ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics SAM

Quantized tensor FEM for multiscale problems: diffusion problems in two and three dimensions

V. Kazeev and I. Oseledets and M. Rakhuba and Ch. Schwab

Research Report No. 2020-33

June 2020
Latest revision: February 2022

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

QUANTIZED TENSOR FEM FOR MULTISCALE PROBLEMS: DIFFUSION PROBLEMS IN TWO AND THREE DIMENSIONS *

V. KAZEEV †, I. OSELEDETS ‡, M. RAKHUBA §, AND CH. SCHWAB ¶

Dedicated to the memory of Assyr Abdulle

Abstract. Homogenization in terms of multiscale limits transforms a multiscale problem with $n+1$ asymptotically separated microscales posed on a physical domain $D \subset \mathbb{R}^d$ into a one-scale problem posed on a product domain of dimension $(n+1)d$ by introducing n so-called "fast variables". This procedure allows to convert $n+1$ scales in d physical dimensions into a single-scale structure in $(n+1)d$ dimensions. We prove here that both the original, physical multiscale problem and the corresponding high-dimensional, one-scale limiting problem can be efficiently treated numerically with the recently developed *quantized tensor-train finite-element method* (QTT-FEM).

The QTT-FE approximation consists in restricting approximation and computation to sequences of nested subspaces, each of which is a tensor product of two factors of low dimension (rank), within a vast, but generic "virtual" (background) discretization space. In practice, these subspaces are determined *iteratively and data-adaptively at runtime* bypassing any "offline precomputation". For theoretical analysis, low-dimensional subspaces are constructed analytically to bound the tensor ranks against the error tolerance.

We consider a model linear elliptic multiscale problem in several physical dimensions and show, theoretically and experimentally, that both (i) the solution of the associated high-dimensional one-scale problem and (ii) the approximation to the solution of the multiscale problem induced thereby admit efficient QTT-FE approximations. These problems can therefore be numerically solved in a scale-robust fashion by standard (low-order) PDE discretizations combined with state-of-the-art general-purpose solvers for tensor-structured linear systems. Specifically, we prove the existence of QTT-FE approximations with an upper bound on the tensor ranks growing no faster than algebraically with respect to $\log \epsilon^{-1}$ and independent of the scale parameters, where ϵ is the target accuracy for the approximation of the solution and of its gradient. In numerical experiments, we verify the theoretical rank bounds and computationally investigate the dependence of the complexity of the solutions on the number n of microscales.

Key words. Multiscale problems, low-rank approximation, low-rank tensors, multilevel structure, data-driven discretization, matrix product states, tensor train.

AMS subject classifications. 15A69, 35B27, 65N15, 65N22, 65N30, 65N50.

1. Introduction.

The efficient numerical solution of mathematical models of *physical processes with multiple scales* has undergone a rapid development during recent years. Several classes of computational approaches have been put forward which aim, usually through selective and sparing access of the microscopic structure of the problem, to reconstruct numerically the "effective", macroscopic or "homogenized" behavior of the solution. In the context of finite-element discretizations, these methodologies are referred to as *multiscale FEM* (MsFEM). In a broader context, several such computational approaches for the numerical approximation of multiscale differential equation models have been developed:

*Submitted to the editors DATE. This work was initiated while VK and IO were visiting the Institute for Mathematical Research (FIM) of ETH Zürich in 2017 (IO) and 2018 (VK), and performed in part while the authors participated in the thematic programme "Numerical Analysis of Complex PDE Models in the Sciences" of the Erwin Schrödinger International Institute for Mathematics and Physics (ESI), Vienna, Austria, in June–August 2018. The work of MR was performed while he was a postdoc at ETH Zürich. VK was partially supported by the Austrian Science Fund (FWF) under project F65 "Taming Complexity in Partial Differential Systems".

† Faculty of Mathematics, University of Vienna, Kolingasse 14–16, 1090 Vienna, Austria (vladimir.kazeev@univie.ac.at).

‡ Skolkovo Institute of Science and Technology, Novaya Ul. 100, 143025 Skolkovo, Moscow Region, Russia and Institute of Numerical Mathematics, Russian Academy of Sciences, Gubkina St. 8, 119333, Moscow, Russia (i.oseledets@skoltech.ru).

§ HSE University, Pokrovsky Boulevard 11, Moscow, 109028 Russia (mrakhuba@hse.ru).

¶ Seminar for Applied Mathematics, ETH Zurich, Rämistrasse 101, 8092 Zurich, Switzerland. (christoph.schwab@sam.math.ethz.ch)

generalized FEM (gFEM) [6], *hierarchic multiscale methods* (HMM) [1], *local, orthogonal decompositions* (LOD) [5], to name a few. For a comprehensive discussion, we refer to [32, 1, 5] and to the references therein.

In these approaches, the solution of the correct macroscopic, or "upscaled" mathematical model is numerically approximated by *extracting microscopic information in the form of the solutions of "localized", "micro" problems*, with suitable (artificial) boundary conditions. This can be achieved by the mentioned methods in (essentially optimal) *numerical complexity that is independent of the microscopic length scale of the problem*, under the rather optimistic *provision that exact solutions of the localized problems are available at negligible cost*. Once the macroscopic solution has been computed, *postprocessing techniques* allow for *localized numerical recovery of the microscopic structure of the physical solution*. Solving the localized problems is necessary to take into account the effect of the microscale structure; it is, however, not necessarily trivial. Considering that essential computation "offline" or "free of computational cost" is therefore debatable in general and especially in the settings where microscales and microstructure evolve during computation. The latter may be relevant, for example, in Newton-type iterations for nonlinear problems or in time stepping for evolution problems.

In the present paper, we consider multilevel low-rank tensor approximation for a class of linear second-order multiscale diffusion problems of the following form:

$$\nabla^{\mathsf{T}} A^{\varepsilon} \nabla u^{\varepsilon} = f \text{ on } D \quad \text{and} \quad u^{\varepsilon} = 0 \text{ on } \partial D \tag{1.1}$$

with n microscales controlled by a scale parameter $\varepsilon > 0$. The two key ingredients of the approach, to which we refer as *QTT-FE*, are as follows. First, a generic, low-order "background" finite-element (FE) space, so vast as to resolve the microscale behavior of the solution, is considered. Second, approximation is restricted to a subset of the background finite-element space that is efficiently parametrized by the low-rank multilevel tensor decomposition known as the *quantized tensor train* (QTT) decomposition [60, 43] and as the *matrix product states* (MPS) representation in computational quantum physics [72, 71, 70, 65]. The QTT-FE *approximation* and the associated QTT-FE *method* for the numerical solution of PDEs rely on certain sequences of nested subspaces of tensor-product form and of low rank. These can be constructed using analytical tools for the purpose of analysis or computed in a fully data-driven way in practice.

For multiscale problems, such as (1.1), the QTT-FE approximation and the associated QTT-FE *method* aim to numerically resolve the macroscopic and microscopic behavior of the solution *simultaneously and throughout the physical domain* by immediately solving the original multiscale model. In particular, the QTT-FE *method* is entirely "online", i.e., in contrast to the mentioned gFEM, LOD and MsFEM [5, 33, 6, 1], it does not involve any "offline" computation of problem-adapted subspaces (basis functions) or any "offline" solution of localized auxiliary problems. The *mathematical analysis* of tensor approximation, which we provide in the present paper, however, relies on several intermediate approximations and the associated assumptions on the microscale structure.

Under suitable assumptions, including that of *asymptotic scale separation*, the microscale structure of the solutions of such problems is known to be represented by $(n+1)$-scale limits (see, e.g., [55, 3, 4]). These limits lead to *high-dimensional, possibly anisotropic elliptic boundary value problems*, which are independent of the scale parameter ε and are posed on a Cartesian product of the physical domain D and of n *microscale cells* $Y_1, \dots, Y_n$. The dependence on the microscale variables, taking value in the cells, represents the microscale structure of solution of the multiscale problem. The $(n+1)$-scale limits hence *trade the physical scale* of the microscale structure *for the high dimensionality* of the limit problem.

This idea, proposed for problems with n scales in [66], has been developed in the context of *sparse tensor FEM* multiscale diffusion problems in [30] and, subsequently,

for elasticity and electromagnetics [29, 73, 74, 75, 18, 67]. In particular, *algebraic convergence rates independent of the scale parameter with weak or no dependence on the number of n of fast variables* were established. The implementation of the developed sparse-tensor FEM discretizations of the high-dimensional limits requires, however, the explicit derivation of the PDEs which describe the $(n+1)$-scale limits. This may not be feasible, in particular, for nonlinear multiscale problems, even though the existence of $(n+1)$-scale limits has been established mathematically.

1.1. Contributions. In the present paper, we extend the earlier work [46, 38] for the case of a single spatial dimension ($d = 1$) and consider the *quantized tensor-train finite-element* (QTT-FE) approximation. We develop a *mathematical analysis* of the low-rank QTT-FE [35, 40, 52] approximation to the solutions of the class of linear elliptic second-order multiscale problems specified in Section 2 below, with diffusion coefficients depending on $n+1$ separated scales, i.e., in the classical setting of $(n+1)$-scale homogenization. We also mention the work [42], where multiscale-diffusion problems in two spatial dimensions ($d = 2$) were treated but using low-rank matrix structure with respect to the two spatial variables and not the multilevel low-rank tensor structure, as in [46, 38].

Here, we prove first that the QTT-FE approach allows for *root-exponentially convergent* numerical approximations of the solution to the one-scale limit problem. We achieve that by analytically constructing low-rank approximations with an upper bound on the ranks that grows no faster than algebraically with respect to $L \simeq \log \epsilon^{-1}$ and is independent of the scale parameter ε. Here, ϵ is the target accuracy for the H^1 approximation of the scale-interaction functions, which constitute the solution. From these approximations, we construct their counterparts approximating the solutions of the corresponding family of multiscale problems.

The idea of approximating the multiscale problem by reapproximating the homogenized problem (via $(n+1)$-scale limits as in [4, 20]), proposed for elliptic multiscale problems in [66], was exploited in the context of *sparse grid* approximations [30, 29, 28]. Our present perspective extends beyond that, as the adaptive QTT-FE *approximation* and the QTT-FE *method* can completely bypass the homogenization procedure and operate entirely on the physical domain, adaptively accessing the fine-scale information of the PDE. The numerical approximations computed by this approach are better adapted to the data, are more efficient than the particular approximations constructed analytically in our proofs through the reapproximation of one-scale limit problems and do not suffer from the accuracy limitation of homogenization. This is confirmed by our numerical results, reported in Section 5, that were obtained by such a practical computational QTT-FEM algorithm based on the Julia package `TensorRefinement.jl` [36].

1.2. Structure of the present paper. In Section 2, we describe the n-scale homogenization problem, and present in particular the QTT discretization of this problem in the physical domain in Section 2.1. The emphasis in Section 2 is to present the n-scale problem and its quantized, tensor-formatted discretization entirely in the physical domain. Section 3 presents the asymptotic analysis of the n-scale solution by the so-called unfolding method: the asymptotic limit of the physical problem is described by a high-dimensional one-scale problem. To this end, we recapitulate from [55, 3, 4] results on reiterated homogenization for linear elliptic multiscale problems, which are required in the ensuing numerical analysis of the QTT-FE approach.

In Section 4, the solution of the one-scale limit problem is approximated by tensor products of algebraic and trigonometric polynomials, and these products, in turn, are reapproximated by low-order (but also low-rank) finite-element functions. Finally, corresponding approximations to the solutions of multiscale problems with $\varepsilon > 0$ are derived.

Section 5 presents numerical experiments with model multiscale problems, for

which the QTT-FE structure of approximate solutions computed numerically is investigated. Finally, in Section 6, we discuss our results and their possible generalizations. The Appendix contains a few proofs postponed due to their technicality.

2. Model elliptic multiscale problem. We consider a bounded "physical" domain $D \subset \mathbb{R}^d$ (with which, for notational convenience, we associate the macroscale $\varepsilon_0 \equiv 1$) and a moderate number $n \in \mathbb{N}$ of microscales $\varepsilon_1, \dots, \varepsilon_n$, which we assume to be positive functions of a positive scale parameter ε.

We assume furthermore that there exist n *unit cells* $Y_1, \dots, Y_n$ such that D is partitioned into a union of translations of $\varepsilon_1 Y_1$ and each Y_{i-1} with $i \in \{2, \dots, n\}$ is partitioned into a union of translations of $\varepsilon_i Y_i$. We deal in particular with the case of $Y_1, \dots, Y_n = (0,1)^d$ in the present paper, while more sophisticated constructions may be used to model, e.g., perforated media. For notational convenience, we set $\boldsymbol{Y}_0 = \{0\}$ and $\boldsymbol{Y}_i = Y_1 \times \cdots \times Y_i$ for each $i \in \{1, \dots, n\}$.

To formulate a multiscale diffusion problem on D, we consider a matrix function A defined on $D \times \boldsymbol{Y}_n$, which therefore depends on a macroscale ("slow") variable and on n microscale ("fast") variables. We will consider multiscale diffusion coefficients A^ε induced by functions satisfying the following Assumption 2.1. Here and throughout, we use $C_\#(\boldsymbol{Y}_n)$ to denote the space of functions that are continuous on $\overline{\boldsymbol{Y}_n}$ and Y_i-periodic with respect to the ith variable for each $i \in \{1, \dots, n\}$.

ASSUMPTION 2.1. *$A \in L^\infty(D; C_\#(\boldsymbol{Y}_n; \mathbb{R}^{d\times d}_{\mathrm{sym}}))$ is essentially bounded and uniformly positive definite with constants Γ and γ: $\gamma \le \xi^\mathsf{T} A(x, \boldsymbol{y}_n)\, \xi \le \Gamma$ for every unit vector $\xi \in \mathbb{R}^d$, a.e. $x \in D$ and all $\boldsymbol{y}_n \in \boldsymbol{Y}_n$.*

For every $\varepsilon > 0$, a function A satisfying Assumption 2.1 induces a multiscale coefficient $A^\varepsilon \in L^\infty(D; \mathbb{R}^{d\times d}_{\mathrm{sym}})$ as follows:

$$A^\varepsilon(x) = A\Big(x, \frac{x}{\varepsilon_1}, \dots, \frac{x}{\varepsilon_n}\Big) \quad \text{for a.e.} \quad x \in D\,. \tag{2.1}$$

With such a coefficient, we consider the following model variational problem on $\mathring{V} = H^1_0(D)$:

$$\text{find } u^\varepsilon \in \mathring{V} \text{ such that } \quad \int_D (\nabla v)^\mathsf{T} A^\varepsilon\, \nabla u^\varepsilon = \int_D f v \quad \text{for all} \quad v \in \mathring{V}\,, \tag{2.2}$$

where $f \in L^2(D)$ is a forcing term. Assumption 2.1 and the Lax–Milgram theorem guarantee that this problem has a unique solution, which satisfies the stability bound

$$|u^\varepsilon|_{H^1(D)} \le \gamma^{-1} \sup_{v\in\mathring{V}\setminus\{0\}} \frac{|f(v)|}{|v|_{H^1(D)}} \le C\,\gamma^{-1} \sup_{v\in\mathring{V}\setminus\{0\}} \frac{|f(v)|}{\|v\|_{L^2(D)}} = C\,\gamma^{-1}\|f\|_{L^2(D)}\,,$$

where C is the classical Poincaré constant for D.

Although the forcing term f is assumed to be independent of the scale parameter ε for simplicity, we hasten to add that all results that follow admit a straightforward generalization to the case when f exhibits a microscale structure analogous to the one expressed by (2.1).

2.1. Low-rank tensor multilevel discretization. Starting from this section, we restrict the general setting to the case of $D = Y_1 = \cdots = Y_n = (0,1)^d$ for simplicity. In this section, we give an explicit construction of the low-rank tensor multilevel parametrization of functions, such as the solutions of multiscale problems of the form (2.2). Regarding the cases when D and $Y_1, \dots, Y_n$ are domains of more complex geometry, we only remark that they can be handled using the approach of [35, Section 5].

We start with defining, in Sections 2.1.1 and 2.1.2, the underlying *virtual grids* and the associated *background finite-element spaces* based on partitioning D into 2^{dL}

Cartesian-product elements, where $L \in \mathbb{N}$ is a fixed number of levels associated with the iterative uniform hierarchical partitioning of D.

Subsequently, in Section 2.1.3, we present the multilevel TT-MPS decomposition, which renders the aforementioned grids and finite-element spaces *virtual* in the sense that they are to be involved in computations only in a low-parametric representation. Finally, in Section 2.1.4, we elaborate on the relation of the TT-MPS decomposition to low-rank matrix representation and to subspace representation, which is crucial for the choice of this particular decomposition and for the subspace-factorization language that we use for our proofs thereafter.

2.1.1. Finite-element discretization on the unit interval. Let $L \in \mathbb{N}$ be a fixed number of levels of the low-rank tensor multilevel discretization to be constructed. For $[0,1]$, we consider a uniform partition into 2^L subintervals with the nodes

$$t_j^L = j \cdot 2^{-L} \quad \text{with} \quad j \in \{0, \dots, 2^L\}\,. \tag{2.3}$$

These points constitute a partitioning of $(0,1)$, inducing the *nodal continuous piecewise-linear* functions φ_j^L with $j \in \{0, \dots, 2^L\}$, defined by $\varphi_j^L(t_{j'}^L) = \delta_{jj'}$ for all $j, j' \in \{0, \dots, 2^L\}$, and the *piecewise-constant* functions $\bar{\varphi}_i^L$ with $i \in \{1, \dots, 2^L\}$, defined by $\bar{\varphi}_i^L|_{(t_{i'-1}^L)(t_{i'}^L)} = \delta_{ii'}$ for all $i, i' \in \{1, \dots, 2^L\}$.

Using the functions defined above, we introduce the following notations for finite-element spaces:

$$\bar{U}^L = \operatorname{span}\{\bar{\varphi}_j^L\}_{j=1}^{2^L}\,, \quad U^L = \operatorname{span}\{\varphi_j^L\}_{j=0}^{2^L} \quad \text{and} \quad \mathring{U}^L = \operatorname{span}\{\varphi_j^L\}_{j=1}^{2^L-1}\,. \tag{2.4}$$

The subspace $\mathring{U}^L \subset U^L$ accommodates the homogeneous essential boundary conditions of the problem (2.2).

To obtain coefficients of finite-element approximations with respect to these bases, we will use the *analysis operators* $\bar{\Phi}^L \colon L^2(0,1) \to \mathbb{C}^{2^L}$ and $\Phi^L \colon H^1(0,1) \to \mathbb{C}^{2^L}$ defined as follows: for all $w \in L^2(0,1)$, $v \in H^1(0,1)$ and $i \in \{1, \dots, 2^L\}$, we set

$$(\bar{\Phi}^L w)_i = 2^L \int_{t_{i-1}^L}^{t_i^L} w \quad \text{and} \quad (\Phi^L v)_i = v(t_i^L)\,. \tag{2.5}$$

Here and throughout, $\mathbb{C}$ denotes the field of complex numbers, which appears in our analysis because we express polynomial approximation in terms of Fourier approximation. Throughout this Section 2, $\mathbb{C}$ can be safely replaced with $\mathbb{R}$.

2.1.2. Finite-element discretization on D. Tensorizing the univariate basis functions defined above, we obtain d-variate basis functions that span the corresponding finite-element spaces:

$$\bar{V}^L = \bigotimes_{k=1}^{d} \bar{U}^L \subset L^2(D)\,, \quad V^L = \bigotimes_{k=1}^{d} U^L \subset V \quad \text{and} \quad \mathring{V}^L = \bigotimes_{k=1}^{d} \mathring{U}^L = V^L \cap \mathring{V}\,. \tag{2.6}$$

As in the case of the finite-element spaces (2.4) of univariate functions, the definition of $\mathring{V}^L \subset V^L$ serves to accommodate the boundary conditions of the problem (2.2). Classical approximation bounds (see, e.g., [19]) give

$$\inf_{v^L \in \mathring{V}^L} \|v - v^L\|_{H^1(D)} \le C\, 2^{-\alpha L}\, \|v\|_{H^{1+\alpha}(D)} \quad \text{for all} \quad v \in H^{1+\alpha}(D)\,, \tag{2.7}$$

where $\alpha \in (0,1]$ is a fractional order of Sobolev smoothness and $C > 0$ is a constant that depends on α but not on L.

Since the solution u^ε of (2.2) may exhibit algebraic singularities at the boundary of D due to the interplay of the domain's geometry, boundary conditions and diffusion coefficient, $u^\varepsilon \in H^{1+\alpha}(D)$ may hold only for α significantly less than one. To efficiently approximate such solutions in low-rank form, we follow [35, 40, 52] in using the multilevel QTT format for the low-rank separation of the indices associated with different levels and, for example, not different physical variables. This consists in applying the isomorphies

$$\bigotimes_{k=1}^{d} \mathbb{C}^{2^L} \simeq \bigotimes_{k=1}^{d}\bigotimes_{\ell=1}^{L} \mathbb{C}^2 \simeq \mathbb{C}^{2^{dL}} \simeq \bigotimes_{\ell=1}^{L}\bigotimes_{k=1}^{d} \mathbb{C}^2 \simeq \bigotimes_{\ell=1}^{L} \mathbb{C}^{2^d}, \tag{2.8}$$

so that the 2^{dL} degrees of freedom of V^L in (2.6) are indexed by L indices corresponding to the L levels of discretization, each taking 2^d values that enumerate the elements of the corresponding factor on the right-hand side of (2.8).

To use the separation of variables associated with levels (and not with the "physical" variables) in low-rank tensor approximation, we need a notational device to apply the isomorphies (2.8) and explicitly form these level variables. To this end, we define $\boldsymbol{\Pi}^L$ with $L \in \mathbb{N}$ as the permutation matrix of order 2^{dL} satisfying

$$\boldsymbol{\Pi}^L(\, i_{1,1}\,, \ldots,\, i_{d,1}, \ldots\ldots,\, i_{1,L}\,, \ldots,\, i_{d,L};\quad i_{1,1}\,, \ldots,\, i_{1,L}, \ldots\ldots,\, i_{d,1}\,, \ldots,\, i_{d,L}) = 1 \tag{2.9}$$

for all $i_{k\ell} \in \{1,2\}$ with $k \in \{1,\ldots,d\}$ and $\ell \in \{1,\ldots,L\}$.

Under the permutation realized by $\boldsymbol{\Pi}^L$, each "physical" index i_k with $k \in \{1,\ldots,d\}$, taking value in $\{1,\ldots,2^L\}$ and corresponding to the kth "physical" dimension, is decomposed into its L binary bits $i_{k,\ell}$ with $\ell \in \{1,\ldots,L\}$. A standard choice of isomorphism is given by

$$i_k = 1 + 2^{L-1}\cdot(i_{k,1}-1) + \cdots + 2^0\cdot(i_{k,L}-1) \tag{2.10}$$

for all $i_k \in \{1,\ldots,2^L\}$ and $i_{k,1},\ldots,i_{k,L} \in \{1,2\}$. The resulting dL binary bits are then reordered in groups of L, and the indices within each group $\ell \in \{1,\ldots,L\}$ can be merged into an index j_ℓ taking value in $\{1,\ldots,2^d\}$ and corresponding to the ℓth factor on the right-hand side of (2.8). The respective isomorphism is

$$j_\ell = 1 + 2^{d-1}\cdot(i_{1,\ell}-1) + \cdots + 2^0\cdot(i_{d,\ell}-1) \tag{2.11}$$

for all $j_\ell \in \{1,\ldots,2^d\}$ and $i_{1,\ell},\ldots,i_{d,\ell} \in \{1,2\}$.

The elements of $\mathring{V}^L$ and $\bar{V}^L$ can be parametrized by their coefficients extracted using the analysis operators

$$\bar{\Psi}^L = \boldsymbol{\Pi}^L \bigotimes_{k=1}^{d} \bar{\Phi}^L \colon L^2(D) \to \mathbb{C}^{2^{dL}} \quad \text{and} \quad \Psi^L = \boldsymbol{\Pi}^L \bigotimes_{k=1}^{d} \Phi^L \colon \big(H^1(0,1)\big)^{\otimes d} \to \mathbb{C}^{2^{dL}}. \tag{2.12}$$

Note that the restriction of Ψ^L to $\mathring{V}^L$ is not surjective. This lack of surjectivity stems from that we choose to use nested finite-element spaces $\mathring{V}^L$ with $L \in \mathbb{N}$ given by (2.6) but represent every function from $\mathring{V}^L$ with $L \in \mathbb{N}$ by 2^{dL} values instead of $(2^L-1)^d$, while the extraneous parameters are set equal to zero in agreement with the boundary conditions of the problem (2.2).

2.1.3. Low-rank tensor parametrization of functions. Consider a tensor[1] $\boldsymbol{u} \in \mathbb{C}^{n_1\cdots n_L}$ with $L \in \mathbb{N}$ dimensions and mode sizes $n_1,\ldots,n_L \in \mathbb{N}$. The vector $\boldsymbol{u}$ is

[1] It is standard to say that a "tensor" with dimensions $n_1,\ldots,n_L$ is an L-dimensional array, has size $n_1 \times \cdots \times n_L$ and is an element of $\mathbb{C}^{n_1\times\cdots\times n_L}$ (for complex-valued tensors). For convenience, we choose to work here with "long vectors" instead, calling them tensors. These are "vectorizations" of tensors in the usual meaning.

said to be represented in the *matrix-product state* (MPS) representation [72, 71, 70, 65], or in the *tensor-train* (TT) decomposition [62, 61], by

$$\boldsymbol{u}_{j_1,\dots,j_L} = \sum_{\alpha_1}^{r_1} \cdots \sum_{\alpha_{L-1}}^{r_{L-1}} U_1(\alpha_0, j_1, \alpha_1) \cdot U_2(\alpha_1, j_2, \alpha_2) \cdots U_L(\alpha_{L-1}, j_L, \alpha_L) \tag{2.13}$$

for all $j_\ell \in \{1, \dots, n_\ell\}$ with $\ell \in \{1, \dots, L\}$, where we use $\alpha_0 \equiv 1 \equiv \alpha_L$ and $r_0 = 1 = r_L$ for notational convenience. The arrays $U_1, \dots, U_L$ are called *factors* or *cores* (*sites* in the MPS literature), and the integer parameters $r_1, \dots, r_{L-1}$, governing the number of entries of the factors, are called *ranks* (referred to as *"bond dimensions"* in the MPS literature). The factors $U_1, \dots, U_L$ correspond to the indices $j_1, \dots, j_L$ separated in (2.13); they are multiplied in a specific way and, in particular, in a fixed order, so that the latter partly defines the decomposition.

To place our approach in a broader context, we note that the TT-MPS representation (2.13) is one of many tensor decompositions [27, 24, 26, 45], or *tensor networks* [57].

In the present paper, we use the TT-MPS representation as a *multilevel* tensor decomposition [68] for parametrizing discretizations with $L \in \mathbb{N}$ levels of iterative uniform hierarchical partitioning of the physical domain D. This means that the tensors we consider and represent in the form (2.13) have L dimensions of size

$$n_1 = \ldots = n_L = 2^d \,,$$

representing the levels of the discretization in the sense of (2.8) and (2.10)–(2.11). In the context of the TT-MPS decomposition, this has been known in the literature as the *quantized tensor-train (QTT) decomposition* [60, 43, 44, 45, 2].

Specifically in our setting, the decomposition of the form (2.13) is used to represent, in terms of the factors $U_1, \dots, U_L$, the unique finite-element function $u^L \in \mathring{V}^L$ or $u^L \in \bar{V}^L$ (see Section 2.1.2) such that, respectively, $\boldsymbol{u} = \mathring{\Psi}^L u^L$ or $\boldsymbol{u} = \bar{\Psi}^L u^L$ holds. The reason for parametrizing a function u^L from a low-order, generic finite-element space, such as $\mathring{V}^L$ or $\bar{V}^L$, via the QTT representation is as follows: the *number of effective parameters*, i.e., the total number

$$N = \sum_{\ell=1}^{L} r_{\ell-1}\, n_\ell\, r_\ell \tag{2.14}$$

of entries of all the factors $U_1, \dots, U_L$, may be moderate (in particular, much smaller than 2^{dL}) even when L is large. The number of effective parameters can be bounded from above using the *maximum rank*. We use (2.13) for $n_1 = \cdots = n_L = 2^d$, which gives

$$N \le 2^d L\, r_{\max}^2 \quad \text{for} \quad r_{\max} = \max\{r_0, \dots, r_L\} \,. \tag{2.15}$$

The goal of the present paper is to show, theoretically and experimentally, that the QTT representation, see (2.13) with (2.8), is well-suited for approximating the exact solution u^ε of (2.2). Specifically, for a fixed choice (i) of quasi-optimality constants $\tilde{C}_L$ with $L \in \mathbb{N}$, (ii) of the scale parameter $\varepsilon > 0$, (iii) of the desired accuracy $\delta > 0$ and (iv) of the number $L \in \mathbb{N}$ of levels such that

$$\tilde{C}_L \inf_{v^L \in \mathring{V}^L} \|v^L - u^\varepsilon\|_{H^1(D)} \le \delta \,, \tag{2.16}$$

we prove (in the analysis presented here) the existence of, and compute (in numerical experiments), functions $u^{\varepsilon,L}, v_1^{\varepsilon,L}, \dots, v_d^{\varepsilon,L} \in \bar{V}^L$ represented *exactly* by low-rank QTT decompositions of the form (2.13) for the tensors $\boldsymbol{u}, \boldsymbol{v}_1^L, \dots, \boldsymbol{v}_d^L$ given

by $\boldsymbol{u} = \bar{\Psi}^L u^{\varepsilon,L}$ and $\boldsymbol{v}_k^L = \bar{\Psi}^L v_k^{\varepsilon,L}$ for $k \in \{1,\dots,d\}$ and such that the tuple $\bar{u}^{\varepsilon,L} = (u^{\varepsilon,L}, v_1^{\varepsilon,L}, \dots, v_d^{\varepsilon,L})$ satisfies the quasi-optimal error bound

$$\|\bar{u}^{\varepsilon,L} - (u^\varepsilon, \nabla u^\varepsilon)\|_{L^2(D)^{d+1}} \le \tilde{C}_L \inf_{v^L \in \mathring{V}^L} \|v^L - u^\varepsilon\|_{H^1(D)} . \tag{2.17}$$

For such decompositions, we address the dependence of the maximum rank $r_{\max}$ (2.15) and of the numbers N of effective parameters (2.14) on the microscales $\varepsilon_1, \dots, \varepsilon_n$. The rationale for considering, for $L \in \mathbb{N}$, the approximation of u^ε and of its derivatives in $\bar{V}^L$ instead of approximating u^ε in $\mathring{V}^L$ with $L \in \mathbb{N}$ is presented in Section 2.2 below. Nodal basis functions from spaces similar to $\mathring{V}^L$ with $L \in \mathbb{N}$ are, however, used in our analysis for intermediate approximations related to the one-scale limit problem (3.6).

To illustrate the potential of our approach, we remark here that the complexity of low-rank approximations to the solution of the multiscale problem (2.2) may be drastically lower than $2^{dL} = \dim \bar{V}^L \simeq \dim \mathring{V}^L$ when L takes large values for $\mathring{V}^L$ or $\bar{V}^L$ to afford high accuracy. In our numerical experiments for $d = 2$ dimensions, we consider discretizations with up to $L = 50$ levels. and *compute* approximations that are quasi-optimal for each L by discretizing and *solving* the multiscale problem. For the finest microscale $\varepsilon_n = 2^{-40}$, the approximations computed have the number N of effective parameters not exceeding $20,000$ to $100,000$ and the maximum rank $r_{\max}$ not exceeding 20 to 50 (depending on the specific problem considered; see Figures 5.3 and 5.4). These values of N are to be contrasted with 2^{dL}, which reaches the value of $2^{100} \approx 10^{30}$. The reason for the efficiency of the TT-MPS decomposition is its inherent relation to subspace approximation and its reliance on well-established matrix algorithms. We elaborate on this point in Section 2.1.4.

2.1.4. TT-MPS representation and subspace factorization. Let us revisit the TT-MPS decomposition (2.13) of a tensor $\boldsymbol{u} \in \mathbb{C}^{n_1 \cdots n_L}$. For each $\ell \in \{1, \dots, L-1\}$, consider the ℓth *unfolding matrix* $\mathcal{U}_\ell \in \mathbb{C}^{(n_1 \cdots n_\ell) \times (n_{\ell+1} \cdots n_L)}$ of $\boldsymbol{u}$, formed from $\boldsymbol{u}$ by merging the first ℓ dimensions into a row dimension and the remaining $L - \ell$ dimensions, into a column dimension:

$$(\mathcal{U}_\ell)_{j_1,\dots,j_\ell;\, j_{\ell+1},\dots,j_L} = \boldsymbol{u}_{j_1,\dots,j_\ell,j_{\ell+1},\dots,j_L} \tag{2.18}$$

for all $j_k \in \{1, \dots, n_k\}$ with $k \in \{1, \dots, L\}$. Then (2.13) implies

$$\operatorname{rank} \mathcal{U}_\ell \le r_\ell \tag{2.19}$$

for every $\ell \in \{1, \dots, L-1\}$.

On the other hand, the decomposition (2.13) implies that, for every $\ell \in \{1, \dots, L-1\}$,

$$\boldsymbol{u} \in \mathscr{L}_\ell^L \otimes \mathbb{C}^{n_{\ell+1} \cdots n_L} \quad \text{with} \quad \mathscr{L}_\ell \subset \mathbb{C}^{n_1 \cdots n_\ell} \quad \text{such that} \quad \dim \mathscr{L}_\ell \le r_\ell \tag{2.20}$$

holds for $\mathscr{L}_\ell = \mathsf{im}\, \mathscr{V}_\ell$, where "$\mathsf{im}$" denotes the image (range) of a matrix, and

$$\boldsymbol{u} \in \mathbb{C}^{n_1 \cdots n_\ell} \otimes \mathscr{M}_\ell^L \quad \text{with} \quad \mathscr{M}_\ell \subset \mathbb{C}^{n_{\ell+1} \cdots n_L} \quad \text{such that} \quad \dim \mathscr{M}_\ell \le r_\ell \tag{2.21}$$

holds for $\mathscr{M}_\ell = \mathsf{im}\, \mathscr{W}_\ell$.

Condition (2.19) shows that (2.13) with every single $\ell \in \{1, \dots, L-1\}$, *is* a rank-$r_\ell$ factorization of the ℓth unfolding matrix $\mathcal{U}_\ell$ of $\boldsymbol{u}$. This is often referred to as the *separation of variables* in $\boldsymbol{u}$; namely, that of the first ℓ indices from the remaining $L - \ell$ indices.

On the other hand, conditions (2.20) and (2.21) with every single $\ell \in \{1, \dots, L-1\}$ express the same separation of variables in the language of tensor-product subspaces. These subspaces are given in terms of the factors $U_1, \dots, U_L$, are *highly adapted* to $\boldsymbol{u}$ and cannot be chosen *a priori*, without knowing $\boldsymbol{u}$.

We note that the converse is also true: if, for some $r_1, \dots, r_{L-1} \in \mathbb{N}$, a tensor $\boldsymbol{u}$ satisfies (2.19), (2.20) or (2.21) for $\ell \in \{1, \dots, L-1\}$, then it has a decomposition of the form (2.13) with ranks $r_1, \dots, r_{L-1}$ and some factors $U_1, \dots, U_L$. For details, we refer the reader to [26, 71, 70, 62, 61]. Here, in the case of the QTT decomposition, i.e., of (2.13) with $n_1 = \cdots = n_L = 2^d$ and (2.8), it is convenient to equivalently express the existence of a low-rank representation in the language of the factorization of subspaces, which we now introduce.

DEFINITION 2.2 (multilevel factorization of subspaces). *Let $d \in \mathbb{N}$ and $L \in \mathbb{N}$. We say that a subspace S of $\bar{V}^L$ (or of V^L)* factorizes at level $\ell \in \{0, \dots, L\}$ with rank $r \in \mathbb{N}_0$ *if there exists a subspace $\mathscr{L} \subset \mathbb{C}^{d\ell}$ such that $\dim \mathscr{L} \le r$ and $\bar{\Psi}^L(S) \subset \mathscr{L} \otimes \mathbb{C}^{2^{d(L-\ell)}}$ (or, respectively, $\Psi^L(S) \subset \mathscr{L} \otimes \mathbb{C}^{2^{d(L-\ell)}}$).*

As follows from the above discussion, for $n_1 = \cdots = n_L = 2^d$, factors $U_1, \dots, U_L$ satisfying (2.13) with ranks $r_1, \dots, r_{L-1}$ exist if and only if $\operatorname{span}\{\boldsymbol{u}\}$ factorizes at levels $1, \dots, L-1$ with ranks $r_1, \dots, r_{L-1}$ in the sense of Definition 2.2. Besides, if $m \in \mathbb{N}$, $\boldsymbol{u}_1, \dots, \boldsymbol{u}_m \in \mathbb{C}^{2^{dL}}$ and every $\boldsymbol{u}_i$ with $i \in \{1, \dots, m\}$ has a decomposition of the form (2.13) with $n_1 = \cdots = n_L = 2^d$ and ranks $r_1^{(i)}, \dots, r_{L-1}^{(i)}$, then $\operatorname{span}\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_m\}$ factorizes at levels $\ell = 1, \dots, L-1$ with ranks $\sum_{i=1}^m r_1^{(i)}, \dots, \sum_{i=1}^m r_{L-1}^{(i)}$.

REMARK 2.3. *Many elementary functions (exponential, polynomial, trigonometric, etc.) exhibit [25, 44] the low-rank structure defined by (2.13), with or without (2.8). In the present paper, we use this property of algebraic polynomials and of exponential functions in the form stated in Propositions 4.7 and 4.8 below. Together with suitable regularity results, such properties can be used to prove that the solutions of PDEs possess the low-rank structure of the same type, meaning that* some *choice of $L-1$ subspaces satisfying (2.20) or (2.21) of dimensions moderately depending on the desired approximation accuracy is readily available via classical approximation. This type of argument has been employed, e.g., in [25, 39, 35, 40, 38, 52], where the approximability of the solutions of certain PDEs with accuracy converging root-exponentially with respect to the number N of effective parameters was established.*

REMARK 2.4. *The TT-MPS decomposition comes with robust arithmetic: not only can basic linear operations (addition, multiplication, etc.) be performed immediately on the factors (such as $U_1, \dots, U_L$ above) without ever representing any of the $n_1 \cdots n_L$ individual entries of $\boldsymbol{u}$, but also decompositions with excessive ranks can be efficiently and reliably truncated using well-established algorithms for low-rank matrix approximation. In particular, approximation with given ranks that is quasi-optimal with respect to the Frobenius norm of tensors (equivalently, with respect to the vector ℓ^2 norm of function coefficients) is achieved by what is known as* Schmidt decomposition *in the MPS literature [70, Lemma 1] and [65, Section 4.1.3] and as* TT-SVD *and* TT rounding *algorithms in the TT literature [62, Algorithms 1 and 2]. This truncation inherits adaptivity from the matrix SVD and may be seen as a way of implicitly constructing $L-1$ tensor-product subspaces (2.20) adapted to $\boldsymbol{u}$. These are constructed* computationally *and are typically superior to those constructed analytically using classical approximation techniques in approximability* proofs *(see Remark 2.3 above).*

2.2. Discrete multiscale problem and low-rank tensor parametrization.

For every $L \in \mathbb{N}$, we consider the following discretization of the problem (2.2):

$$\text{find } u^{\varepsilon,L} \in \mathring{V}^L \text{ such that } \quad \int_D (\nabla v^L)^\mathsf{T} A^\varepsilon \nabla u^{\varepsilon,L} = \int_D f v^L \quad \text{for all} \quad v^L \in \mathring{V}^L . \tag{2.22}$$

As for the original problem, Assumption 2.1 and the Lax–Milgram theorem guarantee that the above discretization has a unique solution. By Céa's lemma,

$$\|u^{\varepsilon,L} - u^\varepsilon\|_{H^1(D)} \le \Gamma\gamma^{-1} \inf_{v^L \in \mathring{V}^L} \|v^L - u^\varepsilon\|_{H^1(D)} \le C\,\Gamma\gamma^{-1}\, 2^{-\alpha L} \|u^\varepsilon\|_{H^{1+\alpha}(D)} , \tag{2.23}$$

where C is the constant appearing in the approximation bound (2.7).

For any fixed $\varepsilon > 0$, the number L of levels may be increased in order to achieve any desired accuracy of approximating the solution of the multiscale problem (2.2) in the sense of (2.17). In the idealized setting when the discretization (2.22) can be solved exactly or another quasi-optimal projection onto $\bar{V}^L$ can be constructed exactly for any $L \in \mathbb{N}$, arbitrarily low errors can be achieved due to (2.23) and (2.7) respectively. The presence of the microscale delays the convergence of the discrete solutions to the exact solution: the factor $\|u^\varepsilon\|_{H^{1+\alpha}(D)}$ in (2.23) and (2.7) is approximately compensated by a number of levels proportional to $\lambda_n = \log_2 \varepsilon_n^{-1}$, which is required to start resolving the microscale structure. Then a suitable number of additional levels allows to achieve any desired accuracy in exact arithmetic.

The coefficient $\boldsymbol{u}^{\varepsilon,L} = \Psi^L u^{\varepsilon,L}$ of the *exact discrete* solution is characterized by the Galerkin optimality condition in the form of a linear system

$$\boldsymbol{A}^{\varepsilon,L}\, \boldsymbol{u}^{\varepsilon,L} = \boldsymbol{f}^L \tag{2.24}$$

with appropriately defined data: a symmetric matrix $\boldsymbol{A}^{\varepsilon,L}$ and a right-hand side vector $\boldsymbol{f}^L$. These can be accurately represented with low ranks in the QTT decomposition, see (2.13) with (2.8). Specifically, accurate low-rank approximations of the right-hand side vector may be constructed by classical approximation using the regularity of f, as we mention in the discussion following Remark 2.3 (the stipulation of Assumption 4.1 is sufficient). The same is true for the entries of the diffusion coefficient. Finally, the matrix $\boldsymbol{A}^{\varepsilon,L}$ can be represented in a low-rank form of matrices corresponding to (2.13); see, e.g., [41] and [10, Section 2.3].

For large L, however, linear systems of the form (2.24) are ill-conditioned and cannot be solved accurately *in finite-precision computations*. Besides, numerical solvers for linear systems in the TT-MPS representation rely on frequent low-rank truncation (rounding, see Remark 2.4 above), which allows to control the associated error only in the ℓ^2 norm of the coefficient tensor. Controlling the H^1 accuracy of the corresponding finite-element function under such truncation is impossible in practice due to the ill-conditioning of the basis introduced for $\mathring{V}^L$ in Section 2.1.2 with respect to the H^1 norm. As a result, the discrete solution $u^{\varepsilon,L}$ cannot be accurately computed in terms of $\boldsymbol{u}^{\varepsilon,L} = \Psi^L u^{\varepsilon,L}$ for practically relevant values of L. An effectual remedy to this in the form of a tensor-structured BPX-type preconditioner was developed in [10].

Literature on numerical tensor-structured solvers for linear systems is vast; we only mention here several works: on tensorized solvers [50, 14, 12, 11], which are based on implementing standard numerical methods with the use of TT-MPS arithmetic, alternating-optimization methods [31, 64, 63, 59], which rely on the multilinearity of tensor decompositions, and Riemannian-optimization methods [69, 49], which exploit the geometry of sets of tensors defined by hard rank constraints.

3. Reiterated homogenization and high-dimensional one-scale limit.

For analysis, instead of the original multiscale problem (2.2), we consider a one-scale high-dimensional limit problem posed in (3.6) in this section. The derivation of the one-scale limit problem goes back to [55, 3, 4]. It is obtained from the original multiscale problem (2.2) by homogenization, analyzed for a single ($n = 1$) microscale in [15, 13, 34, 54, 55, 3], and for $n > 1$ microscales by iteration in [4]. For a general discussion, we refer to [23]. For every $\varepsilon > 0$, the solution of the one-scale limit problem induces an approximation of the solution of the corresponding multiscale problem. We invoke this approximation only for analysis, as an auxiliary approximation.

3.1. One-scale high-dimensional limit problem.

To formulate reiterated homogenization, we consider the following assumption, of which Assumption 2.1 is a particular case with $i = n$ and $A_n = A$. We recall that, for a space W of functions with domain Y, the notation $W_\#$ denotes the subspace of Y-periodic functions of W.

ASSUMPTION 3.1 (on coefficient A_i with $i \in \{0,\dots,n\}$ microscales, with ellipticity and continuity constants γ and Γ). $A_i \in L^\infty(D; C_\#(\boldsymbol{Y}_i; \mathbb{R}^{d\times d}_{\mathrm{sym}}))$ *is uniformly positive definite with constants* Γ *and* γ*:* $\gamma \le \xi^\mathsf{T} A_i(x, \boldsymbol{y}_i)\,\xi \le \Gamma$ *for every unit vector* $\xi \in \mathbb{R}^d$*, a.e.* $x \in D$ *and all* $\boldsymbol{y}_i \in \boldsymbol{Y}_i$.

For each step $i \in \{1,\dots,n\}$ of homogenization, we define

$$\begin{aligned} V_i &= L^2\big(D \times \boldsymbol{Y}_{i-1}, H^1_\#(Y_i)/_\mathbb{R}\big) \simeq L^2(D) \otimes L^2(Y_1) \otimes \cdots \otimes L^2(Y_{i-1}) \otimes H^1_\#(Y_i)/_\mathbb{R}\,, \\ W_i &= L^\infty(D \times \boldsymbol{Y}_{i-1}, H^1_\#(Y_i)/_\mathbb{R})\,, \end{aligned} \tag{3.1}$$

where $H^1_\#(Y_i)/_\mathbb{R}$ is the closed subspace of $H^1_\#(Y_i)$ that consists of functions with vanishing mean, and consider the Cartesian-product space

$$\boldsymbol{V}_i = \mathring{V} \times V_1 \times \cdots \times V_i \tag{3.2}$$

endowed with the inner product $\langle\cdot,\cdot\rangle_{\boldsymbol{V}_i}$ given by

$$\langle \boldsymbol{\psi}, \boldsymbol{\phi}\rangle_{\boldsymbol{V}_i} = \sum_{|\alpha|=1} \langle \partial^\alpha \psi_0, \partial^\alpha \phi_0\rangle_{L^2(D)} + \sum_{j=1}^{i} \sum_{|\alpha_j|=1} \langle \partial_j^{\alpha_j} \psi_j, \partial_j^{\alpha_j} \phi_j\rangle_{L^2(D\times \boldsymbol{Y}_j)} \tag{3.3}$$

for all $\boldsymbol{\psi} = (\psi_0, \psi_1, \dots, \psi_i), \boldsymbol{\phi} = (\phi_0, \phi_1, \dots, \phi_i) \in \boldsymbol{V}_i$. We denote the norm induced by $\langle\cdot,\cdot\rangle_{\boldsymbol{V}_i}$ with $\|\cdot\|_{\boldsymbol{V}_i}$. Here and throughout, the symbol ∂^α with $\alpha \in \mathbb{N}_0^d$ denotes the differentiation of functions with respect to the first d-dimensional variable, taking value in D, indicated by the multi-index α, whereas ∂_j^α with $j \in \{1,\dots,n\}$ and $\alpha \in \mathbb{N}_0^d$ denotes differentiation with respect to the $(j+1)$th d-dimensional variable, taking value in Y_j, according to the multi-index α. Further, we define a bilinear form $\mathsf{B}_i\colon \boldsymbol{V}_i \times \boldsymbol{V}_i \to \mathbb{R}$:

$$\mathsf{B}_i(\boldsymbol{\psi}, \boldsymbol{\phi}) = \int_{D\times \boldsymbol{Y}_i} \Big(\nabla\psi_0 + \sum_{j=1}^{i} \nabla_j \psi_j\Big)^\mathsf{T} A_i \Big(\nabla\phi_0 + \sum_{j=1}^{i} \nabla_j \phi_j\Big) \tag{3.4}$$

for all $\boldsymbol{\psi} = (\psi_0, \psi_1, \dots, \psi_i), \boldsymbol{\phi} = (\phi_0, \phi_1, \dots, \phi_i) \in \boldsymbol{V}_i$, where A_i is a matrix function satisfying Assumption 3.1 with i microscales and with positive constants γ and Γ and ∇_j with $j \in \{1,\dots,n\}$ denotes differentiation with respect to the $(j+1)$th d-dimensional variable, taking value in Y_j. Then the bilinear form B_i is continuous and coercive: the inequalities

$$\gamma\,\|\boldsymbol{\phi}\|^2_{\boldsymbol{V}_i} \le \mathsf{B}_i(\boldsymbol{\phi}, \boldsymbol{\phi}) \quad \text{and} \quad \mathsf{B}_i(\boldsymbol{\psi}, \boldsymbol{\phi}) \le \Gamma\,\|\boldsymbol{\psi}\|_{\boldsymbol{V}_i}\|\boldsymbol{\phi}\|_{\boldsymbol{V}_i} \tag{3.5}$$

hold for all $\boldsymbol{\psi}, \boldsymbol{\phi} \in \boldsymbol{V}_i$. Then, since $f \in L^2(D)$, the problem of finding $\boldsymbol{u} \in \boldsymbol{V}_i$ such that

$$\mathsf{B}_i(\boldsymbol{u}, \boldsymbol{\phi}) = \int_D f\phi_0 \quad \text{for all} \quad \boldsymbol{\phi} = (\phi_0, \phi_1, \dots, \phi_i) \in \boldsymbol{V}_i \tag{3.6}$$

has a unique solution $\boldsymbol{u} = (u_0, u_1, \dots, u_i)$ (by the Lax–Milgram theorem). We remark that the bilinear forms $\mathsf{B}_1, \dots, \mathsf{B}_n$ in (3.6) satisfy property (3.5) with constants uniform with respect to the scale parameter ε.

The problem (3.6) with $i = n$ microscales, representing the result of n iterations of homogenization applied to the original multiscale problem (2.2).

For notational convenience, we introduce

$$v_i = \sum_{j=0}^{i} \nabla_j u_j \quad \text{for} \quad i \in \{1,\dots,n\}\,. \tag{3.7}$$

In the next section, we discuss the approximation of the solution u^ε of the multiscale problem (2.2) by u_0 and of ∇u^ε, using v_n.

3.2. Approximation of the multiscale problem via the one-scale limit problem. In the analysis of homogenization, the limit for $\varepsilon \to 0$ is considered under the additional assumption that the microscales converge to zero and *asymptotically separate*:

$$\lim_{\varepsilon\to 0} \frac{\varepsilon_i}{\varepsilon_{i-1}} = 0 \quad \text{for} \quad i = 1, \dots, n\,, \tag{3.8}$$

where we use convention $\varepsilon_0 \equiv 1$. Then the one-scale limit problem (3.6) approximates the multiscale problem in the following sense.

The physical solution u^ε, including the oscillations in its gradient induced by the multiscale structure of the diffusion coefficient (2.1), can be approximated in terms of the solution of the one-scale high-dimensional limit problem. Specifically, the error of approximating u^ε with $\tilde{u}^\varepsilon$ defined by

$$\tilde{u}^\varepsilon(x) = u_0(x) + \sum_{i=1}^{n} \varepsilon_i u_i\Big(x, \frac{x}{\varepsilon_1}, \dots, \frac{x}{\varepsilon_i}\Big) \quad \text{for all} \quad x \in D\,, \tag{3.9}$$

which is a homogenized solution with microscale correctors, converges to zero as $\varepsilon \to 0$ under additional smoothness assumptions.

PROPOSITION 3.2 (Theorem 2.14 in [4]). *Assume that the solution $(u_0, u_1, \dots, u_n)$ of the one-scale limit problem (3.6) satisfies $u_0 \in C^1(\overline{D})$ and $u_i \in C^1(\overline{D}, C^1_\#(\boldsymbol{Y}_i))$ for all $i \in \{1, \dots, n\}$. Then $\tilde{u}^\varepsilon - u^\varepsilon \to 0$ in $H^1(D)$ as $\varepsilon \to 0$.*

For the convergence claimed in Proposition 3.2, the rate can be quantified in the case of a single ($n = 1$) microscale as follows.

PROPOSITION 3.3 (Proposition 3.5 and the subsequent discussion in [30]). *For $n = 1$, assume that $A \in C^\infty(\overline{D}, C^\infty_\#(Y_1))^{d\times d}_{\mathrm{sym}}$ and that the homogenized solution u_0 belongs to $H^2(D)$. Then*

$$\|u^\varepsilon - \tilde{u}^\varepsilon\|_{H^1(D)} \le C\varepsilon_1^{\frac{1}{2}}\,. \tag{3.10}$$

The constant C is independent of ε_1 but depends on u_0 and u_1.

The homogenization approximation (3.9), with pointwise correctors, is standard [15, 13, 34] but requires the additional smoothness from the solution $\boldsymbol{u} = (u_0, u_1, \dots, u_n) \in \boldsymbol{V}_n$ of the one-scale limit problem. In [56, 55, 22, 21], an alternative homogenization approximation constructed using *unfolding* and *averaging operators* was proposed, also in the case of a single ($n = 1$) microscale. An obvious advantage of this alternative homogenization approximation is that it is well defined in terms of the solution $\boldsymbol{u} = (u_0, u_1) \in \boldsymbol{V}_1$ of the one-scale limit problem without additional smoothness assumptions.

DEFINITION 3.4 (see Definitions 2.1 and 2.16 in [21]). *For any $\varepsilon_1 > 0$, the corresponding* unfolding *operator $\mathcal{T}^{\varepsilon_1}\colon L^2(D) \to L^2(D \times Y_1)$ and the* averaging *operator $\mathcal{U}^{\varepsilon_1}\colon L^2(D \times Y_1) \to L^2(D)$ are defined by*

$$(\mathcal{T}^{\varepsilon_1}\phi)(x, y) = \phi\Big(\varepsilon_1\Big[\frac{x}{\varepsilon_1}\Big] + \varepsilon_1 y\Big)$$

for a.e. $(x, y) \in D \times Y_1$ and every $\phi \in L^2(D)$, where ϕ is extended by zero outside its domain, and by

$$(\mathcal{U}^{\varepsilon_1}\Phi)(x) = |Y_1|^{-1} \int_{Y_1} \Phi\Big(\varepsilon_1\Big[\frac{x}{\varepsilon_1}\Big] + \varepsilon_1 z, \Big\{\frac{x}{\varepsilon_1}\Big\}\Big)\,\mathrm{d}z$$

for a.e. $x \in D$ and every $\Phi \in L^2(D \times Y_1)$. Here, the notations "$[\,\cdot\,]$" and "$\{\,\cdot\,\}$" are used for the integral and fractional parts of the real argument.

The application of the unfolding and averaging operators to vector-valued functions (such as gradients) is meant componentwise.

Certain basic properties of the unfolding and averaging operators were analyzed in [21]. In particular, by [21, Proposition 2.17], the operator $\mathcal{U}^{\varepsilon_1}\colon L^2(D\times Y_1)\to L^2(D)$ is continuous and has norm $|Y_1|^{-1/2}$.

The following result is obtained by bounding the L^2 norm of u_0-u^ε using [30, Proposition 3.5] with the subsequent discussion and [30, Lemma 3.8] and by bounding the L^2 norm of $\mathcal{U}^\varepsilon v_1-\nabla u^\varepsilon$ using the argument given in the proof of [30, Theorem 3.9].

PROPOSITION 3.5. *For $n=1$, the following bound holds under the hypothesis of Proposition 3.3:*

$$\|(u_0,\mathcal{U}^{\varepsilon_1}v_1)-(u^\varepsilon,\nabla u^\varepsilon)\|_{L^2(D)^{d+1}}\le C\varepsilon_1^{\frac{1}{2}}\,. \tag{3.11}$$

The constant C is independent of ε_1 but depends on u_0 and u_1.

In the case when A^ε does not depend on the macroscale variable, analogous bounds were justified in [21] under the assumption $u_0\in H^{1+\alpha}(D)$ with $\alpha\in(\frac{1}{2},1)$.

To the best of our knowledge, no results analogous to those of Propositions 3.3 and 3.5 are available for problems with $n>1$ microscales.

Following [30], we introduce in this section n-microscale unfolding and averaging operators, which generalize those introduced [21, Definitions 2.1 and 2.16] in the case of a single ($n=1$) microscale (see Definition 3.4 above).

DEFINITION 3.6. *For every $i\in\{1,\dots,n\}$, we define the unfolding and averaging operators corresponding to the ith microscale as follows:*

$$\begin{aligned}\mathcal{T}_i^\varepsilon&=\mathcal{T}^{\varepsilon_i}\otimes\mathsf{id}_{Y_{i+1}\times\cdots\times Y_n}\colon L^2(D\times Y_{i+1}\times\cdots\times Y_n)\to L^2(D\times Y_i\times\cdots\times Y_n)\,,\\ \mathcal{U}_i^\varepsilon&=\mathcal{U}^{\varepsilon_i}\otimes\mathsf{id}_{Y_{i+1}\times\cdots\times Y_n}\colon L^2(D\times Y_i\times\cdots\times Y_n)\to L^2(D\times Y_{i+1}\times\cdots\times Y_n)\,.\end{aligned}$$

Further, for every $i\in\{1,\dots,n\}$, the n-microscale unfolding and averaging operators are defined as

$$\mathcal{T}^\varepsilon=\mathcal{T}_1^\varepsilon\circ\cdots\circ\mathcal{T}_n^\varepsilon\colon L^2(D)\to L^2(D\times\boldsymbol{Y}_n)$$

and

$$\mathcal{U}^\varepsilon=\mathcal{U}_n^\varepsilon\circ\cdots\circ\mathcal{U}_1^\varepsilon\colon L^2(D\times\boldsymbol{Y}_n)\to L^2(D)\,.$$

The definition of the n-microscale unfolding and averaging operators is given above in the form of the iterated composition of those corresponding to individual scales, which allows to carry out the analysis of averaging in Section 4.3.5 in a similar, iterated fashion. We note that the above definition is equivalent to [30, Definitions 3.11 and 3.12].

Certain basic properties of the n-microscale unfolding and averaging operators follow from those of the single-scale operators. In particular, for each $i\in\{1,\dots,n\}$, the operator $\mathcal{U}_i^\varepsilon\colon L^2(D\times Y_i\times\cdots\times Y_n)\to L^2(D\times Y_{i+1}\times\cdots\times Y_n)$ is continuous and has norm $|Y_i|^{-1/2}$. As a result,

$$\|\mathcal{U}^\varepsilon(\Phi-\widetilde{\Phi})\|_{L^2(D)}\le\|\Phi-\widetilde{\Phi}\|_{L^2(D\times\boldsymbol{Y}_n)}\quad\text{for all}\quad\Phi,\widetilde{\Phi}\in L^2(D\times\boldsymbol{Y}_n)\,. \tag{3.12}$$

As in [21], one can show that the solution u^ε of the multiscale problem (2.2) under the scale-separation condition (3.8) satisfies

$$\lim_{\varepsilon\to0}\|\mathcal{T}^\varepsilon\nabla u^\varepsilon-v_n\|_{L^2(D\times\boldsymbol{Y}_n)}=0\,. \tag{3.13}$$

Using the folding operator $\mathcal{U}^\varepsilon$, we can state an analog of (3.10) for several microscales, showing that the scale-interaction functions $u_1,\dots,u_n$ in (3.13) describe, to the leading order, the oscillations of the functions u^ε with $\varepsilon>0$ as they approach the weak limit u^0.

LEMMA 3.7. *Under the scale-separation condition (3.8), for the multiscale problem (2.2) we have* $\nabla u^\varepsilon - \mathcal{U}^\varepsilon v_n \to 0$ *in* $L^2(D)$ *as* $\varepsilon \to 0$.

For a proof, we refer to [21, Theorem 6.1] for the case $n = 1$ of a single microscale and [21, Remark 7.5] regarding the case of $n > 1$ microscales. When $u_0, u_1, \dots, u_n$ are sufficiently smooth, this result can be inferred from the corrector result stated in Proposition 3.2.

3.3. Recurrence for scale-interaction functions. Let $i \in \{1, \dots, n\}$ and assume that A_i is a matrix function satisfying Assumption 3.1 with i microscales and positive constants γ and Γ. Then the limit problem (3.6), posed on $D \times \boldsymbol{Y}_i$, is well posed and has a unique solution.

Assume that $\xi \in \mathbb{R}^d$ is a unit vector. For a.e. $(x, \boldsymbol{y}_{i-1}) \in D \times \boldsymbol{Y}_{i-1}$, define a bilinear form $\mathsf{b}_i(x, \boldsymbol{y}_{i-1}, \cdot, \cdot) : H^1_\#(Y_i)/\mathbb{R} \times H^1_\#(Y_i)/\mathbb{R} \to \mathbb{R}$ and a linear form $\mathsf{f}_i(x, \boldsymbol{y}_{i-1}, \xi, \cdot) \colon H^1_\#(Y_i)/\mathbb{R} \to \mathbb{R}$ as follows:

$$\begin{aligned} \mathsf{b}_i(x, \boldsymbol{y}_{i-1}, \psi, \phi) &= \int_{Y_i} (\nabla\psi)^\mathsf{T} A_i(x, \boldsymbol{y}_{i-1}, \cdot)\, \nabla\phi\,, \\ \mathsf{f}_i(x, \boldsymbol{y}_{i-1}, \xi, \phi) &= -\int_{Y_i} \xi^\mathsf{T} A_i(x, \boldsymbol{y}_{i-1}, \cdot)\, \nabla\phi \end{aligned} \tag{3.14}$$

for all $\psi, \phi \in H^1_\#(Y_i)/\mathbb{R}$. Then the following holds for a.e. $(x, \boldsymbol{y}_{i-1}) \in D \times \boldsymbol{Y}_{i-1}$.

First, the assumption regarding A_i results in the continuity and ellipticity of $\mathsf{b}_i(x, \boldsymbol{y}_{i-1}, \cdot, \cdot)$: for all $\psi, \phi \in H^1_\#(Y_i)/\mathbb{R}$, $\mathsf{b}_i(x, \boldsymbol{y}_{i-1}, \psi, \phi) \le \Gamma |\psi|_{H^1(Y_i)} |\phi|_{H^1(Y_i)}$ and $\mathsf{b}_i(x, \boldsymbol{y}_{i-1}, \phi, \phi) \ge \gamma \|\phi\|^2_{H^1_\#(Y_i)/\mathbb{R}}$. Second, by the same argument, the linear form $\mathsf{f}_i(x, \boldsymbol{y}_{i-1}, \xi, \cdot)$ is continuous:

$$\big|\mathsf{f}_i(x, \boldsymbol{y}_{i-1}, \xi, \phi)\big| \le \Gamma \|\phi\|_{H^1_\#(Y_i)/\mathbb{R}} \quad \text{for all} \quad \phi \in H^1_\#(Y_i)/\mathbb{R}\,.$$

By the Lax–Milgram theorem, the problem of finding $\mathsf{w}_\xi(x, \boldsymbol{y}_{i-1}, \cdot) \in H^1_\#(Y_i)/\mathbb{R}$ such that

$$\mathsf{b}_i(x, \boldsymbol{y}_{i-1}, \mathsf{w}_\xi(x, \boldsymbol{y}_{i-1}, \cdot), \phi) = \mathsf{f}_i(x, \boldsymbol{y}_{i-1}, \xi, \phi) \quad \text{for all} \quad \phi \in H^1_\#(Y_i)/\mathbb{R}\,. \tag{3.15}$$

admits a unique solution, which satisfies $\|\mathsf{w}_\xi(x, \boldsymbol{y}_{i-1}, \cdot)\|_{H^1_\#(Y_i)/\mathbb{R}} \le \gamma^{-1}\Gamma$.

Let $\xi_1, \dots, \xi_d$ be the columns of the identity matrix I of order d. Being valid for a.e. $(x, \boldsymbol{y}_{i-1}) \in D \times \boldsymbol{Y}_{i-1}$ and every unit vector $\xi \in \mathbb{R}^d$, the above argument defines $w_i \in W_i^d$ whose components $w_{ik} \in W_i$ with $k \in \{1, \dots, d\}$ are given by $w_{ik}(x, \boldsymbol{y}_{i-1}, y_i) = \mathsf{w}_{\xi_k}(x, \boldsymbol{y}_{i-1}, y_i)$ for a.e. $(x, \boldsymbol{y}_{i-1}, y_i) \in D \times \boldsymbol{Y}_{i-1} \times Y_i$ and for each $k \in \{1, \dots, d\}$. Note that w_i is also an element of V_i^d. Furthermore, it is the only element of V_i^d such that

$$\int_{D \times \boldsymbol{Y}_i} \big(I + \mathcal{J}_i w_i\big)\, A_i\, \nabla_i \phi = 0 \tag{3.16}$$

for all $\phi \in V_i$. Here, $\mathcal{J}_i$ denotes the differential operator returning the Jacobi matrix with respect to the last variable (varying in Y_i), as a function of all variables (taking value in $D \times \boldsymbol{Y}_i$). Since $A_i \in L^\infty(D; C_\#(\boldsymbol{Y}_i; \mathbb{R}^{d\times d}_{\mathrm{sym}}))$, one can define $A_{i-1} \in L^\infty(D; C_\#(\boldsymbol{Y}_{i-1}; \mathbb{R}^{d\times d}_{\mathrm{sym}}))$ by setting

$$\begin{aligned} A_{i-1}(x, \boldsymbol{y}_{i-1}) &= \int_{Y_i} \big(I + \mathcal{J}_i w_i(x, \boldsymbol{y}_{i-1}, \cdot)\big) A_i(x, \boldsymbol{y}_{i-1}, \cdot)\, \big(I + \mathcal{J}_i w_i(x, \boldsymbol{y}_{i-1}, \cdot)\big)^\mathsf{T} \\ &= \int_{Y_i} \big(I + \mathcal{J}_i w_i(x, \boldsymbol{y}_{i-1}, \cdot)\big) A_i(x, \boldsymbol{y}_{i-1}, \cdot) \end{aligned} \tag{3.17}$$

for a.e. $x \in D$ and for all $\boldsymbol{y}_{i-1} \in \boldsymbol{Y}_{i-1}$. By [54], see also [15, Chapter 1, Theorem 3.9],the matrix function A_{i-1}, to which we refer as *upscaled coefficient*,

satisfies Assumption 3.1 with $i-1$ microscales *and with the identical positive constants* γ *and* Γ. The corresponding problem (3.6), involving i variables, is therefore well posed and has a unique solution $(u_0,\dots,u_{i-1})\in \boldsymbol{V}_{i-1}$.

Since $u_{i-1}\in V_{i-1}$, we have $\nabla_{i-1}u_{i-1}\in L^2(D\times \boldsymbol{Y}_{i-1})^d$. On the other hand, we have noted that $w_i\in W_i^d$, so we can define $u_i\in V_i$ by setting

$$u_i(x,\boldsymbol{y}_{i-1},\cdot\,)=\big(w_i(x,\boldsymbol{y}_{i-1},\cdot\,)\big)^{\mathsf{T}}\,\nabla_{i-1}u_{i-1}(x,\boldsymbol{y}_{i-1})\quad\text{in}\quad H^1_{\#}(Y_i)/_{\mathbb{R}} \tag{3.18}$$

for a.e. $x\in D$ and $\boldsymbol{y}_{i-1}\in \boldsymbol{Y}_{i-1}$. Indeed, this entails that $u_i(x,\boldsymbol{y}_{i-1},\cdot\,)$ has the gradient

$$\nabla_i u_i(x,\boldsymbol{y}_{i-1},\cdot\,)=\mathcal{J}_i w_i(x,\boldsymbol{y}_i)\,\nabla_{i-1}u_{i-1}(x,\boldsymbol{y}_{i-1})\quad\text{in}\quad L^2(Y_i) \tag{3.19}$$

for a.e. $x\in D$ and $\boldsymbol{y}_{i-1}\in \boldsymbol{Y}_{i-1}$, so that the bound $\|u_i\|_{V_i}\lesssim \|w_i\|_{W_i^d}\,\|u_{i-1}\|_{V_{i-1}}$ holds (with a constant implied in "$\lesssim$" determined by the choice of a norm for W_i^d). This results in $(u_0,\dots,u_{i-1},u_i)\in \boldsymbol{V}_i$ and, as one verifies using (3.16) and (3.17), also in that this tuple solves the problem (3.6) with $i+1$ variables.

Applying the above argument iteratively, we obtain the "effective" macroscopic diffusion coefficient $A_0\in L^\infty(D;\mathbb{R}^{d\times d}_{\mathrm{sym}})$:

$$A_0=\int_{Y_1}\cdots\int_{Y_n}\big(I+\mathcal{J}_1w_1\big)\cdots\big(I+\mathcal{J}_nw_n\big)\,A\,, \tag{3.20}$$

which satisfies Assumption 3.1 with zero microscales and with ellipticity and continuity constants γ and Γ. The "effective" problem for the homogenized limit u_0 reads: find $u_0\in \mathring{V}$ such that

$$\int_D(\nabla\phi)^{\mathsf{T}}A_0\,\nabla u_0=\int_D f\phi \tag{3.21}$$

for every $\phi\in\mathring{V}$.

Then the solution $(u_0,\dots,u_n)\in \boldsymbol{V}_n$ of the limit problem (3.6) with $n+1$ variables can be solved using the recursion (3.18), so that the scale-interaction functions u_i and the sums of their gradients given by (3.7) satisfy

$$u_i=w_i^{\mathsf{T}}v_{i-1}\quad\text{and}\quad v_i=\big(I+\mathcal{J}_iw_i\big)^{\mathsf{T}}v_{i-1}=\big(I+\mathcal{J}_iw_i\big)^{\mathsf{T}}\cdots\big(I+\mathcal{J}_1w_1\big)^{\mathsf{T}}\nabla u_0 \tag{3.22}$$

in V_i and $L^2(D\times\boldsymbol{Y}_i)^d$ respectively.

3.4. Approximate recurrence for scale-interaction functions. To obtain low-rank tensor-structured approximations of $(u_0,u_1,\dots,u_n)\in \boldsymbol{V}_n$, we use the following approximation scheme with a discretization parameter $L\in\mathbb{N}$. For every $i=1,\dots,n$, we approximate w_i and $\mathcal{J}_iw_i$ by w_i^L and J_i^L in W_i^d and $L^\infty(D\times\boldsymbol{Y}_{i-1},L^2(Y_i))^{d\times d}$ respectively. Assuming that u_0 and ∇u_0 are approximated by u_0^L and v_0^L in V and $L^2(D)^d$ respectively, we follow (3.22) to define the corresponding approximations u_i^L and v_i^L to u_i and v_i with $i\in\{1,\dots,n\}$: in V_i and $L^2(D\times\boldsymbol{Y}_i)^d$ respectively, we set

$$u_i^L=\big(w_i^L\big)^{\mathsf{T}}v_{i-1}^L\quad\text{and}\quad v_i^L=\big(I+J_i^L\big)^{\mathsf{T}}v_{i-1}^L=\big(I+J_i^L\big)^{\mathsf{T}}\cdots\big(I+J_1^L\big)^{\mathsf{T}}v_0^L\,. \tag{3.23}$$

The associated errors can be represented by telescoping sums: for example,

$$\begin{aligned}v_i-v_i^L=&\big(I+\mathcal{J}_iw_i\big)^{\mathsf{T}}\cdots\big(I+\mathcal{J}_1w_1\big)^{\mathsf{T}}\big(v_0-v_0^L\big)\\&+\sum_{j=1}^{i}\Big\{\prod_{m=j+1}^{i}\big(I+\mathcal{J}_mw_m\big)^{\mathsf{T}}\Big\}\big(\mathcal{J}_jw_j-J_j^L\big)^{\mathsf{T}}\Big\{\prod_{m=1}^{j-1}\big(I+J_m^L\big)^{\mathsf{T}}\Big\}v_0^L\end{aligned}$$

for every $i \in \{1,\dots,n\}$, where sums and products over empty ranges are to be omitted. Assuming that the errors $w_i - w_i^L$, $\mathcal{J}_i w_i - J_i^L$ and $v_0 - v_0^L$ are bounded, respectively, in W_i^d, $L^\infty(D \times \boldsymbol{Y}_{i-1}, L^2(Y_i))^{d\times d}$ and $L^2(D)^d$ uniformly with respect to $L \in \mathbb{N}$ and $i \in \{1,\dots,n\}$, we obtain, with a positive equivalence constant independent of the discretization parameter $L \in \mathbb{N}$, the bounds

$$\big\|v_i - v_i^L\big\|_{L^2(D\times \boldsymbol{Y}_i)^d} \lesssim \|v_0 - v_0^L\|_{L^2(D)^d} + \sum_{j=1}^{i} \|\mathcal{J}_j w_j - J_j^L\|_{L^\infty(D\times \boldsymbol{Y}_{i-1}, L^2(Y_i))^{d\times d}} \quad (3.24)$$

and

$$\begin{aligned}\|u_i - u_i^L\|_{V_i} \lesssim\ & \|v_0 - v_0^L\|_{L^2(D)^d} \\ & + \sum_{j=1}^{i-1} \|\mathcal{J}_j w_j - J_j^L\|_{L^\infty(D\times \boldsymbol{Y}_{i-1}, L^2(Y_i))^{d\times d}} + \|w_i - w_i^L\|_{W_i} \quad (3.25)\end{aligned}$$

for $i \in \{1,\dots,n\}$. In Section 4.3.3, we construct particular approximations w_i^L, J_i^L, u_i^L and v_i^L with $i \in \{1,\dots,n\}$ and $L \in \mathbb{N}$ in the finite-element spaces specified in Section 4.1.2.

4. Approximability under the assumption of analyticity. In the present section, we investigate regularity and approximability of $u_0, u_1, \dots, u_n$. With the aim of establishing convergence rates and (quantized) tensor rank bounds which are independent of the scales, we impose additional assumptions on the data D, A and f. Specifically, we consider a *tensor-product physical domain* and *analytic data*.

The first set of additional assumptions consists in the following.

ASSUMPTION 4.1. *For every ε and $i \in \{0,1,\dots,n\}$, we have $\varepsilon_i = 2^{-\lambda_i}$ with $\lambda_i \in \mathbb{N}$ depending on ε (we set $\lambda_0 \equiv 0$ for notational convenience). For the physical domain and the unit cells, we have $D = Y_1 = \cdots = Y_n = (0,1)^d$. The diffusion coefficient A is analytic and one-periodic with respect to each of the last nd scalar variables on $\overline{D \times \boldsymbol{Y}_n}$. The right-hand side f is analytic on $\overline{D \times \boldsymbol{Y}_n}$.*

Assumption 4.1 allows to prove that the solution of the one-scale high-dimensional limiting problem can be approximated by finite-element functions of tensor ranks that are logarithmic in accuracy. This implies that the solution of the one-scale high-dimensional limiting problem admits an infinite sequence of approximations that converge exponentially with respect to the number of parameters used to represent them.

4.1. Low-order finite-element approximation. In this section, we extend the construction of finite-element spaces given in Section 2.1.1 to address the boundary conditions of the high-dimensional problem (3.6) and establish main approximation results. As stated in Assumption 4.1, we consider the case $D = Y_1 = \cdots = Y_n = (0,1)^d$.

4.1.1. Low-order approximation on an interval. For every $L \in \mathbb{N}$, we will use the *analysis operators* introduced in (2.5) to extract the coefficients of finite-element approximations belonging to U^L and $\bar{U}^L$. To construct such approximations, we will use the following projection operators, $\pi^L \colon H^1(0,1) \to U^L$ and $\bar{\pi}^L \colon L^2(0,1) \to \bar{U}^L$. The first we define as the operator of continuous, piecewise-linear Lagrange interpolation at the nodes given in (2.3), in the basis of φ_j^L with $j \in \{0,1,\dots,2^L\}$. The second operator we define as the operator of piecewise-constant L^2 approximation onto the span of $\bar{\varphi}_i^L$ with $j \in \{1,\dots,2^L\}$, which are defined in Section 2.1.1. Note that $(\pi^L u)' = \bar{\pi}^L u'$ for every $u \in H^1(0,1)$. Finally, both the projection operators can be expressed in terms of the analysis operators defined in (2.5): for all $u \in H^1(0,1)$, and $w \in L^2(0,1)$, we have

$$\pi^L u = \sum_{i=0}^{2^L} (\Phi^L u)_i \, \varphi_i^L \quad \text{and} \quad \bar{\pi}^L w = \sum_{j=1}^{2^L} (\bar{\Phi}^L w)_i \, \bar{\varphi}_i^L \,. \quad (4.1)$$

In the following proposition, we summarize classical bounds for the projection operators π^L and $\bar{\pi}^L$ for $L \in \mathbb{N}$.

PROPOSITION 4.2. *For all $v \in C[0,1] \cap C^2(0,1)$, $w \in C[0,1] \cap C^1(0,1)$ and $L \in \mathbb{N}$, the projections $\pi^L v$ and $\bar{\pi}^L w$ satisfy the error bounds*

$$\|v - \pi^L v\|_{L^\infty(0,1)} \le 2^{-2L-3} \|v''\|_{L^\infty(0,1)}, \quad \|(v - \pi^L v)'\|_{L^\infty(0,1)} \le 2^{-L} \|v''\|_{L^\infty(0,1)},$$
$$\|w - \bar{\pi}^L w\|_{L^\infty(0,1)} \le 2^{-L} \|w'\|_{L^\infty(0,1)}$$

and the stability bounds

$$\|\pi^L v\|_{L^\infty(0,1)} \le \|v\|_{L^\infty(0,1)}, \quad \|(\pi^L v)'\|_{L^\infty(0,1)} \le \|v'\|_{L^\infty(0,1)},$$
$$\|\bar{\pi}^L w\|_{L^\infty(0,1)} \le \|w\|_{L^\infty(0,1)}.$$

4.1.2. Low-order approximation on $D \times \boldsymbol{Y}_i$. For every $L \in \mathbb{N}$, from the univariate bases defined above, we obtain by tensorization finite-element spaces

$$\bar{V}^L = \bigotimes_{k=1}^d \bar{U}^L \subset L^2(D), \quad V^L = \bigotimes_{k=1}^d U^L \subset L^2(D), \quad \mathring{V}^L = \bigotimes_{k=1}^d \mathring{U}^L = V^L \cap H_0^1(D) \tag{4.2}$$

of d-variate functions and the corresponding projection operators $\bar{\Pi}^L \colon L^2(D) \to \bar{V}^L$ and $\Pi^L \colon \big(H^1(0,1)\big)^{\otimes d} \to V^L$ given by

$$\bar{\Pi}^L = \bigotimes_{k=1}^d \bar{\pi}^L \quad \text{and} \quad \Pi^L = \bigotimes_{k=1}^d \pi^L. \tag{4.3}$$

Further, for all $i \in \{0, \dots, n\}$ and $L \in \mathbb{N}$, we define

$$\bar{V}_i^L = \big(\bar{V}^L\big)^{\otimes(i+1)} \quad \text{and} \quad V_i^L = \bar{V}_{i-1}^L \otimes V^L \tag{4.4}$$

and the corresponding operators $\bar{\Pi}_i^L \colon L^2(D \times \boldsymbol{Y}_i) \to \bar{V}_i^L$ and $\Pi_i^L \colon L^2(D \times \boldsymbol{Y}_{i-1}) \otimes \big(H^1(0,1)\big)^{\otimes d} \to V_i^L$ given by

$$\bar{\Pi}_i^L = \big(\bar{\Pi}^L\big)^{\otimes(i+1)} \quad \text{and} \quad \Pi_i^L = \bar{\Pi}_{i-1}^L \otimes \Pi^L. \tag{4.5}$$

In particular, we have $\bar{V}_0^L = \bar{V}^L$, $V_0^L = V^L$, $\bar{\Pi}_0^L = \bar{\Pi}^L$ and $\Pi_0^L = \Pi^L$.

The following accuracy bounds for $\bar{\Pi}_i^L$ and Π_i^L with $i \in \{1, \dots, n\}$ and $L \in \mathbb{N}$ can be derived from Proposition 4.2.

LEMMA 4.3. *Let $i \in \{0, \dots, n\}$ and $\|\cdot\|_\infty$ denote $\|\cdot\|_{L^\infty(D \times \boldsymbol{Y}_i)}$. Assume that $v \in C^1(\overline{D \times \boldsymbol{Y}_i})$ and $w \in C^3(\overline{D \times \boldsymbol{Y}_i})$. Then the following error bounds hold for all $L \in \mathbb{N}$ and $k \in \{1, \dots, d\}$:*

$$\|v - \bar{\Pi}_i^L v\|_\infty \le 2^{-L} \sum_{j'=0}^{i} \sum_{k'=1}^{d} \|\partial_{j'k'} w\|_\infty,$$
$$\|w - \Pi_i^L w\|_\infty \le 2^{-L} \sum_{j'=0}^{i-1} \sum_{k'=1}^{d} \|\partial_{j'k'} w\|_\infty + 2^{-2L-3} \sum_{k'=1}^{d} \|\partial_{ik'}^2 w\|_\infty,$$
$$\|\partial_{ik}(w - \Pi_i^L w)\|_\infty \le 2^{-L} \sum_{j'=0}^{i-1} \sum_{k'=1}^{d} \|\partial_{j'k'} \partial_{ik} w\|_\infty + 2^{-L} \sum_{k'=1}^{d} \|\partial_{ik'}^2 \partial_{ik} w\|_\infty.$$

We give a proof of Lemma 4.3 in the Appendix.

For all $i \in \{0, \dots, n\}$ and $L \in \mathbb{N}$, the projections produced by the operators $\bar{\Pi}_i^L$ and Π_i^L, defined by (4.5), can be parametrized by the coefficients

extracted using the analysis operators $\bar{\Psi}_i^L\colon L^2(D\times \boldsymbol{Y}_i) \to \mathbb{C}^{2^{(i+1)dL}}$ and $\Psi_i^L\colon L^2(D\times \boldsymbol{Y}_{i-1})\otimes \big(H^1(0,1)\big)^{\otimes d} \to \mathbb{C}^{2^{(i+1)dL}}$ given by

$$\bar{\Psi}_i^L = \bigotimes_{j=0}^{i} \bar{\Psi}^L \quad \text{and} \quad \Psi_i^L = \bar{\Psi}_{i-1}^L \otimes \Psi^L \tag{4.6}$$

for $i\in\{0,\dots,n\}$, so that, in particular, $\bar{\Psi}_0^L = \bar{\Psi}^L$ and $\Psi_0^L = \Psi^L$.

4.2. High-order approximation.

4.2.1. High-order approximation on an interval. By $\widetilde{T}_\alpha$ with $\alpha\in\mathbb{N}_0$, we denote the Chebyshev polynomials of the first kind orthogonal on $(0,1)$:

$$\widetilde{T}_\alpha(x) = \cos\big\{n\arccos(2x-1)\big\} \quad \text{for all} \quad x\in(0,1) \quad \text{and} \quad \alpha\in\mathbb{N}_0\,, \tag{4.7}$$

so that the orthogonality property holds with respect to the weight function ω given by

$$\omega(x) = 1/\sqrt{x\,(1-x)} \quad \text{for all} \quad x\in(0,1)\,. \tag{4.8}$$

Specifically, we have

$$\langle \widetilde{T}_\alpha, \widetilde{T}_{\alpha'}\rangle_{L^2_\omega(0,1)} = \int_0^1 \omega\,\widetilde{T}_\alpha\,\widetilde{T}_{\alpha'} = \delta_{\alpha\alpha'}\,\|\widetilde{T}_\alpha\|^2_{L^2_\omega(0,1)} \quad \text{for all} \quad \alpha,\alpha'\in\mathbb{N}_0\,, \tag{4.9}$$

where $\|\widetilde{T}_0\|^2_{L^2_\omega(0,1)} = \pi$ and $\|\widetilde{T}_\alpha\|^2_{L^2_\omega(0,1)} = \dfrac{\pi}{2}$ for all $\alpha\in\mathbb{N}$.

Further, we consider the complex exponentials $\widehat{T}_\alpha$ with $\alpha\in\mathbb{Z}$ defined as follows:

$$\widehat{T}_\alpha(x) = \exp(2\pi\mathsf{i}\alpha x) \quad \text{for all} \quad x\in(0,1) \quad \text{and} \quad \alpha\in\mathbb{Z}\,. \tag{4.10}$$

These are also orthogonal on $(0,1)$:

$$\langle \widehat{T}_\alpha, \widehat{T}_{\alpha'}\rangle_{L^2(0,1)} = \int_0^1 \widehat{T}_\alpha^*\,\widehat{T}_{\alpha'} = \delta_{\alpha\alpha'} \quad \text{for all} \quad \alpha,\alpha'\in\mathbb{Z}\,. \tag{4.11}$$

We will use the following notation for the spaces of univariate algebraic and trigonometric polynomials of degree at most $p\in\mathbb{N}_0$:

$$\mathscr{P}_p = \operatorname{span}\{\widetilde{T}_\alpha\}_{\alpha=0}^p \quad \text{and} \quad \mathscr{P}_{\#p} = \operatorname{span}\{\widehat{T}_\alpha\}_{\alpha=-p}^p\,, \tag{4.12}$$

where the span is meant with respect to the field $\mathbb{C}$.

We will use polynomial approximations obtained by the following orthogonal projections onto $\mathcal{P}_p$ and $\mathcal{P}_{\#p}$ with $p\in\mathbb{N}_0$:

$$\pi_p = \frac{1}{\pi}\,\widetilde{T}_0\,\langle\widetilde{T}_0,\cdot\rangle_{L^2_\omega(0,1)} + \frac{2}{\pi}\sum_{\alpha=1}^{p}\widetilde{T}_\alpha\,\langle\widetilde{T}_\alpha,\cdot\rangle_{L^2_\omega(0,1)}\colon L^2_\omega(0,1)\to\mathscr{P}_p\,,$$

$$\pi_{\#p} = \widehat{T}_0\,\langle\widehat{T}_0,\cdot\rangle_{L^2(0,1)} + \sum_{\pm\alpha=1}^{p}\widehat{T}_\alpha\,\langle\widehat{T}_\alpha,\cdot\rangle_{L^2(0,1)}\colon L^2(0,1)\to\mathscr{P}_{\#p}\,.$$

4.2.2. High-order approximation on $\overline{D\times\boldsymbol{Y}_i}$. For every $i\in\{1,\dots,n\}$, denoting by id the identity transformation of $\mathbb{C}^{id}$, let us define the following tensor-product operators:

$$\mathit{\Pi}_{i,p} = \pi_p^{\otimes d}\otimes\pi_{\#p}^{\otimes id}\colon L^2_{\omega^{\otimes d}\otimes\mathsf{id}}(D\times\boldsymbol{Y}_i)\to\mathscr{P}_p^{\otimes d}\otimes\mathscr{P}_{\#p}^{\otimes id}$$

for all $p\in\mathbb{N}_0$. Here, ω denotes the weight function (4.8). The following lemma verifies that, when applied to analytic functions, these operators yield approximations that

converge exponentially with respect to $p \in \mathbb{N}_0$. Proofs for the following two classical results are provided in the Appendix.

LEMMA 4.4. *Assume that $i \in \{1,\dots,n\}$ and $w \in V_i$ is analytic and one-periodic with respect to each of the last id scalar variables on $\overline{D \times \boldsymbol{Y}_i}$. Let $\epsilon_0 > 0$. Then there exist positive constants C and c such that, for any $\epsilon \in (0,\epsilon_0)$ and for $p = \lceil c \log \epsilon^{-1} \rceil$, the following bounds hold for all $k \in \{1,\dots,d\}$ and $j \in \{1,\dots,i\}$:*

$$\begin{gathered}\|w - \Pi_{i,p}\, w\|_{L^\infty(D\times\boldsymbol{Y}_i)} \le C\epsilon\,, \qquad \|\partial_k(w - \Pi_{i,p}\, w)\|_{L^\infty(D\times\boldsymbol{Y}_i)} \le C\epsilon p^2\,, \\ \|\partial_{jk}(w - \Pi_{i,p}\, w)\|_{L^\infty(D\times\boldsymbol{Y}_i)} \le C\,\epsilon p\,.\end{gathered} \tag{4.13}$$

LEMMA 4.5. *Let the assumptions of Lemma 4.4 hold and $\|\cdot\|_\infty$ denote $\|\cdot\|_{L^\infty(D\times\boldsymbol{Y}_i)}$. Then there exist positive constants C and c such that, for any $L \in \mathbb{N}$ and for $p = \lceil cL \rceil$, the following bounds hold for every $k \in \{1,\dots,d\}$:*

$$\|\partial_{ik}(w - \Pi_i^L \Pi_{i,p}\, w)\|_\infty \le C\, p^2\, 2^{-L} \quad \text{and} \quad \|\partial_{ik} w - \bar{\Pi}_i^L \partial_{ik} \Pi_{i,p}\, w\|_\infty \le C\, p^2\, 2^{-L}\,.$$

4.3. Low-rank tensor approximation. In this section, for $i \in \{1,\dots,n\}$ and $L \in \mathbb{N}$, we consider

$$w_i^L = (\Pi_i^L \Pi_{i,p_L} w_{ik})_{k=1}^d \in \big(V_i^L\big)^d \quad \text{and} \quad J_i^L = (\bar{\Pi}_i^L \partial_{ik} \Pi_{i,p_L} w_{ik'})_{k',k=1}^d \in \big(\bar{V}_i^L\big)^{d\times d} \tag{4.14}$$

with a suitable $p_L \in \mathbb{N}$ as approximations to w_i and $\mathcal{J}_i w_i$, where w_i is the solution of (3.16). Then the approximation scheme (3.23) produces $u_i^L \in V_i^L$ and $v_i^L \in (\bar{V}_i^L)^d$.

Section 4.3.1 relates the error of the approximation scheme (3.23), bounded by (3.24)–(3.25), to the error of w_i^L and J_i^L as approximations to w_i and $\mathcal{J}_i w_i$ for all $i \in \{1,\dots,n\}$ and $L \in \mathbb{N}$.

In Section 4.3.2, the error bounds proved in Section 4.3.1 are followed by a quantized tensor-rank analysis, which is based on auxiliary definitions and rank bounds which are also provided in Section 4.3.2.

The analysis is based on the following assumption regarding the low-rank TT-MPS approximability of the solution u_0 of (3.21) and of its gradient $v_0 = \nabla u_0$.

ASSUMPTION 4.6 (the low-rank approximability of the homogenized solution u_0, with $\alpha_0 \in (0,1]$ and with positive constants C_0, ν_0 and c_0). *There exist $u_0^L \in \bar{V}^L$ and $v_0^L \in (\bar{V}^L)^d$ with $L \in \mathbb{N}$ such that*

$$\|(u_0, v_0) - (u_0^L, v_0^L)\|_{L^2(D)^{d+1}} \le C_0\, 2^{-\alpha L} \quad \text{for each} \quad L \in \mathbb{N} \tag{4.15}$$

and the subspace $\mathcal{W}^L = \operatorname{span}\{u_0^L, v_{0,1}^L, \dots, v_{0,d}^L\}$ factorizes at every level $\ell \in \{1,\dots,L-1\}$ with rank not exceeding $c_0 L^{\nu_0 d}$.

The purpose of Sections 4.3.3 and 4.3.5 is to bound the growth with respect to $L \in \mathbb{N}$ of the maximum ranks (2.15) for the coefficients of $u_1^L, \dots, u_n^L$ and $\mathcal{U}^\varepsilon v_n^L$ under Assumption 4.6, which we justify in Section 4.3.6 in the case of $d = 2$ physical dimensions.

4.3.1. Accuracy of the approximation scheme. Under Assumption 4.1, differentiating the equation expressing the cell problem (3.16) in the strong form for $i \in \{1,\dots,n\}$, one verifies that the solutions $w_i \in V_i^d$ with $i \in \{1,\dots,n\}$ are analytic and therefore satisfy the assumption of Lemma 4.4. This gives that, with a positive constant c, for any $L \in \mathbb{N}$ and for

$$p_L = \lceil cL \rceil\,, \tag{4.16}$$

the approximations w_i^L and J_i^L defined by (4.14) satisfy the error bounds

$$\|w_i - w_i^L\|_{W_i} \lesssim L^2\, 2^{-L} \quad \text{and} \quad \|\mathcal{J}_i w_i - J_i^L\|_{L^\infty(D\times\boldsymbol{Y}_{i-1}, L^2(Y_i))^{d\times d}} \lesssim L^2\, 2^{-L}$$

The equivalence holds with a positive constant that is independent of $L \in \mathbb{N}$ and $i \in \{1, \dots, n\}$. Then the bounds (3.24)–(3.25) for the approximation scheme (3.23) show that the resulting approximations $u_i^L \in V_i$ and $v_i^L \in L^2(D \times \boldsymbol{Y}_i)^d$ satisfy the bounds

$$\|u_i - u_i^L\|_{V_i}\,,\ \big\|v_i - v_i^L\big\|_{L^2(D\times \boldsymbol{Y}_i)^d} \lesssim \|v_0 - v_0^L\|_{L^2(D)^d} + L^2\, 2^{-L} \tag{4.17}$$

with a positive equivalence constant independent of $L \in \mathbb{N}$ and $i \in \{1, \dots, n\}$.

4.3.2. Auxiliary subspaces and results. For all $L \in \mathbb{N}$ and $p \in \mathbb{N}_0$, we will use the following notation for the sets of tensors obtained by evaluating d-variate algebraic and trigonometric polynomials of maximum degree at most p on a uniform tensor-product grid with 2^L nodes in each variable:

$$\mathcal{P}_p^{L,d} = \Psi^L\, \mathcal{P}_p^{\otimes d} \subset \mathbb{C}^{2^{dL}} \quad \text{and} \quad \mathcal{P}_{\#p}^{L,d} = \bar{\Psi}^L\, \mathcal{P}_{\#p}^{\otimes d} \subset \mathbb{C}^{2^{dL}}\,. \tag{4.18}$$

Let us extend (4.12) and (4.18) by introducing, for all $p \in \mathbb{N}_0$, $L \in \mathbb{N}$ and $\lambda \in \mathbb{Z}$,

$$\mathcal{P}_{\#p,\lambda} = \operatorname{span}\big\{\widehat{T}_\alpha(2^\lambda\,\cdot\,)\big\}_{\alpha=-p}^{p} \quad \text{and} \quad \mathcal{P}_{\#p,\lambda}^{L,d} = \bar{\Psi}^L\, \mathcal{P}_{\#p,\lambda}^{\otimes d} \subset \mathbb{C}^{2^{dL}}\,. \tag{4.19}$$

We will use several results, stated below, to analyze the low-rank structure of the approximations $u_i^L \in V_i$ and $v_i^L \in L^2(D \times \boldsymbol{Y}_i)^d$ with $i \in \{1, \dots, n\}$ and $L \in \mathbb{N}$, defined by (3.23) and (4.14).

PROPOSITION 4.7. *For all $p \in \mathbb{N}_0$ and $L \in \mathbb{N}$, the subspaces $\bar{\Pi}^L \mathcal{P}_p^{\otimes d} \subset \bar{V}^L$ and $\Pi^L \mathcal{P}_p^{\otimes d} \subset V^L$ factorize with rank $(p+1)^d$ at every level $\ell \in \{1, \dots, L-1\}$ in the sense of Definition 2.2: $\mathcal{P}_p^{L,d} \subset \mathcal{P}_p^{\ell,d} \otimes \mathcal{P}_p^{L-\ell,d}$.*

The proof of Proposition 4.7 follows trivially from the binomial formula applied to the standard basis of monomials.

An immediate consequence of Proposition 4.7 is that the tensor of the values of any d-variate polynomial of maximum degree at most $p \in \mathbb{N}$ at any tensor-product uniform grid with 2^L entries in each dimension can be represented in the QTT decomposition with ranks not exceeding $(p+1)^d$. This was originally shown, in the case of $d = 1$, in [25, Corollary 13]. The language of space factorization, which we adopt here, is different from that of [58, 60, 44, 25]; we use it here to mostly avoid lengthy expressions for explicit basis expansions with numerous indices associated with nodes of tensor-product grids. We refer to Section 2.1.4 for an exposition of the subspace interpretation of the TT-MPS decomposition.

The additional notation (4.19) allows to state the following analog of Proposition 4.7 for trigonometric polynomials, which is an immediate consequence of the separability of the exponential function.

PROPOSITION 4.8. *For all $p \in \mathbb{N}_0$ and $L \in \mathbb{N}$, the subspaces $\bar{\Pi}^L \mathcal{P}_{\#p}^{\otimes d} \subset \bar{V}^L$ and $\Pi^L \mathcal{P}_{\#p}^{\otimes d} \subset V^L$ factorize with rank $(2p+1)^d$ at every level $\ell \in \{1, \dots, L-1\}$ in the sense of Definition 2.2: $\mathcal{P}_{\#p}^{L,d} \subset \mathcal{P}_{\#p}^{\ell,d} \otimes \mathcal{P}_{\#p,-\ell}^{L-\ell,d}$.*

4.3.3. Approximation of the one-scale high-dimensional limit problem. The approximations given by (3.23) and (4.14) realize arbitrary accuracy and belong to tensor products of spaces of algebraic and trigonometric polynomials:

$$w_i^L \in \bar{\Pi}^L \mathcal{P}_{p_L}^{\otimes d} \otimes \bigotimes_{j=1}^{i-1} \bar{\Pi}^L \mathcal{P}_{\#p_L}^{\otimes d} \otimes \Pi^L \mathcal{P}_{\#p_L}^{\otimes d} \tag{4.20}$$

for all $L \in \mathbb{N}$ and $i \in \{1, \dots, n\}$. For every $L \in \mathbb{N}$, we consider the subspaces

$$\mathcal{W}^L = \mathcal{W}^L = \operatorname{span}\{u_0^L, v_{0,1}^L, \dots, v_{0,d}^L\} \subset \bar{V}^L \quad \text{and} \quad \mathcal{S}_0^L = \mathcal{W}^L \odot \bar{\Pi}^L \mathcal{P}_{p_L}^{\otimes d} \subset \bar{V}^L\,. \tag{4.21}$$

Here and throughout, the sign "$\odot$" between two spaces denotes the operation of constructing the subspace that consists of all pointwise products of all pairs of elements

from the respective spaces. Assuming that $\mathscr{W}^L$ factorizes with rank $r \in \mathbb{N}$ at every level $\ell \in \{0,\dots,L\}$ in the sense of Definition 2.2, we obtain from (4.21) and Proposition 4.7 that $\mathcal{S}_0^L$ factorizes with rank $r_0 = r(p_L+1)^d$ at every level $\ell \in \{0,\dots,L\}$. In particular, with $\ell = L$, we obtain $\dim \mathcal{S}_0^L \le r_0$.

For all $L \in \mathbb{N}$ and $i \in \{1,\dots,n\}$, relations (4.20) and (3.23) with (4.14) result in

$$v_i^L \in (\bar{\mathbb{Q}}_i^L)^d \quad \text{with} \quad \bar{\mathbb{Q}}_i^L = \mathcal{S}_0^L \otimes \bigotimes_{j=1}^{i} \bar{\Pi}^L \mathcal{P}^{\otimes d}_{\#(i+1-j)p_L} \subset \bar{V}_i^L \tag{4.22}$$

and

$$u_i^L \in \mathbb{Q}_i^L = \mathcal{S}_0^L \otimes \bigotimes_{j=1}^{i} \bar{\Pi}^L \mathcal{P}^{\otimes d}_{\#(i+1-j)p_L} \otimes \Pi^L \mathcal{P}^{\otimes d}_{\#p_L} \subset V_i^L\,. \tag{4.23}$$

THEOREM 4.9. *Let Assumptions 2.1 and 4.1 hold and $(u_0,u_1,\dots,u_n) \in \boldsymbol{V}_n$ be the solution of (3.6). Consider $v_0^L \in (\bar{V}_i^L)^d$ with $L \in \mathbb{N}$ satisfying Assumption 4.6 with $\alpha_0 \in (0,1]$ and positive constants C_0, c_0, ν_0. Then the approximations $u_i^L \in V_i^L$ with $L \in \mathbb{N}$ and $i \in \{1,\dots,n\}$ given by (3.23) and (4.14) satisfy the following with $\tilde{\alpha} \in (0,\alpha_0)$ (which may be chosen arbitrarily close to α_0) and with positive constants $\tilde{C}$ and $\tilde{c}$ independent of L.*

For all $L \in \mathbb{N}$ and $i \in \{1,\dots,n\}$, the bound $\|u_i - u_i^L\|_V \le \tilde{C}\,2^{-\tilde{\alpha}L}$ holds and the coefficient tensor $\Psi_i^L u_i^L$ has a decomposition of the form (2.13) with $(i+1)L$ levels and with ranks bounded from above by $\tilde{c}L^{(n+1+\nu_0)d}$.

Proof. Consider $i \in \{1,\dots,n\}$ and $L \in \mathbb{N}$. The claimed accuracy bound follows from (4.17).

To bound the first L ranks, we note that the subspace $\mathscr{W}^L$ defined in (4.21) factorizes at every level $\ell \in \{1,\dots,L\}$ with rank not exceeding $c_0 L^{\nu_0 d}$ by Assumption 4.6. Then, by the argument following (4.21), $\mathcal{S}_0^L$ factorizes at every level $\ell \in \{0,\dots,L\}$ with rank not exceeding $c_0 L^{\nu_0 d}(p_L+1)^d$.

To bound the other ranks, we now consider $j \in \{1,\dots,i\}$ and $\ell \in \{1,\dots,L\}$ and factorize $\mathbb{Q}_i^L$ so that the first factor is a subspace of $\mathbb{C}^{2^{jdL+d\ell}}$. This subspace corresponds to the finest ℓ levels of the jth microscale and all levels of all coarser scales. The dimension of this subspace majorates the corresponding rank of $\Psi_i^L u_i^L \in \Psi_i^L(\mathbb{Q}_i^L)$.

Applying Proposition 4.8, we obtain $\mathcal{P}^{L,d}_{\#(i+1-j)p_L} \subset \mathcal{P}^{\ell,d}_{\#(i+1-j)p_L} \otimes \mathcal{P}^{L-\ell,d}_{\#(i+1-j)p_L,-\ell}$, where the dimension of both the factors is $(2(i+1-j)p_L+1)^d$. Then we have $\Psi_i^L(\mathbb{Q}_i^L) \subset \mathscr{L} \otimes \mathscr{M}$ with $\mathscr{L} = \Psi_i^L(\mathcal{S}_0^L) \otimes \bigotimes_{m=1}^{j-1} \mathcal{P}^{L,d}_{\#(i+1-m)p_L} \otimes \mathcal{P}^{\ell,d}_{\#(i+1-j)p_L}$ and $\mathscr{M} = \mathcal{P}^{L-\ell,d}_{\#(i+1-j)p_L,-\ell} \otimes \bigotimes_{m=j+1}^{i} \mathcal{P}^{L,d}_{\#(i+1-m)p_L}$. Using (4.16), we bound the dimension of the first factor as follows:

$$\begin{aligned}\dim \mathscr{L} &\le (\dim \mathcal{S}_0^L)\dim \mathcal{P}^{\ell,d}_{\#(i+1-j)p_L} \prod_{m=1}^{j-1} \dim \mathcal{P}^{L,d}_{\#(i+1-m)p_L} \\ &\le c_0 L^{\nu_0 d}(p_L+1)^d \prod_{m=1}^{j} (2(i+1-m)p_L+1)^d \le \tilde{c}L^{(n+1+\nu_0)d}\end{aligned}$$

with a suitable positive constant $\tilde{c}$ that depends on on n, d, as well as on c and c_0,ν_0 appearing in (4.16) and in Assumption 4.6 but not on L. ∎

4.3.4. Approximation of the homogenized gradient. In this section, we analyze the low-rank structure and accuracy of $\mathcal{U}^\varepsilon v_n^\Lambda$ with $\Lambda \in \mathbb{N}$ as an approximation to $\mathcal{U}^\varepsilon v_n$. Index Λ is used because, as we show below, $\mathcal{U}^\varepsilon v_n^\Lambda \in \bar{V}^{\lambda_n+\Lambda}$, and we choose to reserve L for the (total) number of levels of the resulting approximation: $L = \lambda_n + \Lambda$. As in Section 4.3.3, we develop our analysis here under Assumptions 2.1, 4.1 and 4.6.

We start with the matrix representation of the averaging operators, defined in Section 3.2. To this end, for all $\Lambda \in \mathbb{N}$ and $i \in \{1, \dots, n\}$, we define the matrices

$$\begin{aligned}
\boldsymbol{E}_i^\Lambda &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{\otimes d(\lambda_{i-1}+\Lambda)} \otimes \begin{pmatrix} 1 \\ 1 \end{pmatrix}^{\otimes d(\lambda_i - \lambda_{i-1})} && \in \mathbb{R}^{d(\lambda_i+\Lambda)\times d(\lambda_{i-1}+\Lambda)}, \\
\boldsymbol{M}_i^\Lambda &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{\otimes d\lambda_i} \otimes 2^{-d\Lambda} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}^{\otimes d\Lambda} && \in \mathbb{R}^{d(\lambda_i+\Lambda)\times d(\lambda_i+\Lambda)}, \\
\boldsymbol{F}_i^\Lambda &= \begin{pmatrix} 1 \\ 1 \end{pmatrix}^{\otimes d\lambda_i} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{\otimes d\Lambda} && \in \mathbb{R}^{d(\lambda_i+\Lambda)\times d\Lambda}.
\end{aligned} \tag{4.24}$$

These matrices represent the averaging operator $\mathcal{U}^{\varepsilon_i}$ corresponding to microscale ε_i as follows.

LEMMA 4.10. *Let $\Lambda \in \mathbb{N}$ and $i \in \{1, \dots, n\}$. Consider $v \in \bar{V}^{\lambda_{i-1}+\Lambda}$ and $w \in \bar{V}^\Lambda$. Then $\mathcal{U}^{\varepsilon_i}(v \otimes w) \in \bar{V}^{\lambda_i+\Lambda}$ and*

$$(\bar{\Psi}^{\lambda_i+\Lambda} \circ \mathcal{U}^{\varepsilon_i})(v \otimes w) = (\boldsymbol{M}_i^\Lambda \boldsymbol{E}_i^\Lambda \bar{\Psi}^{\lambda_{i-1}+\Lambda} v) \odot (\boldsymbol{F}_i^\Lambda \bar{\Psi}^\Lambda w). \tag{4.25}$$

Proof. Applying Definition 3.4, we obtain $\mathcal{U}^{\varepsilon_i}(v \otimes w) = \bar{v} \cdot \tilde{w} \in \bar{V}^{\lambda_i+\Lambda} \subset L^2(D)$, where $\bar{v}, \tilde{w} \in \bar{V}^{\lambda_i+\Lambda}$ are given by

$$\bar{v}(x) = \int_{Y_i} v\Big(\varepsilon_i \Big[\frac{x}{\varepsilon_i}\Big] + \varepsilon_i z\Big)\,\mathrm{d}z \quad \text{and} \quad \tilde{w}(x) = w\Big(\Big\{\frac{x}{\varepsilon_i}\Big\}\Big) \quad \text{for all } x \in D.$$

Then we trivially have

$$(\bar{\Psi}^{\lambda_i+\Lambda} \circ \mathcal{U}^{\varepsilon_i})(v \otimes w) = (\bar{\Psi}^{\lambda_i+\Lambda} \bar{v}) \odot (\bar{\Psi}^{\lambda_i+\Lambda} \tilde{w}). \tag{4.26}$$

Let us set $\boldsymbol{v}^\Lambda = \bar{\Psi}^{\lambda_{i-1}+\Lambda} v$ and $\boldsymbol{w}^\Lambda = \bar{\Psi}^\Lambda w$.

First, we observe that we have $v \in \bar{V}^{\lambda_i+\Lambda}$ and $\bar{\Psi}^{\lambda_i+\Lambda} v = \boldsymbol{E}_i^\Lambda \boldsymbol{v}^\Lambda$. This is due to that $\boldsymbol{E}_i^\Lambda$ is the matrix of the embedding $\bar{V}^{\lambda_{i-1}+\Lambda} \to \bar{V}^{\lambda_i+\Lambda}$ (with respect to the tensor-product bases obtained from the univariate ones defined in Section 2.1.1). Note that the order of the Kronecker factors in (4.24) agrees with the use of transposition (2.9) in (2.12): the factors can be split into contiguous groups of d factors corresponding in a one-to-one fashion to the levels of the uniform hierarchical partitioning of D.

Further, the multiplication of $\boldsymbol{E}_i^\Lambda \boldsymbol{v}^\Lambda$ by $\boldsymbol{M}_i^\Lambda$ amounts to averaging $2^{d\Lambda}$ entries within each of the $2^{d\lambda_i}$ contiguous blocks of the vector $\boldsymbol{E}_i^\Lambda \boldsymbol{v}^\Lambda$. Again, the order of the Kronecker factors in (4.24) agrees with the use of transposition (2.9) in (2.12): in terms of transforming v, the above averaging of the coefficient vector is nothing else than averaging v over the translations of $\varepsilon_i Y_i$ that partition D. This gives $\bar{\Psi}^{\lambda_i+\Lambda} \bar{v} = \boldsymbol{M}_i^\Lambda \boldsymbol{E}_i^\Lambda \boldsymbol{v}^\Lambda$.

Finally, the multiplication of $\boldsymbol{w}^\Lambda$ by $\boldsymbol{F}_i^\Lambda$ amounts to translating the coefficient periodically $2^{d\lambda_i}$ times, so that $\bar{\Psi}^{\lambda_i+\Lambda} \tilde{w} = \boldsymbol{F}_i^\Lambda \boldsymbol{w}^\Lambda$.

Combining the above observations with (4.26), we obtain $(\bar{\Psi}^{\lambda_i+\Lambda} \circ \mathcal{U}^{\varepsilon_i})(v \otimes w) = (\boldsymbol{M}_i^\Lambda \boldsymbol{E}_i^\Lambda \boldsymbol{v}^\Lambda) \odot (\boldsymbol{F}_i^\Lambda \boldsymbol{w}^\Lambda)$. ∎

COROLLARY 4.11. *Let $\Lambda \in \mathbb{N}$, $i \in \{1, \dots, n\}$ and $\ell \in \{0, \dots, \lambda_i + \Lambda\}$. The following statements on subspace factorization at level ℓ in the sense of Definition 2.2 hold.*

- **a)** *For $\ell < \min\{\lambda_{i-1} + \Lambda, \lambda_i\}$, consider a subspace $S \subset \bar{V}^{\lambda_{i-1}+\Lambda}$ factorizing at level ℓ with rank $r \in \mathbb{N}$. Then the subspace $\mathcal{U}^{\varepsilon_i}(S \otimes \bar{V}^\Lambda)$ factorizes at level ℓ with rank r.*
- **b)** *For $\min\{\lambda_{i-1} + \Lambda, \lambda_i\} \le \ell \le \lambda_i$, consider a subspace $S \subset \bar{V}^{\lambda_{i-1}+\Lambda}$. Then the subspace $\mathcal{U}^{\varepsilon_i}(S \otimes \bar{V}^\Lambda)$ factorizes at level ℓ with rank $\dim S$.*
- **c)** *For $\ell > \lambda_i$, consider a subspace $T \subset \bar{V}^\Lambda$ factorizing at level with rank $r \in \mathbb{N}$ $\ell - \lambda_i$ and a subspace $S \subset \bar{V}^{\lambda_{i-1}+\Lambda}$. Then the subspace $\mathcal{U}^{\varepsilon_i}(S \otimes T)$ factorizes at level ℓ with rank $r \cdot \dim S$.*

Proof. In each part and for each ℓ, the corresponding subspace factorization follows from equation (4.25) of Lemma 4.10. The claim regarding the dimension of the resulting subspace follows immediately from the Kronecker-product form of the matrices defined in (4.24). ∎

Let us consider the subspace $\mathcal{S}_0^\Lambda$ defined in (4.21). For $i = 1, \dots, n$, we iteratively define the subspace

$$\mathcal{S}_i^\Lambda = \mathcal{U}^{\varepsilon_i}\big(\mathcal{S}_{i-1}^\Lambda \otimes \bar{\Pi}^\Lambda \mathcal{P}^{\otimes d}_{\#(n+1-i)p_\Lambda}\big) \subset \bar{V}^{\lambda_i+\Lambda}, \tag{4.27}$$

where $\mathcal{U}^{\varepsilon_i}$ is given by Definition 3.4. From (4.22) with $i = n$, recalling Definition 3.6, we derive

$$\begin{aligned}\mathcal{U}^\varepsilon v_n^\Lambda &= \mathcal{U}_n^\varepsilon \circ \cdots \circ \mathcal{U}_1^\varepsilon\,(v_n^\Lambda)\\ &= \mathcal{U}^{\varepsilon_n} \circ (\mathcal{U}^{\varepsilon_{n-1}} \otimes \mathsf{id}_{Y_n}) \circ \cdots \circ (\mathcal{U}^{\varepsilon_1} \otimes \mathsf{id}_{Y_2\times\cdots\times Y_n})\,(v_n^\Lambda) \in \mathcal{S}_n^\Lambda\end{aligned}$$

We will now assume that the subspace $\mathscr{W}^\Lambda$ factorizes with rank $r \in \mathbb{N}$ at every level $\ell \in \{0, \dots, \Lambda\}$ in the sense of Definition 2.2 and establish, for all $i \in \{0, \dots, n\}$ and $\ell \in \{1, \dots, \lambda_i + \Lambda - 1\}$, the factorization of $\mathcal{S}_i^\Lambda$ at level ℓ with rank

$$R_{i\ell} = \begin{cases} r_0 & \text{if } \lambda_0 < \ell \le \lambda_1,\\ \vdots & \vdots\\ r_{i-1} & \text{if } \lambda_{i-1} < \ell \le \lambda_i,\\ r_i & \text{if } \lambda_i < \ell \le \lambda_i + \Lambda - 1,\end{cases} \tag{4.28}$$

where $r_0 = r(p_\Lambda + 1)^d$ and $r_i = r_{i-1}(2(n+1-i)p_\Lambda + 1)^d$ for each $i \in \{1, \dots, n\}$.

First, as we noted in the discussion following (4.21), $\mathcal{S}_0^\Lambda$ factorizes at every level $\ell \in \{0, \dots, \lambda_0 + \Lambda\}$ with rank $R_{0,\ell} = r_0$, and $\dim \mathcal{S}_0^\Lambda \le r_0$.

Next, consider $i \in \{1, \dots, n\}$ and assume that the subspace $\mathcal{S}_{i-1}^\Lambda$ factorizes at every level $\ell \in \{0, \dots, \lambda_{i-1} + \Lambda\}$ with rank $R_{i-1,\ell}$. Then, by Corollary 4.11 and Proposition 4.8, the subspace $\mathcal{S}_i^\Lambda$ factorizes at every level $\ell \in \{0, \dots, \lambda_i + \Lambda\}$ with rank $R_{i\ell}$. By induction, the subspace $\mathcal{S}_n^\Lambda$ factorizes at every level $\ell \in \{0, \dots, \lambda_n + \Lambda\}$, with rank $R_{n\ell}$. This implies (see Section 2.1.4) that $\bar{\Psi}^\Lambda \mathcal{U}_n^\varepsilon v_n^\Lambda \in \mathbb{C}^{2^{d(\lambda_n+\Lambda)}}$ has a representation of the form (2.13) with ranks $R_{n\ell}$, $\ell = 1, \dots, \lambda_n + \Lambda - 1$. The maximum rank $r_{\max}$ (2.15) and the number N of effective parameters (2.14) of this representation can be bounded as follows:

$$r_{\max} \le r_n \le r\,(p_\Lambda + 1)^d\,(2(n+1)p_\Lambda + 1)^{nd} \le \tilde{c}\,\Lambda^\theta \tag{4.29}$$

with $\theta = (n + 1 + \nu_0)d$ and

$$N \le 2^d(\lambda_1 - 1)r_0^2 + \sum_{i=2}^{n} 2^d(\lambda_i - \lambda_{i-1})r_{i-1}^2 + 2^d(\Lambda+1)r_n^2 \le \tilde{c}^2\,(\lambda_n + \Lambda)\,\Lambda^{2\theta} \le \tilde{c}^2\,L^\kappa\,. \tag{4.30}$$

with $\kappa = 2\theta = 2(n + 1 + \nu_0)d$. Here, as above, $L = \lambda_1 + \Lambda$ is the (total) number of discretization levels. In (4.29) and (4.30), $r_0, \dots, r_n$ are bounded for the choice (4.16) of p_Λ and under Assumption 4.6. The constant $\tilde{c}$ therefore depends on n, d, as well as on c and c_0, ν_0 appearing in (4.16) and in Assumption 4.6 but not on Λ or $\varepsilon_1, \dots, \varepsilon_n$. However crude the final bounds of (4.29) and (4.30) are, they show that $r_{\max}$ and N grow at most algebraically with respect to Λ and also, more crudely, with respect to $L = \lambda_1 + \Lambda$.

Summarizing the above discussion and combining the error bound of Assumption 4.6 with (4.16), (4.17), and (3.12), we obtain the following result.

THEOREM 4.12. *Let Assumptions 2.1, 4.1 and 4.6 hold. Consider the solution* $(u_0, u_1, \dots, u_n) \in \boldsymbol{V}_n$ *of* (3.6). *Then the approximations* $\mathcal{U}^\varepsilon v_n^\Lambda \in (\bar{V}^{\lambda_n+\Lambda})^d$ *with*

$\Lambda \in \mathbb{N}$, defined by (3.23) and (4.14), where v_n is defined by (3.7), satisfy the following with a constant $\tilde{\alpha} \in (0, \alpha_0)$ (which may be chosen arbitrarily close to α_0) and with positive constants $\tilde{C}$ and $\tilde{c}$ independent of $\Lambda \in \mathbb{N}$ and $\varepsilon > 0$.

For all $\Lambda \in \mathbb{N}$, the error bound $\|\mathcal{U}^\varepsilon v_n - \mathcal{U}^\varepsilon v_n^\Lambda\|_{L^2(D)^d} \leq \tilde{C}\, 2^{-\tilde{\alpha}\Lambda}$ holds and the coefficient tensor $\bar{\Psi}^{\lambda_n+\Lambda}\mathcal{U}^\varepsilon v_n^\Lambda$ has a decomposition of the form (2.13) with ranks bounded from above by $\tilde{c}\Lambda^\theta$, where $\theta = (n+1+\nu_0)d$.

Note that Theorem 4.12 only bounds the error of approximating the "homogenized" gradient $\mathcal{U}^\varepsilon v_n$ with a low-rank function $\mathcal{U}^\varepsilon v_n^\Lambda$. This is also why it involves the scale parameter only via the averaging and evaluation operators. Whenever the convergence of the homogenization error $\|\mathcal{U}^\varepsilon v_n - \nabla u^\varepsilon\|_{L^2(D)^d}$ as $\varepsilon \to 0$ can be quantified, Theorem 4.12 leads to a bound for $\|\mathcal{U}^\varepsilon v_n^\Lambda - \nabla u^\varepsilon\|_{L^2(D)^d}$, as we elaborate in the next section.

4.3.5. Approximation of the multiscale problem with a single microscale. To the best of our knowledge, results on the rate of convergence of the homogenization error are available only in the case of a single ($n = 1$) microscale $\varepsilon_1 = 2^{-\lambda_1}$ with $\lambda_1 \in \mathbb{N}$ and under additional smoothness assumptions on u_0 (see Proposition 3.5), so we choose this case to illustrate the results of Section 4.3.4.

Assuming that $\|\mathcal{U}^{\varepsilon_1} v_1 - \nabla u^\varepsilon\|_{L^2(D)^{d+1}} \leq C\varepsilon_1^s$ for any $\varepsilon_1 > 0$ with $s \in (0, 1]$ and with a positive constant C, both independent of ε_1, and using the notations from the statement of Theorem 4.12, we obtain

$$\begin{aligned}\|\mathcal{U}^\varepsilon v_1^\Lambda - \nabla u^\varepsilon\|_{L^2(D)^d} &\leq \|\mathcal{U}^\varepsilon v_1^\Lambda - \mathcal{U}^\varepsilon v_1\|_{L^2(D)^d} + \|\mathcal{U}^\varepsilon v_1 - \nabla u^\varepsilon\|_{L^2(D)^d} \\ &\leq \tilde{C}\, 2^{-\tilde{\alpha}\Lambda} + C\varepsilon_1^s = \tilde{C}\, 2^{-\tilde{\alpha}\Lambda} + C\, 2^{-s\lambda_1} \leq (\tilde{C} + C)\, 2^{-\tilde{\alpha}\Lambda}\,. \quad (4.31)\end{aligned}$$

The last inequality is obtained under the assumption

$$\Lambda \leq \frac{s}{\tilde{\alpha}}\lambda_1\,, \quad (4.32)$$

which allows to absorb the term bounding the error of homogenization.

First of all, we note that, *for any fixed $\varepsilon_1 > 0$*, our analysis yields *no* asymptotic error bounds for the low-rank approximation of ∇u^ε and does *not* estimate the rate at which the error converges to zero with respect to the number N of effective parameters. This is an artifact of our method of analysis. The error of homogenization which remains fixed once ε_1 has been fixed, precludes the derivation of an ε_1-uniform bound on the maximum rank in terms of accuracy.

Crucial for the meaningful interpretation of the bounds (4.29) and (4.31) is that the constants involved are all independent of $\Lambda \in \mathbb{N}$ and $\lambda_1 = \log_2 \varepsilon_1^{-1} \in \mathbb{N}$. As a result, the bounds hold for the whole sequence of low-rank finite-element approximations indexed by $\Lambda \in \mathbb{N}$ and $\lambda_1 \in \mathbb{N}$ such that (4.32). Any desired accuracy ϵ of approximation can be realized by finite-element functions with the corresponding number $\Lambda \simeq \log_2 \epsilon^{-1}$ of discretization levels added to the first $\lambda_1 = \log_2 \varepsilon_1^{-1}$ levels, for all multiscale problems with $\lambda_1 \in \mathbb{N}$ sufficiently large to satisfy (4.32). At the same time, the ranks of these approximations are bounded from above (4.29) uniformly with respect to $\varepsilon_1 = 2^{-\lambda_1}$ and grow no faster than algebraically with respect to Λ, and ϵ therefore decays root-exponentially with respect to the maximum rank $r_{\max}$ (namely, exponentially with respect to $\sqrt[\theta]{r_{\max}}$).

4.3.6. Physical domains in two dimensions: approximation of functions with corner singularities. In this section, we consider the case of $d = 2$ physical dimensions. By invoking a result from [35, 40], we show that Assumption 4.6 on the homogenized solution u_0 and its gradient $v) = \nabla u_0$ holds in that case. Regarding the case of $d = 3$ dimensions, we mention that an analogous analysis is currently available for point singularities [52]. That analysis, however, does not cover all types of singularities that u_0 and v_0 may exhibit when $d = 3$, which is why we restrict our exposition to the case of $d = 2$.

In order to quantify analytic regularity of the solution u_0 of the homogenized equation in a physical domain $\Omega \subset \mathbb{R}^2$, we use spaces of functions defined on Ω that are analytic on the closure of Ω except for a finite number of corner points where algebraic singularities of a certain order may occur. The analytic regularity of the homogenized solution u_0, expressed in terms of corner-weighted Sobolev spaces in Ω allows us to leverage exponential rank bounds of QTT-FEM from [35, 40]. To define the corner-weighted Sobolev spaces, with any set Θ of a finite number of distinct points in $\mathbb{R}^2$, we associate the weight function χ_Θ given by

$$\chi_\Theta(x) = \prod_{\vartheta\in\Theta} \|x-\vartheta\|_2 \quad \text{for all} \quad x \in \mathbb{R}^2\,, \tag{4.33}$$

where $\|\cdot\|_2$ denotes the Euclidean norm on $\mathbb{R}^2$.

To quantify the analytic regularity of solutions to the high-dimensional one-scale problem, we use weighted Sobolev spaces and associated countably normed classes as introduced in [48, 47, 8, 7, 9] and denoted here by $H^{m,\ell}_{\Theta,\beta}(\Omega)$ and $C^\ell_{\Theta,\beta}(\Omega)$ with $\ell \in \{1,2\}$, $m \in \{0,1,\dots,\ell\}$ and $\beta \in [0,1)$, where $\Omega \subset \mathbb{R}^2$ is a polygonal domain and Θ is a set of $S \in \mathbb{N}$ distinct points in $\overline{\Omega}$.

To express analytic regularity in a form which is suitable to establish exponential QTT-FE approximation rates, we use corner-weighted Sobolev spaces:

$$H^{m,0}_{\Theta,\beta}(\Omega) = \big\{u\colon \Omega\to\mathbb{R}:\ \chi_\Theta^{\beta+|\alpha|}\,\partial^\alpha u \in L^2(\Omega) \quad \text{if} \quad 0\le|\alpha|\le m\big\}$$

for all $\ell \ge 0$ and

$$H^{m,\ell}_{\Theta,\beta}(\Omega) = \big\{u\in H^{\ell-1}(\Omega):\ \chi_\Theta^{\beta+|\alpha|-\ell}\,\partial^\alpha u \in L^2(\Omega) \quad \text{if} \quad 0\le|\alpha|\le m\big\}$$

for all $m \ge \ell \ge 1$, where the differentiation is understood in the weak sense. By setting

$$|u|^2_{H^{m,\ell}_{\Theta,\beta}(\Omega)} = \sum_{|\alpha|=m} \|\chi_\Theta^{\beta+m-\ell}\,\partial^\alpha u\|^2_{L^2(\Omega)} \quad \text{for all} \quad u \in H^{m,\ell}_{\Theta,\beta}(\Omega), \tag{4.34}$$

we introduce $|\cdot|_{H^{m,\ell}_{\Theta,\beta}(\Omega)}$, a seminorm on $H^{m,\ell}_{\Theta,\beta}(\Omega)$. Also, by setting

$$\|u\|^2_{H^{m,0}_{\Theta,\beta}(\Omega)} = \sum_{k=0}^{m} |u|^2_{H^{m,0}_{\Theta,\beta}(\Omega)} \quad \text{for} \quad u\in H^{m,0}_{\Theta,\beta}(\Omega),\quad m\ge 0,$$

$$\|u\|^2_{H^{m,\ell}_{\Theta,\beta}(\Omega)} = \|u\|^2_{H^{\ell-1}(\Omega)} + \sum_{k=\ell}^{m} |u|^2_{H^{k,\ell}_{\Theta,\beta}(\Omega)} \quad \text{for} \quad u\in H^{m,\ell}_{\Theta,\beta}(\Omega),\quad m\ge\ell\ge 1,$$

we define $\|\cdot\|^2_{H^{m,\ell}_{\Theta,\beta}(\Omega)}$, a norm on $H^{m,\ell}_{\Theta,\beta}(\Omega)$ for any $\ell, m \in \mathbb{N}_0$ such that $m \ge \ell$.

Definition 4.13 (analyticity of a function with point algebraic singularities, with positive constants M and ρ). *Let $\beta \in [0,1)$, $\Omega \subset \mathbb{R}^2$ be a polygonal domain, Θ be a finite set of distinct points in $\overline{\Omega}$ and $\ell \in \{1,2\}$. Then $u \in C^\ell_{\Theta,\beta}(\Omega)$ if $u \in H^{\ell,\ell}_{\Theta,\beta}(\Omega)$ and there exist positive constants M and ρ such that, for all $\alpha \in \mathbb{N}_0^2$ with $|\alpha| \ge \ell-1$,*

$$\sup_{x\in\Omega} \chi_\Theta^{\beta+|\alpha|-\ell+1}(x)\,\big|\partial^\alpha u(x)\big| \le M\rho^{|\alpha|}|\alpha|!\,.$$

The following result is a consequence of [7, Theorems 3.4–3.5] for the iterated-homogenization scheme of (3.16)–(3.19).

Proposition 4.14. *Assume that Θ is the set of vertices of the unit square $D = (0,1)^2$. Let Assumptions 2.1 and 4.1 hold. Then the solution u_0 of the homogenized problem (3.21) satisfies $u_0 \in C^2_{\Theta,\beta}(D)$ with some $\beta \in [0,1)$.*

Note that this statement remains valid for an arbitrary curvilinear polygon with piecewise-analytic boundary [7]. However, even in the setting of Assumption 4.1, the exponent β depends on the transformation diagonalizing the diffusion coefficient at the vertices of D and can be estimated in terms of the spectral bounds γ and Γ.

We will now combine the weighted-analyticity statement of Proposition 4.14 with rank bounds for the QTT-FE approximation of functions from $C^2_{\Theta,\beta}(D)$ in [35, 40].

THEOREM 4.15. *Assume that $\beta \in [0,1)$ and Θ is the set of the vertices of D. Let $u_0 \in V \cap C^2_{\Theta,\beta}(D)$, $\nu_0 = 1$ and $\alpha_0 \in (0, 1-\beta)$. Then there exist positive constants C_0, c_0 and functions $u_0^L \in \bar{V}^L$ and $v_0^L \in (\bar{V}^L)^d$ with $L \in \mathbb{N}$ satisfying Assumption 4.6 (with constants $\alpha_0, C_0, \nu_0, c_0$).*

Proof. The claim follows from either of [35, Theorem 5.3.7] and [40, Theorem 5.16]. Indeed, there exist positive constants C, c_0 and $\hat{u}_0^L \in \mathring{V}^L$ with $L \in \mathbb{N}$ such that $\|u_0 - \hat{u}_0^L\|_{H^1(D)} \leq C L^3\, 2^{-(1-\beta)L}$ and the ranks of $\hat{u}_0^L$ are bounded from above by $r_L = \lceil c_0 L^2 \rceil$ for each $L \in \mathbb{N}$.

For each $L \in \mathbb{N}$, we consider $u_0^L \in \bar{V}^L$ and $v_0^L \in (\bar{V}^L)^d$ defined by $u_0^L = (\bar{\pi}^L)^{\otimes d}\, u_0$ and $v_{0,k}^L = \big((\bar{\pi}^L)^{\otimes(k-1)} \otimes \mathsf{id} \otimes (\bar{\pi}^L)^{\otimes(d-k)}\big)\, \partial_k u_0 \in \bar{V}^L$ for all $k \in \{1, \dots, d\}$. Bounds analogous to those of Proposition 4.2 yield $\|u_0^L - \hat{u}_0^L\|_{L^2(D)} \lesssim 2^{-L} |\hat{u}_0^L|_{H^1(D)}$ and $\|v_0^L - \nabla \hat{u}_0^L\|_{L^2(D)^d} \lesssim 2^{-L} |\hat{u}_0^L|_{H^1(D)}$ with equivalence constants independent of $L \in \mathbb{N}$. Then the triangle inequality gives the error bound of Assumption 4.6 with $\nu_0 = 1$ and a suitable positive constant C_0. Further, the action of the operators $\bar{\pi}^L$ and ∂_k, up to scaling, consists in adding to and subtracting from the coefficient tensor its single-position shift along the respective dimension; see, e.g, [10, equations (12a)–(12e)]. This operation preserves the piecewise-polynomial structure used to establish rank bounds in [35, Lemma 4.6.1 and Corollary 4.6.2] and in [40, Lemma 5.13 and Corollary 5.14]. Inspecting those proofs, one concludes that the rank analysis given there applies verbatim to u_0^L and v_0^L with each $L \in \mathbb{N}$: for every $\ell \in \{0, 1, \dots, L\}$, we have that the subspace $\operatorname{span}\{u_0^L, v_{0,1}^L, \dots, v_{0,d}^L\} \subset \bar{V}^L$ factorizes at level ℓ with rank at most r_L, so that these approximations satisfy Assumption 4.6 with $\nu_0 = 2$, the constants $\alpha_0 \in (0, 1-\beta)$ and c_0 appearing above and with a suitable positive constant C_0. ∎

5. Numerical experiments. In the case of $d = 2$ dimensions, we give a parallel numerical study of two series of multiscale problems, with analytic and with discontinuous diffusion coefficients. For the reader's convenience, we juxtapose in pairs all plots corresponding to these two series, placing them, within each pair, on the **left** and on the **right** respectively. As above, we work with the physical domain $D = (0,1)^2$, n microscales and the unit cells $Y_1 = \cdots = Y_n = (0,1)^2$.

First, we define Y-periodic functions $b_{\mathbf{a}}, b_{\mathbf{d}} \colon \mathbb{R}^2 \to \mathbb{R}$ by setting

$$b_{\mathbf{a}}(y_1, y_2) = \sin^2(\pi y_1)\, \sin^2(\pi y_2) \quad \text{and} \quad b_{\mathbf{d}}(y_1, y_2) = \begin{cases} 1 & \text{if } y_1, y_2 \in \left(\frac{1}{4}, \frac{3}{4}\right), \\ 0 & \text{otherwise,} \end{cases} \tag{5.1}$$

for all $y_1, y_2 \in [0,1]$. The multiscale diffusion tensors considered below in this section involve the above two functions composed with argument rescaling so as to introduce microscales and the affine function $a_0 \colon D \to \mathbb{R}$ defined by

$$a_0(x_1, x_2) = 1 + \frac{x_1}{2} + \frac{x_2}{2} \quad \text{for all} \quad x_1, x_2 \in (0,1)\,, \tag{5.2}$$

which we use to introduce macroscale dependence.

The diffusion coefficients are constructed as follows. For a number $n \in \mathbb{N}_0$ of microscales $\varepsilon_1, \dots, \varepsilon_n$ and fixed oscillation amplitudes $\mu_1, \dots, \mu_n \in (0,1)$, we define $a_{\mathbf{a}}, a_{\mathbf{d}} \colon D \times Y_1 \times \cdots \times Y_n \to \mathbb{R}$:

$$a_{\mathbf{a}} = a_0 \otimes (1 - \mu_1 b_{\mathbf{a}}) \otimes \cdots \otimes (1 - \mu_n b_{\mathbf{a}}) \quad \text{and} \quad a_{\mathbf{d}} = a_0 \otimes (1 - \mu_1 b_{\mathbf{d}}) \otimes \cdots \otimes (1 - \mu_n b_{\mathbf{d}})\,. \tag{5.3}$$

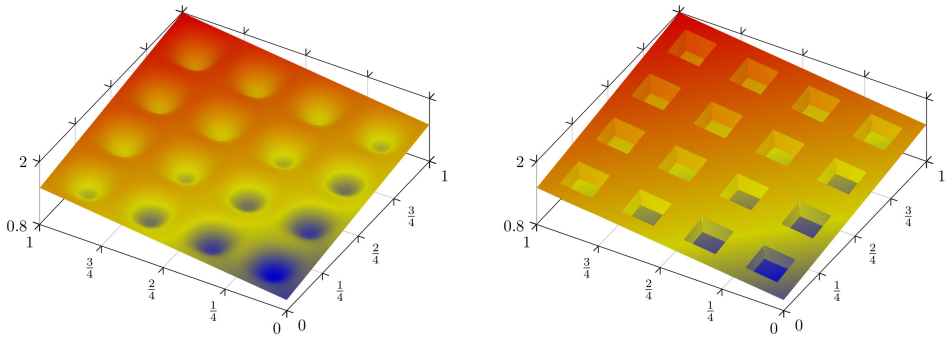

Figure 5.1: The graphs of the scalar diffusion coefficients $a^\varepsilon_{\mathbf{a}}$ (**left**) and $a^\varepsilon_{\mathbf{d}}$ (**right**) with a single ($n = 1$) microscale $\varepsilon_1 = 2^{-2} = \frac{1}{4}$, and with oscillation amplitude $\mu_1 = 0.2$, see (5.1) and (5.4) Here, $\lambda = -\log_2 \varepsilon_1 = 2$ is the index of the level at which the factors introduced via $b_{\mathbf{a}}$ and $b_{\mathbf{d}}$ oscillate. The diffusion coefficients used in our experiments involve up to $n = 5$ such microscales, specified in (5.8) and (5.9), the finest one being $\varepsilon_5 = 2^{-40} \approx 9 \cdot 10^{-13}$.

Then we consider $A_{\mathbf{a}}, A_{\mathbf{d}} \colon D \times Y_1 \times \cdots \times Y_n \to \mathbb{R}$ given by $A_{\mathbf{a}} = I \cdot a_{\mathbf{a}}$ and $A_{\mathbf{d}} = I \cdot a_{\mathbf{d}}$, where I is the identity matrix of order two. Note that $A_{\mathbf{a}}$ satisfies both Assumptions 2.1 and 4.1, whereas $A_{\mathbf{d}}$ fails to satisfy either (namely, the continuity and analyticity conditions) due to jump discontinuities with respect to the "fast" variables. Nevertheless, the functions $A_{\mathbf{a}}$ and $A_{\mathbf{d}}$ induce by (2.1) the respective diffusion coefficients[2] $A^\varepsilon_{\mathbf{a}}, A^\varepsilon_{\mathbf{d}} \in L^\infty(D; \mathbb{R}^{d\times d}_{\mathrm{sym}})$. These are given by $A^\varepsilon_{\mathbf{a}} = I \cdot a^\varepsilon_{\mathbf{a}}$ and $A^\varepsilon_{\mathbf{d}} = I \cdot a^\varepsilon_{\mathbf{d}}$, where $a^\varepsilon_{\mathbf{a}}, a^\varepsilon_{\mathbf{d}} \in L^\infty(D)$ are the scalar diffusion coefficients defined by

$$\begin{aligned} a^\varepsilon_{\mathbf{a}}(x) &= a_0(x) \cdot \Big(1 - \mu_1 b_{\mathbf{a}}\Big(\frac{x}{\varepsilon_1}\Big)\Big) \cdots \Big(1 - \mu_n b_{\mathbf{a}}\Big(\frac{x}{\varepsilon_n}\Big)\Big) \quad \text{and} \\ a^\varepsilon_{\mathbf{d}}(x) &= a_0(x) \cdot \Big(1 - \mu_1 b_{\mathbf{d}}\Big(\frac{x}{\varepsilon_1}\Big)\Big) \cdots \Big(1 - \mu_n b_{\mathbf{d}}\Big(\frac{x}{\varepsilon_n}\Big)\Big) \quad \text{for a.e. } x \in D. \end{aligned} \tag{5.4}$$

The graphs of the scalar diffusion coefficients $a^\varepsilon_{\mathbf{a}}$ and $a^\varepsilon_{\mathbf{d}}$ with a single ($n = 1$) microscale $\varepsilon_1 = \frac{1}{4}$ and oscillation amplitude $\mu_1 = 0.2$ are shown in Figure 5.1.

Both $A^\varepsilon_{\mathbf{a}}$ and $A^\varepsilon_{\mathbf{d}}$ satisfy the uniform lower and upper bounds stipulated in Assumption 2.1 with the constants $\gamma = \prod_{i=1}^n (1 - \mu_i)$ and $\Gamma = 2$. The corresponding multiscale problems of the form (2.2) with the right-hand side function $f = 1 \in L^2(D)$, which are the weak formulations of the boundary-value problems

$$\begin{cases} \nabla^{\mathsf{T}} a^\varepsilon_{\mathbf{a}} \nabla u^\varepsilon = 1 & \text{on } D, \\ u^\varepsilon = 0 & \text{on } \partial D \end{cases} \qquad \text{and} \qquad \begin{cases} \nabla^{\mathsf{T}} a^\varepsilon_{\mathbf{d}} \nabla u^\varepsilon = 1 & \text{on } D, \\ u^\varepsilon = 0 & \text{on } \partial D, \end{cases} \tag{5.5}$$

are therefore well-posed, as we discussed in Section 2. These are the two series of multiscale problems that we study numerically in this section.

5.1. Analytic coefficients (*left plots*). We note that the coefficient $a_{\mathbf{a}}$ is analytic with respect to all variables for any number n of microscales, and the corresponding problems all satisfy Assumption 4.1, under which we analyzed homogenization approximations to the exact solutions obtained via the one-scale limit problem.

Let us denote by Θ the set of the corners of D and consider the corresponding weight function χ_Θ given by (4.33). The solution u^ε is analytic on D but exhibits a

[2] In the experiments presented here, we immediately treat multiscale problems. The microscales $\varepsilon_1, \ldots, \varepsilon_n$ are therefore fixed and no longer depend on any scale parameter. The superscript ε serves purely as a notational device, distinguishing, in particular, $A^\varepsilon_{\mathbf{a}}, A^\varepsilon_{\mathbf{d}}, a^\varepsilon_{\mathbf{a}}, a^\varepsilon_{\mathbf{d}}$ from $A_{\mathbf{a}}, A_{\mathbf{d}}, a_{\mathbf{a}}, a_{\mathbf{d}}$.

singularity at every corner $\vartheta \in \Theta$. Specifically, u^ε is the product of $\chi_\Theta^2 \log \chi_\Theta$ and of a function analytic on $\overline{D}$; see [17, Theorem 13 in §8 of Chapter 2] or [51, 48, 47]. This means, first of all, that $u^\varepsilon \in H^{1+\alpha}(D)$ holds for $\alpha \leq 1$ but not for $\alpha > 1$. In particular, the quasi-optimality bound (2.23) then holds with $\alpha = 1$. Furthermore, we have $u^\varepsilon \in C^2_{\Theta,0}(D)$ in the sense of Definition 4.13 for any fixed positive microscales $\varepsilon_1, \ldots, \varepsilon_n$, but the constants in the definition cannot be chosen independently of the microscales.

In the present setting, the upscaled coefficients (3.17) all inherit isotropy from $A^\varepsilon_{\mathbf{a}}$, so that the homogenized coefficient A_0 given by (3.20) is isotropic. Together with the boundary conditions of (5.5), this implies [8, 9] that $u_0 \in H^2(D)$ and, furthermore, $u_0 \in C^2_{\Theta,0}(D)$. By Theorem 4.15, u_0 has an infinite sequence of low-rank approximations satisfying Assumption 4.6.

Finally, we note that in the case of a single ($n = 1$) microscale the hypothesis of Proposition 3.5 is satisfied and the bound (3.11) holds. The discussion given in Section 4.3.5 therefore applies to this case.

5.2. Discontinuous coefficients (*right plots*). The coefficient $a_{\mathbf{d}}$, on the other hand, exhibits jump discontinuities with respect to the "fast" variables for any *positive* number n of microscales, and the corresponding problems therefore all fail to satisfy Assumption 4.1 and to be covered by the analysis given in the present paper.

The discontinuous coefficients give rise to solutions of low regularity. For any positive $\varepsilon_1, \ldots, \varepsilon_n$, the jump discontinuities in the diffusion coefficient $a^\varepsilon_{\mathbf{d}}$ induce jump discontinuities in ∇u^ε, leading, in particular, to $u^\varepsilon \notin H^{1+\alpha}(D)$ for any $\alpha \geq \frac{1}{2}$. Locally, on every $\Omega \subset D$ such that $a^\varepsilon_{\mathbf{d}}$ is constant on Ω, we have $u^\varepsilon \in H^{1+\alpha}(\Omega)$ if and only if $\alpha < \alpha_*$, where $\alpha_* \geq \frac{2}{3}$ is a critical exponent determined, in our setting, by the ratio of the values of $a^\varepsilon_{\mathbf{d}}$ on the two sides of $\partial\Omega$. For details, we refer to the work [16], which covers the discontinuous coefficients we consider here. Since the interfaces across which $a^\varepsilon_{\mathbf{d}}$ (and hence of ∇u^ε) is discontinuous are aligned with the partitioning underlying the finite-element spaces V^L and $\bar{V}^L$ when $L \geq \lambda_n$ (see Section 2.1 and Assumption 4.1), a convergence bound analogous to (2.23) holds with any $\alpha \in (0, \alpha_*)$, in a suitable broken norm and with a positive constant on the right-hand side.

As our numerical results demonstrate below, the solutions of multiscale problems with diffusion coefficients $a^\varepsilon_{\mathbf{d}}$ have low-rank approximations as efficient as the solutions of the problems with the analytic coefficients $a^\varepsilon_{\mathbf{a}}$.

5.3. Experiment design. We present below a numerical study of the low-rank approximability (in the sense of Section 2.1) of the weak solutions u^ε of the problems (5.5) with $n \in \{0, \ldots, 5\}$ microscales given by (5.8) and (5.9) and with the corresponding diffusion coefficients (5.4). Since the exact solutions are not available in closed form, we approximated them by solving sufficiently fine discretizations of the problems. The numerical experiments were performed[3] in Julia, on a desktop computer, with a peak RAM usage of 6Gb, using the package `TensorRefinement.jl` [36] and a numerical method that was developed independently from the present paper and will be presented in a forthcoming publication [37]. For a brief overview of the key approaches for solving linear systems in the TT-MPS representation that are currently available in the literature, we refer to Section 2.2 above. For each problem (5.5) (with every number n of microscales and every diffusion coefficient we considered), the following was done.

First, an approximate solution $\widetilde{u}^\varepsilon$ of a discretization with 51 levels was computed in the QTT representation (2.13), (2.8). Second, for a selected range of $L \in \{2, \ldots, 50\}$, a projection $\widetilde{u}^{\varepsilon,L}$ of $(\widetilde{u}^\varepsilon, \nabla\widetilde{u}^\varepsilon)$ onto $(V^L)^3$ and a low-rank approximation $\bar{u}^{\varepsilon,L} \in (V^L)^3$ of $\widetilde{u}^{\varepsilon,L}$ such that

$$\|\bar{u}^{\varepsilon,L} - (\widetilde{u}^\varepsilon, \nabla\widetilde{u}^\varepsilon)\|_{L^2(D)^3} \leq q \, \|\widetilde{u}^{\varepsilon,L} - (\widetilde{u}^\varepsilon, \nabla\widetilde{u}^\varepsilon)\|_{L^2(D)^3} \tag{5.6}$$

[3] The code and data of the numerical experiments will be made publicly available at github.com/vladimirkazeev/kors2.

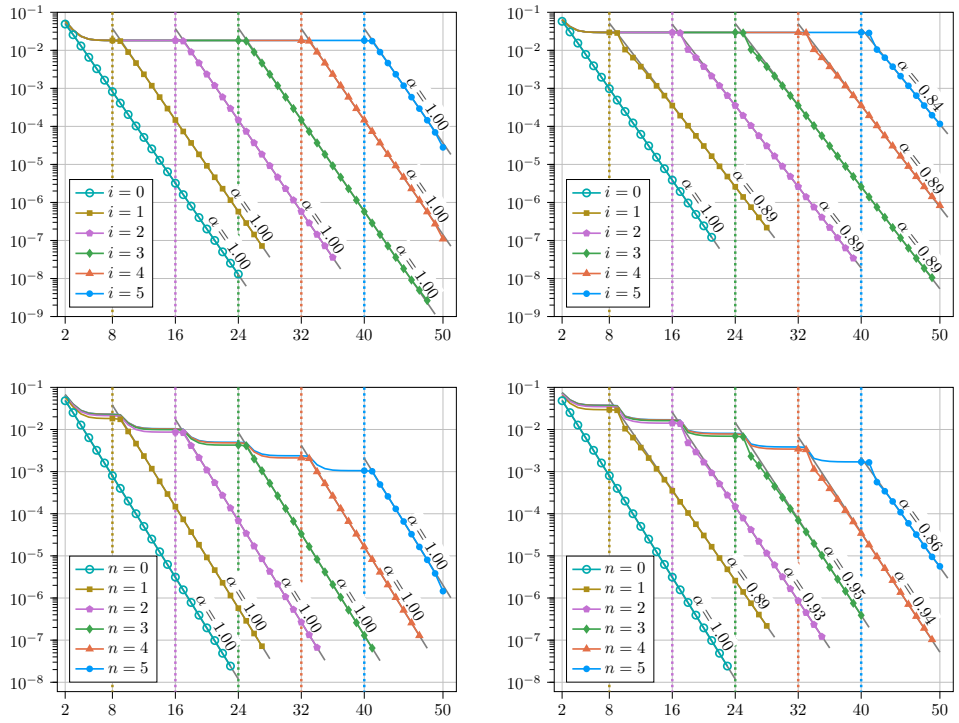

Figure 5.2: Multiscale problems (5.5): the projection error $\|\widetilde{u}^{\varepsilon,L} - (\widetilde{u}^{\varepsilon}, \nabla\widetilde{u}^{\varepsilon})\|_{L^2(D)^3}$ (vertical axes) against the number L of discretization levels (horizontal axes).
Top. Problems with *a single microscale* each, see Section 5.4.
Bottom. Problems with *up to five microscales* each, see Section 5.5.
Left. Diffusion coefficients $a^{\varepsilon}_{\mathbf{a}}$. **Right.** Diffusion coefficients $a^{\varepsilon}_{\mathbf{d}}$.
For every $i \in \{1, \ldots, 5\}$, the markers for the experiments with the finest microscale 2^{-8i} are not shown for $L < 8i$. For each $i \in \{1, \ldots, 5\}$, the microscale 2^{-8i} is depicted by a dotted vertical line $L = 8i$ of the color corresponding to the experiments with that microscale as the finest one. Each reference line (gray) represents the dependence $C\,2^{-\alpha L}$ with the value of α indicated along the line and with a positive constant C.

were computed for $q = \frac{3}{2}$. By the triangle inequality, the low-rank QTT approximation $\bar{u}^{\varepsilon,L} \in (V^L)^3$ satisfies the following accuracy bound:

$$\|\bar{u}^{\varepsilon,L} - (u^{\varepsilon}, \nabla u^{\varepsilon})\|_{L^2(D)^3} \leq \frac{3}{2}\,\|\widetilde{u}^{\varepsilon,L} - (\widetilde{u}^{\varepsilon}, \nabla\widetilde{u}^{\varepsilon})\|_{L^2(D)^3} + \|\widetilde{u}^{\varepsilon} - u^{\varepsilon}\|_{H^1(D)}\,. \tag{5.7}$$

For each problem, an optimal *a posteriori* error estimate and the dependence of the right-hand side of (5.6) on L were used to select the range of L for each problem so as to ensure that For our choice of the range of L for each problem, the first term on the right-hand side of (5.7) dominates the second and serves as a suitable proxy for the error we aim to track, given on the left-hand side of (5.7). This means that $\widetilde{u}^{\varepsilon}$ sufficiently accurately approximates the exact solution u^{ε} and the exact low-rank structure of $\bar{u}^{\varepsilon,L}$ therefore sufficiently accurately reflects the low-rank approximability of $(u^{\varepsilon}, \nabla u^{\varepsilon})$.

5.4. A single microscale (*top plots*). In this series of experiments, we consider a single ($n = 1$) fixed microscale ε_1 and the corresponding oscillation amplitude μ_1 set as follows:

$$\varepsilon_1 = 2^{-\lambda_1} \quad \text{with} \quad \mu_1 = \frac{1}{2} \quad \text{for} \quad \lambda_1 = 8i\,, \qquad \text{where} \quad i \in \{0, \ldots, 5\}\,. \tag{5.8}$$

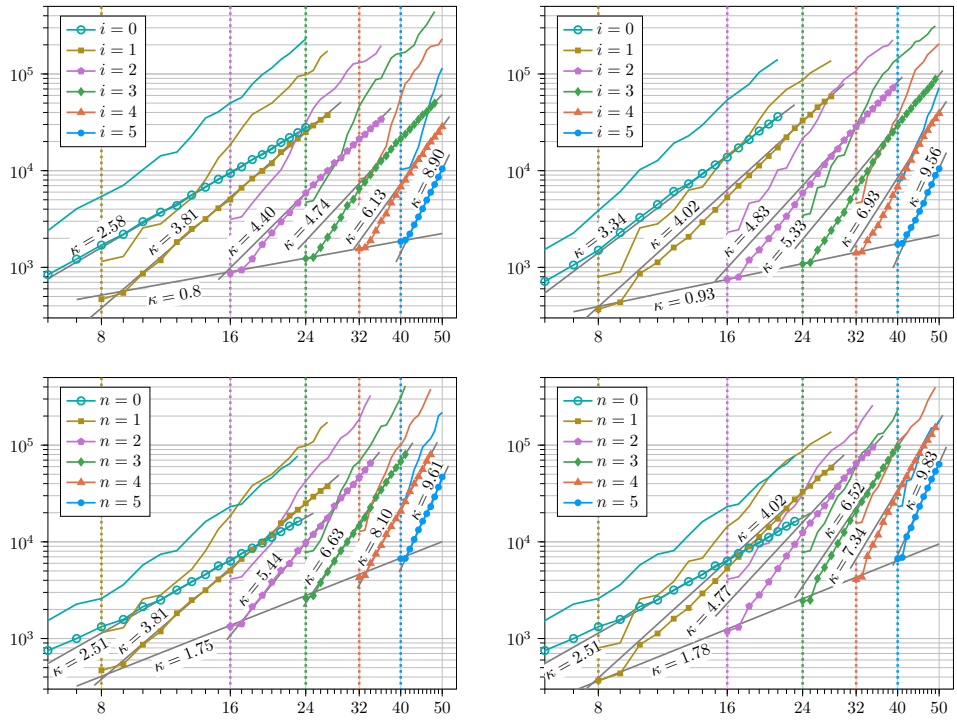

Figure 5.3: Multiscale problems (5.5): the number N of effective parameters (2.14) of the QTT representation of $\bar{u}^{\varepsilon,L}$ (vertical axes) against the number L of discretization levels (horizontal axes).
Top. Problems with *a single microscale* each, see Section 5.4.
Bottom. Problems with *up to five microscales* each, see Section 5.5.
Left. Diffusion coefficients $a^{\varepsilon}_{\mathbf{a}}$. **Right.** Diffusion coefficients $a^{\varepsilon}_{\mathbf{d}}$.
For each $i \in \{1, \dots, 5\}$, the microscale 2^{-8i} is depicted by a dotted vertical line $L = 8i$ of the color corresponding to the experiments with that microscale as the finest one. Each reference line (gray) represents the dependence $N = c\, L^{\kappa}$ with the value of κ indicated along the line and with a positive constant c. For each $i \in \{1, \dots, 5\}$, the solid line without markers and in the respective color shows the bound (2.15) for N in terms of the maximum rank $r_{\max}$ obtained experimentally (the latter is plotted in Figures 5.4 and 5.5).

In the case of $i = 0$, which we include for comparison, we have $\lambda_1 = \lambda_0 \equiv 1$ (the microscale coalesces with the macroscale), so that $a^{\varepsilon}_{\mathbf{a}} = a_0(1 - \frac{1}{2}b_{\mathbf{a}})$ and $a^{\varepsilon}_{\mathbf{d}} = a_0(1 - \frac{1}{2}b_{\mathbf{d}})$.

5.5. Several microscales (*bottom plots*). In this series of experiments, we consider $n \in \{0, \dots, 5\}$ fixed microscales $\varepsilon_1, \dots, \varepsilon_n$ and corresponding oscillation amplitudes $\mu_1, \dots, \mu_n$ set as follows:

$$\varepsilon_i = 2^{-\lambda_i} \quad \text{with} \quad \mu_i = 2^{-i} \quad \text{for} \quad \lambda_i = 8i\,, \qquad \text{where} \quad i \in \{1, \dots, 5\}\,. \tag{5.9}$$

In the case of $n = 0$, which we include for comparison, we have $a^{\varepsilon}_{\mathbf{a}} = a^{\varepsilon}_{\mathbf{d}} = a_0$.

5.6. Discussion. The error $\|\widetilde{u}^{\varepsilon,L} - (\widetilde{u}^{\varepsilon}, \nabla\widetilde{u}^{\varepsilon})\|_{L^2(D)^3}$ is shown in Figure 5.2. Consistently with the discussion regarding discrete solutions given in Section 2.2 above, this error decays exponentially with respect to L, i.e., algebraically with respect to the step 2^{-L} of the virtual grid (2.4). The exponent α reflects the Sobolev smoothness of the solution. In the experiments with analytic diffusion

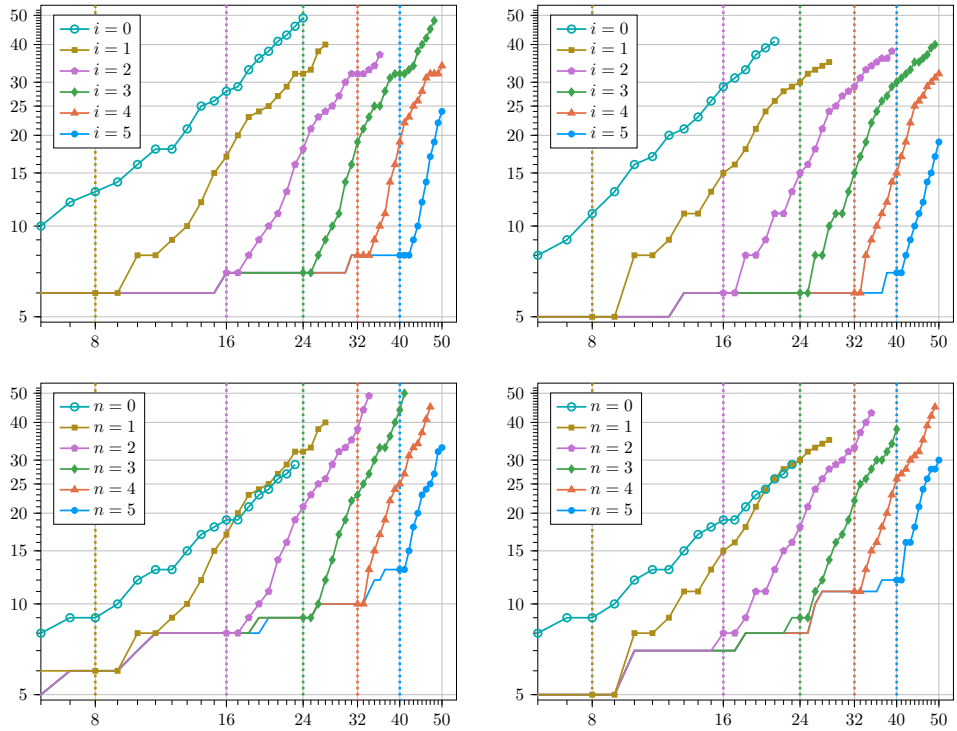

Figure 5.4: Multiscale problems (5.5): the maximum rank (2.15) $r_{\max}$ of the QTT representation of $\bar{u}^{\varepsilon,L}$ (vertical axes). against the number L of discretization levels (horizontal axes)
Top. Problems with *a single microscale* each, see Section 5.4.
Bottom. Problems with *up to five microscales* each, see Section 5.5.
Left. Diffusion coefficients $a^{\varepsilon}_{\mathbf{a}}$. **Right.** Diffusion coefficients $a^{\varepsilon}_{\mathbf{d}}$.

coefficients, we observe $\alpha = 1$, which is in agreement with the discussion given in Section 5.1. In the experiments with discontinuous coefficients, we observe $\alpha \approx 0.89$ in the case of a single microscale ($n = 1$), which agrees with Section 5.2. Indeed, by [16, equations (21a) and (21b)], the bound for the index of local Sobolev smoothness is $\alpha_* = \frac{4}{\pi} \tan^{-1} \frac{\sqrt{5}}{\sqrt{7}} \approx 0.8934$ when the diffusion coefficient jumps by a factor of two. In the case of $n > 1$ microscales (the bottom right plot), that ratio varies locally, from one interface to another, between $1 - \mu_1$ and $1 - \mu_n$, which results in convergence with intermediate values of α.

Figures 5.3 and 5.4 show how the number N of parameters (2.14) and the maximum rank (2.15) $r_{\max}$ of the computed representations (2.13) of $\bar{u}^{\varepsilon,L} \in (V^L)^3$ depend on L for each problem. For N, we observe

$$N \leq C \, \lambda_n^{\kappa_0} \left(\frac{L}{\lambda_n} \right)^{\kappa} \tag{5.10}$$

with positive constants C and κ_0 that do *not* depend on the problem within each series of experiments and with a positive constant κ that does. For the experiments with discontinuous coefficients, Figure 5.3 shows that the slopes decrease as L increases. That is consistent with our understanding that none of our experiments is in the asymptotic regime for N corresponding to large L, which requires $L \gg \lambda_n$. Nevertheless, the experimentally observed bound (5.10) is of practical relevance since our experiments push the number L of levels to the limit imposed by the double-precision floating-point format. Figure 5.3 shows also, in solid colored lines

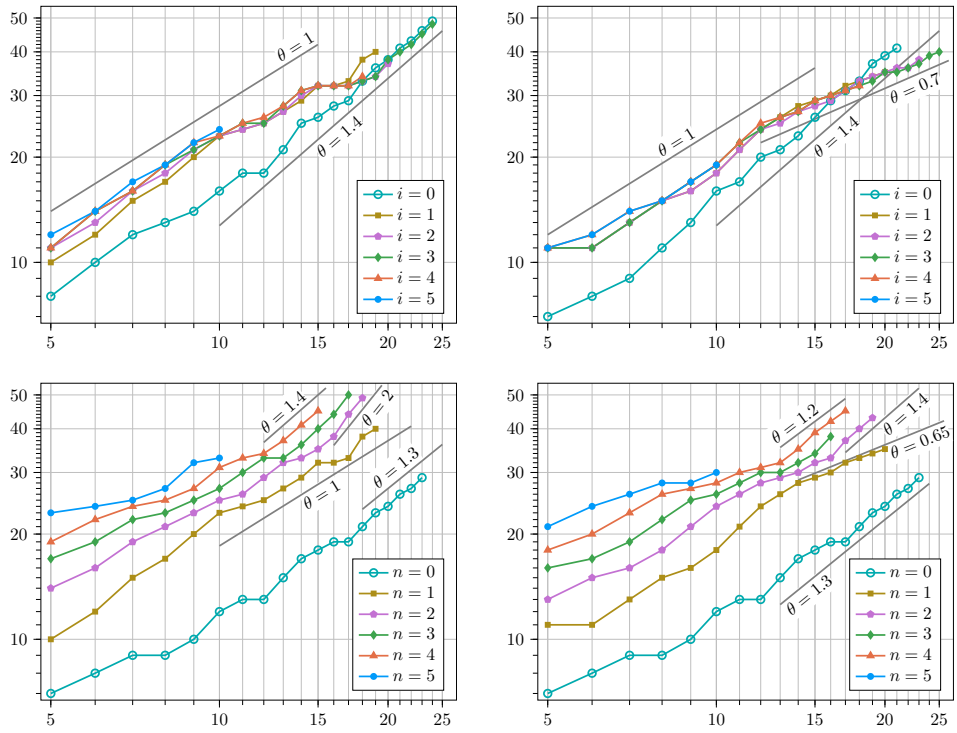

Figure 5.5: Multiscale problems (5.5): the maximum rank (2.15) $r_{\max}$ of the QTT representation of $\bar{u}^{\varepsilon,L}$ (vertical axes) against the number Λ of discretization levels beyond the finest microscale (horizontal axes)
Top. Problems with *a single microscale* each, see Section 5.4.
Bottom. Problems with *up to five microscales* each, see Section 5.5.
Left. Diffusion coefficients $a^{\varepsilon}_{\mathbf{a}}$. **Right.** Diffusion coefficients $a^{\varepsilon}_{\mathbf{d}}$.
Each reference line (gray) represents the dependence $r_{\max} = c\,\Lambda^{\theta}$ with the value of θ indicated along the line and with a positive constant c.

without markers, the bound (2.15) for N in terms of L and $r_{\max}$ for the QTT decompositions obtained experimentally, which is rather crude compared to the actual values of N.

Figure 5.5 reports the dependence of the maximum QTT rank $r_{\max}$ (2.15) of the representations of $\bar{u}^{\varepsilon,L} \in (V^L)^3$ on $\Lambda = L - \lambda_n$ for each problem, which is precisely what we analyzed theoretically in Sections 4.3.4 and 4.3.5.

Specifically, for each problem, we observe $r_{\max} \leq C\Lambda^{\theta}$ with positive constants C and θ depending on the problem. We note that in the experiments with one microscale ($n = 1$, top plots), the values of $r_{\max}$ are almost the same for each Λ independently of λ_1. Qualitatively, this behavior agrees with the theoretical bound stated in Theorem 4.12. However, for none of the problems considered does θ exceed two, and our theoretical value $\theta = 2n + 4$ is noticeably larger. Such a discrepancy between theoretical bounds and experimentally observed dependencies is remarkable but not at all unexpected or uncommon, see Remarks 2.3 and 2.4 above and, for example, [40, Section 6.2.2].

Furthermore, in the case of $n > 1$, the exponential decay (5.9) of μ_i with respect to the microscale index i allows to adapt the polynomial degree for each microscale variable by letting the respective values of c in (4.16) decrease linearly with respect to the microscale index. This has a positive effect even for the approximation of $\mathcal{U}^{\varepsilon} v_n$ by analytical techniques we constructed in our theoretical analysis. In addition, the

product form (5.4) of the diffusion coefficients allows to strengthen (4.22) and (4.23) in the sense that, for each $i \in \{1, \dots, n\}$, the functions w_i^L (4.14), and v_i^L, u_i^L are also of product form. Their factors corresponding to $D, Y_1, \dots, Y_i$ can then be approximated independently, which means that the rank bound developed in Section 4.3.4 can be improved in this particular setting.

Finally, comparing Figures 5.4 and 5.5 for the problems with one microscale ($n = 1$), we note that the dependence of $r_{\max}$ on L in the pre-asymptotic regime of $L \simeq \lambda_1$ is described as algebraic growth with respect to $\Lambda = L - \lambda_1$ more accurately (and with an exponent independent of λ_1) than with respect to L. For any positive C and θ, the dependence $r_{\max} = C\Lambda^\theta$ corresponds to a straight line in a plot of $\log r_{\max}$ against $\log \Lambda$, which becomes a curved line in a plot of $\log r_{\max}$ against $\log L$. While the asymptotic slope for $L \to \infty$ equals θ for both the lines, in the pre-asymptotic regime, the slope of the latter line grows linearly with respect to λ_1 at any fixed $\Lambda > 0$:

$$\frac{\mathrm{d} \log \Lambda^\theta}{\mathrm{d} \log L} = \theta \left(1 + \frac{\lambda_1}{\Lambda}\right). \tag{5.11}$$

This elementary calculation explains the growth of the pre-asymptotic slopes in the log-log plots of N and $r_{\max}$ against L (the top plots in Figures 5.3 and 5.5) with respect to λ_1.

6. Conclusions and outlook. We analyzed theoretically and studied numerically the approximability of the solution of a linear elliptic second-order multiscale diffusion problem in a multilevel tensor decomposition. In our analysis, as the first of several intermediate approximations, we used the approximation of the solution and of its derivatives obtained by homogenization. The homogenization procedure was represented by a high-dimensional but one-scale limit problem, for which we assumed the analyticity of the data. The resulting analyticity of the solution of the limit problem allowed to construct approximations of accuracy converging exponentially with respect to a discretization parameter and of tensor ranks bounded algebraically with respect to the same discretization parameter. That led us to an analogous result for the corresponding solution of the original, multiscale problem, where the above discretization parameter takes the role of the number of discretization levels beyond the finest microscale. The behavior of the tensor ranks predicted by our analysis is confirmed in half of our numerical experiments.

The other half of our numerical experiments extends beyond the strong analyticity assumptions, which are necessary for the particular way of constructing QTT-FE approximations used in the theoretical analysis presented here. Those experiments involve jump discontinuities in the diffusion coefficient and clearly indicate that analogous theoretical results can be expected for non-analytic data. Another example is the setting of a perforated medium, when the unit cells have "holes". In both these settings, the gradient v_n (3.7) exhibits singularities on inner interfaces within the respective unit cells or on the boundaries thereof. The corresponding generalization of one-scale limit problems for perforated media is given in [20]. The remainder of our approximation scheme can be adapted accordingly, with suitable spaces replacing the factors on the right-hand side of (4.20). Regularity results for parametric unit-cell problems in countably normed spaces are available (for $n = 1$ microscale and $d = 2$ space dimensions) in [53]. When combined with the QTT-FE approximations developed in [40, 35] and with the treatment of complex geometry developed in [35, Sections 3 and 5], these can be expected to lead to accuracy and rank bounds of QTT-FE approximation analogous to those obtained here.

Furthermore, similar results for homogenization hold for other settings, such as the so-called reticulated structures and lattice materials (see the survey [21] and the references therein) and for other types of PDEs (see, e.g., [73, 74, 75, 18, 67] and the references therein). The corresponding development of QTT-FE approaches for

these problem classes is a natural extension of the present analysis. Additionally, we point out that high-dimensional one-scale limit problems with the same tensor structure as those considered here arise also for certain *non-periodic multiscale problems*, which fall into the class of the so-called *homogenization structures*, as proposed by Nguetseng in [56]. Finally, we emphasize also that analogous homogenization results are available for *nonlinear problems with multiple scales*; we refer to [29] and to the references therein for further details. The results of the present paper indicate that such problems can also be solved efficiently by QTT-FE discretization in combination with a nonlinear solver.

7. Appendix.

Proof of Lemma 4.3.

Proof. Let id denote the identity transformation with respect to a scalar variable ranging in $(0,1)$. For all $L \in \mathbb{N}$ and $k \in \{1,\dots,d\}$, the errors bounded by the claim can be represented by telescoping sums as follows:

$$v - \bar{\Pi}_i^L v = \sum_{k'=1}^{d} \bar{\Pi}_{i-1}^L \otimes \left\{ \left(\bigotimes_{k=1}^{k'-1} \pi^L \right) \otimes \left(\mathsf{id} - \pi^L\right) \otimes \mathsf{id}^{\otimes(d-k')} \right\} v$$
$$+ \sum_{j'=1}^{i} \sum_{k'=1}^{d} \bar{\Pi}_{j'-1}^L \otimes \left\{ \left(\bigotimes_{k=1}^{k'-1} \bar{\pi}^L \right) \otimes \left(\mathsf{id} - \bar{\pi}^L\right) \otimes \mathsf{id}^{\otimes(d-k')} \right\} \otimes \mathsf{id}^{\otimes(i-j')d} \otimes \mathsf{id}^{\otimes d}\, v\,,$$

$$w - \Pi_i^L w = \sum_{k'=1}^{d} \bar{\Pi}_{i-1}^L \otimes \left\{ \left(\bigotimes_{k=1}^{k'-1} \pi^L \right) \otimes \left(\mathsf{id} - \pi^L\right) \otimes \mathsf{id}^{\otimes(d-k')} \right\} w$$
$$+ \sum_{j'=1}^{i} \sum_{k'=1}^{d} \bar{\Pi}_{j'-1}^L \otimes \left\{ \left(\bigotimes_{k=1}^{k'-1} \bar{\pi}^L \right) \otimes \left(\mathsf{id} - \bar{\pi}^L\right) \otimes \mathsf{id}^{\otimes(d-k')} \right\} \otimes \mathsf{id}^{\otimes(i-j')d} \otimes \mathsf{id}^{\otimes d}\, w\,,$$

$$\partial_{ik}(w - \Pi_i^L w) = \sum_{k'=1}^{d} \bar{\Pi}_{i-1}^L \otimes \left\{ \left(\bigotimes_{k=1}^{k'-1} \pi^L \right) \otimes \partial_{ik}\left(\mathsf{id} - \pi^L\right) \otimes \mathsf{id}^{\otimes(d-k')} \right\} w$$
$$+ \sum_{j'=1}^{i} \sum_{k'=1}^{d} \bar{\Pi}_{j'-1}^L \otimes \left\{ \left(\bigotimes_{k=1}^{k'-1} \bar{\pi}^L \right) \otimes \left(\mathsf{id} - \bar{\pi}^L\right) \otimes \mathsf{id}^{\otimes(d-k')} \right\} \otimes \mathsf{id}^{\otimes(i-j')d} \otimes \mathsf{id}^{\otimes d} \partial_{ik} w\,.$$

Applying Proposition 4.2 to these representations, we obtain the claimed bounds. ∎

Proof of Lemma 4.4.

Proof. The exponentials and shifted Chebyshev polynomials defined by (4.10) and (4.7) form orthogonal bases in the spaces $L^2(0,1)$ and $L^2_\omega(0,1)$ respectively, where ω is the Chebyshev weight function given by (4.8). It follows from the assumption that $w \in L^2_{\omega^{\otimes d}\otimes \mathsf{id}}(D \times \boldsymbol{Y}_i)$, so that w can be represented by the following absolutely convergent series:

$$w = \sum_{\alpha \in \mathbb{N}_0^d} \sum_{\beta_1 \in \mathbb{Z}^d} \cdots \sum_{\beta_i \in \mathbb{Z}^d} c_{\alpha,\beta_1,\dots,\beta_i} \left(\bigotimes_{k=1}^{d} \widetilde{T}_{\alpha_k} \right) \otimes \left(\bigotimes_{j=1}^{i} \bigotimes_{k=1}^{d} \widehat{T}_{\beta_{jk}} \right) \quad \text{in} \quad L^2_{\omega^{\otimes d}\otimes \mathsf{id}}(D \times \boldsymbol{Y}_i)\,. \tag{7.1}$$

Let us set $\kappa_0 = 1$, $\kappa_\alpha = \kappa_{-\alpha} = 2\,(-1)^\alpha$ for each $\alpha \in \mathbb{N}$ and $\kappa_\alpha = \kappa_{\alpha_1} \cdots \kappa_{\alpha_d}$ for all $\alpha \in \mathbb{Z}^d$. Then, due to (4.9) and (4.11), the coefficients of the series (7.1) can be expressed as follows:

$$c_{\alpha,\beta_1,\dots,\beta_i} = \frac{\kappa_\alpha}{\pi^d} \left\langle \left(\bigotimes_{k=1}^{d} \widetilde{T}_{\alpha_k} \right) \otimes \left(\bigotimes_{j=1}^{i} \bigotimes_{k=1}^{d} \widehat{T}_{\beta_{jk}} \right),\, w \right\rangle_{L^2_{\omega^{\otimes d}\otimes \mathsf{id}}(D \times \boldsymbol{Y}_i)} \tag{7.2}$$

for all $\alpha \in \mathbb{N}_0^d$ and $\beta_1, \dots, \beta_i \in \mathbb{Z}^d$.

The entire function $\varkappa\colon \mathbb{C} \to \mathbb{C}$ given by $\varkappa(\zeta) = (1 - \cos 2\pi\zeta)/2$ for all $\zeta \in \mathbb{C}$ bijectively maps each of the intervals $(0, 1/2)$ and $(1/2, 1)$ onto $(0, 1)$, with opposite orientations. Then, introducing $\mathfrak{X} = \varkappa^{\otimes d} \otimes \mathsf{id}\colon \mathbb{C}^{(i+1)d} \to \mathbb{C}^{(i+1)d}$, we can substitute $\mathfrak{X}$ in (7.2) to express the coefficients of w as follows:

$$c_{\alpha,\beta_1,\dots,\beta_i} = \sum_{\sigma\in\{\pm1\}^d} \widehat{c}_{\sigma\odot\alpha,\beta_1,\dots,\beta_i} \quad \text{for all } \alpha \in \mathbb{N}_0^d \text{ and } \beta_1, \dots, \beta_i \in \mathbb{Z}^d\,, \tag{7.3}$$

where "$\odot$" denotes the component-wise multiplication of multi-indices (so that $\sigma\odot\alpha = (\sigma_1\alpha_1, \dots, \sigma_d\,\alpha_d)$ for any $\sigma \in \{\pm1\}^d$ and $\alpha \in \mathbb{N}_0^d$) and

$$\widehat{c}_{\beta_0,\beta_1,\dots,\beta_i} = \kappa_{\beta_0} \Big\langle \bigotimes_{j=0}^{i} \bigotimes_{k=1}^{d} \widehat{T}_{\beta_{jk}},\, w \circ \mathfrak{X} \Big\rangle_{L^2(D\times \boldsymbol{Y}_i)} \quad \text{for all } \beta_0, \beta_1, \dots, \beta_i \in \mathbb{Z}^d\,. \tag{7.4}$$

For every $\delta > 0$, the function $\varkappa$ bijectively maps $\mathcal{S}_\delta = \big\{\xi - \mathrm{i}\eta\colon \xi \in (0,1)\,, \eta \in (0,\delta)\big\} \subset \mathbb{C}$ onto $\mathcal{E}_\delta = \big\{(1 - a_\eta \cos 2\pi\xi)/2 - \mathrm{i}\,(b_\eta \sin 2\pi\xi)/2\colon \xi \in (0,1)\,, \eta \in (0,\delta)\big\}$, where $a_\eta = \cosh 2\pi\eta$ and $b_\eta = \sinh 2\pi\eta$ for every $\eta > 0$. Note that $\mathcal{E}_\delta \cup ((1-a_\delta)/2, 1]$ is the image of the standard open Bernstein ellipse with parameter $\rho = e^{2\pi\delta}$ (with foci ±1 and semi-axes a_δ and b_δ) under the affine mapping $\mathbb{C} \ni z \mapsto (1 - z)/2 \in \mathbb{C}$. Since the function w is analytic on $\overline{D \times \boldsymbol{Y}_i}$ by assumption, it admits analytic continuation to an open neighborhood of $\overline{D \times \boldsymbol{Y}_i}$. Specifically, for some $\delta_{i0}, \delta_{i1}, \dots, \delta_{ii} > 0$, it has a unique continuous extension to $\overline{\mathcal{G}_i}$, where $\mathcal{G}_i = \mathcal{E}_{\delta_{i0}}^d \times \mathcal{S}_{\delta_{ij}}^{id}$, that is holomorphic on $\mathcal{G}_i$. We identify the original function w with this extension and set $M_i = \sup_{z\in\mathcal{G}_i} |w(z)|$. For the domain $\mathcal{D}_i = \mathcal{S}_{\delta_{ij}}^{(i+1)d}$, we have $\mathcal{G}_i = \mathfrak{X}(\mathcal{D}_i)$ and $\sup_{\zeta\in\mathcal{D}_i} |(w \circ \mathfrak{X})(\zeta)| = M_i$. Furthermore, $w\circ\mathfrak{X}$ is holomorphic on $\mathcal{D}_i$, continuous on $\overline{\mathcal{D}_i} = \mathfrak{X}(\overline{\mathcal{D}_i})$ and one-periodic with respect to each of its $(i+1)d$ variables. Using these properties and applying the Cauchy–Goursat theorem for the domain $\mathcal{D}_i$, we obtain

$$\widehat{c}_{\beta_0,\beta_1,\dots,\beta_i} = \kappa_\alpha \int\limits_{\times_{j=0}^{i}[-\mathrm{i}\delta_{ij},\,1-\mathrm{i}\delta_{ij}]^d} \cdots \int \Big(\bigotimes_{j=0}^{i} \bigotimes_{k=1}^{d} \widehat{T}^*_{\beta_{jk}}\Big)\,(w \circ \mathfrak{X})$$

and hence $|\widehat{c}_{\beta_0,\beta_1,\dots,\beta_i}| \le M_i \kappa_\alpha \exp(-\sum_{j=0}^{i} 2\pi\delta_{ij}|\beta_j|\,)$ for all $\beta_0, \beta_1, \dots, \beta_i \in \mathbb{N}_0^d$. Then (7.3) gives

$$|c_{\alpha,\beta_1,\dots,\beta_i}| \le 2^d M_i \kappa_\alpha \exp\Big(-\sum_{j=0}^{i} 2\pi\delta_{ij}|\beta_j|\,\Big) \tag{7.5}$$

for all $\alpha \in \mathbb{N}_0^d$ and $\beta_1, \dots, \beta_i \in \mathbb{Z}^d$.

Now we set $\delta_* = \min\{\delta_{i0}, \delta_{i1}, \dots, \delta_{ii}\}$ and verify the claimed bounds for $c = (2\pi\delta_*)^{-1}$, $p = \lceil c \log \epsilon^{-1} \rceil$ and a suitable positive constant C. Let $\mathcal{I}_0 = \{0, 1, \dots, p-1\}$, $\mathcal{J}_0 = \{0, \pm1, \dots, \pm(p-1)\}$ and $\mathcal{I}_1 = \mathbb{N}_0 \setminus \mathcal{I}_0$, $\mathcal{J}_1 = \mathbb{Z} \setminus \mathcal{J}_0$. Using the product index sets $\boldsymbol{\mathcal{I}}_\mu = \mathcal{I}_{\mu_1} \times \cdots \times \mathcal{I}_{\mu_d}$ and $\boldsymbol{\mathcal{J}}_\mu = \mathcal{J}_{\mu_1} \times \cdots \times \mathcal{J}_{\mu_d}$ with $\mu \in \{0,1\}^d$, we can recast the expansion (7.1) in $L^2_{\omega^{\otimes d}\otimes\mathsf{id}}(D \times \boldsymbol{Y}_i)$ as follows:

$$w = \sum_{m=0}^{(i+1)d} \sum_{\substack{\mu,\nu_1,\dots,\nu_i\in\{0,1\}^d\colon \\ |\mu|+\sum_{j=1}^{i}|\nu_j|=m}} \sum_{\substack{\alpha\in\boldsymbol{\mathcal{I}}_\mu \\ \beta_1\in\boldsymbol{\mathcal{J}}_{\nu_1} \\ \cdots \\ \beta_i\in\boldsymbol{\mathcal{J}}_{\nu_i}}} c_{\alpha,\beta_1,\dots,\beta_i} \Big(\bigotimes_{k=1}^{d} \widetilde{T}_{\alpha_k}\Big) \otimes \Big(\bigotimes_{j=1}^{i} \bigotimes_{k=1}^{d} \widehat{T}_{\beta_{jk}}\Big)\,. \tag{7.6}$$

In the right-hand side of (7.6), the term of the outer sum corresponding to $m = 0$ is

$\Pi_{i,p}\, w$, and the remainder can be bounded using (7.5):

$$\|w - \Pi_{i,p}\, w\|_{L^\infty(D\times \boldsymbol{Y}_i)} \le \sum_{m=1}^{(i+1)d} \sum_{\substack{\mu,\nu_1,\dots,\nu_i\in\{0,1\}^d:\\ |\mu|+\sum_{j=1}^i|\nu_j|=m}} \sum_{\substack{\alpha\in\boldsymbol{\mathcal{I}}_\mu\\ \beta_1\in\boldsymbol{\mathcal{J}}_{\nu_1}\\ \cdots\\ \beta_i\in\boldsymbol{\mathcal{J}}_{\nu_i}}} |c_{\alpha,\beta_1,\dots,\beta_i}|$$
$$\le \frac{2^d M_i\, 2^{d-1}\, 2^{id}}{(1-\lambda)^{(i+1)d}} \sum_{m=1}^{(i+1)d} \epsilon^m \binom{(i+1)d}{m} \le C_0\,\epsilon\,, \quad (7.7)$$

where $\lambda = e^{-2\pi\delta_*} \in (0,1)$ and $C_0 = 2^d M_i (i+1) d\, 2^{(i+1)d-1}(1+\epsilon_0)^{(i+1)d-1}/(1-\lambda)^{(i+1)d} > 0$. This gives the first of the bounds (4.13) with any constant $C \ge C_0$, selected independently of ϵ.

For derivatives of the shifted Chebyshev polynomials and exponentials, we have $\|\widetilde{T}'_\alpha\|_{L^\infty(0,1)} = 2\alpha^2$ for all $\alpha \in \mathbb{N}_0$ and $\|\widehat{T}'_\beta\|_{L^\infty(0,1)} = 2\pi|\beta|$ for all $\beta \in \mathbb{Z}$. Note that there exist positive constants γ_1 and γ_2 such that $\sum_{\beta=r}^\infty \beta\lambda^\beta \le \gamma_1(1-\lambda)^{-1} r \sum_{\beta=r}^\infty \lambda^\beta$ and $\sum_{\beta=r}^\infty \beta^2\lambda^s \le \gamma_2(1-\lambda)^{-2} r^2 \sum_{\beta=r}^\infty \lambda^\beta$ for any $r \in \mathbb{N}_0$. Using this, we obtain, as in (7.7), the following inequalities:

$$\|\partial_k(w - \Pi_{i,p}\, w)\|_{L^\infty(D\times\boldsymbol{Y}_i)} \le \frac{2\gamma_2 C_0 \epsilon p^2}{(1-\lambda)^2}\,,\ \|\partial_{jk}(w - \Pi_{i,p}\, w)\|_{L^\infty(D\times\boldsymbol{Y}_i)} \le \frac{2\pi\gamma_1 C_0\epsilon p}{1-\lambda}\,.$$

for all $k \in \{1,\dots,d\}$ and $j \in \{1,\dots,i\}$. This proves the last two of the bounds (4.13) with a suitable positive constant C independent of $\epsilon \in (0,\epsilon_0)$. □

7.1. Proof of Lemma 4.5.

Proof. Let $L \in \mathbb{N}$. Using the triangle inequality, we bound the errors as follows:

$$\begin{aligned} \|\partial_{ik}(w - \Pi_i^L \Pi_{i,p}\, w)\|_\infty &\le \|\partial_{ik}(\mathsf{id} - \Pi_{i,p})\, w\|_\infty + \|\partial_{ik}(\mathsf{id} - \Pi_i^L)\Pi_{i,p}\, w\|_\infty\,,\\ \|\partial_{ik} w - \bar\Pi_i^L \partial_{ik}\Pi_{i,p} w\|_\infty &\le \|\partial_{ik}(\mathsf{id} - \Pi_{i,p})\, w\|_\infty + \|(\mathsf{id} - \bar\Pi_i^L)\partial_{ik}\Pi_{i,p}\, w\|_\infty \end{aligned} \quad (7.8)$$

for every $k \in \{1,\dots,d\}$. By Lemma 4.4, there exist positive constants C_0 and c such that, for $p = \lceil cL\rceil$, we have

$$\|\partial_k(\mathsf{id} - \Pi_{i,p})\, w\|_\infty \le C_0\, p^2 2^{-L}\,, \quad \|\partial_{jk}(\mathsf{id} - \Pi_{i,p})\, w\|_\infty \le C_0\, p\, 2^{-L}\,. \quad (7.9)$$

Certain derivatives of $\Pi_{i,p}\, w$ can be bounded in terms of first-order derivatives of $\Pi_{i,p}\, w$ using the Bernstein's inequality for trigonometric polynomials. Applying it together with the bounds (7.9) and Lemma 4.3, we obtain

$$\begin{aligned} \|\partial_{ik}(\mathsf{id} - \Pi_i^L)\Pi_{i,p}\, w\|_\infty &\le 2^{-L} \sum_{k'=1}^d \|\partial_{k'}\partial_{ik}\Pi_{i,p}\, w\|_\infty\\ &\quad + 2^{-L}\sum_{j'=1}^{i-1}\sum_{k'=1}^d \|\partial_{j'k'}\partial_{ik}\Pi_{i,p}\, w\|_\infty + 2^{-L}\sum_{k'=1}^d \|\partial_{ik'}^2\partial_{ik}\Pi_{i,p}\, w\|_\infty\\ &\le 2^{-L}\sum_{k'=1}^d 2\pi p\, \|\partial_{k'}\Pi_{i,p}\, w\|_\infty + 2^{-L}\sum_{j'=1}^{i-1}\sum_{k'=1}^d 2\pi p\, \|\partial_{j'k'}\Pi_{i,p}\, w\|_\infty\\ &\quad + 2^{-L}\sum_{k'=1}^d (2\pi p)^2\, \|\partial_{ik}\Pi_{i,p}\, w\|_\infty \le 2^{-L}\sum_{k'=1}^d 2\pi p\, \big\{\|\partial_{k'} w\|_\infty + C_0\, p^2 2^{-L}\big\}\\ &\quad + 2^{-L}\sum_{j'=1}^{i} (2\pi p)^2\, \big\{\|\partial_{jk} w\|_\infty + C_0\, p^2 2^{-L}\big\} \le C_1\, p^2\, 2^{-L} \end{aligned} \quad (7.10)$$

for every $k \in \{1, \dots, d\}$ with a positive constant C_1 independent of L. The same approach leads to the bound

$$\begin{aligned}
\|(\mathsf{id} - \bar{\Pi}_i^L)\partial_{ik}\Pi_{i,p}\, w\|_\infty &\le 2^{-L}\sum_{k'=1}^{d}\|\partial_{k'}\partial_{ik}\Pi_{i,p}w\|_\infty + 2^{-L}\sum_{j'=1}^{i}\sum_{k'=1}^{d}\|\partial_{j'k'}\partial_{ik}\Pi_{i,p}w\|_\infty \\
&\le 2^{-L}\sum_{k'=1}^{d} 2\pi p\,\|\partial_{k'}\Pi_{i,p}w\|_\infty + 2^{-L}\sum_{j'=1}^{i}\sum_{k'=1}^{d} 2\pi p\,\|\partial_{j'k'}\Pi_{i,p}w\|_\infty \\
&\le 2^{-L}\sum_{k'=1}^{d} 2\pi p\left\{\|\partial_{k'}w\|_\infty + C_0\,p^2\,2^{-L}\right\} + 2^{-L}\sum_{j'=1}^{i}\sum_{k'=1}^{d} 2\pi p\left\{\|\partial_{j'k'}w\|_\infty + C_0\,p\,2^{-L}\right\} \\
&\le C_2\,p^2\,2^{-L} \qquad (7.11)
\end{aligned}$$

for every $k \in \{1, \dots, d\}$ with a positive constant C_2 independent of L. Combining inequalities (7.10) and (7.11) with (7.9) and (7.8), we obtain the claimed error bounds with $C = C_0 + \max\{C_1, C_2\}$. □

REFERENCES

[1] A. Abdulle, W. E, B. Engquist, and E. Vanden-Eijnden, *The heterogeneous multiscale method*, Acta Numer., 21 (2012), pp. 1–87, https://doi.org/10.1017/S0962492912000025.

[2] M. Ali and A. Nouy, *Approximation with tensor networks. Part I: Approximation spaces*, 2020, http://arxiv.org/abs/2007.00118. arXiv preprint 2007.00118.

[3] G. Allaire, *Homogenization and two-scale convergence*, SIAM Journal on Mathematical Analysis, 23 (1992), pp. 1482–1518, https://doi.org/10.1137/0523084.

[4] G. Allaire and M. Briane, *Multiscale convergence and reiterated homogenisation*, Proceedings of the Royal Society of Edinburgh: Section A. Mathematics, 126 (1996), pp. 297–342, https://doi.org/10.1017/S0308210500022757.

[5] R. Altmann, P. Henning, and D. Peterseim, *Numerical homogenization beyond scale separation*, Acta Numerica, 30 (2021), p. 1–86, https://doi.org/10.1017/S0962492921000015.

[6] I. Babuska and R. Lipton, *Optimal local approximation spaces for generalized finite element methods with application to multiscale problems*, Multiscale Model. Simul., 9 (2011), pp. 373–406, https://doi.org/10.1137/100791051,

[7] I. Babuška and B. Guo, *The h-p version of the finite element method for domains with curved boundaries*, SIAM Journal on Numerical Analysis, 25 (1988), pp. 837–861, https://doi.org/10.1137/0725048.

[8] I. Babuška and B. Guo, *Regularity of the solution of elliptic problems with piecewise analytic data. Part I. boundary value problems for linear elliptic equation of second order*, SIAM Journal on Mathematical Analysis, 19 (1988), pp. 172–203, https://doi.org/10.1137/0519014.

[9] I. Babuška and B. Guo, *Regularity of the solution of elliptic problems with piecewise analytic data. II: The trace spaces and application to the boundary value problems with nonhomogeneous boundary conditions*, SIAM Journal on Mathematical Analysis, 20 (1989), pp. 763–781, https://doi.org/10.1137/0520054.

[10] M. Bachmayr and V. Kazeev, *Stability and preconditioning of elliptic PDEs with low-rank multilevel structure*, Foundations of Computational Mathematics, (2020), https://doi.org/10.1007/s10208-020-09446-z.

[11] M. Bachmayr and R. Schneider, *Iterative methods based on soft thresholding of hierarchical tensors*, Foundations of Computational Mathematics, 17 (2017), pp. 1037–1083, https://doi.org/10.1007/s10208-016-9314-z.

[12] M. Bachmayr, R. Schneider, and A. Uschmajew, *Tensor networks and hierarchical tensors for the solution of high-dimensional partial differential equations*, Foundations of Computational Mathematics, 16 (2016), pp. 1423–1472, https://doi.org/10.1007/s10208-016-9317-9.

[13] N. Bakhvalov and G. Panasenko, *Homogenisation: Averaging Processes in Periodic Media*, vol. 36 of Mathematics and its Applications, Springer Netherlands, 1989, https://doi.org/10.1007/978-94-009-2247-1.

[14] J. Ballani and L. Grasedyck, *A projection method to solve linear systems in tensor format*, Numerical Linear Algebra with Applications, (2012), https://doi.org/10.1002/nla.1818.

[15] A. Bensoussan, J.-L. Lions, and G. Papanicolaou, *Asymptotic analysis for periodic structures*, vol. 5 of Studies in Mathematics and its Applications, North-Holland Publishing Co., Amsterdam-New York, 1978.

[16] G. Birkhoff, *Angular singularities of elliptic problems*, Journal of Approximation Theory, 6 (1972), pp. 215–230, https://doi.org/10.1016/0021-9045(72)90075-5.

[17] G. Birkhoff and R. E. Lynch, *Numerical Solution of Elliptic Problems*, Society for Industrial and Applied Mathematics, 1984, https://doi.org/10.1137/1.9781611970869.
[18] V. T. Chu and V. H. Hoang, *High dimensional finite elements for two-scale Maxwell wave equations*, Journal of Computational and Applied Mathematics, 375 (2020), p. 112756, https://doi.org/10.1016/j.cam.2020.112756.
[19] P. G. Ciarlet, *The finite element method for elliptic problems*, North-Holland, Amsterdam, New York, Oxford, 1978.
[20] D. Cioranescu, A. Damlamian, P. Donato, G. Griso, and R. Zaki, *The periodic unfolding method in domains with holes*, SIAM J. Math. Anal., 44 (2012), pp. 718–760, https://doi.org/10.1137/100817942.
[21] D. Cioranescu, A. Damlamian, and G. Griso, *The periodic unfolding method in homogenization*, SIAM Journal on Mathematical Analysis, 40 (2008), pp. 1585–1620, https://doi.org/10.1137/080713148.
[22] D. Cioranescu, A. Damlamian, and G. Griso, *The Periodic Unfolding Method in Homogenization*, GAKUTO International Series, Gakkotosho, Tokyo, Japan, 2009, pp. 1–35, http://perso-math.univ-mlv.fr/users/damlamian.alain/CAPE_TOWN-FINAL.pdf.
[23] B. Engquist and P. E. Souganidis, *Asymptotic and numerical homogenization*, Acta Numerica, 17 (2008), p. 147–190, https://doi.org/10.1017/S0962492906360011.
[24] L. Grasedyck, *Hierarchical singular value decomposition of tensors*, SIAM Journal on Matrix Analysis and Applications, 31 (2010), pp. 2029–2054, https://doi.org/10.1137/090764189.
[25] L. Grasedyck, *Polynomial approximation in Hierarchical Tucker format by vector-tensorization*, Preprint 308, Institut für Geometrie und Praktische Mathematik, RWTH Aachen, April 2010, http://www.igpm.rwth-aachen.de/Download/reports/pdf/IGPM308_k.pdf.
[26] W. Hackbusch, *Tensor Spaces and Numerical Tensor Calculus*, vol. 42 of Springer Series in Computational Mathematics, Springer, 2012, https://doi.org/10.1007/978-3-642-28027-6.
[27] W. Hackbusch and S. Kühn, *A new scheme for the tensor representation*, Journal of Fourier Analysis and Applications, 15 (2009), pp. 706–722, https://doi.org/10.1007/s00041-009-9094-9.
[28] H. Harbrecht and C. Schwab, *Sparse tensor finite elements for elliptic multiple scale problems*, Comput. Methods Appl. Mech. Engrg., 200 (2011), pp. 3100–3110, https://doi.org/10.1016/j.cma.2011.07.001.
[29] V. H. Hoang, *Sparse finite element method for periodic multiscale nonlinear monotone problems*, Multiscale Model. Simul., 7 (2008), pp. 1042–1072, https://doi.org/10.1137/070698683.
[30] V. H. Hoang and C. Schwab, *High-dimensional finite elements for elliptic problems with multiple scales*, Multiscale Model. Simul., 3 (2004/05), pp. 168–194, https://doi.org/10.1137/030601077.
[31] S. Holtz, T. Rohwedder, and R. Schneider, *The alternating linear scheme for tensor optimization in the Tensor Train format*, SIAM Journal on Scientific Computing, 34 (2012), pp. A683–A713, https://doi.org/10.1137/100818893.
[32] T. Y. Hou, *Multiscale computations for flow and transport in porous media*, in Multi-scale phenomena in complex fluids, vol. 12 of Ser. Contemp. Appl. Math. CAM, World Sci. Publishing, Singapore, 2009, pp. 175–285, https://doi.org/10.1142/9789814273268_0003.
[33] T. Y. Hou, X.-H. Wu, and Z. Cai, *Convergence of a multiscale finite element method for elliptic problems with rapidly oscillating coefficients*, Math. Comp., 68 (1999), pp. 913–943, https://doi.org/10.1090/S0025-5718-99-01077-7.
[34] V. V. Jikov, S. M. Kozlov, and O. A. Oleinik, *Homogenization of Differential Operators and Integral Functionals*, Springer-Verlag, 1994, http://www.springer.com/book/9783642846618.
[35] V. Kazeev, *Quantized tensor-structured finite elements for second-order elliptic PDEs in two dimensions*, PhD thesis, SAM, ETH Zurich, ETH Dissertation No. 23002, 2015, https://doi.org/10.3929/ethz-a-010554062.
[36] V. Kazeev, *TensorRefinement.jl, a Julia package for low-rank computations with functions*, 2018–2022, http://github.com/TensorRefinement/TensorRefinement.jl.
[37] V. Kazeev, *Low-rank tensor refinement for second-order elliptic PDEs*, in preparation, http://mat.univie.ac.at/~kazeev/pub.
[38] V. Kazeev, I. Oseledets, M. Rakhuba, and C. Schwab, *QTT-finite-element approximation for multiscale problems I: model problems in one dimension*, Advances in Computational Mathematics, 43 (2017), pp. 411–442, https://doi.org/10.1007/s10444-016-9491-y.
[39] V. Kazeev and C. Schwab, *Approximation of singularities by quantized-tensor FEM*, Proceedings in Applied Mathematics and Mechanics, 15 (2015), pp. 743–746, https://doi.org/10.1002/pamm.201510353, http://dx.doi.org/10.1002/pamm.201510353.
[40] V. Kazeev and C. Schwab, *Quantized tensor-structured finite elements for second-order elliptic PDEs in two dimensions*, Numer. Math., 138 (2018), pp. 133–190, https://doi.org/10.1007/s00211-017-0899-1.

[41] V. A. Kazeev and B. N. Khoromskij, *Low-rank explicit QTT representation of the Laplace operator and its inverse*, SIAM Journal on Matrix Analysis and Applications, 33 (2012), pp. 742–758, https://doi.org/10.1137/100820479.

[42] B. Khoromskij and S. Repin, *Rank structured approximation method for quasi-periodic elliptic problems*, Computational Methods in Applied Mathematics, 17 (2017), pp. 457–477, https://doi.org/10.1515/cmam-2017-0014.

[43] B. N. Khoromskij, $\mathcal{O}(d \log n)$*-quantics approximation of* n*-d tensors in high-dimensional numerical modeling*, Preprint 55, Max Planck Institute for Mathematics in the Sciences, September 2009, http://www.mis.mpg.de/publications/preprints/2009/prepr2009-55.html.

[44] B. N. Khoromskij, $\mathcal{O}(d \log n)$*-quantics approximation of* n*-d tensors in high-dimensional numerical modeling*, Constructive Approximation, 34 (2011), pp. 257–280, https://doi.org/10.1007/s00365-011-9131-1.

[45] B. N. Khoromskij, *Tensor Numerical Methods in Scientific Computing*, vol. 19 of Radon Series on Computational and Applied Mathematics, De Gruyter, 2018, https://doi.org/10.1515/9783110365917.

[46] B. N. Khoromskij and S. I. Repin, *A fast iteration method for solving elliptic problems with quasiperiodic coefficients*, Russian Journal of Numerical Analysis and Mathematical Modelling, 30 (2015), pp. 329–344, https://doi.org/10.1515/rnam-2015-0030.

[47] V. A. Kondrat'ev, *Boundary value problems for elliptic equations in conical regions*, Soviet Mathematics, 4 (1963), pp. 1600–1602.

[48] V. A. Kondrat'ev, *Boundary problems for elliptic equations with conical or angular points*, Transactions of the Moscow Mathematical Society, 16 (1967), pp. 209–292, http://mi.mathnet.ru/eng/mmo186.

[49] D. Kressner, M. Steinlechner, and B. Vandereycken, *Preconditioned low-rank Riemannian optimization for linear systems with tensor product structure*, SIAM Journal on Scientific Computing, 38 (2016), pp. A2018–A2044, https://doi.org/10.1137/15M1032909.

[50] D. Kressner and C. Tobler, *Low-rank tensor Krylov subspace methods for parametrized linear systems*, SIAM Journal on Matrix Analysis and Applications, 32 (2011), pp. 1288–1316, https://doi.org/10.1137/100799010.

[51] R. S. Lehman, *Developments at an analytic corner of solutions of elliptic partial differential equations*, Journal of Mathematics and Mechanics, 8 (1959), pp. 727–760, http://www.jstor.org/stable/24900685.

[52] C. Marcati, M. Rakhuba, and C. Schwab, *Tensor rank bounds for point singularities in* $\mathbb{R}^3$, Research Report 68, Seminar for Applied Mathematics, ETH Zurich, 2019, to appear in Adv. Comp. Math., http://www.sam.math.ethz.ch/reports/2019/68.

[53] A.-M. Matache and J. M. Melenk, *Two-scale regularity for homogenization problems with nonsmooth fine scale geometry*, Math. Models Methods Appl. Sci., 13 (2003), pp. 1053–1080, https://doi.org/10.1142/S0218202503002817.

[54] F. Murat and L. Tartar, *H-Convergence*, Birkhäuser Boston, Boston, MA, 1997, p. 21–43, https://doi.org/10.1007/978-1-4612-2032-9_3.

[55] G. Nguetseng, *A general convergence result for a functional related to the theory of homogenization*, SIAM Journal on Mathematical Analysis, 20 (1989), pp. 608–623, https://doi.org/10.1137/0520043.

[56] G. Nguetseng, *Homogenization structures and applications. I*, Z. Anal. Anwendungen, 22 (2003), pp. 73–107, https://doi.org/10.4171/ZAA/1133.

[57] R. Orús, *A practical introduction to tensor networks: Matrix product states and projected entangled pair states*, Annals of Physics, 349 (2014), pp. 117–158, https://doi.org/10.1016/j.aop.2014.06.013.

[58] I. Oseledets, *Approximation of matrices with logarithmic number of parameters*, Doklady Mathematics, 80 (2009), pp. 653–654, https://doi.org/10.1134/S1064562409050056.

[59] I. Oseledets and S. Dolgov, *Solution of linear systems and matrix inversion in the TT-format*, SIAM Journal on Scientific Computing, 34 (2012), pp. A2718–A2739, https://doi.org/10.1137/110833142.

[60] I. V. Oseledets, *Approximation of* $2^d \times 2^d$ *matrices using tensor decomposition*, SIAM Journal on Matrix Analysis and Applications, 31 (2010), pp. 2130–2145, https://doi.org/10.1137/090757861.

[61] I. V. Oseledets, *Tensor Train decomposition*, SIAM Journal on Scientific Computing, 33 (2011), pp. 2295–2317, https://doi.org/10.1137/090752286.

[62] I. V. Oseledets and E. E. Tyrtyshnikov, *Breaking the curse of dimensionality, or how to use SVD in many dimensions*, SIAM Journal on Scientific Computing, 31 (2009), pp. 3744–3759, https://doi.org/10.1137/090748330.

[63] S. Östlund and S. Rommer, *Thermodynamic limit of density matrix renormalization*, Phys. Rev. Lett., 75 (1995), pp. 3537–3540, https://doi.org/10.1103/PhysRevLett.75.3537.

[64] T. Rohwedder and A. Uschmajew, *Local convergence of alternating schemes for optimization of convex problems in the TT format*, Preprint 112, DFG-Schwerpunktprogramm 1324, August 2012, http://www.dfg-spp1324.de/download/preprints/preprint112.pdf.

[65] U. Schollwöck, *The density-matrix renormalization group in the age of matrix product states*, Annals of Physics, 326 (2011), pp. 96–192, https://doi.org/10.1016/j.aop.2010.09.012.

[66] C. Schwab, *High-dimensional finite elements for elliptic problems with multiple scales and stochastic data*, in Proc. Intern. Congress of Mathematicians Vol. III, Higher Education Press Beijing, 2002, pp. 727–734.

[67] W. C. Tan and V. H. Hoang, *High dimensional finite elements for time-space multiscale parabolic equations*, Advances in Computational Mathematics, 45 (2019), pp. 1291–1327, https://doi.org/10.1007/s10444-018-09657-7.

[68] E. E. Tyrtyshnikov, *Tensor approximations of matrices generated by asymptotically smooth functions*, Sbornik: Mathematics, 194 (2003), pp. 941–954, https://doi.org/10.1070/SM2003v194n06ABEH000747.

[69] A. Uschmajew and B. Vandereycken, *The geometry of algorithms using hierarchical tensors*, Linear Algebra and its Applications, 439 (2013), pp. 133–166, https://doi.org/10.1016/j.laa.2013.03.016.

[70] F. Verstraete and J. I. Cirac, *Matrix product states represent ground states faithfully*, Phys. Rev. B, 73 (2006), p. 094423, https://doi.org/10.1103/PhysRevB.73.094423.

[71] G. Vidal, *Efficient classical simulation of slightly entangled quantum computations*, Phys. Rev. Lett., 91 (2003), pp. 147902–147905, https://doi.org/10.1103/PhysRevLett.91.147902.

[72] S. R. White, *Density matrix formulation for quantum renormalization groups*, Phys. Rev. Lett., 69 (1992), pp. 2863–2866, https://doi.org/10.1103/PhysRevLett.69.2863.

[73] B. Xia and V. H. Hoang, *High dimensional finite elements for multiscale wave equations*, Multiscale Model. Simul., 12 (2014), pp. 1622–1666, https://doi.org/10.1137/120902409.

[74] B. Xia and V. H. Hoang, *High-dimensional finite element method for multiscale linear elasticity*, IMA J. Numer. Anal., 35 (2015), pp. 1277–1314, https://doi.org/10.1093/imanum/dru033.

[75] B. Xia and V. H. Hoang, *Sparse tensor finite elements for elastic wave equation with multiple scales*, J. Comput. Appl. Math., 282 (2015), pp. 179–214, https://doi.org/10.1016/j.cam.2015.01.001.