MULTISCALE MODEL. SIMUL.
Vol. 13, No. 3, pp. 812–828

BAYESIAN NUMERICAL HOMOGENIZATION*

HOUMAN OWHADI†

Abstract. Numerical homogenization, i.e., the finite-dimensional approximation of solution spaces of PDEs with arbitrary rough coefficients, requires the identification of accurate basis elements. These basis elements are oftentimes found after a laborious process of scientific investigation and plain guesswork. Can this identification problem be facilitated? Is there a general recipe/decision framework for guiding the design of basis elements? We suggest that the answer to the above questions could be positive based on the reformulation of numerical homogenization as a Bayesian inference problem in which a given PDE with rough coefficients (or multiscale operator) is excited with noise (random right-hand side/source term) and one tries to estimate the value of the solution at a given point based on a finite number of observations. We apply this reformulation to the identification of bases for the numerical homogenization of arbitrary integro-differential equations and show that these bases have optimal recovery properties. In particular we show how rough polyharmonic splines can be rediscovered as the optimal solution of a Gaussian filtering problem.

Key words. numerical homogenization, Bayesian inference, Bayesian numerical analysis, coarse graining, polyharmonic splines, Gaussian filtering

AMS subject classifications. 41A15, 34E13, 62C10, 60H30

DOI. 10.1137/140974596

1. Bayesian numerical analysis. This paper is inspired by a curious (and, perhaps, overlooked) link between Bayesian inference and numerical analysis [21], known as Bayesian numerical analysis [21, 63, 49, 50], that can be traced back to Poincaré's course on probability theory [62]. We will recall Diaconis' compelling example [21] as an illustration of this link.

Let $f : [0, 1] \to \mathbb{R}$ be a given function and assume that we are interested in the numerical approximation of $\int_0^1 f(t)\,dt$. The Bayesian approach to this quadrature problem is to (1) put a prior (probability distribution) on continuous functions $\mathcal{C}[0, 1]$, (2) calculate f at $x_1, x_2, \ldots, x_n$ (to obtain the data $(f(x_1), \ldots, f(x_n))$), (3) compute a posterior, and (4) estimate $\int_0^1 f(t)\,dt$ by the Bayes rule.

If the prior on $\mathcal{C}[0, 1]$ is that of a Brownian motion (i.e., $f(t) = B_t$, where B_t is a Brownian motion and B_0 is normal), then $\mathbb{E}\big[f(x)\big|f(x_1), \ldots, f(x_n)\big]$ is the piecewise linear interpolation of f between the points $x_1, \ldots, x_n$, and one rediscovers the trapezoidal quadrature rule.

If the prior on $\mathcal{C}[0, 1]$ is that of the first integral of a Brownian motion (i.e., $f(t) \sim \int_0^t B_s\,ds$), then the posterior $\mathbb{E}\big[f(x)\big|f(x_1), \ldots, f(x_n)\big]$ is the cubic spline interpolant, and integrating k times yields splines of order $2k + 1$. Although this link has led to the identification of new quadrature rules for numerical integration [49], it appears to have remained little known, and our paper is prompted by the question of the existence of a similar link between Bayesian inference and numerical homogenization.

*Received by the editors June 26, 2014; accepted for publication (in revised form) May 5, 2015; published electronically July 14, 2015. The author gratefully acknowledges support from the Air Force Office of Scientific Research under award FA9550-12-1-0389 (Scientific Computation of Optimal Statistical Estimators) and the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research, through the Exascale Co-Design Center for Materials in Extreme Environments (ExMatEx, LANL contract DE-AC52-06NA25396, Caltech subcontract 273448).
http://www.siam.org/journals/mms/13-3/97459.html

†Computing & Mathematical Sciences, California Institute of Technology, Pasadena, CA 91125 (owhadi@caltech.edu).

As a prototypical example, consider the numerical homogenization of the PDE

$$\begin{cases} -\operatorname{div}\Big(a(x)\nabla u(x)\Big) = g(x), \quad x \in \Omega,\, g \in L^2(\Omega), \\ u = 0 \quad \text{on} \quad \partial\Omega, \end{cases} \tag{1.1}$$

where Ω is a bounded subset of $\mathbb{R}^d$ with piecewise Lipschitz boundary and a is a symmetric, uniformly elliptic $d \times d$ matrix on Ω with entries in $L^\infty(\Omega)$.

Recall that numerical homogenization concerns the approximation of the solution space of (1.1) with a finite-dimensional space. Although classical homogenization concepts [12, 46, 64, 20, 61, 44] might be present in some instances of this problem [42, 43, 3, 26, 2, 37, 15, 16, 33, 11, 32, 1], one of the main objectives of numerical homogenization is to achieve a numerical approximation of the solution space of (1.1) with arbitrary rough coefficients, i.e., in particular, without the assumptions found in classical homogenization, such as scale separation, ergodicity at fine scales, and ϵ-sequences of operators. In this situation, piecewise linear finite elements can perform arbitrarily badly [10], and the numerical approximation of the solution space involves the identification of accurate basis elements adapted to the microstructure $a(x)$ [70, 9, 6, 57, 58, 56, 29, 28, 5, 4, 48, 18, 17, 19, 13, 7, 59, 27, 45, 39].

As for the identification of quadrature rules in numerical analysis, the identification of accurate basis elements in numerical homogenization has been based on a difficult process of scientific investigation. Let us now turn our attention to the Bayesian approach to this problem. An immediate question is, Where do we place the prior? (1) If the prior is placed on u, then posterior values do not see (depend on) the microstructure. (2) If the prior is placed on a, then the microstructure becomes random, whereas our purpose is the numerical homogenization of a given deterministic microstructure. Let us also note that the randomization of the microstructure, as investigated by polynomial chaos approximation/stochastic expansion methods [36, 35, 72, 8, 31, 22], does not lead to the simplification seen after homogenization but to increased complexity with the dimension of input stochastic variables [66, 14] (although stochastic expansion methods have been used successfully to beat Monte-Carlo sampling, they do not lead to averaging results seen in homogenization). (3) If the prior is placed on g, then the noise propagates through the microstructure and the posterior value of u contains that information.

This observation motivates us to place the prior on the source term g in (1.1), e.g., replace it by white noise (i.e., a centered Gaussian field $\xi(x)$ on Ω with covariance function $\delta(x-y)$), and consider the stochastic PDE

$$\begin{cases} -\operatorname{div}\Big(a(x)\nabla u(x)\Big) = \xi(x), \quad x \in \Omega, \\ u = 0 \quad \text{on} \quad \partial\Omega. \end{cases} \tag{1.2}$$

Observe that the solution (1.2) at the point x, $u(x)$, is a random variable, and its best (mean squared) approximation given $u(x_1), \dots, u(x_N)$ (the values of the solution of (1.2) at the points $x_1, \dots, x_N$ form the data) is its conditional expectation $\mathbb{E}\big[u(x)\big|u(x_1), \dots, u(x_N)\big]$. One result of this paper is that

$$\mathbb{E}\big[u(x)\big|u(x_1), \dots, u(x_N)\big] = \sum_{i=1}^N u(x_i)\phi_i(x), \tag{1.3}$$

where the functions ϕ_i are rough polyharmonic splines (RPS) [60] which have been identified as accurate basis elements for the numerical homogenization of (1.1) having noteworthy variational, optimal recovery, and localization properties. The discovery of these RPS has required a significant amount of work and trial and error, but here they are identified after a single step of Bayesian conditioning.

This observation motivates us to investigate what the same process of Bayesian conditioning would give under different priors and under observations other than the values of u at individual points (we will consider data formed by the values of a finite number of linear functions of u). In particular, we will use this link between Bayesian inference and numerical homogenization to identify bases for the numerical homogenization of arbitrary linear integro-differential equations. Our purpose is to show that this link is generic and could in principle be used, beyond numerical homogenization, as a guiding principle for the coarse-graining of multiscale systems. The Bayesian approach to this problem is to (1) put a prior on the degrees of freedom of the system, (2) select a finite number of coarse variables, and (3) compute the posterior value of the state of the system conditioned on the coarse variables.

2. General setup.

Let $\mathcal{L}$ and $\mathcal{B}$ be linear integro-differential operators on Ω and $\partial\Omega$ such that (1) $(\mathcal{L},\mathcal{B}) : \mathcal{H}(\Omega) \to \mathcal{H}_{\mathcal{L}}(\Omega) \times \mathcal{H}_{\mathcal{B}}(\partial\Omega)$, where $\mathcal{H}(\Omega)$, $\mathcal{H}_{\mathcal{L}}(\Omega)$, and $\mathcal{H}_{\mathcal{B}}(\partial\Omega)$ are Hilbert spaces of generalized functions on Ω and $\partial\Omega$, and (2) $\mathcal{H}_{\mathcal{L}}(\Omega)$ contains $L^2(\Omega)$ and $\mathcal{H}(\Omega)$ is contained in $L^2(\Omega)$.

Consider the integro-differential equation

$$\begin{cases} \mathcal{L}u(x) = g(x), & x \in \Omega, \\ \mathcal{B}u = 0 \quad \text{on} & \partial\Omega. \end{cases} \tag{2.1}$$

As with (1.1) the numerical homogenization of (2.1) will require the assumption that g belongs to a strict subspace of $\mathcal{H}_{\mathcal{L}}(\Omega)$.

We will assume that $\mathcal{L}$ and $\mathcal{B}$ are such that (2.1) admits (1) a unique solution in $\mathcal{H}(\Omega)$, and (2) a Green's function G. Recall that G is defined as the solution of

$$\begin{cases} \mathcal{L}G(x,y) = \delta(x-y), & x \in \Omega, \\ \mathcal{B}G(x,y) = 0 \quad \text{for} & x \in \partial\Omega, \end{cases} \tag{2.2}$$

where $\delta(\cdot - y)$ is the delta mass of Dirac at the point y.

Example 2.1. Note that for the prototypical example (1.1) we have

$$\mathcal{L}u(x) := -\operatorname{div}\big(a(x)\nabla u(x)\big) \quad \text{and} \quad \mathcal{B}u(x) = u(x). \tag{2.3}$$

Our purpose is to identify a good basis for the numerical homogenization or coarse-graining of (2.1).

3. Bayesian numerical homogenization.

Our Bayesian approach to the numerical homogenization of (2.1) is to replace the source term g by a Gaussian field ξ. More precisely, we introduce ξ, a centered Gaussian field on Ω with covariance function

$$\Lambda(x,y) := \mathbb{E}\big[\xi(x)\xi(y)\big], \tag{3.1}$$

and consider the stochastic integro-differential equation

$$\begin{cases} \mathcal{L}u(x) = \xi(x), & x \in \Omega, \\ \mathcal{B}u = 0 \quad \text{on} & \partial\Omega. \end{cases} \tag{3.2}$$

PROPOSITION 3.1. *The solution of* (3.2) *is a Gaussian field on* Ω *whose covariance function* $\Gamma(x,y) := \mathbb{E}\big[u(x)u(y)\big]$ *is*

$$\Gamma(x,y) = \int_{\Omega^2} G(x,z)\Lambda(z,z')G(y,z')\,dz\,dz'. \tag{3.3}$$

Remark 3.2. Denote $(\mathcal{L}^*,\mathcal{B}^*)$ as the adjoint of $(\mathcal{L},\mathcal{B})$ with respect to the (scalar) product defined on $\mathcal{H}(\Omega)$ by $\big\langle u,v\big\rangle_{L^2} := \int_\Omega u(x)v(x)\,dx$. Observe that $G(y,x)$ (the transpose of $G(x,y)$ with respect to the scalar product $\big\langle\cdot,\cdot\big\rangle_{L^2}$) is the Green's function of $(\mathcal{L}^*,\mathcal{B}^*)$ (the complex conjugation of the Green's function is not required to define its adjoint because the scalar product is bilinear and not sesquilinear). Observe that if ξ is white noise (i.e., $\Lambda(x-y)=\delta(x-y)$), then

$$\Gamma(x,y) = \int_{\Omega} G(x,z)G(y,z)\,dz, \tag{3.4}$$

which is the kernel of $\mathcal{L}^*\mathcal{L}$, i.e., $\mathcal{L}^*\mathcal{L}\Gamma(x,y)=\delta(x-y)$.

Proof. Since $\mathcal{L}$ and $\mathcal{B}$ are linear operators, u is a linear function of ξ and is therefore a Gaussian field. Moreover, its covariance function is given by

$$\begin{aligned}\Gamma(x,y) = \mathbb{E}\big[u(x)u(y)\big] &= \mathbb{E}\left[\int_{\Omega^2} G(x,z)\xi(z)G(y,z')\xi(z')\right]\,dz\,dz' \\ &= \int_{\Omega^2} G(x,z)G(y,z')\mathbb{E}\big[\xi(z)\xi(z')\big]\,dz\,dz',\end{aligned} \tag{3.5}$$

which finishes the proof. ∎

Remark 3.3. Beyond Bayesian homogenization, equations with random right-hand side can also be of interest in practical applications, for instance, in the modeling of the electrostatics in nanoscale field-effect sensors, where fluctuations arise from random charge concentrations [41].

3.1. On the choice of the noise. We will show that the choice of the noise Λ can be determined by the regularity of the source term g in the right-hand side of (2.1). More precisely, if ξ is white noise ($\Lambda(x,y)=\delta(x-y)$), then the resulting accuracy estimates will be obtained under the assumption that $g\in L^2(\Omega)$ and as a function of $\|g\|_{L^2(\Omega)}$.

If ξ is not white noise (i.e., if its covariance function is not $\delta(x-y)$), then we assume that there exist two linear integro-differential operators $\mathcal{L}_\Lambda$ and $\mathcal{B}_\Lambda$ such that ξ is the stochastic solution of the following equation with white noise ξ' as the source term:

$$\begin{cases}\mathcal{L}_\Lambda\xi(x)=\xi'(x), & x\in\Omega,\\ \mathcal{B}_\Lambda\xi=0 \quad \text{on} \quad \partial\Omega.\end{cases} \tag{3.6}$$

In what follows, if ξ is not white noise, then we assume it to be obtained as in (3.6), and the resulting accuracy estimates will be obtained under the assumption that $\mathcal{L}_\Lambda g\in L^2(\Omega)$ and as a function of $\|\mathcal{L}_\Lambda g\|_{L^2(\Omega)}$. A prototypical example corresponds to the situation where ξ is obtained as the regularization of white noise via a power of the Laplace–Dirichlet operator on Ω, and this allows us to identify optimal recovery bases under the assumption that $g\in H^s(\Omega)$ with $s\geq 0$ or $s<0$.

3.2. Identification of basis elements via conditioning. Let N be a strictly positive integer. Our Bayesian approach is based on the conditioning of the solution of (3.2) posterior to the observation of N linear functions of $u(x)$, expressed as

$$\int_{\Omega} u(x)\psi_i(x)\,dx, \quad i \in \{1,\dots,N\}, \tag{3.7}$$

where $\psi_1,\dots,\psi_N$ are N linearly independent generalized functions (distributions) on Ω such that for all i

$$\int_{\Omega^2} \psi_i(x)\Gamma(x,y)\psi_i(y)\,dx\,dy < \infty. \tag{3.8}$$

Examples of ψ_i include masses of Dirac ($\psi_i(x) = \delta(x-x_i)$), indicator functions of subsets of Ω, and elements of $L^1(\Omega)$. Let Θ be the $N\times N$ symmetric matrix defined by

$$\Theta_{i,j} := \int_{\Omega^2} \psi_i(x)\Gamma(x,y)\psi_j(y)\,dx\,dy. \tag{3.9}$$

Note that (3.8) implies that if u is the solution of (3.2), then

$$\Psi := \Big(\int_{\Omega} u(x)\psi_1(x)\,dx,\dots,\int_{\Omega} u(x)\psi_N(x)\,dx\Big) \tag{3.10}$$

is a well-defined center Gaussian random vector with covariance matrix Θ.

We will assume from now on that the covariance function (3.1) is not degenerate in the sense that for $f \in \mathcal{H}(\Omega)$,

$$\|f\|_{\Lambda}^2 := \int_{\Omega} f(x)\Lambda(x,y)f(y)\,dx\,dy \tag{3.11}$$

is zero if and only if f is the null function. Note that if ξ is obtained via (3.6), then $\|f\|_{\Lambda}^2 = \|\mathcal{L}_{\Lambda}^{-1}f\|_{L^2(\Omega)}^2$ (writing $\mathcal{L}_{\Lambda}^{-1}f$ as the solution of $\mathcal{L}_{\Lambda}u = f$ in Ω with $\mathcal{B}_{\Lambda}u = 0$ on $\partial\Omega$), and the nondegeneracy of Λ is equivalent to that of the operator $\mathcal{L}_{\Lambda}$.

LEMMA 3.4. *The $N\times N$ matrix Θ is symmetric positive definite. Furthermore, for all $l \in \mathbb{R}^N$,*

$$l^T\Theta l = \|v\|_{\Lambda}^2, \tag{3.12}$$

where v is the solution of

$$\begin{cases} \mathcal{L}^* v(x) = \sum_{j=1}^N l_j\psi_j(x) & \text{for } x \in \Omega,\\ \mathcal{B}^* v(x) = 0 & \text{for } x \in \partial\Omega. \end{cases} \tag{3.13}$$

Proof. We obtain from (3.3) that for $l \in \mathbb{R}^N$

$$l^T\Theta l = \int_{\Omega^2} \Big(\int_{\Omega}\sum_{i=1}^N \psi_i(x)G(x,z)\,dx\Big)\Lambda(z,z')\Big(\int_{\Omega}\sum_{j=1}^N \psi_j(y)G(y,z')\,dy\Big)\,dz\,dz'. \tag{3.14}$$

Write

$$v(x) := \sum_{i=1}^N l_i \int_{\Omega} G(y,x)\psi_i(y)\,dy. \tag{3.15}$$

Since $G(\cdot, x)$ is the Green's function of the adjoint operator (Remark 3.2), it follows that v is the solution of (3.13) and $\|v\|_\Lambda^2 = l^T\Theta l$, which implies that Θ is symmetric positive definite. Indeed if Θ is not positive definite, then there would exist a nonzero vector $l \in \mathbb{R}^N$ such that $\Theta l = 0$. This would imply that $\|v\|_\Lambda = 0$, which is a contradiction since (3.13) has a nonzero solution (since $l \neq 0$ and the ψ_i are linearly independent). ∎

Our motivation for using Gaussian noise in (3.2) lies in the fact that for Gaussian fields, conditional expected values can be computed via linear projection. Henceforth our approach is also akin to Gaussian filtering for numerical homogenization, and the following theorem shows that this approach allows for the identification of a (projection) basis ϕ_i.

THEOREM 3.5. *Let u be the solution of* (3.2), *and let Ψ be defined by* (3.10); *then*

$$\mathbb{E}\big[u(x)\big|\Psi\big] = \sum_{i=1}^{N} \Psi_i \phi_i(x), \tag{3.16}$$

with

$$\Psi_i := \int_\Omega u(y)\psi_i(y)\, dy, \tag{3.17}$$

and

$$\phi_i(x) := \sum_{j=1}^{N} \Theta^{-1}_{i,j} \int_\Omega \Gamma(x,y)\psi_j(y)\, dy. \tag{3.18}$$

Furthermore, $u(x)$ conditioned on the value of Ψ is a Gaussian random variable with mean (3.16) *and variance*

$$\sigma(x)^2 = \Gamma(x,x) - \sum_{i,j=1}^{N} \Theta^{-1}_{i,j} \int_\Omega \Gamma(x,y)\psi_j(y)\, dy \int_\Omega \Gamma(x,y)\psi_i(y)\, dy. \tag{3.19}$$

Proof. Let

$$u_\Psi(x) := \mathbb{E}\big[u(x)\big|\Psi\big]. \tag{3.20}$$

Since u and Ψ belong to the same Gaussian space, it follows that u_Ψ is a linear function of Ψ obtained by minimizing the mean squared error

$$\mathbb{E}\Big[\big(u(x) - c\cdot\Psi\big)^2\Big] = \Gamma(x,x) - 2\sum_{i=1}^{N} c_i \int_\Omega \Gamma(x,y)\psi_i(y)\, dy + \sum_{i,j=1}^{N} c_i c_j \Theta_{i,j}, \tag{3.21}$$

with respect to $c \in \mathbb{R}^N$, where Θ is defined by (3.9). We conclude the proof by identifying the minimizer in c, using Lemma 3.4 for the invertibility of Θ, and noting that (3.19) is simply (3.21) at the minimum in c. ∎

Example 3.1. If $\mathcal{L}$ and $\mathcal{B}$ correspond to the prototypical example (1.1) (see also Example 2.1), if ξ is white noise (i.e., if its covariance matrix is $\Lambda(x,y) = \delta(x-y)$), and if the observable functions are masses of Dirac at points $x_i \in \Omega$ (and $d \leq 3$ which is required for (3.8)), then Theorem 3.5 implies (1.3) and the basis elements ϕ_i are the RPS elements of [60], which are a generalization of polyharmonic splines to PDEs with rough coefficients. Recall that polyharmonic splines can be traced back to the seminal work of Harder and Desmarais [40] and Duchon [23, 24, 25].

Note also that according to Theorem 3.5 the process of Bayesian conditioning gives us the whole posterior distribution of $u(x)$ and not only its (conditional) expected value. In particular, the distribution of $u(x)$ conditioned on $u(x_1), \dots, u(x_N)$ is a Gaussian random variable with mean (1.3) and variance

$$\sigma^2(x) = \Gamma(x,x) - \sum_{i,j=1}^{N} \Theta^{-1}_{i,j} \Gamma(x,x_j)\Gamma(x,x_i), \tag{3.22}$$

and this observation can be used to compute the probability of deviation of the RPS interpolation from $u(x)$ by a given margin and guide the addition of interpolation points (note that $\sigma^2(x) = 0$ at the interpolation points $x_1, \dots, x_N$).

Remark 3.6. We will show in Theorem 5.1 that $\sigma(x)$ also controls the pointwise error between the solution of the original integro-differential equation (2.1) and the approximation $\sum_{i=1}^N \phi_i(x) \int_\Omega u(y)\psi_i(y)\,dy$.

4. Variational properties of basis elements. In this section we will show that, as for RPS [60], the basis elements ϕ_i from Bayesian inference have remarkable variational and optimal recovery properties that can be used (1) for their practical computation, and (2) for the derivation of accuracy estimates.

4.1. White Gaussian noise. In this subsection we will assume that ξ is white noise (i.e., $\Lambda(x,y) = \delta(x-y)$). Define

$$V := \big\{ \phi \in \mathcal{H}(\Omega) \big| \mathcal{L}\phi \in L^2(\Omega) \text{ and } \mathcal{B}\phi = 0 \text{ on } \partial\Omega \big\}, \tag{4.1}$$

and let $\langle \cdot, \cdot \rangle$ be the (scalar) product on V defined by, for $u, v \in V$,

$$\langle u, v \rangle := \int_\Omega \big(\mathcal{L}u(x)\big)\big(\mathcal{L}v(x)\big)\,dx. \tag{4.2}$$

Note in particular that $\langle v, v \rangle = 0$ if and only if $v = 0$. Then we write

$$\|v\|_V := \langle v, v \rangle^{\frac{1}{2}}, \tag{4.3}$$

the corresponding norm (note that $\|v\|_V$ is a norm on V because $\|v\|_V = 0$ and $v \in V$ imply $\mathcal{L}v = 0$ in Ω and $\mathcal{B}v = 0$ on $\partial\Omega$, which leads to $v = 0$ by the nondegeneracy of the operator $\mathcal{L}$).

THEOREM 4.1. *If $\Gamma(x,x) < \infty$, then for $v \in V$ and $x \in \Omega$,*

$$\big|v(x)\big| \leq \big(\Gamma(x,x)\big)^{\frac{1}{2}} \|v\|_V, \tag{4.4}$$

and the space V with the reproducing kernel $\Gamma(x,y)$ forms a reproducing kernel Hilbert space. In particular, for all $v \in V$,

$$\big\langle v, \Gamma(\cdot, x) \big\rangle = v(x). \tag{4.5}$$

Proof. Theorem 4.1 is a direct consequence of the fact that

$$\Big\langle v, \int_\Omega \Gamma(\cdot, y) f(y) dy \Big\rangle = \int_\Omega v(y) f(y)\,dy \tag{4.6}$$

and (by Cauchy–Schwartz inequality and $\langle \Gamma(\cdot,x), \int_\Omega \Gamma(\cdot,x)\rangle = \Gamma(x,x)$)

$$\langle v, \Gamma(\cdot,x)\rangle \le \langle v,v\rangle^{\frac{1}{2}} \big(\Gamma(x,x)\big)^{\frac{1}{2}}. \qquad \square \tag{4.7}$$

Define

$$V_i := \Big\{ \phi \in V \,\Big|\, \int_\Omega \phi(x)\psi_i(x)\,dx = 1 \text{ and } \int_\Omega \phi(x)\psi_j(x)\,dx = 0 \\ \text{for } j \in \{1,\dots,N\} \text{ such that } j \neq i \Big\}, \tag{4.8}$$

and consider the following optimization problem over V_i:

$$\begin{cases} \text{Minimize } \langle \phi,\phi\rangle \\ \text{subject to } \phi \in V_i. \end{cases} \tag{4.9}$$

PROPOSITION 4.2. *V_i is a nonempty closed affine subspace of V. Problem* (4.9) *is a strictly convex quadratic optimization problem over V_i. The unique minimizer of* (4.9) *is ϕ_i as defined by* (3.18).

Proof. Let us first prove that $\phi_i \in V_i$. Let

$$\theta_i(x) := \int_\Omega \Gamma(x,y)\psi_i(y)\,dy. \tag{4.10}$$

First observe that for all $i \in \{1,\dots,N\}$,

$$\mathcal{L}\theta_i(x) = \int_\Omega G(y,x)\psi_i(y)\,dy, \tag{4.11}$$

and $\mathcal{B}\theta_i(x) = 0$ on $\partial\Omega$. Noting that $\|\mathcal{L}\theta_i\|^2_{L^2(\Omega)} = \Theta_{i,i}$, we deduce from (3.8) that $\theta_i \in V$. We conclude from (3.18) and Lemma 3.4 that $\phi_i \in V$. Now observe that (3.9) implies that

$$\int_\Omega \phi_i(x)\psi_j(x) = (\Theta^{-1}\cdot\Theta)_{i,j} = \delta_{i,j}, \tag{4.12}$$

where $\delta_{i,i} = 1$ and $\delta_{i,j} = 0$ for $j \neq i$. We conclude that $\phi_i \in V_i$, which implies that V_i is nonempty (it is easy to check that it is a closed affine subspace of V).

Now let us prove that problem (4.9) is a strictly convex optimization problem over V_i. Let $v, w \in V_i$ such that $v \neq w$. We write for $\lambda \in [0,1]$

$$f(\lambda) := \langle v + \lambda(w-v), v + \lambda(w-v)\rangle, \tag{4.13}$$

and we need to show that $f(\lambda)$ is a strictly convex function. Observing that

$$f(\lambda) = \langle v,v\rangle + 2\lambda\langle v, w-v\rangle + \lambda^2\langle v-w, v-w\rangle \tag{4.14}$$

and noting that $\langle v-w, v-w\rangle > 0$ (otherwise one would have $v = w$), we deduce that f is strictly convex in λ. We conclude (see, for example, [30, Proposition 1.2, p. 35]) that problem (4.9) is a strictly convex optimization problem over V_i and that it admits a unique minimizer in V_i. We will postpone the proof of the fact that ϕ_i is the minimizer of (4.9) to the proof of Theorem 4.6. $\square$

Remark 4.3. It is important to note that in practical (numerical) applications, each element ϕ_i would be obtained by solving the quadratic optimization problem

(4.9) rather than through the representation formula (3.18) because the identification of Γ in (3.18) is more expensive than solving the linear systems associated with (4.9) (inverting a matrix is more expensive than solving a linear system). Note also that if u is the (stochastic) solution of (3.2), then ϕ_i is also equal to the expected value of $u(x)$ conditioned on $\int_\Omega u(x)\psi_i(x) = 1$ and $\int_\Omega u(x)\psi_j(x) = 0$ for $j \neq i$, i.e.,

$$\phi_i(x) = \mathbb{E}\Big[u(x)\Big|\int_\Omega u(x)\psi_i(x) = 1 \text{ and } \int_\Omega u(x)\psi_j(x) = 0 \text{ for } j \neq i\Big]. \tag{4.15}$$

Remark 4.4. A simple calculation allows us to show that ϕ_i is also the solution of the following nested equations:

$$\begin{cases} \mathcal{L}\phi_i(x) = \chi_i(x), & x \in \Omega, \\ \mathcal{B}\phi_i = 0 \quad \text{on} & \partial\Omega, \end{cases} \tag{4.16}$$

$$\begin{cases} \mathcal{L}^*\chi_i(x) = \sum_{j=1}^N \Theta^{-1}_{i,j}\psi_j(x), & x \in \Omega, \\ \mathcal{B}^*\chi_i(x) = 0 \quad \text{on} & \partial\Omega. \end{cases} \tag{4.17}$$

Remark 4.5. Another simple calculation allows us to show that ϕ_i is also the solution of the following nested equations:

$$\begin{cases} \mathcal{L}\phi_i(x) = \chi_i(x), & x \in \Omega, \\ \mathcal{B}\phi_i = 0 \quad \text{on} & \partial\Omega, \\ \int_\Omega \phi_i(x)\psi_j(x)\,dx = \delta_{i,j} \quad \text{for} & j \in \{1,\dots,N\}, \end{cases} \tag{4.18}$$

$$\begin{cases} \mathcal{L}^*\chi_i(x) = \sum_{j=1}^N c_j\psi_j(x), & x \in \Omega, \\ \mathcal{B}^*\chi_i(x) = 0 \quad \text{on} & \partial\Omega, \end{cases} \tag{4.19}$$

where $c \in \mathbb{R}^N$ is an unknown vector determined by the third equation in (4.18).

Write V_0 as the subset of V defined by

$$V_0 := \Big\{v \in V : \int_\Omega v(x)\psi_i(x)\,dx = 0 \quad \forall i \in \{1,\dots,N\}\Big\}. \tag{4.20}$$

THEOREM 4.6. *The following hold true:*

- *The basis ϕ_i is orthogonal to V_0 with respect to the product $\langle\cdot,\cdot\rangle$, i.e.,*

$$\langle \phi_i, v\rangle = 0 \quad \forall i \in \{1,\dots,N\} \text{ and } \forall v \in V_0. \tag{4.21}$$

- *$\sum_{i=1}^N w_i\phi_i$ is the unique minimizer of $\langle v,v\rangle$ over all $v \in V$ such that $\int_\Omega v(x)\psi_i(x)\,dx = w_i$.*
- *For all $i \in \{1,\dots,N\}$ and for all $v \in V$,*

$$\langle \phi_i, v\rangle = \sum_{j=1}^N \Theta^{-1}_{i,j} \int_\Omega v(x)\psi_j(x)\,dx. \tag{4.22}$$

- *For all $i,j \in \{1,\dots,N\}$,*

$$\langle \phi_i, \phi_j\rangle = \Theta^{-1}_{i,j}. \tag{4.23}$$

Remark 4.7. Theorem 4.6 and its proof are analogous to the optimal property of strictly conditionally positive definite kernels [69] when used as interpolant solutions of the optimal recovery problem [38].

Proof. We have, using (4.10), (3.18), and (4.11),

$$\langle \phi_i, v\rangle = \sum_{j=1}^{N} \Theta^{-1}_{i,j} \int_\Omega \mathcal{L}\theta_j(x)\mathcal{L}v(x)\,dx = \sum_{j=1}^{N} \Theta^{-1}_{i,j} \int_\Omega \psi_j(y)v(y)\,dy = 0, \tag{4.24}$$

which implies (4.21), (4.22), and (4.23).

Let $w \in \mathbb{R}^N$ and $\phi_w := \sum_{i=1}^N w_i\phi_i$. Let $v \in V$ such that $\int_\Omega v(x)\psi_i(x)\,dx = w_i$ for all $i \in \{1,\dots,N\}$. Since $\phi_w - v \in V_0$, it follows that

$$\langle v, v\rangle = \langle \phi_w, \phi_w\rangle + \langle v-\phi_w, v-\phi_w\rangle. \tag{4.25}$$

It follows that $\sum_{i=1}^N w_i\phi_i$ is the unique minimizer of $\langle v, v\rangle$ over all $v \in V$ such that $\int_\Omega v(x)\psi_i(x)\,dx = w_i$. Note that this also implies that ϕ_i is the minimizer of (4.9). □

4.2. Nonwhite Gaussian noise. If ξ is not white noise (i.e., $\Lambda(x,y) \neq \delta(x-y)$), then Theorem 4.1, Theorem 4.6, and Proposition 4.2 remain true, provided that the definitions of the space V and scalar product $\langle \cdot,\cdot\rangle$ are changed to

$$V := \{\phi \in \mathcal{H}(\Omega) \,|\, \mathcal{L}_\Lambda\mathcal{L}\phi \in L^2(\Omega),\ \mathcal{B}\phi = 0 \text{ and } \mathcal{B}_\Lambda\mathcal{L}\phi = 0 \text{ on } \partial\Omega\}, \tag{4.26}$$

$$\langle u, v\rangle := \int_\Omega \big(\mathcal{L}_\Lambda\mathcal{L}u(x)\big)\big(\mathcal{L}_\Lambda\mathcal{L}v(x)\big)\,dx, \tag{4.27}$$

where $\mathcal{L}_\Lambda$ and $\mathcal{B}_\Lambda$ are defined as in (3.6).

5. Accuracy of the basis elements ϕ_i.

5.1. Pointwise estimates. Let $\|v\|_V$ be defined as in (4.3).

THEOREM 5.1. *Assume that $\Gamma(x,x) < \infty$. Let $v \in V$. It holds true that for $x \in \Omega$,*

$$\left| v(x) - \sum_{i=1}^{N} \phi_i(x)\Big(\int_\Omega v(y)\psi_i(y)\,dy\Big)\right| \leq \sigma(x)\|v\|_V, \tag{5.1}$$

where $\sigma^2(x)$ is the variance of $u(x)$ (solution of (3.2)) conditioned on $\int_\Omega u(y)\psi_1(y)\,dy$, $\dots, \int_\Omega u(y)\psi_N(y)\,dy$ as defined by (3.19). In particular, if u is the solution of the original integro-differential equation (2.1), then

$$\left| u(x) - \sum_{i=1}^{N} \phi_i(x)\Big(\int_\Omega u(y)\psi_i(y)\,dy\Big)\right| \leq \sigma(x)\|g\|_{L^2(\Omega)} \tag{5.2}$$

if ϕ_i, σ are derived from white noise, and

$$\left| u(x) - \sum_{i=1}^{N} \phi_i(x)\Big(\int_\Omega u(y)\psi_i(y)\,dy\Big)\right| \leq \sigma(x)\|\mathcal{L}_\Lambda g\|_{L^2(\Omega)} \tag{5.3}$$

if ϕ_i, σ are derived from the noise with covariance function Λ described in (3.6).

Proof. Let $v \in V$ and $x \in \Omega$. Using the reproducing kernel property of Theorem 4.1, we obtain that

$$\left|v(x)-\sum_{i=1}^{N}\phi_i(x)\int_\Omega v(y)\psi_i(y)\,dy\right| = \left|\left\langle v, \Gamma(\cdot,x)-\sum_{i=1}^{N}\phi_i(x)\int_\Omega \Gamma(\cdot,y)\psi_i(y)\,dy\right\rangle\right|. \tag{5.4}$$

Therefore, using Cauchy–Schwartz inequality,

$$\left|v(x)-\sum_{i=1}^{N}\phi_i(x)\int_\Omega v(y)\psi_i(y)\,dy\right| \leq \|v\|_V \left\|\Gamma(\cdot,x)-\sum_{i=1}^{N}\phi_i(x)\int_\Omega \Gamma(\cdot,y)\psi_i(y)\,dy\right\|_V. \tag{5.5}$$

We conclude by expanding the right-hand side of (5.5) and using the definition $\phi_i(x) = \sum_{j=1}^{N}\Theta^{-1}_{i,j}\int_\Omega \Gamma(x,y)\psi_i(y)\,dy$. ∎

Remark 5.2. $\sigma^2(x)$ is also known as the power function in radial basis function interpolation [69, 34]. The proof of Theorem 5.1 is similar to the one used to derive local error estimates for radial basis function interpolation of scattered data (see [71], in which $\sigma^2(x)$ was referred to as the Kriging function, terminology coming from geostatistics [47]).

5.2. $\mathcal{H}(\Omega)$-norm estimates. Let V_0 be the subset of V defined by (4.20). Write

$$\rho(V_0) := \sup_{v\in V_0}\frac{\|v\|_{\mathcal{H}(\Omega)}}{\|v\|_V}, \tag{5.6}$$

where $\|.\|_{\mathcal{H}(\Omega)}$ is the natural norm associated with the space on which the operator $\mathcal{L}$ is defined.

THEOREM 5.3. *We have for all $v \in V$*

$$\left\|v-\sum_{i=1}^{N}\phi_i\left(\int_\Omega v(y)\psi_i(y)\,dy\right)\right\|_{\mathcal{H}(\Omega)} \leq \rho(V_0)\|v\|_V, \tag{5.7}$$

and $\rho(V_0)$ is the smallest constant for which (5.7) *holds for all $v \in V$.*

Proof. Write $v_\Psi(x) := \sum_{i=1}^{N}\phi_i(x)(\int_\Omega v(y)\psi_i(y)\,dy)$.

Observing that $v - v_\Psi$ belongs to V_0 implies that

$$\|v-v_\Psi\|_{\mathcal{H}(\Omega)} \leq \rho(V_0)\big\langle v-v_\Psi, v-v_\Psi\big\rangle^{\frac{1}{2}}. \tag{5.8}$$

Theorem 4.6 implies that

$$\big\langle v,v\big\rangle = \big\langle v_\Psi, v_\Psi\big\rangle + \big\langle v-v_\Psi, v-v_\Psi\big\rangle, \tag{5.9}$$

which leads to

$$\big\langle v-v_\Psi, v-v_\Psi\big\rangle = \big\langle v,v\big\rangle - \big\langle v_\Psi,v_\Psi\big\rangle \leq \big\langle v,v\big\rangle, \tag{5.10}$$

which concludes the proof. ∎

Remark 5.4. Observe that Theorem 5.3 implies that if u is the solution of the original integro-differential equation (2.1) and ϕ_i, σ are derived from white noise, then

$$\left\|u-\sum_{i=1}^{N}\phi_i\left(\int_\Omega u(y)\psi_i(y)\,dy\right)\right\|_{\mathcal{H}(\Omega)} \leq \rho(V_0)\|g\|_{L^2(\Omega)}. \tag{5.11}$$

Similarly, if ϕ_i, σ are derived from the noise with covariance function Λ described in (3.6), then

$$\left\| u - \sum_{i=1}^{N} \phi_i \left(\int_\Omega u(y)\psi_i(y)\,dy \right) \right\|_{\mathcal{H}(\Omega)} \leq \rho(V_0) \|\mathcal{L}_\Lambda g\|_{L^2(\Omega)}. \tag{5.12}$$

Example 5.1. If $\mathcal{L}$ and $\mathcal{B}$ correspond to the prototypical example (1.1) (Example 2.1), if ξ is white noise, and if the observable functions are masses of Dirac at points $x_i \in \Omega$ (and $d \leq 3$), then [60]

$$\rho(V_0) \leq CH, \tag{5.13}$$

where C depends only on $\lambda_{\min}(a), \lambda_{\max}(a)$ and where $\lambda_{\max}(a) := \sup_{x\in\Omega, l\neq 0} l^T a(x) l/|l|^2$, $\lambda_{\min}(a) := \inf_{x\in\Omega, l\neq 0} l^T a(x) l/|l|^2$, and H is the mesh norm

$$H := \sup_{x\in\Omega} \min_i \|x - x_i\|, \tag{5.14}$$

and

$$\left\| u - \sum_{i=1}^{N} \phi_i(x) u(x_i) \right\|_{\mathcal{H}_0^1(\Omega)} \leq CH \big\| \operatorname{div}(a\nabla u) \big\|_{L^2(\Omega)}. \tag{5.15}$$

Let us also recall that the proof of (5.13) is based on the following Poincaré inequality (Lemma 3.1 of [60]).

LEMMA 5.5 (see [60, Lemma 3.1]). *Let $d \leq 3$ and B_1 be the open ball of center 0 and radius 1. There exists a finite strictly positive constant $C_{\lambda_{\min}(a),\lambda_{\max}(a)}$ such that for all $v \in \mathcal{H}^1(B_1)$ such that $\operatorname{div}(a\nabla v) \in L^2(B_1)$ it holds true that*

$$\|v - v(0)\|_{L^2(B_1)}^2 \leq C_{\lambda_{\min}(a),\lambda_{\max}(a)} \Big(\|\nabla v\|_{L^2(B_1)}^2 + \big\| \operatorname{div}(a\nabla v) \big\|_{L^2(B_1)}^2 \Big). \tag{5.16}$$

Proof. We will recall the proof of this lemma (as presented in [60, Lemma 3.1]) for the sake of completeness. The proof is per absurdum. Note that since $d \leq 3$ the assumptions $v \in \mathcal{H}^1(B_1)$ and $\operatorname{div}(a\nabla v) \in L^2(B_1)$ imply the Hölder continuity of v in B_1. Assume that (5.16) does not hold. Then there exist a sequence v_n and a sequence a'_n whose maximum and minimum eigenvalues are uniformly bounded by $\lambda_{\min}(a)$ and $\lambda_{\max}(a)$ (we need to introduce that sequence because we want the constant in (5.16) to depend only on $d, \lambda_{\min}(a), \lambda_{\max}(a)$) such that

$$\|v_n - v_n(0)\|_{L^2(B_1)}^2 > n \Big(\|\nabla v_n\|_{L^2(B_1)}^2 + \big\| \operatorname{div}(a'_n \nabla v_n) \big\|_{L^2(B_1)}^2 \Big). \tag{5.17}$$

Letting $w_n = \frac{v_n - v_n(0)}{\|v_n - v_n(0)\|_{L^2(B_1)}}$, we obtain that $w_n(0) = 0$, $\|w_n\|_{L^2(B_1)} = 1$, and

$$\|\nabla w_n\|_{L^2(B_1)}^2 + \big\| \operatorname{div}(a'_n \nabla w_n) \big\|_{L^2(B_1)}^2 < \frac{1}{n}. \tag{5.18}$$

Since

$$\|w_n\|_{\mathcal{H}^1(B_1)} < 1 + \frac{1}{n} \leq 2, \tag{5.19}$$

it follows that there exist a subsequence w_{n_j} and a $w \in \mathcal{H}^1(B_1)$ such that $w_{n_j} \rightharpoonup w$ weakly in $\mathcal{H}^1(B_1)$ and $\nabla w_{n_j} \rightharpoonup \nabla w$ weakly in $L^2(B_1)$. Using $\|\nabla w_n\|_{L^2(B_1)} \leq 1/n$, we deduce that $\nabla w = 0$, which implies that w is a constant in B_1. Since by the Rellich–Kondrachov theorem the embedding $\mathcal{H}^1(B_1) \subset L^2(B_1)$ is compact, it follows from (5.19) that $w_{n_j} \to w$ strongly in $L^2(B_1)$, which (using $\|w_n\|_{L^2(B_1)} = 1$) implies that $\|w\|_{L^2(B_1)} = 1$. Now (5.19), together with the fact that $\big\|\operatorname{div}(a'_n \nabla w_n)\big\|^2_{L^2(B_1)}$ is uniformly bounded and that $d \leq 3$, implies that w_n is uniformly Hölder continuous on $B(0, \frac{1}{2})$ (see, for instance, [65]). This implies that w is continuous in $B(0, \frac{1}{2})$ and that $w(0) = 0$. This contradicts the fact that w is a constant in B_1 with $\|w\|_{L^2(B_1)} = 1$. ∎

Example 5.2. If $\mathcal{L}$ and $\mathcal{B}$ correspond to the prototypical example (1.1) (Example 2.1), if ξ is white noise, and if the observable functions are indicator functions of Voronoi cells around points in $x_i \in \Omega$ or of tetrahedra of a regular tessellation of the points $x_i \in \Omega$, then (5.13) remains valid as a simple consequence of localized Poincaré inequalities. Indeed for $v \in V_0$, writing C_i as the Voronoi cells at the points $x_i \in \Omega$, we have (assuming Ω is the union of those Voronoi cells)

$$\|v\|^2_{\mathcal{H}^1_0(\Omega)} = \int_\Omega v(x)\big(-\operatorname{div}(a(x)\nabla v(x))\big)\,dx \leq \|v\|_{L^2(\Omega)} \big\|\operatorname{div}(a\nabla v)\big\|_{L^2(\Omega)}, \tag{5.20}$$

and we conclude by applying Poincaré's inequality to the L^2-norm of v within each cell C_i, i.e.,

$$\|v\|^2_{L^2(\Omega)} = \sum_i \|v\|^2_{L^2(C_i)} \leq CH^2 \sum_i \|\nabla v\|^2_{L^2(C_i)} = CH^2 \|\nabla v\|^2_{L^2(\Omega)}. \tag{5.21}$$

We will give the last example as a theorem.

THEOREM 5.6. *Let $\mathcal{L}$ and $\mathcal{B}$ be as in the prototypical example* (1.1) *(Example* 2.1*), and let ξ be white noise. Let $\psi_1, \dots, \psi_N$ be linearly independent generalized probability densities on Ω with (possibly overlapping) support* $\operatorname{support}(\psi_i)$. *Define*

$$H := \sup_{x\in\Omega} \min_i \sup_{y \in \operatorname{support}(\psi_i)} \|x - y\|. \tag{5.22}$$

Then, it holds true that

$$\rho(V_0) \leq CH, \tag{5.23}$$

where C depends only on $\lambda_{\min}(a)$ and $\lambda_{\max}(a)$. Henceforth, for $u \in V$,

$$\Big\| u - \sum_{i=1}^N \phi_i(x) \int_\Omega u(y)\psi_i(y)\,dy \Big\|_{\mathcal{H}^1_0(\Omega)} \leq CH \big\|\operatorname{div}(a\nabla u)\big\|_{L^2(\Omega)}. \tag{5.24}$$

Remark 5.7. Observe that if for all i the support of ψ_i is contained in a ball of center x_i and radius H', then

$$H \leq H' + \sup_{x\in\Omega} \min_i \|x - x_i\|; \tag{5.25}$$

in particular, if the points x_i have mesh norm H'' (see (5.14)), then $H \leq H' + H''$.

Proof. The proof of (5.23) is simply based on the observation that if $v \in V_0$, then (since $\int_\Omega v(x)\psi_i(x)\,dx = 0$) there exist N points $y_1, \dots, y_N$ such that $v(y_i) = 0$ and the mesh norm of those points is bounded by H. Therefore we can apply the result of Example 5.1. ∎

6. Pseudoalgorithm. A simple pseudoalgorithmic description of the proposed framework for the numerical homogenization of (2.1) is as follows:

1. Select N linearly independent (measurement) functions $\psi_1, \dots, \psi_N$ in $L^2(\Omega)$.
2. Let ξ in (3.2) be a Gaussian field of mean 0 and covariance function $\Lambda(x, y)$ (assumed to be nondegenerate, i.e., such that there exists an inverse covariance function $\Lambda^{-1}(x, y)$ with $\int_{\Omega^2} \Lambda(x, y)\Lambda^{-1}(y, z)\, dy = \delta(x - z)$).
3. The basis functions $\phi_1, \dots, \phi_N$ for the numerical homogenization of (2.1) are identified as (writing u as the solution of (3.2) and $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ if $i \neq j$) the deterministic functions

$$\phi_i(x) = \mathbb{E}\left[u(x) \middle| \int_\Omega u(x)\psi_j(x)\, dx = \delta_{i,j} \text{ for } j = 1, \dots, N\right]. \tag{6.1}$$

4. Each ϕ_i can also be identified as the unique minimizer of

$$\begin{cases} \text{Minimize } \int_{\Omega^2} (\mathcal{L}u(x))\Lambda^{-1}(x, y)(\mathcal{L}u(y))\, dx\, dy \\ \text{subject to } \phi \in \mathcal{H}(\Omega) \text{ and } \int_\Omega \phi(x)\psi_j(x)\, dx = \delta_{i,j} \text{ for } j = 1, \dots, N. \end{cases} \tag{6.2}$$

5. Under appropriate choices of the measurement functions ψ_i and the covariance function $\Lambda(x, y)$, the basis functions ϕ_i can be computed by localizing the optimization problems (6.2) to subdomains of Ω.

7. Statistical decision theory and practical applications. Another motivation for exploring Bayesian approximations of the solution space lies in the decision theory/game theory approach to numerical homogenization. In this approach one looks at the numerical homogenization problem (1.1) as a repeated game where player B chooses a function θ of the linear measurements (data) $\int_\Omega u(x)\psi_1(x)\, dx, \dots, \int_\Omega u(x)\psi_N(x)\, dx$ and player A chooses a source term g in the unit ball of $L^2(\Omega)$. These two choices combine and form an error term

$$\mathcal{E}(\theta, g) = \left\| u - \theta\left(\int_\Omega u(x)\psi_1(x)\, dx, \dots, \int_\Omega u(x)\psi_N(x)\, dx \right) \right\|_{L^2(\Omega)}. \tag{7.1}$$

Player B's objective is to minimize the error (7.1), while player A's objective is to maximize it. A surprising result stemming from a generalization [51] of Wald's decision theory [68] and Von Neumann's game theory [67] is that, although such games are deterministic, under weak regularity conditions, the optimal strategy for player A is to play at random by placing an optimal probability distribution π_A on the set of candidates for g, and, similarly, the best strategy for player B is to assume that player A is playing at random and to use a function θ living in the Bayesian class (obtained by placing a prior π_B on the set of candidates for g and conditioning with respect to the measurements $\int_\Omega u(x)\psi_i(x)\, dx$).

Although the estimator employed by player B may be called Bayesian, the game described here is not (i.e., the choice of player A might be distinct from that of player B), and player B must solve a min-max optimization problem over π_A and π_B to identify an optimal prior distribution for the Bayesian estimator (a careful choice of the prior also appears to be important due to the possible high sensitivity of posterior distributions [55, 54, 53, 52]).

We refer the reader to [51] for (1) the complete description of the generalization of the Bayesian framework described here to the decision theory/information game formulation (described above), and (2) practical (including numerical) applications of that generalized framework to the problems of finding numerical homogenization bases and fast solvers for (1.1). In that generalization, optimal numerical homogenization base functions are obtained by selecting the prior distribution of ξ (in (1.2)) to be that of a Gaussian field with mean zero and covariance function the operator (1.1) (i.e., such that for $f \in H_0^1(\Omega)$, $\int_\Omega f(x)\xi(x)\,dx$ is a Gaussian random variable of mean zero and variance $\int_\Omega (\nabla f(x))^T a(x) \nabla f(x)\,dx$). In particular, [51] shows how the identification of an optimal distribution for ξ (in the Gaussian class) leads to the (automated) discovery of multigrid and multiresolution solvers for PDEs with rough coefficients.

Acknowledgments. The author thanks Dongbin Xiu, Lei Zhang, and Guillaume Bal for stimulating discussions, and Leonid Berlyand for comments on the manuscript. The author also thanks two anonymous referees for valuable comments and suggestions.

REFERENCES

[1] A. Abdulle and M. J. Grote, *Finite element heterogeneous multiscale method for the wave equation*, Multiscale Model. Simul., 9 (2011), pp. 766–792.

[2] A. Abdulle and C. Schwab, *Heterogeneous multiscale FEM for diffusion problems on rough surfaces*, Multiscale Model. Simul., 3 (2005), pp. 195–220.

[3] G. Allaire and R. Brizzi, *A multiscale finite element method for numerical homogenization*, Multiscale Model. Simul., 4 (2005), pp. 790–812.

[4] T. Arbogast and K. J. Boyd, *Subgrid upscaling and mixed multiscale finite elements*, SIAM J. Numer. Anal., 44 (2006), pp. 1150–1171.

[5] T. Arbogast, C.-S. Huang, and S.-M. Yang, *Improved accuracy for alternating-direction methods for parabolic equations based on regular and mixed finite elements*, Math. Models Methods Appl. Sci., 17 (2007), pp. 1279–1305.

[6] I. Babuška, G. Caloz, and J. E. Osborn, *Special finite element methods for a class of second order elliptic problems with rough coefficients*, SIAM J. Numer. Anal., 31 (1994), pp. 945–981.

[7] I. Babuska and R. Lipton, *Optimal local approximation spaces for generalized finite element methods with application to multiscale problems*, Multiscale Model. Simul., 9 (2011), pp. 373–406.

[8] I. Babuška, F. Nobile, and R. Tempone, *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM Rev., 52 (2010), pp. 317–355.

[9] I. Babuška and J. E. Osborn, *Generalized finite element methods: Their performance and their relation to mixed methods*, SIAM J. Numer. Anal., 20 (1983), pp. 510–536.

[10] I. Babuška and J. E. Osborn, *Can a finite element method perform arbitrarily badly?*, Math. Comp., 69 (2000), pp. 443–462.

[11] G. Bal and W. Jing, *Corrector theory for MSFEM and HMM in random media*, Multiscale Model. Simul., 9 (2011), pp. 1549–1587.

[12] A. Bensoussan, J.-L. Lions, and G. Papanicolaou, *Asymptotic Analysis for Periodic Structures*, North Holland, Amsterdam, New York, 1978.

[13] L. Berlyand and H. Owhadi, *Flux norm approach to finite dimensional homogenization approximations with non-separated scales and high contrast*, Arch. Ration. Mech. Anal., 198 (2010), pp. 677–721.

[14] M. Bieri and C. Schwab, *Sparse high order FEM for elliptic sPDEs*, Comput. Methods Appl. Mech. Engrg., 198 (2009), pp. 1149–1170.

[15] X. Blanc, C. Le Bris, and P.-L. Lions, *Une variante de la théorie de l'homogénéisation stochastique des opérateurs elliptiques*, C. R. Math. Acad. Sci. Paris, 343 (2006), pp. 717–724.

[16] X. Blanc, C. Le Bris, and P.-L. Lions, *Stochastic homogenization and random lattices*, J. Math. Pures Appl. (9), 88 (2007), pp. 34–63.

[17] L. V. Branets, S. S. Ghai, S. L. Lyons, and X.-H. Wu, *Challenges and technologies in reservoir modeling*, Commun. Comput. Phys., 6 (2009), pp. 1–23.

[18] L. A. Caffarelli and P. E. Souganidis, *A rate of convergence for monotone finite difference approximations to fully nonlinear, uniformly elliptic PDEs*, Comm. Pure Appl. Math., 61 (2008), pp. 1–17.

[19] C.-C. Chu, I. G. Graham, and T. Y. Hou, *A new multiscale finite element method for high-contrast elliptic interface problems*, Math. Comp., 79 (2010), pp. 1915–1955.

[20] E. De Giorgi, *Sulla convergenza di alcune successioni d'integrali del tipo dell'aera*, Rend. Mat. (6), 8 (1975), pp. 277–294.
[21] P. Diaconis, *Bayesian numerical analysis*, in Statistical Decision Theory and Related Topics, IV, Vol. 1 (West Lafayette, IN, 1986), Springer, New York, 1988, pp. 163–175.
[22] A. Doostan and H. Owhadi, *A non-adapted sparse approximation of PDEs with stochastic inputs*, J. Comput. Phys., 230 (2011), pp. 3015–3034.
[23] J. Duchon, *Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces*, RAIRO Anal. Numér., 10 (1976), pp. 5–12.
[24] J. Duchon, *Splines minimizing rotation-invariant semi-norms in Sobolev spaces*, in Constructive Theory of Functions of Several Variables (Proc. Conf., Math. Res. Inst., Oberwolfach, 1976), Lecture Notes in Math. 571, Springer, Berlin, 1977, pp. 85–100.
[25] J. Duchon, *Sur l'erreur d'interpolation des fonctions de plusieurs variables par les D^m-splines*, RAIRO Anal. Numér., 12 (1978), pp. 325–334, vi.
[26] W. E and B. Engquist, *The heterogeneous multiscale methods*, Commun. Math. Sci., 1 (2003), pp. 87–132.
[27] Y. Efendiev, J. Galvis, and X. Wu, *Multiscale finite element and domain decomposition methods for high-contrast problems using local spectral basis functions*, J. Comput. Phys., 230 (2011), pp. 937–955.
[28] Y. Efendiev, V. Ginting, T. Hou, and R. Ewing, *Accurate multiscale finite element methods for two-phase flow simulations*, J. Comput. Phys., 220 (2006), pp. 155–174.
[29] Y. Efendiev and T. Hou, *Multiscale finite element methods for porous media flows and their applications*, Appl. Numer. Math., 57 (2007), pp. 577–596.
[30] I. Ekeland and R. Témam, *Convex Analysis and Variational Problems*, Classics Appl. Math. 28, SIAM, Philadelphia, 1999.
[31] M. S. Eldred, *Design under uncertainty employing stochastic expansion methods*, Int. J. Uncertain. Quantif., 1 (2011), pp. 119–146.
[32] B. Engquist, H. Holst, and O. Runborg, *Multi-scale methods for wave propagation in heterogeneous media*, Commun. Math. Sci., 9 (2011), pp. 33–56.
[33] B. Engquist and P. E. Souganidis, *Asymptotic and numerical homogenization*, Acta Numer., 17 (2008), pp. 147–190.
[34] G. E. Fasshauer, *Meshfree methods*, in Handbook of Theoretical and Computational Nanotechnology, American Scientific Publishers, Valencia, CA, 2005.
[35] R. Ghanem, *Ingredients for a general purpose stochastic finite elements implementation*, Comput. Methods Appl. Mech. Engrg., 168 (1999), pp. 19–34.
[36] R. Ghanem and S. Dham, *Stochastic finite element analysis for multiphase flow in heterogeneous porous media*, Transp. Porous Media, 32 (1998), pp. 239–262.
[37] A. Gloria, *An analytical framework for the numerical homogenization of monotone elliptic operators and quasiconvex energies*, Multiscale Model. Simul., 5 (2006), pp. 996–1043.
[38] M. Golomb and H. F. Weinberger, *Optimal approximation and error bounds*, in On Numerical Approximation. Proceedings of a Symposium, Madison, 1958, R. E. Langer, ed., Publication No. 1 of the Mathematics Research Center, U.S. Army, the University of Wisconsin, The University of Wisconsin Press, Madison, WI, 1959, pp. 117–190.
[39] L. Grasedyck, I. Greff, and S. Sauter, *The AL basis for the solution of elliptic problems in heterogeneous media*, Multiscale Model. Simul., 10 (2012), pp. 245–258.
[40] R. L. Harder and R. N. Desmarais, *Interpolation using surface splines*, J. Aircraft, 9 (1972), pp. 189–191.
[41] C. Heitzinger and C. Ringhofer, *Multiscale modeling of fluctuations in stochastic elliptic PDE models of nanosensors*, Commun. Math. Sci., 12 (2014), pp. 401–421.
[42] T. Y. Hou and X. H. Wu, *A multiscale finite element method for elliptic problems in composite materials and porous media*, J. Comput. Phys., 134 (1997), pp. 169–189.
[43] T. Y. Hou, X.-H. Wu, and Z. Cai, *Convergence of a multiscale finite element method for elliptic problems with rapidly oscillating coefficients*, Math. Comp., 68 (1999), pp. 913–943.
[44] S. M. Kozlov, *The averaging of random operators*, Mat. Sb. (N.S.), 109(151) (1979), pp. 188–202, 327.
[45] A. Målqvist and D. Peterseim, *Localization of elliptic multiscale problems*, Math. Comp., 83 (2014), pp. 2583–2603.
[46] F. Murat and L. Tartar, *H-convergence*, in Séminaire d'Analyse Fonctionnelle et Numérique, Université d'Alger, Algiers, Algeria, 1978.
[47] D. E. Myers, *Kriging, co-Kriging, radial basis functions and the role of positive definiteness*, Comput. Math. Appl., 24 (1992), pp. 139–148.
[48] J. Nolen, G. Papanicolaou, and O. Pironneau, *A framework for adaptive multiscale methods for elliptic problems*, Multiscale Model. Simul., 7 (2008), pp. 171–196.

[49] A. O'Hagan, *Bayes-Hermite quadrature*, J. Statist. Plann. Inference, 29 (1991), pp. 245–260.

[50] A. O'Hagan, *Some Bayesian numerical analysis*, in Bayesian Statistics, 4 (Peñíscola, 1991), Oxford University Press, New York, 1992, pp. 345–363.

[51] H. Owhadi, *Multigrid with Rough Coefficients and Multiresolution Operator Decomposition from Hierarchical Information Games*, preprint, arXiv:1503.03467v2 [math.NA], 2015.

[52] H. Owhadi and C. Scovel, *Qualitative Robustness in Bayesian Inference*, preprint, arXiv:1411.3984v2 [math.ST], 2014.

[53] H. Owhadi and C. Scovel, *Brittleness of Bayesian inference and new Selberg formulas*, Commun. Math. Sci., to appear; available online from http://arxiv.org/abs/1304.7046.

[54] H. Owhadi, C. Scovel, and T. J. Sullivan, *Brittleness of Bayesian inference under finite information in a continuous world*, Electron. J. Stat., 9 (2015), pp. 1–79.

[55] H. Owhadi, C. Scovel, and T. J. Sullivan, *On the brittleness of Bayesian inference*, SIAM Rev., to appear.

[56] H. Owhadi and L. Zhang, *Homogenization of parabolic equations with a continuum of space and time scales*, SIAM J. Numer. Anal., 46 (2007), pp. 1–36.

[57] H. Owhadi and L. Zhang, *Metric-based upscaling*, Comm. Pure Appl. Math., 60 (2007), pp. 675–723.

[58] H. Owhadi and L. Zhang, *Homogenization of the acoustic wave equation with a continuum of scales*, Comput. Methods Appl. Mech. Engrg., 198 (2008), pp. 397–406.

[59] H. Owhadi and L. Zhang, *Localized bases for finite-dimensional homogenization approximations with nonseparated scales and high contrast*, Multiscale Model. Simul., 9 (2011), pp. 1373–1398.

[60] H. Owhadi, L. Zhang, and L. Berlyand, *Polyharmonic homogenization, rough polyharmonic splines and sparse super-localization*, ESAIM Math. Model. Numer. Anal., 48 (2014), pp. 517–552.

[61] G. C. Papanicolaou and S. R. S. Varadhan, *Boundary value problems with rapidly oscillating random coefficients*, in Random Fields, Vol. I, II (Esztergom, 1979), Colloq. Math. Soc. János Bolyai 27, North-Holland, Amsterdam, New York, 1981, pp. 835–873.

[62] H. Poincaré, *Calcul des probabilités*, Georges Carrés, Paris, 1896.

[63] J. E. H. Shaw, *A quasirandom approach to integration in Bayesian statistics*, Ann. Statist., 16 (1988), pp. 895–914.

[64] S. Spagnolo, *Sulla convergenza di soluzioni di equazioni paraboliche ed ellittiche*, Ann. Scuola Norm. Sup. Pisa (3), 22 (1968), pp. 571–597; errata, ibid. (3), 22 (1968), p. 673.

[65] G. Stampacchia, *Èquations elliptiques du second ordre à coefficients discontinus*, Séminaire Jean Leray, No. 3 (1963–1964), pp.1–77.

[66] R. A. Todor and C. Schwab, *Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients*, IMA J. Numer. Anal., 27 (2007), pp. 232–261.

[67] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ, 1944.

[68] A. Wald, *Statistical decision functions which minimize the maximum risk*, Ann. of Math. (2), 46 (1945), pp. 265–280.

[69] H. Wendland, *Scattered Data Approximation*, Cambridge Monogr. Appl. Comput. Math. 17, Cambridge University Press, Cambridge, UK, 2005.

[70] C. D. White and R. N. Horne, *Computing absolute transmissibility in the presence of fine-scale heterogeneity*, in Proceedings of the SPE Symposium on Reservoir Simulation, 1987, 16011.

[71] Z. M. Wu and R. Schaback, *Local error estimates for radial basis function interpolation of scattered data*, IMA J. Numer. Anal., 13 (1993), pp. 13–27.

[72] D. Xiu, *Fast numerical methods for stochastic computations: A review*, Commun. Comput. Phys., 5 (2009), pp. 242–272.