to rank deficient K we assume F to have full row rank, and define A as the right inverse of F. The routine AUGMENTBASIS with $\epsilon = 0$ hierarchically constructs such F and A. With positive ϵ, it adds a compression of the covariance matrix and (4.5) only holds approximately.

We consider the following hierarchically constructed $(e_m)_{m\in\Theta} \subset H(\gamma)$. Let $p = (p_i)_{i\in\Lambda}$ be a sequence in D, and let $(e_m)_{m\in\Theta_0}$ be an arbitrary finite orthonormal system in $H(\gamma)$ in the linear hull of $(k_{p_i})_{i\in\Lambda_0}$ for a finite set $\Lambda_0 \subset \Lambda$. For all $\ell \in \mathbb{N}$ let $\Lambda_\ell \subset \Lambda$ be a finite set, and let $(e_m)_{m\in\Theta_\ell}$ be the sequence in $H(\gamma)$ constructed by applying AUGMENTBASIS to $(e_m)_{m\in\Theta_{\ell'},\ell'\leqslant\ell-1}$ as in Proposition 4.2 with $\epsilon = \epsilon_\ell$. We define $\Theta := \bigcup_{\ell\in\mathbb{N}_0}\Theta_\ell$.

For all $\ell \in \mathbb{N}_0$ let P_ℓ be the orthonormal projection in $H(\gamma)$ onto the span of e_m for $m \in \Theta_{\ell'}$ with $\ell' \leqslant \ell$. It has the form

$$P_\ell h = \sum_{\ell'=0}^{\ell}\sum_{m\in\Theta_{\ell'}} (h,e_m)_{H(\gamma)}e_m, \quad h \in H(\gamma). \tag{4.6}$$

For all $\ell \in \mathbb{N}_0$ let

$$K_\ell := K|_{\Lambda_\ell\times\Lambda_\ell} = [(k_{p_i},k_{p_j})_{H(\gamma)}]_{i,j\in\Lambda_\ell} \tag{4.7}$$

be the submatrix of K from (4.4) on the index set Λ_ℓ. Using P_ℓ we define the approximation

$$\tilde{K}_\ell := [(P_\ell k_{p_i}, P_\ell k_{p_j})_{H(\gamma)}]_{i,j\in\Lambda_\ell} \tag{4.8}$$

to K_ℓ. Note that

$$\big[\tilde{K}_\ell\big]_{ij} = (k_{p_i},k_{p_j})_{H(\gamma)} = \sum_{\ell'=0}^{\ell}\sum_{m\in\Theta_{\ell'}} e_m(p_i)e_m(p_j), \tag{4.9}$$

so $\tilde{K}_\ell$ is computable directly from e_m for $m \in \Theta_{\ell'}$, $\ell' \leqslant \ell$.

LEMMA 4.4 Using the notation from AUGMENTBASIS,

$$\tilde{K}_1 = \big(F^{\mathrm{T}}F\big)|_{\Lambda_1\times\Lambda_1} = C^{\mathrm{T}}C + VDV^{\mathrm{T}}. \tag{4.10}$$

Proof. By Proposition 4.2 and (4.6),

$$\left[F^{\mathrm{T}}F\right]_{ij} = \Big[\sum_{m\in\Theta_0\cup\Theta_1}(k_{p_i},e_m)_{H(\gamma)}(k_{p_j},e_m)_{H(\gamma)}\Big]_{ij} = [(P_1k_{p_i},P_1k_{p_j})_{H(\gamma)}]_{ij},$$

where $i, j \in \Lambda_0 \cup \Lambda_1$. We compute

$$F^{\mathrm{T}}F = \begin{pmatrix} F_0^{\mathrm{T}} & 0 \\ C^{\mathrm{T}} & VD^{1/2}\end{pmatrix}\begin{pmatrix} F_0 & C \\ 0 & D^{1/2}V^{\mathrm{T}}\end{pmatrix} = \begin{pmatrix} F_0^{\mathrm{T}}F_0 & F_0^{\mathrm{T}}C \\ C^{\mathrm{T}}F_0 & C^{\mathrm{T}}C + VDV^{\mathrm{T}}\end{pmatrix}.$$

Therefore, for all $i, j \in \Lambda_1$, $[\tilde{K}_1]_{ij} = [F^{\mathrm{T}}F]_{ij} = [C^{\mathrm{T}}C + VDV^{\mathrm{T}}]_{ij}$. □

PROPOSITION 4.5 For all $\ell \in \mathbb{N}$ and all $b, c \in \mathbb{R}^{\Lambda_\ell} \subset \mathbb{R}^{\Lambda}$,

$$\left|\langle R_\gamma\, b\delta_{\boldsymbol{p}}, c\delta_{\boldsymbol{p}}\rangle - b^{\mathrm{T}}\tilde{K}_\ell c\right| \leqslant \epsilon_\ell \|b\|_{\ell^2}\|c\|_{\ell^2}\,. \tag{4.11}$$

Proof. We assume without loss of generality that $\ell = 1$. By (2.20), $\langle R_\gamma\, b\delta_{\boldsymbol{p}}, c\delta_{\boldsymbol{p}}\rangle = b^{\mathrm{T}}K_1c$. Therefore, using Lemma 4.4 and the definition of ϵ_ℓ,

$$\left|\langle R_\gamma\, b\delta_{\boldsymbol{p}}, c\delta_{\boldsymbol{p}}\rangle - b^{\mathrm{T}}\tilde{K}_1c\right| = \left|b^{\mathrm{T}}\big(K_1 - C^{\mathrm{T}}C - VDV^{\mathrm{T}}\big)c\right| \leqslant \epsilon_\ell\|b\|_{\ell^2}\|c\|_{\ell^2}.$$ □

COROLLARY 4.6 For all $\ell \in \mathbb{N}$ and all $i, j \in \Lambda_\ell$,

$$|k(p_i,p_j) - (P_\ell k_{p_i}, P_\ell k_{p_j})_{H(\gamma)}| \leqslant \epsilon_\ell. \tag{4.12}$$

Proof. The claim follows from Proposition 4.5 and (2.13). □

COROLLARY 4.7 For all $\ell \in \mathbb{N}$ and all $i \in \Lambda_\ell$,

$$\|k_{p_i} - P_\ell k_{p_i}\|_{H(\gamma)} \leqslant \sqrt{\epsilon_\ell}, \tag{4.13}$$

Proof. Using orthogonality of the projection P_ℓ and Corollary 4.6 with $j = i$,

$$\|k_{p_i} - P_\ell k_{p_i}\|^2_{H(\gamma)} = \|k_{p_i}\|^2_{H(\gamma)} - \|P_\ell k_{p_i}\|^2_{H(\gamma)} \leqslant \epsilon_\ell.$$ □

THEOREM 4.8 If $p = (p_i)_{i\in\Lambda}$ is dense in D, $\bigcup_{\ell\in\mathbb{N}_0}\Lambda_\ell = \Lambda$ and $\epsilon_\ell \to 0$ as $\ell \to \infty$, then $(e_m)_{m\in\Theta}$ is an orthonormal basis of $H(\gamma)$.

Proof. Orthonormality of $(e_m)_{m\in\Theta}$ follows from Proposition 4.2. To prove density it suffices by Proposition 2.9 to show that k_x is in the span of $(e_m)_{m\in\Theta}$ for all $x \in D$.

Let $x \in D$. Then there is a sequence $(i_n)_{n\in\mathbb{N}}$ in Λ such that $x_n := p_{i_n} \to x$ and $i_n \neq i_{n'}$ if $n \neq n'$. Since $\|\delta_{x_n}\|_{C(D)^*} = 1$ for all n, Lemma 2.3 implies $k_{x_n} \to k_x$ in $H(\gamma)$. For all $n \in \mathbb{N}$ let $\ell_n \in \mathbb{N}_0$ with $i_n \in \Lambda_{\ell_n}$. Then $\ell_n \to \infty$ since Λ_ℓ is finite for all $\ell \in \mathbb{N}_0$. Due to Corollary 4.7, $P_{\ell_n}k_{x_n} \to k_x$ by the assumption $\epsilon_\ell \to 0$. □

4.2 *Gram–Schmidt orthogonalization in the Cameron–Martin space*

We consider AUGMENTBASIS with $\epsilon = 0$ for $\Lambda_1 = \{j\}$, where j is assumed to be an element of $\Lambda \setminus \Lambda_0$. In this case, AUGMENTBASIS reduces to Gram–Schmidt orthogonalization in the Cameron–Martin space and can be found in Pugachev (1965, Chapter 9). An explicit formulation is given in GRAMSCHMIDTSTEP.

GRAMSCHMIDTSTEP$[k, p, \Lambda_0, j, A_0, F_0] \mapsto [A, F]$

$k_0 \longleftarrow [k(p_i, p_j)]_{i \in \Lambda_0}$
$c \longleftarrow A_0^{\mathrm{T}} k_0$ // i.e. $c_i = e_i(p_j) = (e_i, k_{p_j})_{H(\gamma)}$
$d \longleftarrow k(p_j, p_j) - c^{\mathrm{T}} c$ // i.e. $d = k(p_j, p_j) - \sum_{i \in \Lambda_0} e_i(p_j)^2$
if $d = 0$ then
$A \longleftarrow \begin{pmatrix} A_0 \\ 0 \end{pmatrix}$
$F \longleftarrow \begin{pmatrix} F_0 & c \end{pmatrix}$
else // i.e. $d > 0$
$s \longleftarrow \sqrt{d}$
$A \longleftarrow \begin{pmatrix} A_0 & -s^{-1} A_0 c \\ 0 & s^{-1} \end{pmatrix}$
$F \longleftarrow \begin{pmatrix} F_0 & c \\ 0 & s \end{pmatrix}$
end

REMARK 4.9 Let $(p_i)_{i=1}^n \subset D$ such that $(k_{p_i})_{i=1}^n$ are linearly independent, and let $(e_i)_{i=1}^n$ be constructed by recursive application of GRAMSCHMIDTSTEP. Furthermore, let K be the covariance matrix of $(\delta_{p_i})_{i=1}^n$, and let F be the last output of GRAMSCHMIDTSTEP. By Proposition 4.2,

$$F^{\mathrm{T}} F = \left[\sum_{m=1}^{n} (k_{p_i}, e_m)_{H(\gamma)} (k_{p_j}, e_m)_{H(\gamma)} \right]_{i,j=1}^{n} = [(k_{p_i}, k_{p_j})_{H(\gamma)}]_{i,j=1}^{n} = K.$$

Furthermore, it follows by induction that F is an upper triangular matrix with positive diagonal entries. Therefore, F is the right Cholesky factor of K.

REMARK 4.10 The basis functions $(e_m)_{m \in \mathbb{N}}$ can be characterized independently of each other. For a fixed sequence $(p_i)_{i \in \mathbb{N}}$ in D, e_m is in the span of $(k_{p_i})_{i=1}^m$ and orthogonal to k_{p_i} for all $i \leqslant m - 1$. This defines e_m uniquely up to a scalar factor, which is determined by normalization in $H(\gamma)$.

REMARK 4.11 Assuming that the covariance kernel can be evaluated in unit time, the computational cost of GRAMSCHMIDTSTEP is dominated by the matrix–vector multiplication. Therefore, the nth step requires $\mathcal{O}(n^2)$ arithmetic operations, and the construction of the first n basis functions $(e_m)_{m=1}^n$ has complexity $\mathcal{O}(n^3)$.

4.3 *Hierarchic spectral decomposition of the covariance matrix*

As a second particular case of AUGMENTBASIS, we assume that the sets Λ_ℓ are nested,

$$\Lambda_0 \subset \Lambda_1 \subset \cdots \subset \Lambda_\ell \subset \Lambda_{\ell+1} \subset \cdots \subset \Lambda, \tag{4.14}$$

and, in view of Theorem 4.8, $\Lambda = \bigcup_{\ell=0}^{\infty} \Lambda_\ell$. For example, $(p_i)_{i\in\Lambda_\ell}$ could be the nodes of nested grids of D. In this setting Proposition 4.5 applies to all nodes p_i visited up to level ℓ for any $\ell \in \mathbb{N}$.

The routine AUGMENTBASIS is customized to (4.14) in AUGMENTSPECTRALBASIS, which assumes $\Lambda_0 \subset \Lambda_1$. In particular, AUGMENTSPECTRALBASIS takes into account Remark 4.1.

AUGMENTSPECTRALBASIS$[k, p, \Lambda_0, \Lambda_1, A_0] \mapsto [A, F]$

$K \longleftarrow [k(p_i, p_j)]_{i,j\in\Lambda_1}$

$K_0 \longleftarrow K|_{\Lambda_0\times\Lambda_1}$

$C \longleftarrow A_0^{\mathrm{T}} K_0$

$[V, D] \longleftarrow \mathtt{eig}_\epsilon(K - C^{\mathrm{T}}C)$

$B = \begin{pmatrix} B_0 \\ B_1 \end{pmatrix} \longleftarrow V D^{-1/2}$ // $B_0 = B|_{\Lambda_0\times\Lambda_1}$ and $B_1 = B|_{(\Lambda_1\setminus\Lambda_0)\times\Lambda_1}$

$A \longleftarrow \begin{pmatrix} A_0 & B_0 - A_0 C B \\ 0 & B_1 \end{pmatrix}$

$F \longleftarrow \begin{pmatrix} C \\ D^{1/2} V^{\mathrm{T}} \end{pmatrix}$

REMARK 4.12 A common but seemingly naive approximation to the Karhunen–Loève series consists of computing some or all eigenpairs of the covariance matrix K, e.g. on the vertices of a finite element mesh, in place of more precise representations of eigenfunctions of the covariance operator. Let $p = (p_i)_{i\in\Lambda}$ be a finite sequence in D. Then this procedure is similar to AUGMENTSPECTRALBASIS with $\Lambda_0 = \varnothing$ and $\Lambda_1 = \Lambda$. However, we provide a different interpretation. Instead of being approximations to the eigenfunctions of the covariance operator the eigenvectors of the covariance matrix define an orthonormal system in $H(\gamma)$ in their own right,

$$e_m := \lambda_m^{-1/2} \sum_{i\in\Lambda} v_i^m k_{p_i}, \tag{4.15}$$

where $v^m = (v_i^m)_{i\in\Lambda}$ is a normalized eigenvector of K with eigenvalue λ_m. Thus, the functions e_m are defined on all of D not just at the discrete points $(p_i)_{i\in\Lambda}$. This decouples their construction from any other discretization of D, such as a finite element mesh.

REMARK 4.13 The computational cost of AUGMENTSPECTRALBASIS depends on the number of nodes p_i, and the number of basis functions e_m. Suppose we apply AUGMENTSPECTRALBASIS recursively to construct $(e_m)_{m\in\Theta}$, using n_ℓ nodes $(p_i)_{i\in\Lambda_\ell}$ to construct m_ℓ new basis functions in the ℓth call. Ignoring for the moment the cost of computing eigenvalues and eigenvectors and abbreviating $M_\ell = \sum_{i=1}^{\ell} m_i$, the ℓth call of AUGMENTSPECTRALBASIS uses $\mathcal{O}(n_\ell^2 m_\ell M_\ell)$ arithmetic operations.

For example, let $n_\ell = 2^{d\ell}$. If the eigenvalues of $K - C^{\mathrm{T}}C$ in each call of the algorithm converge exponentially it is sufficient to take $m_\ell \sim \ell$, in which case M_ℓ is of the order of ℓ^2. The resulting complexity of the ℓth step of the construction, disregarding the spectral decomposition, is $\mathcal{O}(2^{2d\ell}\ell^3)$, and the cost of the first ℓ steps is dominated by that of the ℓth step alone. We note that this is less than $\mathcal{O}(2^{3d\ell})$ required by GRAMSCHMIDTSTEP; see Remark 4.11. Moreover, the evaluation of the covariance kernel already uses $\mathcal{O}(2^{2d\ell})$ operations, so the complexity is almost optimal in this example.

To this we need to add the cost of computing m_ℓ eigenpairs of a symmetric $n_\ell \times n_\ell$ matrix in each step. Since one sweep of a Jacobi iteration or one step of a Lanczos method uses $\mathcal{O}(n_\ell^2)$ operations for a dense covariance matrix, the cost of constructing spectral decompositions for generic covariances is comparable to that of the rest of the algorithm.

5. Numerical examples

5.1 *Exponential kernels*

We consider covariance kernels of the form

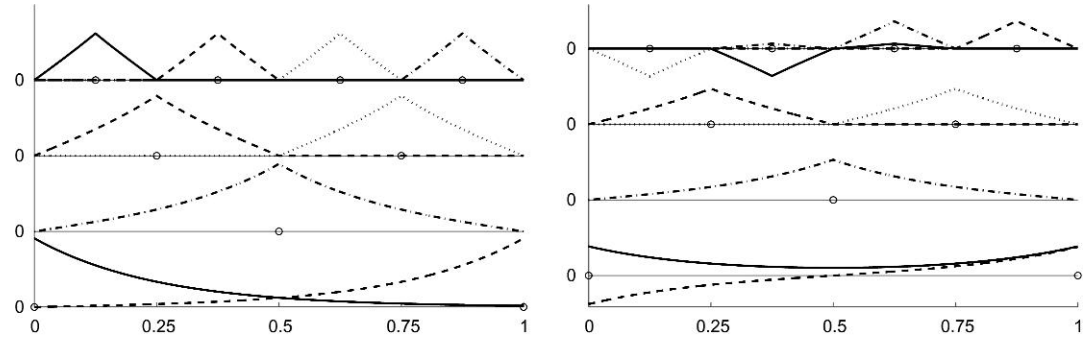
$$k(x, y) := e^{-|x-y|^\alpha/\lambda^\alpha}, \quad x, y \in [0, 1]^d, \tag{5.1}$$

with parameters $\alpha \in [1, 2)$ and $\lambda > 0$, on the d-dimensional unit hypercube $D := [0, 1]^d$. We assume throughout that $\lambda = 1/4$.

Figures 2 and 3 show a few basis functions generated by the Gram–Schmidt method from Section 4.2 and the hierarchic spectral method from Section 4.3 in one dimension, with $\alpha = 1$ and $\alpha = 1.5$, respectively. The sets $(p_i)_{i\in\Lambda_\ell}$ are hierarchic dyadic grids,

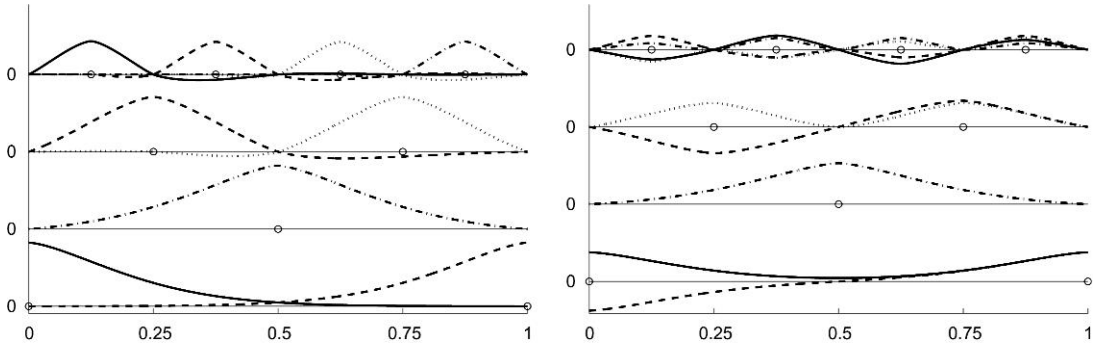
$$(p_i)_{i\in\Lambda_\ell} = \left\{ i2^{-\ell}; i = 0, 1, \ldots, 2^\ell \right\}. \tag{5.2}$$

The new points on each level are marked in the plots by circles. The Gram–Schmidt method adds the points from left to right. The functions generated by Gram–Schmidt for $\alpha = 1$, shown in Fig. 2, coincide



(a) Gram–Schmidt (b) hierarchic spectral

FIG. 2. The first few functions for $\alpha = 1$.



(a) Gram–Schmidt (b) hierarchic spectral

FIG. 3. The first few functions for $\alpha = 1.5$.

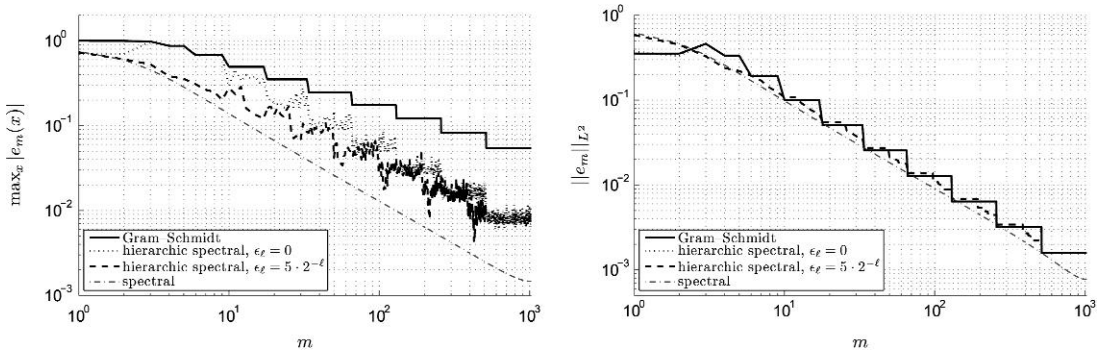
with those derived in Section 3.1. In particular, by Proposition 3.2, they have compact support. This is not the case for the functions generated by the hierarchic spectral method; see Fig. 2. For $\alpha = 1.5$, though the basis functions generated by both methods have global support, those generated by Gram–Schmidt appear to be more localized in Fig. 3.

In higher dimensions the dyadic grids are given by

$$(p_i)_{i\in\Lambda_\ell} = \left\{ i2^{-\ell}; i \in \{0, 1, \ldots, 2^\ell\}^d \right\}, \tag{5.3}$$

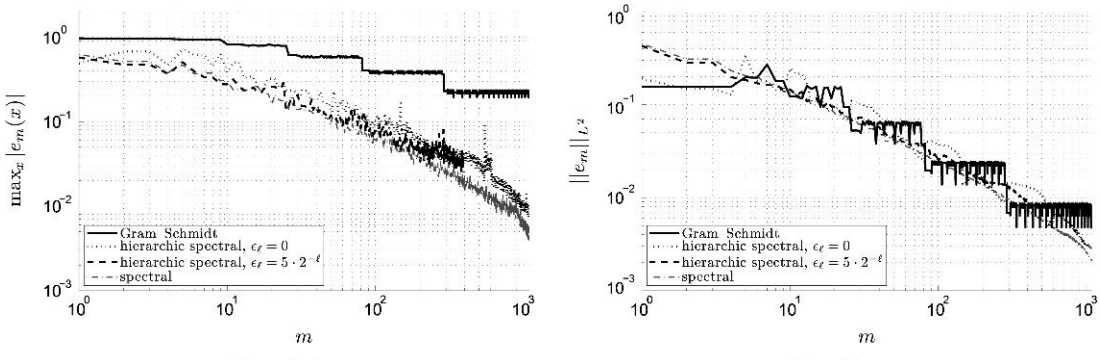
and the Gram–Schmidt method adds points in lexicographic order.

The decay of the $C(D)$ and $L^2(D)$ norms of the basis functions generated by Gram–Schmidt and the hierarchic spectral method for $\alpha = 1$ are plotted in Figs 4 and 5 for one and two dimensions, respectively. Figure 6 shows the same for $\alpha = 1.5$ and $d = 1$. Both norms are approximated on uniform grids in D containing 4096 points. The decay is compared to that of the spectral basis computed directly on the finest level. For the hierarchic spectral method, tolerances $\epsilon_\ell = 0$ and $\epsilon_\ell = 5.2^{-\alpha\ell}$ are used. In the latter case, the hierarchic spectral method generates fewer basis functions than the other methods.



(a) $C(D)$ (b) $L^2(D)$

FIG. 4. Decay of the basis functions for $\alpha = 1$ and $d = 1$.



(a) $C(D)$ (b) $L^2(D)$

FIG. 5. Decay of the basis functions for $\alpha = 1$ and $d = 2$.

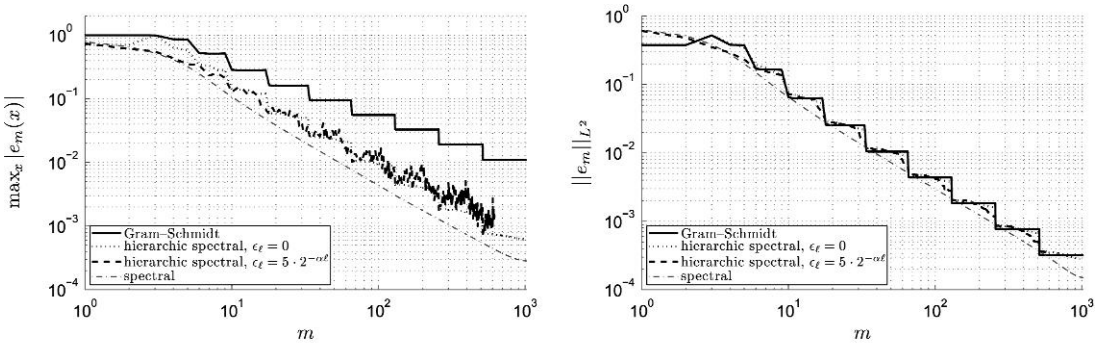


FIG. 6. Decay of the basis functions for $\alpha = 1.5$ and $d = 1$.

In each case, the rate of decay in the $L^2(D)$-norm is the same for all basis functions considered. It is approximately

$$\|e_m\|_{L^2(D)} \sim m^{-\alpha}. \tag{5.4}$$

A staircase effect is apparent, in particular for the Gram–Schmidt method. This is due to the uneven spacing of points between levels of the hierarchic dyadic grids. Within each level, the position of a node relative to all previous nodes is very similar for all nodes. Between levels the distances scale by a factor of two. In two and three dimensions the functions constructed by the hierarchic spectral method display jumps in the opposite direction between levels. At these points the covariance matrix is refined, and the subsequent basis functions correspond to eigenvectors with large eigenvalues of the difference between the coarse and fine covariance matrices.

The rate of decay in $C(D)$ coincides with that in $L^2(D)$ for the hierarchic spectral method. However, the decay of the functions generated by the Gram–Schmidt method is slower in $C(D)$ than in $L^2(D)$. For $\alpha = 1$ and $d = 1$ this is shown in Example 3.7.

The slower decay in $C(D)$ seems to be the cost for the better localization in space of the basis functions generated by the Gram–Schmidt method. For $\alpha = 1$ and $d = 1$ these functions have compact support, and the size of the support decreases at the rate m^{-1}. In other cases, though their supports are not compact, the functions generated by Gram–Schmidt are still almost local. Figure 7 illustrates the decay of basis functions generated by the Gram–Schmidt method for $\alpha = 1.5$ and $d = 1$. Figure 7(a) shows the measures of level sets

$$\{x \in D; |e_m(x)| \geqslant \epsilon \max_{y \in D} |e_m(y)|\} \tag{5.5}$$

for four different values of ϵ. At least for larger values of ϵ the measure of these level sets decreases approximately as m^{-1}. Figure 7(b) plots some basis functions in logarithmic scale. Apparently, at higher levels of the hierarchic dyadic grids there is a fast initial decay, followed by decay comparable to that of the kernel but at a much lower level.

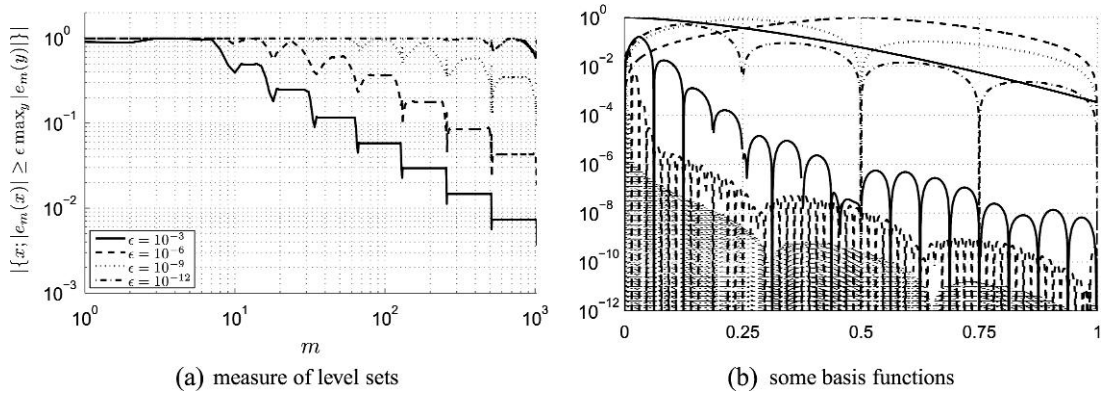


FIG. 7. Localization of functions generated by Gram–Schmidt for $\alpha = 1.5$, $d = 1$.

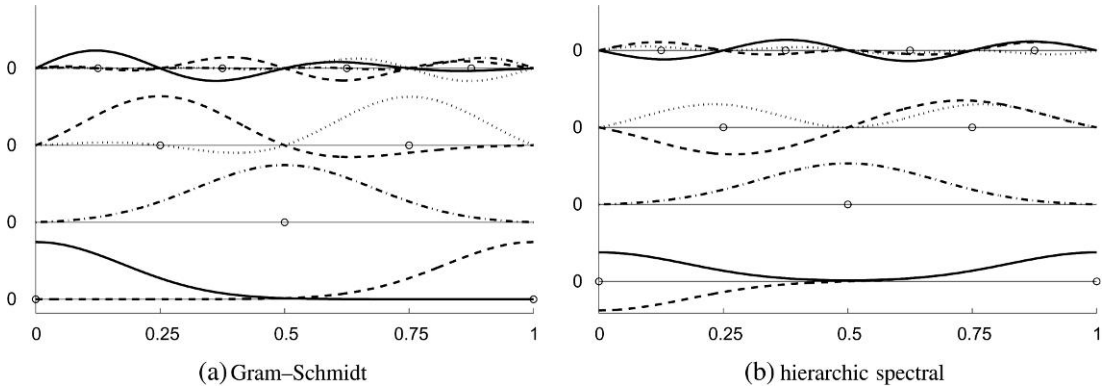


FIG. 8. The first few functions for a Gaussian kernel.

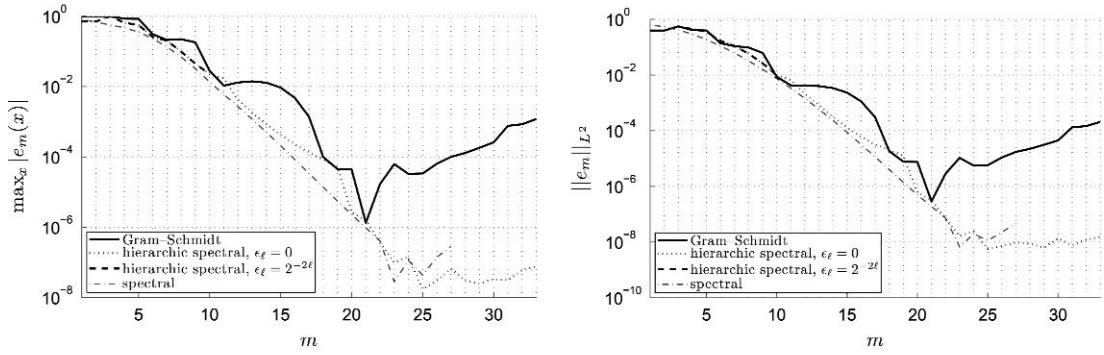
5.2 *Gaussian kernels*

The Gaussian kernel on $D = [0, 1]^d$ with correlation length $\lambda > 0$ is

$$k(x, y) := e^{-|x-y|^2/\lambda^2}, \quad x, y \in D. \tag{5.6}$$

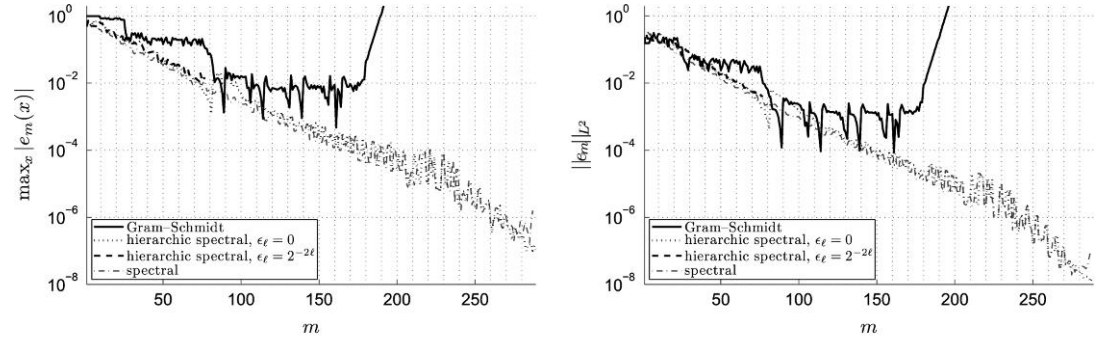
The main difference between the Gaussian kernel (5.6) and the exponential kernels (5.1) with $1 \leqslant \alpha < 2$ is that the eigenvalues of the covariance operator associated to the Gaussian kernel decay exponentially, as opposed to algebraic decay for exponential kernels. Again, we set λ to $1/4$ in all computations. Figure 8 shows the first few basis functions generated by the Gram–Schmidt method and the hierarchic spectral method in one dimension.

The decay of the basis functions generated by both of the above methods for Gaussian kernels is plotted in Figs 9 and 10 in one and two dimensions, respectively. These are compared to the spectral basis computed directly on the finest level. In all cases, the decay is exponential, with approximately the same rate. However, the Gram–Schmidt method suffers from an instability, limiting its convergence. It can be stabilized by replacing the Cholesky decomposition with a suitable generalization of a pivoted Cholesky depcomposition, as was studied in Harbrecht *et al.* (2010) for finite sets of points.



(a) $C(D)$ (b) $L^2(D)$

FIG. 9. Decay of the basis functions for $d = 1$.



(a) $C(D)$ (b) $L^2(D)$

FIG. 10. Decay of the basis functions for $d = 2$.

The hierarchic spectral method is shown with tolerances $\epsilon_\ell = 0$ and $\epsilon_\ell = 2^{-2\ell}$. In the latter case, only very few basis functions e_m are constructed, and their norms are very close to those of the basis constructed with no truncation. Figure 11 shows the number of new basis functions constructed on each level, i.e. after each refinement of the covariance matrix. The Gram–Schmidt method and the hierarchic spectral method with $\epsilon_\ell = 0$ construct one basis function for each point at which the covariance kernel is evaluated. Accordingly, the number of new basis functions per level increases exponentially. Conversely, the hierarchic spectral method with positive tolerances $\epsilon_\ell = 2^{-2\ell}$ constructs far fewer basis functions. The number of new basis functions constructed on each level seems to be bounded independently of the level.

5.3 *Spherical covariance kernels*

The spherical covariance kernel is given by

$$k(x, y) := \sigma^2 \rho_3 \left(\frac{|x - y|}{\lambda} \right) \tag{5.7}$$

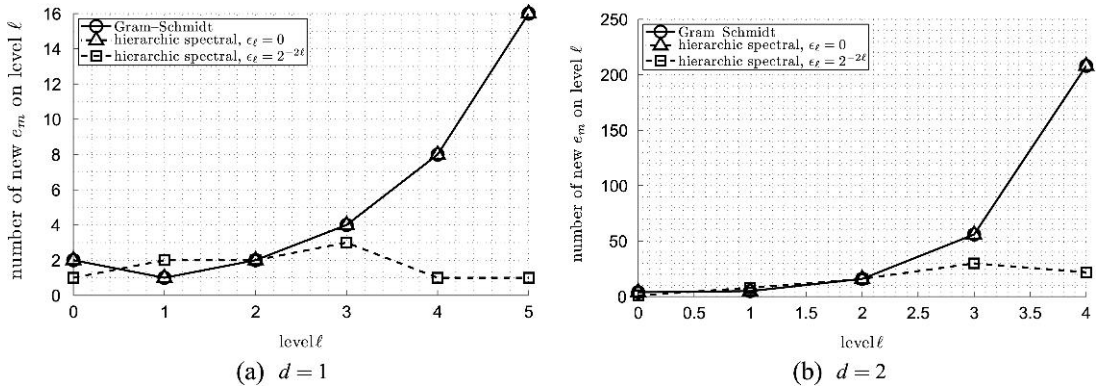


FIG. 11. Number of new basis functions per level.

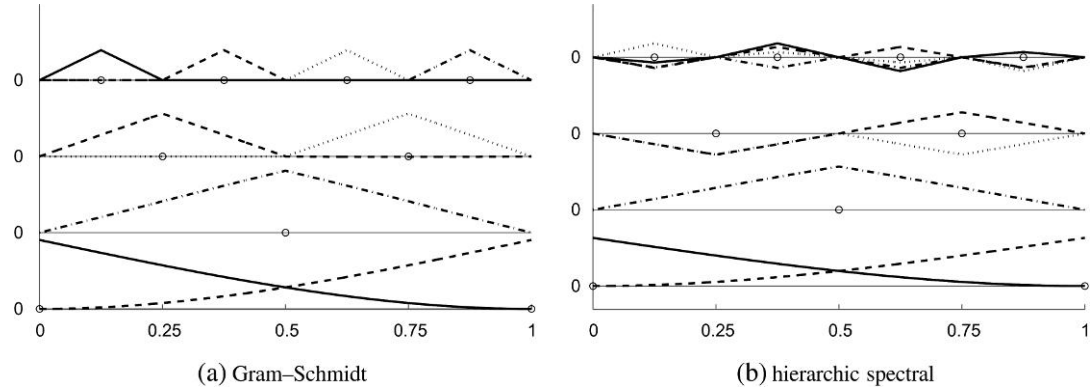


FIG. 12. The first few functions for a spherical covariance.

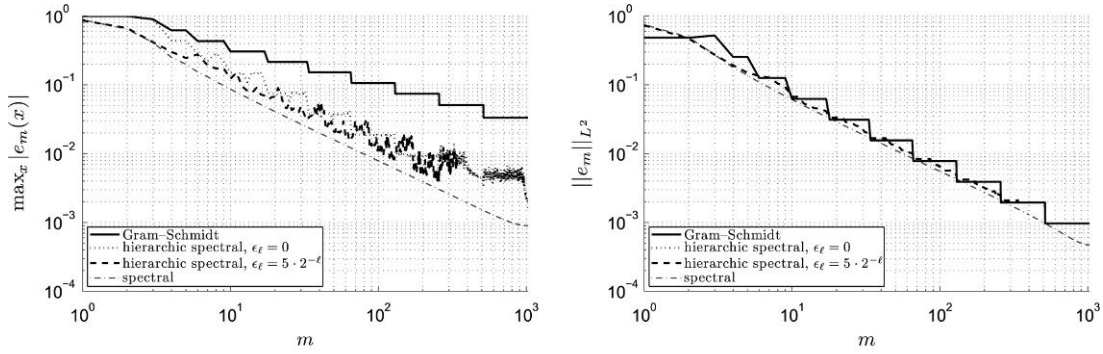
in three dimensions or less, with positive parameters σ and λ, where

$$\rho_3(z) := 1 - \frac{3}{2}z + \frac{1}{2}z^3, \quad z \in [0, 1], \tag{5.8}$$

and $\rho_3(z) := 0$ if $z \geqslant 1$ is, up to a scale factor, the volume of intersection of two spheres with diameter 1 and midpoints separated by z. Similar constructions exist for balls of dimension different from three.

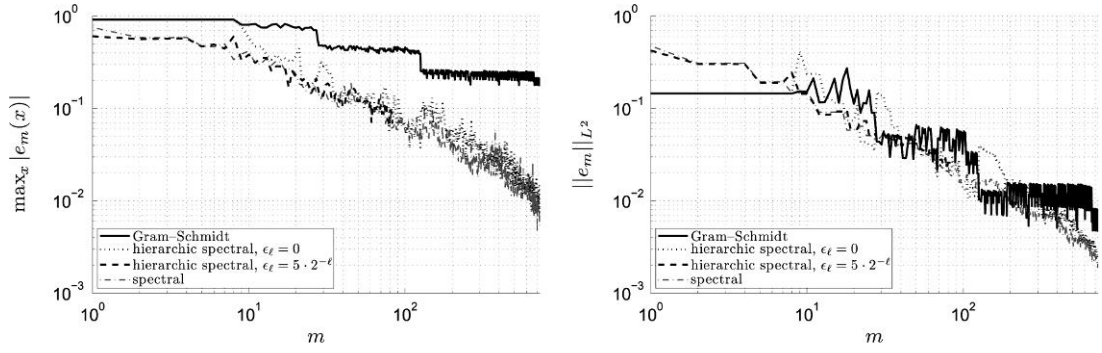
We consider the covariance kernel (5.7) on $[0, 1]^d$, $d \in \{1, 2, 3\}$, with $\sigma = \lambda = 1$. We use the hierarchic dyadic grids $(p_i)_{i \in \Lambda_\ell}$ defined in Section 5.1. The first few basis functions generated by the Gram–Schmidt and hierarchic spectral methods are shown in Fig. 12.

Figures 13 and 14 show the decay of the basis functions generated by Gram–Schmidt and the hierarchic spectral method in one and three dimensions, respectively. The behaviour is very similar to that of the exponential covariance kernel with $\alpha = 1$, discussed in Section 5.1.



(a) $C(D)$ (b) $L^2(D)$

FIG. 13. Decay of the basis functions for a spherical covariance and $d = 1$.



(a) $C(D)$ (b) $L^2(D)$

FIG. 14. Decay of the basis functions for a spherical covariance and $d = 3$.

6. Conclusions

Orthonormal bases of the Cameron–Martin space of a Gaussian measure on $C(D)$ can be constructed explicitly, without resorting to eigenfunctions of the covariance operator. Their construction uses only the covariance kernel, which is readily available in many stochastic models. No assumptions are made on the structure of the kernel.

The covariance kernel is evaluated on an unstructured discrete set of points. Generally, one basis function can be computed for every evaluation point. The general algorithm for constructing such bases can be formulated on the level of numerical linear algebra involving the covariance matrix, and, as such, is amenable to implementation. In exact arithmetic it constructs exact representations of the basis functions.

The bases constructed in this manner are hierarchic. Elements computed on an initial coarse grid of evaluation points are left unchanged when additional basis functions are computed on the same grid or on a refined grid. This flexible construction of the basis elements may lend itself to adaptive or iterative methods that require representations of a Gaussian field with various degrees of accuracy. The representation can be refined locally by selectively adding evaluation points of the covariance kernel. Also, the

initial basis functions represent the most dominant components of the random field, and constructing these on a coarse grid of evaluation points leads to simple representations of these functions in terms of the covariance kernel at only a few points.

Numerical experiments indicate that the decay of the basis functions in $L^2(D)$ is comparable to that of the square root of the eigenvalues of the covariance operator, which is known to be optimal. This is confirmed by explicit computations in the case of an exponential covariance kernel in one dimension. The hierarchic spectral method for constructing basis functions is particularly close to optimal in this respect, and stable in the case of ill-conditioned covariance matrices, e.g. resulting from Gaussian covariance kernels.

Bases constructed by Gram–Schmidt orthogonalization in the Cameron–Martin space are often spatially localized. In some settings, e.g. an exponential covariance kernel in one dimension or the Wiener measure, the basis functions have compact supports with diameter tending to zero. Furthermore, the basis functions can be characterized independently of each other as functions in certain finite dimensional spaces with given zeros. Thus, if the covariance kernel is modified the basis functions can be updated independently of each other.

Acknowledgements

The author wishes to thank Hans-Jörg Starkloff for his insightful suggestions and also Andrea Barth and Christoph Schwab for their remarks.

Funding

The Swiss National Science Foundation (200021-120290/1).

References

BABUŠKA, I., TEMPONE, R. & ZOURARIS, G. E. (2004) Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J. Numer. Anal.*, **42**, 800–825 (electronic).

BIERI, M., ANDREEV, R. & SCHWAB, C. (2009) Sparse tensor discretization of elliptic SPDEs. *SIAM J. Sci. Comput.*, **31**, 4281–4304.

BOGACHEV, V. I. (1998) *Gaussian Measures*, Mathematical Surveys and Monographs, vol. 62. Providence, RI: American Mathematical Society, p. xii, p. 433

DIETRICH, C. R. & NEWSAM, G. N. (1997) Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. *SIAM J. Sci. Comput.*, **18**, 1088–1107.

DIGGLE, P. J. & RIBEIRO, J. P. J. (2007) *Model-Based Geostatistics*, Springer Series in Statistics. New York: Springer, p. xiv, p. 228.

ELLIOTT, J. F. W., HORNTROP, D. J. & MAJDA, A. J. (1997) A Fourier-wavelet Monte Carlo method for fractal random fields. *J. Comput. Phys.*, **132**, 384–408.

ELLIOTT, J. F. W. & MAJDA, A. J. (1994) A wavelet Monte Carlo method for turbulent diffusion with many spatial scales. *J. Comput. Phys.*, **113**, 82–111.

FERNIQUE, X. (1970) Intégrabilité des vecteurs Gaussiens. *C. R. Acad. Sci. Paris Sér. A-B*, **270**, A1698–A1699.

FRAUENFELDER, P., SCHWAB, C. & TODOR, R. A. (2005) Finite elements for elliptic problems with stochastic coefficients. *Comput. Methods Appl. Mech. Eng.*, **194**, 205–228.

GRAHAM, I. G., KUO, F. Y., NUYENS, D., SCHEICHL, R. & SLOAN, I. H. (2010) Quasi-Monte Carlo methods for computing flow in random porous media. *Technical Report 4/10*. Department of Mathematical Sciences, Bath Institute For Complex Systems, Bath.

HARBRECHT, H., PETERS, M. & SCHNEIDER, R. (2010) On the low-rank approximation by the pivoted Cholesky decomposition. *Technical Report 2010-32*. SimTech Cluster of Excellence, Universität Stuttgart, Stuttgart.

JAIN, N. C. & KALLIANPUR, G. (1970) Norm convergent expansions for Gaussian processes in Banach spaces. *Proc. Am. Math. Soc.*, **25**, 890–895.

KARATZAS, I. & SHREVE, S. E. (1991) *Brownian Motion and Stochastic Calculus. Graduate Texts in Mathematics*, vol. 113, 2nd edn. New York: Springer, p. xxiv, p. 470.

KUELBS, J. (1971) Expansions of vectors in a Banach space related to Gaussian measures. *Proc. Am. Math. Soc.*, **27**, 364–370.

LANG, A. & POTTHOFF, J. (2009) Fast simulation of Gaussian random fields. Monte Carlo Methods Appl. (to appear) arXiv:math.NA.1105.2737.

MATTHIES, H. G. & KEESE, A. (2005) Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Comput. Methods Appl. Mech. Eng.*, **194**, 1295–1331.

NOBILE, F., TEMPONE, R. & WEBSTER, C. G. (2008) A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, **46**, 2309–2345.

PAGE, R. D. L. (1972) Note relating Bochner integrals and reproducing kernels to series expansions on a Gaussian Banach space. *Proc. Am. Math. Soc.*, **32**, 285–288.

PUGACHEV, V. S. (1965) *Theory of Random Functions and its Application to Control Problems*. International Series of Monographs on Automation and Automatic Control, vol. 5. Oxford: Pergamon Press, p. xvii, p. 833. (Revised translation by O. M. Blunn. Translation edited by N. L. Johnson).

RIPLEY, B. D. (1987) *Stochastic Simulation*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. New York: John Wiley, p. xiv, p. 237.

SCHWAB, C. & TODOR, R. A. (2006) Karhunen–Loève approximation of random fields by generalized fast multipole methods. *J. Comput. Phys.*, **217**, 100–122.

STEIN, M. L. (1999) *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics. New York: Springer, p. xviii, p. 247.

TODOR, R. A. & SCHWAB, C. (2007) Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients. *IMA J. Numer. Anal.*, **27**, 232–261.

VAKHANIA, N. N., TARIELADZE, V. I. & CHOBANYAN, S. A. (1987) *Probability Distributions on Banach Spaces*. Mathematics and its Applications (Soviet Series), vol. 14. Dordrecht: D. Reidel Publishing, p. xxvi, p. 482. (Translated from the Russian and with a preface by Wojbor A. Woyczynski).

WAN, X. & KARNIADAKIS, G. E. (2009) Solving elliptic problems with non-Gaussian spatially-dependent random coefficients. *Comput. Methods Appl. Mech. Eng.*, **198**, 1985–1995.