MULTISCALE MODEL. SIMUL.
Vol. 20, No. 1, pp. 188–219

MULTISCALE ELLIPTIC PDE UPSCALING AND FUNCTION APPROXIMATION VIA SUBSAMPLED DATA*

YIFAN CHEN† AND THOMAS Y. HOU†

Abstract. There is an intimate connection between numerical upscaling of multiscale PDEs and scattered data approximation of heterogeneous functions: the coarse variables selected for deriving an upscaled equation (in the former) correspond to the sampled information used for approximation (in the latter). As such, both problems can be thought of as recovering a target function based on some coarse data that are either artificially chosen by an upscaling algorithm or determined by some physical measurement process. The purpose of this paper is then to study, under such a setup and for a specific elliptic problem, how the lengthscale of the coarse data, which we refer to as the subsampled lengthscale, influences the accuracy of recovery, given limited computational budgets. Our analysis and experiments identify that reducing the subsampling lengthscale may improve the accuracy, implying a guiding criterion for coarse-graining or data acquisition in this computationally constrained scenario, especially leading to direct insights for the implementation of the Gamblets method in the numerical homogenization literature. Moreover, reducing the lengthscale to zero may lead to a blow-up of approximation error if the target function does not have enough regularity, suggesting the need for a stronger prior assumption on the target function to be approximated. We introduce a singular weight function to deal with it, both theoretically and numerically. This work sheds light on the interplay of the lengthscale of coarse data, the computational costs, the regularity of the target function, and the accuracy of approximations and numerical simulations.

Key words. subsampled data, multiscale PDEs, numerical upscaling, function approximation, exponential decay, localization

AMS subject classifications. 65D07, 65N15, 65N30, 35A35, 35J25, 65D05

DOI. 10.1137/20M1372214

1. Introduction.

1.1. Background and context.

In this paper, we are interested in studying a common approach for solving the following two categories of problems.

1.1.1. Problem 1: Numerical upscaling.

The aim of this problem is to identify the coarse scale solution of a multiscale PDE via solving an upscaled equation for coarse variables. As a prototypical example, in $\Omega = [0,1]^d$, consider the elliptic equation for $u \in H_0^1(\Omega), f \in L^2(\Omega)$, and $\mathcal{L} = -\nabla \cdot (a\nabla\cdot)$:

$$\begin{cases} \mathcal{L}u = f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega\,, \end{cases} \tag{1.1}$$

where the rough coefficient $a(x)$ satisfies $0 < a_{\min} \le a(x) \le a_{\max} < \infty$ for $x \in \Omega$. Suppose we select the upscaled data of the solution: $[u, \phi_i], i \in I$, where ϕ_i is some *measurement function* that is often localized in space, I is an index set, and $[\cdot,\cdot]$ denotes the standard L^2 inner product. Then, the task is to derive an effective model for these upscaled variables and use them to approximate the solution of the PDE.

*Received by the editors October 9, 2020; accepted for publication (in revised form) October 28, 2021; published electronically February 24, 2022.
https://doi.org/10.1137/20M1372214
Funding: This research is in part supported by NSF grants DMS-1912654 and DMS-1907977. The first author is partly supported by the Caltech Kortchak Scholar Program.
†Applied and Computational Mathematics, Caltech, Pasadena, CA 91125 USA (yifanc@caltech.edu, hou@cms.caltech.edu).

1.1.2. Problem 2: Scattered data approximation. This problem aims to recover a function u (assume it has an underlying PDE model as (1.1)) based on sampled data $[u, \phi_i], i \in I$. Here we intentionally use the same notation for the sampled data as that of the upscaled data in Problem 1 to make an explicit connection. We will also often call $[u, \phi_i], i \in I$ the coarse data in both problems.

1.1.3. A common approach. Problem 1 is a standard task in multiscale PDE computations, while Problem 2 has more of its background from data scientific investigations. Despite their different origins, there is an approach that solves and connects the two—study of this method is the focus of the present paper.

To motivate the method, we start from Problem 1: a natural and ideal approach for getting the coarse data is to multiply the equation with the set of *basis functions*,

$$\text{span } \{\psi_i\}_{i\in I} = \text{span } \{\mathcal{L}^{-1}\phi_i\}_{i\in I}\,,$$

so that $[\psi_i, f], i \in I$, after an integration by parts, matches the target $[u, \phi_i], i \in I$.

Phrased in the language of Galerkin's method, $\{\psi_i\}_{i\in I}$ will constitute the test space; furthermore, one needs to select a trial space V (with the same dimension) in order to get the ultimate numerical approximation of u. As such, this viewpoint has interpreted Problem 1 as a special case of Problem 2, of recovering u, from $[u, \phi_i], i \in I$, via choosing a space V. Often and conveniently, the trial space $V = \text{span } \{\psi_i\}_{i\in I}$ is chosen to be the same as the test space. Under such a choice and after selecting a suitable representative basis $\{\psi_i\}_{i\in I}$ of the linear space V so that $[\psi_i, \phi_j] = \delta_{ij}$, we can write the final solution in a concise form:

$$u^{\text{ideal}} := \sum_{i\in I} [u, \phi_i]\psi_i\,. \tag{1.2}$$

It is the ideal solution (here, "ideal" means that we have not accounted for the computational cost yet) in this setting, both to numerical upscaling and scattered data approximation. In practice, the basis function ψ_i can have global support, and we need a localization step for efficient computation.

As a special case in numerical upscaling, if we choose ϕ_i to be piecewise linear tent functions, then we get the ideal localized orthogonal decomposition (LOD) method [23]; if ϕ_i is set to be piecewise constant functions, then we obtain the Gamblet method in [26]. In their contexts, localization of $\{\psi_i\}_{i\in I}$ is achieved via an exponential decay property, and a provable accuracy guarantee has been established by controlling the coarse-graining error of using u^{ideal} to approximate u and the localization error of computing $\{\psi_i\}_{i\in I}$, respectively.

1.1.4. Our goals. The purpose of this paper is twofold.

- On the numerical upscaling side, we contribute a further discussion to this family of upscaling methods, concentrating on the fundamental role of a *subsampled lengthscale* (defined in the next subsection) in choosing $\{\phi_i\}_{i\in I}$, with its highly nontrivial consequence on the localization of $\{\psi_i\}_{i\in I}$ and the solution accuracy of u. We will get a novel trade-off between approximation and localization regarding the subsampled scale.
- On the function approximation side, the above recovery method takes advantage of the underlying physical model (1.1), combining the merits of data and physics. In addition to contributing a detailed analysis of accuracy and comparisons to numerical upscaling, we will pay close attention to the regime where the subsampled lengthscale is small and approaches zero, in which

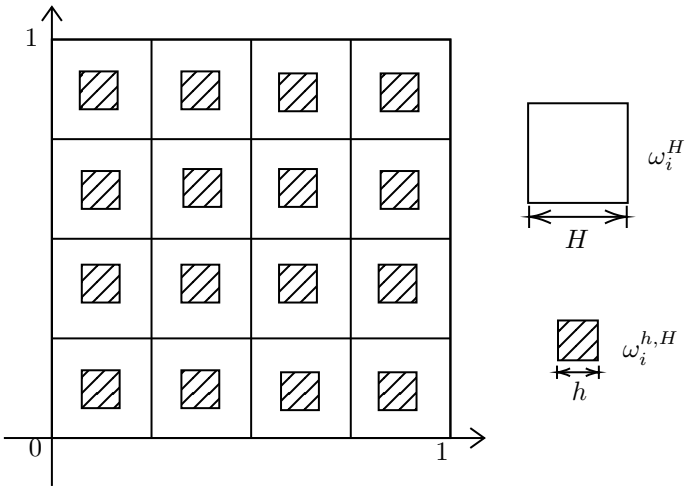

FIG. 1. *Illustration of subsampled data:* $H = 1/4, h = 1/10$.

we provide some numerical evidence that exemplifies and extends our earlier work on function approximation via subsampled data [5].

Our detailed contributions are outlined in subsection 1.4.

1.2. Subsampled lengthscales. We begin by introducing the concept of *subsampled data*. For a demonstration of ideas, we work on the domain $\Omega = [0,1]^d$, and it is decomposed uniformly into cubes with side length H; this becomes our coarse grid. Let I be the index set of these cubes such that its cardinality $|I| = 1/H^d$. The measurement function $\phi_i^{h,H}$ (we use superscripts now for notational convenience) for each $i \in I$ is set to be the (L^1 normalized) indicator function of a cube with side length $0 < h \leq H$, centered in the corresponding cube with side length H; see Figure 1 for a two-dimensional (2D) example.[1] For each $i \in I$, these two cubes are denoted by ω_i^H and $\omega_i^{h,H}$, respectively; we assume they are closed sets, i.e., their boundaries are included. We will call H the coarse lengthscale, and h is the *subsampled lengthscale*.

The consideration of this subsampled lengthscale is natural from the perspective of both function approximation and numerical upscaling. In the former scenario, the measurement data of a field function in physics is often the macroscopic averaged quantity, taking a similar form as $[u, \phi_i^{h,H}]$ for some $h \leq H$. In the latter problem, we have the freedom to choose the upscaled information of the multiscale PDEs, so taking a free parameter h in the approach enables us to analyze the algorithm's behavior more thoroughly. Later on, we will see that the parameter h has a nontrivial influence on the subsequent localization and accuracy of the approximation.

Note that the choice of ω_i^H and $\omega_i^{h,H}$ being cubes here is for convenience of analysis only; results in this paper will generalize easily to regular domains with other shapes.

1.3. Basis functions and localization. Before outlining our main contributions (which are in the next subsection), we make precise here the definition of the basis functions and their localization. Per the discussion in subsection 1.1 and especially

[1] For illustration, the cube $\omega_i^{h,H}$ in the figure is centered in ω_i^H. However, the relative position of the two cubes is not important in our analysis; see the proofs of Theorems 2.1 and 2.3. The key is that the subsampled Poincaré inequality developed in [5] does not depend on the relative position of the subdomain and the domain.

the formula (1.2), the basis function $\psi_i^{h,H}$ (we add the superscripts for notational clarity) is the solution of the variational problem

$$\begin{aligned} \psi_i^{h,H} = \operatorname{argmin}_{\psi\in H_0^1(\Omega)} \quad & \|\psi\|^2_{H_a^1(\Omega)} \\ \text{subject to} \quad & [\psi, \phi_j^{h,H}] = \delta_{i,j} \quad \text{for} \quad j \in I\,, \end{aligned} \tag{1.3}$$

where we have used the notation $\|\psi\|^2_{H_a^1(\Omega)} := \int_\Omega a|\nabla\psi|^2$. This formulation is a consequence of the two properties that are mentioned in subsection 1.1:

$$\text{(I) span } \{\psi_i^{h,H}\}_{i\in I} = \text{span } \{\mathcal{L}^{-1}\phi_i^{h,H}\}_{i\in I} \quad \text{and} \quad \text{(II) } [\psi_i^{h,H}, \phi_j^{h,H}] = \delta_{ij}\,.$$

For ease of computation, in practice we will solve a localized version of (1.3) instead:

$$\begin{aligned} \psi_i^{h,H,l} = \operatorname{argmin}_{\psi\in H_0^1(\mathrm{N}^l(\omega_i^H))} \quad & \|\psi\|^2_{H_a^1(\mathrm{N}^l(\omega_i^H))} \\ \text{subject to} \quad & [\psi, \phi_j^{h,H}] = \delta_{i,j} \quad \text{for} \quad j \in I\,, \end{aligned} \tag{1.4}$$

where $l \in \mathbb{N}$ is called the *oversampled layer*. We have $\mathrm{N}^0(\omega_i^H) = \omega_i^H$, and recursively,

$$\mathrm{N}^l(\omega_i^H) := \bigcup \{\omega_j^H, j \in I : \omega_j^H \cap \mathrm{N}^{l-1}(\omega_i^H) \neq \emptyset\}\,. \tag{1.5}$$

Then, the level-l localized solution for Problem 2 is

$$u^{\mathrm{loc},l} := \sum_{i\in I} [u, \phi_i^{h,H}]\psi_i^{h,H,l}\,. \tag{1.6}$$

By abuse of notation, we will equate $u^{\mathrm{loc},\infty} = u^{\mathrm{ideal}}$. The energy error and L^2 error of this localized solution are written as

$$\begin{aligned} e_1^{h,H,l}(a,u) &= \|u - u^{\mathrm{loc},l}\|_{H_a^1(\Omega)}\,, \\ e_0^{h,H,l}(a,u) &= \|u - u^{\mathrm{loc},l}\|_{L^2(\Omega)}\,. \end{aligned} \tag{1.7}$$

For Problem 1, we also get a solution $\tilde{u}^{\mathrm{loc},l}$ by using the localized basis functions $\{\psi_i^{h,H,l}\}_{i\in I}$ and the Galerkin method. This solution is different from $u^{\mathrm{loc},l}$ in general, unless $l = \infty$, i.e., in the ideal case. The corresponding energy error and L^2 error of $\tilde{u}^{\mathrm{loc},l}$ are denoted by $\tilde{e}_1^{h,H,l}(a,u)$ and $\tilde{e}_0^{h,H,l}(a,u)$.

We call $u^{\mathrm{loc},l}$ the *recovery solution* of Problem 2 and $\tilde{u}^{\mathrm{loc},l}$ the *Galerkin solution* of Problem 1. The computation costs of the two solutions are different—the former only requires solving the basis functions, while the latter also needs to solve an upscaled equation. Their errors in the solution are called the recovery error and the Galerkin error, respectively.

Under the above setup, our precise goal in this paper is to understand how the recovery error and the Galerkin error depend on the following three factors:

1. the coarse scale H and subsampled lengthscale h;
2. the oversampled layer l (corresponded to the computational budget);
3. the regularity of function u (in function approximation, it is given as prior information; in multiscale PDEs, it is influenced by the right-hand-side f).

Note that the regularity of a function is also intimately connected to the dimension parameter d.

1.4. Our contributions. In the first part of this work, we consider the finite regime of the subsampled lengthscale, i.e., h is a strictly positive number.

- We provide numerical experiments and theoretical analysis of these recovery and Galerkin errors. We show that for a fixed h/H, if $l = O(\log(1/H))$, then both energy errors are of $O(H)$ and both L^2 errors are of $O(H^2)$.
- Further, we decompose the error into two parts: the approximation error of the ideal solution and the localization error. We demonstrate that there is a competition between the two. Roughly, reducing h worsens the former, while improving the latter, for a fixed H and l. This leads to a novel trade-off that was not investigated before—choosing an appropriate h can benefit the final accuracy.
- Moreover, there appears a fundamental difference between $e_0^{h,H,l}(a,u)$ and the other three errors, when $d \geq 2$. For a fixed l and h/H, the former remains bounded as $H \to 0$, while the other three blow up. We characterize this phenomenon both theoretically and numerically.

In the second part of this work, we consider the small limit regime of h. When $d \geq 2$, the error estimates in the first part blow up as $h \to 0$. To remedy this issue in the context of scattered data approximation, we propose to use a singular weight function in the algorithm. The weight function puts more importance on the subsampled data and avoids the degeneracy, given the target function has improved the regularity property around these data. Numerical experiments and theoretical analysis are presented to offer a quantitative explanation of this phenomenon.

1.5. Related works. We review the related works below.

1.5.1. Numerical upscaling. There is a vast literature on numerical upscaling of multiscale PDEs. For our context, i.e., elliptic PDEs with rough coefficients, rigorous theoretical results include generalized finite element methods [1, 2], harmonic coordinates [28], LOD [23, 15, 18, 10, 14, 22], Gamblets related approaches [29, 30, 25, 26, 17, 27], and generalizations of multiscale finite element methods [16, 8, 20, 12, 6, 7]. Among them, the ones most related to this paper are LOD and Gamblets; the connection has been explained in subsection 1.1.3. Indeed, in Gamblets [26, 27], the author has formulated the framework in the perspective of optimal recovery, bridging numerical upscaling to game-theoretical approaches and Gaussian process regressions for function recovery. This formulation connects our Problem 1 and Problem 2 in subsection 1.1.

A main component in LOD and Gamblets is the localization problem—the ideal multiscale basis functions need to be localized for efficient computation. In this paper, our localization strategy, as outlined in subsection 1.3, follows from the one in [23, 26]. The main difference is that our measurement function $\phi_i^{h,H}$ contains a subsampled lengthscale parameter, which makes the analysis more delicate. Moreover, in addition to showing a trade-off between approximation errors and localization errors regarding the oversampling parameter l, our setup allows us to discover another trade-off regarding the subsampled lengthscale h—a good choice of h can improve the algorithm in [23, 26]. We also remark that the work [21] has considered a similar algorithm for convection-dominated diffusion equations, where h is fixed to be the small scale grid size, but the analysis there did not reveal the trade-off here.

1.5.2. Function approximation. Function approximation via scattered data is a classical problem in numerical analysis (interpolation), statistics (nonparametric regression), and machine learning (supervised learning). For the type of scattered

data, the most frequently considered one is the pointwise data [33]. The subsampled data introduce an additional small scale parameter h and are generalizations to pointwise data. Our earlier work [5] performed some analysis on this aspect and provides some theoretical foundation for this paper. The multiscale basis functions constructed for the subsampled data allow us to capture the heterogeneous behaviors of the target function.

The method in subsection 1.1.3 connects to the graph Laplacian approach in semisupervised learning. In the machine learning literature, the degeneracy issue of graph Laplacians has long been studied, and various approaches have been proposed to remedy this issue. Among them, the one that is most related to this paper is the weighted graph Laplacian method [31, 4], which puts more weights around the labeled data to avoid degeneracy. The second part of this work presents some analysis for this type of idea in the context of numerical analysis.

1.6. Organization. The rest of this paper is organized as follows. In section 2 we discuss the regime that $0 < h \leq H$. We present numerical experiments and theoretical analysis of these Galerkin errors in numerical upscaling and recovery errors in function approximation. In section 3, we consider the regime $h \to 0$, a case that degeneracy may occur. We use a singular weight function to deal with this issue both numerically and theoretically. Section 4 contains all the proofs in this paper. We summarize, discuss, and conclude this paper in section 5.

2. Finite regime of subsampled lengthscales. In this section, we study the finite regime of h, i.e., $0 < h \leq H$. We start with the ideal solution u^{ideal}, or equivalently $u^{\text{loc},\infty}$, and then move to the localized solution $u^{\text{loc},l}$ and $\tilde{u}^{\text{loc},l}$ for finite l. Experiments are presented first, followed with theoretical analysis. Special attention is paid to the dependence of accuracy on the coarse scale H, subsampled lengthscale h, and when in the localized case, the oversampling parameter l.

2.1. Experiments: Ideal solution. In this subsection, we perform a numerical study of the effect of h in $e_1^{h,H,\infty}(a,u)$ and $e_0^{h,H,\infty}(a,u)$ for $d = 1$ and 2, respectively.

In this ideal case, the recovery solution and Galerkin solution are the same, and in our computation, we directly solve a PDE to get these solutions. Theoretical analysis of these numerical results is given in subsection 2.2.

2.1.1. One-dimensional example. We consider the domain $\Omega = [0,1]$. The rough coefficient $a(x)$ is a sample drawn from the random field

$$\xi = 1 + 0.5 \times \sin\left(\sum_{k=1}^{100} \eta_k \cos(kx) + \zeta_k \sin(kx)\right), \tag{2.1}$$

where $\eta_k, \zeta_k, 1 \leq k \leq 100$, are independent and identically distributed random variables uniformly distributed in $[-0.5, 0.5]$; see the upper left of Figure 2 for a single realization. The right-hand-side f is drawn from the Gaussian process $\mathcal{N}(0, (-\Delta)^{-0.5-\delta})$ for $\delta = 10^{-2}$, which guarantees $f \in H^t(\Omega)$ for any $t < \delta$ but not $t \geq \delta$; see the upper right of Figure 2 for a single realization of this process. Note that this setup of f ensures that it is roughly an element in $L^2(\Omega)$ and has no apparent higher regularity. This is important because we do not want f to be too regular to influence the results, as our focus is on $f \in L^2(\Omega)$.

In the lower part of Figure 2, we output the energy errors and L^2 errors of the ideal solution, $e_1^{h,H,\infty}(a,u)$ and $e_0^{h,H,\infty}(a,u)$, for $H = 2^{-2}, 2^{-3}, \ldots, 2^{-7}$ and the subsampled ratio $h/H = 1, 1/2, 1/4, 1/8$. The grid size we use to discretize the operator is set to be 2^{-11}. These two figures lead to the following observations:

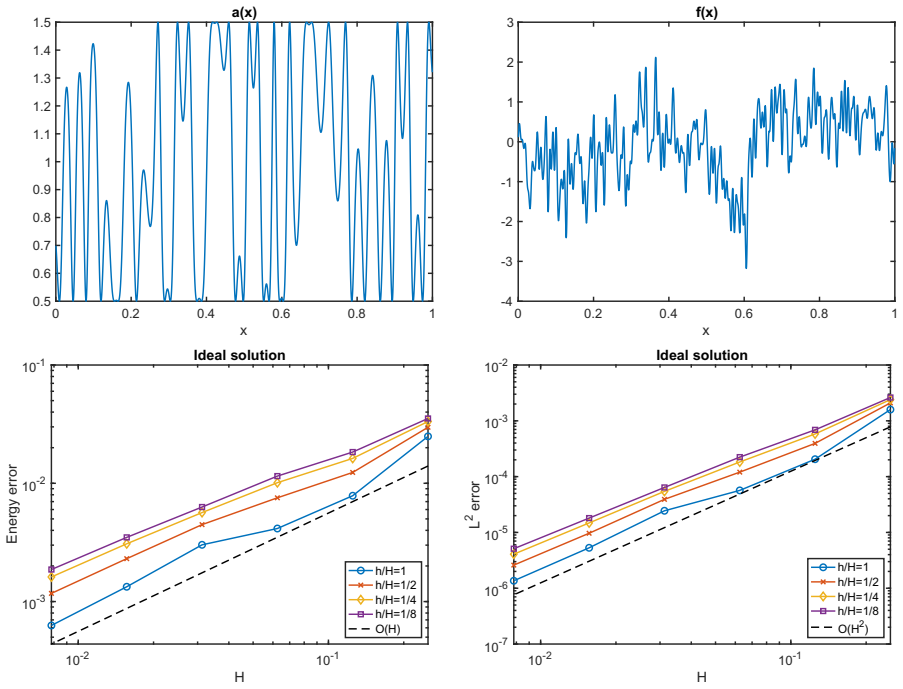

FIG. 2. *1D example, ideal solution. Upper left: $a(x)$; upper right: $f(x)$; lower left: energy error; lower right: L^2 error.*

- For the ideal solution, the energy error decays linearly with respect to the coarse scale H, while the L^2 error decays quadratically.
- Decreasing h leads to a decrease of accuracy.

In the next subsection, we move to a 2D example to further confirm these observations.

2.1.2. Two-dimensional example. We consider $\Omega = [0,1]^2$. The coefficient $a(x)$ is chosen as

$$a(x) = \frac{1}{6}\left(\frac{1.1+\sin(2\pi x_1/\epsilon_1)}{1.1+\sin(2\pi x_2/\epsilon_1)} + \frac{1.1+\sin(2\pi x_2/\epsilon_2)}{1.1+\cos(2\pi x_1/\epsilon_2)} + \frac{1.1+\cos(2\pi x_1/\epsilon_3)}{1.1+\sin(2\pi x_2/\epsilon_3)} + \frac{1.1+\sin(2\pi x_2/\epsilon_4)}{1.1+\cos(2\pi x_1/\epsilon_4)} + \frac{1.1+\cos(2\pi x_1/\epsilon_5)}{1.1+\sin(2\pi x_2/\epsilon_5)} + \sin\left(4x_1^2x_2^2\right) + 1\right), \tag{2.2}$$

where $\epsilon_1 = 1/5$, $\epsilon_2 = 1/13$, $\epsilon_3 = 1/17$, $\epsilon_4 = 1/31$, $\epsilon_5 = 1/65$. For the right-hand side, we sample two independent one-dimensional processes in the last subsection, denoted by $f_1(x_1)$ and $f_2(x_2)$, and we set $f(x) = f_1(x_1)f_2(x_2)$. This guarantees $f \in H^t(\Omega)$ for any $t < \delta$ but not $t \geq \delta$ in two dimensions.

In the upper part of Figure 3, we output $a(x)$ and a single realization of $f(x)$. The lower part depicts $e_1^{h,H,\infty}(a,u)$ and $e_0^{h,H,\infty}(a,u)$ for $H = 2^{-2}, 2^{-3}, \ldots, 2^{-6}$ and the subsampled ratio $h/H = 1, 3/4, 1/2, 1/4$. The grid size we use to discretize the operator is set to be 2^{-8}. These two figures yield the same conclusions as those in the one-dimensional case.

2.2. Analysis: Ideal solution. In this subsection, we move to the theoretical analysis of the ideal solution, to better understand the above empirical observations.

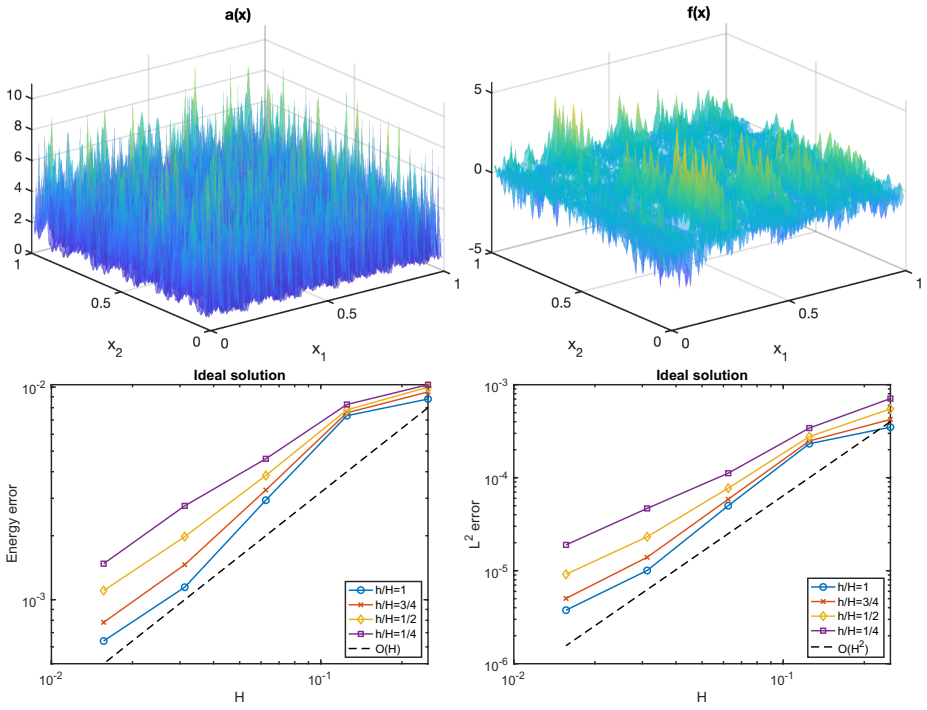

FIG. 3. *2D example, ideal solution. Upper left: $a(x)$; upper right: $f(x)$; lower left: energy error; lower right: L^2 error.*

For this purpose, we use our earlier results in function approximation via subsampled data [5]. Especially, Theorem 3.3 in [5] implies the following result.

THEOREM 2.1. *For the ideal solution, it holds that*

$$e_1^{h,H,\infty}(a,u) \leq \frac{1}{\sqrt{a_{\min}}} C_1(d) H \rho_{2,d}\left(\frac{H}{h}\right) \|\mathcal{L}u\|_{L^2(\Omega)}\,; \tag{2.3}$$

$$e_0^{h,H,\infty}(a,u) \leq \frac{1}{a_{\min}} C_1(d)^2 H^2 \left(\rho_{2,d}\left(\frac{H}{h}\right)\right)^2 \|\mathcal{L}u\|_{L^2(\Omega)}\,, \tag{2.4}$$

where $C_1(d)$ is a constant that depends on the dimension d only, and for $p, d \geq 1$, the function $\rho_{p,d} : \mathbb{R}_+ \to \mathbb{R}_+$ is defined as

$$\rho_{p,d}(t) = \begin{cases} 1, & d < p, \\ (\log(1+t))^{\frac{d-1}{d}}, & d = p, \\ t^{\frac{d-p}{p}}, & d > p\,. \end{cases} \tag{2.5}$$

In Theorem 2.1, we get the upper bound of $e_1^{h,H,\infty}(a,u)$ and $e_0^{h,H,\infty}(a,u)$. The dependence of this upper bound on h is determined by the function $\rho_{2,d}$. Note that it is a nondecreasing function, so as h decreases, for a fixed H, the ratio H/h increases, and the upper bound will also increase. One exception is when $d = 1$, the upper bound remains constant when h changes, and it is still finite even when h approaches 0. This phenomenon is in sharp contrast with the case $d \geq 2$, where as $h \to 0$, the upper bound blows up to infinity.

The above theoretical implications match what we have observed in the experiments—reducing h leads to a decrease of accuracy, in both $d = 1$ and $d = 2$; moreover, the deterioration of accuracy is more severe in $d = 2$ than $d = 1$.

Therefore, if one is adopting the ideal solution, without considering computational costs, then we would recommend choosing $h = H$, which achieves the best of both worlds with a theoretical guarantee and practical performance.

Remark 2.2. Applying the above recommendation ($h = H$) is straightforward in the context of numerical upscaling—we can choose the suitable upscaled coarse variables. Nevertheless, for scattered data approximation, the data acquisition step also matters. Our analysis suggests that for the sake of accuracy (in the case there is no burden of computational costs), it could be a good idea to make the lengthscale of the coarse data larger; this provides guidance for data collection in such a scenario.

2.3. Experiments: Localized solution. Solving the ideal solution can be computationally expensive due to the global optimization problem (1.3). This is also why we stop at $H = 2^{-6}$ and do not decrease H further in the previous 2D experiments. For better practical algorithms, in this subsection, we move to the localized solution. We start with the numerical experiments for one and two dimensions, followed by theoretical analysis. In these experiments, we use the same functions $a(x)$ and $f(x)$ as in the ideal case.

In the localized scenario, the Galerkin solution in numerical upscaling and the recovery solution in scattered data approximation are different. Thus, we will compute them separately and compare the results. More precisely, for the Galerkin solution, we use the localized basis functions in the Galerkin framework to solve the PDE; for the recovery solution, it is simpler—once the basis functions are computed, we readily get the recovery solution by using the available subsampled data and the formula (1.6). For both cases, the ground truth solution u is given as a solution to a PDE.

2.3.1. One-dimensional example. We consider the 1D model in subsection 2.1.1. We compute the Galerkin errors $\tilde{e}_1^{h,H,l}(a,u)$ and $\tilde{e}_0^{h,H,l}(a,u)$ and the recovery errors $e_1^{h,H,l}(a,u)$ and $e_0^{h,H,l}(a,u)$ for $H = 2^{-2}, 2^{-3}, \ldots, 2^{-7}$, $h/H = 1, 1/2, 1/4, 1/8$, and $l = 2, 4$. The grid size we use to discretize the operator is set to be 2^{-11}.

In Figure 4, the oversampling parameter $l = 2$. The upper part depicts the energy and L^2 errors of the Galerkin solution, while the lower part corresponds to that of the recovery solution. From the figure, we observe the following facts:

- Due to localization, the error line of $h/H = 1, 1/2, 1/4$ finally turns up as we make H very small, deviating from what we have observed in the ideal solution. This implies the localization error matters a lot.
- Among the four choices, the case $h/H = 1/8$ that corresponds to the smallest h behaves the best for small H. It appears that decreasing h may suppress the localization error to a certain extent.
- The L^2 error of the recovery solution is more stable and accurate compared to the Galerkin solution, when H is small. Especially, there is no obvious blow-up as H becomes small.

Next, we increase the oversampling parameter to $l = 4$ and output the same set of observables in Figure 5. Now, only the case $h/H = 1$ leads to a turning up of the error line, while the other three cases lead to similar error lines as the ideal solution. The best choice among the four becomes $h/H = 1/2$. Thus, as l increases, the localized solution is approaching the ideal one, and choosing a larger h would be good.

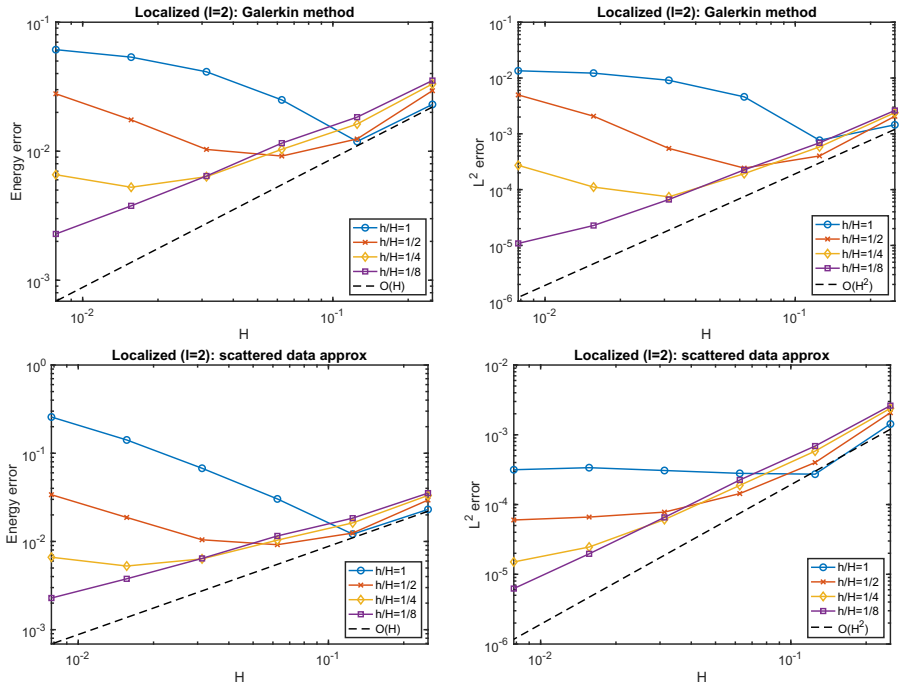

FIG. 4. *1D example, localized solution $l = 2$. Upper left: $\tilde{e}_1^{h,H,l}(a,u)$; upper right: $\tilde{e}_0^{h,H,l}(a,u)$; lower left: $e_1^{h,H,l}(a,u)$; lower right: $e_0^{h,H,l}(a,u)$.*

2.3.2. Two-dimensional example. In this subsection, we move to a 2D example that corresponds to the the ideal case in subsection 2.1.2. As before, we compute the Galerkin errors $\tilde{e}_1^{h,H,l}(a,u)$ and $\tilde{e}_0^{h,H,l}(a,u)$ and the recovery errors $e_1^{h,H,l}(a,u)$ and $e_0^{h,H,l}(a,u)$ for $H = 2^{-2}, 2^{-3}, \ldots, 2^{-8}$, $h/H = 1, 3/4, 1/2, 1/4$, and $l = 2, 4$. The grid size we use to discretize the operator is set to be 2^{-10}.

We start with $l = 2$, in Figure 6. Our observations are as follows:

- All the error lines deviate from the desired $O(H)$ or $O(H^2)$ line to some extent, and among the four choices, the ratio $h/H = 1/2$ performs the best when H is small.
- Compared to the 1D example, the localization errors in 2D are larger, since the deviation from the desired $O(H)$ or $O(H^2)$ line is more apparent.
- The error line exhibits a turning up behavior even for very small $h/H = 1/4$. That means in the 2D case, small h can also lead to large overall errors. This observation indeed matches our theory for the ideal solution, as $\rho_{2,d}(H/h)$ in Theorem 2.1 will blow up as $h \to 0$, when $d = 2$.
- When H is small, the L^2 error of the recovery solution in the scattered data approximation is more accurate than the Galerkin solution in numerical upscaling. This phenomenon has also been observed in the 1D example.

Then, we increase the oversampling parameter to $l = 4$. The results are output in Figure 7. We observe a better accuracy and more stable behavior of the error lines compared to $l = 2$. Now the best among the four ratios becomes $h/H = 3/4$. Moreover, the relative behaviors of the three cases $h/H = 3/4, 1/2, 1/4$ are very similar to that in the ideal solution, indicating that when $l = 4$, the localization error may be small compared to the approximation error of the ideal solution.

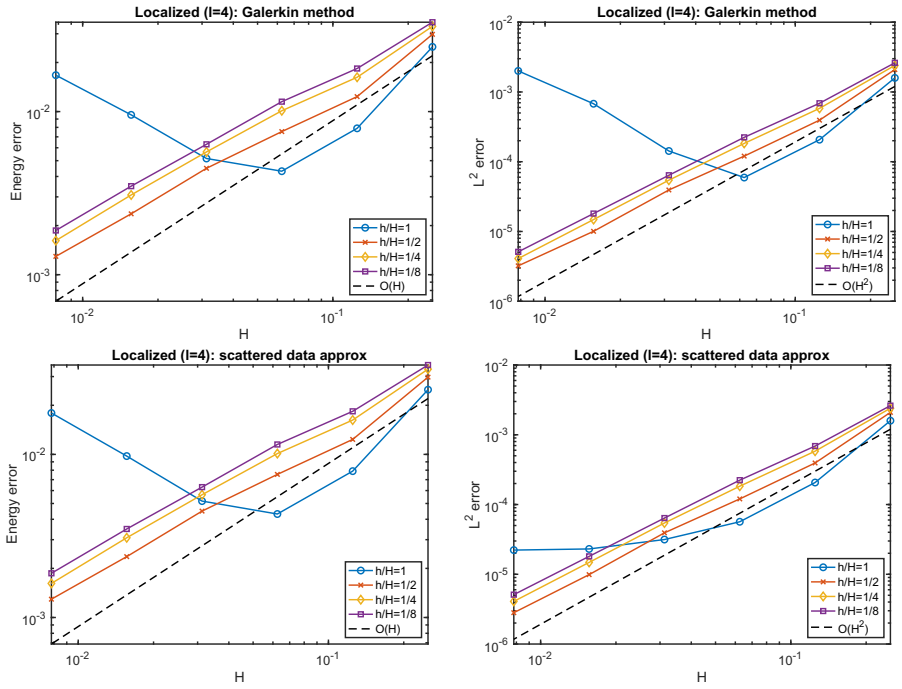

FIG. 5. *1D example, localized solution $l = 4$. Upper left: $\tilde{e}_1^{h,H,l}(a,u)$; upper right: $\tilde{e}_0^{h,H,l}(a,u)$; lower left: $e_1^{h,H,l}(a,u)$; lower right: $e_0^{h,H,l}(a,u)$.*

2.4. Analysis: Localized solution. In this subsection, we provide some theoretical analysis for the localized solution. To begin with, we summarize the main observations in the numerical experiments that we want to understand more deeply in our theoretical study.

1. The error lines of the localized solution, $e_1^{h,H,l}(a,u), \tilde{e}_1^{h,H,l}(a,u)$, and also $\tilde{e}_0^{h,H,l}(a,u)$, turn up when H is small, if l is fixed.
2. The localization error appears to become smaller as h decreases—for the overall error of the localized solution, there seems to be a competition between the approximation error of the ideal solution (which increases as h decreases) and the localization error (which decreases as h decreases). The strength of the competition depends on the oversampling parameter l.
3. The L^2 error of the recovery solution is smaller compared to that of the Galerkin solution, i.e., $\tilde{e}_0^{h,H,l}(a,u)$ appears to be larger than $e_0^{h,H,l}(a,u)$, and for the latter, it does not blow up as H becomes small.

We will provide a reasonable theoretical explanation of these observations. First, we introduce several useful notations.

2.4.1. Notations. For any function $v \in H_0^1(\Omega)$, we write

$$\mathsf{P}^{h,H,l} v = \sum_{i \in I} [v, \phi_i^{h,H}] \psi_i^{h,H,l}. \tag{2.6}$$

Moreover, we use the convention $\mathsf{P}^{h,H} v = \sum_{i \in I} [v, \phi_i^{h,H}] \psi_i^{h,H}$. These definitions lead to the relation $\mathsf{P}^{h,H,l} \psi_i^{h,H} = \psi_i^{h,H,l}$, which connects the ideal and localized basis functions.

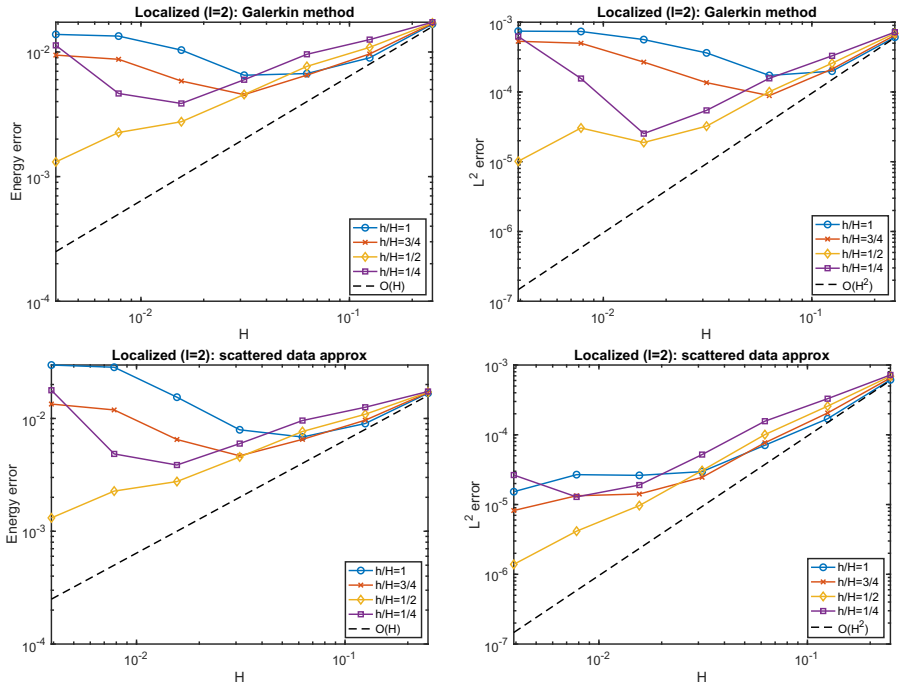

FIG. 6. *2D example, localized solution $l = 2$. Upper left: $\tilde{e}_1^{h,H,l}(a,u)$; upper right: $\tilde{e}_0^{h,H,l}(a,u)$; lower left: $e_1^{h,H,l}(a,u)$; lower right: $e_0^{h,H,l}(a,u)$.*

Since we are mainly interested in how the error depends on h, H, l, and u, we use $A \lesssim B$ (resp., $A \gtrsim B$) to denote the condition $A \leq CB$ (resp., $A \geq CB$) for some constant C independent of h, H, l, and u. If we have both $A \lesssim B$ and $A \gtrsim B$, then we will write $A \simeq B$. We use $\langle \cdot, \cdot \rangle_a$ to denote the a-weighted inner product in $H_0^1(\Omega)$, i.e., $\langle u, v \rangle_a := \int_\Omega a \nabla u \cdot \nabla v$.

2.4.2. Analysis. To analyze the error of localized solutions, we first use the triangle inequality:

$$
\begin{aligned}
e_1^{h,H,l}(a,u) &= \|u - \mathsf{P}^{h,H,l}u\|_{H_a^1(\Omega)} \\
&\leq \|u - \mathsf{P}^{h,H}u\|_{H_a^1(\Omega)} + \|\mathsf{P}^{h,H}u - \mathsf{P}^{h,H,l}u\|_{H_a^1(\Omega)} \\
&\lesssim H\rho_{2,d}(\frac{H}{h})\|\mathcal{L}u\|_{L^2(\Omega)} + \|\mathsf{P}^{h,H}u - \mathsf{P}^{h,H,l}u\|_{H_a^1(\Omega)}\,,
\end{aligned} \tag{2.7}
$$

where in the last inequality, we have used the estimate for the ideal solution. The second part $\|\mathsf{P}^{h,H}u - \mathsf{P}^{h,H,l}u\|_{H_a^1(\Omega)}$ is the localization error. Our main goal is to estimate this part of error. For this purpose, we have Theorem 2.3 below.

THEOREM 2.3. *The following results hold:*

1. *(Inverse estimate) For any $v \in \operatorname{span}\{\psi_i^{h,H}\}_{i\in I}$ and in each $\omega_j^{h,H}$, $j \in I$, we have the estimate*

$$
\|\nabla \cdot (a\nabla v)\|_{L^2(\omega_j^{h,H})} \leq \frac{\sqrt{a_{\max}}C_2(d)}{h}\|v\|_{H_a^1(\omega_j^{h,H})}\,,
$$

where $C_2(d)$ is a constant that depends on d only.

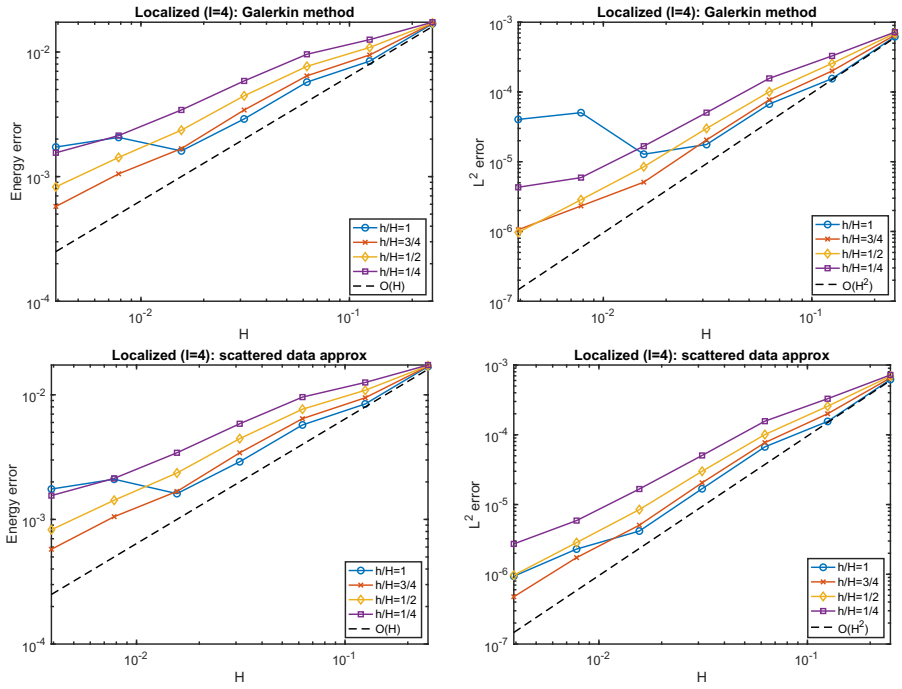

FIG. 7. *2D example, localized solution $l = 4$. Upper left: $\tilde{e}_1^{h,H,l}(a,u)$; upper right: $\tilde{e}_0^{h,H,l}(a,u)$; lower left: $e_1^{h,H,l}(a,u)$; lower right: $e_0^{h,H,l}(a,u)$.*

2. *(Exponential decay) For each $i \in I$ and $k \in \mathbb{N}$, we have*

$$\|\psi_i^{h,H}\|^2_{H^1_a(\Omega\setminus \mathrm{N}^k(\omega_i^H))} \leq (\beta(h,H))^k \, \|\psi_i^{h,H}\|^2_{H^1_a(\Omega)}, \tag{2.8}$$

where

$$\beta(h,H) = \frac{C_0(d)\sqrt{\frac{a_{\max}}{a_{\min}}}\left(C_1(d)\rho_{2,d}(\frac{H}{h}) + C_1(d)C_2(d)\frac{h}{H}\right)}{C_0(d)\sqrt{\frac{a_{\max}}{a_{\min}}}\left(C_1(d)\rho_{2,d}(\frac{H}{h}) + C_1(d)C_2(d)\frac{h}{H}\right) + 1}. \tag{2.9}$$

Here, $C_0(d)$ is a universal constant dependent on d, $C_1(d)$ is the constant in Theorem 2.1, while $C_2(d)$ is the constant in the inverse estimate.

3. *(Norm estimate) Suppose for each $i \in I$, $\phi_i^{h,H}$ is L^1 normalized in the sense that $\|\phi_i^{h,H}\|_{L^1(\omega_i^{h,H})} = 1$; then the following estimate holds:*

$$\|\psi_i^{h,H}\|_{H^1_a(\Omega)} \lesssim \frac{1}{\rho_{2,d}(\frac{H}{h})} H^{d/2-1}. \tag{2.10}$$

4. *(Localization error per basis function) For each $i \in I$, it holds that*

$$\begin{aligned} &\|\psi_i^{h,H} - \psi_i^{h,H,l}\|_{H^1_a(\Omega)} \\ &\lesssim H^{d/2-1} \min\left\{(\beta(h,H))^{l/2}, \frac{1}{\rho_{2,d}(\frac{H}{h})}\right\}. \end{aligned} \tag{2.11}$$

5. *(Overall localization error) The following error estimate holds:*
(2.12)
$$\begin{aligned}&\|\mathsf{P}^{h,H}u-\mathsf{P}^{h,H,l}u\|_{H^1_a(\Omega)}\\ \lesssim&\min\left\{\left(\beta(h,H)\right)^{l/2}\rho_{2,d}\left(\frac{H}{h}\right),1\right\}\times\min\left\{\frac{l^{d/2}}{H},\frac{1}{H^{d/2+1}\rho_{2,d}(\frac{H}{h})}\right\}\|u\|_{L^\infty(\Omega)}.\end{aligned}$$

6. *(Overall recovery error) Suppose $d\le 3$. For the energy recovery error, we have*
(2.13)
$$\begin{aligned}e_1^{h,H,l}(a,u)\lesssim\Bigg(&H\rho_{2,d}\left(\frac{H}{h}\right)+\min\left\{\left(\beta(h,H)\right)^{l/2}\rho_{2,d}\left(\frac{H}{h}\right),1\right\}\\&\times\min\left\{\frac{l^{d/2}}{H},\frac{1}{H^{d/2+1}\rho_{2,d}(\frac{H}{h})}\right\}\Bigg)\|\mathcal{L}u\|_{L^2(\Omega)},\end{aligned}$$
and for the L^2 recovery error, we have
(2.14)
$$\begin{aligned}e_0^{h,H,l}(a,u)\lesssim\Bigg(&\left(H\rho_{2,d}\left(\frac{H}{h}\right)\right)^2+\min\left\{1,H\rho_{2,d}\left(\frac{H}{h}\right)\right\}\\&\times\min\left\{\left(\beta(h,H)\right)^{l/2}\rho_{2,d}\left(\frac{H}{h}\right),1\right\}\\&\times\min\left\{\frac{l^{d/2}}{H},\frac{1}{H^{d/2+1}\rho_{2,d}(\frac{H}{h})}\right\}\Bigg)\|\mathcal{L}u\|_{L^2(\Omega)}.\end{aligned}$$

7. *(Overall Galerkin error) Suppose $d\le 3$. The energy Galerkin error is upper bounded by the energy recovery error: $\tilde{e}_1^{h,H,l}(a,u)\le e_1^{h,H,l}(a,u)$. For the L^2 Galerkin error, we have*
(2.15)
$$\begin{aligned}\tilde{e}_0^{h,H,l}(a,u)\lesssim\Bigg(&H\rho_{2,d}\left(\frac{H}{h}\right)+\min\left\{\left(\beta(h,H)\right)^{l/2}\rho_{2,d}\left(\frac{H}{h}\right),1\right\}\\&\times\min\left\{\frac{l^{d/2}}{H},\frac{1}{H^{d/2+1}\rho_{2,d}(\frac{H}{h})}\right\}\Bigg)^2\|\mathcal{L}u\|_{L^2(\Omega)}.\end{aligned}$$

2.4.3. Implications. Before we move to the proof part, let us first discuss the implications of this theorem. We focus on the localization error in the final estimates.

- Fix an l and the ratio H/h. Due to (2.13) and (2.15), the localization error parts in $e_1^{h,H,l}(a,u)$, $\tilde{e}_1^{h,H,l}(a,u)$, and $\tilde{e}_0^{h,H,l}(a,u)$ will blow up as H goes to 0. In contrast, due to (2.14), the localization error in $e_0^{h,H,l}(a,u)$ remains bounded in this limit. Indeed, it is bounded by
$$\begin{aligned}&H\rho_{2,d}\left(\frac{H}{h}\right)\times\left(\beta(h,H)\right)^{l/2}\rho_{2,d}\left(\frac{H}{h}\right)\times\frac{l^{d/2}}{H}\|\mathcal{L}u\|_{L^2(\Omega)}\\ \le\,&l^{d/2}\left(\beta(h,H)\right)^{l/2}\left(\rho_{2,d}\left(\frac{H}{h}\right)\right)^2\|\mathcal{L}u\|_{L^2(\Omega)},\end{aligned}$$
which does not blow up as $H\to 0$. This reveals a different behavior of $e_0^{h,H,l}(a,u)$ compared to the other three errors, which have been observed in our experiments. Our analysis explains this phenomenon.

- For $e_1^{h,H,l}(a,u)$, our analysis shows that there is a competition between the approximation error of the ideal solution, $H\rho_{2,d}(\frac{H}{h})$ (we omit $\|\mathcal{L}u\|_{L^2(\Omega)}$ for simplicity), and the localization error

$$\min\left\{(\beta(h,H))^{l/2}\,\rho_{2,d}\left(\frac{H}{h}\right),1\right\}\times\min\left\{\frac{l^{d/2}}{H},\frac{1}{H^{d/2+1}\rho_{2,d}(\frac{H}{h})}\right\}.$$

  Fix an H and l. When $d\geq 2$, since $\lim_{h\to 0}\rho_{2,d}(\frac{H}{h})=\infty$, we have that as $h\to 0$, the approximation error goes to infinity, while the localization error goes to zero. When $d=1$, both parts of the errors remain bounded as $h\to 0$, and thus the competition is less pronounced; this matches what we have observed in our 1D experiments—the effect of reducing h is not as large as in our 2D example.

  The existence of competition implies that in general, there should be a value of h that leads to the best error for the fixed H and l. Because the localization error decreases as l increases, this optimal value would also increase for a larger l, as observed in our experiments.

  The above phenomenon also applies to other errors, i.e., the recover L^2 error $e_0^{h,H,l}(a,u)$ and the Galerkin errors $\tilde{e}_1^{h,H,l}(a,u)$ and $\tilde{e}_0^{h,H,l}(a,u)$.
- If we fix H/h, and want to have an overall error of $O(H)$ (for energy error) or $O(H^2)$ (for the L^2 error), then our estimates show that

$$l=O\left(\frac{\log H}{\log\beta(h,H)}\right)$$

  suffices for this goal. Note that $\beta(h,H)$ can be treated as a constant (less than 1) when H/h is fixed, so generally $l=O(\log(1/H))$ is enough. Moreover, our experiments demonstrate that we could do much better in practice—a constant value of $l=2$ or 4 behaves well for a wide range of H and h.

The three points above explain the questions that we raised at the beginning of subsection 2.4.

Remark 2.4. Though the presence of "min" in many places of our estimates complicates the formula, they play critical roles in the above explanations, since we need to choose the correct term inside the "min" to get the desired conclusion.

Remark 2.5. In Theorem 2.3, the basis function $\psi_i^{h,H}$ has an exponential decay property; see (2.8). The localization error should heavily depend on the decay rate, so obtaining a tight bound of this rate is important here. In our analysis, we get the rate $\beta(h,H)$, which contains a term $\rho_{2,d}(H/h)$ that increases as h decreases (when $d\geq 2$), and a term h/H that decreases while h decreases. The two mixed components may suggest a nonmonotone behavior of the decay rate. Moreover, when $h\to 0$, we get $\beta(h,H)\to 1$, so the decay appears to deteriorate eventually for small h. On the other hand, it seems intuitive that once h is small, the measurement region $\omega_i^{h,H}$ becomes more localized, and then the decay shall be amplified. To understand this problem better, we conduct a numerical experiment as follows. For the coefficient $a(x)$ in (2.2) and $H=2^{-5}$, we compute the relative localization error $\frac{\|\psi_i^{h,H}-\psi_i^{h,H,l}\|^2_{H^1_a(\Omega)}}{\|\psi_i^{h,H}\|^2_{H^1_a(\Omega)}}$ for $h=2^{-5},2^{-6},\ldots,2^{-10}$ and $l=0,1,2,\ldots,5$. The index i is selected so that ω_i^H is centered in the domain Ω. The result is shown in Figure 8.

From the figure, we observe that there is indeed a nonmonotone behavior with respect to h in the relative localization error. Among these choices of h and l, we only

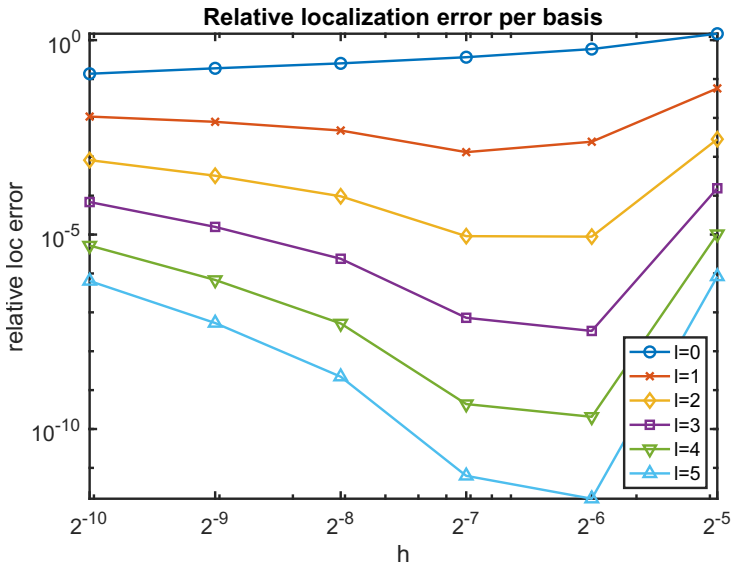

FIG. 8. *Relative localization error per basis function.*

see a monotone tendency for $l = 0$. For other l, the value h that leads to the minimal relative localization error increases as l increases. For the $a(x)$ and H considered, we can see $h/H = 1/2, 1/4$ lead to small errors in general, which also explains that this choice of h works quite well in our previous experiments. Overall, the above investigation suggests that our bound on the exponential decay and localization error can reasonably predict the behavior in practice. The decay is truly subtle regarding the small parameter h.

Remark 2.6. Our current result does not provide explicit clues on how to choose h according to l and H to achieve the best accuracy. Nonetheless, our experiments have shown that usually $h/H = 3/4$ or $1/2$ behaves well, across a wide range of $H = 2^{-8}, 2^{-7}, \ldots, 2^{-2}$ and $l = 2, 4$, in the 2D problems. Providing more guidance on this aspect, either numerically or theoretically, is left as future work.

2.4.4. Proof strategy. The results in Theorem 2.3 are presented progressively. Our proofs will start from the first and move forward one by one to the seventh. We summarize the main ideas below, together with their connections to existing results in the literature. The detailed proof is in subsection 4.1.

1. The inverse estimate is obtained due to a scaling argument—that is why we have the subsampled scale h here. (Subsection 4.1.1)
2. Based on the inverse estimate and the subsampled Poincaré inequality (see Proposition 2.5 in [5]), we can establish the exponential decay property via a Caccioppoli type of argument. The logical line of our proof here is similar to that of the original LOD method (Lemma 3.4 in [23]) and Gamblets (Theorem 3.9 in [26]), while now we need to be careful to make every estimate adaptive to the small scale parameter h. (Subsection 4.1.2)
3. For the norm estimate, we construct critical examples whose energy norm leads to a desired upper bound. The critical example here is similar to the one we used before to prove the optimality of the subsampled Poincaré inequality (see Proposition 2.6 in [5]). This type of profile has also been studied in the context of semisupervised learning; see Theorem 2 in [24]. (Subsection 4.1.3)

4. The localization error per basis function is established by combining the exponential decay estimate and the norm estimate. Our results contain two parts inside the "min" operation. The idea of proving the first part is similar to that of Lemma 3.4 in [23]. The second part is a direct application of the norm estimate. Both parts are important. The first part captures the exponential decay property, while the second part captures the behavior with respect to small h—when $d \geq 2$, this estimate implies the localization error per basis function vanishes as h goes to 0. (Subsection 4.1.4)
5. To move from the localization error per basis function to the overall localization error, we also proceed in two directions. The first one follows the idea of proving Lemma 3.5 in [23], leading to an upper bound of $O(l^{d/2}/H)$, which remains bounded as $h \to 0$. On the other hand, we can use a simple triangle inequality, which yields an estimate of $O\left(1/\left(H^{d/2+1}\rho_{2,d}(\frac{H}{h})\right)\right)$, which is worse in the power of H than the first one, but can capture the limit as $h \to 0$, i.e., it vanishes as $h \to 0$. The combination of the two leads to the final estimate (Subsection 4.1.5)
6. It is straightforward to go from overall localization error to the energy recovery error by a triangle inequality. For the L^2 recovery error, we can bound it through the energy error in two ways, with or without using the subsampled Poincaré inequality. This leads to a further "min" operation in the final estimate. (Subsection 4.1.6)
7. The energy Galerkin error is upper bounded by the energy recover error according to the Galerkin orthogonality. The L^2 Galerkin error is obtained by the standard Aubin–Nitsche trick. (Subsection 4.1.7)

3. Small limit regime of subsampled lengthscales. In the last section, we have made a detailed study of the recovery error and Galerkin error with respect to h, H, and l. We observe that there is a deterioration of accuracy as h becomes small, especially for $d \geq 2$—the benefit of small localization errors by a very small h is overwhelmed by the curse of induced large approximation errors. Due to this reason, in our experiments, we choose the ratio h/H to be not too small—we select $h/H \geq 1/8$ in one dimension and $h/H \geq 1/4$ in two dimensions. Our theoretical analysis also collaborates with these observations, as the function $\rho_{2,d}(H/h)$ that appears in the error estimate will blow up as $h/H \to 0$ for $d \geq 2$.

Therefore, we are advised not to use a very small h. While this is a practical suggestion in the problem of numerical upscaling, since we have the freedom of choosing the upscaled variables and thus can avoid this pathological phenomenon, in the problem of scattered data approximation, we may not have such flexibility due to the prevalent physical constraints for data measurements. As we often encounter recovery problems in high dimensions with scattered data that possibly have a very small lengthscale, e.g., pointwise data, it is natural to ask that whether we could get an accurate recovery even in the $h \to 0$ regime. The analysis above implies that this goal is not achievable in general for the model problem we have considered. Thus, we need to put stronger assumptions on the function u to be approximated.

Since the degeneracy of accuracy for $d \geq 2$ can be partially attributed to the low regularity of the target function u, that is, when $d \geq 2$, functions in $H^1(\Omega)$ may not have a well-defined pointwise value (according to the Sobolev embedding theorem [11]), a natural idea is to assume u to be more regular. There has been some work in which u is assumed to be in $W^{k,2}(\Omega)$ for some larger k [34]; this assumption ensures the continuity of the function. Alternatively, one can assume $u \in W^{1,p}(\Omega)$ and increase p—when $p > d$, the degeneracy issue disappears; see [9, 32, 19, 3].

The above assumptions of better regularity on u, via increasing either k or p, require to modify the recovery algorithm substantially—in the former, the basis functions are obtained by replacing the $H^1_a(\Omega)$ norm in (1.3) by a high order norm, similar to the polyharmonic splines and their rough version [30]; in the latter, the recovery function is obtained by minimizing the $W^{1,p}(\Omega)$ norm subject to the observed data.

Here, to stick to the formulation (1.3) and thus the main theme of this paper, we consider improving the regularity via choosing a singular weight function $a(x)$. Naturally, in order to make the recovery nondegenerate regarding a vanishing h, we need to put more importance on the coarse data of a small lengthscale h. Thus, we could assume the function is "nearly flat" around the data location by using a singular $a(x)$ such that $\int_\Omega a|\nabla u|^2 < \infty$—this guarantees the information content of coarse data even for very small h. We will make this intuition more quantitative in this section.

3.1. Numerical experiment. As before, we start with a numerical experiment. We choose $d = 2$ and $\Omega = [0,1]^2$. The ground truth function u is depicted in the upper left of Figure 9. The coarse scale $H = 2^{-2}$, and suppose for now we collect subsampled data with lengthscale $h = H/2 = 2^{-3}$; the grid size h_g is set to be 2^{-7}. In the upper right of Figure 9, we plot the ideal recovery solution by using $a(x) = 1$, the subsampled data $[u, \phi_i^{h,H}], i \in I$, and the ideal basis functions $\{\psi_i^{h,H}\}_{i\in I}$. We observe that to a certain extent, the recovery solution can capture the large scale property of u.

Then, we decrease the subsampled lengthscale—we choose $h = 2^{-4} \cdot H = 2^{-6}$. The recovery solution obtained by solving (1.3) with $a(x) = 1$ is in the lower left

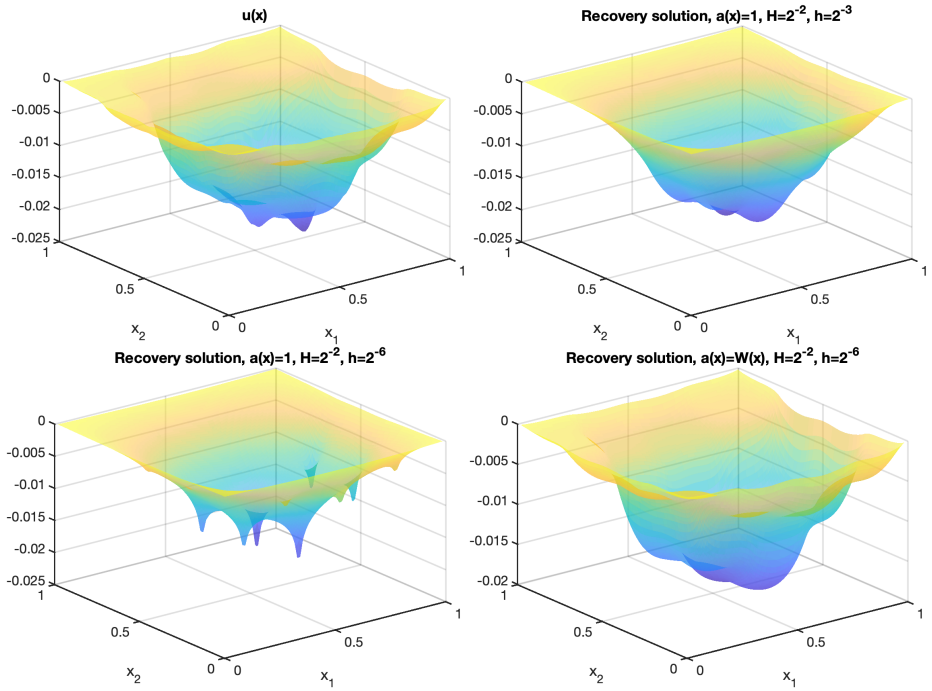

FIG. 9. *Upper left: $u(x)$; upper right: recovery solution, $h/H = 1/2$ and $a(x) = 1$; lower left: recovery solution, $h/H = 1/2^4$ and $a(x) = 1$; lower right: recovery solution, $h/H = 1/2^4$ and $a(x) = W(x)$.*

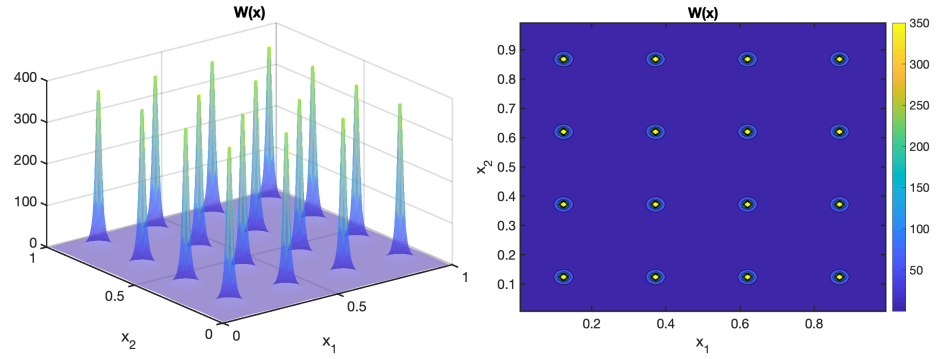

FIG. 10. *Left: figure of $W(x)$; right: contour of $W(x)$.*

of Figure 9. The degeneracy issue becomes apparent—there are many spikes in the recovery solution, and the locations of these spikes are the data positions. This confirms our understanding that a small h leads to a degenerate recovery.

Now, we define a weight function as follows. For each local patch $\omega_i^H, i \in I$, its center is denoted by $x_i \in \omega_i^H$. We write $X^H = \bigcup_{i=1}^I \{x_i^H\}$ and $\mathsf{d}(x, X^H)$ is the Euclidean distance from x to the set X^H. The weight function is defined as

$$W(x) = \left(\frac{H}{\mathsf{d}(x, X^H)}\right) \log^2 \left(1 + \frac{H}{\mathsf{d}(x, X^H)}\right). \tag{3.1}$$

It is singular at the center of our subsampled data; see Figure 10. In the lower right of Figure 9, we we construct the recovery solution by solving (1.3) with $a(x) = W(x)$. To avoid numerical instability in the experiment, we use a regularized version of the singular weight as follows:

$$W(x; h_g) = \left(\frac{H}{\max\{h_g, \mathsf{d}(x, X^H)\}}\right) \log^2 \left(1 + \frac{H}{\max\{h_g, \mathsf{d}(x, X^H)\}}\right), \tag{3.2}$$

where h_g is the grid size. From the figure, we observe that the recovery solution appears much better than the one based on $a(x) = 1$. It captures most of the large scale behaviors. Moreover, it is visually smoother—due to the singular weight function, the impact of the subsampled data does propagate to other points in the domain.

Remark 3.1. The idea of function recovery based on a weight function that puts more importance around the data regions has been used in semisupervised learning and image processing [31], through using a weighted graph Laplacian. Recently, the work [4] proposed a properly weighted Laplacian that attains a well-defined continuous limit. Our earlier work [5] also discussed a similar weighted discovery. In the next subsection, we will provide some theoretical analysis of this recovery based on results in [5], assuming $u(x)$ belongs to a weighted function space.

3.2. Analysis: Weighted inequality. For simplicity, in dimension $d \geq 2$, we consider the following class of weight functions:

$$W_{\gamma,H}(x) = \left(\frac{H}{\mathsf{d}(x, X^H)}\right)^{d-2+\gamma}, \tag{3.3}$$

where $\gamma > 0$. Indeed, the additional log term in (3.1) only makes the problem easier, since it makes the function blow up even faster.

We use the same notation as in subsection 2.4.1. Then, we have the following theorem.

THEOREM 3.2. *Let $d \geq 2$ and $\gamma > 0$. Fix an H, and we choose $a(x) = W_{\gamma,H}(x)$. Then the following results hold:*

1. *If $\|u\|_{H^1_a(\Omega)} < \infty$, then the L^2 error of the ideal solution satisfies*

$$e_0^{h,H,\infty}(a,u) \lesssim C(\gamma) H \|u\|_{H^1_a(\Omega)} . \tag{3.4}$$

2. *If $-\nabla\cdot(a\nabla u) = f \in L^2(\Omega)$, then the energy error of the ideal solution satisfies*

$$e_1^{h,H,\infty}(a,u) \lesssim C(\gamma) H \|f\|_{L^2(\Omega)} , \tag{3.5}$$

and the L^2 error satisfies

$$e_0^{h,H,\infty}(a,u) \lesssim C(\gamma) H^2 \|f\|_{L^2(\Omega)} . \tag{3.6}$$

Here, $C(\gamma)$ represents a positive constant that depends on γ only and can vary its value from place to place.

The proof is deferred to subsection 4.2. We observe from the theorem that the upper bound of the accuracy is independent of the subsampled scale h, which implies that it is still valid in the small h limit. This is in sharp contrast with the estimates in Theorem 2.1, where the upper bound blows up as $h \to 0$. The key here is the use of a singular weight function that puts more importance on the subsampled data.

We also use a numerical experiment to demonstrate this theorem. We choose $d = 2$, $\Omega = [0,1]^2$, and $H = 2^{-2}$. The parameter $\gamma = 1$. We use the mechanism in subsection 2.1.2 to generate a right-hand side $f \in L^2(\Omega)$, and u solves

$$-\nabla \cdot (W_{\gamma,H} \nabla u) = f .$$

The grid size is set to be 2^{-8}. We choose $h = 2^{-3}, 2^{-4}, \ldots, 2^{-7}$. For each h, we collect the data $[u, \phi_i^{h,H}], i \in I$, and compute the ideal recovery solutions by solving (1.3) with $a(x) = 1$ and $a(x) = W_{\gamma,H}(x)$, respectively. We output the $H_0^1(\Omega)$ and $L^2(\Omega)$ error of these recovery solutions in Figure 11.

From this figure, we observe that the recovery errors using $a(x) = 1$ will increase as h decrease, while those using $a(x) = W_{\gamma,H}(x)$ lead to a flattened curve with respect to h. This matches our theoretical predictions. Since in this example the dimension $d = 2$, the blow-up rate predicted by Theorem 2.1 is only logarithmic, so even though h is very small, the overall accuracy is still not too bad.

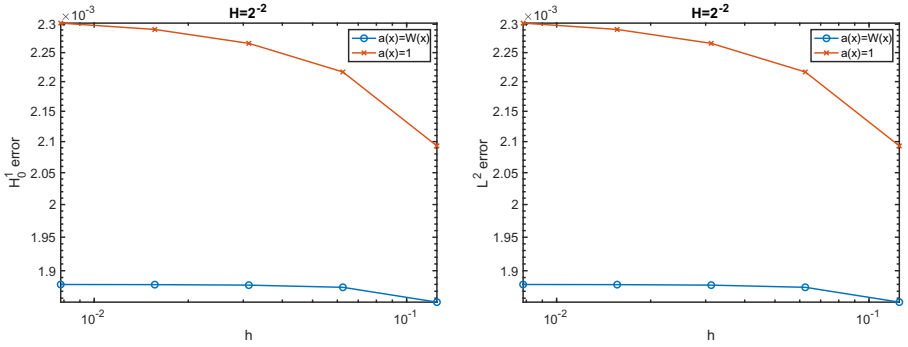

FIG. 11. *The $H_0^1(\Omega)$ and $L^2(\Omega)$ errors for different h, using constant $a(x)$ or singular weighted $a(x)$. Left: $H_0^1(\Omega)$ error; right: $L^2(\Omega)$ error.*

4. Proofs. This section provides all the proofs in this paper.

4.1. Proof of Theorem 2.3. There are seven subresults in this theorem. We prove them one by one.

4.1.1. Inverse estimate. In the domain $\omega_j^{h,H}$, we have $\nabla \cdot (a\nabla v) = c_i \phi_i^{h,H}$ for some $c_i \in \mathbb{R}$. Let $v = v_1 + v_2$ such that

$$\nabla \cdot (a\nabla v_1) = \nabla \cdot (a\nabla v) = c_i \phi_i^{h,H} \text{ in } \omega_j^{h,H}, \quad v_1|_{\partial\omega_j^{h,H}} = 0\,,$$

and for the second part,

$$\nabla \cdot (a\nabla v_2) = 0 \text{ in } \omega_j^{h,H}, \quad v_2|_{\partial\omega_j^{h,H}} = v|_{\partial\omega_j^{h,H}}\,.$$

We have the orthogonality $\int_{\omega_j^{h,H}} a\nabla v_1 \cdot \nabla v_2 = 0$. Thus, it holds that

$$\|v\|_{H_a^1(\omega_j^{h,H})} \geq \|v_1\|_{H_a^1(\omega_j^{h,H})}\,. \tag{4.1}$$

For v_1, we use the elliptic estimate:

$$\|v_1\|_{H_a^1(\omega_j^{h,H})} \geq \frac{1}{\sqrt{a_{\max}}} \|\nabla \cdot (a\nabla v_1)\|_{H^{-1}(\omega_j^{h,H})} = \frac{1}{\sqrt{a_{\max}}} \|c_j \phi_j^{h,H}\|_{H^{-1}(\omega_j^{h,H})}\,.$$

By a scaling argument, we obtain

$$\|\phi_j^{h,H}\|_{L^2(\omega_j^{h,H})} \leq \frac{C_2(d)}{h} \|\phi_j^{h,H}\|_{H^{-1}(\omega_j^{h,H})}$$

for a constant $C_2(d)$ dependent on d. Then, it follows that

$$\|v_1\|_{H_a^1(\omega_j^{h,H})} \geq \frac{h}{\sqrt{a_{\max}} C_2(d)} \|c_j \phi_j^{h,H}\|_{L^2(\omega_j^{h,H})} = \frac{h}{\sqrt{a_{\max}} C_2(d)} \|\nabla \cdot (a\nabla v)\|_{L^2(\omega_j^{h,H})}\,. \tag{4.2}$$

Combining (4.1) and (4.2), we arrive at the desired result:

$$\|\nabla \cdot (a\nabla v)\|_{L^2(\omega_j^{h,H})} \leq \frac{\sqrt{a_{\max}} C_2(d)}{h} \|v\|_{H_a^1(\omega_j^{h,H})}\,.$$

4.1.2. Exponential decay. Fix $i \in I$. For ease of notation, we will write $\psi_i^{h,H}$ by ψ, and $\mathrm{N}^k(\omega_i^H)$ by S_k in this proof.

First, we choose a cut-off function η with value 0 in S_k and value 1 in S_{k+1}^c such that it satisfies $\eta \geq 0$ and $\|\nabla\eta\|_\infty \leq C_0(d)/H$ for some universal constant $C_0(d)$ dependent on d. An example of η could be

$$\eta(x) = \frac{\mathrm{dist}(x, S_k)}{\mathrm{dist}(x, S_k) + \mathrm{dist}(x, S_{k+1}^c)}\,.$$

Then, we obtain the relation

$$\|\psi\|_{H_a^1(\Omega\setminus S_{k+1})}^2 = \int_{\Omega\setminus S_{k+1}} \nabla\psi \cdot a\nabla\psi \leq \int_\Omega \eta\nabla\psi \cdot a\nabla\psi\,. \tag{4.3}$$

Using some algebra, we have

$$\begin{aligned}\eta\nabla\psi\cdot a\nabla\psi &= \nabla(\eta\psi)\cdot a\nabla\psi-(\nabla\eta)\cdot a\psi\nabla\psi\\ &= \nabla\cdot(\eta\psi a\nabla\psi)-\eta\psi\nabla\cdot(a\nabla\psi)-(\nabla\eta)\cdot a\psi\nabla\psi\,.\end{aligned}$$

Integrating the above formula in Ω and applying the divergence theorem yields

$$\int_\Omega \eta\nabla\psi\cdot a\nabla\psi \le \left|\int_\Omega -\eta\psi\nabla\cdot(a\nabla\psi)\right| + \left|\int_\Omega(\nabla\eta)\cdot\psi a\nabla\psi\right|\,. \tag{4.4}$$

For the first term in (4.4), we have

$$\begin{aligned}\int_\Omega -\eta\psi\nabla\cdot(a\nabla\psi) &\overset{(a)}{=} \sum_{\omega_j^H\subset S_{k+1}\setminus S_k}\int_{\omega_j^H} -\eta\psi\nabla\cdot(a\nabla\psi)\\ &\overset{(b)}{=} \sum_{\omega_j^H\subset S_{k+1}\setminus S_k}\int_{\omega_j^{h,H}} -\eta\psi\nabla\cdot(a\nabla\psi)\\ &\overset{(c)}{=} \sum_{\omega_j^H\subset S_{k+1}\setminus S_k}\int_{\omega_j^{h,H}} -(\eta-\eta(x_j))\,\psi\nabla\cdot(a\nabla\psi)\\ &\overset{(d)}{\le} \sum_{\omega_j^H\subset S_{k+1}\setminus S_k}\frac{C_0(d)h}{H}\|\psi\|_{L^2(\omega_j^{h,H})}\|\nabla\cdot(a\nabla\psi)\|_{L^2(\omega_j^{h,H})}\,,\end{aligned} \tag{4.5}$$

where

- in (a), we have used the fact that η is supported in $\Omega\setminus S_k$; moreover, in $\Omega\setminus S_{k+1}$, $\eta=1$ and $\nabla\cdot(a\nabla\psi)=\sum_j c_j\phi_j^{h,H}$ for some $c_j\in\mathbb{R}$, and we have relied on the property $\int_{\omega_j^H}\phi_j^{h,H}\psi=0$ for $\omega_j^H\in\Omega\setminus S_{k+1}$;
- in (b), we have used the fact that $\phi_j^{h,H}$ is supported in $\omega_j^{h,H}$;
- in (c), we have relied on the fact $\int_{\omega_j^{h,H}}\phi_j^{h,H}\psi=0$ for $\omega_j^{h,H}\in\Omega\setminus S_k$ so we can subtract η by the constant $\eta(x_j)$ for x_j being the center of $\omega_j^{h,H}$;
- in (d) we have used the gradient bound on η and the Cauchy–Schwarz inequality.

For the term $\|\nabla\cdot(a\nabla\psi)\|_{L^2(\omega_j^{h,H})}$, we apply the inverse estimate established earlier, which leads to

$$\begin{aligned}(4.5) &\le \frac{C_0(d)h}{H}\frac{\sqrt{a_{\max}}C_2(d)}{h}\sum_{\omega_j^H\subset S_{k+1}\setminus S_k}\|\psi\|_{L^2(\omega_j^{h,H})}\|\psi\|_{H_a^1(\omega_j^{h,H})}\\ &\overset{(e)}{\le} \frac{C_0(d)h}{H}\sqrt{a_{\max}}C_2(d)C_1(d)\sum_{\omega_j^H\subset S_{k+1}\setminus S_k}\|\nabla\psi\|_{L^2(\omega_j^{h,H})}\|\psi\|_{H_a^1(\omega_j^{h,H})}\\ &\overset{(f)}{\le} \frac{C_0(d)C_1(d)C_2(d)h\sqrt{a_{\max}}}{H}\|\nabla\psi\|_{L^2(S_{k+1}\setminus S_k)}\|\psi\|_{H_a^1(S_{k+1}\setminus S_k)}\\ &\le \frac{C_0(d)C_1(d)C_2(d)h}{H}\sqrt{\frac{a_{\max}}{a_{\min}}}\|\psi\|^2_{H_a^1(S_{k+1}\cap S_k^c)}\,,\end{aligned}$$

where in (e), we have used the Poincaré inequality, based on the fact $\int_{\omega_j^{h,H}}\psi\phi_j^{h,H}=0$. The constant in the Poincaré inequality can be chosen the same as the one in Theorem

2.1, i.e., $C_1(d)$; for details see Proposition 2.5 and Theorem 3.3 in [5]. Step (f) is by the Cauchy–Schwarz inequality.

For the second term in (4.4), we have

$$\begin{aligned}\int_\Omega (\nabla\eta)\cdot\psi a\nabla\psi &= \int_{S_{k+1}\setminus S_k} (\nabla\eta)\cdot\psi a\nabla\psi\\ &= \sum_{\omega_j^H\subset S_{k+1}\setminus S_k}\int_{\omega_j^H}(\nabla\eta)\cdot\psi a\nabla\psi\\ &\le \frac{C_0(d)\sqrt{a_{\max}}}{H}\sum_{\omega_j^H\subset S_{k+1}\setminus S_k}\|\psi\|_{L^2(\omega_j^H)}\|\psi\|_{H^1_a(\omega_j^H)}\\ &\overset{(g)}{\le} \frac{C_0(d)\sqrt{a_{\max}}}{H}\sum_{\omega_j^H\subset S_{k+1}\setminus S_k} H\rho_{2,d}\left(\frac{H}{h}\right)C_1(d)\|\nabla\psi\|_{L^2(\omega_j^H)}\|\psi\|_{H^1_a(\omega_j^H)}\\ &\le C_0(d)C_1(d)\rho_{2,d}\left(\frac{H}{h}\right)\sqrt{\frac{a_{\max}}{a_{\min}}}\|\psi\|^2_{H^1_a(S_{k+1}\setminus S_k)},\end{aligned}$$

where in step (g), we have used the subsampled Poincaré inequality (Proposition 2.5 in [5]) and the fact $\int_{\omega_j^H}\phi_j^{h,H}\psi = 0$.

Combining the estimates of the two terms and (4.3), we get

$$\|\psi\|^2_{H^1_a(\Omega\setminus S_{k+1})} \le C_0(d)\left(C_1(d)\rho_{2,d}\left(\frac{H}{h}\right)+C_1(d)C_2(d)\frac{h}{H}\right)\sqrt{\frac{a_{\max}}{a_{\min}}}\|\psi\|^2_{H^1_a(S_{k+1}\setminus S_k)}.$$

Writing $\|\psi\|^2_{H^1_a(S_{k+1}\setminus S_k)} = \|\psi\|^2_{H^1_a(\Omega\setminus S_k)} - \|\psi\|^2_{H^1_a(\Omega\setminus S_{k+1})}$, we then arrive at

$$\|\psi\|^2_{H^1_a(\Omega\setminus S_{k+1})} \le \beta(h,H)\|\psi\|^2_{H^1_a(\Omega\setminus S_k)} \le \cdots \le (\beta(h,H))^{k+1}\|\psi\|^2_{H^1_a(\Omega)},$$

where

$$\beta(h,H) = \frac{C_0(d)\sqrt{\frac{a_{\max}}{a_{\min}}}\left(C_1(d)\rho_{2,d}(\frac{H}{h})+C_1(d)C_2(d)\frac{h}{H}\right)}{C_0(d)\sqrt{\frac{a_{\max}}{a_{\min}}}\left(C_1(d)\rho_{2,d}(\frac{H}{h})+C_1(d)C_2(d)\frac{h}{H}\right)+1}.$$

4.1.3. Norm estimate. Let us recall the definition of $\psi_i^{h,H}$ and $\psi_i^{h,H,l}$ for $l=0$:

$$\begin{aligned}\psi_i^{h,H} = \operatorname{argmin}_{\psi\in H^1_0(\Omega)} \quad & \|\psi\|^2_{H^1_a(\Omega)}\\ \text{subject to} \quad & [\psi,\phi_j^{h,H}] = \delta_{i,j} \quad \text{for} \quad j\in I,\end{aligned} \tag{4.6}$$

$$\begin{aligned}\psi_i^{h,H,0} = \operatorname{argmin}_{\psi\in H^1_0(\omega_i^H)} \quad & \|\psi\|^2_{H^1_a(\omega_i^H)}\\ \text{subject to} \quad & [\psi,\phi_i^{h,H}] = 1.\end{aligned} \tag{4.7}$$

Clearly, $\|\psi_i^{h,H}\|_{H^1_a(\Omega)} \le \|\psi_i^{h,H,0}\|_{H^1_a(\omega_i^H)}$ so it suffices to estimate the latter. Without loss of generality, we can assume ω_i^H is centered at 0, so that $\omega_i^{h,H} = [-h/2,h/2]^d$ and $\omega_i^H = [-H/2,H/2]^d$.

First, we choose $v\in H^1_0(\omega_i^H)$ to be a cut-off function that equals 1 in $[-H/4,H/4]^d$ and equals 0 outside ω_i^H. Moreover, $v\ge 0$ and $\|\nabla v\|_\infty \lesssim 1/H$. Then, we have

$$[v,\phi_i^{h,H}] = \frac{1}{h^d}\int_{[-h/2,h/2]^d} v \simeq 1$$

and

$$\|v\|^2_{H^1_a(\omega_i^H)} \lesssim \int_{\omega_i^H} |\nabla v|^2 \lesssim H^d \cdot \frac{1}{H^2} \lesssim H^{d-2}\,.$$

Define $w = v/[v, \phi_i^{h,H}]$; then w satisfies the constraint in (4.7), and $\|w\|_{H^1_a(\omega_i^H)} \lesssim H^{d/2-1}$, which leads to $\|\psi_i^{h,H,0}\|_{H^1_a(\omega_i^H)} \lesssim H^{d/2-1}$. Thus, the case $d = 1$ is proved.

Second, we deal with the case $d = 2$. Suppose $h \le H/2$, and we choose

$$v(x) = \begin{cases} 1 - \dfrac{\log\big(1 + \frac{4|x|}{h}\big)}{\log\big(1 + \frac{H}{h}\big)}, & |x| \le \dfrac{H}{4}, \\ 0, & |x| > \dfrac{H}{4}\,. \end{cases}$$

We have $v(x) \le 1$, and for $|x| \le h/4$, $v(x) \ge 1 - \frac{\log(2)}{\log(3)} \gtrsim 1$. Therefore, it holds that

$$[v, \phi_i^{h,H}] = \frac{1}{h^d} \int_{[-h/2,h/2]^d} v \simeq 1\,.$$

Then, we calculate the energy norm of v as follows:

$$\begin{aligned} \|v\|^2_{H^1_a(\omega_i^H)} &\lesssim \frac{1}{\log^2(1 + \frac{H}{h})} \int_{B(0,H/4)} \Big(\frac{1}{h + 4|x|}\Big)^2 \,\mathrm{d}x \\ &\lesssim \frac{1}{\log^2(1 + \frac{H}{h})} \int_0^{H/4} \frac{r}{(4r + h)^2} \,\mathrm{d}r\,. \end{aligned}$$

We write $\int_0^{H/4} \frac{r}{(4r+h)^2}\,\mathrm{d}r = \int_0^{h/2} \frac{r}{(4r+h)^2}\,\mathrm{d}r + \int_{h/2}^{H/4} \frac{r}{(4r+h)^2}\,\mathrm{d}r \lesssim \int_0^{h/2} \frac{1}{h}\,\mathrm{d}r + \int_{h/2}^{H/4} \frac{1}{r}\,\mathrm{d}r \lesssim \log\big(1 + \frac{H}{h}\big)$. Thus, it follows that

$$\|v\|_{H^1_a(\omega_i^H)} \lesssim \left(\frac{1}{\log\big(1 + \frac{H}{h}\big)}\right)^{1/2} = \frac{1}{\rho_{2,d}(\frac{H}{h})}\,.$$

This concludes the proof for the case $h \le H/2$. When $h > H/2$, we use the result in the first step $\|v\|_{H^1_a(\omega_i^H)} \lesssim H^{d/2-1} \lesssim 1 \lesssim \frac{1}{\rho_{2,d}(\frac{H}{h})}$. The case $d = 2$ is proved.

Finally, when $d \ge 3$, we choose v in a similar fashion as in the first step, such that $v = 1$ in $[-h/4, h/4]^d$ and $v = 1$ outside $[-h/2, h/2]^d$. Moreover, $v \ge 0$ and $\|\nabla v\|_\infty \lesssim 1/h$. Following the same argument in the first step, we will arrive at

$$\|\psi_i^{h,H,0}\|_{H^1_a(\omega_i^H)} \lesssim h^{d/2-1} = \frac{1}{\rho_{2,d}\big(\frac{H}{h}\big)} H^{d/2-1}\,,$$

which completes the proof.

4.1.4. Localization per basis function. We define a space

$$V^{h,H} := \{v \in H^1_0(\Omega) : [v, \phi_j^{h,H}] = 0, j \in I\}\,.$$

Then, by the optimality of $\psi_i^{h,H}$ and $\psi_i^{h,H,l}$ in their corresponding optimization problems, we have $\langle \psi_i^{h,H}, v\rangle_a = 0$ for any $v \in V^{h,H}$ and $\langle \psi_i^{h,H,l}, v\rangle_a = 0$ for any $v \in V^{h,H} \bigcap H^1_0(\mathrm{N}^l(\omega_i^H))$. Thus, $\langle \psi_i^{h,H} - \psi_i^{h,H,l}, v\rangle_a = 0$ for any $v \in V^{h,H} \bigcap H^1_0(\mathrm{N}^l(\omega_i^H))$.

Then, we define $\chi_i^{h,H} = \psi_i^{h,H} - \psi_i^{h,H,0}$ and $\chi_i^{h,H,l} = \psi_i^{h,H,l} - \psi_i^{h,H,0}$. We have $\psi_i^{h,H} - \psi_i^{h,H,l} = \chi_i^{h,H} - \chi_i^{h,H,l}$ and $\chi_i^{h,H,l} \in V^{h,H} \bigcap H_0^1(\mathrm{N}^l(\omega_i^H))$.

Based on the above fact and the orthogonality, we get

$$\begin{aligned} \|\psi_i^{h,H} - \psi_i^{h,H,l}\|_{H_a^1(\Omega)}^2 &= \|\chi_i^{h,H} - \chi_i^{h,H,l}\|_{H_a^1(\Omega)}^2 \\ &\leq \|\chi_i^{h,H} - v\|_{H_a^1(\Omega)}^2 \end{aligned} \tag{4.8}$$

for any $v \in V^{h,H} \bigcap H_0^1(\mathrm{N}^l(\omega_i^H))$. We take

$$v = \eta\chi_i^{h,H} - \mathsf{P}^{h,H,0}(\eta\chi_i^{h,H}),$$

where η is a cut-off function that equals 1 in $\mathrm{N}^{l-1}(\omega_i^H)$ and equals 0 outside $\mathrm{N}^l(\omega_i^H)$. Moreover, $\eta \geq 0$ and $\|\nabla\eta\|_\infty \lesssim 1/H$. This v belongs to $V^{h,H} \bigcap H_0^1(\mathrm{N}^l(\omega_i^H))$ because both $\eta\chi_i^{h,H}$ and $\mathsf{P}^{h,H,0}(\eta\chi_i^{h,H})$ belong to $H_0^1(\mathrm{N}^l(\omega_i^H))$, and by definition, $[\eta\chi_i^{h,H} - \mathsf{P}^{h,H,0}(\eta\chi_i^{h,H}), \phi_j^{h,H}] = 0, j \in I$. Then, it follows that

$$\begin{aligned} \|\chi_i^{h,H} - v\|_{H_a^1(\Omega)}^2 &= \|(1-\eta)\chi_i^{h,H} - \mathsf{P}^{h,H,0}\left(\eta\chi_i^{h,H}\right)\|_{H_a^1(\Omega)}^2 \\ &= \|(1-\eta)\chi_i^{h,H} - \mathsf{P}^{h,H,0}\left((1-\eta)\chi_i^{h,H}\right)\|_{H_a^1(\Omega)}^2, \end{aligned} \tag{4.9}$$

where we have used the fact $\mathsf{P}^{h,H,0}\chi_i^{h,H} = 0$. To move further, we need to use the following lemma.

LEMMA 4.1. *The operator $\mathsf{P}^{h,H,0}$ is stable under the norm $\|\cdot\|_{H_a^1(\Omega)}$. More precisely, we have that for any $w \in H_0^1(\Omega)$ it holds that*

$$\|\mathsf{P}^{h,H,0}w\|_{H_a^1(\Omega)} \lesssim \|w\|_{H_a^1(\Omega)}.$$

Proof of Lemma 4.1. By definition, $\psi_i^{h,H,0}$ is supported in ω_i^H, and $\mathsf{P}^{h,H,0}w = \sum_{i\in I}[w, \phi_i^{h,H}]\psi_i^{h,H,0}$. Thus, we have

$$\begin{aligned} \|w - \mathsf{P}^{h,H,0}w\|_{H_a^1(\Omega)}^2 &= \sum_{i\in I}\int_{\omega_i^H} a\left|\nabla(w - [w,\phi_i^{h,H}]\psi_i^{h,H,0})\right|^2 \\ &\leq \sum_{i\in I}\int_{\omega_i^H} a\left|\nabla w\right|^2 = \|w\|_{H_a^1(\Omega)}^2, \end{aligned} \tag{4.10}$$

where we have used the fact that in each ω_i^H, it holds that

$$\int_{\omega_i^H} a\nabla(w - [w,\phi_i^{h,H}]\psi_i^{h,H,0}) \cdot \nabla\psi_i^{h,H,0} = 0,$$

according to the definition of $\psi_i^{h,H,0}$. Equation (4.10) implies $\mathsf{P}^{h,H,0}$ is stable. ∎

Using Lemma 4.1, we proceed as follows:

$$\begin{aligned} (4.9) &\lesssim \|(1-\eta)\chi_i^{h,H}\|_{H_a^1(\Omega)}^2 \\ &= \int_{S_l\setminus S_{l-1}} a^2|(\nabla\eta)\chi_i^{h,H}|^2 + \int_{S_l\setminus S_{l-1}} a^2|\eta\nabla\chi_i^{h,H}| + \|\chi_i^{h,H}\|_{H_a^1(\Omega\setminus S_l)}^2, \end{aligned} \tag{4.11}$$

where we have used the notation $S_l = \mathrm{N}^l(\omega_i^H)$. For the first term in (4.11), we have

$$\begin{aligned}\int_{S_l\setminus S_{l-1}} a^2|(\nabla\eta)\chi_i^{h,H}|^2 &= \sum_{\omega_j^H\subset S_l\setminus S_{l-1}}\int_{\omega_j^H} a^2|(\nabla\eta)\chi_i^{h,H}|^2 \\ &\lesssim \sum_{\omega_j^H\subset S_l\setminus S_{l-1}} \frac{1}{H^2}\cdot H^2\left(\rho_{2,d}(\frac{H}{h})\right)^2 \|\chi_i^{h,H}\|^2_{H_a^1(\omega_j^H)} \\ &= \left(\rho_{2,d}(\frac{H}{h})\right)^2 \|\chi_i^{h,H}\|^2_{H_a^1(S_l\setminus S_{l-1})}\,.\end{aligned} \tag{4.12}$$

In the above inequality, we have used the gradient bound of η, the subsampled Poincare inequality (due to the property $[\chi_i^{h,H},\phi_j^{h,H}]=0$). Therefore, we obtain

$$\begin{aligned}(4.11) &\lesssim \left(1+\left(\rho_{2,d}\left(\frac{H}{h}\right)\right)^2\right)\|\chi_i^{h,H}\|^2_{H_a^1(S_l\setminus S_{l-1})} + \|\chi_i^{h,H}\|^2_{H_a^1(\Omega\setminus S_l)} \\ &\lesssim \left(1+\left(\rho_{2,d}\left(\frac{H}{h}\right)\right)^2\right)\|\chi_i^{h,H}\|^2_{H_a^1(\Omega\setminus S_{l-1})}\,.\end{aligned} \tag{4.13}$$

Using the fact $\|\chi_i^{h,H}\|^2_{H_a^1(\Omega\setminus S_{l-1})} = \|\psi_i^{h,H}\|^2_{H_a^1(\Omega\setminus S_{l-1})}$, the exponential decay property, and the norm estimate of $\psi_i^{h,H}$, we finally obtain

$$\|\psi_i^{h,H}-\psi_i^{h,H,l}\|_{H_a^1(\Omega)} \lesssim H^{d/2-1}\left(\beta(h,H)\right)^{l/2}\left(1+\frac{1}{\rho_{2,d}(\frac{H}{h})}\right).$$

On the other hand, we have

$$\|\psi_i^{h,H}-\psi_i^{h,H,l}\|_{H_a^1(\Omega)} \le \|\psi_i^{h,H}\|_{H_a^1(\Omega)} + \|\psi_i^{h,H,l}\|_{H_a^1(\Omega)} \lesssim H^{d/2-1}\frac{1}{\rho_{2,d}(\frac{H}{h})}\,,$$

due to the norm estimate established before. Thus, finally we obtain

$$\|\psi_i^{h,H}-\psi_i^{h,H,l}\|_{H_a^1(\Omega)} \lesssim H^{d/2-1}\cdot\min\left\{\left(\beta(h,H)\right)^{l/2}\left(1+\frac{1}{\rho_{2,d}(\frac{H}{h})}\right), \frac{1}{\rho_{2,d}(\frac{H}{h})}\right\}.$$

Note that $1\le 1+\frac{1}{\rho_{2,d}(\frac{H}{h})}\le 1+\frac{1}{\rho_{2,d}(1)}$; we could further simplify the upper bound by

$$\|\psi_i^{h,H}-\psi_i^{h,H,l}\|_{H_a^1(\Omega)} \lesssim H^{d/2-1}\cdot\min\left\{\left(\beta(h,H)\right)^{l/2}, \frac{1}{\rho_{2,d}(\frac{H}{h})}\right\}.$$

4.1.5. Overall localization error. Let $w = \mathsf{P}^{h,H}u - \mathsf{P}^{h,H,l}u$; then

$$\|w\|^2_{H_a^1(\Omega)} = \sum_{i\in I}[u,\phi_i^{h,H}]\left\langle w,\psi_i^{h,H}-\psi_i^{h,H,l}\right\rangle_a. \tag{4.14}$$

For each i, to deal with the term $\langle w,\psi_i^{h,H}-\psi_i^{h,H,l}\rangle_a$, we introduce a cut-off function η that equals 0 in $\mathrm{N}^l(\omega_i^H)$ and equals 1 in $\Omega\setminus\mathrm{N}^{l+1}(\omega_i^H)$; moreover, $\eta\ge 0$ and $\|\nabla\eta\|_\infty \lesssim 1/H$. We define

$$v = \sum_{\omega_j^H\subset\Omega\setminus\mathrm{N}^l(\omega_i^H)}[\eta w,\phi_j^{h,H}]\psi_j^{h,H,0}\in H_0^1(\Omega\setminus\mathrm{N}^l(\omega_i^H))\,.$$

Then $\eta w - v \in V^{h,H} \bigcap H_0^1(\Omega \backslash \mathrm{N}^l(\omega_i^H))$. Thus, we have $\langle \eta w - v, \psi_i^{h,H} - \psi_i^{h,H,l} \rangle = 0$ because $\eta w - v$ has a different support with that of $\psi_i^{h,H,l}$, and $\langle \psi_i^{h,H}, v \rangle_a = 0$ for any $v \in V^{h,H}$; see the first paragraph in subsection 4.1.4. Therefore, we get

$$\begin{aligned} &\left\langle w, \psi_i^{h,H} - \psi_i^{h,H,l} \right\rangle_a \\ &= \left\langle w - \eta w + v, \psi_i^{h,H} - \psi_i^{h,H,l} \right\rangle_a \\ &\leq \left(\|(1-\eta)w\|_{H_a^1(\mathrm{N}^l(\omega_i^H))} + \|v\|_{H_a^1(\mathrm{N}^{l+1}(\omega_i^H)\backslash \mathrm{N}^l(\omega_i^H))} \right) \|\psi_i^{h,H} - \psi_i^{h,H,l}\|_{H_a^1(\Omega)}, \end{aligned} \tag{4.15}$$

where we have used the fact that v is supported in $\mathrm{N}^{l+1}(\omega_i^H)\backslash \mathrm{N}^l(\omega_i^H)$. Then, by construction of v, we have $\|v\|_{H_a^1(\mathrm{N}^{l+1}(\omega_i^H)\backslash \mathrm{N}^l(\omega_i^H))} \lesssim \|\eta w\|_{H_a^1(\mathrm{N}^{l+1}(\omega_i^H)\backslash \mathrm{N}^l(\omega_i^H))}$; the proof of this property is similar to that of Lemma 4.1. Now, by using the fact $[w, \phi_j^{h,H}] = 0$ and the subsampled Poincaré inequality, we obtain

$$\|(1-\eta)w\|_{H_a^1(\mathrm{N}^l(\omega_i^H))} + \|\eta w\|_{H_a^1(\mathrm{N}^{l+1}(\omega_i^H)\backslash \mathrm{N}^l(\omega_i^H))} \lesssim \rho_{2,d}\left(\frac{H}{h}\right) \|w\|_{H_a^1(\mathrm{N}^{l+1}(\omega_i^H))}.$$

Therefore, $\langle w, \psi_i^{h,H} - \psi_i^{h,H,l} \rangle_a \lesssim \rho_{2,d}(\frac{H}{h}) \|w\|_{H_a^1(\mathrm{N}^{l+1}(\omega_i^H))} \|\psi_i^{h,H} - \psi_i^{h,H,l}\|_{H_a^1(\Omega)}$. Then combining this estimate with (4.14), we arrive at

$$\begin{aligned} \|w\|_{H_a^1(\Omega)}^2 &\lesssim \rho_{2,d}\left(\frac{H}{h}\right) \sum_{i \in I} [u, \phi_i^{h,H}] \|w\|_{H_a^1(\mathrm{N}^{l+1}(\omega_i^H))} \|\psi_i^{h,H} - \psi_i^{h,H,l}\|_{H_a^1(\Omega)} \\ &\lesssim \rho_{2,d}\left(\frac{H}{h}\right) \|u\|_{L^\infty(\Omega)} l^{d/2} \|w\|_{H_a^1(\Omega)} \left(\sum_{i \in I} \|\psi_i^{h,H} - \psi_i^{h,H,l}\|_{H_a^1(\Omega)}^2 \right)^{1/2}, \end{aligned} \tag{4.16}$$

where the last step is by the Cauchy–Schwarz inequality. Combining the above estimate with the result in the last subsection (notice that the cardinality of I is $1/H^d$), we get

$$\|w\|_{H_a^1(\Omega)} \lesssim \min\left\{ (\beta(h,H))^{l/2} \rho_{2,d}\left(\frac{H}{h}\right), 1 \right\} \cdot \frac{l^{d/2}}{H} \|u\|_{L^\infty(\Omega)}. \tag{4.17}$$

On the other hand, we can also bound

$$\begin{aligned} \|w\|_{H_a^1(\Omega)} &\leq \sum_{i \in I} |[u, \phi_i^{h,H}]| \cdot \|\psi_i^{h,H} - \psi_i^{h,H,l}\|_{H_a^1(\Omega)} \\ &\lesssim \|u\|_{L^\infty(\Omega)} H^{-d} \cdot H^{d/2-1} \cdot \min\left\{ (\beta(h,H))^{l/2}, \frac{1}{\rho_{2,d}(\frac{H}{h})} \right\} \\ &\lesssim \min\left\{ (\beta(h,H))^{l/2} \rho_{2,d}\left(\frac{H}{h}\right), 1 \right\} \cdot \frac{1}{H^{d/2+1} \rho_{2,d}(\frac{H}{h})} \|u\|_{L^\infty(\Omega)}. \end{aligned} \tag{4.18}$$

Therefore, we can write

$$\|w\|_{H_a^1(\Omega)} \lesssim \min\left\{ (\beta(h,H))^{l/2} \rho_{2,d}\left(\frac{H}{h}\right), 1 \right\} \cdot \min\left\{ \frac{l^{d/2}}{H}, \frac{1}{H^{d/2+1} \rho_{2,d}(\frac{H}{h})} \right\} \|u\|_{L^\infty(\Omega)}. \tag{4.19}$$

4.1.6. Overall recovery error. When $d \leq 3$, we have $\|u\|_{L^\infty(\Omega)} \lesssim \|\mathcal{L}u\|_{L^2(\Omega)}$; for details see Theorems 8.22 and 8.29 in [13]. Combining the estimates in (2.7) and

(2.12) leads to the estimate of the energy recovery error. For the L^2 recovery error, similar to (2.7), we have

$$e_0^{h,H,l}(a,u) \lesssim \left(H\rho_{2,d}\left(\frac{H}{h}\right)\right)^2 \|\mathcal{L}u\|_{L^2(\Omega)} + \|\mathsf{P}^{h,H}u - \mathsf{P}^{h,H,l}u\|_{L^2(\Omega)}\,. \tag{4.20}$$

The second term $\|\mathsf{P}^{h,H}u - \mathsf{P}^{h,H,l}u\|_{L^2(\Omega)}$ is the L^2 localization error. We can simply bound it by

$$\|\mathsf{P}^{h,H}u - \mathsf{P}^{h,H,l}u\|_{L^2(\Omega)} \le \|\mathsf{P}^{h,H}u - \mathsf{P}^{h,H,l}u\|_{H^1_a(\Omega)}\,. \tag{4.21}$$

On the other hand, notice that since $[\mathsf{P}^{h,H}u - \mathsf{P}^{h,H,l}u, \phi_i^{h,H}] = 0$ for any $i \in I$, we can use the subsampled Poincaré inequality so that

$$\begin{aligned}\|\mathsf{P}^{h,H}u - \mathsf{P}^{h,H,l}u\|^2_{L^2(\Omega)} &= \sum_{i\in I}\int_{\omega_i^H} |\mathsf{P}^{h,H}u - \mathsf{P}^{h,H,l}u|^2\\ &\lesssim \left(H\rho_{2,d}\left(\frac{H}{h}\right)\right)^2 \int_{\omega_i^H} a|\nabla(\mathsf{P}^{h,H}u - \mathsf{P}^{h,H,l}u)|^2\\ &= \left(H\rho_{2,d}\left(\frac{H}{h}\right)\right)^2 \|\mathsf{P}^{h,H}u - \mathsf{P}^{h,H,l}u\|^2_{H^1_a(\Omega)}\,.\end{aligned} \tag{4.22}$$

Therefore, we obtain

$$\|\mathsf{P}^{h,H}u - \mathsf{P}^{h,H,l}u\|_{L^2(\Omega)} \le \min\left\{1, H\rho_{2,d}\left(\frac{H}{h}\right)\right\} \|\mathsf{P}^{h,H}u - \mathsf{P}^{h,H,l}u\|_{H^1_a(\Omega)}\,. \tag{4.23}$$

Using the estimate of the energy error, we arrive at the final estimate.

4.1.7. Overall Galerkin error. The estimate for the energy Galerkin error is straightforward due to the Galerkin orthogonality. The L^2 error is estimated using the standard Aubin–Nitsche trick in finite element theory, which leads to square of the energy error. This completes the proof.

4.2. Proof of Theorem 3.2. We start with the first case, i.e., $\|u\|_{H^1_a(\Omega)} < \infty$. By definition,

$$e_0^{h,H,\infty}(a,u) = \|u - \mathsf{P}^{h,H}u\|_{L^2(\Omega)}\,.$$

We have the relation $[u - \mathsf{P}^{h,H}u, \phi_j^{h,H}] = 0$ for any $j \in I$. Thus, using the weighted Poincaré inequality in [5] (Theorem 4.3 and Example 1), we can estimate the error as follows:

$$\begin{aligned}\|u - \mathsf{P}^{h,H}u\|^2_{L^2(\Omega)} &= \sum_{i\in I} \|u - \mathsf{P}^{h,H}u\|^2_{L^2(\omega_i^H)}\\ &\lesssim C(\gamma)^2 H^2 \sum_{i\in I} \|u - \mathsf{P}^{h,H}u\|^2_{H^1_a(\omega_i^H)}\\ &\lesssim C(\gamma)^2 H^2 \|u\|^2_{H^1_a(\Omega)}\,,\end{aligned} \tag{4.24}$$

where in the last step, we have used the fact that $\|u - \mathsf{P}^{h,H}u\|_{H^1_a(\Omega)} \le \|u\|_{H^1_a(\Omega)}$ due to the energy orthogonality. The first case is proved.

For the second case, by energy orthogonality of the recovery, we get

$$e_1^{h,H,\infty}(a,u) \le \|u-v\|_{H_a^1(\Omega)} \tag{4.25}$$

for any $v \in \text{span}\,\{\psi_i^{h,H}\}_{i\in I}$. We can write $v = \mathcal{L}^{-1}(\sum_{i\in I} c_i \phi_i^{h,H})$ for some c_i. Then, it holds that

$$\begin{aligned} \|u-v\|_{H_a^1(\Omega)}^2 &= [u-v, \mathcal{L}(u-v)] \\ &= [u-v, f - \sum_{i\in I} c_i \phi_i^{h,H}] \\ &= \sum_{i\in I} \int_{\omega_i^H} (u-v)(f - c_i\phi_i^{h,H})\,. \end{aligned} \tag{4.26}$$

We choose $c_i = \int_{\omega_i^H} f$, so that

$$\begin{aligned} \sum_{i\in I} \int_{\omega_i^H} (u-v)(f - c_i\phi_i^{h,H}) &= \sum_{i\in I} \int_{\omega_i^H} \left(u - v - \int_{\omega_i^H} (u-v)\phi_i^{h,H}\right) f \\ &\lesssim C(\gamma) \sum_{i\in I} H \|u-v\|_{H_a^1(\omega_i^H)} \|f\|_{L^2(\omega_i^H)} \\ &\le C(\gamma) H \|u-v\|_{H_a^1(\Omega)} \|f\|_{L^2(\Omega)}\,, \end{aligned} \tag{4.27}$$

where in the second inequality, we use the Cauchy–Schwarz inequality and the weighted Poincaré inequality (Theorem 4.3 and Example 1 in [5]). Thus, finally we get $\|u - v\|_{H_a^1(\Omega)} \lesssim C(\gamma) H \|f\|_{L^2(\Omega)}$, which implies the desired energy error estimate. The L^2 error estimate is obtained by using the standard Aubin–Nitsche trick in the finite element theory.

5. Concluding remarks.

We summarize, discuss, and conclude this paper in this section.

5.1. Summary.

In this paper, we performed a detailed study of a specific approach that connects the problem of numerical upscaling and function approximation, in the context that the target function is a solution to some multiscale elliptic PDEs with rough coefficients. Our main focus is on a subsampled lengthscale that appears in the coarse data of both problems. We investigated, both numerically and theoretically, the effect of h on the recovery errors (for function approximation) and Galerkin errors (for numerical upscaling), given no computational constraints (ideal solution) or limited computational budgets (localized solution with a finite l), and given different regularity assumptions on the target function ($a(x) \in L^\infty(\Omega)$ or a singular $a(x)$). Our results imply that

- there is a trade-off between approximation errors (of ideal solutions) and localization errors (due to finite l) regarding the subsampled lengthscale h, in addition to the oversampling parameter l;
- due to the finite l caused by our limited computational budget, the Galerkin solution and recovery solution are different in general; the former behaves better in the energy accuracy, while the latter stands out in the L^2 accuracy;
- when the target function is "nearly flat" around the data locations, the subsampled data with a very small h can still contain much coarse scale information. Thus, we would recommend taking our measurements there as a first choice.

The more quantitative descriptions of these main results are established by our numerical experiments and analytic studies based on tools such as the finite element theory, the subsampled Poincaré inequality, and weighted inequalities.

5.2. Discussions. There could be multiple future directions:

- A better understanding of the trade-off regarding h and l: how to choose optimal l and h adaptively with respect to u or f. Our current results do not address this question fully.
- Other localization strategies: our localization in subsection 1.3 follows from that in [23, 26], and there are other possibilities, for example, the one in [15] or that in [18], which lead to error estimates that do not blow up as $H \to 0$. It is of interest to understand how the subsampled lengthscale influences the accuracy in that context.
- Other measurement functions: as we mentioned earlier in subsection 1.2, the choice of $\phi_i^{h,H}$ to be indicator functions in subsampled cubes is only for simplicity of analysis. Thus, results in this paper could be generalized to other types of subsampled measurement functions, for example, subsampled finite element tent functions.
- Generalization to high order models: the approach in subsection 1.1.3 applies to a general operator $\mathcal{L}$ that can be high order elliptic operators. This also connects to our discussion in subsection 3 regarding a high order model to avoid the degeneracy issues. It is of interest to study the effect of h, l and also the order of the operator $\mathcal{L}$ simultaneously on the recovery and Galerkin errors.
- Coupling of two problems: we have considered a common approach that connects two class of problems. A natural question is about a hybrid model: suppose we have the domain Ω split into two smaller domains Ω_1 and Ω_2. In Ω_1, we have a multiscale PDE $\mathcal{L}u = f$ with known f, and in Ω_2 we have some subsampled data $[u, \phi_i], i \in I$. How shall we take advantage of the PDE model in Ω_1 and the measured data in Ω_2 to recover an accurate u? This can be a very fundamental problem in combining physics and data science.

5.3. Conclusion. Overall, we have explored the connection between numerical upscaling for multiscale PDEs and scattered data approximation for heterogeneous functions, focusing on the roles of a subsampled lengthscale h and the localization parameter l. We believe it sheds light on the interplay of the lengthscale of coarse data, the computational costs, the regularity of the target function, and the accuracy of approximations and numerical simulations.

REFERENCES

[1] 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.

[2] 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.

[3] J. Calder, *Consistency of Lipschitz learning with infinite unlabeled data and finite labeled data*, SIAM J. Math. Data Sci., 1 (2019), pp. 780–812.

[4] J. Calder and D. Slepčev, *Properly-weighted graph Laplacian for semi-supervised learning*, Appl. Math. Optim., 82 (2019).

[5] Y. Chen and T. Y. Hou, *Function approximation via the subsampled Poincaré inequality*, Discrete Contin. Dyn. Syst. Ser. A, 41 (2021), pp. 169–199. https://doi.org/10.3934/dcds.2020296, 2020.

[6] Y. Chen, T. Y. Hou, and Y. Wang, *Exponential convergence for multiscale linear elliptic PDEs via adaptive edge basis functions*, Multiscale Model. Simul., 19 (2021), pp. 980–1010.
[7] Y. Chen, T. Y. Hou, and Y. Wang, *Exponentially Convergent Multiscale Methods for High Frequency Heterogeneous Helmholtz Equations*, preprint, arXiv:2105.04080, 2021.
[8] E. T. Chung, Y. Efendiev, and W. T. Leung, *Constraint energy minimizing generalized multiscale finite element method*, Comput. Methods Appl. Mech. Engrg., 339 (2018), pp. 298–319.
[9] A. El Alaoui, X. Cheng, A. Ramdas, M. J. Wainwright, and M. I. Jordan, *Asymptotic behavior of l_p-based Laplacian regularization in semi-supervised learning*, in Proceedings of the Conference on Learning Theory, 2016, pp. 879–906.
[10] C. Engwer, P. Henning, A. Målqvist, and D. Peterseim, *Efficient implementation of the localized orthogonal decomposition method*, Comput. Methods Appl. Mech. Engrg., 350 (2019), pp. 123–153.
[11] L. Evans, *Partial Differential Equations*, Graduate Stud. Math. 19, AMS, Providence, RI, 2010.
[12] S. Fu, E. Chung, and G. Li, *Edge multiscale methods for elliptic problems with heterogeneous coefficients*, J. Comput. Phys., 396 (2019), pp. 228–242.
[13] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Springer, New York, 2015.
[14] M. Hauck and D. Peterseim, *Super-localization of Elliptic Multiscale Problems*, preprint, arXiv:2107.13211, 2021.
[15] P. Henning and D. Peterseim, *Oversampling for the multiscale finite element method*, Multiscale Model. Simul., 11 (2013), pp. 1149–1175.
[16] 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.
[17] T. Y. Hou and P. Zhang, *Sparse operator compression of higher-order elliptic operators with rough coefficients*, Res. Math. Sci., 4 (2017).
[18] R. Kornhuber, D. Peterseim, and H. Yserentant, *An analysis of a class of variational multiscale methods based on subspace decomposition*, Math. Comp., 87 (2018), pp. 2765–2774.
[19] R. Kyng, A. Rao, S. Sachdeva, and D. A. Spielman, *Algorithms for Lipschitz learning on graphs*, in Proceedings of the Conference on Learning Theory, 2015, pp. 1190–1223.
[20] G. Li, *On the convergence rates of GMSFEMs for heterogeneous elliptic problems without oversampling techniques*, Multiscale Model. Simul., 17 (2019), pp. 593–619.
[21] G. Li, D. Peterseim, and M. Schedensack, *Error analysis of a variational multiscale stabilization for convection-dominated diffusion equations in two dimensions*, IMA J. Numer. Anal., 38 (2018), pp. 1229–1253.
[22] R. Maier, *A high-order approach to elliptic multiscale problems with general unstructured coefficients*, SIAM J. Numer. Anal., 59 (2021), pp. 1067–1089.
[23] A. Målqvist and D. Peterseim, *Localization of elliptic multiscale problems*, Math. Comput., 83 (2014), pp. 2583–2603.
[24] B. Nadler, N. Srebro, and X. Zhou, *Semi-supervised learning with the graph laplacian: The limit of infinite unlabelled data*, in Advances in Neural Information Processing Systems, 22, 2009, pp. 1330–1338.
[25] H. Owhadi, *Bayesian numerical homogenization*, Multiscale Model. Simul., 13 (2015), pp. 812–828.
[26] H. Owhadi, *Multigrid with rough coefficients and multiresolution operator decomposition from hierarchical information games*, SIAM Rev., 59 (2017), pp. 99–149.
[27] H. Owhadi and C. Scovel, *Operator-Adapted Wavelets, Fast Solvers, and Numerical Homogenization: From a Game Theoretic Approach to Numerical Approximation and Algorithm Design*, Cambridge Monogr. Appl. Comput. Math. 35, Cambridge University Press, Cambridge, UK, 2019.
[28] H. Owhadi and L. Zhang, *Metric-based upscaling*, Comm. Pure Appl. Math., 60 (2007), pp. 675–723.
[29] 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.
[30] 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.
[31] Z. Shi, S. Osher, and W. Zhu, *Weighted nonlocal laplacian on interpolation from sparse data*, J. Sci. Comput., 73 (2017), pp. 1164–1177.

[32] D. Slepcev and M. Thorpe, *Analysis of p-laplacian regularization in semisupervised learning*, SIAM J. Math. Anal., 51 (2019), pp. 2085–2120.
[33] H. Wendland, *Scattered Data Approximation*, Cambridge Monogr. Appl. Comput. Math. 17, Cambridge University Press, Cambridge, UK, 2004.
[34] X. Zhou and M. Belkin, *Semi-supervised learning by higher order regularization*, in Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, 2011, pp. 892–900.