

Multigrid with Rough Coefficients and Multiresolution Operator Decomposition from Hierarchical Information Games*

Houman Owhadi[†]

Abstract. We introduce a near-linear complexity (geometric and meshless/algebraic) multigrid/multiresolution method for PDEs with rough (L^∞) coefficients with rigorous a priori accuracy and performance estimates. The method is discovered through a decision/game theory formulation of the problems of (1) identifying restriction and interpolation operators, (2) recovering a signal from incomplete measurements based on norm constraints on its image under a linear operator, and (3) gambling on the value of the solution of the PDE based on a hierarchy of nested measurements of its solution or source term. The resulting elementary gambles form a hierarchy of (deterministic) basis functions of $H_0^1(\Omega)$ (gamblets) that (1) are orthogonal across subscales/subbands with respect to the scalar product induced by the energy norm of the PDE, (2) enable sparse compression of the solution space in $H_0^1(\Omega)$, and (3) induce an orthogonal multiresolution operator decomposition. The operating diagram of the multigrid method is that of an inverted pyramid in which gamblets are computed locally (by virtue of their exponential decay) and hierarchically (from fine to coarse scales) and the PDE is decomposed into a hierarchy of independent linear systems with uniformly bounded condition numbers. The resulting algorithm is parallelizable both in space (via localization) and in bandwidth/subscale (subscales can be computed independently from each other). Although the method is deterministic, it has a natural Bayesian interpretation under the measure of probability emerging (as a mixed strategy) from the information game formulation, and multiresolution approximations form a martingale with respect to the filtration induced by the hierarchy of nested measurements.

Key words. multigrid, multiresolution, game theory, wavelets, information based complexity, decision theory

AMS subject classifications. 68T99, 65N55, 65F99, 65N75, 62C99, 42C40, 60G42, 68Q25

DOI. 10.1137/15M1013894

Contents

I Introduction	100
1.1 Scientific Discovery as a Decision Theory Problem	100
1.2 Outline of the Paper	102

*Received by the editors March 27, 2015; accepted for publication (in revised form) March 4, 2016; published electronically February 7, 2017.

<http://www.siam.org/journals/sirev/59-1/M101389.html>

Funding: This work was supported by the Air Force Office of Scientific Research and the DARPA EQUiPS Program under awards FA9550-12-1-0389 (Scientific Computation of Optimal Statistical Estimators) and FA9550-16-1-0054 (Computational Information Games) 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).

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

1.3	On the Correspondence between Statistical Inference and Numerical Approximation	104
2	Linear Algebra with Incomplete Information	106
2.1	The Recovery Problem	106
2.2	Player I's Mixed Strategy	106
2.3	Variational/Optimal Recovery Properties and Approximation Error . .	107
2.4	Energy Norm Estimates and Selection of the Prior	108
2.5	Impact and Selection of the Measurement Matrix Φ	108
3	Numerical Homogenization and Design of the Interpolation Operator in the Continuous Case	110
3.1	Information Game and Gamblets	110
3.2	Optimal Recovery Properties	111
3.3	Optimal Accuracy of the Recovery	112
3.4	Transfer Property and Selection of the Measurement Functions	112
3.5	Exponential Decay of Gamblets	114
3.6	Localization of the Basis Elements	116
4	Multiresolution Operator Decomposition	118
4.1	Hierarchy of Nested Measurement Functions	118
4.2	Hierarchy of Nested Gamblets and Multiresolution Approximations . .	118
4.3	Nested Games and Martingale/Multiresolution Decomposition	119
4.4	Interpolation and Restriction Matrices/Operators	121
4.5	Nested Computation of the Interpolation and Stiffness Matrices	121
4.6	Multiresolution Gamblets	122
4.7	Multiresolution Operator Inversion	123
4.8	Uniformly Bounded Condition Numbers across Subscales/Subbands .	125
4.9	Well-Conditioned Relaxation across Subscales	128
4.10	Hierarchical Localization and Error Propagation across Scales	128
5	The Algorithm, Its Implementation, and Its Complexity	132
5.1	The Initialization of the Algorithm	132
5.2	Exact Gamblet Transform and Multiresolution Operator Inversion . .	133
5.3	Fast Gamblet Transform/Solve	139
5.4	Complexity vs. Accuracy of Algorithm 2 and Choice of the Localization Radii ρ_k	139
	References	144

I. Introduction.

I.1. Scientific Discovery as a Decision Theory Problem. The process of scientific discovery is often based on intuition, trial and error, and plain guesswork. This article is motivated by the question of the existence of a rational decision framework that could be used to facilitate/guide this process or turn it, to some degree, into an algorithm. In exploring this question, we will consider the problem of finding a method for solving (up to a prespecified level of accuracy) PDEs with rough (L^∞) coefficients as fast as possible with the following prototypical PDE (and its possible

discretization over a fine mesh) as an example:

$$(1.1) \quad \begin{cases} -\operatorname{div}(a(x)\nabla u(x)) = g(x), & x \in \Omega; g \in L^2(\Omega) \text{ or } g \in H^{-1}(\Omega), \\ u = 0 & \text{on } \partial\Omega, \end{cases}$$

where Ω is a bounded subset of \mathbb{R}^d (of arbitrary dimension $d \in \mathbb{N}^*$) with piecewise Lipschitz boundary and a is a symmetric, uniformly elliptic $d \times d$ matrix with entries in $L^\infty(\Omega)$ such that, for all $x \in \Omega$ and $l \in \mathbb{R}^d$,

$$(1.2) \quad \lambda_{\min}(a)|l|^2 \leq l^T a(x)l \leq \lambda_{\max}(a)|l|^2.$$

Although multigrid methods [41, 16, 50, 51, 101] are now well known as being the fastest for solving elliptic boundary problems and have been successfully generalized to other types of PDEs and computational problems [123], their convergence rate can be severely affected by the lack of regularity of the coefficients [37, 114]. Furthermore, although significant progress has been achieved in the development of multigrid methods that are, to some degree, robust with respect to mesh size and lack of smoothness (we refer, in particular, to algebraic multigrid [94], multilevel finite element splitting [124], hierarchical basis multigrid [9, 24], multilevel preconditioning [106], stabilized hierarchical basis methods [107, 109, 110], energy minimization [65, 114, 122, 121, 108], and homogenization based methods [37, 34]), the design of multigrid methods that are provably robust with respect to rough (L^∞) coefficients has remained an open problem of practical importance [17].

Alternative hierarchical strategies for the resolution of (1.1) are (1) wavelet based methods [18, 15, 3, 29, 38] (2) the fast multipole method [49], and (3) hierarchical matrices [52, 11]. Although methods based on (classical) wavelets achieve a multiresolution compression of the solution space of (1.1) in L^2 , and although approximate wavelets and approximate L^2 projections can stabilize hierarchical basis methods [109, 110], their applications to (1.1) are limited by the facts that (a) the underlying wavelets can perform arbitrarily badly [7] in their $H_0^1(\Omega)$ approximation of the solution space, and (b) the operator (1.1) does not preserve the orthogonality between subscales/subbands with classical wavelets. The fast multipole method and hierarchical matrices exploit the property that submatrices of the inverse discrete operator are of low rank away from the diagonal. This low rank property can be rigorously proven for (1.1) (based on the approximation of its Green's function by sums of products of harmonic functions [10]) and leads to provable convergence (with rough coefficients), up to the prespecified level of accuracy ϵ in the L^2 -norm, in $\mathcal{O}(N \ln^6 N \ln^{2d+2} \frac{1}{\epsilon})$ operations (see [10] and [11, Thms. 2.33 and 4.28]). Can the problem of finding a fast solver for (1.1) be, to some degree, reformulated as an uncertainty quantification/decision theory problem that could, to some degree, be solved as such in an automated fashion? Can discovery be computed? Although these questions may seem unorthodox their answer appears to be positive: this paper shows that this reformulation is possible and leads to a multigrid/multiresolution method/algorithm solving (1.1), up to the prespecified level of accuracy ϵ in the H^1 -norm (i.e., finding u^{app} such that $\|u - u^{\text{app}}\|_{H_0^1(\Omega)} \leq \epsilon \|g\|_{H^{-1}(\Omega)}$ for an arbitrary g decomposed over N degrees of freedom), in $\mathcal{O}(N \ln^{3d} (\max(\frac{1}{\epsilon}, N^{1/d}))$) operations (for $\epsilon \sim N^{-1/d}$, the hierarchical matrix method achieves ϵ -accuracy in the L^2 norm in $\mathcal{O}(N \ln^{2d+8} N)$ operations and the proposed multiresolution method achieves ϵ -accuracy in the H^1 norm in $\mathcal{O}(N \ln^{3d} N)$ operations). For subsequent solves (i.e., if (1.1) needs to be solved for more than one g), the proposed multiresolution method achieves accuracy

$\epsilon \approx N^{-\frac{1}{2}}$ in the H^1 -norm in $\mathcal{O}(N \ln^{d+1} N)$ operations (we refer to subsection 5.4 and, in particular, to Table 1 for a detailed complexity analysis of the proposed method, which can also achieve sublinear complexity if one only requires L^2 -approximations).

The core mechanism supporting the complexity of the method presented here is the fast decomposition of $H_0^1(\Omega)$ into a direct sum of linear subspaces that are orthogonal (or near-orthogonal) with respect to the energy scalar product and over which (1.1) has uniformly bounded condition numbers. It is, to some degree, surprising that this decomposition can be achieved in near-linear complexity and not in the complexity of an eigenspace decomposition. Naturally [84], this decomposition can be applied to the fast simulation of the wave and parabolic equations associated with (1.1) or with its fast diagonalization.

The essential step behind the automation of the discovery/design of scalable numerical solvers is the observation that fast computation requires repeated computation with partial information (and limited resources) over hierarchies of levels of complexity and the reformulation of this process as that of playing underlying hierarchies of adversarial information games [111, 112].

Although the problem of finding a fast solver for (1.1) may appear to be disconnected from that of finding statistical estimators or making decisions from data sampled from an underlying unknown probability distribution, the proposed game theoretic reformulation is, to some degree, analogous to the one developed in Wald's decision theory [113], evidently influenced by Von Neumann's game theory [111, 112] (the generalization of worst-case uncertainty quantification analysis [83] to sample data/model uncertainty requires an analogous game theoretic formulation [80]; see also [79] for how the underlying calculus could be used to guide the discovery of new Selberg identities). We also refer to section 1.3 for a review of the correspondence between statistical inference and numerical approximation.

1.2. Outline of the Paper. The essential difficulty in generalizing the multigrid concept to PDEs with rough coefficients lies in the fact that the interpolation (down-scaling) and restriction (up-scaling) operators are, a priori, unknown. Indeed, in this situation, piecewise linear finite elements can perform arbitrarily badly [7] and the design of the interpolation operator requires the identification of accurate basis elements adapted to the microstructure $a(x)$.

This identification problem has also been the essential difficulty in numerical homogenization [117, 6, 4, 18, 59, 33, 85, 17]. Although inspired by classical homogenization ideas and concepts (such as oscillating test functions [68, 36, 35], cell problems/correctors and effective coefficients [13, 90, 1, 72, 39, 46], harmonic coordinates [61, 6, 4, 76, 12, 2, 85], and compactness by compensation [100, 45, 67, 14]), an essential goal of numerical homogenization has been the numerical approximation of the solution space of (1.1) with arbitrary rough coefficients [85], i.e., in particular, without the assumptions found in classical homogenization, such as scale separation, ergodicity at fine scales, and ϵ -sequences of operators (otherwise, the resulting method could lack robustness to rough coefficients, even under the assumption that coefficients are stationary [8]). Furthermore, to envisage applications to multigrid methods, the computation of these basis functions must also be provably localized [5, 86, 64, 48, 87, 58] and compatible with nesting strategies [87]. In [77] it was shown that this process of identification (of accurate basis elements for numerical homogenization) could, in principle, be guided through its reformulation as a Bayesian inference problem in which the source term g in (1.1) is replaced by noise ξ and one tries to estimate the value of the solution at a given point based on a finite number

of observations. In particular it was found that rough polyharmonic splines [87] and polyharmonic splines [54, 30, 31, 32] can be rediscovered as solutions of Gaussian filtering problems. This paper is inspired by the suggestion that this link between numerical homogenization and Bayesian inference (and the link between numerical quadrature and Bayesian inference [92, 27, 97, 74, 75]) are not coincidences, but particular instances of mixed strategies for underlying information games, and that optimal or near-optimal methods could be obtained by identifying such games and their optimal strategies.

The process of identification of these games starts with the (information based complexity [119]) notion that computation can only be done with partial information. For instance, since the operator (1.1) is infinite-dimensional, one cannot directly compute with $u \in H_0^1(\Omega)$ but only with finite-dimensional *features* of u . An example of such finite-dimensional features is the m -dimensional vector $u_m := (\int_{\Omega} u\phi_1, \dots, \int_{\Omega} u\phi_m)$ obtained by integrating the solution u of (1.1) against m test/measurement functions $\phi_i \in L^2(\Omega)$. However, to achieve an accurate approximation of u through computation with u_m one must fill the information gap between u_m and u (i.e., construct an interpolation operator giving u as a function of u_m). We will, therefore, reformulate the identification of this interpolation operator as a non-cooperative (min max) game where Player I chooses the source term g (1.1) in an admissible set/class (e.g., the unit ball of $L^2(\Omega)$) and Player II is shown u_m and must approximate u from these incomplete measurements. Using the energy norm

$$(1.3) \quad \|u\|_a^2 := \int_{\Omega} \nabla u^T(x) a(x) \nabla u(x) dx$$

to quantify the accuracy of the recovery and calling u^* Player I's bet (on the value of u), the objective of Player I is to maximize the approximation error $\|u - u^*\|_a$, while the objective of Player II is to minimize it. A remarkable result from game theory (as developed by Von Neumann [111], Von Neumann and Morgenstern [112], and Nash [70]) is that optimal strategies for deterministic zero sum finite games are mixed (i.e., randomized) strategies. Although the information game described above is zero sum, it is not finite. Nevertheless, as in Wald's decision theory [113], under sufficient regularity conditions it can be made compact and therefore approximable by a finite game. Therefore, although the information game described above is purely deterministic (and has no a priori connection to statistical estimation), under compactness (and continuity of the loss function), the best strategy for Player I is to play at random by placing a probability distribution π_I on the set of candidates for g (and select g as a sample from π_I), and the optimal strategy for Player II is to place a probability distribution π_{II} on the set of candidates for g and approximate the solution of (1.1) by the expectation of u (under π_{II} used as a prior distribution) conditioned on the measurements $\int_{\Omega} u\phi_i$.

Although the estimator employed by Player II may be called Bayesian, the game described here is not (i.e., the choice of Player I might be distinct from that of Player II) and Player II must solve a min max optimization problem over π_I and π_{II} 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 [81, 79, 82]). Although solving the min max problem over π_I and π_{II} may be one way of determining the strategy of Player II, it will not be the method employed here. We will instead analyze the error of Player II's approximation as a function of Player II's prior and the source term g picked by Player I. Furthermore, to

preserve the linearity of the calculations we will restrict Player II's decision space (the set of possible priors π_{II}) to Gaussian priors on the source term g . Since the resulting analysis is independent of the structure of (1.1) and depends solely on its linearity, we will first perform this investigation, in section 2, in the algebraic framework of linear systems of equations, identify Player II's optimal mixed strategy, and show that it is characterized by deterministic optimal recovery and accuracy properties. The mixed strategy identified in section 2 will then be applied in section 3 to the numerical homogenization of (1.1) and the discovery of interpolation interpolators. In particular, it will be shown that the resulting elementary gambles form a set of deterministic basis functions (gamblets) characterized by (1) optimal recovery and accuracy properties, (2) exponential decay (enabling their localized computation), and (3) robustness to high contrast.

To compute rapidly, the game presented above must not be limited to filling the information gap between $u_m \in \mathbb{R}^m$ and $u \in H_0^1(\Omega)$. This game must be played (and repeated) over hierarchies of levels of complexity (e.g., one must fill information gaps between \mathbb{R}^4 and \mathbb{R}^{16} , then \mathbb{R}^{16} and \mathbb{R}^{64} , etc.). We will, therefore, in section 4, consider the (hierarchical) game where Player I chooses the right-hand side of (1.1) and Player II must (iteratively) gamble on the value of its solution based on a hierarchy of nested measurements of u (from coarse to fine measurements). Under Player II's mixed strategy (identified in section 2 and used in section 3), the resulting sequence of multiresolution approximations forms a martingale. Conditioning and the independence of martingale increments lead to the hierarchy of nested interpolation operators and to the multiresolution orthogonal decomposition of (1.1) into independent linear systems of uniformly bounded condition numbers. The resulting elementary gambles (gamblets) (1) form a hierarchy of nested basis functions leading to the orthogonal decomposition (in the scalar product of the energy norm) of $H_0^1(\Omega)$, (2) enable the sparse compression of the solution space of (1.1), (3) can be computed and stored in near-linear complexity by solving a nesting of linear systems with uniformly bounded condition numbers, and (4) enable the computation of the solution of (1.1) (or its hyperbolic or parabolic analogues) in near-linear complexity. The implementation and complexity of the algorithm are discussed in section 5 with numerical illustrations.

1.3. On the Correspondence between Statistical Inference and Numerical

Approximation. As exposed by Diaconis [27], the investigation of the correspondence between statistical inference and numerical approximation can be traced back to Poincaré's course in probability theory [92]. It is useful to recall Diaconis's compelling example [27] as an illustration of this connection. Let $f : [0, 1] \rightarrow \mathbb{R}$ 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, \dots, x_n (to obtain the data $(f(x_1), \dots, f(x_n))$), (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 $\mathbb{E}[f(x)|f(x_1), \dots, f(x_n)]$ is the piecewise linear interpolation of f between the points x_1, \dots, 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 $\mathbb{E}[f(x)|f(x_1), \dots, f(x_n)]$ is the cubic spline interpolant and integrating k times yields splines of order $2k + 1$.

Subsequent to Poincaré's early discovery [92], Sul'din [102] and (in particular) Larkin [62] initiated the systematic investigation of the correspondence between con-

conditioning Gaussian measures/processes and numerical approximation. As noted by Larkin [62], despite Sard's introduction of probabilistic concepts in the theory of linear approximation [95], and Kimeldorf and Wahba's exposition [60] of the correspondence between Bayesian estimation and spline smoothing/interpolation, the application of probabilistic concepts and techniques to numerical integration/approximation "attracted little attention among numerical analysts" (perhaps due to the counterintuitive nature of the process of randomizing a known function). However, a natural framework for understanding this process of randomization can be found in the pioneering works of Woźniakowski [118], Packel [88], and Traub, Wasilkowski, and Woźniakowski [104] on information based complexity [71, 119], the branch of computational complexity that studies the complexity of approximating continuous mathematical operations with discrete and finite operations up to a specified level of accuracy. Indeed, the concept that numerical implementation requires computation with partial information and limited resources emerges naturally from information based complexity, where it is also augmented by concepts of contaminated and priced information associated with, for example, truncation errors and the cost of numerical operations. In this framework, the performance of an algorithm operating on incomplete information can be analyzed in the usual worst case setting or the average case (randomized) setting [93, 73] with respect to the missing information. Although the measure of probability (on the solution space) employed in the average case setting may be arbitrary, as observed by Packel [88], if that measure is chosen carefully (as the solution of a game theoretic problem), then the average case setting can be interpreted as lifting a (worst case) min max problem (where saddle points of pure strategies do not, in general, exist) to a min max problem over mixed (randomized) strategies (where saddle points do exist [111, 112]). As exposed by Diaconis [27] (see also Shaw [97]), the randomized setting also establishes a correspondence between numerical analysis and Bayesian inference, providing a natural framework for the statistical description of numerical errors (in which confidence intervals can be derived from posterior distributions). Furthermore [89, 27], classical min max numerical quadrature rules can be formulated as solutions of Bayesian inference problems with carefully chosen priors [27] and, as shown by Hagan [74, 75], this correspondence can be exploited to discover new and useful numerical quadrature rules. As envisioned by Skilling [99], by placing a (carefully chosen) probability distribution on the solution space of an ODE and conditioning on quadrature points, one obtains a posterior distribution on the solution whose mean may coincide with classical numerical integrators such as Runge–Kutta methods [96]. As shown in [23] the statistical approach is particularly well suited for chaotic dynamical systems for which deterministic worst case error bounds may provide little information. While in [99, 96, 23] the probability distribution is directly placed on the solutions space, for PDEs [77] argues that the prior distribution must be placed on source terms (or on the image space of an integro-differential operator) and propagated/filtered through the inverse operator to reflect the structure of the solution space. In particular, [77] shows that this process of filtering noise with the inverse operator, when combined with conditioning, produces accurate finite element basis functions for the solution space whose deterministic worst case errors can be bounded by standard deviation errors using the reproducing kernel structure of the covariance function of the filtered Gaussian field. As already witnessed in [23, 96, 77, 56, 55, 19, 25], it is natural to expect that the possibilities offered by combining numerical uncertainties/errors with model uncertainties/errors in a unified framework will stimulate a resurgence of the statistical inference approach to numerical analysis.

2. Linear Algebra with Incomplete Information.

2.1. The Recovery Problem. The problem of identifying interpolation operators for (1.1) is equivalent (after discretization or in the algebraic setting) to that of recovering or approximating the solution of a linear system of equations from an incomplete set of measurements (coarse variables) given known norm constraints on the image of the solution.

Let $n \geq 2$ and A be a known real invertible $n \times n$ matrix. Let b be an unknown element of \mathbb{R}^n . Our purpose is to approximate the solution x of

$$(2.1) \quad Ax = b$$

based on the information that (1) x solves

$$(2.2) \quad \Phi x = y,$$

where Φ (the measurement matrix) is a known, rank m , $m \times n$ real matrix such that $m < n$ and y (the measurement vector) is a known vector of \mathbb{R}^m , and (2) the norm $b^T T^{-1} b$ of b is known or bounded by a known constant (e.g., $b^T T^{-1} b \leq 1$), where T^{-1} is a known positive definite $n \times n$ matrix (with T^{-1} being the identity matrix as a prototypical example). Observe that since $m < n$, the measurements (2.2) are, a priori, not sufficient to recover the exact value x .

As described in section 1, by formulating this recovery problem as a (noncooperative) information game (where Player I chooses b and Player II chooses an approximation x^* of x based on the observation Φx), one (Player II) is naturally lead to search for a mixed strategy in the Bayesian class by placing a prior distribution on b . The purpose of this section is to analyze the resulting approximation error and select the prior distribution accordingly. To preserve the linearity (i.e., simplicity and computational efficiency) of calculations we will restrict Player II's decision space to Gaussian priors.

2.2. Player I's Mixed Strategy. We will, therefore, in the first step of the analysis, replace b in (2.1) by ξ , a centered Gaussian vector of \mathbb{R}^n with covariance matrix Q (which may be distinct from T) and consider the stochastic linear system

$$(2.3) \quad AX = \xi.$$

The Bayesian answer (a mixed strategy for Player II) to the recovery problem of section 2 is to approximate x by the conditional expectation $\mathbb{E}[X|\Phi X = y]$.

THEOREM 2.1. *The solution X of (2.3) is a centered Gaussian vector of \mathbb{R}^n with covariance matrix*

$$(2.4) \quad K = A^{-1}Q(A^{-1})^T.$$

Furthermore, X conditioned on the value $\Phi X = y$ is a Gaussian vector of \mathbb{R}^n with mean $\mathbb{E}[X|\Phi X = y] = \Psi y$ and of covariance matrix K^Φ , where Ψ is the $n \times m$ matrix

$$(2.5) \quad \Psi := K\Phi^T(\Phi K\Phi^T)^{-1}$$

and K^Φ is the rank $n-m$ positive $n \times n$ symmetric matrix defined by $K^\Phi := K - \Psi\Phi K$.

Proof. (2.4) simply follows from $X = A^{-1}\xi$. Since X is a Gaussian vector, $\mathbb{E}[X|\Phi X = y] = \Psi y$, where Ψ is an $n \times m$ matrix minimizing the mean squared error $\mathbb{E}[|X - M\Phi X|^2]$ over all $n \times m$ matrices M . We have $\mathbb{E}[|X - M\Phi X|^2] = \text{Trace}[K] + \text{Trace}[M\Phi K\Phi^T M^T] - 2\text{Trace}[\Phi K M]$, whose minimum is achieved for $M = \Psi$ as defined by (2.5). The covariance matrix of X given $\Phi X = y$ is then obtained by observing that for $v \in \mathbb{R}^n$, $v^T K^\Phi v = \mathbb{E}[|v^T X - v^T \Psi \Phi X|^2] = v^T K v - v^T \Psi \Phi K v$. \square

2.3. Variational/Optimal Recovery Properties and Approximation Error. For an $n \times n$ symmetric positive definite matrix M , let $\langle \cdot, \cdot \rangle_M$ be the (scalar) product on \mathbb{R}^n defined as follows: for $u, v \in \mathbb{R}^n$, $\langle u, v \rangle_M := u^T M v$ and write $\|v\|_M := \langle v, v \rangle_M^{\frac{1}{2}}$ as the corresponding norm. When M is the identity matrix, we write $\langle u, v \rangle$ and $\|v\|$ as the corresponding scalar product and norm. For a linear subspace V of \mathbb{R}^n we write $P_{V,M}$ for the orthogonal projections onto V with respect to the scalar product $\langle \cdot, \cdot \rangle_M$. For a (possibly rectangular) matrix B we write $\text{Im}(B)$ as the image (range) of B and $\text{Ker}(B)$ as the null space of B . For an integer n , let I_n be the $n \times n$ identity matrix.

THEOREM 2.2. For $w \in \mathbb{R}^m$, Ψw is the unique minimizer of the quadratic problem

$$(2.6) \quad \begin{cases} \text{Minimize} & \langle v, v \rangle_{K^{-1}} \\ \text{subject to} & \Phi v = w \text{ and } v \in \mathbb{R}^n. \end{cases}$$

In particular, $v = \Psi y$, the Bayesian approximation of the solution of (2.1), is the unique minimizer of $\|Av\|_{Q^{-1}}$ under the measurement constraints $\Phi v = y$. Furthermore, it also holds true that (1) $\Phi\Psi = I_m$, (2) $\text{Im}(\Psi)$ is the orthogonal complement of $\text{Ker}(\Phi)$ with respect to the product $\langle \cdot, \cdot \rangle_{K^{-1}}$, and (3) $\Psi\Phi = P_{\text{Im}(K\Phi^T), K^{-1}}$ and $I_n - \Psi\Phi = P_{\text{Ker}(\Phi), K^{-1}}$.

Proof. First observe that (2.5) implies that $\Phi\Psi = I_m$, where I_m is the identity $m \times m$ matrix. Therefore, $\Phi(\Psi w) = w$. Note that (2.5) implies that for all $z \in \mathbb{R}^m$, $\langle \Psi z, v \rangle_{K^{-1}} = z^T (\Phi K \Phi^T)^{-1} \Phi v$. Therefore, if $v \in \text{Ker}(\Phi)$, then $\langle \Psi z, v \rangle_{K^{-1}} = 0$ for all $z \in \mathbb{R}^m$. Conversely, if $\langle \Psi z, v \rangle_{K^{-1}} = 0$ for all $z \in \mathbb{R}^m$, then v must belong to $\text{Ker}(\Phi)$. Since the dimension of $\text{Im}(\Psi)$ is m and that of $\text{Ker}(\Phi)$ is $n - m$, we conclude that $\text{Im}(\Psi)$ is the orthogonal complement $\text{Ker}(\Phi)$ with respect to the product $\langle \cdot, \cdot \rangle_{K^{-1}}$ and, in particular,

$$(2.7) \quad \langle \Psi w, v \rangle_{K^{-1}} = 0 \quad \forall w \in \mathbb{R}^m \text{ and } \forall v \in \mathbb{R}^n \text{ such that } \Phi v = 0.$$

Let $w \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$ such that $\Phi v = w$. Since $\Psi w - v \in \text{Ker}(\Phi)$, it follows from (2.7) that $\langle v, v \rangle_{K^{-1}} = \langle \Psi w, \Psi w \rangle_{K^{-1}} + \langle v - \Psi w, v - \Psi w \rangle_{K^{-1}}$. Therefore, Ψw is the unique minimizer of $\langle v, v \rangle_{K^{-1}}$ over all $v \in \mathbb{R}^n$ such that $\Phi v = w$. Now consider $f \in \mathbb{R}^n$; since $\text{Im}(\Psi) = \text{Im}(K\Phi^T)$ and $\text{Im}(\Psi)$ is the orthogonal complement of $\text{Ker}(\Phi)$ with respect to the product $\langle \cdot, \cdot \rangle_{K^{-1}}$, there exist a unique $z \in \mathbb{R}^m$ and a unique $g \in \text{Ker}(\Phi)$ such that $f = K\Phi^T z + g$. Since $\Psi\Phi = K\Phi^T (\Phi K \Phi^T)^{-1} \Phi$, it follows that $\Psi\Phi f = K\Phi^T z$ and $(I_n - \Psi\Phi)f = g$. We conclude by observing that $g = P_{\text{Ker}(\Phi), K^{-1}} f$. \square

THEOREM 2.3. For $v \in \mathbb{R}^n$, $w^* = \Phi v$ is the unique minimizer of $\|v - \Psi w\|_{K^{-1}}$ over all $w \in \mathbb{R}^m$. In particular, $\|v - \Psi\Phi v\|_{K^{-1}} = \min_{z \in \mathbb{R}^m} \|v - K\Phi^T z\|_{K^{-1}}$ and if x is the solution of the original equation (2.1), then $\|x - \Psi y\|_{K^{-1}} = \min_{w \in \mathbb{R}^m} \|x - \Psi w\|_{K^{-1}} = \min_{z \in \mathbb{R}^m} \|x - K\Phi^T z\|_{K^{-1}}$.

Proof. The proof follows by observing that $v - \Psi\Phi v$ belongs to the null space of Φ which, from Theorem 2.2, is the orthogonal complement of the image of Ψ with respect to the scalar product defining the norm $\|\cdot\|_{K^{-1}}$. Observe also that the image of Ψ is equal to that of $K\Phi^T$. \square

Remark 1. Observe that, from Theorem 2.2, $v - \Psi\Phi v$ spans the null space of Φ , and $\|v\|_{K^{-1}}^2 = \|v - \Psi\Phi v\|_{K^{-1}}^2 + \|\Psi\Phi v\|_{K^{-1}}^2$. Therefore, if D is a symmetric positive definite $n \times n$ matrix, then $\sup_{v \in \mathbb{R}^n} \|v - \Psi\Phi v\|_D / \|v\|_{K^{-1}} = \sup_{v \in \mathbb{R}^n, \Phi v = 0} \|v\|_D / \|v\|_{K^{-1}}$. In particular, if x is the solution of (2.1) and y the vector in (2.2), then $\|x -$

$\Psi y \|_D / \|b\|_{Q^{-1}} \leq \sup_{v \in \mathbb{R}^n, \Phi v=0} \|v\|_D / \|v\|_{K^{-1}}$ and the right-hand side is the smallest constant for which the inequality holds (for all b).

Remark 2. A simple calculation (based on the reproducing Kernel property $\langle v, K_{\cdot, i} \rangle_{K^{-1}} = v_i$) shows that if x is the solution of (2.1) and y the vector in (2.2), then $|(x - \Psi y)_i| \leq (K_{i, i}^\Phi)^{\frac{1}{2}} \|b\|_{Q^{-1}}$, i.e., the variance of the i th entry of the solution of the stochastic system (2.3) conditioned on $\Phi X = y$ controls the accuracy of the approximation of the i th entry of the solution of the deterministic system (2.1). In that sense, the role of K^Φ is analogous to that of the power function in radial basis function interpolation [116, 40] and that of the Kriging function [120] in geostatistics [69].

2.4. Energy Norm Estimates and Selection of the Prior. We will from now on assume that A is symmetric positive definite. Observe that in this situation the energy norm $\|\cdot\|_A$ is of practical significance for quantifying the approximation error and Theorem 2.3 leads to the estimate $\|x - \Psi y\|_{K^{-1}} = \min_{z \in \mathbb{R}^m} \|Q^{-\frac{1}{2}} b - Q^{-\frac{1}{2}} A^{\frac{1}{2}} K^{\frac{1}{2}} \Phi^T z\|$, which simplifies to the energy norm estimate expressed by Corollary 2.4 under the choice $Q = A$ (note that $K^{-1} = A$ under that choice).

COROLLARY 2.4. *If A is symmetric positive definite and $Q = A$, then for $v \in \mathbb{R}^n$, $\|v - \Phi v\|_A = \min_{z \in \mathbb{R}^m} \|v - A^{-1} \Phi^T z\|_A$. Therefore, if x is the solution of (2.1) and y the vector in (2.2), then $\|x - \Psi y\|_A = \min_{w \in \mathbb{R}^m} \|x - \Psi w\|_A = \min_{z \in \mathbb{R}^m} \|x - A^{-1} \Phi^T z\|_A$. In particular,*

$$(2.8) \quad \|x - \Psi y\|_A = \min_{z \in \mathbb{R}^m} \|A^{-\frac{1}{2}} b - A^{-\frac{1}{2}} \Phi^T z\|.$$

Remark 3. Therefore, according to Corollary 2.4, if $Q = A$, then Ψy is the Galerkin approximation of x , i.e., the best approximation of x in the $\|\cdot\|_A$ -norm in the image of Ψ (which is equal to the image of $A^{-1} \Phi^T$). This is interesting because Ψy is obtained without prior knowledge of b .

Corollary 2.4 and Remark 3 motivate us to select $Q = A$ as the covariance matrix of the Gaussian prior distribution (mixed strategy of Player II).

2.5. Impact and Selection of the Measurement Matrix Φ . It is natural to wonder how good this recovery strategy is (under the choice $Q = A$) compared to the best possible function of y and how the approximation error is impacted by the measurement matrix Φ . If the energy norm is used to quantify accuracy, then the recovery problem can be expressed as finding the function θ of the measurements y minimizing the (worst case) approximation error $\inf_{\theta} \sup_{\|b\| \leq 1} \|x - \theta(y)\|_A / \|b\|$ with $x = A^{-1} b$ and $y = \Phi A^{-1} b$. Writing $0 < \lambda_1(A) \leq \dots \leq \lambda_n(A)$ as the eigenvalues of A in increasing order, and a_1, \dots, a_n as the corresponding eigenvectors, it is easy to obtain that (1) the best choice for Φ would correspond to measuring the projection of x on $\text{span}\{a_1, \dots, a_m\}$ and would lead to the worst approximation error $1/\sqrt{\lambda_{m+1}}$, and (2) the worst choice would correspond to measuring the projection of x on a subspace orthogonal to a_1 and would lead to the worst approximation error $1/\sqrt{\lambda_1}$. Under the decision $Q = A$ the minimal value of (2.8) is also $1/\sqrt{\lambda_{m+1}}$ and is achieved for $\text{Im}(\Phi^T) = \text{span}\{a_1, \dots, a_m\}$, and the maximal value of (2.8) is $1/\sqrt{\lambda_1}$ and is achieved when $\text{Im}(\Phi^T)$ is orthogonal to a_1 . The following theorem, which is a direct application of (2.8) and the estimate derived in [53, p. 10] (see also [66]), shows that the subset of measurement matrices that are not *nearly optimal* is of small measure if the rows of Φ^T are sampled independently on the unit sphere of \mathbb{R}^n .

THEOREM 2.5. *If Φ is an $n \times m$ matrix with i.i.d. $\mathcal{N}(0, 1)$ (Gaussian) entries, $Q = A$, x is the solution of the original equation (2.1), and $2 \leq p$, then with probability at least $1 - 3p^{-p}$, $\|x - \Psi y\|_A / \|b\| \leq (1 + 9\sqrt{m + p\sqrt{n}}) / \sqrt{\lambda_{m+1}}$.*

Although the randomization of the measurement matrix [42, 63, 43, 78] can be an efficient strategy in compressed sensing [105, 21, 20, 28, 44, 22] and in singular value decomposition/low rank approximation [53], we will not use this strategy here because the design of the interpolation operator presents the (added) difficulty of approximating the eigenvectors associated with the smallest eigenvalues of A rather than those associated with the largest ones. Furthermore, Ψ has to be computed efficiently and the dependence of the approximation constant in Theorem 2.5 on n and m can be problematic if sharp convergence estimates are to be obtained. We will instead select the measurement matrix based on the transfer property introduced in [14] and given in a discrete context in the following theorem.

THEOREM 2.6. *If A is symmetric positive definite, $Q = A$, and x is the solution of the original equation (2.1), then for any symmetric positive definite matrix B , we have*

$$(2.9) \quad \inf_{v \in \mathbb{R}^n} \sqrt{\frac{v^T B v}{v^T A v}} \min_{z \in \mathbb{R}^m} \|b - \Phi^T z\|_{B^{-1}} \leq \|x - \Psi y\|_A \leq \sup_{v \in \mathbb{R}^n} \sqrt{\frac{v^T B v}{v^T A v}} \min_{z \in \mathbb{R}^m} \|b - \Phi^T z\|_{B^{-1}}.$$

Proof. Corollary 2.4 implies that if x is the solution of the original equation (2.1), then $\|x - \Psi y\|_A = \min_{z \in \mathbb{R}^m} \|b - \Phi^T z\|_{A^{-1}}$. We finish the proof by observing that if A and B are symmetric positive definite matrices such that $\alpha_1 B \leq A \leq \alpha_2 B$ for some constants $\alpha_1, \alpha_2 > 0$, then $\alpha_2^{-1} B^{-1} \leq A^{-1} \leq \alpha_1^{-1} B^{-1}$. \square

Therefore, according to Theorem 2.6, once a good measurement matrix Φ has been identified for a symmetric positive definite matrix B such that $\alpha_1 B \leq A$, the same measurement matrix can be used for A at the cost of an increase in the bound on the error by the multiplicative factor $\alpha_1^{-1/2}$. As a prototypical example, one may consider a (stiffness) matrix A obtained from a finite element discretization of the PDE (1.1) and B may be the stiffness matrix of the finite element discretization of the Laplace Dirichlet PDE

$$(2.10) \quad -\Delta u'(x) = g(x) \text{ on } \Omega \text{ with } u' = 0 \text{ on } \partial\Omega,$$

obtained from the same finite elements (e.g., piecewise-linear nodal basis functions over the same fine mesh \mathcal{T}_h). Using the energy norm (1.3), Theorem 2.6 and Remark 3 imply the following proposition

PROPOSITION 2.7. *Let u_h (resp., u'_h) be the finite element approximation of the solution u of (1.1) (resp., the solution u' of (2.10)) over the finite nodal elements of \mathcal{T}_h . Let u_H (resp., u'_H) be the finite element approximation of the solution u of (1.1) (resp., the solution u' of (2.10)) over linear space spanned by the rows of $A^{-1}\Phi^T$ (resp., over the linear space spanned by the rows of $B^{-1}\Phi^T$). It holds true that*

$$(2.11) \quad \frac{1}{\sqrt{\lambda_{\max}(a)}} \|u'_h - u'_H\|_{H_0^1(\Omega)} \leq \|u_h - u_H\|_a \leq \frac{1}{\sqrt{\lambda_{\min}(a)}} \|u'_h - u'_H\|_{H_0^1(\Omega)}.$$

Observe that the right-hand side of (2.11) does not depend on $\lambda_{\max}(a)$, therefore, if $\lambda_{\min}(a) = 1$, then the error bound on $\|u_h - u_H\|_a$ does not depend on the contrast of a (i.e., $\lambda_{\max}(a)/\lambda_{\min}(a)$).

3. Numerical Homogenization and Design of the Interpolation Operator in the Continuous Case. We will now generalize the results and continue the analysis of section 2 in the continuous case and design the interpolation operator for (1.1) in the context of numerical homogenization.

3.1. Information Game and Gamblets. As in section 2 we will identify the interpolation operator (that will be used for the multigrid algorithm) through a noncooperative game formulation where Player I chooses the source term g (1.1) and Player II tries to approximate the solution u of (1.1) based on a finite number of measurements $(\int_{\Omega} u \phi_i)_{1 \leq i \leq m}$ obtained from linearly independent test functions $\phi_i \in L^2(\Omega)$. As in section 2, this game formulation motivates the search for a mixed strategy for Player II that can be expressed by replacing the source term g with noise ξ . We will therefore consider the stochastic PDE (SPDE)

$$(3.1) \quad \begin{cases} -\operatorname{div}(a(x)\nabla v(x)) = \xi(x), & x \in \Omega, \\ v = 0 & \text{on } \partial\Omega, \end{cases}$$

where Ω and a are the domain and conductivity of (1.1). As in section 2, to preserve the computational efficiency of the interpolation operator we will assume that ξ is a centered Gaussian field on Ω . The decision space of Player II is therefore the covariance function of ξ . Write as \mathcal{L} the differential operator $-\operatorname{div}(a\nabla)$ with zero Dirichlet boundary condition mapping $H_0^1(\Omega)$ onto $H^{-1}(\Omega)$. Motivated by the analysis (Remark 3) of section 2.4 (which can be reproduced in the continuous case) we will select the covariance function of ξ (Player II's decision) to be \mathcal{L} . Therefore, under that choice, for all $f \in H_0^1(\Omega)$, $\int_{\Omega} f(x)\xi(x) dx$ is a Gaussian random variable with mean 0 and variance $\int_{\Omega} f \mathcal{L} f = \|f\|_a^2$, where $\|f\|_a$ is the energy norm of f defined in (1.3). Introducing the scalar product on $H_0^1(\Omega)$ defined by

$$(3.2) \quad \langle v, w \rangle_a := \int_{\Omega} (\nabla v)^T a \nabla w,$$

recall that if (e_1, e_2, \dots) is an orthonormal basis of $(H_0^1(\Omega), \|\cdot\|_a)$ diagonalizing \mathcal{L} , then ξ can formally be represented as $\xi = \sum_{i=1}^{\infty} (\mathcal{L}e_i)X_i$ (where the X_i are i.i.d. $\mathcal{N}(0, 1)$ random variables) and, therefore, ξ can also be identified as the linear isometry mapping $H_0^1(\Omega)$ onto a Gaussian space and $f = \sum_{i=1}^{\infty} \langle f, e_i \rangle_a e_i$ onto $\int_{\Omega} f(x)\xi(x) dx = \sum_{i=1}^{\infty} \langle f, e_i \rangle_a X_i$.

Observe also that [77], if ξ' is white noise on Ω (i.e., a Gaussian field with covariance function $\delta(x-y)$), then ξ can be represented as $\xi = \mathcal{L}^{-\frac{1}{2}}\xi'$. Furthermore [77, Prop. 3.1], the solution of (3.1) is a Gaussian field with covariance function $G(x, y)$ (where G is the Green's function of the PDE (1.1), i.e., $\mathcal{L}G(x, y) = \delta(x-y)$ with $G(x, y) = 0$ for $y \in \partial\Omega$).

Let \mathcal{F} be the σ -algebra generated by the random variables $\int_{\Omega} v(x)\phi_i$ for $i \in \{1, \dots, m\}$ (with v solution of (3.1)). We will identify the interpolation basis elements by conditioning the solution of (3.1) on \mathcal{F} . Observe that the covariance matrix of the measurement vector $(\int_{\Omega} v(x)\phi_i)_{1 \leq i \leq m}$ is the $m \times m$ symmetric matrix Θ defined by

$$(3.3) \quad \Theta_{i,j} := \int_{\Omega^2} \phi_i(x)G(x, y)\phi_j(y) dx dy.$$

Note that for $l \in \mathbb{R}^m$, $l^T \Theta l = \|w\|_a^2$, where w is the solution of (1.1) with right-hand side $g = \sum_{i=1}^m l_i \phi_i$. Therefore (since the test functions ϕ_i are linearly independent), Θ

is positive definite and we will write Θ^{-1} as its inverse. Write $\delta_{i,j}$ as the Kronecker's delta ($\delta_{i,i} = 1$ and $\delta_{i,j} = 0$ for $i \neq j$).

THEOREM 3.1. *Let v be the solution of (3.1). It holds true that*

$$(3.4) \quad \mathbb{E}[v(x)|\mathcal{F}] = \sum_{i=1}^m \psi_i(x) \int_{\Omega} v(y)\phi_i(y) dy,$$

where the functions $\psi_i \in H_0^1(\Omega)$ are defined by

$$(3.5) \quad \psi_i(x) := \mathbb{E}\left[v(x) \middle| \int_{\Omega} v(y)\phi_j(y) dy = \delta_{i,j}, j \in \{1, \dots, m\}\right]$$

and admit the representation formula

$$(3.6) \quad \psi_i(x) = \sum_{j=1}^m \Theta_{i,j}^{-1} \int_{\Omega} G(x,y)\phi_j(y) dy.$$

Furthermore, the distribution of v conditioned on \mathcal{F} is that of a Gaussian field with mean (3.4) and covariance function $\Gamma(x,y) = G(x,y) + \sum_{i,j=1}^m \psi_i(x)\psi_j(y)\Theta_{i,j} - \sum_{i=1}^m \psi_i(x) \int_{\Omega} G(y,z)\phi_i(z) dz - \sum_{i=1}^m \psi_i(y) \int_{\Omega} G(x,z)\phi_i(z) dz$.

Proof. The proof is similar to that of [77, Thm. 3.5]. The identification of the covariance function follows from the expansion of $\Gamma(x,y) = \mathbb{E}[(v(x) - \mathbb{E}[v(x)|\mathcal{F}])(v(y) - \mathbb{E}[v(y)|\mathcal{F}])]$. Note that (3.6) proves that $\psi_i \in H_0^1(\Omega)$. \square

Since, according to (3.5) and the discussion preceding (3.1), each ψ_i is an elementary gamble (bet) on value of the solution of (1.1) given the information $\int_{\Omega} \phi_j u = \delta_{i,j}$ for $j = 1, \dots, m$, we will refer to the basis functions $(\psi_i)_{1 \leq i \leq m}$ as *gamblets*. According to (3.4), once gamblets have been identified, they form a basis for betting on the value of the solution of (1.1) given the measurements $(\int_{\Omega} \phi_j u)_{1 \leq j \leq m}$.

3.2. Optimal Recovery Properties. Although gamblets admit the representation formula (3.6), we will not use it for their practical (numerical) computation. Instead, we will work with variational properties inherited from the conditioning of the Gaussian field v . To guide our intuition, note that since \mathcal{L} is the precision function (inverse of the covariance function) of v , the conditional expectation of v can be identified by minimizing $\int_{\Omega} \psi \mathcal{L} \psi$ given measurement constraints. This observation motivates us to consider, for $i \in \{1, \dots, m\}$, the quadratic optimization problem

$$(3.7) \quad \begin{cases} \text{Minimize} & \|\psi\|_a \\ \text{subject to} & \psi \in H_0^1(\Omega) \text{ and } \int_{\Omega} \phi_j \psi = \delta_{i,j} \text{ for } j = 1, \dots, m, \end{cases}$$

where $\|\psi\|_a$ is the energy norm of ψ defined in (1.3).

The following theorem shows that (3.7) can be used to identify ψ_i and that gamblets are characterized by optimal (variational) recovery properties.

THEOREM 3.2. *It holds true that (1) the optimization problem (3.7) admits a unique minimizer ψ_i defined by (3.5) and (3.6), (2) for $w \in \mathbb{R}^m$, $\sum_{i=1}^m w_i \psi_i$ is the unique minimizer of $\|\psi\|_a$ subject to $\int_{\Omega} \psi(x)\phi_j(x) = w_j$ for $j \in \{1, \dots, m\}$, and (3) (using the scalar product defined in (3.2)) $\langle \psi_i, \psi_j \rangle_a = \Theta_{i,j}^{-1}$.*

Proof. Let $w \in \mathbb{R}^m$ and $\psi_w = \sum_{i=1}^m w_i \psi_i$ with ψ_i defined as in (3.6). The definition of Θ implies that $\int_{\Omega} \psi_w(x) \phi_j(x) = w_j$ for $j \in \{1, \dots, m\}$. Furthermore, we obtain by integration by parts that for all $\varphi \in H_0^1(\Omega)$, $\langle \psi_w, \varphi \rangle_a = \sum_{i,j=1}^m w_i \Theta_{i,j}^{-1} \int_{\Omega} \phi_j \varphi$. Therefore, if $\psi \in H_0^1(\Omega)$ is such that $\int_{\Omega} \psi(x) \phi_j(x) = w_j$ for $j \in \{1, \dots, m\}$, then $\langle \psi_w, \psi - \psi_w \rangle_a = 0$ and

$$(3.8) \quad \|\psi\|_a^2 = \|\psi_w\|_a^2 + \|\psi - \psi_w\|_a^2,$$

which finishes the proof of the optimality of ψ_i and ψ_w . \square

3.3. Optimal Accuracy of the Recovery.

Define

$$(3.9) \quad u^*(x) := \sum_{i=1}^m \psi_i(x) \int_{\Omega} u(y) \phi_i(y) dy,$$

where u is the solution of (1.1) and ψ_i are the gamblets defined by (3.5) and (3.6). Note that u^* corresponds to Player II's bet on the value of u given the measurements $(\int_{\Omega} u(y) \phi_i(y) dy)_{1 \leq i \leq m}$. In particular, if v is the solution of (3.1), then

$$(3.10) \quad u^*(x) = \mathbb{E}[v(x) | \int_{\Omega} v(y) \phi_i(y) dy = \int_{\Omega} u(y) \phi_i(y) dy].$$

For $\phi \in H^{-1}(\Omega)$ write $\mathcal{L}^{-1}\phi$ as the solution of (1.1) with $g = \phi$. The following theorem shows that u^* is the best approximation (in the energy norm) of u in $\text{span}\{\mathcal{L}^{-1}\phi_i : i \in \{1, \dots, m\}\}$.

THEOREM 3.3. *Let u be the solution of (1.1), with u^* as defined in (3.9) and (3.10). It holds true that*

$$(3.11) \quad \|u - u^*\|_a = \inf_{\psi \in \text{span}\{\mathcal{L}^{-1}\phi_i : i \in \{1, \dots, m\}\}} \|u - \psi\|_a.$$

Proof. By Theorem 3.1, $\text{span}\{\mathcal{L}^{-1}\phi_i : i \in \{1, \dots, m\}\} = \text{span}\{\psi_1, \dots, \psi_m\}$ and (3.11) follows from the fact that $\int_{\Omega} (u - u^*) \phi_j = 0$ for all j implies that $u - u^*$ is orthogonal to $\text{span}\{\psi_1, \dots, \psi_m\}$ with respect to the scalar product $\langle \cdot, \cdot \rangle_a$. \square

3.4. Transfer Property and Selection of the Measurement Functions. We will now select the measurement (test) functions ϕ_i by extending the result of Proposition 2.7 to the continuous case. For V a finite-dimensional linear subspace of $H^{-1}(\Omega)$, define

$$(3.12) \quad (\text{div } a \nabla)^{-1} V := \text{span}\{(\text{div } a \nabla)^{-1} \phi : \phi \in V\},$$

where $(\text{div } a \nabla)^{-1} \phi$ is the solution of (1.1) with $g = -\phi$. Similarly, define $\Delta^{-1} V := \text{span}\{\Delta^{-1} \phi : \phi \in V\}$, where $\Delta^{-1} \phi$ is the solution of (2.10) with $g = -\phi$.

PROPOSITION 3.4. *If u and u' are the solutions of (1.1) and (2.10) (with the same right-hand side g) and V is a finite-dimensional linear subspace of $H^{-1}(\Omega)$, then*

$$(3.13) \quad \frac{1}{\sqrt{\lambda_{\max}(a)}} \inf_{v \in \Delta^{-1} V} \|u' - v\|_{H_0^1(\Omega)} \leq \inf_{v \in (\text{div } a \nabla)^{-1} V} \|u - v\|_a \\ \leq \frac{1}{\sqrt{\lambda_{\min}(a)}} \inf_{v \in \Delta^{-1} V} \|u' - v\|_{H_0^1(\Omega)}.$$

Proof. Write G as the Green's function of (1.1) and G^* as the Green's function of (2.10). Observe that for $f \in V$ and $v = (\operatorname{div} a \nabla)^{-1} f$, $\|u - v\|_a^2 = \int_{\Omega^2} (g(x) - f(x))G(x, y)(g(y) - f(y)) dx dy$. The monotonicity of Green's function as a quadratic form (see, for instance, [12, Lemma 4.13]) implies $\int_{\Omega^2} (g(x) - f(x))G(x, y)(g(y) - f(y)) dx dy \leq \frac{1}{\lambda_{\min}(a)} \int_{\Omega^2} (g(x) - f(x))G^*(x, y)(g(y) - f(y)) dx dy$ (with a similar inequality on the left-hand side), which concludes the proof. \square

This extension, which is also directly related to the transfer property of the flux-norm (introduced in [14] and generalized in [103]; see also [115]), allows us to select accurate finite-dimensional bases for the approximation of the solution space of (1.1).

CONSTRUCTION 3.5. Let $(\tau_i)_{1 \leq i \leq m}$ be a partition of Ω such that each τ_i is Lipschitz, convex, and of diameter at most H . Let $(\phi_i)_{1 \leq i \leq m}$ be elements of $L^2(\Omega)$ such that for each i , the support of ϕ_i is contained in the closure of τ_i and $\int_{\tau_i} \phi_i \neq 0$.

PROPOSITION 3.6. Let $(\phi_i)_{1 \leq i \leq m}$ be the elements of Construction 3.5 and let u be the solution of (1.1). If $V = \operatorname{span}\{\phi_i : 1 \leq i \leq m\}$, then

$$(3.14) \quad \inf_{v \in (\operatorname{div} a \nabla)^{-1} V} \|u - v\|_a \leq CH \|g\|_{L^2(\Omega)}$$

with $C = (\pi \sqrt{\lambda_{\min}(a)})^{-1} (1 + \max_{1 \leq i \leq m} (\frac{1}{|\tau_i|} \frac{\int_{\tau_i} \phi_i^2}{(\int_{\tau_i} \phi_i)^2})^{\frac{1}{2}})$ (writing $|\tau_i|$ as the volume of τ_i).

Proof. Using Proposition 3.4 it is sufficient to complete the proof when a is the constant identity matrix. Let u' be the solution of (2.10) and $v \in \Delta^{-1} V$. Note that $\Delta v = \sum_{i=1}^m c_i \phi_i$, therefore, $\|u' - v\|_{H_0^1(\Omega)}^2 = -\int_{\Omega} (u' - v)(g - \sum_{i=1}^m c_i \phi_i)$. Taking $c_i = \int_{\tau_i} g / \int_{\tau_i} \phi_i$ we obtain that $\int_{\tau_i} (g - \sum_{j=1}^m c_j \phi_j) = 0$ and, writing $|\tau_i|$ as the volume of τ_i , $\|u' - v\|_{H_0^1(\Omega)}^2 = -\sum_{i=1}^m \int_{\tau_i} (u' - v - \frac{1}{|\tau_i|} \int_{\tau_i} (u' - v))(g - \sum_{j=1}^m c_j \phi_j)$, which by Poincaré's inequality (see [91] for the optimal constant $1/\pi$ used here) lead to $\|u' - v\|_{H_0^1(\Omega)}^2 \leq \frac{H}{\pi} \sum_{i=1}^m (\int_{\tau_i} |\nabla(u' - v)|^2)^{\frac{1}{2}} (\int_{\tau_i} (g - \sum_{j=1}^m c_j \phi_j)^2)^{\frac{1}{2}}$. Therefore, by using the Cauchy-Schwarz inequality and simplifying, $\|u' - v\|_{H_0^1(\Omega)} \leq \frac{H}{\pi} \|g - \sum_{i=1}^m c_i \phi_i\|_{L^2(\Omega)}$. Now, since each ϕ_i has support in τ_i , we have $\|\sum_{i=1}^m c_i \phi_i\|_{L^2(\Omega)}^2 = \sum_{i=1}^m (\int_{\tau_i} g)^2 \frac{\int_{\tau_i} \phi_i^2}{(\int_{\tau_i} \phi_i)^2} \leq \|g\|_{L^2(\Omega)}^2 \max_{1 \leq i \leq m} \frac{1}{|\tau_i|} \frac{\int_{\tau_i} \phi_i^2}{(\int_{\tau_i} \phi_i)^2}$, which concludes the proof. \square

The value of the constant C in Proposition 3.6 motivates us to modify Construction 3.5 as follows.

CONSTRUCTION 3.7. Let $(\phi_i)_{1 \leq i \leq m}$ be the elements constructed in Construction 3.5 under the additional assumptions that (a) each ϕ_i is equal to one on τ_i and zero elsewhere, and (b) there exists $\delta \in (0, 1)$ such that for each $i \in \{1, \dots, m\}$, τ_i contains a ball of diameter δH .

Let $(\phi_i)_{1 \leq i \leq m}$ be as in Construction 3.7. Note that the additional assumption (a) implies that the constant C in Proposition 3.6 is equal to $2/(\pi \sqrt{\lambda_{\min}(a)})$. Assumption (b) will be used for localization purposes in sections 3.5 and 3.6 (and is not required for Theorem 3.8). The following theorem is a direct consequence of Proposition 3.6 and Theorem 3.3.

THEOREM 3.8. If u is the solution of (1.1) and $(\psi_i)_{i=1}^m$ are the gamblets identified in (3.5), (3.7), and (3.6), then

$$(3.15) \quad \inf_{v \in \operatorname{span}\{\psi_1, \dots, \psi_m\}} \|u - v\|_a \leq \frac{2}{\pi \sqrt{\lambda_{\min}(a)}} H \|g\|_{L^2(\Omega)}$$

and the minimum in the left-hand side of (3.15) is achieved for $v = u^*$ defined in (3.9) and (3.10).

Remark 4. The assumption of convexity of the subdomains τ_i is only used to derive sharper constants via Poincaré's inequality for convex domains (without it, approximation error bounds remain valid after multiplication by π). Similarly, the transfer property can be used to derive constructions that are distinct from Constructions 3.5 and 3.7.

Remark 5. Gamblets defined via the constrained energy minimization problem (3.7) are analogous to the *energy minimizing bases* of [65, 114, 122, 121] and, in particular, [108]. However, they form a different set of basis functions when global constraints are added: the (total) *energy minimizing bases* of [65, 114, 122, 121, 108] are defined by minimizing the total energy $\sum_i \|\psi_i\|_a^2$ subject to the constraint $\sum_i \psi_i(x) = 1$ related to the local preservation of constants. Numerical experiments [122] suggest that total energy minimizing basis functions could lead to an $\mathcal{O}(\sqrt{H})$ convergence rate (with rough coefficients). Note that (3.7) is also analogous to the constrained minimization problems associated with polyharmonic splines [54, 30, 31, 32, 87], which can be recovered with a Gaussian prior (on ξ) with covariance function $\delta(x-y)$ (corresponding to exciting (3.1) with white noise). We suspect that the basis functions obtained in the orthogonal decomposition of [64] can also be recovered via the variational formulation (3.7) by identifying the null space of the Clement quasi-interpolation operator with that of appropriately chosen measurement functions ϕ_i .

3.5. Exponential Decay of Gamblets. Theorems 3.2 and 3.3 show that the gamblets ψ_i have optimal recovery properties analogous to the discrete case of Theorem 2.2 and Corollary 2.4. However, one might wonder why one should compute these gamblets rather than the elements $(\operatorname{div} a \nabla)^{-1} \phi_i$ since they span the same linear space (by the representation formula (3.6)). The answer lies in the fact that each gamblet ψ_i decays exponentially as a function of the distance from the support of ϕ_i , and its computation can therefore be localized to a subdomain of diameter $\mathcal{O}(H \ln \frac{1}{H})$ without impacting the order of accuracy (3.15). Consider the construction 3.7. Let ψ_i be defined as in Theorem 3.2 and let x_i be an element of τ_i . Write $B(x, r)$ as the ball of center x and radius r .

THEOREM 3.9 (exponential decay of the basis elements ψ_i). *It holds true that*

$$(3.16) \quad \int_{\Omega \cap (B(x_i, r))^c} (\nabla \psi_i)^T a \nabla \psi_i \leq e^{1 - \frac{r}{iH}} \int_{\Omega} (\nabla \psi_i)^T a \nabla \psi_i$$

with $l = 1 + (e/\pi) \sqrt{\lambda_{\max}(a)/\lambda_{\min}(a)} (1 + 2^{\frac{3}{2}} (2/\delta)^{1+d/2})$ (where e is Euler's number).

Proof. Let $k, l \in \mathbb{N}^*$ and $i \in \{1, \dots, m\}$. Let S_0 be the union of all the domains τ_j that are contained in the closure of $B(x_i, klH) \cap \Omega$, let S_1 be the union of all the domains τ_j that are contained in the closure of $(B(x_i, (k+1)lH))^c \cap \Omega$, and let $S^* = S_0^c \cap S_1^c \cap \Omega$ be the union of the remaining elements τ_j not contained in S_0 or S_1 . Let η be the function on Ω defined by $\eta(x) = \operatorname{dist}(x, S_0) / (\operatorname{dist}(x, S_0) + \operatorname{dist}(x, S_1))$. Observe that (1) $0 \leq \eta \leq 1$, (2) η is equal to zero on S_0 , (3) η is equal to one on S_1 , and (4) $\|\nabla \eta\|_{L^\infty(\Omega)} \leq \frac{1}{iH}$. Observe that $-\int_{\Omega} \eta \psi_i \operatorname{div}(a \nabla \psi_i) = \int_{\Omega} \nabla(\eta \psi_i) a \psi_i = \int_{\Omega} \eta (\nabla \psi_i)^T a \nabla \psi_i + \int_{\Omega} \psi_i (\nabla \eta)^T a \nabla \psi_i$. Therefore, $\int_{S_1} (\nabla \psi_i)^T a \nabla \psi_i \leq I_1 + I_2$ with

$$(3.17) \quad I_1 = \|\nabla \eta\|_{L^\infty(\Omega)} \left(\sum_{\tau_j \subset S^*} \int_{\tau_j} \psi_i^2 \right)^{\frac{1}{2}} \left(\int_{S^*} (\nabla \psi_i)^T a \nabla \psi_i \right)^{\frac{1}{2}} \sqrt{\lambda_{\max}(a)}$$

and $I_2 = -\int_{\Omega} \eta \psi_i \operatorname{div}(a \nabla \psi_i)$. By (3.6), $-\operatorname{div}(a \nabla \psi_i)$ is piecewise constant and equal to $\Theta_{i,j}^{-1}$ on τ_j . By the constraints of (3.7), $\int_{\tau_j} \psi_i = 0$ for $i \neq j$. Therefore (writing η_j the volume average of η over τ_j), we have

$$(3.18) \quad I_2 \leq - \sum_{\tau_j \subset S_1 \cup S^*} \int_{\tau_j} (\eta - \eta_j) \psi_i \operatorname{div}(a \nabla \psi_i) \leq \frac{1}{l} \sum_{\tau_j \subset S^*} \|\psi_i\|_{L^2(\tau_j)} \|\operatorname{div}(a \nabla \psi_i)\|_{L^2(\tau_j)}.$$

We will now need the following lemma.

LEMMA 3.10. *If $v \in \operatorname{span}\{\psi_1, \dots, \psi_m\}$, then*

$$(3.19) \quad \|\operatorname{div}(a \nabla v)\|_{L^2(\Omega)} \leq H^{-1} \|v\|_a (\lambda_{\max}(a) 2^{5+d} / \delta^{2+d})^{\frac{1}{2}}.$$

Proof. Let $c \in \mathbb{R}^m$ and $v = \sum_{i=1}^m c_i \psi_i$. Observing that $-\operatorname{div}(a \nabla v) = \sum_{i=1}^m c_i \Theta_{i,j}^{-1}$ in τ_j and using the decomposition $\|\operatorname{div}(a \nabla v)\|_{L^2(\Omega)}^2 = \sum_{i=1}^m \|\operatorname{div}(a \nabla v)\|_{L^2(\tau_j)}^2$, we obtain that

$$(3.20) \quad \|\operatorname{div}(a \nabla v)\|_{L^2(\Omega)}^2 = \sum_{j=1}^m \left(\sum_{i=1}^m c_i \Theta_{i,j}^{-1} \right)^2 |\tau_j|.$$

Furthermore, v can be decomposed over τ_j as $v = v_1 + v_2$, where v_1 solves $-\operatorname{div}(a \nabla v_1) = \sum_{i=1}^m c_i \Theta_{i,j}^{-1}$ in τ_j with $v_1 = 0$ on $\partial\tau_j$, and v_2 solves $-\operatorname{div}(a \nabla v_2) = 0$ in τ_j with $v_2 = v$ on $\partial\tau_j$. Using the notation $|\xi|_a^2 = \xi^T a \xi$, observe that $\int_{\tau_j} |\nabla v|_a^2 = \int_{\tau_j} |\nabla v_1|_a^2 + \int_{\tau_j} |\nabla v_2|_a^2$. Furthermore, $\int_{\tau_j} |\nabla v_1|_a^2 = \sum_{i=1}^m c_i \Theta_{i,j}^{-1} \int_{\tau_j} v_1$. Writing G_j as the Green's function of the operator $-\operatorname{div}(a \nabla \cdot)$ with Dirichlet boundary condition on $\partial\tau_j$, note that $\int_{\tau_j} v_1 = (\sum_{i=1}^m c_i \Theta_{i,j}^{-1}) \int_{\tau_j} G_j(x, y) dx dy$. Using the monotonicity of the Green's function as a quadratic form (as in the proof of Proposition 3.4), we have $\int_{\tau_j} G_j(x, y) dx dy \geq \frac{1}{\lambda_{\max}(a)} \int_{\tau_j} G_j^*(x, y) dx dy$, where G_j^* is the Green's function of the operator $-\Delta$ with Dirichlet boundary condition on $\partial\tau_j$. Recall that $2 \int_{\tau_j} G_j^*(x, y) dy$ is the mean exit time (from τ_j) of a Brownian motion started from x , and the mean exit time of a Brownian motion started from x to exit a ball of center x and radius r is r^2 (see, for instance, [12]). Since τ_j contains a ball of diameter δH , it follows that $2 \int_{\tau_j} G_j^*(x, y) dx dy \geq (\delta H/4)^{2+d} V_d$ (where V_d is the volume of the d -dimensional unit ball). Therefore (using $|\tau_j| \leq V_d (H/2)^d$ and simplification),

$$(3.21) \quad \int_{\tau_j} |\nabla v_1|_a^2 \geq \left(\sum_{i=1}^m c_i \Theta_{i,j}^{-1} \right)^2 |\tau_j| H^2 \delta^{2+d} / (2^{5+d} \lambda_{\max}(a)),$$

which finishes the proof after taking the sum over j . □

Now observe that since $\int_{\tau_j} \psi_i = 0$ for $i \neq j$, we obtain, using Poincaré's inequality (with the optimal constant of [91]), that $\|\psi_i\|_{L^2(\tau_j)} \leq \|\nabla \psi_i\|_{L^2(\tau_j)} H / \pi$. Therefore, combining (3.17), (3.18), and the result of Lemma 3.10, we obtain after simplification

$$(3.22) \quad \int_{S_1} (\nabla \psi_i)^T a \nabla \psi_i \leq \frac{1}{\pi l} \sqrt{\lambda_{\max}(a) / \lambda_{\min}(a)} (1 + 2^{\frac{3}{2}} (2/\delta)^{1+d/2}) \int_{S^*} (\nabla \psi_i)^T a \nabla \psi_i.$$

Taking $l \geq \frac{\varepsilon}{\pi} \sqrt{\lambda_{\max}(a) / \lambda_{\min}(a)} (1 + 2^{\frac{3}{2}} (2/\delta)^{1+d/2})$ and enlarging the integration domain on the right-hand side, we obtain $\int_{S_1} (\nabla \psi_i)^T a \nabla \psi_i \leq e^{-1} \int_{S^* \cup S_1} (\nabla \psi_i)^T a \nabla \psi_i$. We conclude the proof via straightforward iteration on k . □

3.6. Localization of the Basis Elements. Theorem 3.9 allows us to localize the construction of basis elements ψ_i as follows. For $r > 0$, let S_r be the union of the subdomains τ_j intersecting $B(x_i, r)$ (recall that x_i is an element of τ_i) and let $\psi_i^{\text{loc},r}$ be the minimizer of the quadratic problem

$$(3.23) \quad \begin{cases} \text{Minimize} & \int_{S_r} (\nabla \psi)^T a \nabla \psi \\ \text{subject to} & \psi \in H_0^1(S_r) \text{ and } \int_{S_r} \phi_j \psi = \delta_{i,j} \text{ for } j \text{ such that } \tau_j \subset S_r. \end{cases}$$

We will naturally identify $\psi_i^{\text{loc},r}$ with its extension to $H_0^1(\Omega)$ by setting $\psi_i^{\text{loc},r} = 0$ outside of S_r . From now on, to simplify the expression of constants, we will assume without loss of generality that the domain is rescaled so that $\text{diam}(\Omega) \leq 1$.

THEOREM 3.11. *It holds true that*

$$(3.24) \quad \|\psi_i - \psi_i^{\text{loc},r}\|_a \leq C e^{-\frac{r}{2lH}},$$

where l is defined in Theorem 3.9, $C = (\lambda_{\max}(a)/\sqrt{\lambda_{\min}(a)})H^{-\frac{d}{2}-2}2^{2d+9}/(\sqrt{V_d}\delta^{d+2})$, and V_d is the volume of the d -dimensional unit ball.

Proof. We will need the following lemma.

LEMMA 3.12. *It holds true that*

$$(3.25) \quad \|\psi_i\|_a \leq (H\delta)^{-\frac{d}{2}-1} \sqrt{\lambda_{\max}(a)} 2^{\frac{3}{2}d+2} (V_d)^{-\frac{1}{2}},$$

where V_d is the volume of the d -dimensional unit ball and

$$(3.26) \quad |\langle \psi_i, \psi_j \rangle_a| \leq e^{-\frac{r_{i,j}}{2lH}} H^{-2-d} \lambda_{\max}(a) 2^{\frac{5d+11}{2}} / (V_d \delta^{d+2}),$$

where l is the constant of Theorem 3.9 and $r_{i,j}$ is the distance between τ_i and τ_j .

Proof. Since τ_i contains a ball $B(x_i, \delta H/2)$ of center $x_i \in \tau_i$ and diameter $\delta H/2$, there exists a piecewise-differentiable function η equal to 1 on $B(x_i, \delta H/4)$, equal to 0 on $(B(x_i, \delta H/2))^c$, and such that $0 \leq \eta \leq 1$ with $\|\nabla \eta\|_{L^\infty(\Omega)} \leq \frac{4}{H\delta}$. Since $\psi = \eta / (\int_{\tau_i} \eta)$ satisfies the constraints of the minimization problem (3.7), we have $\|\psi_i\|_a \leq \|\psi\|_a$, which proves (3.25). Theorem 3.2 implies that $\langle \psi_i, \psi_j \rangle_a = \Theta_{i,j}^{-1}$. Observing that $-\text{div}(a\nabla \psi_i)$ is piecewise constant and equal to $\Theta_{i,j}^{-1}$ on τ_j and applying (3.21) (with $v = \psi_i$ and using $\int_{\tau_j} |\nabla v_1|_a^2 \leq \int_{\tau_j} |\nabla v|_a^2$), we obtain that

$$(3.27) \quad |\Theta_{i,j}^{-1}| \leq (\lambda_{\max}(a) 2^{5+d} / (\delta^{2+d} |\tau_j|))^{1/2} H^{-1} \left(\int_{\tau_j} (\nabla \psi_i)^T a \nabla \psi_i \right)^{1/2},$$

which leads to (3.26) by the exponential decay obtained in Theorem 3.9 and (3.25). \square

Let us now prove Theorem 3.11. Let S_0 be the union of the subdomains τ_j not contained in S_r and let S_1 be the union of the subdomains τ_j that are at distance at least H from S_0 (for $S_0 = \emptyset$ the proof is trivial, so we may assume that $S_0 \neq \emptyset$; similarly, it is no restriction to assume that $S_1 \neq \emptyset$). Let η be the function on Ω defined by $\eta(x) = \text{dist}(x, S_0) / (\text{dist}(x, S_0) + \text{dist}(x, S_1))$. Observe that is a piecewise-differentiable function on Ω such that (1) η is equal to one on S_1 and zero on S_0 , (2) $\|\nabla \eta\|_{L^\infty(\Omega)} \leq \frac{1}{H}$, and (3) $0 \leq \eta \leq 1$. Since $\psi_i^{\text{loc},r}$ satisfies the constraints of (3.7), we have from (3.8),

$$(3.28) \quad \|\psi_i - \psi_i^{\text{loc},r}\|_a^2 = \|\psi_i^{\text{loc},r}\|_a^2 - \|\psi_i\|_a^2.$$

Let $\psi_k^{i,r}$ be the minimizer of $\int_{S_r} (\nabla\psi)^T a \nabla\psi$ subject to $\psi \in H_0^1(S_r)$ and $\int_{S_r} \phi_j \psi = \delta_{k,j}$ for $\tau_j \subset S_r$. Write $w_j = \int_{\Omega} \eta\psi_i \phi_j$. Let $\psi_w^{i,r} := \sum_{j=1}^m w_j \psi_j^{i,r}$. Noting that $\psi_w^{i,r} = \psi_i^{\text{loc},r} + \sum_{\tau_j \subset S^*} w_j \psi_j^{i,r}$, where S^* is the union of $\tau_j \subset S_r$ not contained in S_1 , and using property (3) of Theorem 3.2 (with $\Theta_{k,k'}^{i,-1} = \int_{S_r} (\nabla\psi_k^{i,r})^T a \nabla\psi_{k'}^{i,r}$), it follows that

$$(3.29) \quad \|\psi_w^{i,r}\|_a^2 = \|\psi_i^{\text{loc},r}\|_a^2 + \left\| \sum_{\tau_j \subset S^*} w_j \psi_j^{i,r} \right\|_a^2 + 2 \sum_{\tau_j \subset S^*} \Theta_{i,j}^{i,-1} w_j.$$

Noting that $\eta\psi_i \in H_0^1(S_r)$, Theorem 3.2 implies that $\|\psi_w^{i,r}\|_a \leq \|\eta\psi_i\|_a$, which, combined with (3.29) and (3.28), leads to $\|\psi_i - \psi_i^{\text{loc},r}\|_a^2 \leq \|\eta\psi_i\|_a^2 - \|\psi_i\|_a^2 - 2 \sum_{\tau_j \subset S^*} \Theta_{i,j}^{i,-1} w_j$ and (using $\|\eta\psi_i\|_a^2 - \|\psi_i\|_a^2 \leq \int_{S^*} \nabla(\eta\psi_i)^T a \nabla(\eta\psi_i)$)

$$(3.30) \quad \|\psi_i - \psi_i^{\text{loc},r}\|_a^2 \leq \int_{S^*} \nabla(\eta\psi_i)^T a \nabla(\eta\psi_i) + 2 \left| \sum_{\tau_j \subset S^*} \Theta_{i,j}^{i,-1} w_j \right|.$$

Now observe $\frac{1}{2} \int_{S^*} \nabla(\eta\psi_i)^T a \nabla(\eta\psi_i) \leq \int_{\Omega \cap (B(x_i, r-2H))^c} (\nabla\psi_i)^T a \nabla\psi_i + \frac{\lambda_{\max}(a)}{H^2} \int_{S^*} |\psi_i|^2$. Applying Poincaré’s inequality we obtain $\int_{S^*} |\psi_i|^2 \leq \frac{1}{\pi^2} H^2 \sum_{\tau_j \subset S^*} \int_{\tau_j} |\nabla\psi_i|^2$ (since $\int_{\tau_j} \psi_i = 0$ for $\tau_j \subset S^*$) and $\int_{S^*} |\psi_i|^2 \leq \frac{H^2}{\pi^2 \lambda_{\min}(a)} \int_{\Omega \cap (B(x_i, r-2H))^c} (\nabla\psi_i)^T a \nabla\psi_i$. Combining these equations with the exponential decay of Theorem 3.9 we deduce

$$(3.31) \quad \int_{S^*} \nabla(\eta\psi_i)^T a \nabla(\eta\psi_i) \leq 2 \left(1 + \lambda_{\max}(a) / (\pi^2 \lambda_{\min}(a)) \right) e^{1 - \frac{r-2H}{lH}} \|\psi_i\|_a^2.$$

Similarly, using the Cauchy–Schwarz and Poincaré inequalities, we have for $\tau_j \subset S^*$, $|w_j| \leq |\tau_j|^{\frac{1}{2}} \|\psi_i\|_{L^2(\tau_j)} \leq |\tau_j|^{\frac{1}{2}} (\int_{\tau_j} (\nabla\psi_i)^T a (\nabla\psi_i))^{\frac{1}{2}} / \sqrt{\lambda_{\min}(a)}$ and $|\sum_{\tau_j \subset S^*} \Theta_{i,j}^{i,-1} w_j| \leq |\sum_{\tau_j \subset S^*} (\Theta_{i,j}^{i,-1})^2 |\tau_j|^{\frac{1}{2}} (\int_{S^*} (\nabla\psi_i)^T a (\nabla\psi_i) / \lambda_{\min}(a))^{\frac{1}{2}}|$. Using (3.27) we obtain that $|\sum_{\tau_j \subset S^*} (\Theta_{i,j}^{i,-1})^2 |\tau_j|^{\frac{1}{2}}| \leq (\lambda_{\max}(a) 2^{5+d} / \delta^{2+d})^{\frac{1}{2}} H^{-1} (\int_{S^*} (\nabla\psi_i^{i,r})^T a \nabla\psi_i^{i,r})^{\frac{1}{2}}$, which by the exponential decay of Theorem (3.9) (and $\|\psi_i\|_a \leq \|\psi_i^{i,r}\|_a$) leads to

$$(3.32) \quad \left| \sum_{\tau_j \subset S^*} \Theta_{i,j}^{i,-1} w_j \right| \leq \left(\frac{\lambda_{\max}(a) 2^{5+d}}{\lambda_{\min}(a) \delta^{2+d}} \right)^{\frac{1}{2}} H^{-1} \|\psi_i^{i,r}\|_a^2 e^{1 - \frac{r-2H}{lH}}.$$

Using (3.25) to bound $\|\psi_i^{i,r}\|_a$ and combining (3.32) with (3.31) and (3.30) concludes the proof. \square

The following theorem shows that gamblots preserve the $\mathcal{O}(H)$ rate of convergence (in the energy norm) after localization to subdomains of size $\mathcal{O}(H \ln(1/H))$. They can therefore be used as localized basis functions in numerical homogenization [5, 86, 64, 87]. Section 4 will show that they can also be computed hierarchically at near-linear complexity.

THEOREM 3.13. *Let u be the solution of (1.1) and $(\psi_1^{\text{loc},r})_{1 \leq i \leq m}$ be the localized gamblots identified in (3.23), then for $r \geq H(C_1 \ln \frac{1}{H} + C_2)$ we have*

$$(3.33) \quad \inf_{v \in \text{span}\{\psi_1^{\text{loc},r}, \dots, \psi_m^{\text{loc},r}\}} \|u - v\|_a \leq \frac{1}{\sqrt{\lambda_{\min}(a)}} H \|g\|_{L^2(\Omega)}.$$

The constants are $C_1 = (d + 4)l$ and $C_2 = 2l \ln \left(\frac{\lambda_{\max}(a)}{\lambda_{\min}(a)} \frac{2^{\frac{3}{2}d+11}}{\delta^{d+2}} \right)$, where l is the constant of Theorem 3.11. Furthermore, the inequality (3.33) is achieved for $v = \sum_{i=1}^m \psi_i^{\text{loc},r} \int_{\Omega} u \phi_i$.

Proof. Let $v_1 := \sum_{i=1}^m c_i \psi_i$ and $v_2 = \sum_{i=1}^m c_i \psi_i^{\text{loc},r}$ with $c_i = \int_{\Omega} u \phi_i$. Theorem 3.8 implies that $\|u - v_1\|_a \leq 2/(\pi \sqrt{\lambda_{\min}(a)}) H \|g\|_{L^2(\Omega)}$. Observe that $\|u - v_2\|_a \leq \|u - v_1\|_a + \|v_1 - v_2\|_a$ and $\|v_1 - v_2\|_a \leq \max_i \|\psi_i - \psi_i^{\text{loc},r}\|_a \sum_{i=1}^m |c_i|$. Using Poincaré's inequality $\|u\|_{L^2(\Omega)} \leq \text{diam}(\Omega) \|\nabla u\|_{L^2(\Omega)}$ (with $\text{diam}(\Omega) \leq 1$) we obtain $\sum_{i=1}^m |c_i| \leq \int_{\Omega} |u| \leq \|g\|_{L^2(\Omega)} 2^{-d/2} V_d^{1/2} / \lambda_{\min}(a)$. We conclude using Theorem 3.11 to bound $\max_i \|\psi_i - \psi_i^{\text{loc},r}\|_a$. \square

4. Multiresolution Operator Decomposition. Building on the analysis of section 3, we will now gamble on the approximation of the solution of (1.1) based on measurements performed at different levels of resolution. The resulting hierarchical (and nested) games will then be used to derive a multiresolution decomposition of (1.1) (orthogonal across subscales) and a near-linear complexity multiresolution algorithm with a priori error bounds.

4.1. Hierarchy of Nested Measurement Functions. In order to define the hierarchy of games we will first define a hierarchy of nested measurement functions.

DEFINITION 4.1. *We say that \mathcal{I} is an index tree of depth q if it is a finite set of q -tuples of the form $i = (i_1, \dots, i_q)$ with $1 \leq i_1 \leq m_0$ and $1 \leq i_j \leq m_{(i_1, \dots, i_{j-1})}$ for $j \geq 2$, where m_0 and $m_{(i_1, \dots, i_{j-1})}$ are strictly positive integers. For $1 \leq k \leq q$ and $i = (i_1, \dots, i_q) \in \mathcal{I}$, we write $i^{(k)} := (i_1, \dots, i_k)$ and $\mathcal{I}^{(k)} := \{i^{(k)} : i \in \mathcal{I}\}$. For $k \leq k' \leq q$ and $j = (j_1, \dots, j_{k'}) \in \mathcal{I}^{(k')}$, we write $j^{(k)} := (j_1, \dots, j_k)$. For $i \in \mathcal{I}^{(k)}$ and $k \leq k' \leq q$, we write $i^{(k,k')}$ the set of elements $j \in \mathcal{I}^{(k')}$ such that $j^{(k)} = i$.*

CONSTRUCTION 4.2. *Let \mathcal{I} be an index tree of depth q . Let $\delta \in (0, 1)$ and $0 < H_q < \dots < H_1 < 1$. Let $(\tau_i^{(k)}, k \in \{1, \dots, q\}, i \in \mathcal{I}^{(k)})$ be a collection of subsets of Ω such that (1) for $1 \leq k \leq q$, $(\tau_i^{(k)}, i \in \mathcal{I}^{(k)})$ is a partition of Ω such that each $\tau_i^{(k)}$ is a Lipschitz, convex subset of Ω of diameter at most H_k and contains a ball of diameter δH_k , and (2) the sequence of partitions is nested, i.e., for $k \in \{1, \dots, q-1\}$ and $i \in \mathcal{I}^{(k)}$, $\tau_i^{(k)} := \cup_{j \in i^{(k,k+1)}} \tau_j^{(k+1)}$.*

As in Remark 4, the assumption of convexity of the subdomains $\tau_i^{(k)}$ is not necessary to the results presented here and is only used to derive sharper/simpler constants. Let $\phi_i^{(k)}$ be the indicator function of the set $\tau_i^{(k)}$ (i.e., $\phi_i^{(k)} = 1$ if $x \in \tau_i^{(k)}$ and $\phi_i^{(k)} = 0$ if $x \notin \tau_i^{(k)}$). Note that the nesting of the domain decomposition implies that of the measurement functions, i.e., for $k \in \{1, \dots, q-1\}$ and $i \in \mathcal{I}^{(k)}$,

$$(4.1) \quad \phi_i^{(k)} = \sum_{j \in \mathcal{I}^{(k+1)}} \pi_{i,j}^{(k,k+1)} \phi_j^{(k+1)},$$

where $\pi^{(k,k+1)}$ is the $\mathcal{I}^{(k)} \times \mathcal{I}^{(k+1)}$ matrix defined by $\pi_{i,j}^{(k,k+1)} = 1$ if $j \in i^{(k,k+1)}$ and $\pi_{i,j}^{(k,k+1)} = 0$ if $j \notin i^{(k,k+1)}$. We will assume without loss of generality that $\|\phi_i^{(k)}\|_{L^2(\Omega)}^2 = |\tau_i^{(k)}|$ is constant in i (for the general case, rescale/renormalize each $\phi_i^{(k)}$ and the entries of $\pi^{(k,k+1)}$ by the corresponding multiplicative factors; we will keep track of the dependence of some of the constants on $\max_{i,j} |\tau_i^{(k)}|/|\tau_j^{(k)}|$).

4.2. Hierarchy of Nested Gambles and Multiresolution Approximations. Let us now consider the problem of recovering the solution of (1.1) based on the nested measurements $(\int_{\Omega} u \phi_i^{(k)})_{i \in \mathcal{I}^{(k)}}$ for $k \in \{1, \dots, q\}$. As in section 3 we are lead to investigate the mixed strategy (for Player II) expressed by replacing the source term g with

a centered Gaussian field with covariance function $\mathcal{L} = -\operatorname{div}(a\nabla)$. Under that mixed strategy, Player II's bet on the value of the solution of (1.1), given the measurements $(\int_{\Omega} u(y)\phi_i^{(k)}(y) dy)_{i \in \mathcal{I}^{(k)}}$, is (see section 3.3)

$$(4.2) \quad u^{(k)}(x) := \sum_{i \in \mathcal{I}^{(k)}} \psi_i^{(k)}(x) \int_{\Omega} u(y)\phi_i^{(k)}(y) dy,$$

where (see Theorem 3.2), for $k \in \{1, \dots, q\}$ and $i \in \mathcal{I}^{(k)}$, $\psi_i^{(k)}$ is the minimizer of

$$(4.3) \quad \begin{cases} \text{Minimize} & \|\psi\|_a \\ \text{subject to} & \psi \in H_0^1(\Omega) \text{ and } \int_{\Omega} \phi_j^{(k)}\psi = \delta_{i,j} \text{ for } j \in \mathcal{I}^{(k)}. \end{cases}$$

Define $\mathfrak{V}^{(q+1)} := H_0^1(\Omega)$ and, for $k \in \{1, \dots, q\}$,

$$(4.4) \quad \mathfrak{V}^{(k)} := \operatorname{span}\{\psi_i^{(k)} \mid i \in \mathcal{I}^{(k)}\}.$$

By Theorem 3.1 $\operatorname{span}\{\psi_i^{(k)} \mid i \in \mathcal{I}^{(k)}\} = \operatorname{span}\{\mathcal{L}^{-1}\phi_i^{(k)} \mid i \in \mathcal{I}^{(k)}\}$, and the nesting (4.1) of the measurement functions implies the nesting of the spaces $\mathfrak{V}^{(k)}$. The following theorem (a direct application of Theorems 3.3 and 3.8) shows that $u^{(k)}$ is the best (energy norm) approximation of the solution of (1.1) in $\mathfrak{V}^{(k)}$.

THEOREM 4.3. *It holds true that (1) for $k \in \{1, \dots, q\}$, $\mathfrak{V}^{(k)} \subset \mathfrak{V}^{(k+1)}$ and $\mathfrak{V}^{(k)} = \operatorname{span}\{\mathcal{L}^{-1}\phi_i^{(k)} \mid i \in \mathcal{I}^{(k)}\}$, and (2) if u is the solution of (1.1) and $u^{(k)}$ is defined in (4.2), then*

$$(4.5) \quad \|u - u^{(k)}\|_a = \inf_{v \in \mathfrak{V}^{(k)}} \|u - v\|_a \leq \frac{2}{\pi\sqrt{\lambda_{\min}(a)}} H_k \|g\|_{L^2(\Omega)}.$$

4.3. Nested Games and Martingale/Multiresolution Decomposition. As in section 3 we consider the mixed strategy (for Player II) expressed by replacing the source term g with a centered Gaussian field with covariance function \mathcal{L} . Under this mixed strategy, Player II's bet (4.2) on the value of the solution of (1.1), given the measurements $(\int_{\Omega} u(y)\phi_i^{(k)}(y) dy)_{i \in \mathcal{I}^{(k)}}$, can also be obtained by conditioning the solution v of the SPDE (3.1) (see (3.10)), i.e.,

$$(4.6) \quad u^{(k)}(x) = \mathbb{E} \left[v(x) \mid \int_{\Omega} v(y)\phi_i^{(k)}(y) dy = \int_{\Omega} u(y)\phi_i^{(k)}(y) dy, i \in \mathcal{I}^{(k)} \right].$$

Furthermore, each gamblet $\psi_i^{(k)}$ represents Player II's bet on the value of the solution of (1.1) given the measurements $\int_{\Omega} u(y)\phi_j^{(k)}(y) dy = \delta_{i,j}$, i.e.,

$$(4.7) \quad \psi_i^{(k)} = \mathbb{E} \left[v \mid \int_{\Omega} v(y)\phi_j^{(k)}(y) dy = \delta_{i,j}, j \in \mathcal{I}^{(k)} \right].$$

Now consider the nesting of noncooperative games where Player I chooses g in (1.1) and Player II is shown the measurements $(\int_{\Omega} u\phi_i^{(k)})_{i \in \mathcal{I}^{(k)}}$, step by step in a hierarchical manner, from coarse ($k = 1$) to fine ($k = q$), and must, at each step k of the game, gamble on the value of solution u . The following theorem and (4.6) show that the resulting sequence of approximations $u^{(k)}$ forms the realization of a martingale with independent increments.

THEOREM 4.4. Let \mathcal{F}_k be the σ -algebra generated by the random variables $(\int_{\Omega} v(x)\phi_i^{(k)})_{i \in \mathcal{I}^{(k)}}$ and

$$(4.8) \quad v^{(k)}(x) := \mathbb{E}[v(x)|\mathcal{F}_k] = \sum_{i \in \mathcal{I}^{(k)}} \psi_i^{(k)}(x) \int_{\Omega} v(y)\phi_i^{(k)}(y) dy.$$

It holds true that (1) $\mathcal{F}_1, \dots, \mathcal{F}_q$ forms a filtration, i.e., $\mathcal{F}_k \subset \mathcal{F}_{k+1}$, (2) for $x \in \Omega$, $v^{(k)}(x)$ is a martingale with respect to the filtration $(\mathcal{F}_k)_{k \geq 1}$, i.e., $v^{(k)}(x) = \mathbb{E}[v^{(k+1)}(x)|\mathcal{F}_k]$, and (3) $v^{(1)}$ and the increments $(v^{(k+1)} - v^{(k)})_{k \geq 1}$ are independent Gaussian fields.

Proof. The nesting (4.1) of the measurement functions implies $\mathcal{F}_k \subset \mathcal{F}_{k+1}$, and $(\mathcal{F}_k)_{k \geq 1}$ is therefore filtration. The fact that $v^{(k)}$ is a martingale follows from $v^{(k)} = \mathbb{E}[v|\mathcal{F}_k]$. Since $v^{(1)}$ and the increments $(v^{(k+1)} - v^{(k)})_{k \geq 1}$ are Gaussian fields belonging to the same Gaussian space, their independence is equivalent to zero covariance, which follows from the martingale property, i.e., for $k \geq 1$, $\mathbb{E}[v^{(1)}(v^{(k+1)} - v^{(k)})] = \mathbb{E}[\mathbb{E}[v^{(1)}(v^{(k+1)} - v^{(k)})|\mathcal{F}_k]] = \mathbb{E}[v^{(1)}\mathbb{E}[(v^{(k+1)} - v^{(k)})|\mathcal{F}_k]] = 0$ and for $k > j \geq 1$, $\mathbb{E}[(v^{(j+1)} - v^{(j)})(v^{(k+1)} - v^{(k)})] = \mathbb{E}[(v^{(j+1)} - v^{(j)})\mathbb{E}[(v^{(k+1)} - v^{(k)})|\mathcal{F}_k]] = 0$. \square

Remark 6. Theorem 4.4 enables the application of classical results concerning martingales to the numerical analysis of $v^{(k)}$ (and $u^{(k)}$). In particular, (1) martingale (concentration) inequalities can be used to control the fluctuations of $v^{(k)}$, (2) optimal stopping times can be used to derive optimal strategies for stopping numerical simulations based on loss functions mixing computation costs with the cost of imperfect decisions, and (3) taking $q = \infty$ in the construction of the basis elements $\psi_i^{(k)}$ (with a sequence H_k decreasing toward 0) and using the martingale convergence theorem imply that, for all $\varphi \in C_0^\infty(\Omega)$, $\int_{\Omega} v^{(k)}\varphi \rightarrow \int_{\Omega} v\varphi$ as $k \rightarrow \infty$ (a.s. and in L^1).

The independence of the increments $v^{(k+1)} - v^{(k)}$ is related to the following orthogonal multiresolution decomposition of the operator (1.1). For $\mathfrak{V}^{(k)}$ defined as in (4.4) and for $k \in \{2, \dots, q+1\}$, let $\mathfrak{W}^{(k)}$ be the orthogonal complement of $\mathfrak{V}^{(k-1)}$ within $\mathfrak{V}^{(k)}$ with respect to the scalar product $\langle \cdot, \cdot \rangle_a$. Write \oplus_a as the orthogonal direct sum with respect to the scalar product $\langle \cdot, \cdot \rangle_a$. Note that by Theorem 4.3, $u^{(k)}$ defined by (4.2) is the finite element solution of (1.1) in $\mathfrak{V}^{(k)}$ (in particular, we will write $u^{(q+1)} = u$).

THEOREM 4.5. It holds true that (1) for $k \in \{2, \dots, q+1\}$,

$$(4.9) \quad \mathfrak{V}^{(k)} = \mathfrak{V}^{(1)} \oplus_a \mathfrak{W}^{(2)} \oplus_a \dots \oplus_a \mathfrak{W}^{(k)},$$

(2) for $k \in \{1, \dots, q\}$, $u^{(k+1)} - u^{(k)}$ belongs to $\mathfrak{W}^{(k+1)}$ and

$$(4.10) \quad u = u^{(1)} + (u^{(2)} - u^{(1)}) + \dots + (u^{(q)} - u^{(q-1)}) + (u - u^{(q)})$$

is the orthogonal decomposition of u in $H_0^1(\Omega) = \mathfrak{V}^{(1)} \oplus_a \mathfrak{W}^{(2)} \oplus_a \dots \oplus_a \mathfrak{W}^{(q)} \oplus_a \mathfrak{W}^{(q+1)}$, and (3) $u^{(k+1)} - u^{(k)}$ is the finite element solution of (1.1) in $\mathfrak{W}^{(k+1)}$.

Proof. Observe that since the $\mathfrak{V}^{(k)}$ are nested (see Theorem 4.3), $u^{(k+1)} - u^{(k)}$ belongs to $\mathfrak{V}^{(k+1)}$. Furthermore (by property (1) of Theorem 4.3 and integration by parts), for $i \in \mathcal{I}^{(k)}$, $\langle u^{(k+1)} - u^{(k)}, \psi_i^{(k)} \rangle_a$ belongs to $\text{span}\{\int_{\Omega} (u^{(k+1)} - u^{(k)})\phi_i^{(k)} \mid i \in \mathcal{I}^{(k)}\}$. Finally, (4.2), the constraints of (4.3), and the nesting property (4.1) imply that for $i \in \mathcal{I}^{(k)}$, $\int_{\Omega} (u^{(k+1)} - u^{(k)})\phi_i^{(k)} = \sum_{j \in \mathcal{I}^{(k+1)}} \pi_{i,j}^{(k,k+1)} \int_{\Omega} u\phi_j^{(k+1)} - \int_{\Omega} u\phi_i^{(k)} = 0$, which implies that $u^{(k+1)} - u^{(k)}$ belongs to $\mathfrak{W}^{(k+1)}$. \square

4.4. Interpolation and Restriction Matrices/Operators. Since the spaces $\mathfrak{V}^{(k)}$ are nested there exists a $\mathcal{I}^{(k)} \times \mathcal{I}^{(k+1)}$ matrix $R^{(k,k+1)}$ such that for $1 \leq k \leq q - 1$ and $i \in \mathcal{I}^{(k)}$,

$$(4.11) \quad \psi_i^{(k)} = \sum_{j \in \mathcal{I}^{(k+1)}} R_{i,j}^{(k,k+1)} \psi_j^{(k+1)}.$$

We will refer to $R^{(k,k+1)}$ as the restriction matrix and to its transpose $R^{(k+1,k)} := (R^{(k,k+1)})^T$ as the interpolation/prolongation matrix. The following theorem shows that (see Figure 1) $R_{i,j}^{(k,k+1)}$ is Player II's best bet on the value of $\int_{\Omega} u \phi_j^{(k+1)}$ given the information that $\int_{\Omega} u \phi_s^{(k)} = \delta_{i,s}$, $s \in \mathcal{I}^{(k)}$.

THEOREM 4.6. *It holds true that for $i \in \mathcal{I}^{(k)}$ and $j \in \mathcal{I}^{(k+1)}$,*

$$R_{i,j}^{(k,k+1)} = \int_{\Omega} \psi_i^{(k)} \phi_j^{(k+1)} = \mathbb{E} \left[\int_{\Omega} v(y) \phi_j^{(k+1)}(y) dy \mid \int_{\Omega} v(y) \phi_l^{(k)}(y) dy = \delta_{i,l}, l \in \mathcal{I}^{(k)} \right].$$

Proof. The first equality is obtained by integrating (4.11) against $\phi_j^{(k+1)}$ and using the constraints satisfied by $\psi_j^{(k+1)}$ in (4.3). For the second equality, observe that since \mathcal{F}_k is a filtration, we can replace v in the representation formula (4.7) by $v^{(k)}$ (as defined by the right-hand side of (4.8)) and obtain $\psi_i^{(k)}(x) = \sum_{j \in \mathcal{I}^{(k+1)}} \psi_j^{(k+1)}(x) \mathbb{E} \left[\int_{\Omega} v(y) \phi_j^{(k+1)}(y) dy \mid \int_{\Omega} v(y) \phi_l^{(k)}(y) dy = \delta_{i,l}, l \in \mathcal{I}^{(k)} \right]$, which corresponds to (4.11). \square

4.5. Nested Computation of the Interpolation and Stiffness Matrices. Let v be the solution of (3.1). Observe that $(\int_{\Omega} v(x) \phi_i^{(k)})_{i \in \mathcal{I}^{(k)}}$ is a Gaussian vector with (symmetric, positive definite) covariance matrix $\Theta^{(k)}$ defined by, for $i, j \in \mathcal{I}^{(k)}$,

$$(4.12) \quad \Theta_{i,j}^{(k)} := \int_{\Omega^2} \phi_i^{(k)}(x) G(x, y) \phi_j^{(k)}(y) dx dy.$$

As in (3.3), $\Theta^{(k)}$ is invertible and we write $\Theta^{(k),-1}$ as its inverse. Observe that, as in Theorem 3.2, $\psi_i^{(k)}$ admits the representation formula

$$(4.13) \quad \psi_i^{(k)}(x) = \sum_{j \in \mathcal{I}^{(k)}} \Theta_{i,j}^{(k),-1} \int_{\Omega} G(x, y) \phi_j^{(k)}(y) dy.$$

Observe that, as in Theorem 3.2, $\Theta^{(k),-1} = A^{(k)}$, where $A^{(k)}$ is the (symmetric, positive definite) stiffness matrix of the elements $\psi_i^{(k)}$, i.e., for $i, j \in \mathcal{I}^{(k)}$,

$$(4.14) \quad A_{i,j}^{(k)} := \langle \psi_i^{(k)}, \psi_j^{(k)} \rangle_a.$$

Write $\pi^{(k+1,k)}$ as the transpose of the matrix $\pi^{(k,k+1)}$ (defined below (4.1)) and $I^{(k)}$ as the $\mathcal{I}^{(k)} \times \mathcal{I}^{(k)}$ identity matrix. The following theorem enables the hierarchical/nested computation of $A^{(k)}$ from $A^{(k+1)}$.

THEOREM 4.7. *For $b \in \mathbb{R}^{\mathcal{I}^{(k)}}$, $R^{(k+1,k)}b$ is the (unique) minimizer $c \in \mathbb{R}^{\mathcal{I}^{(k+1)}}$ of*

$$(4.15) \quad \begin{cases} \text{Minimize} & c^T A^{(k+1)} c \\ \text{subject to} & \pi^{(k,k+1)} c = b. \end{cases}$$

Furthermore, $R^{(k,k+1)} \pi^{(k+1,k)} = \pi^{(k,k+1)} R^{(k+1,k)} = I^{(k)}$, $R^{(k,k+1)} = A^{(k)} \pi^{(k,k+1)} \Theta^{(k+1)}$,

$\Theta^{(k)} = \pi^{(k,k+1)}\Theta^{(k+1)}\pi^{(k+1,k)}$, and

$$(4.16) \quad A^{(k)} = R^{(k,k+1)}A^{(k+1)}R^{(k+1,k)}.$$

Proof. Using the decompositions (4.11) and (4.1) in $\int_{\Omega} \phi_j^{(k)} \psi_i^{(k)} = \delta_{i,j}$ leads to $R^{(k,k+1)}\pi^{(k+1,k)} = I^{(k)}$. Using (4.13) and (4.1) to expand $\psi_i^{(k)}$, Theorem 4.6 leads to $R^{(k,k+1)} = A^{(k)}\pi^{(k,k+1)}\Theta^{(k+1)}$. Using (4.1) to expand $\phi_i^{(k)}$ and $\phi_j^{(k)}$, in (4.12) leads to $\Theta^{(k)} = \pi^{(k,k+1)}\Theta^{(k+1)}\pi^{(k+1,k)}$. Using (4.11) to expand $\psi_i^{(k)}$ and $\psi_j^{(k)}$ in (4.14) leads to (4.16). Let $b \in \mathbb{R}^{\mathcal{I}^{(k)}}$. Theorem 3.2 implies that $\sum_{i \in \mathcal{I}^{(k)}} b_i \psi_i^{(k)}$ is the unique minimizer of $\|v\|_a^2$ subject to $v \in H_0^1(\Omega)$ and $\int_{\Omega} \phi_j^{(k)} v = b_j$ for $j \in \mathcal{I}^{(k)}$. Since $\mathfrak{W}^{(k)} \subset \mathfrak{W}^{(k+1)}$ and since the minimizer is in $\mathfrak{W}^{(k)}$, the minimization over $v \in H_0^1(\Omega)$ can be reduced to $v \in \mathfrak{W}^{(k+1)}$ of the form $v = \sum_{i \in \mathcal{I}^{(k+1)}} c_i \psi_i^{(k+1)}$, which, after using (4.1) to expand the constraint $\int_{\Omega} \phi_j^{(k)} v = b_j$, corresponds to (4.15). \square

4.6. Multiresolution Gamblets. The interpolation and restriction operators are sufficient to derive a multigrid method for solving (1.1). To design a multiresolution algorithm we need to continue the analysis and identify basis functions for the subspaces $\mathfrak{W}^{(k)}$. For $k = 2, \dots, q$, let $\mathcal{J}^{(k)}$ be the finite set of k -tuples of the form $i = (i_1, \dots, i_k)$ with $1 \leq i_1 \leq m_0$, $1 \leq i_j \leq m_{(i_1, \dots, i_{j-1})}$ for $2 \leq j \leq k-1$, and $1 \leq i_k \leq m_{(i_1, \dots, i_{k-1})} - 1$, where the integers m_{\cdot} are the same as those defining the index tree \mathcal{I} . For a matrix M , write $\text{Im}(M)$ and $\text{Ker}(M)$ as its image and kernel.

LEMMA 4.8. For $k = 2, \dots, q$, let $W^{(k)}$ be a $\mathcal{J}^{(k)} \times \mathcal{I}^{(k)}$ matrix such that $\text{Im}(W^{(k),T}) = \text{Ker}(\pi^{(k-1,k)})$. It holds true that the elements $(\chi_i^{(k)})_{i \in \mathcal{J}^{(k)}} \in \mathfrak{W}^{(k)}$ defined as

$$(4.17) \quad \chi_i^{(k)} := \sum_{j \in \mathcal{I}^{(k)}} W_{i,j}^{(k)} \psi_j^{(k)}$$

form a basis of $\mathfrak{W}^{(k)}$.

Proof. Since $\mathcal{L}\mathfrak{W}^{(k-1)} = \text{span}\{\phi_i^{(k-1)} \mid i \in \mathcal{I}^{(k-1)}\}$, $w \in \mathfrak{W}^{(k)}$ belongs to $\mathfrak{W}^{(k)}$ if and only if $\int_{\Omega} \phi_j^{(k-1)} w = 0$ for all $j \in \mathcal{I}^{(k-1)}$, which, taking $w = \chi_i^{(k)}$ and using (4.1), translates to $(\pi^{(k-1,k)}W^{(k),T})_{j,i} = 0$. Writing $|\mathcal{J}^{(k)}|$ as the number of elements of $\mathcal{J}^{(k)}$ (which is equal to the dimension of $\mathfrak{W}^{(k)}$), observe that $|\mathcal{J}^{(k)}| = |I^{(k)}| - |I^{(k-1)}|$. Therefore, $\text{Im}(W^{(k),T}) = \text{Ker}(\pi^{(k-1,k)})$ also implies that the $|\mathcal{J}^{(k)}|$ elements $\chi_i^{(k)}$ are linearly independent and, therefore, form a basis of $\mathfrak{W}^{(k)}$. \square

Remark 7. Observe that since $0 = \langle \psi_i^{(k-1)}, \chi_j^{(k)} \rangle_a = (R^{(k-1,k)}A^{(k)}W^{(k),T})_{i,j}$, it also holds true that $\text{Im}(W^{(k),T}) = \text{Ker}(R^{(k-1,k)}A^{(k)})$ and $\text{Im}(A^{(k)}W^{(k),T}) = \text{Ker}(R^{(k-1,k)})$.

From now on we choose, for each $k \in \{2, \dots, q\}$, a $\mathcal{J}^{(k)} \times \mathcal{I}^{(k)}$ matrix $W^{(k)}$ as in Lemma 4.8. This choice is not unique and to enable fast multiplication by $W^{(k)}$ (or its transpose) we require that for $(j, i) \in \mathcal{J}^{(k)} \times \mathcal{I}^{(k)}$, $W_{j,i}^{(k)} = 0$ if $j^{(k-1)} \neq i^{(k-1)}$. Therefore, the construction of $W^{(k)}$ requires, for each $s \in \mathcal{I}^{(k-1)}$, the specification of a number $m_s - 1$ of m_s -dimensional vectors $W_{(s,1),(s,\cdot)}^{(k)}, \dots, W_{(s,m_s-1),(s,\cdot)}^{(k)}$ that are linearly independent and orthogonal to the m_s -dimensional vector $(1, 1, \dots, 1, 1)$. We propose two simple constructions.

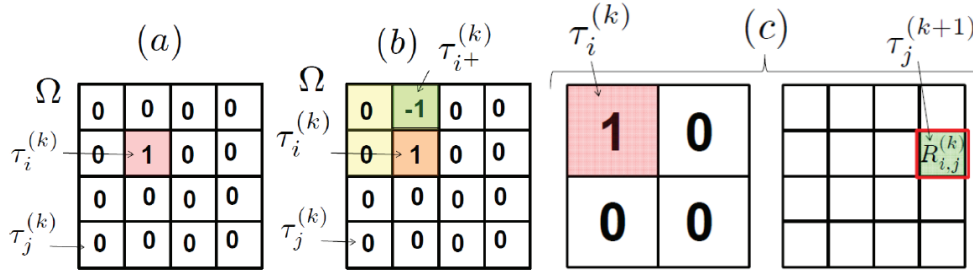


Fig. 1 If $(\tau_s^{(k)}, s \in \mathcal{I}^{(k)})$ is a nested rectangular partition of Ω , then (a) $\psi_i^{(k)}$ is Player II's best bet on the value of the solution u of (1.1) given $\int_{\tau_j^{(k)}} u = \delta_{i,j}$ for $j \in \mathcal{I}^{(k)}$, (b) $\chi_i^{(k)}$ is Player II's best bet on u given $\int_{\tau_j^{(k)}} u = \delta_{i,j} - \delta_{i^+,j}$ for $j \in \mathcal{I}^{(k)}$, and (c) $R_{i,j}^{(k,k+1)}$ is Player II's best bet on $\int_{\tau_j^{(k+1)}} u$ given $\int_{\tau_j^{(k)}} u = \delta_{i,j}$ for $j \in \mathcal{I}^{(k)}$.

CONSTRUCTION 4.9. For $k \in \{2, \dots, q\}$, choose $W^{(k)}$ such that (1) $W_{j,i}^{(k)} = 0$ for $(j, i) \in \mathcal{J}^{(k)} \times \mathcal{I}^{(k)}$ with $j^{(k-1)} \neq i^{(k-1)}$ and (2) for $s \in \mathcal{I}^{(k-1)}$, $t \in \{1, \dots, m_s - 1\}$, and $t' \in \{1, \dots, m_s\}$, $W_{(s,t),(s,t')}^{(k)} = \delta_{t,t'} - \delta_{t+1,t'}$.

For $k \in \{2, \dots, q\}$ and $i = (i_1, \dots, i_{k-1}, i_k) \in \mathcal{J}^{(k)}$, define $i^+ := (i_1, \dots, i_{k-1}, i_k + 1)$ and observe that under Construction 4.9,

$$(4.18) \quad \chi_i^{(k)} = \psi_i^{(k)} - \psi_{i^+}^{(k)},$$

whose game-theoretic interpretation is provided in Figure 1.

For the second construction we need the following lemma whose proof is trivial.

LEMMA 4.10. Let $U^{(n)}$ be the sequence of $n \times n$ matrices defined (1) for $n = 2$ by $U_{1,\cdot}^{(2)} = (1, -1)$ and $U_{2,\cdot}^{(2)} = (1, 1)$ and (2) iteratively for $n \geq 2$ by $U_{i,j}^{(n+1)} = U_{i,j}^{(n)}$ for $1 \leq i, j \leq n$, $U_{n+1,j}^{(n+1)} = 1$ for $1 \leq j \leq n + 1$, $U_{i,n+1}^{(n+1)} = 0$ for $1 \leq i \leq n - 1$, and $U_{n,n+1}^{(n+1)} = -n$. Then, for $n \geq 2$, the rows of $U^{(n)}$ are orthogonal, $U_{n,j}^{(n)} = 1$ for $1 \leq j \leq n$, and we write $\bar{U}^{(n)}$ for the corresponding orthonormal matrix obtained by renormalizing the rows of $U^{(n)}$.

CONSTRUCTION 4.11. For $k \in \{2, \dots, q\}$, choose $W^{(k)}$ such that (1) $W_{j,i}^{(k)} = 0$ for $(j, i) \in \mathcal{J}^{(k)} \times \mathcal{I}^{(k)}$ with $j^{(k-1)} \neq i^{(k-1)}$ and (2) for $s \in \mathcal{I}^{(k-1)}$ and $t \in \{1, \dots, m_s - 1\}$ and $t' \in \{1, \dots, m_s\}$, $W_{(s,t),(s,t')}^{(k)} = \bar{U}_{t,t'}^{(m_s)}$ (where $\bar{U}^{(m_s)}$ is defined in Lemma 4.10).

Observe that under Construction 4.11, (1) the complexity of constructing $W^{(k)}$ is $|\mathcal{I}^{(k-1)}| \times m_s^2$ and (2) $W^{(k)}W^{(k),T} = J^{(k)}$, where $J^{(k)}$ is the $\mathcal{J}^{(k)} \times \mathcal{J}^{(k)}$ identity matrix.

4.7. Multiresolution Operator Inversion. We will now use the basis functions $\psi_i^{(1)}$ and $\chi_i^{(k)}$ to perform the multiresolution inversion of (1.1). Let $B^{(k)}$ be the $\mathcal{J}^{(k)} \times \mathcal{J}^{(k)}$ (stiffness) matrix $B_{i,j}^{(k)} = \langle \chi_i^{(k)}, \chi_j^{(k)} \rangle_a$ and observe that

$$(4.19) \quad B^{(k)} = W^{(k)}A^{(k)}W^{(k),T}.$$

Observe that $B^{(k)}$ is positive symmetric definite and write $B^{(k),-1}$ as its inverse. Let

$\bar{\pi}^{(k,k+1)}$ be the $\mathcal{I}^{(k)} \times \mathcal{I}^{(k+1)}$ matrix defined by

$$(4.20) \quad \bar{\pi}_{i,j}^{(k,k+1)} = \pi_{i,j}^{(k,k+1)} / (\pi^{(k,k+1)} \pi^{(k+1,k)})_{i,i}.$$

Using the notations of Definition 4.1 note that $(\pi^{(k,k+1)} \pi^{(k+1,k)})_{i,i} = m_i$. Let $D^{(k,k-1)}$ be the $\mathcal{J}^{(k)} \times \mathcal{I}^{(k-1)}$ matrix defined as

$$(4.21) \quad D^{(k,k-1)} := -B^{(k),-1} W^{(k)} A^{(k)} \bar{\pi}^{(k,k-1)}$$

and write $D^{(k-1,k)} := D^{(k,k-1),T}$ as its transpose.

THEOREM 4.12. *It holds true that for $k \in \{1, \dots, q-1\}$ and $i \in \mathcal{I}^{(k)}$,*

$$(4.22) \quad \psi_i^{(k)} = \sum_{l \in \mathcal{I}^{(k+1)}} \bar{\pi}_{i,l}^{(k,k+1)} \psi_l^{(k+1)} + \sum_{j \in \mathcal{J}^{(k+1)}} D_{i,j}^{(k,k+1)} \chi_j^{(k+1)}.$$

In particular,

$$(4.23) \quad R^{(k,k+1)} = \bar{\pi}^{(k,k+1)} + D^{(k,k+1)} W^{(k+1)}.$$

Proof. For $s \in \mathcal{I}^{(k)}$, write $\bar{\psi}_s^{(k)} := \sum_{l \in \mathcal{I}^{(k+1)}} \bar{\pi}_{s,l}^{(k,k+1)} \psi_l^{(k+1)}$ and $\bar{\mathfrak{Y}}^{(k)} := \text{span}\{\bar{\psi}_s^{(k)} \mid s \in \mathcal{I}^{(k)}\}$. Let $x \in \mathbb{R}^{\mathcal{I}^{(k)}}$, $y \in \mathbb{R}^{\mathcal{J}^{(k+1)}}$, and

$$(4.24) \quad \psi = \sum_{s \in \mathcal{I}^{(k)}} x_s \bar{\psi}_s^{(k)} + \sum_{j \in \mathcal{J}^{(k+1)}} y_j \chi_j^{(k+1)}.$$

If $\psi = 0$, then integrating ψ against $\phi_i^{(k)}$ for $i \in \mathcal{I}^{(k)}$ (and observing that $\int_{\Omega} \phi_i^{(k)} \bar{\psi}_s^{(k)} = \delta_{i,s}$) implies $x = 0$ and $y = 0$. Therefore, the elements $\bar{\psi}_s^{(k)}, \chi_j^{(k+1)}$ form a basis for $\bar{\mathfrak{Y}}^{(k)} + \mathfrak{W}^{(k+1)}$. Observing that $\dim(\mathfrak{Y}^{(k+1)}) = \dim(\bar{\mathfrak{Y}}^{(k)}) + \dim(\mathfrak{W}^{(k+1)})$, we deduce that $\mathfrak{Y}^{(k+1)} = \bar{\mathfrak{Y}}^{(k)} + \mathfrak{W}^{(k+1)}$. Therefore, since $\mathfrak{Y}^{(k)} \subset \mathfrak{Y}^{(k+1)}$, $\psi_i^{(k)}$ can be decomposed as in (4.24). The constraints $\int_{\Omega} \phi_s^{(k)} \psi_i^{(k)} = \delta_{i,s}$ lead to $x_s = \delta_{i,s}$. The orthogonality between ψ and $\mathfrak{W}^{(k+1)}$ leads to the equations $\langle \psi, \chi_j^{(k+1)} \rangle_a = 0$ for $j \in \mathcal{J}^{(k+1)}$, i.e., $\sum_{l \in \mathcal{I}^{(k+1)}} \bar{\pi}_{i,l}^{(k,k+1)} \langle \psi_l^{(k+1)}, \chi_j^{(k+1)} \rangle_a + \sum_{j' \in \mathcal{J}^{(k+1)}} y_{j'} \langle \chi_{j'}^{(k+1)}, \chi_j^{(k+1)} \rangle_a = 0$, which translates into $W^{(k+1)} A^{(k+1)} \bar{\pi}_{i,i}^{(k+1,k)} + B^{(k+1)} y$, that is, (4.22). Plugging (4.17) into (4.22) and comparing with (4.11) leads to (4.23). \square

Let g be the right-hand side of (1.1). For $k \in \{1, \dots, q\}$, let $g^{(k)}$ be the $|\mathcal{I}^{(k)}|$ -dimensional vector defined by $g_i^{(k)} = \int_{\Omega} \psi_i^{(k)} g$ for $i \in \mathcal{I}^{(k)}$. Observe that $g^{(k)}$ can be computed iteratively using

$$(4.25) \quad g^{(k)} = R^{(k,k+1)} g^{(k+1)}.$$

For $k \in \{2, \dots, q\}$, let $w^{(k)}$ be the $|\mathcal{J}^{(k)}|$ -dimensional vector defined as the solution of

$$(4.26) \quad B^{(k)} w^{(k)} = W^{(k)} g^{(k)}.$$

Furthermore, let $U^{(1)}$ be the $|\mathcal{I}^{(1)}|$ -dimensional vector defined as the solution of

$$(4.27) \quad A^{(1)} U^{(1)} = g^{(1)}.$$

According to following theorem, which is a direct consequence of Theorem 4.5, the solution of (1.1) can be computed at any scale by solving the decoupled linear systems (4.26) and (4.27).

THEOREM 4.13. For $k \in \{2, \dots, q\}$, let $u^{(k)}$ be the finite element solution of (1.1) in $\mathfrak{V}^{(k)}$. It holds true that $u^{(k)} - u^{(k-1)} = \sum_{i \in \mathcal{I}^{(k)}} w_i^{(k)} \chi_i^{(k)}$ and, in particular,

$$(4.28) \quad u^{(k)} = \sum_{i \in \mathcal{I}^{(1)}} U_i^{(1)} \psi_i^{(1)} + \sum_{k'=2}^k \sum_{i \in \mathcal{I}^{(k')}} w_i^{(k')} \chi_i^{(k')}.$$

4.8. Uniformly Bounded Condition Numbers across Subscales/Subbands.

Taking $q = \infty$ in Theorem 4.5, the construction of the basis elements $\psi_i^{(k)}$ leads to the multiresolution orthogonal decomposition

$$(4.29) \quad H_0^1(\Omega) = \mathfrak{V}^{(1)} \bigoplus_a^{\infty} \mathfrak{W}^{(i)}.$$

In that sense, the basis elements $\psi_i^{(k)}$ and $\chi_i^{(k)}$ could be seen as a generalization of wavelets to the orthogonal decomposition of $H_0^1(\Omega)$ (rather than $L^2(\Omega)$) adapted to the solution space of the PDE (1.1). We will now show that this orthogonal decomposition induces a subscale decomposition of the operator $-\text{div}(a\nabla)$ into layered subbands of increasing frequencies. Moreover, the condition number of the operator $-\text{div}(a\nabla)$ restricted to each subspace $\mathfrak{W}^{(k)}$ will be shown to be uniformly bounded if H_{k-1}/H_k is uniformly bounded (e.g., if H_k is a geometric sequence). Write $H_0 := 1$ and let δ be defined as in Construction 4.2.

THEOREM 4.14. If $k \in \{1, \dots, q\}$ and $v \in \mathfrak{V}^{(k)}$, then

$$(4.30) \quad \frac{\delta^{1+d/2}}{\sqrt{\lambda_{\max}(a)} 2^{5/2+d/2}} H_k \leq \frac{\|v\|_a}{\|\text{div}(a\nabla v)\|_{L^2(\Omega)}}.$$

If $k = 1$ and $v \in \mathfrak{V}^{(1)}$, or $k \in \{2, \dots, q\}$ and $v \in \mathfrak{W}^{(k)}$, then

$$(4.31) \quad \frac{\|v\|_a}{\|\text{div}(a\nabla v)\|_{L^2(\Omega)}} \leq \frac{1}{\sqrt{\lambda_{\min}(a)}} H_{k-1}.$$

Proof. (4.30) is a direct consequence of Lemma 3.10. For $k = 1$ (4.31) is a simple consequence of Poincaré’s inequality. Let $k \in \{2, \dots, q\}$. $\mathfrak{V}^{(k)} = \mathfrak{V}^{(k-1)} \oplus_a \mathfrak{W}^{(k)}$ and Theorem 4.3 imply $\sup_{v \in \mathfrak{W}^{(k)}} \frac{\|v\|_a}{\|\text{div}(a\nabla v)\|_{L^2(\Omega)}} \leq \sup_{v \in \mathfrak{V}^{(k)}} \inf_{v' \in \mathfrak{V}^{(k-1)}} \frac{\|v-v'\|_a}{\|\text{div}(a\nabla v)\|_{L^2(\Omega)}} \leq \frac{2}{\pi \sqrt{\lambda_{\min}(a)}} H_{k-1}$. \square

Write $|c|$ as the Euclidean norm of c and for $k \in \{1, \dots, q\}$ let

$$(4.32) \quad \underline{\gamma}_k := \inf_{c \in \mathbb{R}^{\mathcal{I}^{(k)}}} \frac{\|\sum_{i \in \mathcal{I}^{(k)}} c_i \phi_i^{(k)}\|_{L^2(\Omega)}^2}{|c|^2} \quad \text{and} \quad \bar{\gamma}_k := \sup_{c \in \mathbb{R}^{\mathcal{I}^{(k)}}} \frac{\|\sum_{i \in \mathcal{I}^{(k)}} c_i \phi_i^{(k)}\|_{L^2(\Omega)}^2}{|c|^2}.$$

Write $|\tau|$ as the volume of a set τ and note that $\bar{\gamma}_k \leq \max_{i \in \mathcal{I}^{(k)}} |\tau_i^{(k)}|$ and $\underline{\gamma}_k \geq \min_{i \in \mathcal{I}^{(k)}} |\tau_i^{(k)}|$; therefore, $\bar{\gamma}_k/\underline{\gamma}_k \leq \delta^{-d}$.

For a given matrix M , write $\text{Cond}(M) := \sqrt{\lambda_{\max}(M^T M)}/\sqrt{\lambda_{\min}(M^T M)}$ as its condition number.

THEOREM 4.15. It holds true that

$$(4.33) \quad \text{Cond}(A^{(1)}) \leq \frac{1}{H_1^2} \frac{\lambda_{\max}(a) 2^{5+d}}{\lambda_{\min}(a) \delta^{2+2d}}$$

and, for $k \in \{2, \dots, q\}$,

$$(4.34) \quad \text{Cond}(B^{(k)}) \leq \left(\frac{H_{k-1}}{H_k}\right)^2 \left(\frac{\lambda_{\max}(a)}{\lambda_{\min}(a)}\right)^2 \frac{2^{11+2d}}{\delta^{4+7d}\pi^2} \text{Cond}(W^{(k)}W^{(k),T}).$$

Furthermore, $\text{Cond}(W^{(k)}W^{(k),T}) = 1$ under Construction 4.11 and $\text{Cond}(W^{(k)}W^{(k),T}) \leq 2(H_{k-1}/(\delta H_k))^{2d}$ under Construction 4.9.

Proof. Let $k \in \{1, \dots, q\}$ and $c \in \mathbb{R}^{\mathcal{I}^{(k)}}$. Write $v = \sum_{i \in \mathcal{I}^{(k)}} c_i \psi_i^{(k)}$. Observing that $\|v\|_a^2 = c^T A^{(k)} c$ and $\|\text{div}(a \nabla v)\|_{L^2(\Omega)}^2 = \|\sum_{i \in \mathcal{I}^{(k)}} (A^{(k)} c)_i \phi_i^{(k)}\|_{L^2(\Omega)}^2 \geq \gamma_k |A^{(k)} c|^2$, (4.30) implies that $\gamma_k H_k^2 \delta^{2+d} / (\lambda_{\max}(a) 2^{5+d}) \leq c^T A^{(k)} c / |A^{(k)} c|^2$, which, after taking the minimum in c , leads to (for $k \geq 1$)

$$(4.35) \quad \lambda_{\max}(A^{(k)}) \leq \lambda_{\max}(a) 2^{5+d} / (H_k^2 \delta^{2+d} \gamma_k),$$

and for $k \geq 2$ (using (4.19))

$$(4.36) \quad \lambda_{\max}(B^{(k)}) \leq \lambda_{\max}(W^{(k)}W^{(k),T}) \lambda_{\max}(a) 2^{5+d} / (H_k^2 \delta^{2+d} \gamma_k).$$

Similarly, for $k = 1$ (4.31) leads to $\lambda_{\min}(A^{(1)}) \geq \lambda_{\min}(a) / \bar{\gamma}_1$. Now let us consider $k \in \{2, \dots, q\}$ and $c \in \mathbb{R}^{\mathcal{J}^{(k)}}$. Write $w = \sum_{i \in \mathcal{J}^{(k)}, j \in \mathcal{I}^{(k)}} c_i W_{i,j}^{(k)} \psi_j^{(k)}$. (4.17) and (4.19) imply that $\|w\|_a^2 = c^T B^{(k)} c$ and, using (4.32), $\|\text{div}(a \nabla w)\|_{L^2(\Omega)} = \|\sum_{i \in \mathcal{J}^{(k)}, j \in \mathcal{I}^{(k)}} (A^{(k)} W^{(k),T} c)_j \phi_j^{(k)}\|_{L^2(\Omega)} \leq \bar{\gamma}_k |A^{(k)} W^{(k),T} c|^2$. Observing that $w \in \mathfrak{W}^{(k)}$, (4.31) implies that $\frac{c^T B^{(k)} c}{c^T W^{(k)} (A^{(k)})^2 W^{(k),T} c} \leq \bar{\gamma}_k \frac{1}{\lambda_{\min}(a)} H_{k-1}^2$. Taking $c = B^{(k),-1} y$ for $y \in \mathbb{R}^{\mathcal{J}^{(k)}}$, we deduce that $\frac{y^T B^{(k),-1} y}{|A^{(k)} W^{(k),T} B^{(k),-1} y|^2} \leq \bar{\gamma}_k \frac{1}{\lambda_{\min}(a)} H_{k-1}^2$. Writing $N^{(k)} = -A^{(k)} W^{(k),T} B^{(k),-1}$, we have obtained that

$$(4.37) \quad \lambda_{\min}(a) / (H_{k-1}^2 \bar{\gamma}_k \lambda_{\max}(N^{(k),T} N^{(k)})) \leq \lambda_{\min}(B^{(k)}).$$

For $k \in \{2, \dots, q\}$ let $P^{(k)} := \pi^{(k,k-1)} R^{(k-1,k)}$. Using $R^{(k-1,k)} = A^{(k-1)} \pi^{(k-1,k)} \Theta^{(k)}$ and $\pi^{(k-1,k)} \Theta^{(k)} \pi^{(k,k-1)} = \Theta^{(k-1)}$ (Theorem 4.7), we obtain that $(P^{(k)})^2 = P^{(k)}$, i.e., $P^{(k)}$ is a projection. Write $\|P^{(k)}\|_{\text{Ker}(\pi^{(k-1,k)})} := \sup_{x \in \text{Ker}(\pi^{(k-1,k)})} |P^{(k)} x| / |x|$.

LEMMA 4.16. *It holds true that, for $k \in \{2, \dots, q\}$,*

$$(4.38) \quad \lambda_{\max}(N^{(k),T} N^{(k)}) \leq \frac{1 + \|P^{(k)}\|_{\text{Ker}(\pi^{(k-1,k)})}^2}{\lambda_{\min}(W^{(k)}W^{(k),T})}.$$

Proof. Since $\text{Im}(W^{(k),T})$ and $\text{Im}(\pi^{(k,k-1)})$ are orthogonal and $\dim(\mathbb{R}^{\mathcal{I}^{(k)}}) = \dim(\text{Im}(W^{(k),T})) + \dim(\text{Im}(\pi^{(k,k-1)}))$, for $x \in \mathbb{R}^{\mathcal{I}^{(k)}}$ there exists a unique $y \in \mathbb{R}^{\mathcal{J}^{(k)}}$ and $z \in \mathbb{R}^{\mathcal{I}^{(k-1)}}$ such that $x = W^{(k),T} y + \pi^{(k,k-1)} z$ and $|x|^2 = |W^{(k),T} y|^2 + |\pi^{(k,k-1)} z|^2$. Observe that $W^{(k)} x = W^{(k)} W^{(k),T} y$ (since $W^{(k)} \pi^{(k,k-1)} = 0$) and $R^{(k-1,k)} x = R^{(k-1,k)} W^{(k),T} y + z$ (since $R^{(k-1,k)} \pi^{(k,k-1)} = I^{(k-1)}$ from Theorem 4.7). Therefore, $|x|^2 = |W^{(k),T} y|^2 + |P^{(k)}(x - W^{(k),T} y)|^2$ with $y = (W^{(k)} W^{(k),T})^{-1} W^{(k)} x$. Let $v \in \mathbb{R}^{\mathcal{J}^{(k)}}$. Taking $x = A^{(k)} W^{(k),T} v$ and observing that $P^{(k)} x = 0$ (since $R^{(k-1,k)} A^{(k)} W^{(k),T} = 0$ from the $\langle \cdot, \cdot \rangle_a$ -orthogonality between $\mathfrak{V}^{(k-1)}$ and $\mathfrak{W}^{(k)}$) leads to $|A^{(k)} W^{(k),T} v|^2 = |W^{(k),T} y|^2 + |P^{(k)} W^{(k),T} y|^2$ with $y = (W^{(k)} W^{(k),T})^{-1} B^{(k)} v$. Therefore, $|A^{(k)} W^{(k),T} v|^2 \leq (1 + \|P^{(k)}\|_{\text{Ker}(\pi^{(k-1,k)})}^2) \frac{|B^{(k)} v|^2}{\lambda_{\min}(W^{(k)} W^{(k),T})}$, which concludes the proof after taking $v = B^{(k),-1} v'$ and maximizing the left-hand side over $|v'| = 1$. \square

LEMMA 4.17. *Writing $\|M\|_2 := \sup_x |Mx|/x$ as the spectral norm, we have*

$$(4.39) \quad \|P^{(k)}\|_{\text{Ker}(\pi^{(k-1,k)})}^2 \leq \|\pi^{(k,k-1)}A^{(k-1)}\pi^{(k-1,k)}\|_2 \sup_{x \in \text{Ker}(\pi^{(k-1,k)})} \frac{x^T \Theta^{(k)} x}{x^T x}.$$

Proof. Let $x \in \text{Ker}(\pi^{(k-1,k)})$. Using $P^{(k)} = \pi^{(k,k-1)}A^{(k-1)}\pi^{(k-1,k)}\Theta^{(k)}$ we obtain that $|P^{(k)}x| = \|\pi^{(k,k-1)}A^{(k-1)}\pi^{(k-1,k)}(\Theta^{(k)})^{\frac{1}{2}}\|_2 |(\Theta^{(k)})^{\frac{1}{2}}x|$. Observing that for $M = \pi^{(k-1,k)}(\Theta^{(k)})^{\frac{1}{2}}$ we have $MM^T = \Theta^{(k-1)}$ and for $N = \pi^{(k,k-1)}A^{(k-1)}\pi^{(k-1,k)}(\Theta^{(k)})^{\frac{1}{2}}$ we have $\lambda_{\max}(N^TN) = \lambda_{\max}(NN^T)$, we deduce that $\|\pi^{(k,k-1)}A^{(k-1)}\pi^{(k-1,k)}(\Theta^{(k)})^{\frac{1}{2}}\|_2^2 = \|\pi^{(k,k-1)}A^{(k-1)}\pi^{(k-1,k)}\|_2^2$ and conclude by taking the supremum over $x \in \text{Ker}(\pi^{(k-1,k)})$. \square

LEMMA 4.18. *It holds true that*

$$(4.40) \quad \sup_{x \in \text{Ker}(\pi^{(k-1,k)})} \frac{x^T \Theta^{(k)} x}{x^T x} \leq H_{k-1}^2 \frac{\bar{\gamma}_k^2}{\gamma_k \pi^2 \lambda_{\min}(a)}.$$

Proof. Let $y \in \mathbb{R}^{\mathcal{I}^{(k)}}$ and $\alpha \in \mathbb{R}$. Let $x = \alpha W^{(k),T}y$. Write $\phi = \sum_{i \in \mathcal{I}^{(k)}} x_i \phi_i^{(k)}$ and $\psi = (-\text{div}(a\nabla \cdot))^{-1}\phi$. Observe that $\|\psi\|_a^2 = x^T \Theta^{(k)} x \geq \alpha y^T W^{(k)} \Theta^{(k)} W^{(k),T} y$. Using $\int_{\Omega} \phi_i^{(k)} \phi_l^{(k)} = 0$ for $i \neq l$ and selecting $\alpha = \|\phi_i^{(k)}\|_{L^2(\Omega)}^{-2}$ (assuming, without loss of generality, that $\|\phi_i^{(k)}\|_{L^2(\Omega)}^2 = |\tau_i^{(k)}|$ is constant in i , for the general case, rescale each $\phi_i^{(k)}$ by a multiplicative constant), we obtain that for $j \in \mathcal{I}^{(k-1)}$, $\int_{\Omega} \phi \phi_j^{(k-1)} = \sum_{i \in \mathcal{I}^{(k)}} x_i \|\phi_i^{(k)}\|_{L^2(\Omega)}^2 \pi_{j,i}^{(k-1,k)} = (\pi^{(k-1,k)} W^{(k),T} y)_j = 0$. Therefore, since $\|\psi\|_a^2 = \int_{\Omega} \phi \psi$, we have for $\psi' \in \text{span}\{\phi_i^{(k-1)} \mid i \in \mathcal{I}^{(k-1)}\}$, $\|\psi\|_a^2 = \int_{\Omega} \phi(\psi - \psi') \leq \|\phi\|_{L^2(\Omega)} \|\psi - \psi'\|_{L^2(\Omega)}$. Choosing $\psi' = \sum_{i \in \mathcal{I}^{(k-1)}} \phi_i^{(k-1)} \int_{\Omega} \psi \phi_i^{(k-1)} / \|\phi_i^{(k-1)}\|_{L^2(\Omega)}^2$ we obtain (via the Poincaré and Cauchy–Schwarz inequalities as in the proof of Proposition 3.6) that $\|\psi - \psi'\|_{L^2(\Omega)} \leq H_{k-1} \|\psi\|_a / (\pi \sqrt{\lambda_{\min}(a)})$ and deduce $\|\psi\|_a \leq H_{k-1} \|\phi\|_{L^2(\Omega)} / (\pi \sqrt{\lambda_{\min}(a)})$. Observing that $\|\phi\|_{L^2(\Omega)}^2 \leq |x|^2 \bar{\gamma}_k$ and $\gamma_k \leq \alpha^{-1} \leq \bar{\gamma}_k$, we summarize and obtain that $y^T W^{(k)} \Theta^{(k)} W^{(k),T} y \leq H_{k-1}^2 |x|^2 \bar{\gamma}_k^2 / (\pi^2 \lambda_{\min}(a)) \leq H_{k-1}^2 |W^{(k),T} y|^2 \bar{\gamma}_k^2 / (\gamma_k \pi^2 \lambda_{\min}(a))$, which concludes the proof of the lemma (since $\text{Ker}(\pi^{(k-1,k)}) = \text{Im}(W^{(k),T})$). \square

Observing that $\|\pi^{(k,k-1)}A^{(k-1)}\pi^{(k-1,k)}\|_2 \leq \lambda_{\max}(\pi^{(k,k-1)}\pi^{(k-1,k)})\lambda_{\max}(A^{(k-1)})$ and using (4.35), we derive from Lemmas 4.17 and 4.18 that

$$(4.41) \quad \|P^{(k)}\|_{\text{Ker}(\pi^{(k-1,k)})}^2 \leq \lambda_{\max}(\pi^{(k,k-1)}\pi^{(k-1,k)}) \frac{\bar{\gamma}_k^2 2^{5+d} \lambda_{\max}(a)}{\gamma_k \gamma_{k-1} \delta^{2+d} \pi^2 \lambda_{\min}(a)}.$$

Observing that $\pi^{(k,k-1)}\pi^{(k-1,k)}$ is block diagonal and using the notations of Definition 4.1, we have $\lambda_{\max}(\pi^{(k,k-1)}\pi^{(k-1,k)}) = \max_{j \in \mathcal{I}^{(k-1)}} \sup_{x \in \mathbb{R}^{m_j}} |\sum_{i=1}^{m_j} x_i|^2 / |x|^2 = \max_{j \in \mathcal{I}^{(k-1)}} m_j$. Noting that a set $\tau_j^{(k-1)}$ can contain at most $(\max_{j \in \mathcal{I}^{(k-1)}} |\tau_j^{(k-1)}|) / (\min_{i \in \mathcal{I}^{(k)}} |\tau_i^{(k)}|)$ subsets $\tau_i^{(k)}$, we have

$$(4.42) \quad \max_{j \in \mathcal{I}^{(k-1)}} m_j \leq (H_{k-1} / (\delta H_k))^d$$

and conclude that $\lambda_{\max}(\pi^{(k,k-1)}\pi^{(k-1,k)}) \leq (H_{k-1} / (\delta H_k))^d$. Therefore, (4.37) and Lemma 4.16 imply, after simplification, that

$$(4.43) \quad \lambda_{\min}(B^{(k)}) \geq \frac{\lambda_{\min}(a)}{H_{k-1}^2 \bar{\gamma}_k} \lambda_{\min}(W^{(k)}W^{(k),T}) \frac{H_k^d \gamma_{k-1} \gamma_k \delta^{2+2d} \pi^2 \lambda_{\min}(a)}{H_{k-1}^d \bar{\gamma}_k^2 2^{6+d} \lambda_{\max}(a)}.$$

Recalling that $\bar{\gamma}_k/\gamma_k \leq \delta^{-d}$, using $\bar{\gamma}_k/\gamma_{k-1} \leq H_k^d/(H_{k-1}\delta)^d$, and summarizing, we conclude the proof of (4.33) and (4.34). Recall that under Construction 4.11 we have $W^{(k)}W^{(k),T} = J^{(k)}$, which implies $\text{Cond}(W^{(k)}W^{(k),T}) = 1$. Under Construction 4.9, $W^{(k)}W^{(k),T}$ is block diagonal with, for $j \in \mathcal{I}^{(k-1)}$, diagonal blocks corresponding to $(m_j - 1) \times (m_j - 1)$ matrices $M^{(m_j-1)}$ such that (1) for $n = 1$ and $x \in \mathbb{R}$, $x^T M^{(1)}x = 2x^2$, (2) for $n = 2$ and $x \in \mathbb{R}^2$, $x^T M^{(2)}x = x_1^2 + (x_2 - x_1)^2 + x_2^2$, and (3) for $n \geq 3$ and $x \in \mathbb{R}^n$, $x^T M^{(n)}x = x_1^2 + \sum_{i=1}^{n-2} (x_i - x_{i+1})^2 + x_n^2$. Note that for all $n \geq 1$, $\lambda_{\max}(M^{(n)}) \leq 3$. Furthermore, for $n \geq 3$ ($n \leq 2$ is trivial), introducing the variables $y_2 = x_2 - x_1, \dots, y_n = x_n - x_{n-1}$, we obtain that $x^T M^{(n)}x = x_1^2 + y_2^2 + \dots + y_n^2 + x_n^2$ and $|x|^2 = x_1^2 + (x_1 + y_2)^2 + \dots + (x_1 + y_2 + \dots + y_n)^2 \leq (x^T M^{(n)}x)n(n+1)/2$. Therefore, $\lambda_{\min}(M^{(n)}) \geq 2/(n(n+1))$. We conclude that under Construction 4.9, $\text{Cond}(W^{(k)}W^{(k),T}) \leq \max_{j \in \mathcal{I}^{(k-1)}} 3(m_j - 1)m_j/2$ and bound m_j as in (4.42). \square

4.9. Well-Conditioned Relaxation across Subscales. If H_k is a geometric sequence, or if H_{k-1}/H_k is uniformly bounded, then, by Theorem (4.15), the linear systems (4.26) and (4.27) entering into the calculation of the gamblets $\chi_i^{(k)}$ (and therefore $\psi_i^{(k)}$) and the subband/subscale solutions $u^{(1)}$ and $u^{(k+1)} - u^{(k)}$ have uniformly bounded condition numbers (in particular, these condition numbers are bounded independently from the mesh size/resolution and the regularity of $a(x)$). Therefore, these systems can be solved efficiently using iterative methods. One such method is the conjugate gradient (CG) method [57]. Recall [98] that the application of the CG method to a linear system $Ax = b$ (where A is an $n \times n$ symmetric positive definite matrix) with initial guess $x^{(0)}$ yields a sequence of approximations $x^{(l)}$ satisfying (writing $|e|_A^2 := e^T A e$) $|x - x^{(l)}|_A \leq 2 \left(\frac{\sqrt{\text{Cond}(A)} - 1}{\sqrt{\text{Cond}(A)} + 1} \right)^l |x - x^{(0)}|_A$, where $\text{Cond}(A) := \lambda_{\max}(A)/\lambda_{\min}(A)$. Also recall [98] that the maximum number of iterations required to reduce the error by a factor ϵ ($|x - x^{(l)}|_A \leq \epsilon |x - x^{(0)}|_A$) is bounded by $\frac{1}{2} \sqrt{\text{Cond}(A)} \ln \frac{2}{\epsilon}$ and has complexity (number of required arithmetic operations) $\mathcal{O}(\sqrt{\text{Cond}(A)}N_A)$ (writing N_A as the number of nonzero entries of A).

4.10. Hierarchical Localization and Error Propagation across Scales. Despite the fact that the multiresolution decomposition presented in this section leads to well-conditioned linear systems, the resulting matrices $B^{(k)}$ and $A^{(k)}$ are dense and, to achieve near-linear complexity in the resolution of (1.1), these matrices must be truncated by localizing the computation of the basis functions $\psi_i^{(k)}$ (and therefore $\chi_i^{(k)}$). The approximation error induced by these localization/truncation steps is controlled by the exponential decay of gamblets and the uniform bound on the condition numbers of the matrices $B^{(k)}$. To make this control explicit and derive a bound the size of the localization subdomains we need to quantify the propagation of truncation/localization errors across scales, and this is the purpose of this subsection.

For $k \in \{1, \dots, q\}$, $\rho \geq 1$, and $i \in \mathcal{I}^{(k)}$, we define (1) i^ρ as the subset of indices $j \in \mathcal{I}^{(k)}$ whose corresponding subdomains $\tau_j^{(k)}$ are at distance at most $H_k \rho$ from $\tau_i^{(k)}$ and (2) $S_\rho^i := \cup_{j \in i^\rho} \tau_j^{(k)}$. Let $\rho_1, \dots, \rho_q \geq 1$. For $k \in \{1, \dots, q-1\}$ and $i \in \mathcal{I}^{(k)}$, write $i^{\rho, k+1}$ as the subset of indices $j \in \mathcal{I}^{(k+1)}$ such that $j^{(k)} \in i^\rho$. For $i \in \mathcal{I}^{(q)}$, let $\mathfrak{W}_i^{(q+1), \text{loc}} := H_0^1(S_{\rho_q}^i)$. For $k \in \{1, \dots, q\}$ and $i \in \mathcal{I}^{(k)}$, let $\psi_i^{(k), \text{loc}}$ be the minimizer of

$$(4.44) \quad \text{Minimize } \|\psi\|_a \text{ subject to } \psi \in \mathfrak{W}_i^{(k+1), \text{loc}} \text{ and } \int_{\Omega} \psi \phi_j^{(k)} = \delta_{i,j} \text{ for } j \in i^{\rho_k},$$

where, for $k \in \{1, \dots, q-1\}$ and $i \in \mathcal{I}^{(k)}$, $\mathfrak{W}_i^{(k+1), \text{loc}}$ is defined (via induction) by $\mathfrak{W}_i^{(k+1), \text{loc}} := \text{span}\{\psi_j^{(k+1), \text{loc}} \mid j \in i^{\rho_k, k+1}\}$.

From now on we will assume that $H_k = H^k$ for some $H \in (0, 1)$ (or simply that H_k is uniformly bounded from below and above by H^k). To simplify the presentation, we will also, from now on, write C as any constant that depends only on $d, \Omega, \lambda_{\min}(a), \lambda_{\max}(a)$, or δ (e.g., $2C\lambda_{\max}(a)$ will still be written C). The following theorem allows us to control the localization error propagation across scales. For $k \in \{1, \dots, q\}$, let $A^{(k),\text{loc}}$ be the $\mathcal{I}^{(k)} \times \mathcal{I}^{(k)}$ matrix defined by $A_{i,j}^{(k),\text{loc}} := \langle \psi_i^{(k),\text{loc}}, \psi_j^{(k),\text{loc}} \rangle_a$ and let $\mathcal{E}(k)$ be the (localization) error $\mathcal{E}(k) := (\sum_{i \in \mathcal{I}^{(k)}} \|\psi_i^{(k)} - \psi_i^{(k),\text{loc}}\|_a^2)^{\frac{1}{2}}$.

THEOREM 4.19. *For $k \in \{1, \dots, q-1\}$, we have*

$$(4.45) \quad \mathcal{E}(k) \leq CH^{-\frac{d}{2}} \mathcal{E}(k+1) + Ce^{-\rho_k/C} H^{\frac{d}{2} - (k+1)(d+1)}.$$

Proof. We will need the following lemma summarizing and simplifying some results obtained in Theorem 4.14 when $H_k = H^k$.

LEMMA 4.20. *Let $H_k = H^k$ and $W^{(k)}$ be as in Construction 4.9 or Construction 4.11. It holds true that for $k \in \{q, \dots, 2\}$, (1) $\|W^{(k)}\|_2 \leq \sqrt{3}$, (2) $1/\lambda_{\min}(W^{(k)}W^{(k),T}) \leq CH^{-2d}$, (3) $\|\bar{\pi}^{(k-1,k)}\|_2 \leq CH^{\frac{d}{2}}$, (4) $\|\pi^{(k-1,k)}\|_2 \leq CH^{-d/2}$, (5) $\|R^{(k-1,k)}\|_2 \leq CH^{-d/2}$, (6) $\text{Cond}(B^{(k)}) \leq CH^{-2-2d}$, (7) $\lambda_{\max}(B^{(k)}) \leq CH^{-k(2+d)}$, and (8) $1/\lambda_{\min}(B^{(k)}) \leq CH^{k(2+d)-2-2d}$. Furthermore, (9) $\text{Cond}(A^{(1)}) \leq CH^{-2}$, (10) $1/\lambda_{\min}(A^{(1)}) \leq CH^d$, and, for $k \in \{1, \dots, q\}$, (11) $\lambda_{\max}(A^{(k)}) \leq CH^{-k(2+d)}$.*

Proof. From the proof of Theorem 4.15 we have (1) and $1/\lambda_{\min}(W^{(k)}W^{(k),T}) \leq \max_{i \in \mathcal{I}^{(k-1)}} (m_i - 1)m_i/2$, which implies (2). For (3), noting that $\bar{\pi}_{i,j}^{(k-1,k)} = 0$ if $j^{(k-1)} \neq i$ and $\bar{\pi}_{i,i}^{(k-1,k)} = 1/m_i$, otherwise, we have $\|\bar{\pi}^{(k-1,k)}\|_2 = \max_{i \in \mathcal{I}^{(k-1)}} 1/\sqrt{m_i} \leq CH^{\frac{d}{2}}$. (4) follows from $\lambda_{\max}(\pi^{(k,k-1)}\pi^{(k-1,k)}) = \max_{i \in \mathcal{I}^{(k-1)}} m_i \leq CH^{-d}$. Let us now prove (5). Using (4.23), (4.21), and defining $N^{(k)} = -A^{(k)}W^{(k),T}B^{(k),-1}$ as in Lemma 4.16, we have $R^{(k-1,k)} = \bar{\pi}^{(k-1,k)} + \bar{\pi}^{(k-1,k)}N^{(k)}W^{(k)}$, which leads to $\|R^{(k-1,k)}\|_2 \leq \|\bar{\pi}^{(k-1,k)}\|_2(1 + \|N^{(k)}\|_2\|W^{(k)}\|_2)$. Using Lemma 4.16 and (4.41) we obtain that $\lambda_{\max}(N^{(k),T}N^{(k)}) \leq (1 + C\lambda_{\max}(\pi^{(k,k-1)}\pi^{(k-1,k)})H^d)/\lambda_{\min}(W^{(k)}W^{(k),T})$ and, therefore, $\|N^{(k)}\|_2 \leq CH^{-d}$. Summarizing, we have obtained (5). (6), (7), (8), and (11) follow from Theorem 4.14 and, in particular, (4.36), (4.43), and (4.35). See (4.33) and the proof of Theorem 4.15 for (9) and (10). \square

We will also need the following lemma.

LEMMA 4.21. *Let $k \in \{1, \dots, q-1\}$ and let R be the $\mathcal{I}^{(k)} \times \mathcal{I}^{(k+1)}$ matrix defined by $R_{i,j} = 0$ for $j \in i^{\rho_k, k+1}$ and $R_{i,j} = R_{i,j}^{(k,k+1)}$ for $j \in \mathcal{I}^{(k+1)}/i^{\rho_k, k+1}$. It holds true that $\|R\|_2 \leq CH^{d/2}e^{-\rho_k/C}$.*

Proof. Observe that $\|R\|_2^2 \leq |\mathcal{I}^{(k)}| \max_{i \in \mathcal{I}^{(k)}} \sum_{j \in \mathcal{I}^{(k+1)}/i^{\rho_k, k+1}} |R_{i,j}^{(k,k+1)}|^2$ with $|\mathcal{I}^{(k)}| \leq CH_k^{-d}$. Let $i \in \mathcal{I}^{(k)}$. Using Theorem 4.6 and the Cauchy-Schwarz inequality we have $|R_{i,j}^{(k,k+1)}| \leq \|\psi_i^{(k)}\|_{L^2(\tau_j^{(k+1)})} \|\phi_j^{(k+1)}\|_{L^2(\tau_j^{(k+1)})}$. Therefore, $\sum_{j \in \mathcal{I}^{(k+1)}/i^{\rho_k, k+1}} |R_{i,j}^{(k,k+1)}|^2 \leq CH_{k+1}^d \sum_{j \in \mathcal{I}^{(k+1)}/i^{\rho_k, k+1}} \|\psi_i^{(k)}\|_{L^2(\tau_j^{(k+1)})}^2$. Observe that $\sum_{j \in \mathcal{I}^{(k+1)}/i^{\rho_k, k+1}} \|\psi_i^{(k)}\|_{L^2(\tau_j^{(k+1)})}^2 = \sum_{s \in \mathcal{I}^{(k)}/i^{\rho_k}} \|\psi_i^{(k)}\|_{L^2(\tau_s^{(k)})}^2$. Since $\int_{\tau_s^{(k)}} \psi_i^{(k)} = 0$ for $s \neq i$ we obtain from Poincaré's inequality that $\|\psi_i^{(k)}\|_{L^2(\tau_s)} \leq C\|\nabla\psi_i^{(k)}\|_{L^2(\tau_s^{(k)})}H_k$. Therefore, $\sum_{j \in \mathcal{I}^{(k+1)}/i^{\rho_k, k+1}} |R_{i,j}^{(k,k+1)}|^2 \leq CH_{k+1}^d H_k^2 \sum_{s \in \mathcal{I}^{(k)}/i^{\rho_k}} \|\nabla\psi_i^{(k)}\|_{L^2(\tau_s^{(k)})}^2$. Using Theorem 3.9 we obtain that $\sum_{s \in \mathcal{I}^{(k)}/i^{\rho_k}} \|\nabla\psi_i^{(k)}\|_{L^2(\tau_s^{(k)})}^2 \leq Ce^{-C^{-1}\rho_k} \|\psi_i^{(k)}\|_a^2$.

Using (3.25) we have $\|\psi_i^{(k)}\|_a^2 \leq CH_k^{-d-2}$, therefore, $\sum_{j \in \mathcal{I}^{(k+1)}/i\rho_{k,k+1}} |R_{i,j}^{(k,k+1)}|^2 \leq CH^d e^{-C^{-1}\rho_k}$. \square

Let us now prove Theorem 4.19. We obtain by induction (using the constraints in (4.44)) that for $k \in \{1, \dots, q\}$ and $i \in \mathcal{I}^{(k)}$, $\psi_i^{(k),\text{loc}}$ satisfies the constraints of (4.3). Moreover, (3.8) implies that if ψ satisfies the constraints of (4.3), then $\|\psi\|_a^2 = \|\psi_i^{(k)}\|_a^2 + \|\psi - \psi_i^{(k)}\|_a^2$. Therefore, for $k \in \{2, \dots, q-1\}$, $\psi_i^{(k),\text{loc}}$ is also the minimizer of $\|\psi - \psi_i^{(k)}\|_a$ over functions ψ of the form $\psi = \sum_{j \in i\rho_{k,k+1}} c_j \psi_j^{(k+1),\text{loc}}$ satisfying the constraints of (4.44). Thus, writing $\psi^* := \sum_{j \in i\rho_{k,k+1}} R_{i,j}^{(k,k+1)} \psi_j^{(k+1),\text{loc}}$, we have (since ψ^* satisfies the constraints of (4.44)) $\|\psi_i^{(k),\text{loc}} - \psi_i^{(k)}\|_a \leq \|\psi^* - \psi_i^{(k)}\|_a$. Write $\psi_1 := \sum_{j \in \mathcal{I}^{(k+1)}} R_{i,j}^{(k,k+1)} \psi_j^{(k+1),\text{loc}}$ and $\psi_2 := \sum_{j \in \mathcal{I}^{(k+1)}/i\rho_{k,k+1}} R_{i,j}^{(k,k+1)} \psi_j^{(k+1),\text{loc}}$. Observing that $\psi^* = \psi_1 - \psi_2$, we deduce that $\|\psi_i^{(k),\text{loc}} - \psi_i^{(k)}\|_a^2 \leq 2\|\psi_1 - \psi_i^{(k)}\|_a^2 + 2\|\psi_2\|_a^2$ and, after summing over i , $(\mathcal{E}(k))^2 \leq 2(I_1 + I_2)$ with $I_1 = \sum_{i \in \mathcal{I}^{(k)}} \|\sum_{j \in \mathcal{I}^{(k+1)}} R_{i,j}^{(k,k+1)} (\psi_j^{(k+1)} - \psi_j^{(k+1),\text{loc}})\|_a^2$ and $I_2 = \sum_{i \in \mathcal{I}^{(k)}} \|\sum_{j \in \mathcal{I}^{(k+1)}/i\rho_{k,k+1}} R_{i,j}^{(k,k+1)} \psi_j^{(k+1),\text{loc}}\|_a^2$. Writing S as the $\mathcal{I}^{(k+1)} \times \mathcal{I}^{(k+1)}$ symmetric positive matrix with entries $S_{i,j} = \langle \psi_i^{(k+1)} - \psi_i^{(k+1),\text{loc}}, \psi_j^{(k+1)} - \psi_j^{(k+1),\text{loc}} \rangle_a$, note that $I_1 = \text{Trace}[R^{(k,k+1)} S R^{(k+1,k)}]$. Writing $S^{\frac{1}{2}}$ as the matrix square root of S , observe that for a matrix U , using the cyclic property of the trace, $\text{Trace}[U S U^T] = \text{Trace}[S^{\frac{1}{2}} U^T U S^{\frac{1}{2}}] \leq \lambda_{\max}(U^T U) \text{Trace}[S]$, which, observing that $\text{Trace}[S] = (\mathcal{E}(k+1))^2$ and $\lambda_{\max}(U^T U) = \|U\|_2^2$, implies $I_1 \leq \|R^{(k,k+1)}\|_2^2 (\mathcal{E}(k+1))^2$. Therefore (using Lemma 4.20), we have $\sqrt{I_1} \leq CH^{d/2} \mathcal{E}(k+1)$. Let us now bound I_2 . Let R be defined as in Lemma 4.21. Noting that $\langle \psi_i^{(k+1),\text{loc}}, \psi_j^{(k+1),\text{loc}} \rangle_a = A_{i,j}^{(k+1),\text{loc}}$ we have (as above) $I_2 = \text{Trace}[R A^{(k+1),\text{loc}} R^T] \leq \lambda_{\max}(R^T R) \text{Trace}[A^{(k+1),\text{loc}}]$.

Summarizing and using Lemma 4.21 we deduce that $\mathcal{E}(k) \leq CH^{d/2} \mathcal{E}(k+1) + CH^{d/2} e^{-\rho_k/C} \sqrt{\text{Trace}[A^{(k+1),\text{loc}}]}$. Observing that $\sqrt{\text{Trace}[A^{(k+1),\text{loc}}]} \leq \mathcal{E}(k+1) + \sqrt{\text{Trace}[A^{(k+1)}]}$ and using $\text{Trace}[A^{(k+1)}] \leq CH_{k+1}^{-d} \max_{i \in \mathcal{I}^{(k+1)}} \|\psi_i^{(k+1)}\|_a^2$ and (Lemma 3.12) $\|\psi_i^{(k+1)}\|_a \leq CH_{k+1}^{-\frac{d}{2}-1}$, we conclude the proof of the theorem. \square

Let $u^{(1),\text{loc}}$ be the finite element solution of (1.1) in $\mathfrak{W}^{(1),\text{loc}} := \text{span}\{\psi_j^{(k),\text{loc}} \mid j \in \mathcal{I}^{(1)}\}$. For $k \in \{2, \dots, q\}$, let $W^{(k)}$ be defined as in Construction 4.9 or Construction 4.11. For $i \in \mathcal{J}^{(k)}$, let $\chi_i^{(k),\text{loc}} := \sum_{j \in \mathcal{I}^{(k)}} W_{i,j}^{(k)} \psi_j^{(k),\text{loc}}$. For $k \in \{2, \dots, q\}$, let $u^{(k),\text{loc}} - u^{(k-1),\text{loc}}$ be the finite element solution of (1.1) in $\mathfrak{W}^{(k),\text{loc}} := \text{span}\{\chi_j^{(k),\text{loc}} \mid j \in \mathcal{J}^{(k)}\}$. For $k \in \{2, \dots, q\}$, write $u^{(k),\text{loc}} := u^{(1),\text{loc}} + \sum_{j=2}^k (u^{(j),\text{loc}} - u^{(j-1),\text{loc}})$. Let $B^{(k),\text{loc}}$ be the $\mathcal{J}^{(k)} \times \mathcal{J}^{(k)}$ matrix defined by $B_{i,j}^{(k),\text{loc}} := \langle \chi_i^{(k),\text{loc}}, \chi_j^{(k),\text{loc}} \rangle_a$. Observe that $B^{(k),\text{loc}} = W^{(k)} A^{(k),\text{loc}} W^{(k),T}$. Write for $k \in \{2, \dots, q\}$, $\mathcal{E}(k, \chi) := (\sum_{j \in \mathcal{J}^{(k)}} \|\chi_j^{(k)} - \chi_j^{(k),\text{loc}}\|_a^2)^{\frac{1}{2}}$. The following theorem allows us to control the effect of the localization error on the approximation of the solution of (1.1).

THEOREM 4.22. *It holds true that, for $k \in \{2, \dots, q\}$, (1) $\mathcal{E}(k, \chi) \leq CH^{-d/2} \mathcal{E}(k)$. Furthermore for $k \in \{2, \dots, q\}$ and $\mathcal{E}(k, \chi) \leq C^{-1} H^{-k(1+d/2)+1+d}$, we have (2) $\text{Cond}(B^{(k),\text{loc}}) \leq CH^{-2-2d}$ and (3) $\|u^{(k)} - u^{(k-1)} - (u^{(k),\text{loc}} - u^{(k-1),\text{loc}})\|_a \leq C \mathcal{E}(k, \chi) \|g\|_{H^{-1}(\Omega)} H^{k(1+d/2)-3d-3}$. Similarly, for $\mathcal{E}(1) \leq C^{-1} H^{-d/2}$, we have (4) $\text{Cond}(A^{(1),\text{loc}}) \leq CH^{-2}$ and (5) $\|u^{(1)} - u^{(1),\text{loc}}\|_a \leq C \mathcal{E}(1) \|g\|_{H^{-1}(\Omega)} H^{-2+d/2}$.*

Proof. We will need the following lemma.

LEMMA 4.23. Let χ_1, \dots, χ_m be linearly independent elements of $H_0^1(\Omega)$. Let χ'_1, \dots, χ'_m be another set of linearly independent elements of $H_0^1(\Omega)$. Write $\mathcal{E} := (\sum_{i=1}^m \|\chi_i - \chi'_i\|_a^2)^{\frac{1}{2}}$. Let B (resp., B') be the $m \times m$ matrix defined by $B_{i,j} = \langle \chi_i, \chi_j \rangle_a$ (resp., $B'_{i,j} = \langle \chi'_i, \chi'_j \rangle_a$). Let u_m (resp., u'_m) be the solution of (1.1) in $\text{span}\{\chi_i \mid i = 1, \dots, m\}$ (resp., $\text{span}\{\chi'_i \mid i = 1, \dots, m\}$). It holds true that for $\mathcal{E} \leq \sqrt{\lambda_{\min}(B)}/2$ (1) $\text{Cond}(B') \leq 8 \text{Cond}(B)$, (2) $\|B - B'\|_2 \leq 3\sqrt{\lambda_{\max}(B)}\mathcal{E}$, (3) $\|B^{-1} - (B')^{-1}\|_2 \leq 12\sqrt{\lambda_{\max}(B)}(\lambda_{\min}(B))^{-2}\mathcal{E}$, and (4) $\|u_m - u'_m\|_a \leq C\mathcal{E}\|g\|_{H^{-1}(\Omega)} \frac{\text{Cond}(B)}{\sqrt{\lambda_{\min}(B)}}$.

Proof. For (1) observe that $\sqrt{\lambda_{\max}(B')} = \sup_{|x|=1} \|\sum_{i=1}^m x_i \chi'_i\|_a \leq \sqrt{\lambda_{\max}(B)} + \mathcal{E}$ and $\sqrt{\lambda_{\min}(B')} = \inf_{|x|=1} \|\sum_{i=1}^m x_i \chi'_i\|_a \geq \sqrt{\lambda_{\min}(B)} - \mathcal{E}$. For (2) observe that for $x, y \in \mathbb{R}^m$ with $|x| = |y| = 1$ we have $y^T(B - B')x = \langle \sum_{i=1}^m y_i(\chi_i - \chi'_i), \sum_{i=1}^m x_i \chi_i \rangle_a - \langle \sum_{i=1}^m y_i \chi'_i, \sum_{i=1}^m x_i(\chi_i - \chi'_i) \rangle_a \leq (\sqrt{\lambda_{\max}(B')} + \sqrt{\lambda_{\max}(B)})\mathcal{E}$. (3) follows from (2) and $\|B^{-1} - (B')^{-1}\|_2 \leq \|B - B'\|_2 / (\lambda_{\min}(B)\lambda_{\min}(B'))$. For (4) observe that $u_m = \sum_{i=1}^m w_i \chi_i$ (resp., $u'_m = \sum_{i=1}^m w'_i \chi'_i$), where $w = B^{-1}b$ with $b_i = \int_{\Omega} g \chi_i$ (resp., $w' = (B')^{-1}b'$ with $b'_i = \int_{\Omega} g \chi'_i$). Therefore, $\|u_m - u'_m\|_a \leq |w|\mathcal{E} + |w - w'|\sqrt{\lambda_{\max}(B)}$. $w - w' = B^{-1}(b - b') - B^{-1}(B - B')w'$ leads to $|w - w'| \leq C(\|g\|_{H^{-1}(\Omega)}\mathcal{E} + \|B - B'\|_2|w'|) / \lambda_{\min}(B)$. Using (2), $\lambda_{\min}(B)|w|^2 \leq \|\sum_{i=1}^m w_i \chi_i\|_a^2 \leq \|u\|_a^2 \leq C\|g\|_{H^{-1}(\Omega)}^2$, and $\lambda_{\min}(B')|w'|^2 \leq C\|g\|_{H^{-1}(\Omega)}^2$ we conclude the proof of (4) after simplification. \square

Let us now prove Theorem 4.22. Using $\chi_j^{(k)} - \chi_j^{(k),\text{loc}} = \sum_{i \in \mathcal{I}^{(k)}} W_{j,i}^{(k)}(\psi_i^{(k)} - \psi_i^{(k),\text{loc}})$ and noting that $W_{j,i}^{(k)} = 0$ for $i^{(k-1)} \neq j^{(k-1)}$ we have $(\mathcal{E}(k, \chi))^{(k)} \leq \sum_{j \in \mathcal{J}^{(k)}} (\sum_{i \in \mathcal{I}^{(k)}} (W_{j,i}^{(k)})^2 \sum_{i \in \mathcal{I}^{(k)}, i^{(k-1)}=j^{(k-1)}} \|\psi_i^{(k)} - \psi_i^{(k),\text{loc}}\|_a^2)$. Therefore, $(\mathcal{E}(k, \chi))^{(k)} \leq (\mathcal{E}(k))^{(k)} \max_{i \in \mathcal{I}^{(k-1)}} m_i \max_{j \in \mathcal{J}^{(k)}} \sum_{i \in \mathcal{I}^{(k)}} (W_{j,i}^{(k)})^2$. Observing that (see (4.42)) $\max_{i \in \mathcal{I}^{(k-1)}} m_i \leq 1/(H\delta)^d$ and $\sum_{i \in \mathcal{I}^{(k)}} (W_{j,i}^{(k)})^2 \leq \lambda_{\max}(W^{(k)}W^{(k),T}) \leq 3$ (see Lemma 4.20) we conclude that (1) holds true with $C = (3/\delta^d)^{\frac{1}{2}}$. (2) and (3) are a direct application of Lemmas 4.23 and 4.20. For (3), observe that $u^{(k)} - u^{(k-1)}$ (resp., $u^{(k),\text{loc}} - u^{(k-1),\text{loc}}$) is the finite element solution of (1.1) in $\mathfrak{W}^{(k)}$ (resp., $\mathfrak{W}^{(k),\text{loc}} := \text{span}\{\chi_j^{(k),\text{loc}} \mid j \in \mathcal{J}^{(k)}\}$). The proof of (4) and (5) is similar to that of (2) and (3). \square

THEOREM 4.24. Let $k \in \{1, \dots, q\}$. We have $\mathcal{E}(k) \leq C \sum_{j=k}^q e^{-\rho_j/C} C^{j-k} H^{-\frac{d}{2} + k\frac{d}{2} - j\frac{3d}{2}}$.

Proof. By Theorem 4.19, for $k \in \{1, \dots, q-1\}$, we have $\mathcal{E}(k) \leq a_k + b_k \mathcal{E}(k+1)$ with $a_k = C e^{-\rho_k/C} H^{\frac{d}{2} - (k+1)(d+1)}$ and $b_k = CH^{-\frac{d}{2}}$. Therefore, we obtain by induction that $\mathcal{E}(k) \leq a_k + b_k a_{k+1} + b_k b_{k+1} a_{k+2} + \dots + b_k \dots b_{q-2} a_{q-1} + b_k \dots b_{q-1} \mathcal{E}(q)$. Using Theorem 3.11 we have $\mathcal{E}(q) \leq CH^{-d/2 - q(2+d/2)} e^{-\rho_q/C}$ and obtain the result after simplification. \square

THEOREM 4.25. Let $\epsilon \in (0, 1)$. It holds true that if $\rho_k \geq C((1 + \frac{1}{\ln(1/H)}) \ln \frac{1}{H^k} + \ln \frac{1}{\epsilon})$ for $k \in \{1, \dots, q\}$, then (1) for $k \in \{1, \dots, q\}$ we have $\|u^{(k)} - u^{(k),\text{loc}}\|_a \leq \epsilon \|g\|_{H^{-1}(\Omega)}$ and $\|u - u^{(k),\text{loc}}\|_a \leq C(H^k + \epsilon) \|g\|_{L^2(\Omega)}$, (2) $\text{Cond}(A^{(1),\text{loc}}) \leq CH^{-2}$, and, for $k \in \{2, \dots, q\}$, we have (3) $\text{Cond}(B^{(k),\text{loc}}) \leq CH^{-2-2d}$ and (4) $\|u^{(k)} - u^{(k-1)} - (u^{(k),\text{loc}} - u^{(k-1),\text{loc}})\|_a \leq \frac{\epsilon}{2k^2} \|g\|_{H^{-1}(\Omega)}$.

Proof. Theorems 4.3 and 4.22 imply that the results of Theorem 4.25 hold true if for $k \in \{1, \dots, q\}$, $\mathcal{E}(k) \leq C^{-1} H^{-k(1+d/2) + 7d/2 + 3} \epsilon/k^2$. Using Theorem 4.24 we deduce that the results of Theorem 4.25 hold true if for $k \in \{1, \dots, q\}$ and $k \leq j \leq q$

we have $Ce^{-\rho_j/C} C^{j-k} H^{-\frac{d}{2}+k\frac{d}{2}-j3\frac{d}{2}} \leq H^{-k(1+d/2)+7d/2+3} \epsilon/(k^2 j^2)$. We conclude after simplification. \square

5. The Algorithm, Its Implementation, and Its Complexity.

5.1. The Initialization of the Algorithm. To describe the practical implementation of the algorithm we consider the (finite element) discretized version of (1.1). Let \mathcal{T}_h be a regular fine mesh discretization of Ω of resolution h with $0 < h \ll 1$. Let \mathcal{N} be the set of interior nodes z_i and $N = |\mathcal{N}|$ be the number of interior nodes ($N = \mathcal{O}(h^{-d})$) of \mathcal{T}_h . Write $(\varphi_i)_{i \in \mathcal{N}}$ as a set of regular nodal basis elements (of $H_0^1(\Omega)$) constructed from \mathcal{T}_h such that, for each $i \in \mathcal{N}$, $\text{support}(\varphi_i) \subset B(z_i, C_0 h)$ and for $y \in \mathbb{R}^N$,

$$(5.1) \quad \gamma h^d |y|^2 \leq \left\| \sum_{i \in \mathcal{N}} y_i \varphi_i \right\|_{L^2(\Omega)}^2 \leq \bar{\gamma} h^d |y|^2$$

for some constants $\gamma, \bar{\gamma}, C_0 \approx \mathcal{O}(1)$. In addition to (5.1) the regularity of the finite elements is used to ensure the availability of the inverse Poincaré inequality

$$(5.2) \quad \|\nabla v\|_{L^2(\Omega)} \leq C_1 h^{-1} \|v\|_{L^2(\Omega)}$$

for $v \in \text{span}\{\varphi_i \mid i \in \mathcal{N}\}$ and some constant $C_1 \approx \mathcal{O}(1)$, used to generalize the proof of Theorem 4.14 to the discrete case.

Given $g = \sum_{i \in \mathcal{N}} g_i \varphi_i$ we want to find $u \in \text{span}\{\varphi_i \mid i \in \mathcal{N}\}$ such that for all $j \in \mathcal{N}$,

$$(5.3) \quad \langle \varphi_j, u \rangle_a = \int_{\Omega} \varphi_j g \text{ for all } j \in \mathcal{N}.$$

In practical applications a is naturally assumed to be piecewise constant over the fine mesh (e.g., of constant value in each triangle or square of \mathcal{T}_h) and one purpose of the algorithm is the fast resolution of the linear system (5.3) up to accuracy $\epsilon \in (0, 1)$.

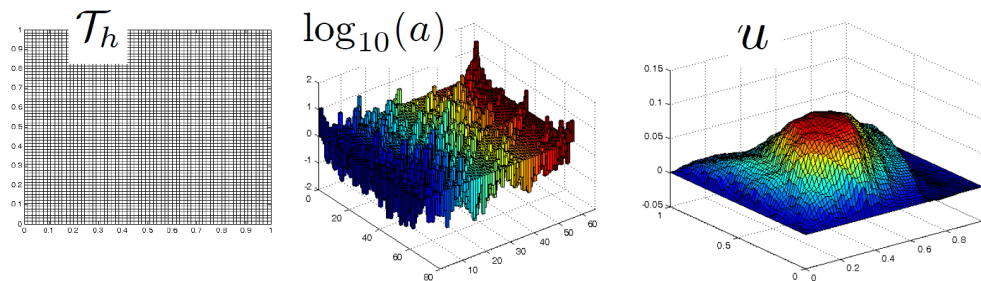


Fig. 2 The (fine) mesh \mathcal{T}_h , a (in \log_{10} scale), and u .

EXAMPLE 5.1. We will illustrate the presentation of the algorithm with a numerical example in which \mathcal{T}_h is a square grid of mesh size $h = (1 + 2^q)^{-1}$ with $q = 6$ and 64×64 interior nodes (see Figure 2). a is piecewise constant on each square of \mathcal{T}_h given by $a(x) = \prod_{k=1}^6 \left(1 + 0.5 \cos\left(2^k \pi \left(\frac{i}{2^q+1} + \frac{j}{2^q+1}\right)\right)\right) \left(1 + 0.5 \sin\left(2^k \pi \left(\frac{j}{2^q+1} - 3\frac{i}{2^q+1}\right)\right)\right)$ for $x \in \left[\frac{i}{2^q+1}, \frac{i+1}{2^q+1}\right] \times \left[\frac{j}{2^q+1}, \frac{j+1}{2^q+1}\right]$. The contrast of a (i.e., when a is scalar, the ratio between its maximum and minimum values) is 1866. The finite element discretization

(5.3) is obtained using continuous nodal basis elements φ_i spanned by $\{1, x_1, x_2, x_1x_2\}$ in each square of \mathcal{T}_h . Writing z_i as the positions of the interior nodes of \mathcal{T}_h , we choose, for our numerical example, $g(x) = \sum_{i \in \mathcal{N}} (\cos(3z_{i,1} + z_{i,2}) + \sin(3z_{i,2}) + \sin(7z_{i,1} - 5z_{i,2}))\varphi_i(x)$.

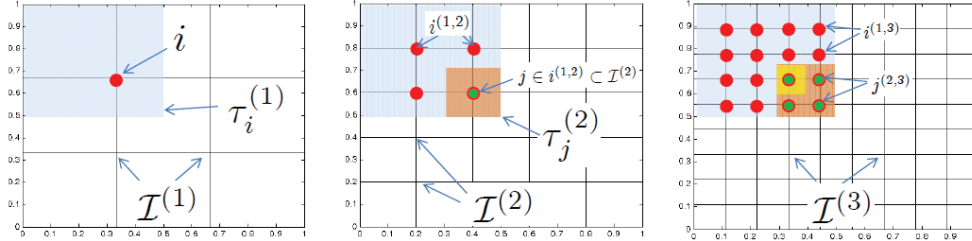


Fig. 3 $\mathcal{I}^{(1)}$, $\mathcal{I}^{(2)}$, and $\mathcal{I}^{(3)}$.

The first step of the proposed algorithm is the construction of the index tree \mathcal{I} of Definition 4.2 describing the domain decomposition of Definition 4.2. To ensure a uniform bound on the condition numbers of the stiffness matrices (4.19), one must select the resolutions H_k to form a geometric sequence (or simply such that H_{k-1}/H_k is uniformly bounded), i.e., $H_k = H^k$ for some $H \in (0, 1)$ (for our numerical example, $H = 1/2$, $q = 6$, and we identify $\mathcal{I}^{(k)}$ as the indices of the interior nodes of a square grid of resolution $(1 + 2^k)^{-1}$ as illustrated in Figure 3). In this construction $H^q = h$ corresponds to the resolution of the fine mesh and each subset $\tau_i^{(q)}$ ($i \in \mathcal{I}^{(q)}$) contains one and only one element of \mathcal{N} (interior node of the fine mesh). Using this one-to-one correspondence we use the elements of $\mathcal{I} = \mathcal{I}^q$ to (re)label the nodal elements $(\varphi_i)_{i \in \mathcal{N}}$ as $(\varphi_i)_{i \in \mathcal{I}}$. The measurement functions $(\phi_i^{(k)})_{i \in \mathcal{I}^{(k)}}$ are then identified (1) by selecting $\phi_i^{(q)} = \varphi_i$ for $i \in \mathcal{I}^{(q)}$ and (2) via the nested aggregation (4.1) of the nodal elements (as is commonly done in AMG), i.e., $\phi_i^{(k)} = \sum_{j \in \mathcal{I}^{(k+1)}} \pi_{i,j}^{(k,k+1)} \phi_j^{(k+1)} = \sum_{j \in i^{(k,k+1)}} \phi_j^{(k+1)}$ for $k \in \{1, \dots, q-1\}$ and $i \in \mathcal{I}^{(k)}$.

Remark 8. We refer to Figure 4 for an illustration of these measurement functions for our numerical example. Note that the support of each $\phi_i^{(k)}$ is only approximately (and not exactly) $\tau_i^{(k)}$ and that the $\phi_i^{(k)}$ are only approximate set functions (and not exact ones). This does not affect the design, accuracy, and localization of the algorithm presented here because the frame inequalities (4.32), and the Poincaré inequalities $\|\sum_{i \in \mathcal{I}^{(k)}} x_i \phi_i^{(k)}\|_{H^{-1}(\Omega)} \leq C H^{k-1} \|\phi_i^{(k)}\|_{L^2(\Omega)}$ $x \in \text{Ker}(\pi^{(k-1,k)})$, hold true. Indeed, (5.1) and Construction 4.2 imply that the frame inequalities (4.32) with $\tilde{\gamma}_k \leq \tilde{\gamma} \delta^{-d}$ and $\underline{\gamma}_k \geq \underline{\gamma} \delta^d$ and the Poincaré inequalities are regularity/homogeneity conditions on the mesh and the aggregated elements. Although a fine mesh has been used to facilitate the presentation of the algorithm, the proposed method is meshless (it only requires the specification of the basis elements $(\varphi_i)_{i \in \mathcal{I}}$).

5.2. Exact Gamblet Transform and Multiresolution Operator Inversion. The near-linear complexity of the proposed multiresolution algorithm (Algorithm 2) is based on three properties: (i) nesting, (ii) uniformly bounded condition numbers, and (iii) localization/truncation based on exponential decay. Truncation/localization levels/subsets are, a priori, functions of the desired level of accuracy $\epsilon \in (0, 1)$ in ap-

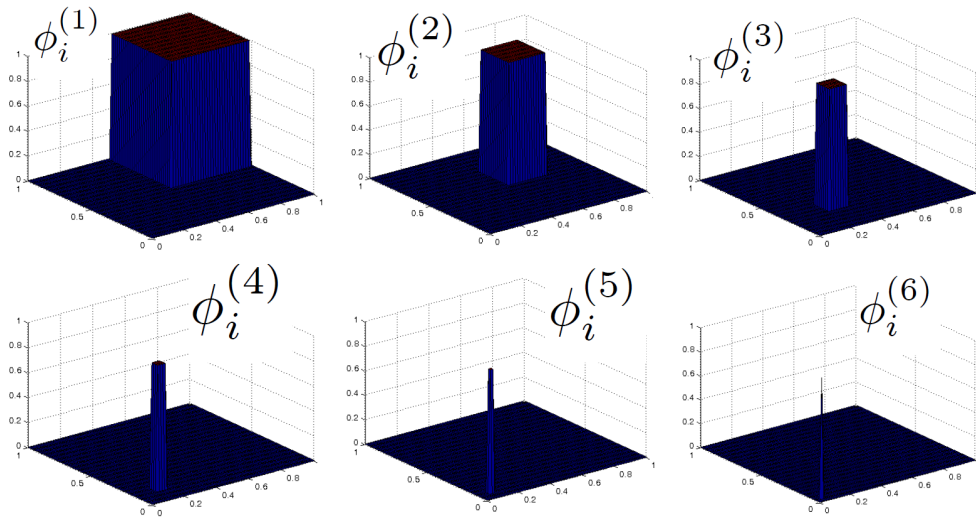


Fig. 4 The functions ϕ_i^k with $k \in \{1, \dots, q\}$ and $q = 6$.

Algorithm 1 Exact gamblet transform/solve.

```

1: For  $i, j \in \mathcal{I}^{(q)}$ ,  $M_{i,j} = \int_{\Omega} \varphi_i \varphi_j$  // Mass matrix
2: For  $i, j \in \mathcal{I}^{(q)}$ ,  $A_{i,j} = \int_{\Omega} (\nabla \varphi_i)^T a \nabla \varphi_j$  // Stiffness matrix
3: Compute  $M^{-1}$  // Mass matrix inversion
4: For  $i \in \mathcal{I}^{(q)}$ ,  $\psi_i^{(q)} = \sum_{j \in \mathcal{I}^{(q)}} M_{i,j}^{-1} \varphi_j$  // Level  $q$  gamblets
5: For  $i \in \mathcal{I}^{(q)}$ ,  $g_i^{(q)} = g_i$  //  $g_i^{(q)} = \int_{\Omega} \psi_i^{(q)} g$  with  $g = \sum_{i \in \mathcal{I}^{(q)}} g_i \varphi_i$ 
6: For  $i, j \in \mathcal{I}^{(q)}$ ,  $A_{i,j}^{(q)} = \int_{\Omega} (\nabla \psi_i^{(q)})^T a \nabla \psi_j^{(q)}$  //  $A^{(q)} = M^{-1} A M^{-1,T}$ 
7: for  $k = q$  to 2 do
8:    $B^{(k)} = W^{(k)} A^{(k)} W^{(k),T}$  // Eq. (4.19)
9:    $w^{(k)} = B^{(k),-1} W^{(k)} g^{(k)}$  // Eq. (4.26)
10:  For  $i \in \mathcal{J}^{(k)}$ ,  $\chi_i^{(k)} = \sum_{j \in \mathcal{I}^{(k)}} W_{i,j}^{(k)} \psi_j^{(k)}$  // Eq. (4.17)
11:   $u^{(k)} - u^{(k-1)} = \sum_{i \in \mathcal{J}^{(k)}} w_i^{(k)} \chi_i^{(k)}$  // Thm. 4.13
12:   $D^{(k,k-1)} = -B^{(k),-1} W^{(k)} A^{(k)} \bar{\pi}^{(k,k-1)}$  // Eq. (4.21)
13:   $R^{(k-1,k)} = \bar{\pi}^{(k-1,k)} + D^{(k-1,k)} W^{(k)}$  // Eq. (4.23)
14:   $A^{(k-1)} = R^{(k-1,k)} A^{(k)} R^{(k,k-1)}$  // Eq. (4.16)
15:  For  $i \in \mathcal{I}^{(k-1)}$ ,  $\psi_i^{(k-1)} = \sum_{j \in \mathcal{I}^{(k)}} R_{i,j}^{(k-1,k)} \psi_j^{(k)}$  // Eq. (4.11)
16:   $g^{(k-1)} = R^{(k-1,k)} g^{(k)}$  // Eq. (4.25)
17: end for
18:  $U^{(1)} = A^{(1),-1} g^{(1)}$  // Eq. (4.27)
19:  $u^{(1)} = \sum_{i \in \mathcal{I}^{(1)}} U_i^{(1)} \psi_i^{(1)}$  // Thm. 4.13
20:  $u = u^{(1)} + (u^{(2)} - u^{(1)}) + \dots + (u^{(q)} - u^{(q-1)})$  // Thm. 4.5 with  $u = u^{(q)}$ 

```

proximating the solution of (5.3), and to distinguish the implementation of localization/truncation (and its consequences) we will first describe this algorithm in its *zero approximation error version* (i.e., $\epsilon = 0$ and without using localization/truncation;

see Algorithm 1). Although this *error-free* version (Algorithm 1) performs the decomposition of the resolution of the linear system (5.3) (whose condition number is of the order of $h^{-d-2} \gg 1$) into the resolutions of a nesting of linear systems with uniformly bounded condition numbers, it is not of near-linear complexity due to the presence of dense matrices. Algorithm 2 achieves near-linear complexity by truncating/localizing the dense matrices appearing in Algorithm 1 (ϵ -accuracy is ensured using the off-diagonal exponential decay of these dense matrices). Let us now describe Algorithm 1 in detail. Lines 1 and 2 correspond to the computation of the (sparse) mass and stiffness matrices of (5.3). Line 4 corresponds to the calculation of level q gamblets $\psi_i^{(q)}$ defined as the minimizer of $\|\psi\|_a$ subject to $\int_{\Omega} \psi \phi_j^{(q)} = \delta_{i,j}$ and $\psi \in \text{span}\{\varphi_l \mid l \in \mathcal{I}\}$; note that since the number of constraints is equal to the number of degrees of freedom of ψ , and since $\int_{\Omega} \varphi_l \phi_j^{(q)} = M_{l,j}$, level q gamblets do not depend on a and are obtained by inverting the mass matrix in line 3 (note that by (5.1), the mass matrix is of $\mathcal{O}(1)$ condition number). Although not done here, one can also initialize the algorithm (and its fast version) with $\psi_i^{(q)} = \varphi_i$ (which is equivalent to using $\sum_{j \in \mathcal{I}^{(q)}} M_{i,j}^{-1} \varphi_j^{(q)}$ as level q measurement functions). Line 5 corresponds to initialization of the vector $g^{(q)}$ introduced above (4.25). Line 6 corresponds to the initialization of the stiffness matrix $A^{(q)}$ introduced in (4.14). The core of the algorithm is the nested computation performed (iteratively from $k = q$ down to $k = 2$) in lines 8 to 16. Note that this nested computation takes $A^{(k)}, g^{(k)}$, and $(\psi_i^{(k)})_{i \in \mathcal{I}^{(k)}}$ as inputs and produces (1) $A^{(k-1)}, g^{(k-1)}$, and $(\psi_i^{(k-1)})_{i \in \mathcal{I}^{(k-1)}}$ as outputs for the next iteration, and (2) the subband $u^{(k)} - u^{(k-1)}$ of the solution and subband gamblets $(\chi_i^{(k)})_{i \in \mathcal{J}^{(k)}}$ (which do not need to be explicitly computed/stored since line 11 is equivalent to $u^{(k)} - u^{(k-1)} = \sum_{i \in \mathcal{I}^{(k)}} (W^{(k),T} w^{(k)})_i \psi_i^{(k)}$). Note also that the gamblets $(\psi_i^{(k)})_{i \in \mathcal{I}^{(k)}}$ and $(\chi_i^{(k)})_{i \in \mathcal{J}^{(k)}}$ can be stored and displayed using the hierarchical structure (4.11). Throughout this section and in what follows we assume that the matrices $W^{(k)}$ are obtained as in Constructions 4.9 or 4.11. Note that the number of nonzero entries of $\pi^{(k-1,k)}$ and $W^{(k)}$ is $\mathcal{O}(|\mathcal{I}^{(k)}|)$ (proportional to H^{-k} in our numerical example). Line 9 corresponds to solving the well-conditioned linear system $B^{(k)} w^{(k)} = W^{(k)} g^{(k)}$ and the $|\mathcal{I}^{(k-1)}|$ well-conditioned linear systems $B^{(k)} D^{(k,k-1)} = -W^{(k)} A^{(k)} \pi^{(k,k-1)}$. Note that by Theorem 4.15 the matrices $B^{(k)}$ have uniformly bounded condition numbers and these linear systems can be solved efficiently using iterative methods (such as the CG method recalled in section 4.9). $u^{(1)}$ is computed in lines 19 and 20 (recall that $A^{(1)}$ is also of uniformly bounded condition number) and the last step of the algorithm is to obtain u via simple addition of the subband/subscale solution $u^{(1)}$ and $(u^{(k)} - u^{(k-1)})_{2 \leq k \leq q}$. Observe that the operating diagram of Algorithm 1 is not a V or W but an inverted pyramid (or a comb). More precisely, the basis functions $\psi_i^{(k)}$ are computed hierarchically from fine to coarse scales. Furthermore, as soon as the elements $\psi_i^{(k)}$ have been computed, they can be applied (independently from the other scales) to the computation of $u^{(k)} - u^{(k-1)}$ (the projection of u onto $\mathfrak{W}^{(k)}$ corresponding to the bandwidth $[H^k, H^{k-1}]$).

EXAMPLE 5.2. We refer to Figures 5 and 6 for an illustration of the gamblets $\psi_i^{(k)}$ and $\chi_j^{(k)}$ corresponding to Example 5.1 with $W^{(k)}$ defined by Construction 4.9. We refer to Figure 7 for an illustration of the exponential decay of the gamblets $\psi_i^{(k)}$. We refer to Figure 8 for an illustration of the condition numbers of $A^{(k)}$ and $B^{(k)}$ (with $W^{(k)}$ still defined by Construction 4.9). Observe that the bound on the condition num-

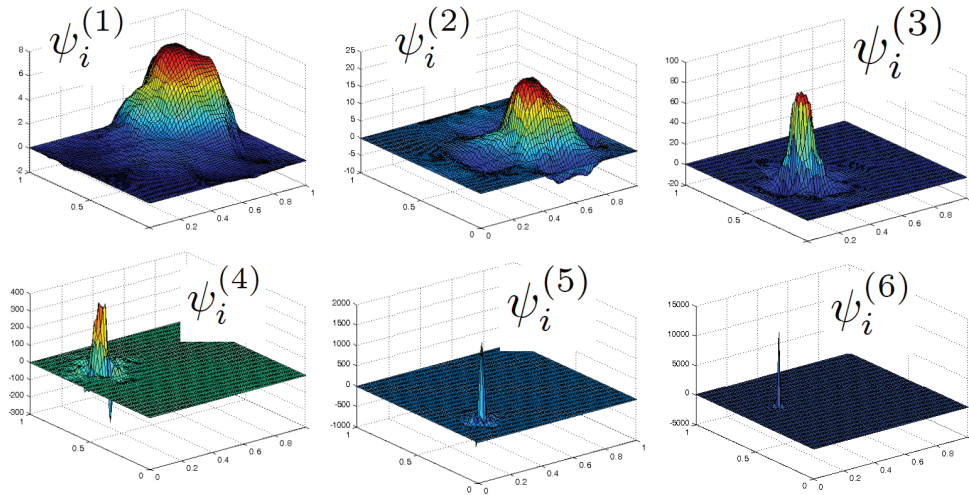


Fig. 5 The basis elements ψ_i^k with $k \in \{1, \dots, 6\}$.

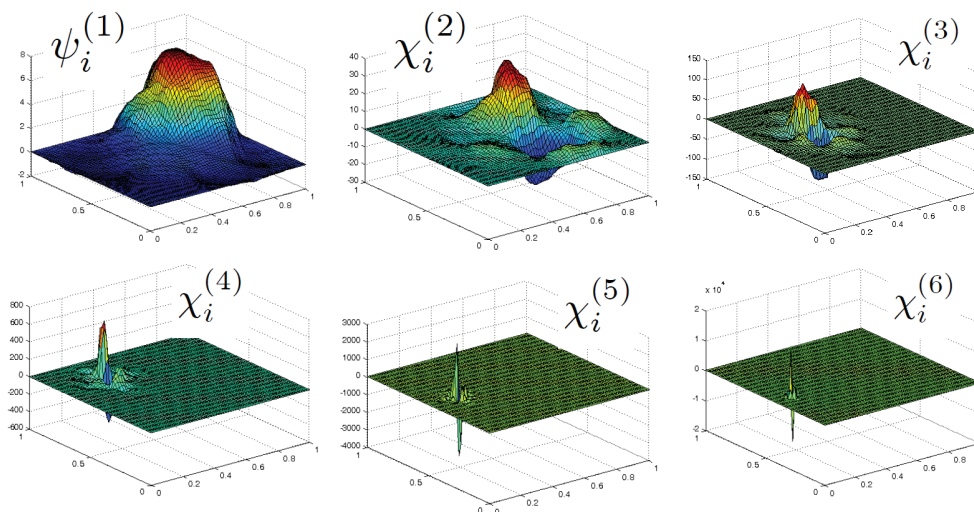


Fig. 6 The basis elements ψ_i^1 and χ_i^k with $k \in \{2, \dots, 6\}$.

bers of $B^{(k)}$ depends on the contrast and the saturation of that bound occurs for smaller values of k under low contrast. We refer to Figure 9 for an illustration of the subband solutions $u^{(1)}, u^{(2)} - u^{(1)}, \dots, u^{(q)} - u^{(q-1)}$ corresponding to Example 5.1. Observe that these (subband) solutions form a multiresolution decomposition of u as a sum of functions characterizing the behavior of u at subscales $[H, 1], [H^2, H], \dots, [H^q, H^{q-1}]$. Once the components $u^{(1)}, u^{(2)} - u^{(1)}, \dots,$ and $u^{(q)} - u^{(q-1)}$ have been computed one obtains, via simple summation, $u^{(1)}, \dots, u^{(q)}$, the finite element approximation of u at resolutions H, H^2, \dots, H^q illustrated in Figure 11. As described in Theorem 4.3, the error of the approximation of u by $u^{(k)}$ is proportional to H^k for $k \in \{1, \dots, q-1\}$. For $k = q$, as illustrated in Figure 11, this approximation error drops down to zero because there is no gap between H^q and the fine mesh (i.e., $\psi_i^{(q)}$ and φ_i span the

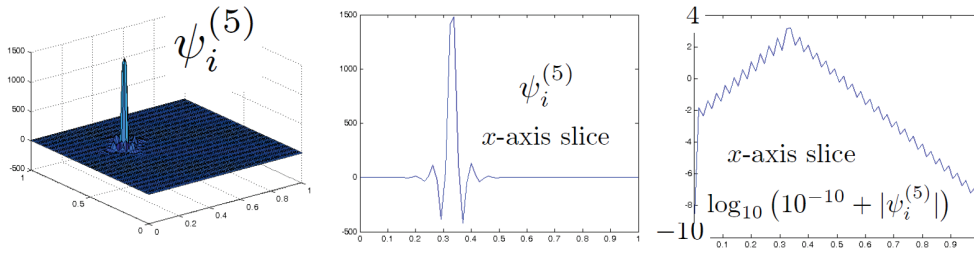


Fig. 7 Exponential decay.

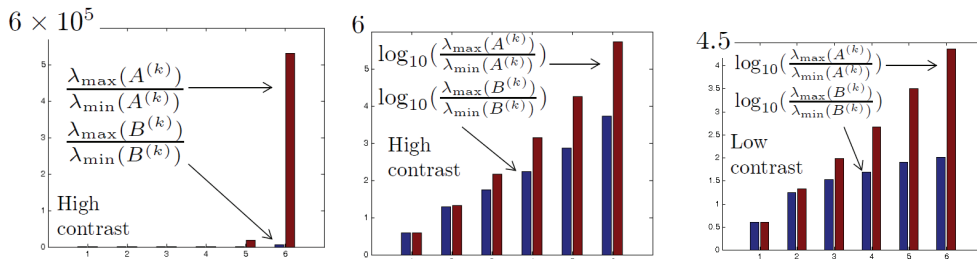


Fig. 8 Condition numbers of $A^{(k)}$ and $B^{(k)}$.

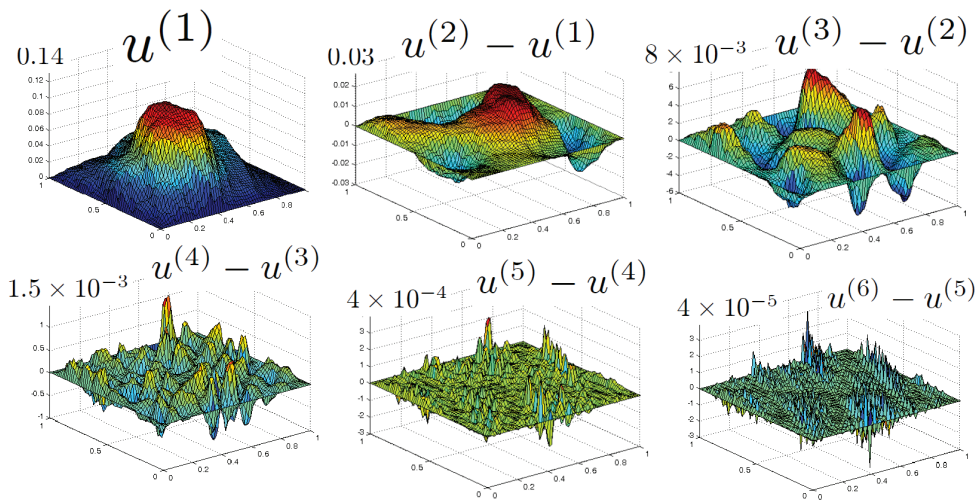


Fig. 9 $u^{(1)}$, $u^{(2)} - u^{(1)}$, ..., and $u^{(q)} - u^{(q-1)}$.

same linear space in the discrete case). Moreover, as illustrated in Figure 10, the representation of u in the basis formed by the functions $\frac{\psi_i^{(1)}}{\|\psi_i^{(1)}\|_\alpha}$ and $\frac{\chi_j^{(k)}}{\|\chi_j^{(k)}\|_\alpha}$ is sparse. Therefore, as illustrated in Figure 11 one can compress u , in this basis, by setting the smallest coefficients to zero without loss in energy norm.

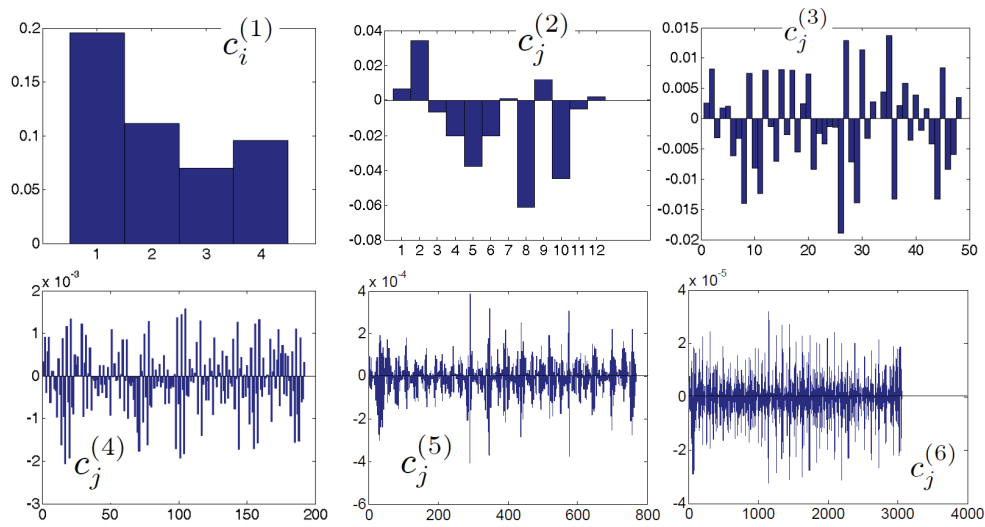


Fig. 10 The coefficients of u in the expansion $u = \sum_i c_i^{(1)} \frac{\psi_i^{(1)}}{\|\psi_i^{(1)}\|_\alpha} + \sum_{k=2}^q \sum_j c_j^{(k)} \frac{\chi_j^{(k)}}{\|\chi_j^{(k)}\|_\alpha}$.

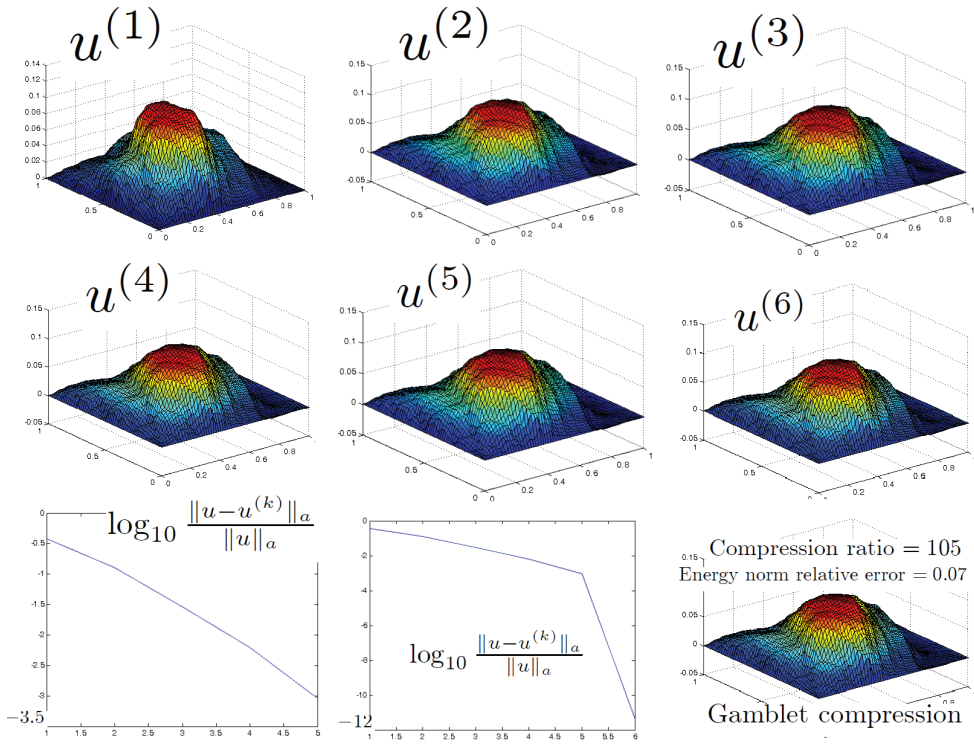


Fig. 11 $u^{(1)}, \dots, u^{(q)}$. Relative approximation error in the energy norm in \log_{10} scale. Compression of u over the basis functions $\psi_i^{(1)}, \chi_i^{(2)}, \dots, \chi_i^{(q)}$ by setting 99% of the smallest coefficients to zero in the decomposition of Figure 10.

5.3. Fast Gamble Transform/Solve. Algorithm 2 achieves near linear complexity (1) in approximating the solution of (5.3) to a given level of accuracy ϵ and (2) in performing an approximate Gamblet transform (sufficient to achieve that level of accuracy). This fast algorithm is obtained by localizing/truncating the linear systems corresponding to lines 3 and 12 in Algorithm 1. We define these localization/truncation steps as follows. For $k \in \{1, \dots, q\}$ and $i \in \mathcal{I}^{(k)}$ define i^ρ as in section 4.10 (i.e., as the subset of indices $j \in \mathcal{I}^{(k)}$ whose corresponding subdomains $\tau_j^{(k)}$ are at distance at most $H_k \rho$ from $\tau_i^{(k)}$).

DEFINITION 5.1. For $i \in \mathcal{I}^{(q)}$, let $M^{(i, \rho_q)}$ be the $i^{\rho_q} \times i^{\rho_q}$ matrix defined by $M_{l,j}^{(i, \rho_q)} = M_{l,j}$ for $l, j \in i^{\rho_q}$. Let $e^{(i, \rho_q)}$ be the $|i^{\rho_q}|$ -dimensional vector defined by $e_j^{(i, \rho_q)} = \delta_{j,i}$ for $j \in i^{\rho_q}$. Let $y^{(i, \rho_q)}$ be the $|i^{\rho_q}|$ -dimensional vector solution of $M^{(i, \rho_q)} y^{(i, \rho_q)} = e^{(i, \rho_q)}$. We define the solution $M_{\cdot,i}^{-1, \rho_q}$ of the localized linear system $M^{i, \rho_q} M_{\cdot,i}^{-1, \rho_q} = \delta_{\cdot,i}$ (line 3 of Algorithm 2) as the i^{ρ_q} -vector given by $M_{j,i}^{-1, \text{loc}} = y_j^{(i, \rho_q)}$ for $j \in i^{\rho_q}$.

Note that the associated gamblet $\psi_i^{(q), \text{loc}}$ (line 4 of Algorithm 2) is also the solution of the problem of finding $\psi \in \text{span}\{\varphi_j \mid j \in i^{\rho_q}\}$ such that $\int_\Omega \psi \varphi_j = \delta_{i,j}$ for $j \in i^{\rho_q}$ (i.e., localizing the computation of the gamblet $\psi_i^{(q)}$ to a subdomain of size $H_q \rho_q$). Line 5 can be replaced by $g_i^{(q), \text{loc}} = g_i$ without loss of accuracy ($g_i^{(q), \text{loc}} = \int_\Omega \psi_i^{(q), \text{loc}} g$ simplifies the presentation of the analysis). line 12 of Algorithm 2 is defined in a similar way as follows.

DEFINITION 5.2. Let $k \in \{2, \dots, q\}$ and B be the positive definite $\mathcal{J}^{(k)} \times \mathcal{J}^{(k)}$ matrix $B^{(k), \text{loc}}$ computed in line 8 of Algorithm 2. For $i \in \mathcal{I}^{(k-1)}$, let $\rho = \rho_{k-1}$ and let i^χ be the subset of indices $j \in \mathcal{J}^{(k)}$ such that $j^{(k-1)} \in i^\rho$ (recall that if $j = (i_1, \dots, i_k) \in \mathcal{J}^{(k)}$, then $j^{(k-1)} := (j_1, \dots, j_{k-1}) \in \mathcal{I}^{(k-1)}$). $B^{(i, \rho)}$ be the $i^\chi \times i^\chi$ matrix defined by $B_{l,j}^{(i, \rho)} = B_{l,j}$ for $l, j \in i^\chi$. Let $b^{(i, \rho)}$ be the $|i^\chi|$ -dimensional vector defined by $b_j^{(i, \rho)} = -(W^{(k)} A^{(k), \text{loc}} \bar{\pi}^{(k, k-1)})_{j,i}$ for $j \in i^\chi$. Let $y^{(i, \rho)}$ be the $|i^\chi|$ -dimensional vector solution of $B^{(i, \rho)} y^{(i, \rho)} = b^{(i, \rho)}$. We define the solution $D^{(k, k-1), \text{loc}}$ of the localized linear system $\text{Inv}(B^{(k), \text{loc}} D^{(k, k-1), \text{loc}} = -W^{(k)} A^{(k), \text{loc}} \bar{\pi}^{(k, k-1)}, \rho_{k-1})$ as the $\mathcal{J}^{(k)} \times \mathcal{I}^{(k-1)}$ sparse matrix given by $D_{j,i}^{(k, k-1), \text{loc}} = 0$ for $j \notin i^\chi$ and $D_{j,i}^{(k, k-1), \text{loc}} = y_j^{(i, \rho)}$ for $j \in i^\chi$. $D^{(k-1, k), \text{loc}}$ (line 13 of Algorithm 2) is then defined as the transpose of $D^{(k, k-1), \text{loc}}$.

Remark 9. Definition 5.2 (line 8 of Algorithm 2) is equivalent to localizing the computation of each gamblet $\psi_i^{(k-1)}$ to a subdomain of size $H_{k-1} \rho_{k-1}$, i.e., the gamblet $\psi_i^{(k-1), \text{loc}}$ computed in line 15 of Algorithm 2 is the solution of (1) the problem of finding ψ in the affine space $\sum_{j \in \mathcal{I}^{(k)}} \bar{\pi}_{i,j}^{(k-1, k)} \psi_j^{(k), \text{loc}} + \text{span}\{\chi_j^{(k), \text{loc}} \mid j^{(k-1)} \in i^{\rho_{k-1}}\}$ such that ψ is $\langle \cdot, \cdot \rangle_a$ orthogonal to $\text{span}\{\chi_j^{(k), \text{loc}} \mid j^{(k-1)} \in i^{\rho_{k-1}}\}$, and (2) the problem of minimizing $\|\psi\|_a$ in $\text{span}\{\psi_l^{(k), \text{loc}} \mid l^{(k-1)} \in i^{\rho_{k-1}}\}$ subject to constraints $\int_\Omega \phi_j^{(k-1)} \psi = \delta_{i,j}$ for $j \in i^{\rho_{k-1}}$.

5.4. Complexity vs. Accuracy of Algorithm 2 and Choice of the Localization Radii ρ_k . The sizes of the localization radii ρ_k (and therefore the complexity of Algorithm 2) depend on whether Algorithm 2 is used as a preconditioner (as is done with AMG) or as a direct solver. Although it is natural to expect the complexity of Algorithm 2 to be significantly smaller if used as a preconditioner (since preconditioning

Algorithm 2 Fast gamblet transform/solve.

1: For $i, j \in \mathcal{I}^{(q)}$, $M_{i,j} = \int_{\Omega} \varphi_i \varphi_j$	// $\mathcal{O}(N)$
2: For $i, j \in \mathcal{I}^{(q)}$, $A_{i,j} = \int_{\Omega} (\nabla \varphi_i)^T a \nabla \varphi_j$	// $\mathcal{O}(N)$
3: $M^{i,\rho_q} M_{i,i}^{-1,\rho_q} = \delta_{i,i}$	// Def. 5.1, Thm. 5.4, $\mathcal{O}(N \rho_q^d \ln \max(\frac{1}{\epsilon}, q))$
4: For $i \in \mathcal{I}^{(q)}$, $\psi_i^{(q),\text{loc}} = \sum_{j \in \mathcal{I}^{\rho_q}} M_{j,i}^{-1,\rho_q} \varphi_j$	// $\mathcal{O}(N \rho_q^d)$
5: For $i \in \mathcal{I}^{(q)}$, $g_i^{(q),\text{loc}} = \int_{\Omega} \psi_i^{(q),\text{loc}} g$	// $\mathcal{O}(N \rho_q^d)$
6: For $i, j \in \mathcal{I}^{(q)}$, $A_{i,j}^{(q),\text{loc}} = \int_{\Omega} (\nabla \psi_i^{(q),\text{loc}})^T a \nabla \psi_j^{(q),\text{loc}}$	// $\mathcal{O}(N \rho_q^{2d})$
7: for $k = q$ to 2 do	
8: $B^{(k),\text{loc}} = W^{(k)} A^{(k),\text{loc}} W^{(k),T}$	// $\mathcal{O}(\mathcal{I}^{(k)} \rho_k^d)$
9: $w^{(k),\text{loc}} = (B^{(k),\text{loc}})^{-1} W^{(k)} g^{(k),\text{loc}}$	// Thm. 5.4, $\mathcal{O}(\mathcal{I}^{(k)} \rho_k^d \ln \frac{1}{\epsilon})$
10: For $i \in \mathcal{J}^{(k)}$, $\chi_i^{(k),\text{loc}} = \sum_{j \in \mathcal{I}^{(k)}} W_{i,j}^{(k)} \psi_j^{(k),\text{loc}}$	// $\mathcal{O}(\mathcal{I}^{(k)} \rho_k^d)$
11: $u^{(k),\text{loc}} - u^{(k-1),\text{loc}} = \sum_{i \in \mathcal{J}^{(k)}} w_i^{(k),\text{loc}} \chi_i^{(k),\text{loc}}$	// $\mathcal{O}(N \rho_k^d)$
12: $\text{Inv}(B^{(k),\text{loc}} D^{(k,k-1),\text{loc}}) = -W^{(k)} A^{(k),\text{loc}} \bar{\pi}^{(k,k-1)}, \rho_{k-1}$	// Def. 5.2, Thm. 5.4, $\mathcal{O}(\mathcal{I}^{(k)} \rho_{k-1}^d \rho_k^d \ln \frac{1}{\epsilon})$
13: $R^{(k-1,k),\text{loc}} = \bar{\pi}^{(k-1,k)} + D^{(k-1,k),\text{loc}} W^{(k)}$	// Def. 5.2, $\mathcal{O}(\mathcal{I}^{(k-1)} \rho_{k-1}^d)$
14: $A^{(k-1),\text{loc}} = R^{(k-1,k),\text{loc}} A^{(k),\text{loc}} R^{(k,k-1),\text{loc}}$	// $\mathcal{O}(\mathcal{I}^{(k-1)} \rho_{k-1}^{2d} \rho_k^d)$
15: For $i \in \mathcal{I}^{(k-1)}$, $\psi_i^{(k-1),\text{loc}} = \sum_{j \in \mathcal{I}^{(k)}} R_{i,j}^{(k-1,k),\text{loc}} \psi_j^{(k),\text{loc}}$	// $\mathcal{O}(\mathcal{I}^{(k-1)} \rho_{k-1}^d)$
16: $g^{(k-1),\text{loc}} = R^{(k-1,k),\text{loc}} g^{(k),\text{loc}}$	// $\mathcal{O}(\mathcal{I}^{(k-1)} \rho_{k-1}^d)$
17: end for	
18: $U^{(1),\text{loc}} = A^{(1),\text{loc},-1} g^{(1),\text{loc}}$	// $\mathcal{O}(\mathcal{I}^{(1)} \rho_1^d \ln \frac{1}{\epsilon})$
19: $u^{(1),\text{loc}} = \sum_{i \in \mathcal{I}^{(1)}} U_i^{(1)} \psi_i^{(1)}$	// $\mathcal{O}(N \rho_1^d)$
20: $u^{\text{loc}} = u^{(1),\text{loc}} + (u^{(2),\text{loc}} - u^{(1),\text{loc}}) + \dots + (u^{(q),\text{loc}} - u^{(q-1),\text{loc}})$	// $\mathcal{O}(Nq)$

requires lower accuracy and therefore smaller localization radii), we will restrict our analysis and presentation to using Algorithm 2 as a direct solver. Note that, when used as a direct solver, Algorithm 2 is parallel both in space (via localization) and in bandwidth/subscale (subscales can be computed independently from each other and $\psi_i^{(k-1),\text{loc}}$ and $u^{(k),\text{loc}} - u^{(k-1),\text{loc}}$ can be resolved in parallel). We will base our analysis on the results of section 4.10 and, in particular, Theorem 4.25. Although obtained in a continuous setting, these results can be generalized to the discrete setting without difficulty. Two small differences are worth mentioning. (1) In this discrete setting, an alternative approach for obtaining localization error bounds in the first step of the algorithm (the computation of the localized gamblets $\psi_i^{(q),\text{loc}}$) is to use the exponential decay property of the inverse of symmetric well-conditioned banded matrices [26]: since M is banded and of uniformly bounded condition number, [26] (see also [11, Thm. 4.10]) implies that $M_{i,j}^{-1}$ decays like $\exp(-\text{dist}(\tau_i^{(q)}, \tau_j^{(q)})/C)$, which guarantees that the bound $\mathcal{E}(q) \leq CH^{-d/2-q(2+d/2)} e^{-\rho_q/C}$ (used in Theorem 4.24) remains valid in the discrete setting. (2) Since the basis functions φ_i are not exact set functions, neither are the resulting aggregates $\phi_i^{(k)}$. This implies that, in the discrete setting, $\int_{\Omega} \psi_i^{(k),\text{loc}} \phi_j^{(k)}$ is not necessarily equal to zero if $\tau_j^{(k)}$ is adjacent to $S_{\rho_k}^i$ (with $j \notin \mathcal{I}^{\rho_k}$, using the notation of section 4.10). This, however, does not prevent the generalization of the proof because the value of $\int_{\Omega} \psi_i^{(k),\text{loc}} \phi_j^{(k)}$ (when $\tau_j^{(k)}$ is adjacent to $S_{\rho_k}^i$) can be controlled via the exponential decay of the basis functions (e.g., as is done in the proof of Theorem 3.11). We will summarize this generalization in the

following theorem (where the constant C depends on the constants $C_1, C_0, \bar{\gamma}$, and γ associated with the finite elements (φ_i) in (5.1), in addition to $d, \Omega, \lambda_{\min}(a), \lambda_{\max}(a)$, and δ).

THEOREM 5.3. *Let u be the solution of the discrete system (5.3). Let $u^{(1),\text{loc}}, u^{(k),\text{loc}} - u^{(k-1),\text{loc}}, u^{\text{loc}}, A^{(k),\text{loc}}$, and $B^{(k),\text{loc}}$ be the outputs of Algorithm 2. Let $u^{(1)}$ and $u^{(k)} - u^{(k-1)}$ be the outputs of Algorithm 1. For $k \in \{2, \dots, q\}$, write $u^{(k),\text{loc}} := u^{(1),\text{loc}} + \sum_{j=2}^k (u^{(j),\text{loc}} - u^{(j-1),\text{loc}})$. Let $\epsilon \in (0, 1)$. It holds true that if $\rho_k \geq C((1 + \frac{1}{\ln(1/H)}) \ln \frac{1}{H^k} + \ln \frac{1}{\epsilon})$ for $k \in \{1, \dots, q\}$, then (1) for $k \in \{1, \dots, q-1\}$ we have $\|u^{(k)} - u^{(k),\text{loc}}\|_{H_0^1(\Omega)} \leq \epsilon \|g\|_{H^{-1}(\Omega)}$ and $\|u^{(k)} - u^{(k),\text{loc}}\|_{H_0^1(\Omega)} \leq C(H^k + \epsilon) \|g\|_{L^2(\Omega)}$, (2) $\text{Cond}(A^{(1),\text{loc}}) \leq CH^{-2}$, and for $k \in \{2, \dots, q\}$ we have (3) $\text{Cond}(B^{(k),\text{loc}}) \leq CH^{-2-2d}$ and (4) $\|u^{(k)} - u^{(k-1)} - (u^{(k),\text{loc}} - u^{(k-1),\text{loc}})\|_{H_0^1(\Omega)} \leq \frac{\epsilon}{2k^2} \|g\|_{H^{-1}(\Omega)}$. Finally, (5) $\|u - u^{\text{loc}}\|_{H_0^1(\Omega)} \leq \epsilon \|g\|_{H^{-1}(\Omega)}$.*

Therefore, according to Theorem 5.3 if the localization radii ρ_k are chosen so that $\rho_k = \mathcal{O}(\ln \max(1/\epsilon, 1/H_k))$ for $k \in \{1, \dots, q\}$, then the condition numbers of the matrices $B^{(k),\text{loc}}$ and $A^{(1),\text{loc}}$ remain uniformly bounded and the algorithm achieves accuracy ϵ in a direct solve. The following theorem shows that the linear systems appearing in lines 3, 9, and 12 of Algorithm 2 do not need to be solved exactly and provide bounds on the accuracy requirements (to simplify notation, we will from now on drop the superscripts of the vectors y and b appearing in Definitions 5.1 and 5.2).

THEOREM 5.4. *The results of Theorem 5.3 remain true if (1) $\rho_k \geq C((1 + \frac{1}{\ln(1/H)}) \ln \frac{1}{H^k} + \ln \frac{1}{\epsilon})$ for $k \in \{1, \dots, q\}$; (2) for each $i \in \mathcal{I}^{(q)}$ the localized linear system $M^{i,\rho_q} y = \delta_{\cdot,i}$ of Definition 5.1 and line 3 of Algorithm 2 is solved up to accuracy $|y - y^{\text{app}}|_{M^{i,\rho_q}} \leq C^{-1} H^{7d/2+3} \epsilon / q^2$ (using the notations of section 4.9, i.e., $|e|_A^2 := e^T A e$, and writing y^{app} as the approximation of y); (3) for $k \in \{2, \dots, q\}$ and each $i \in \mathcal{I}^{(k-1)}$, the localized linear system $B^{(i,\rho)} y = b$ of Definition 5.2 and line 12 of Algorithm 2 is solved up to accuracy $|y - y^{\text{app}}|_{B^{(i,\rho)}} \leq C^{-1} H^{7d/2+4} \epsilon / (k-1)^2$; (4) for $k \in \{2, \dots, q\}$, the linear system $B^{(k),\text{loc}} y = W^{(k)} g^{(k),\text{loc}}$ of line 9 of Algorithm 2 is solved up to accuracy $|y - y^{\text{app}}|_{B^{(k),\text{loc}}} \leq \epsilon \|g\|_{H^{-1}(\Omega)} / (2q)$.*

Proof. From the proof of Theorem 4.25 we need $\mathcal{E}(k) \leq C^{-1} H^{-k(1+d/2)+7d/2+3} \epsilon / k^2$ for $k \in \{1, \dots, q\}$. By the inverse Poincaré inequality (5.2) this inequality is satisfied for $k = q$ for $\|\psi_i^{(q)} - \psi_i^{(q),\text{loc}}\|_{L^2(\Omega)} \leq C^{-1} H^{7d/2+3} \epsilon / q^2$ for each $i \in \mathcal{I}^{(q)}$, which by the definition of M^{i,ρ_q} and line 4 of Algorithm 2 leads to (2). For $k \in \{2, \dots, q\}$ the inequality $\mathcal{E}(k-1) \leq C^{-1} H^{-(k-1)(1+d/2)+7d/2+3} \epsilon / (k-1)^2$ is satisfied if for $i \in \mathcal{I}^{(k-1)}$, $\|\psi_i^{(k-1)} - \psi_i^{(k-1),\text{loc}}\|_a \leq C^{-1} H^{-(k-1)+7d/2+3} \epsilon / (k-1)^2$. Using the notations of Definition 5.2 we have, $\psi_i^{(k-1),\text{loc}} = \sum_{j \in \mathcal{I}^{(k)}} \bar{\pi}_{i,j}^{(k-1,k)} \psi_j^{(k),\text{loc}} + \sum_{j \in i^{\times}} D_{i,j}^{(k-1,k),\text{loc}} \chi_j^{(k),\text{loc}}$ with $\langle \chi_j^{(k),\text{loc}}, \chi_l^{(k),\text{loc}} \rangle_a = B_{j,l}^{(i,\rho)}$, which leads to (3) by lines 15, 13, 10, and 8 of Algorithm 2. For (4) we simply observe that for $y \in \mathcal{J}^{(k)}$, $\|\sum_{i \in \mathcal{J}^{(k)}} (y - y^{\text{app}})_i \chi_i^{(k),\text{loc}}\|_a = |y - y^{\text{app}}|_{B^{(k),\text{loc}}}$ \square

Let us now describe the complexity of Algorithm 2. This complexity depends on the desired accuracy $\epsilon \in (0, 1)$. Lines 1 and 2 correspond to the computation of the (sparse) mass and stiffness matrices of (5.3). Note that since A and M are sparse and banded (of bandwidth $2d = 4$ in our numerical example), this computation is of $\mathcal{O}(N)$ complexity. Line 3 corresponds to the resolution of the localized linear system introduced in Definition 5.1 using M^{i,ρ_q} , the $i^{\rho_q} \times i^{\rho_q}$ submatrix of M . According to Theorem 5.3, the accuracy of each solve must be $|y - y^{\text{app}}|_{M^{i,\rho_q}} \leq C^{-1} H^{7d/2+3} \epsilon / q^2$. Since

$|i^{\rho_q}| = \mathcal{O}(\rho_q^d)$ and since M^{i,ρ_q} is of condition number bounded by that of M , for each i , the linear system of line 3 can be solved efficiently (to accuracy $\mathcal{O}(C^{-1}H^{7d/2+3}\epsilon/q^2)$) using $\mathcal{O}(\rho_q) = \mathcal{O}(\ln \max(\frac{1}{\epsilon}, q))$ iterations of the CG method (c.f. section 4.9) with a cost of $\mathcal{O}(\rho_q^d)$ per iteration, which results in a total cost of $\mathcal{O}(N\rho_q^d \ln \max(\frac{1}{\epsilon}, q))$. Lines 4 and 5 are naturally of complexity $\mathcal{O}(N\rho_q^d)$. Since $A_{i,j}^{(q),\text{loc}} = 0$ if $\tau_i^{(q)}$ and $\tau_j^{(q)}$ are at a distance larger than $2H^q\rho_q$, the complexity of line 6 is $\mathcal{O}(N\rho_q^{2d})$. Note that $A^{(k),\text{loc}}$ and $B^{(k),\text{loc}}$ are banded and of bandwidth $\mathcal{O}(N\rho_k^d)$. It follows that line 8 is of complexity $\mathcal{O}(|\mathcal{I}^{(k)}|\rho_k^d)$. According to Theorem 5.4 the linear system of line 9 needs to be solved up to accuracy $|y - y^{\text{app}}|_{B^{(k),\text{loc}}} \leq \epsilon \|g\|_{H^{-1}(\Omega)}/2$. Since $B^{(k),\text{loc}}$ is of uniformly bounded condition number this can be done using $\mathcal{O}(\ln \frac{1}{\epsilon})$ iterations of the CG method with a cost of $\mathcal{O}(|\mathcal{I}^{(k)}|\rho_k^d)$ per iteration (using $\mathcal{O}(|\mathcal{J}^{(k)}|) = \mathcal{O}(|\mathcal{I}^{(k)}|)$), which results in a total cost of $\mathcal{O}(|\mathcal{I}^{(k)}|\rho_k^d \ln \frac{1}{\epsilon})$ for line 9. Storing the fine mesh values of $u^{(k),\text{loc}} - u^{(k-1),\text{loc}}$ in line 11 costs $\mathcal{O}(N\rho_k^d)$ (since for each node x on the fine mesh only $\mathcal{O}(\rho_k^d)$ localized basis functions contribute to the value of $u^{(k),\text{loc}} - u^{(k-1),\text{loc}}$). According to Theorem 5.4, for each $i \in \mathcal{I}^{(k-1)}$ the linear system $B^{(i,\rho)}y = b$ of line 12 needs to be solved up to accuracy $|y - y^{\text{app}}|_{B^{(i,\rho)}} \leq C^{-1}H^{-k+7d/2+4}\epsilon/(k-1)^2$. Since the matrix $B^{(i,\rho)}$ inherits the uniformly bounded condition number from $B^{(k),\text{loc}}$, this can be done using $\mathcal{O}(\ln \frac{1}{\epsilon})$ iterations of the CG method with a cost of $\mathcal{O}(H^{-d}\rho_{k-1}^d\rho_k^d) = \mathcal{O}(\rho_{k-1}^d\rho_k^d)$ per iteration. This results in a total cost of $\mathcal{O}(|\mathcal{I}^{(k)}|\rho_{k-1}^d\rho_k^d \ln \frac{1}{\epsilon})$ for line 12. We obtain, using the sparsity structures of $D^{(k-1,k),\text{loc}}$ and $R^{(k-1,k),\text{loc}}$, that the complexity of line 13 is $\mathcal{O}(|\mathcal{I}^{(k-1)}|\rho_{k-1}^d H^{-d}) = \mathcal{O}(|\mathcal{I}^{(k-1)}|\rho_{k-1}^d)$ and that of line 14 is $\mathcal{O}(|\mathcal{I}^{(k-1)}|\rho_{k-1}^{2d}\rho_k^d)$. The complexity of lines 15 to 16 is summarized in the display of Algorithm 2 and a simple consequence of the sparsity structure of $R^{(k-1,k),\text{loc}}$. Line 18 is of complexity $\mathcal{O}(|\mathcal{I}^{(1)}|\rho_1^d \ln \frac{1}{\epsilon})$ (using CG as in line 9). As in line 11, storing the values of $u^{(1),\text{loc}}$ costs $\mathcal{O}(N\rho_1^d)$. Finally, obtaining u^{loc} in line 20 costs $\mathcal{O}(Nq)$ (observe that $q = \mathcal{O}(\ln N)$).

Table 1 Complexity of Algorithm 2.

Compute and store $\psi_i^{(k),\text{loc}}, \chi_i^{(k),\text{loc}}, A^{(k),\text{loc}}, B^{(k),\text{loc}}$, and u^{loc} s.t. $\ u - u^{\text{loc}}\ _{H_0^1(\Omega)} \leq \epsilon \ g\ _{H^{-1}(\Omega)}$	$\epsilon \leq H^q$	$\epsilon \geq H^q$
First solve	$N \ln^{3d} \frac{1}{\epsilon}$	$N \ln^{3d} N$
Subsequence solves	$N \ln^{d+1} \frac{1}{\epsilon}$	$N \ln^d N \ln \frac{1}{\epsilon}$
Subsequent solves to compute $u^{(k),\text{loc}}$ s.t. $\ u - u^{(k),\text{loc}}\ _{H_0^1(\Omega)} \leq C\epsilon \ g\ _{L^2(\Omega)}$		$N \ln^{d+1} \frac{1}{\epsilon}$
Subsequent solves to compute the coefficients $c_i^{(k)}$ of $u^{(k),\text{hom}} = \sum_{i \in \mathcal{I}^{(k)}} c_i^{(k)} \psi_i^{(k)}$ s.t. $\ u - u^{(k),\text{loc}}\ _{H_0^1(\Omega)} \leq C\epsilon(\ g\ _{L^2(\Omega)} + \ g\ _{\text{Lip}})$		$\epsilon^{-d} \ln^{d+1} \frac{1}{\epsilon}$
Subsequent solves to compute $u^{(k),\text{hom}}$ s.t. $\ u - u^{(k),\text{hom}}\ _{L^2(\Omega)} \leq C\epsilon(\ g\ _{L^2(\Omega)} + \ g\ _{\text{Lip}})$		$\epsilon^{-d} \ln^{d+1} \frac{1}{\epsilon}$
a periodic/ergodic with mixing length $HP \leq \epsilon$, first solve of $u^{(k),\text{hom}}$ s.t. $\ u - u^{(k),\text{hom}}\ _{L^2(\Omega)} \leq C\epsilon(\ g\ _{L^2(\Omega)} + \ g\ _{\text{Lip}})$		$(N \ln^{3d} N) H^p + \epsilon^{-d} \ln^{d+1} \frac{1}{\epsilon}$

Total Computational Complexity: First Solve. Summarizing, we obtain that the complexity of Algorithm 2, i.e., the cost of computing the gamblets $(\psi_i^{(k),\text{loc}}, \chi_j^{(k),\text{loc}})$, their stiffness matrices $(A^{(k),\text{loc}}, B^{(k),\text{loc}})$, and the approximation u^{loc} such that $\|u - u^{\text{loc}}\|_{H_0^1(\Omega)} \leq \epsilon \|g\|_{H^{-1}(\Omega)}$ is $\mathcal{O}(N(\ln \max(\frac{1}{\epsilon}, N^{\frac{1}{d}}))^{3d})$ (with line 14 being the corresponding complexity bottleneck). The complexity of storing the gamblets $(\psi_i^{(1),\text{loc}}, \chi_j^{(k),\text{loc}})$, and their stiffness matrices $(A^{(1),\text{loc}}, B^{(k),\text{loc}})$ is $\mathcal{O}(N(\ln \max(\frac{1}{\epsilon}, N^{\frac{1}{d}}))^d)$.

Computational Complexity of Subsequent Solves with $g \in H^{-1}(\Omega)$. If (5.3) (i.e., (1.1)) needs to be solved for more than one g , then the gamblets $\psi_i^{(k),\text{loc}}, \chi_i^{(k),\text{loc}}$ and the stiffness matrices $B^{(k),\text{loc}}$ do not need to be recomputed. The cost of subsequent solves is therefore that of line 9, i.e., $\mathcal{O}(N(\ln \max(\frac{1}{\epsilon}, N^{\frac{1}{d}}))^d \ln \frac{1}{\epsilon})$ to achieve the approximation accuracy $\|u - u^{\text{loc}}\|_{H_0^1(\Omega)} \leq \epsilon \|g\|_{H^{-1}(\Omega)}$.

Computational Complexity of Subsequent Solves with $g \in L^2(\Omega)$ and $\epsilon \geq H^q$. If $g \in L^2(\Omega)$ (i.e., if $\|g\|_{L^2(\Omega)}$ is used to express the accuracy of the approximation) and $\epsilon \in [H^k, H^{k-1}]$, then, by Theorem 5.3, $u^{(k),\text{loc}}$ achieves the approximation accuracy $\|u - u^{(k),\text{loc}}\|_{H_0^1(\Omega)} \leq C\epsilon \|g\|_{L^2(\Omega)}$ (i.e., $u^{(j+1),\text{loc}} - u^{(j),\text{loc}}$ does not need to be computed for $j \geq k$) and the corresponding complexity is $\mathcal{O}(N(\ln \frac{1}{\epsilon})^{d+1})$ (if $g \in H^{-1}(\Omega)$, then the energy of the solution can be in the fine scales and $u^{(j+1),\text{loc}} - u^{(j),\text{loc}}$ do need to be computed for $j \geq k$).

Computational Complexity of Subsequent Solves with g Lipschitz Continuous and $\epsilon \geq H^q$. Note that the computational complexity bottleneck for computing the coefficients of $u^{(k),\text{loc}}$ in the basis $(\psi_i^{(k),\text{loc}})$, when $g \in \mathcal{L}^2(\Omega)$ and $\epsilon \in [H^k, H^{k-1}]$, is in the computation of the vectors $g^{(j),\text{loc}}$ for $j > k$. If g is Lipschitz continuous, then let $g_i^{(k),\text{loc}}$ be approximated as $g(x_i^{(k)})$, where $x_i^{(k)}$ is any point in $\tau_i^{(k)}$ without loss of accuracy. Note that this approximation requires (only) $\mathcal{O}(H^{-kd})$ evaluations of g and leads to a corresponding $u^{(k),\text{loc}}$ satisfying $\|u - u^{(k),\text{loc}}\|_a \leq C\epsilon(\|g\|_{L^2(\Omega)} + \|g\|_{\text{Lip}})$ (with $\|g\|_{\text{Lip}} = \sup_{x,y \in \Omega} |g(x) - g(y)|/|x - y|$). Therefore, the computational complexity of subsequent solves to obtain the coefficients $c_i^{(k)}$ in the decomposition $u^{(k),\text{loc}} = \sum_{i \in \mathcal{I}^{(k)}} c_i^{(k)} \psi_i^{(k),\text{loc}}$ is $\mathcal{O}(\epsilon^{-d} (\ln \frac{1}{\epsilon})^{d+1})$ (i.e., independent of N if g is Lipschitz continuous). Of course, obtaining an $H_0^1(\Omega)$ -norm approximation of u with accuracy H^k requires expressing the values of $\psi_i^{(k),\text{loc}}$ (and therefore $u^{(k),\text{loc}}$) on the fine mesh, which leads to a total cost of $\mathcal{O}(N(\ln \frac{1}{\epsilon})^d)$. However, if one is only interested in expressing the values of $u^{(k),\text{loc}}$ on the fine mesh in a subdomain of diameter ϵ , then the resulting complexity is $\mathcal{O}((N\epsilon^d + \epsilon^{-d})(\ln \frac{1}{\epsilon})^d)$.

Computational Complexity of Subsequent L^2 -Approximations with g Lipschitz Continuous and $\epsilon \geq H^q$. Let $(x_i^{(k)})_{i \in \mathcal{I}^{(k)}}$ be points of $(\tau_i^{(k)})_{i \in \mathcal{I}^{(k)}}$ forming a regular coarse mesh of Ω of resolution H^k and write $\varphi_i^{(k)}$ as the corresponding (regular and coarse) piecewise-linear nodal basis elements. If (as in classical homogenization or the heterogeneous multiscale method (HMM)) one is only interested in an L^2 -norm approximation of u with accuracy H^k , then the coefficients $c_i^{(k)}$ defined above are sufficient to obtain the approximation $u^{\text{hom}} = \sum_{i \in \mathcal{I}^{(k)}} \frac{c_i^{(k)}}{\int_{\Omega} \phi_i^{(k)}} \varphi_i^{(k)}$ that satisfies $\|u^{(k),\text{loc}} - u^{\text{hom}}\|_{L^2(\Omega)} \leq CH^k \|u^{(k),\text{loc}}\|_{H_0^1(\Omega)}$ ($\int_{\Omega} u^{\text{hom}} \phi_i^{(k)} = \int_{\Omega} u^{(k),\text{loc}} \phi_i^{(k)}$) and

therefore $\|u - u^{(k),\text{hom}}\|_{L^2(\Omega)} \leq C\epsilon(\|g\|_{L^2(\Omega)} + \|g\|_{\text{Lip}})$. Note that the computational complexity of subsequent solves to obtain u^{hom} is $\mathcal{O}(\epsilon^{-d}(\ln \frac{1}{\epsilon})^{d+1})$.

Total Computational Complexity if a Is Periodic or Ergodic with Mixing Length H^p and $\epsilon \approx H^k$ with $k \geq p$. Under the assumptions of classical homogenization or the HMM [33] (e.g., a is of period H^p or a is ergodic with H^p as mixing length), if the sets $\tau_i^{(k)}$ are chosen to match the period of a and the domain is rescaled so that $1/H$ is an integer, then the entries of $A^{(k)}$ are invariant under periodic translations (or stationary if the medium is ergodic). Therefore, under these assumptions, as in classical homogenization, it is sufficient to limit the computation of gamblets to periodicity cells (or ergodicity cells with a tight control on mixing as in [47]). The resulting cost of obtaining $u^{(k),\text{hom}}$ (in a first solve) such that $\|u^{(k),\text{loc}} - u^{(k),\text{hom}}\|_{L^2(\Omega)} \leq C\epsilon(\|g\|_{L^2(\Omega)} + \|g\|_{\text{Lip}})$ is $\mathcal{O}(N \ln^{3d} N H^p + \epsilon^{-d} \ln^{d+1} \frac{1}{\epsilon})$.

Acknowledgments. The author thanks C. Scovel, L. Zhang, P. B. Bochev, P. S. Vassilevski, J.-L. Cambier, B. Suter, G. Pavliotis, and an anonymous referee for valuable comments and suggestions.

REFERENCES

- [1] A. ABDULLE AND C. SCHWAB, *Heterogeneous multiscale FEM for diffusion problems on rough surfaces*, Multiscale Model. Simul., 3 (2005), pp. 195–220, <https://doi.org/10.1137/030600771>. (Cited on p. 102)
- [2] G. ALLAIRE AND R. BRIZZI, *A multiscale finite element method for numerical homogenization*, Multiscale Model. Simul., 4 (2005), pp. 790–812, <https://doi.org/10.1137/040611239>. (Cited on p. 102)
- [3] A. AVERBUCH, G. BEYLKIN, R. COIFMAN, AND M. ISRAELI, *Multiscale inversion of elliptic operators*, in Signal and Image Representation in Combined Spaces, Wavelet Anal. Appl. 7, Academic Press, San Diego, CA, 1998, pp. 341–359. (Cited on p. 101)
- [4] 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. (Cited on p. 102)
- [5] I. BABUŠKA 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, <https://doi.org/10.1137/100791051>. (Cited on pp. 102, 117)
- [6] 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, <https://doi.org/10.1137/0720034>. (Cited on p. 102)
- [7] I. BABUŠKA AND J. E. OSBORN, *Can a finite element method perform arbitrarily badly?*, Math. Comp., 69 (2000), pp. 443–462. (Cited on pp. 101, 102)
- [8] G. BAL AND W. JING, *Corrector theory for MsFEM and HMM in random media*, Multiscale Model. Simul., 9 (2011), pp. 1549–1587, <https://doi.org/10.1137/100815918>. (Cited on p. 102)
- [9] R. E. BANK, T. F. DUPONT, AND H. YSERENTANT, *The hierarchical basis multigrid method*, Numer. Math., 52 (1988), pp. 427–458, <https://doi.org/10.1007/BF01462238>. (Cited on p. 101)
- [10] M. BEBENDORF, *Efficient inversion of the Galerkin matrix of general second-order elliptic operators with nonsmooth coefficients*, Math. Comp., 74 (2005), pp. 1179–1199, <https://doi.org/10.1090/S0025-5718-04-01716-8>. (Cited on p. 101)
- [11] M. BEBENDORF, *Hierarchical Matrices: A Means to Efficiently Solve Elliptic Boundary Value Problems*, Lect. Notes Comput. Sci. Engrg. 63, Springer-Verlag, Berlin, 2008. (Cited on pp. 101, 140)
- [12] G. BEN AROUS AND H. OWHADI, *Multiscale homogenization with bounded ratios and anomalous slow diffusion*, Comm. Pure Appl. Math., 56 (2003), pp. 80–113, <https://doi.org/10.1002/cpa.10053>. (Cited on pp. 102, 113, 115)
- [13] A. BENSOUSSAN, J. L. LIONS, AND G. PAPANICOLAOU, *Asymptotic Analysis for Periodic Structure*, North Holland, Amsterdam, 1978. (Cited on p. 102)

- [14] 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. (Cited on pp. 102, 109, 113)
- [15] G. BEYLKIN AND N. COULT, *A multiresolution strategy for reduction of elliptic PDEs and eigenvalue problems*, Appl. Comput. Harmon. Anal., 5 (1998), pp. 129–155. (Cited on p. 101)
- [16] A. BRANDT, *Multi-level adaptive technique (MLAT) for fast numerical solutions to boundary value problems*, in Proceedings of the Third International Conference on Numerical Methods in Fluid Mechanics, Lecture Notes in Phys. 18, Springer-Verlag, Berlin, 1973, pp. S-89. (Cited on p. 101)
- [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. (Cited on pp. 101, 102)
- [18] M. E. BREWSTER AND G. BEYLKIN, *A multiresolution strategy for numerical homogenization*, Appl. Comput. Harmon. Anal., 2 (1995), pp. 327–349. (Cited on pp. 101, 102)
- [19] F.-X. BRIOL, C. J. OATES, M. GIROLAMI, M. A. OSBORNE, AND D. SEJDINOVIC, *Probabilistic Integration: A Role for Statisticians in Numerical Analysis?*, preprint, arXiv:1512.00933 [stat.ML], 2015. (Cited on p. 105)
- [20] E. J. CANDÈS, J. ROMBERG, AND T. TAO, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509, <https://doi.org/10.1109/TIT.2005.862083>. (Cited on p. 109)
- [21] E. J. CANDÈS AND T. TAO, *Near-optimal signal recovery from random projections: Universal encoding strategies?*, IEEE Trans. Inform. Theory, 52 (2006), pp. 5406–5425, <https://doi.org/10.1109/TIT.2006.885507>. (Cited on p. 109)
- [22] V. CHANDRASEKARAN, S. SANGHAVI, P. A. PARRILO, AND A. S. WILLSKY, *Rank-sparsity incoherence for matrix decomposition*, SIAM J. Optim., 21 (2011), pp. 572–596, <https://doi.org/10.1137/090761793>. (Cited on p. 109)
- [23] O. A. CHKREBTII, D. A. CAMPBELL, B. CALDERHEAD, AND M. A. GIROLAMI, *Bayesian solution uncertainty quantification for differential equations*, Bayesian Anal., 11 (2016), pp. 1239–1267. (Cited on p. 105)
- [24] E. CHOW AND P. S. VASSILEVSKI, *Multilevel block factorizations in generalized hierarchical bases*, Numer. Linear Algebra Appl., 10 (2003), pp. 105–127, <https://doi.org/10.1002/nla.314>. (Cited on p. 101)
- [25] P. R. CONRAD, M. GIROLAMI, S. SÄRKKÄ, A. STUART, AND K. ZYGALAKIS, *Probability measures for numerical solutions of differential equations*, Statist. Comput., to appear; preprint, arXiv:1506.04592 [stat.ME], 2016. (Cited on p. 105)
- [26] S. DEMKO, W. F. MOSS, AND P. W. SMITH, *Decay rates for inverses of band matrices*, Math. Comp., 43 (1984), pp. 491–499. (Cited on p. 140)
- [27] 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. (Cited on pp. 103, 104, 105)
- [28] D. L. DONOHO, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306, <https://doi.org/10.1109/TIT.2006.871582>. (Cited on p. 109)
- [29] M. DOROBANTU AND B. ENGQUIST, *Wavelet-based numerical homogenization*, SIAM J. Numer. Anal., 35 (1998), pp. 540–559, <https://doi.org/10.1137/S0036142996298880>. (Cited on p. 101)
- [30] 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. (Cited on pp. 103, 114)
- [31] 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. (Cited on pp. 103, 114)
- [32] 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. (Cited on pp. 103, 114)
- [33] W. E AND B. ENGQUIST, *The heterogeneous multiscale methods*, Commun. Math. Sci., 1 (2003), pp. 87–132, <https://projecteuclid.org/euclid.cms/1118150402>. (Cited on pp. 102, 144)
- [34] Y. EFENDIEV, J. GALVIS, AND P. S. VASSILEVSKI, *Spectral element agglomerate algebraic multigrid methods for elliptic problems with high-contrast coefficients*, in Domain Decomposition Methods in Science and Engineering XIX, Lect. Notes Comput. Sci. Eng. 78, Springer, Heidelberg, 2011, pp. 407–414, https://doi.org/10.1007/978-3-642-11304-8_47. (Cited on p. 101)

- [35] 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. (Cited on p. 102)
- [36] Y. EFENDIEV AND T. HOU, *Multiscale finite element methods for porous media flows and their applications*, Appl. Numer. Math., 57 (2007), pp. 577–596. (Cited on p. 102)
- [37] B. ENGQUIST AND E. LUO, *Convergence of a multigrid method for elliptic equations with highly oscillatory coefficients*, SIAM J. Numer. Anal., 34 (1997), pp. 2254–2273, <https://doi.org/10.1137/S0036142995289408>. (Cited on p. 101)
- [38] B. ENGQUIST AND O. RUNBORG, *Wavelet-based numerical homogenization with applications*, in Multiscale and Multiresolution Methods, Lect. Notes Comput. Sci. Eng. 20, Springer, Berlin, 2002, pp. 97–148, https://doi.org/10.1007/978-3-642-56205-1_2. (Cited on p. 101)
- [39] B. ENGQUIST AND P. E. SOUGANIDIS, *Asymptotic and numerical homogenization*, Acta Numer., 17 (2008), pp. 147–190. (Cited on p. 102)
- [40] G. E. FASSHAUER, *Meshfree methods*, in Handbook of Theoretical and Computational Nanotechnology, American Scientific, 2005, pp. 33–97. (Cited on p. 108)
- [41] R. P. FEDORENKO, *A relaxation method of solution of elliptic difference equations*, Ž. Vyčisl. Mat. i Mat. Fiz., 1 (1961), pp. 922–927. (Cited on p. 101)
- [42] A. FRIEZE, R. KANNAN, AND S. VEMPALA, *Fast Monte-Carlo algorithms for finding low-rank approximations*, in Proceedings of the 39th Annual IEEE Symposium on Foundations of Computer Science, IEEE, 1998, pp. 370–378. (Cited on p. 109)
- [43] A. C. GILBERT, S. GUHA, P. INDYK, S. MUTHUKRISHNAN, AND M. STRAUSS, *Near-optimal sparse Fourier representations via sampling*, in Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing, ACM, New York, 2002, pp. 152–161, <https://doi.org/10.1145/509907.509933>. (Cited on p. 109)
- [44] A. C. GILBERT, M. J. STRAUSS, J. A. TROPP, AND R. VERSHYNIN, *One sketch for all: Fast algorithms for compressed sensing*, in STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing, ACM, New York, 2007, pp. 237–246, <https://doi.org/10.1145/1250790.1250824>. (Cited on p. 109)
- [45] E. D. GIORGI, *Sulla convergenza di alcune successioni di integrali del tipo dell'aera*, Rendiconti di Mat., 8 (1975), pp. 277–294. (Cited on p. 102)
- [46] A. GLORIA, *An analytical framework for the numerical homogenization of monotone elliptic operators and quasiconvex energies*, Multiscale Model. Simul., 5 (2006), pp. 996–1043, <https://doi.org/10.1137/060649112>. (Cited on p. 102)
- [47] A. GLORIA, S. NEUKAMM, AND F. OTTO, *Quantification of ergodicity in stochastic homogenization: Optimal bounds via spectral gap on Glauber dynamics*, Invent. Math., 199 (2015), pp. 455–515, <https://doi.org/10.1007/s00222-014-0518-z>. (Cited on p. 144)
- [48] 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, <https://doi.org/10.1137/11082138X>. (Cited on p. 102)
- [49] L. GREENGARD AND V. ROKHLIN, *A fast algorithm for particle simulations*, J. Comput. Phys., 73 (1987), pp. 325–348, [https://doi.org/10.1016/0021-9991\(87\)90140-9](https://doi.org/10.1016/0021-9991(87)90140-9). (Cited on p. 101)
- [50] W. HACKBUSCH, *A fast iterative method for solving Poisson's equation in a general region*, in Numerical Treatment of Differential Equations (Proc. Conf., Math. Forschungsinst., Oberwolfach, 1976), Lecture Notes in Math. 631, Springer, Berlin, 1978, pp. 51–62. (Cited on p. 101)
- [51] W. HACKBUSCH, *Multigrid Methods and Applications*, Springer Ser. Comput. Math. 4, Springer-Verlag, Berlin, 1985, <https://doi.org/10.1007/978-3-662-02427-0>. (Cited on p. 101)
- [52] W. HACKBUSCH, L. GRASEDYCK, AND S. BÖRM, *An introduction to hierarchical matrices*, Math. Bohem., 127 (2002), pp. 229–241. (Cited on p. 101)
- [53] N. HALKO, P. G. MARTINSSON, AND J. A. TROPP, *Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions*, SIAM Rev., 53 (2011), pp. 217–288, <https://doi.org/10.1137/090771806>. (Cited on pp. 108, 109)
- [54] R. L. HARDER AND R. N. DESMARAIS, *Interpolation using surface splines*, J. Aircraft, 9 (1972), pp. 189–191. (Cited on pp. 103, 114)
- [55] P. HENNIG, *Probabilistic interpretation of linear solvers*, SIAM J. Optim., 25 (2015), pp. 234–260, <https://doi.org/10.1137/140955501>. (Cited on p. 105)
- [56] P. HENNIG, M. A. OSBORNE, AND M. GIROLAMI, *Probabilistic numerics and uncertainty in computations*, Proc. A., 471 (2015), 20150142, <https://doi.org/10.1098/rspa.2015.0142>. (Cited on p. 105)
- [57] M. R. HESTENES AND E. STIEFEL, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards, 49 (1952), pp. 409–436. (Cited on p. 128)

- [58] T. Y. HOU AND P. LIU, *Optimal local multi-scale basis functions for linear elliptic equations with rough coefficient*, Discrete Contin. Dyn. Syst., 36 (2016), pp. 4451–4476. (Cited on p. 102)
- [59] 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. (Cited on p. 102)
- [60] G. S. KIMELDORF AND G. WAHBA, *A correspondence between Bayesian estimation on stochastic processes and smoothing by splines*, Ann. Math. Statist., 41 (1970), pp. 495–502. (Cited on p. 105)
- [61] S. M. KOZLOV, *The averaging of random operators*, Mat. Sb. (N.S.), 109 (151) (1979), pp. 188–202, 327. (Cited on p. 102)
- [62] F. M. LARKIN, *Gaussian measure in Hilbert space and applications in numerical analysis*, Rocky Mountain J. Math., 2 (1972), pp. 379–421. (Cited on pp. 104, 105)
- [63] D. LE AND D. S. PARKER, *Using randomization to make recursive matrix algorithms practical*, J. Funct. Program., 9 (1999), pp. 605–624, <http://journals.cambridge.org/article.S0956796899003470>. (Cited on p. 109)
- [64] A. MÁLQVIST AND D. PETERSEIM, *Localization of elliptic multiscale problems*, Math. Comp., 83 (2014), pp. 2583–2603. (Cited on pp. 102, 114, 117)
- [65] J. MANDEL, M. BREZINA, AND P. VANĚK, *Energy optimization of algebraic multigrid bases*, Computing, 62 (1999), pp. 205–228, <https://doi.org/10.1007/s006070050022>. (Cited on pp. 101, 114)
- [66] P.-G. MARTINSSON, V. ROKHLIN, AND M. TYGERT, *A randomized algorithm for the decomposition of matrices*, Appl. Comput. Harmon. Anal., 30 (2011), pp. 47–68, <https://doi.org/10.1016/j.acha.2010.02.003>. (Cited on p. 108)
- [67] F. MURAT, *Compacité par compensation*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4), 5 (1978), pp. 489–507. (Cited on p. 102)
- [68] F. MURAT AND L. TARTAR, *H-convergence*, Séminaire d'Analyse Fonctionnelle et Numérique de l'Université d'Alger, 1978. (Cited on p. 102)
- [69] D. E. MYERS, *Kriging, co-Kriging, radial basis functions and the role of positive definiteness*, Comput. Math. Appl., 24 (1992), pp. 139–148, [https://doi.org/10.1016/0898-1221\(92\)90176-I](https://doi.org/10.1016/0898-1221(92)90176-I). (Cited on p. 108)
- [70] J. NASH, *Non-cooperative games*, Ann. of Math. (2), 54 (1951), pp. 286–295. (Cited on p. 103)
- [71] A. S. NEMIROVSKY, *Information-based complexity of linear operator equations*, J. Complexity, 8 (1992), pp. 153–175, [https://doi.org/10.1016/0885-064X\(92\)90013-2](https://doi.org/10.1016/0885-064X(92)90013-2). (Cited on p. 105)
- [72] J. NOLEN, G. PAPANICOLAOU, AND O. PIRONNEAU, *A framework for adaptive multiscale methods for elliptic problems*, Multiscale Model. Simul., 7 (2008), pp. 171–196, <https://doi.org/10.1137/070693230>. (Cited on p. 102)
- [73] E. NOVAK AND H. WOŹNIAKOWSKI, *Tractability of Multivariate Problems. Volume II: Standard Information for Functionals*, EMS Tracts Math. 12, European Mathematical Society (EMS), Zürich, 2010, <https://doi.org/10.4171/084>. (Cited on p. 105)
- [74] A. O'HAGAN, *Bayes-Hermite quadrature*, J. Statist. Plann. Inference, 29 (1991), pp. 245–260, [https://doi.org/10.1016/0378-3758\(91\)90002-V](https://doi.org/10.1016/0378-3758(91)90002-V). (Cited on pp. 103, 105)
- [75] A. O'HAGAN, *Some Bayesian numerical analysis*, in Bayesian Statistics, 4 (Peñíscola, 1991), Oxford University Press, New York, 1992, pp. 345–363. (Cited on pp. 103, 105)
- [76] H. OWHADI, *Anomalous slow diffusion from perpetual homogenization*, Ann. Probab., 31 (2003), pp. 1935–1969, <https://doi.org/10.1214/aop/1068646372>. (Cited on p. 102)
- [77] H. OWHADI, *Bayesian numerical homogenization*, Multiscale Model. Simul., 13 (2015), pp. 812–828, <https://doi.org/10.1137/140974596>. (Cited on pp. 102, 105, 110, 111)
- [78] H. OWHADI AND E. J. CANDÈS, *Private communication*, 2004. (Cited on p. 109)
- [79] H. OWHADI AND C. SCOVEL, *Brittleness of Bayesian inference and new Selberg formulas*, Commun. Math. Sci., 14 (2016), pp. 83–145. (Cited on pp. 102, 103)
- [80] H. OWHADI AND C. SCOVEL, *Toward Machine Wald*, Handbook for Uncertainty Quantification, Springer International, 2016, to appear; preprint, arXiv:1508.02449 [math.ST], 2015. (Cited on p. 102)
- [81] 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. (Cited on p. 103)
- [82] H. OWHADI, C. SCOVEL, AND T. SULLIVAN, *On the brittleness of Bayesian inference*, SIAM Rev., 57 (2015), pp. 566–582, <https://doi.org/10.1137/130938633>. (Cited on p. 103)
- [83] H. OWHADI, C. SCOVEL, T. J. SULLIVAN, M. MCKERNS, AND M. ORTIZ, *Optimal uncertainty quantification*, SIAM Rev., 55 (2013), pp. 271–345, <https://doi.org/10.1137/10080782X>. (Cited on p. 102)

- [84] H. OWHADI AND L. ZHANG, *Gamblets for Opening the Complexity-Bottleneck of Implicit Schemes for Hyperbolic and Parabolic ODEs/PDEs with Rough Coefficients*, preprint, arXiv:1606.07686 [math.NA], 2016. (Cited on p. 102)
- [85] H. OWHADI AND L. ZHANG, *Metric-based upscaling*, *Comm. Pure Appl. Math.*, 60 (2007), pp. 675–723. (Cited on p. 102)
- [86] 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, <https://doi.org/10.1137/100813968>. (Cited on pp. 102, 117)
- [87] 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, <https://doi.org/10.1051/m2an/2013118>. (Cited on pp. 102, 103, 114, 117)
- [88] E. W. PACKEL, *The algorithm designer versus nature: A game-theoretic approach to information-based complexity*, *J. Complexity*, 3 (1987), pp. 244–257, [https://doi.org/10.1016/0885-064X\(87\)90014-8](https://doi.org/10.1016/0885-064X(87)90014-8). (Cited on p. 105)
- [89] I. PALASTI AND A. RENYI, *On interpolation theory and the theory of games*, *MTA Mat. Kat. Int. Kozl.*, 1 (1956), pp. 529–540. (Cited on p. 105)
- [90] 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, 1981, pp. 835–873. (Cited on p. 102)
- [91] L. E. PAYNE AND H. F. WEINBERGER, *An optimal Poincaré inequality for convex domains*, *Arch. Rational Mech. Anal.*, 5 (1960), pp. 286–292. (Cited on pp. 113, 115)
- [92] H. POINCARÉ, *Calcul des probabilités*, Georges Carrés, Paris, 1896. (Cited on pp. 103, 104)
- [93] K. RITTER, *Average-Case Analysis of Numerical Problems*, *Lecture Notes in Math.* 1733, Springer-Verlag, Berlin, 2000, <https://doi.org/10.1007/BFb0103934>. (Cited on p. 105)
- [94] J. W. RUGE AND K. STÜBEN, *Algebraic multigrid*, in *Multigrid Methods*, *Frontiers Appl. Math.* 3, SIAM, Philadelphia, 1987, pp. 73–130. (Cited on p. 101)
- [95] A. SARD, *Linear Approximation*, AMS, Providence, RI, 1963. (Cited on p. 105)
- [96] M. SCHOBER, D. K. DUVENAUD, AND P. HENNIG, *Probabilistic ODE solvers with Runge-Kutta means*, in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, eds., Curran Associates, Inc., Red Hook, NY, 2014, pp. 739–747. (Cited on p. 105)
- [97] J. E. H. SHAW, *A quasirandom approach to integration in Bayesian statistics*, *Ann. Statist.*, 16 (1988), pp. 895–914, <https://doi.org/10.1214/aos/1176350842>. (Cited on pp. 103, 105)
- [98] J. R. SHEWCHUK, *An Introduction to the Conjugate Gradient Method without the Agonizing Pain*, Tech. report, Carnegie Mellon University, Pittsburgh, PA, 1994. (Cited on p. 128)
- [99] J. SKILLING, *Bayesian solution of ordinary differential equations*, in *Maximum Entropy and Bayesian Methods* (Seattle, 1991), Springer, Dordrecht, The Netherlands, 1992, pp. 23–37. (Cited on p. 105)
- [100] S. SPAGNOLO, *Sulla convergenza di soluzioni di equazioni paraboliche ed ellittiche*, *Ann. Scuola Norm. Sup. Pisa* (3), 22 (1968), pp. 571–597; errata, *Ann. Scuola Norm. Sup. Pisa* (3), 22 (1968), p. 673. (Cited on p. 102)
- [101] K. STÜBEN, *A review of algebraic multigrid*, *J. Comput. Appl. Math.*, 128 (2001), pp. 281–309, [https://doi.org/10.1016/S0377-0427\(00\)00516-1](https://doi.org/10.1016/S0377-0427(00)00516-1). (Cited on p. 101)
- [102] A. V. SUL'DIN, *Wiener measure and its applications to approximation methods*. I, *Izv. Vysš. Učebn. Zaved. Matematika*, 1959 (1959), pp. 145–158. (Cited on p. 104)
- [103] W. SYMES, *Transfer of Approximation and Numerical Homogenization of Hyperbolic Boundary Value Problems with a Continuum of Scales*, Tech Report TR12-20, Rice University, Houston, TX, 2012. (Cited on p. 113)
- [104] J. F. TRAUB, G. W. WASILKOWSKI, AND H. WOŹNIAKOWSKI, *Information-Based Complexity*, *Comput. Sci. Sci. Comput.*, Academic Press, Boston, MA, 1988. (Cited on p. 105)
- [105] J. A. TROPP, *Recovery of short, complex linear combinations via l_1 minimization*, *IEEE Trans. Inform. Theory*, 51 (2005), pp. 1568–1570, <https://doi.org/10.1109/TIT.2005.844057>. (Cited on p. 109)
- [106] P. S. VASSILEVSKI, *Multilevel preconditioning matrices and multigrid V-cycle methods*, in *Robust Multi-grid Methods* (Kiel, 1988), *Notes Numer. Fluid Mech.* 23, Vieweg, Braunschweig, 1989, pp. 200–208. (Cited on p. 101)
- [107] P. S. VASSILEVSKI, *On two ways of stabilizing the hierarchical basis multilevel methods*, *SIAM Rev.*, 39 (1997), pp. 18–53, <https://doi.org/10.1137/S0036144595282211>. (Cited on p. 101)
- [108] P. S. VASSILEVSKI, *General constrained energy minimization interpolation mappings for AMG*, *SIAM J. Sci. Comput.*, 32 (2010), pp. 1–13, <https://doi.org/10.1137/080726252>. (Cited on pp. 101, 114)

- [109] P. S. VASSILEVSKI AND J. WANG, *Stabilizing the hierarchical basis by approximate wavelets. I. Theory*, Numer. Linear Algebra Appl., 4 (1997), pp. 103–126, [https://doi.org/10.1002/\(SICI\)1099-1506\(199703/04\)4:2<103::AID-NLA101>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1099-1506(199703/04)4:2<103::AID-NLA101>3.0.CO;2-J). (Cited on p. 101)
- [110] P. S. VASSILEVSKI AND J. WANG, *Stabilizing the hierarchical basis by approximate wavelets. II. Implementation and numerical results*, SIAM J. Sci. Comput., 20 (1998), pp. 490–514, <https://doi.org/10.1137/S1064827596300668>. (Cited on p. 101)
- [111] J. VON NEUMANN, *Zur Theorie der Gesellschaftsspiele*, Math. Ann., 100 (1928), pp. 295–320, <https://doi.org/10.1007/BF01448847>. (Cited on pp. 102, 103, 105)
- [112] J. VON NEUMANN AND O. MORGENSTERN, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ, 1944. (Cited on pp. 102, 103, 105)
- [113] A. WALD, *Statistical decision functions which minimize the maximum risk*, Ann. of Math. (2), 46 (1945), pp. 265–280. (Cited on pp. 102, 103)
- [114] W. L. WAN, T. F. CHAN, AND B. SMITH, *An energy-minimizing interpolation for robust multigrid methods*, SIAM J. Sci. Comput., 21 (2000), pp. 1632–1649, <https://doi.org/10.1137/S1064827598334277>. (Cited on pp. 101, 114)
- [115] X. WANG, *Transfer-of-Approximation Approaches for Subgrid Modeling*, Ph.D. thesis, Rice University, Houston, TX, 2012. (Cited on p. 113)
- [116] H. WENDLAND, *Scattered Data Approximation*, Cambridge Monogr. Appl. Comput. Math. 17, Cambridge University Press, Cambridge, 2005. (Cited on p. 108)
- [117] C. D. WHITE AND R. N. HORNE, *Computing absolute transmissibility in the presence of finescale heterogeneity*, in SPE Symposium on Reservoir Simulation, Society of Petroleum Engineers, 1987, SPE-16011-MS. (Cited on p. 102)
- [118] H. WOŹNIAKOWSKI, *Probabilistic setting of information-based complexity*, J. Complexity, 2 (1986), pp. 255–269, [https://doi.org/10.1016/0885-064X\(86\)90005-1](https://doi.org/10.1016/0885-064X(86)90005-1). (Cited on p. 105)
- [119] H. WOŹNIAKOWSKI, *What is information-based complexity?*, in Essays on the Complexity of Continuous Problems, Eur. Math. Soc., Zürich, 2009, pp. 89–95, <https://doi.org/10.4171/069-1/5>. (Cited on pp. 103, 105)
- [120] 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, <https://doi.org/10.1093/imanum/13.1.13>. (Cited on p. 108)
- [121] J. XU AND Y. ZHU, *Uniform convergent multigrid methods for elliptic problems with strongly discontinuous coefficients*, Math. Models Methods Appl. Sci., 18 (2008), pp. 77–105. (Cited on pp. 101, 114)
- [122] J. XU AND L. ZIKATANOV, *On an energy minimizing basis for algebraic multigrid methods*, Comput. Vis. Sci., 7 (2004), pp. 121–127, <https://doi.org/10.1007/s00791-004-0147-y>. (Cited on pp. 101, 114)
- [123] I. YAVNEH, *Why multigrid methods are so efficient*, Comput. Sci. Engrg., 8 (2006), pp. 12–22, <https://doi.org/10.1109/MCSE.2006.125>. (Cited on p. 101)
- [124] H. YSERENTANT, *On the multilevel splitting of finite element spaces*, Numer. Math., 49 (1986), pp. 379–412, <https://doi.org/10.1007/BF01389538>. (Cited on p. 101)