ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics SAM

QTT-Finite-Element Approximation For Multiscale Problems

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

Research Report No. 2016-06
January 2016

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

Funding ERC: FP7 Advanced Grant AdG247277
Funding: Russian Science Foundation grant 14-11-00659

QTT-Finite-Element Approximation For Multiscale Problems

Vladimir Kazeev* Ivan Oseledets†‡ Maxim Rakhuba†
Christoph Schwab§

January 25, 2016

Abstract

Tensor-compressed numerical solution of elliptic multiscale-diffusion and high frequency scattering problems is considered. For either problem class, solutions exhibit multiple length scales governed by the corresponding *scale parameter*: the scale of oscillations of the diffusion coefficient or smallest wavelength, respectively. As is well-known, this imposes a scale-resolution requirement on the number of degrees of freedom required to accurately represent the solutions in standard finite-element (FE) discretizations. Low-order FE methods are by now generally perceived unsuitable for high-frequency diffusion coefficients and high wavenumbers, and special techniques have been proposed instead (such as numerical homogenization, heterogeneous multiscale method, oversampling, etc.). They require, in some form, a-priori information on the microstructure of the solution.

We propose the use of *tensor-structured* compressed first-order FE methods for scale resolution without a-priori information. The FE methods are based on principal components dynamically extracted from the FE solution by non-linear, *quantized tensor train* (QTT) decomposition of the system matrix, load and solution vectors. For prototypical model

*Section of Mathematics, University of Geneva, 2–4 rue du Lièvre, case postale 64, 1211 Geneva, Switzerland. vladimir.kazeev@unige.ch.

†Skolkovo Institute of Science and Technology, Novaya St. 100, 143025 Skolkovo, Moscow Region, Russia. {i.oseledets,m.rakhuba}@skoltech.ru.

‡Institute of Numerical Mathematics, Russian Academy of Sciences, Gubkina St. 8, 119333, Moscow, Russia

§Seminar für Angewandte Mathematik, ETH Zurich. Rämistrasse 101, 8092 Zürich, Switzerland. christoph.schwab@sam.math.ethz.ch.

problems, we prove that this approach identifies *effective degrees of freedom* from a uniform "virtual" (i.e. never directly accessed) mesh and its corresponding degrees of freedom (whose number may be prohibitively large). Precisely, solutions of model elliptic homogenization and high frequency acoustic scattering problems are proved to admit QTT-formatted approximations whose number of effective degrees of freedom is robust in the scale parameter and polylogarithmic with respect to the reciprocal of the target Sobolev-norm accuracy $\varepsilon > 0$. No a-priori information on the nature of the problems and intrinsic length scales of the solution is required in the proposed approach.

As a corollary of our analysis, we prove that the Kolmogorov n-widths of solutions sets are exponentially small for analytic data, independent of the problems' scale parameters. That implies, in particular, *robust exponential convergence of reduced basis and MOR techniques*.

Detailed numerical experiments confirm the theoretical bounds.

Keywords: multiscale problems, Helmholtz equation, homogenization, scale resolution, exponential convergence, tensor decompositions, quantized tensor trains .

AMS Subject Classification (2000): 15A69, 35B27, 35J05, 65N15, 65N30.

1 Introduction

Tensor-structured numerical methods for differential equations (PDEs) have received increasing attention in recent years; we refer to the literature survey [15] and the references below. Their primary motivation has been the numerical solution of PDEs on high-dimensional solution and parameter spaces. Such PDEs arise, among others, in applications from quantum chemistry (electron structure calculations) [10, 33] in computational finance (pricing of derivative contracts on baskets of risky assets) [25] and in computational uncertainty quantification [34]. Recent numerical experiments and mathematical results, given in [27, 23, 28], indicate that *tensor-formatted compressed representations can provide exponential convergence of low-order finite-element or finite-difference discretizations*: they can produce numerical solutions which achieve accuracy $\varepsilon > 0$ in Sobolev spaces in polylogarithmic with respect to ε^{-1} work and memory. Analogous numerical results have recently been reported also for elliptic problems with multiple scales in [35], indicating a new computational paradigm in adaptive FE algorithms which are based on combining classical, low order FE discretizations with general tensor-compression results as presented in in [40, 32]. In the present work, we establish *scale-independent, exponential convergence rates* for QTT-structured, continuous, piecewise-linear approximations of solutions to model elliptic multiscale and Helmholtz problems in one space dimension: we prove that the number N of QTT parameters is bounded polylogarithmic with respect to ε^{-1}, *independent of the scale parameter of the solution*, where ε is the corresponding $\mathbb{H}^1$-accuracy, for either type of problem, and without any a-priori information on the type of problem.

As a consequence of our analysis, we also obtain the exponential smallness of Kolmogorov n-widths of the solution sets *independent of the scale parameters*. This implies *scale-robust exponential convergence* of the so-called reduced basis approximations.

The outline of this paper is as follows: in section 2, we introduce the model problems, their variational formulations and the standard FE discretizations. Section 3 provides a short summary of quantized FE approximations. Section 4 introduces the homogenization problem, the asymptotic analysis of its solution and provides scale-separated finite-dimensional approximations with exponential convergence rate bounds which are interesting in their own right. These bounds are subsequently used to prove logarithmic in accuracy QTT rank bounds for the quantized FE solution vectors. Section 5 is devoted to the same programme for the high frequency Helmholtz equation, where we prove QTT rank bounds which are mildly depending on the wavenumber and, again, logarithmic in accuracy. For the practical implementation of quantized tensor train FE methods, analogous rank bounds for the stiffness and mass matrices are required, and we prove this in Section 6. Finally, we provide in section 7 numerical experiments which fully confirm the scale-robust tensor rank bounds.

2 Model problems. Weak formulations. Finite-element discretizations

We shall consider model boundary-value problems for two linear, second-order differential equations in $\mathrm{D} = (0, 1)$: the *multiscale-diffusion (homogenization) equation* and the *Helmholtz equation*. The former arises for homogenization problems with separated scales, and the latter for high-frequency, time-harmonic wave propagation. Both model problems are given by equations of the form

$$-(au')' + cu = f \quad \text{in} \quad \mathrm{D}\,, \tag{2.1}$$

with suitable boundary conditions: either homogeneous Dirichlet conditions or, in the context of wave propagation, so-called "radiating" boundary conditions (see section 2.2 ahead).

In the remainder of this paper, we shall consider FE discretizations of (2.1) which are based on the standard variational formulation of (2.1). In either case, we shall consider a variational space V where we introduce a bilinear (sesquilinear) form $\mathfrak{a}\colon\ V\times V\to\mathbb{R}$ and a linear form $\mathfrak{f}\colon\ V\to\mathbb{R}$. The corresponding variational formulation of (2.1) then reads:

$$\text{Find } u\in V \text{ such that}\quad \mathfrak{a}(u,v)=\mathfrak{f}(v)\quad\text{for all}\quad v\in V. \tag{2.2}$$

The form $\mathfrak{a}$ is assumed to be continuous:

$$|\mathfrak{a}(v,w)|\le\mathfrak{a}^{\sup}\,\|v\|_{\mathbb{H}^1(\mathrm{D})}\,\|w\|_{\mathbb{H}^1(\mathrm{D})}\quad\text{for all}\quad w,v\in V, \tag{2.3}$$

where $\mathfrak{a}^{\sup}=\max\{\|a\|_{\mathbb{L}^\infty(\mathrm{D})},\|c\|_{\mathbb{L}^\infty(\mathrm{D})}\}$ is finite under the assumptions we make below separately for either case.

2.1 Homogenization Problem

For a model homogenization problem, we consider (2.1) with homogeneous Dirichlet boundary conditions,

$$u(0)=0=u(1), \tag{2.4}$$

under the following assumption on the coefficients.

Assumption 2.1. *[homogenization case] Let $\delta\in(0,1)$ and the coefficients in (2.1) be $a=a^\delta$ given by*

$$a^\delta(x)=a_1\left(\frac{x}{\delta}\right)\quad\textit{for all}\quad x\in\mathrm{D} \tag{2.5}$$

and $c=0$, where a_1 is an analytic, 1-periodic function that satisfies the condition

$$\operatorname{ess\,inf}\{a_1(y):y\in\mathrm{Y}\}=a_1^{\inf} \tag{2.6}$$

with a positive constant $a_1^{\inf}$. The source term f is analytic in $\overline{\mathrm{D}}=[0,1]$.

We emphasize the dependence of the solution on δ by denoting it with u^δ. In order to avoid technicalities, we assume that $\delta^{-1}\in\mathbb{N}$, so that the interval D is a multiple of the scaled unit cell $\delta\cdot\mathrm{Y}=(0,\delta)$.

The variational formulation of (2.1) is set in the space $V=\mathbb{H}_0^1(\mathrm{D})$ of functions satisfying homogeneous Dirichlet boundary conditions at the endpoints of D. We also introduce a bilinear form $\mathfrak{a}\colon\ V\times V\to\mathbb{R}$ and a linear form $\mathfrak{f}\colon\ V\to\mathbb{R}$ by setting

$$\mathfrak{a}(w,v)=\int_{\mathrm{D}}(aw'v'+cwv)\quad\text{and}\quad\mathfrak{f}(v)=\int_{\mathrm{D}}fv\quad\text{for all}\quad w,v\in V.$$

Assumption 2.1 implies the continuity (2.3) and coercivity of $\mathfrak{a}$, the latter meaning that there exists $C > 0$ depending on $a_1^{\inf}$ but not on δ such that

$$\mathfrak{a}(v, v) \geq C \|v\|^2_{\mathbb{H}^1(\mathrm{D})} \quad \text{for all} \quad v \in V. \tag{2.7}$$

Then, by the Lax–Milgram lemma, the variational problem (2.2) admits a unique solution for every $f \in \mathbb{H}^{-1}(\mathrm{D})$ and for every δ.

2.2 Helmholtz Problem

In the *Helmholtz case*, we work under the following assumption.

Assumption 2.2. *[Helmholtz case] For a wavenumber $k > 0$, let the coefficients a and c in* (2.1) *be given by*

$$a = 1 \quad \textit{and} \quad c = -k^2. \tag{2.8}$$

The source term f is analytic in $\overline{\mathrm{D}} = [0, 1]$ and $g \in \mathbb{R}$ is a given constant.

Under assumption 2.2, the equation (2.1) reads

$$-u'' - k^2 u = f \quad \text{in} \quad \mathrm{D}. \tag{2.9}$$

We impose the homogeneous Dirichlet boundary condition and the so-called *radiating* boundary condition at the left and right endpoints of D, respectively:

$$u(0) = 0 \quad \text{and} \quad u'(1) - \mathrm{i}ku(1) = g, \tag{2.10}$$

where $\mathrm{i} = \sqrt{-1}$ denotes the imaginary unit.

The solution u of (2.9)–2.10 is complex-valued; for a variational space, we consider $V = \{w \in \mathbb{H}^1(\mathrm{D}; \mathbb{C}) : w(0) = 0\}$. The corresponding sesquilinear and antilinear forms are defined by

$$\mathfrak{a}(w, v) = \int_{\mathrm{D}} w'\overline{v}' - k^2 \int_{\mathrm{D}} w\overline{v} - \mathrm{i}kw(1)\,\overline{v}(1) \quad \text{and} \quad \mathfrak{f}(v) = \int_{\mathrm{D}} f\overline{v} + g\overline{v}(1) \quad \text{for all} \quad w, v \in V. \tag{2.11}$$

The bilinear form $\mathfrak{a}$ satisfies the *inf-sup* condition

$$\alpha = \inf_{w \in V \setminus \{0\}} \sup_{v \in V \setminus \{0\}} \frac{|\mathfrak{a}(w, v)|}{\|w\|_V \|v\|_V} > 0$$

with $\alpha = \mathcal{O}(k^{-1})$ and $\alpha^{-1} = \mathcal{O}(k)$, see, e.g. [20, theorem 4.2]. Also,

$$\sup_{w \in V \setminus \{0\}} \frac{|\mathfrak{a}(w, v)|}{\|w\|_V \|v\|_V} > 0 \quad \text{for all} \quad v \in V \setminus \{0\}.$$

Then (2.2) admits a unique solution ([2, Theorem 2.1]).

We assume given $\{V_\ell\}_{\ell=0}^{\infty}$ a sequence of finite-dimensional spaces $V_\ell \subset V$, $\ell \in \mathbb{N}$, which is dense in V,

$$\lim_{\ell\to\infty} \inf_{v_\ell \in V_\ell} \|v - v_\ell\|_V = 0 \quad \text{for all} \quad v \in V, \tag{2.12}$$

and *inf-sup* stable for the variational form (2.2):

$$\alpha_\ell = \inf_{w_\ell \in V_\ell \setminus \{0\}} \sup_{v_\ell \in V_\ell \setminus \{0\}} \frac{|\mathfrak{a}(w_\ell, v_\ell)|}{\|w_\ell\|_V \|v_\ell\|_V} > 0 \quad \text{for all} \quad \ell \in \mathbb{N}$$

and

$$\sup_{w_\ell \in V \setminus \{0\}} \frac{|\mathfrak{a}(w_\ell, v_\ell)|}{\|w_\ell\|_V \|v_\ell\|_V} > 0 \quad \text{for all} \quad v_\ell \in V_\ell \setminus \{0\}.$$

Then, by [2, Theorem 2.2], for the FE approximations $u_\ell \in V_\ell$, $\ell \geq L_0$, of the solution u of (2.2), which are given by

$$\mathfrak{a}(u_\ell, v_\ell) = \mathfrak{f}(v_\ell) \qquad \text{for all} \quad v_\ell \in V_\ell, \tag{2.13}$$

there holds the following quasi-optimality bound:

$$\|u - u_\ell\|_{\mathbb{H}^1(\mathrm{D})} \leq C_1 \inf_{w^\ell \in V_\ell} \|u - w^\ell\|_{\mathbb{H}^1(\mathrm{D})}, \tag{2.14}$$

where $C_1 = 1 + \mathfrak{a}^{\sup}/\alpha_\ell$.

3 QTT-structured finite-element discretization

3.1 FE spaces: nodal bases, parametrization

We consider the coefficient field $\mathbb{F}$ being either $\mathbb{R}$ or $\mathbb{C}$. We describe the FE spaces V_l, $l \in \mathbb{N}$, that we use below to discretize the solutions of problems of the form (2.2). We shall use $\mu_1, \mu_2 \in \{0, 1\}$ to encode the essential boundary conditions imposed by the construction of the FE spaces: $\mu_1 = 0$ and $\mu_1 = 1$ shall denote that the values at the left endpoint of D are *fixed* and *not fixed* respectively, and μ_2 shall indicate the same for the right endpoint of D.

For every $l \in \mathbb{N}$, we set $n_l = 2^l - \mu_1 - \mu_2$ and $h_l = (n_l + 1)^{-1}$. Then we consider the uniform partition $\mathcal{T}^l$ of D induced by the nodes

$$t_i^l = (i + 1 - \mu_1) h_l \quad \text{with} \quad i \in \mathcal{J}^l, \tag{3.1}$$

where $\mathcal{J}^l = \{\mu_1 - 1, \ldots, 2^l - \mu_2\}$. The number of interior nodes is n_l and the grid size is h_l. For the construction of FE spaces that follows, we call the nodes indexed by $\mathcal{I}_0^l = \{0, \ldots, 2^l - 1\}$ *active*. The boundary nodes $t^l_{\mu_1 - 1} = 0$ and $t^l_{2^l - \mu_2} = 1$ are active for Neumann boundary conditions, i.e., if and only if $\mu_1 = 1$ and $\mu_2 = 1$ respectively hold.

For all $l \in \mathbb{N}$ and $i \in \mathcal{I}^l$, we define $\phi_i^l \in \mathbf{C}(\mathrm{D})$ by requiring linearity on each interval (t^l_{j-1}, t^l_j), $j = \mu_1, \ldots, 2^l - \mu_2$, and the following interpolation condition: $\phi_i^l\big(t^l_{i'}\big) = \delta_{ii'}$ for all $i' \in \mathcal{I}^l$. We define the corresponding FE spaces of continuous piecewise-linear functions over $\mathbb{F}$,

$$\begin{aligned} \mathrm{S}^1(\mathrm{D}, \mathcal{T}^l) &= \operatorname{span}\{\phi_i^l : \; i \in \mathcal{I}^l\} \subset \mathbf{C}(\overline{\mathrm{D}}), \\ \mathrm{S}^1_0(\mathrm{D}, \mathcal{T}^l) &= \operatorname{span}\{\phi_i^l : \; i \in \mathcal{I}^l_0\} \subset \mathbf{C}(\overline{\mathrm{D}}), \end{aligned} \tag{3.2}$$

where the spans are taken with respect to the field $\mathbb{F}$.

For every $u \in \mathbf{C}(\overline{\mathrm{D}})$, we consider the following *set of admissible approximations*:

$$\mathscr{F}_u^l = \mathrm{S}^1_0(\mathrm{D}, \mathcal{T}^l) + \{\mu_1 = 0\}\, u(0) \phi^l_{-1} + \{\mu_2 = 0\}\, u(1) \phi^l_{2^l} \;\subset\; \mathrm{S}^1(\mathrm{D}, \mathcal{T}^l), \tag{3.3}$$

where the boundary terms enter only if the logical expressions enclosed in the corresponding curly brackets are true. We parametrize it by vectors in $\mathbb{F}^{2^l}$ using *analysis* and *synthesis* operators $\mathscr{A}^l \colon \mathbf{C}(\overline{\mathrm{D}}) \to \mathbb{F}^{2^l}$ and $\mathscr{S}^l \colon \mathbb{F}^{2^l} \to \mathrm{S}^1_0(\mathrm{D}, \mathcal{T}^l)$:

$$\begin{aligned} \mathbb{F}^{2^l} \ni \boldsymbol{u}^l = \mathscr{A}^l u \;\leftrightarrow\; u^l = \mathscr{S}^l \boldsymbol{u}^l & \\ + \{\mu_1 = 0\}\, u(0) \phi^l_{-1} + \{\mu_2 = 0\}\, u(1) \phi^l_{2^l} &\in \mathscr{F}_u^l, \end{aligned} \tag{3.4}$$

where

$$(\mathscr{A}^u)_i = u(t_i^l) \quad \text{for all} \quad i \in \mathcal{I}^l_0 \quad \text{and} \quad \mathscr{S}^l \boldsymbol{u}^l = \sum_{i \in \mathcal{I}^l_0} \boldsymbol{u}^l_i \, \phi^l_i. \tag{3.5}$$

In the present paper, we are concerned with FE approximations $u^l \in \mathscr{F}_u^l$ that are *QTT-structured*, i.e. are such that the corresponding coefficient vectors $\boldsymbol{u}^l$ are represented in the QTT format defined in sections 3.2–3.3.

3.2 Tensor Train (TT) representation

We use the *tensor train* (*TT* for short) decomposition, a non-linear, low-parametric approximate representation of multidimensional arrays based on the separation of variables, developed by Oseledets and Tyrtyshnikov [44, 43].

By a d-dimensional $n_1 \times \ldots \times n_d$*-vector* we mean an array indexed by d indices, the range of the kth index being $0, 1, \ldots, n_k - 1$ for $1 \leq k \leq d$. In the literature such arrays are often called *tensors*. However, we distinguish multidimensional *vectors* and *matrices*, which represent functions and operators on spaces of those. We do so due to the difference in their nature and irrespective of the number of mode indices, denoted with d right above, since the latter is subject to change under *quantization* described in Section 3.3.

Let us consider a d-dimensional $n_1 \times \ldots \times n_d$-vector $\boldsymbol{u}$. In the case that, for certain two- and three-dimensional arrays $U_1, U_2, \ldots, U_d$, it admits the representation

$$\begin{aligned}\boldsymbol{u}_{j_1,\ldots,j_d} = \sum_{\alpha_1=1}^{r_1} \ldots \sum_{\alpha_{d-1}=1}^{r_{d-1}} & U_1(j_1, \alpha_1) \\ & \cdot\; U_2(\alpha_1, j_2, \alpha_2) \cdot \ldots \cdot U_{d-1}(\alpha_{d-2}, j_{d-1}, \alpha_{d-1}) \cdot U_d(\alpha_{d-1}, j_d) \quad (3.6)\end{aligned}$$

for $0 \leq j_k \leq n_k - 1$, where $1 \leq k \leq d$, we say that $\boldsymbol{u}$ is represented in the *tensor-train* (TT) decomposition in terms of the core tensors $U_1, U_2, \ldots, U_d$. The summation indices $\alpha_1, \ldots, \alpha_{d-1}$ and limits $r_1, \ldots, r_{d-1}$ on the right-hand side of (3.6) are called, respectively, *rank indices* and *ranks* of the representation. A TT decomposition, exact or approximate, can be constructed via the low-rank representation of a sequence of single matrices; for example, via the SVD. In particular, for every $k = 1, \ldots, d-1$ the representation (3.6) implies a rank-r_k factorization of an *unfolding matrix* $\boldsymbol{U}^{(k)}$ with the entries

$$\boldsymbol{U}^{(k)}_{\frac{\overline{j_1,\ldots,j_k}}{\overline{j_{k+1},\ldots,j_d}}} = \boldsymbol{u}_{j_1,\ldots,j_k,j_{k+1},\ldots,j_d}.$$

Here, the overscore denotes the unfolding of a multi-index into a long scalar index:

$$\overline{j_1, \ldots, j_k} = \sum_{m=1}^{k} j_m \prod_{\ell=m+1}^{k} n_\ell \quad (3.7)$$

for the row index, and similarly for the column index. This renders $\boldsymbol{U}^{(k)}$ a matrix with two “long” indices.

Conversely, if the vector $\boldsymbol{u}$ is such that the unfolding matrices $\boldsymbol{U}^{(1)}, \ldots, \boldsymbol{U}^{(d-1)}$ are of ranks $r_1, \ldots, r_{d-1}$ respectively, then the cores $U_1, U_2, \ldots, U_d$ satisfying (3.6) exist; see Theorem 2.1 in [43]. The ranks of the unfolding matrices are the lowest possible ranks of a TT decomposition of the vector. They are hence referred to as *TT ranks of the vector.*

Another, fundamental, property of the TT representation is that if the unfolding matrices can be approximated with ranks $r_1, \ldots, r_{d-1}$ and accuracies $\varepsilon_1, \ldots, \varepsilon_{d-1}$ in the Frobenius norm, then the vector itself can be approximated in the TT format with ranks $r_1, \ldots, r_{d-1}$ and accuracy $\sqrt{\sum_{k=1}^{d-1} \varepsilon_k^2}$ in the ℓ_2-norm. This underlies a robust and efficient algorithm for the low-rank TT approximation of vectors given in full format or in the TT format with higher ranks. For details see Theorem 2.2 with corollaries and Algorithms 1 and 2 in [43]. In practice it may be essential that the TT representation relies on a certain ordering of the dimensions and *reordering dimensions may affect the numerical values of the TT ranks significantly.*

The multiplication of a vector given in the TT decomposition (3.6) by a d-dimensional $(m_1 \times \ldots \times m_d) \times (n_1 \times \ldots \times n_d)$-matrix $\boldsymbol{A}$ can be performed efficiently if the matrix is represented as follows:

$$\boldsymbol{A}_{\substack{i_1,\ldots,i_d \\ j_1,\ldots,j_d}} = \sum_{\alpha_1=1}^{r_1} \cdots \sum_{\alpha_{d-1}=1}^{r_{d-1}} V_1(i_1, j_1, \alpha_1) \cdot V_2(\alpha_1, i_2, j_2, \alpha_2) \cdot \ldots \cdot V_{d-1}(\alpha_{d-2}, i_{d-1}, j_{d-1}, \alpha_{d-1}) \cdot V_d(\alpha_{d-1}, i_d, j_d) \,. \quad (3.8)$$

The decomposition (3.8) is recognized as *a TT representation of the matrix* $\boldsymbol{A}$; the TT cores $V_1, \ldots, V_d$ are now three- and four-dimensional arrays. The discussion of the efficiency and robustness of the TT decomposition of vectors also applies to the matrix case. Indeed, (3.8) can be interpreted as TT decomposition of a vectorization of $\boldsymbol{A}$, in which the corresponding row and column indices are merged to obtain a d-dimensional $m_1 \cdot n_1 \times \ldots \times m_d \cdot n_d$-vector.

Basic operations of linear algebra with vectors and matrices in the TT format, such as addition, Hadamard and dot products, multi-dimensional contraction, matrix-vector multiplication, etc. are considered in detail in [43]. The use of tensor-structured approximations aims primarily at reducing the complexity of computations and lifting the curse of dimensionality. The TT format achieves this with the storage cost and complexity of basic operations of the TT arithmetic being bounded by dnr^α with $\alpha \in \{2, 3\}$, where $n \geq n_1, \ldots, n_d$ and $r \geq r_1, \ldots, r_{d-1}$. This estimate is formally linear in d; however, the TT ranks $r_1, \ldots, r_{d-1}$ in (3.6) may depend on d and n.

So far there has been increasing, mostly experimental, evidence that in many applications the TT and QTT ranks are moderate, e.g., respectively, at most linear with respect to d and at most polynomial with respect to l, which is crucial for the applicability of TT- and QTT-structured methods. For examples see the papers [19, 5, 9, 11, 24, 29, 36], the extensive survey [15] and later works [3, 1, 37].

3.3 Quantized Tensor Train (QTT) representation

With the aim of further complexity reduction the TT format can be applied to a *quantized* tensor, which leads to the *quantized tensor train (QTT)* format [40, 32, 42]. The use of *quantization* in the context of tensor representations, which dates back to [49], consists in "folding" the vector (matrix) by introducing l_k "virtual" dimensions (levels) corresponding to the k-th "physical" dimension and separating all the dimensions, "physical" and "virtual", just as the former, in a tensor decomposition. For example, if $n_k = 2^l$ with $l \in \mathbb{N}$, the "virtual" dimensions corresponding to a "physical" dimension may correspond the bits of its binary coding:

$$j_k = \overline{j_{k,1}, \ldots, j_{k,l}} = \sum_{q=1}^{l} 2^{q-1} \, j_{k,q} \quad \leftrightarrow \quad (j_{k,1}, \ldots, j_{k,l}) \tag{3.9}$$

with $j_{k,q} \in \{0,1\}$ for every $q = 1, \ldots, l$, cf. (3.7). Binary quantization, which is associated with index transformations of the form (3.9), is central for the remainder of the present paper.

By a *QTT decomposition* of a vector and the *QTT ranks of the decomposition* we mean a TT decomposition of its quantization and the ranks of that TT decomposition. In particular (3.6) and (3.8), with d being replaced with l, also present QTT representations of ranks $r_1, \ldots, r_{l-1}$ of a one-dimensional vector $\boldsymbol{v}$ and of a one-dimensional matrix $\boldsymbol{B}$ with entries $\boldsymbol{v}_{\overline{j_1,\ldots,j_l}} = \boldsymbol{u}_{j_1,\ldots,j_l}$ and $\boldsymbol{B}_{\substack{\overline{i_1,\ldots,i_l} \\ \overline{j_1,\ldots,j_l}}} = \boldsymbol{A}_{\substack{i_1,\ldots,i_l \\ j_1,\ldots,j_l}}$ respectively. In the present paper, for 2^l-component vectors, we shall work with representations of the form

$$\begin{aligned} \boldsymbol{v}_{\overline{j_1,\ldots,j_l}} = \sum_{\alpha_1=1}^{r_1} \ldots \sum_{\alpha_{l-1}=1}^{r_{l-1}} & U_1(j_1, \alpha_1) \\ & \cdot U_2(\alpha_1, j_2, \alpha_2) \cdot \ldots \cdot U_{l-1}(\alpha_{l-2}, j_{l-1}, \alpha_{l-1}) \cdot U_l(\alpha_{l-1}, j_l), \end{aligned} \tag{3.10}$$

cf. (3.6). A representation of the form (3.10) allows to parametrize the vector $\boldsymbol{v}$ by N_l parameters instead of 2^l entries, where

$$N_l = 2r_1 + \sum_{k=2}^{l-1} 2 r_{k-1} r_k + 2 r_{l-1} \leq 2 l R_l^2 \tag{3.11}$$

with $R_l = \max\{r_1, \ldots, r_{l-1}\}$. Note that $R_l = \mathcal{O}(l^\theta)$ with a positive θ implies $N_l = \mathcal{O}(l^{2\theta+1})$.

The structure of basic operation in the TT format and related algorithms, referred to in section 3.2, when applied to quantized vectors, naturally yield the same in the QTT format. Compared to the TT representation, the QTT format seeks to resolve the multilevel structure of the data by splitting the "virtual" dimensions, introduced by quantization and representing the hierarchy of scales.

Example 3.1 (Proposition 1.1 in [32]). *To demonstrate how the quantization reduces complexity of structured data, let us consider the one-dimensional vector* $\boldsymbol{u} = (1, q, \ldots, q^{2^l-1})^\top$ *. This vector has a single "physical" dimension, and its elementwise representation requires storing* 2^l *parameters. However, if we apply the quantization transformation as described above to split the single dimension into* l *virtual levels,* $\boldsymbol{u}$ *is transformed into an* l*-dimensional vector that exhibits a low-parametric structure. Indeed, in terms of the "virtual" indices it is a rank-one Kronecker product of* l *vectors with* 2 *components each:*

$$\boldsymbol{u} = \begin{pmatrix} 1 \\ q^{2^{l-1}} \end{pmatrix} \otimes \begin{pmatrix} 1 \\ q^{2^{l-2}} \end{pmatrix} \otimes \ldots \otimes \begin{pmatrix} 1 \\ q \end{pmatrix},$$

which implies a QTT decomposition of $\boldsymbol{u}$ *with ranks* $1, \ldots, 1$. *Other explicit low-rank examples can be found in [45, 30, 31, 26].*

Note that the *Hierarchical Tensor Representation* [16, 13] itself and combined with *tensorization* [14], a comprehensive exposition of which is given in [17], are closely related counterparts of the TT and QTT formats respectively. Also, the TT representation, in fact, is known as *matrix product states (MPS)* and has been exploited by physicists to describe quantum spin systems theoretically and numerically for at least two decades now [52, 50, 51].

3.4 Basic results on QTT-structured approximation

Proposition 3.2 (Sections 4.1 and 4.2 in [43]). *Assume that* $\boldsymbol{u}$ *and* $\boldsymbol{v}$ *are* d*-dimensional vectors of equal mode sizes, given in TT representations of ranks* $p_1, \ldots, p_{d-1}$ *and* $q_1, \ldots, q_{d-1}$ *respectively. Then for all* $\alpha, \beta \in \mathbb{R}$ *the linear combination* $\alpha\boldsymbol{u} + \beta\boldsymbol{v}$ *has a TT decomposition of ranks* $p_1 + q_1, \ldots, p_{d-1} + q_{d-1}$. *On the other hand, the Hadamard product* $\boldsymbol{u} \odot \boldsymbol{v}$ *can be represented in the TT format with ranks* $p_1 q_1, \ldots, p_{d-1} q_{d-1}$.

Proposition 3.3. *If a vector* $\boldsymbol{u}$ *is given in the TT format with certain ranks, then its diagonalization* $\operatorname{diag} \boldsymbol{u}$*, which is the square matrix whose diagonal is* $\boldsymbol{u}$*, can be represented in the TT format* with the same ranks.

Proof. It is enough to note that under the diagonalization of a vector each TT core is diagonalized with respect to the corresponding mode index. □

Proposition 3.4 (Section 4.3 in [43]). *If matrices* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *have TT representations of ranks* $p_1, \ldots, p_{d-1}$ *and* $q_1, \ldots, q_{d-1}$ *respectively, their product* $\boldsymbol{AB}$*, when it is defined, has a TT representation of ranks* $p_1 q_1, \ldots, p_{d-1} q_{d-1}$.

In the QTT approximation of polynomials, we rely on the following result owing to either of [14, Corollary 13] and [45, Theorem 6].

Proposition 3.5 (QTT structure of a polynomial)**.** *Let $l \in \mathbb{N}$. Assume that a 2^l-component vector $\boldsymbol{u}$ is given by $\boldsymbol{u}_j = P(j)$ for $0 \le j \le 2^l - 1$, where P is a univariate algebraic polynomial of degree $p \in \mathbb{N}$. Then $\boldsymbol{u}$ can be represented in the QTT format with ranks bounded from above by $p+1$.*

A similar result for trigonometric polynomials can be based on the following statement.

Proposition 3.6 (QTT structure of a real-valued trigonometric polynomial)**.** *Let $q, l \in \mathbb{N}$ and $a_k, b_k, \omega_k, \varphi_k \in \mathbb{R}$ for $1 \le k \le q$. Assume that a 2^l-component vector $\boldsymbol{v}$ is given by $\boldsymbol{v}_j = \sum_{k=1}^{q} (a_k \cos(\omega_k\, j + \varphi_k) + b_k \sin(\omega_k\, j + \varphi_k))$ for $j = 0, \ldots, 2^l - 1$. Then $\boldsymbol{v}$ can be represented in the QTT format with ranks bounded from above by $2q$.*

Proof. The proof is analogous to that for the sine function, see [32, Lemma 2.5 B]. Indeed, for $1 \le k \le q$, we may write

$$a_k \cos(\omega_k j + \varphi_k) + b_k \sin(\omega_k j + \varphi_k) = \sqrt{a_k^2 + b_k^2}\ \sin(\omega_k j + \varphi_k + \psi_k),$$

where $\psi_k \in [0, 2\pi)$ is such that $a_k/\sqrt{a_k^2 + b_k^2} = \sin \psi_k$ and $b_k/\sqrt{a_k^2 + b_k^2} = \cos \psi_k$. The claim therefore follows from (a slight generalization of) the proof of [32, Lemma 2.5 B] to the case when the arguments of the sine functions are shifted.

An alternative proof, with a more constructive flavor, can be obtained by a similar generalization of [45, Theorem 7] . Then the statement for a sum of vectors that we recapitulate in Proposition 3.2 proves the claim. Note that this argument, unlike the first one, yields immediately an explicit QTT decomposition of $\boldsymbol{v}$. □

4 QTT-FEM in numerical homogenization

4.1 Asymptotic Expansion of u as $\delta \downarrow 0$

We first discuss some classical asymptotic results from homogenization theory which describe precisely the asymptotic structure of the solution u of (2.1), (2.4), (2.6) under assumption 2.1 as $\delta \downarrow 0$. For our purposes, it will be important to have explicit asymptotic expansions and remainder estimates for their truncation. Once this has been established, we use the quasi-optimality (2.14) of the FE solution and the asymptotic expansion to obtain QTT rank bounds of tensor structured approximations of the FE solution: the QTT ranks of certain finite-element approximations achieving accuracy $\varepsilon = \mathcal{O}(\sqrt{\delta})$ in the $\mathbb{H}^1(\mathrm{D})$ norm are moderate, namely $\mathcal{O}(\log^2 \delta^{-1})$ with a constant *independent of the microscale parameter* δ.

First, let us recall classical results on homogenization, from [4, 21]. Let $w_1 \in \mathbf{C}^2(\mathrm{Y}) \cap \mathbf{C}(\overline{\mathrm{Y}})$ solve the problem

$$\big(a_1(y)(w_1'(y)+1)\big)' = 0 \quad \text{for all} \quad y \in \mathrm{Y} \quad \text{and} \quad \int_{\mathrm{Y}} w_1 = 0, \tag{4.1}$$

which is uniquely solvable under assumption 2.1. The term “limiting solution” $u_0 \in \mathbf{C}^2(\mathrm{D}) \cap \mathbf{C}(\overline{\mathrm{D}})$, shall denote the solution of the *homogenized equation*:

$$-\left(\int_{\mathrm{Y}} \frac{\mathrm{d}y}{a_1(y)}\right)^{-1} u_0''(x) = f(x) \quad \text{for all} \quad x \in \mathrm{D}, \quad u_0(0) = 0 = u_0(1), \tag{4.2}$$

where $f \in \mathbf{C}(\overline{\mathrm{D}})$. Under assumption 2.1, the homogenized equation (4.2) is uniquely solvable.

Classical homogenization techniques treat the problem (2.1), (2.4), (2.5), (2.6) by expanding the solution with respect to powers of δ, see [4, 21]. We are interested in the first-order term, or *first-order corrector*, which reflects oscillations in the solution induced by those in the coefficient and is given by

$$u_1(x,y) = u_0'(x)\, w_1(y) \quad \text{for all} \quad (x,y) \in \mathrm{D} \times \mathrm{Y}. \tag{4.3}$$

One may consider a corresponding first-order expansion of the form

$$U_1^\delta(x) = u_0(x) + \delta\sigma\big(\delta^{-1}\rho(x)\big)\, u_1\Big(x, \frac{x}{\delta}\Big) \quad \text{for all} \quad x \in \mathrm{D}, \tag{4.4}$$

where $\rho(x) = \min\{x, 1-x\}$, for every $x \in \mathrm{D}$, is the distance from x to $\partial\mathrm{D}$ and σ satisfies the following assumption 4.1.

Assumption 4.1. *$\sigma \in \mathbf{C}^1[0,\infty)$ is such that $\sigma(0) = 0$, $0 \le \sigma(t) \le 1$ for all $t \in (0,1)$ and $\sigma(t) = 1$ for all $t \ge 1$.*

The first-order approximation U_1^δ satisfies, due to the factor involving σ, the same boundary conditions as the solutions of the original problem and of the homogenized equation (4.2). Furthermore, it also satisfies the following error bound, see, e.g. [4, chapter 2, §3 and chapter 4, §1, theorem 2] or [21, chapter 1, section 1.4].

Lemma 4.2. *Let assumption 2.1 hold and σ satisfy assumption 4.1. Assume also that the right-hand side f of (2.1) is analytic in $\overline{\mathrm{D}} = [0,1]$. Then there exists a positive constant c such that for every δ with $\delta^{-1} \in \mathbb{N}$, the approximation U_1^δ given by (4.4) satisfies*

$$\|u^\delta - U_1^\delta\|_{\mathbb{H}^1(\mathrm{D})} \le c\sqrt{\delta}. \tag{4.5}$$

Below, we shall use U_1^δ of (4.4) with σ given by

$$\sigma(t) = t(2-t) \quad \text{for all} \quad t \in [0,1], \qquad \sigma(t) = 1 \quad \text{for all} \quad t \in (1,\infty), \tag{4.6}$$

satisfying assumption 4.1.

4.2 Polynomial approximation in D

First, we shall consider approximation in a reference interval $\hat{\mathrm{J}} = (-1,1)$.

For $i \in \mathbb{N}_0$, by L_i we denote the ith Legendre polynomial with the standard normalization: $L_i(1) = 1$ and $\langle L_i, L_i \rangle_{\mathbb{L}^2(\hat{\mathrm{J}})} = (i+\frac{1}{2})^{-1}$.

Definition 4.3. *For every $p \in \mathbb{N}$, we define a quasi-interpolation operator $\hat{\pi}_p \colon \mathbb{H}^1(\hat{\mathrm{J}}) \to \mathbb{P}_p$ by setting*

$$\hat{u}(-1) = \hat{\pi}_p \hat{u}(-1) \quad \textit{and} \quad (\hat{\pi}_p \hat{u})' = \sum_{i=0}^{p-1} c_i L_i$$

for every $\hat{u} \in \mathbb{H}^1(\hat{\mathrm{J}})$, where $c_i = (i+\frac{1}{2}) \langle \hat{u}', L_i \rangle_{\mathbb{L}^2(\hat{\mathrm{J}})}$ for $i = 0, 1, \ldots, p-1$.

For every $p \in \mathbb{N}$, the quasi-interpolation operator $\hat{\pi}_p$ is continuous. Also, by [48, theorem 3.14] or [6, lemma 5], we have the following property

Proposition 4.4. *For every $p \in \mathbb{N}$, $\hat{\pi}_p$ is nodally exact:*

$$\forall \hat{u} \in \mathbb{H}^1(\hat{\mathrm{J}}) : \quad \hat{u}(\pm 1) = \hat{\pi}_p \hat{u}(\pm 1) \, .$$

We shall use the following accuracy and stability bounds.

Proposition 4.5. *Assume that $p \in \mathbb{N}$ and $s \in \mathbb{N}_0$ are such that $s \le p$. Then, for any function $\hat{u} \in \mathbb{H}^{s+1}(\hat{\mathrm{J}})$, the interpolant $\hat{\pi}_p \hat{u}$ satisfies*

$$\begin{aligned}
|\hat{u} - \hat{\pi}_p \hat{u}|^2_{\mathbb{H}^1(\hat{\mathrm{J}})} &\le \frac{(p-s)!}{(p+s)!} \, |\hat{u}|^2_{\mathbb{H}^{s+1}(\hat{\mathrm{J}})}, \\
\|\hat{u} - \hat{\pi}_p \hat{u}\|^2_{\mathbb{L}^2(\hat{\mathrm{J}})} &\le \frac{1}{p\,(p+1)} \frac{(p-s)!}{(p+s)!} \, |\hat{u}|^2_{\mathbb{H}^{s+1}(\hat{\mathrm{J}})}, \\
|\hat{\pi}_p \hat{u}|^2_{\mathbb{H}^2(\hat{\mathrm{J}})} &\le \frac{1}{2}(p^2-1) |\hat{u}|^2_{\mathbb{H}^2(\hat{\mathrm{J}})}.
\end{aligned} \tag{4.7}$$

Proof. A proof of the accuracy bounds can be found, for example, in [48, Corollary 3.15] or [6, Corollary 2]. The stability bound is shown in [28, lemma A-2.2]. $\square$

Lemma 4.6. *For all $\varrho \geq 1$ and $p \in \mathbb{N}$ such that $p \geq \varrho$, we set $s = \lfloor p/\varrho \rfloor$, so that $1 \leq s \leq p$. Then there holds, with $c^2 = e^5/\sqrt{2\pi}$,*

$$(\varrho - 1)^{2s}(s!)^2 \,\frac{(p-s)!}{(p+s)!} \leq c^2\, s\, \exp\!\left(-\frac{2p}{\varrho}\right).$$

Proof. For a proof, see, e.g. [28, lemma A-1.2]. □

Proposition 4.7. *Given a function $\hat{u}$ which is analytic in $\bar{\hat{\mathrm{J}}}$, there exist positive constants b, C and C_m with $m = 0, 1, 2$ such that for every $p \in \mathbb{N}$ the interpolants $\hat{\pi}_p \hat{u}$ satisfy the error bound*

$$\|\hat{u} - \hat{\pi}_p \hat{u}\|_{\mathbb{H}^1(\hat{\mathrm{J}})} \leq C \exp(-bp)$$

and the stability bounds

$$|\hat{\pi}_p \hat{u}|_{\mathbb{H}^m(\hat{\mathrm{J}})} \leq C_m \quad \text{for either} \quad m = 0, 1, \qquad |\hat{\pi}_p \hat{u}|_{\mathbb{H}^2(\hat{\mathrm{J}})} \leq C_2\, p.$$

Proof. Since $\hat{u}$ is analytic in $\hat{\mathrm{J}}$, there exist positive constants M and d such that the inequality

$$\left\|\hat{u}^{(s+1)}\right\|_{\mathbb{L}^\infty(\hat{\mathrm{J}})} \leq M\, d^{s+1}\,(s+1)! \tag{4.8}$$

holds for all $s \in \mathbb{N}$. Consider $p \in \mathbb{N}$ and $\varrho = 1 + d$, and choose $s = \lfloor p/\varrho \rfloor$. Using proposition 4.5, Lemma 4.6 and (4.8), we obtain

$$\begin{aligned}
\|\hat{u} - \hat{\pi}_p \hat{u}\|^2_{\mathbb{H}^1(\hat{\mathrm{J}})} &= \|\hat{u} - \hat{\pi}_p \hat{u}\|^2_{\mathbb{L}^2(\hat{\mathrm{J}})} + |\hat{u} - \hat{\pi}_p \hat{u}|^2_{\mathbb{H}^1(\hat{\mathrm{J}})} \\
&\leq M^2 d^2 \left\{\frac{1}{p(p+1)} + 1\right\} (s+1)^2\, (\varrho - 1)^{2s}\, (s!)^2\, \frac{(p-s)!}{(p+s)!} \\
&\leq M^2 d^2 \left\{\frac{1}{p(p+1)} + 1\right\} (s+1)^2\, c^2\, s\, \exp(-2p/\varrho) \leq C^2 p^3 \exp(-2bp)
\end{aligned}$$

with $b = (1+d)^{-1}$ and a positive constant C independent of p.

The $\mathbb{H}^m$-stability bound with $m = 2$ follows immediately from proposition 4.5, with $m = 1$, from definition 4.9, with $m = 0$, from the triangle inequality and the $\mathbb{L}^2$-error bound. □

The following bound is a particular case of [22, theorem 1].

Proposition 4.8. *Let $p \in \mathbb{N}_0$. Then every algebraic polynomial $u \in \mathbb{P}_p$ satisfies the bound*

$$\|u\|_{\mathbb{L}^\infty(\hat{\mathrm{J}})} \leq \frac{1}{\sqrt{2}}\,(p+1)\,\|u\|_{\mathbb{L}^2(\hat{\mathrm{J}})}.$$

Rescaling yields polynomial approximations on the interval D.

Definition 4.9. *For every $p \in \mathbb{N}$, we define a quasi-interpolation operator $\pi_p \colon \mathbb{H}^1(\mathrm{D}) \to \mathbb{P}_p$ through $\hat{\pi}_p$ given by definition 4.3 by rescaling from* D *to* $\hat{\mathrm{J}}$*:*

$$(\pi_p u) \circ \varphi = \hat{\pi}_p (u \circ \varphi) \quad \textit{in} \quad \mathbb{P}_p \quad \textit{for all} \quad u \in \mathbb{H}^1(\mathrm{D}),$$

where $\varphi(t) = (t+1)/2$ for all $t \in \hat{\mathrm{J}}$.

4.3 Fourier approximation in Y

For every $k \in \mathbb{Z}$, we consider $T_k \colon \mathbb{C} \to \mathbb{C}$ given by $T_k(z) = \exp(2\pi \mathrm{i} k z)$ for all $z \in \mathbb{C}$.

The functions T_k, $k \in \mathbb{Z}$, form an orthonormal basis in $\mathbb{L}^2(\mathrm{Y})$, with $\mathrm{Y} = (0,1)$. For every $w \in \mathbb{L}^2(\mathrm{Y})$,

$$w = \sum_{k \in \mathbb{Z}} c_k T_k \quad \text{in} \quad \mathbb{L}^2(\mathrm{Y}), \tag{4.9}$$

where the Fourier coefficients are given by

$$c_k = \langle w, T_k \rangle_{\mathbb{L}^2(\mathrm{Y})} = \int_{\mathrm{Y}} w \, \overline{T_k} \quad \text{for all} \quad k \in \mathbb{Z}. \tag{4.10}$$

In particular, if w is real-valued, we have $c_k = \overline{c_{-k}}$ for all $k \in \mathbb{N}$ and

$$w = c_0 + \sum_{k \in \mathbb{N}} \{ a_k \, \mathsf{Re}\, T_k + b_k \, \mathsf{Im}\, T_k \} \quad \text{in} \quad \mathbb{L}^2(\mathrm{Y}), \tag{4.11}$$

where $\mathsf{Re}\, T_k$ and $\mathsf{Im}\, T_k$ are cosine and sine functions for each $k \in \mathbb{N}$ and

$$a_k = 2\, \mathsf{Re}\, c_k = \overline{c_{-k}} + c_k, \quad b_k = -2\, \mathsf{Im}\, c_k = \overline{c_{-k}} - c_k \quad \text{for all} \quad k \in \mathbb{N}\,. \tag{4.12}$$

For every $q \in \mathbb{N}$, we set $\mathbb{T}_q = \operatorname{span}\{T_k\}_{k=-q}^{q}$, where the span is taken with respect to the field $\mathbb{C}$.

Definition 4.10. *For every $w \in \mathbb{L}^2(\hat{\mathrm{J}})$ and $q \in \mathbb{N}$, we define a projection operator $\tau_q \colon \mathbb{L}^2(\mathrm{Y}) \to \mathbb{T}_q$ by setting*

$$\tau_q w = \sum_{|k| \le q} c_k T_k$$

where c_k with $k = -q, \ldots, q$ are given by (4.10)*.*

Lemma 4.11. *Let the function w be real-analytic and 1-periodic in $\overline{\mathrm{Y}}$. Then there exist positive constants b, C and C_m with $m = 0, 1, 2$ such that the projections $\tau_q \hat{u}$ with $q \in \mathbb{N}$ satisfy the error bound*

$$\|w - \tau_q w\|_{\mathbb{H}^1(\mathrm{Y})} \leq C q \exp(-bq) \tag{4.13}$$

and the stability bound

$$|\tau_q w|_{\mathbb{H}^m(\mathrm{Y})} \leq C_m q^m \quad \textit{for every} \quad m = 0, 1, 2\ . \tag{4.14}$$

Proof. For $|k| \leq q$, the kth coefficient c_k of $\tau_q w$, as introduced in definition 4.10, is given by (4.10). By the classical power series argument, the integrand, being real-analytic, admits a unique extension to the strip

$$S_\theta = \{z \in \mathbb{C}\colon\ 0 \leq \mathsf{Re}\, z \leq 1,\ |\mathsf{Im}\, z| < \theta/(2\pi)\} \tag{4.15}$$

with some $\theta > 0$ which extension is holomorphic in S_θ and continuous on its closure. Since the integrand is 1-periodic with respect to the real part of z, so is its extension, by uniqueness of this extension. By Cauchy's integral formula, we may therefore shift the path of integration in the integral in (4.10) to $\mathrm{D} + \mathsf{i}(\operatorname{sgn} k)\theta/(2\pi)$. The elementary estimate

$$|T_k(y + \mathsf{i}(\operatorname{sgn} k)\theta/2\pi)| = \exp(-|k|\theta),$$

valid for all $y \in \mathrm{Y}$, results in the bound

$$|c_k| = |c_{-k}| \leq M_\theta \exp(-\theta k) \quad \text{for all} \quad k \in \mathbb{N}, \tag{4.16}$$

where M_θ is the maximum of $|w|$ on S_θ. Then we obtain

$$\begin{aligned} \|w - \tau_q w\|^2_{\mathbb{H}^1(\mathrm{Y})} &= \|w - \tau_q w\|^2_{\mathbb{L}^2(\mathrm{Y})} + |w - \tau_q w|^2_{\mathbb{H}^1(\mathrm{Y})} = \sum_{|k|>q} (1 + k^2)|c_k|^2 \\ &\leq 2M_\theta^2 \exp(-2\theta q) \frac{\exp(-2\theta)}{(1 - \exp(-2\theta))^3} (1 + (q+1)^2) \leq C^2 q^2 \exp(-2\theta q), \end{aligned}$$

where $C^2 = 10 \cdot M_\theta^2 \exp(-2\theta)/(1 - \exp(-2\theta))^3$ depends on θ but not on q.

The stability bounds of (4.14) follow from the orthogonality of the exponential basis and, for $m = 1, 2$, by the differentiation of the truncated expansion. □

The following bound is a particular case of [39, theorem 1].

Proposition 4.12. *Let* $\mathrm{I} \subset \mathbb{Z}$ *be finite and consider a trigonometric polynomial*

$$w = \sum_{k \in \mathrm{I}} c_k T_k \tag{4.17}$$

with $c_k \in \mathbb{C}$ *for every* $k \in \mathrm{I}$. *Then*

$$\|w\|_{\mathbb{L}^\infty(\mathrm{Y})} \leq |\mathrm{I}|^{\frac{1}{2}} \|w\|_{\mathbb{L}^2(\mathrm{Y})}.$$

4.4 Approximation of U_1^δ

Based on proposition 4.7 and lemma 4.11, we now analyze the approximate QTT structure of the first-order asymptotic representation (4.4).

Lemma 4.13. *Assume that* f *is real-valued and analytic on* $\overline{\mathrm{D}}$, *and that* a_1 *is* 1-*periodic, real-valued and analytic in* $\overline{\mathrm{Y}}$ *and* σ *is given by* (4.6).

Then there exist positive constants b, $\hat{C}$ *and* $\hat{C}_2$ *with which, for every* δ *such that* $\delta^{-1} \in \mathbb{N}$, *the function* $\tilde{U}_1^\delta$ *given by*

$$\tilde{U}_1^\delta(x) = (\pi_p u_0)(x) + \delta\sigma\big(\delta^{-1}\rho(x)\big)\,\tilde{u}_1\Big(x, \frac{x}{\delta}\Big) \quad \textit{for all} \quad x \in \mathrm{D} \tag{4.18a}$$

with $\tilde{u}_1$ *given by*

$$\tilde{u}_1(x, y) = (\pi_p u_0')(x)\,(\tau_p w_1)(y) \quad \textit{for all} \quad (x, y) \in \mathrm{D} \times \mathrm{Y} \tag{4.18b}$$

satisfies the following error and stability bounds:

$$\big\|\tilde{U}_1^\delta - U_1^\delta\big\|_{\mathbb{H}^1(\mathrm{D})} \leq \hat{C} p \exp(-bp), \qquad \big|\tilde{U}_1^\delta\big|_{\mathbb{H}^2(\mathrm{D})} \leq \hat{C}_2\, \frac{1}{\delta} p^{5/2}, \tag{4.19}$$

where U_1^δ *is given by* (4.4).

Proof. We observe that the assumption that a_1 being real analytic, and 1-periodic implies that the exists a holomorphic extension (again denoted by a_1 to the complex domain which is 1-periodic.This, in turn, implies that the solution w_1 of the cell problem (4.1) is, likewise, holomorphic and 1-periodic.

By lemma 4.11 and proposition 4.7, there exist positive constants b, C and C_m, $m = 0, 1, 2$, such that, for every $p \in \mathbb{N}$, the approximations $\tilde{u}_0 = \pi_p u_0$, $\tilde{v}_1 = \pi_p u_0'$ and $\tilde{w}_1 = \tau_p w_1$ satisfy the error bounds

$$\|u_0 - \tilde{u}_0\|_{\mathbb{H}^1(\mathrm{D})},\ \|u_0' - \tilde{v}_1\|_{\mathbb{H}^1(\mathrm{D})},\ \|w_1 - \tilde{w}_1\|_{\mathbb{H}^1(\mathrm{D})} \leq C p \exp(-bp) \tag{4.20}$$

and the stability bounds

$$|\tilde{u}_0|_{\mathbb{H}^m(\mathrm{D})},\ |\tilde{v}_1|_{\mathbb{H}^m(\mathrm{D})} \leq C_m p^{\max\{m-1,0\}}, \qquad |\tilde{w}_1|_{\mathbb{H}^m(\mathrm{D})} \leq C_m p^m \tag{4.21}$$

for $m = 0, 1, 2$.

Let $p \in \mathbb{N}$ and $\psi(x) = \sigma\big(\delta^{-1}\rho(x)\big)$ for all $x \in \mathrm{D}$. The error of approximating U_1^δ with $\tilde{U}_1^\delta$ reads

$$\tilde{U}_1^\delta - U_1^\delta = u_0 - \tilde{u}_0 + \delta\psi\big\{u_0' - \tilde{v}_1\big\}w_1\Big(\frac{\cdot}{\delta}\Big) + \delta\psi\tilde{v}_1\{w_1 - \tilde{w}_1\}\Big(\frac{\cdot}{\delta}\Big). \qquad (4.22)$$

Since $|\psi(x)| \le 1$ and $|\psi'(x)| \le 2/\delta$ for all $x \in \mathrm{D}$, the error (4.22) can be estimated in the $\mathbb{H}^1$-norm as follows:

$$\begin{aligned}
\big\|\tilde{U}_1^\delta - U_1^\delta\big\|^2_{\mathbb{H}^1(\mathrm{D})} &= \big\|\tilde{U}_1^\delta - U_1^\delta\big\|^2_{\mathbb{L}^2(\mathrm{D})} + \big|\tilde{U}_1^\delta - U_1^\delta\big|^2_{\mathbb{H}^1(\mathrm{D})} \\
&\le 3\|u_0 - \tilde{u}_0\|^2_{\mathbb{L}^2(\mathrm{D})} + 3|u_0 - \tilde{u}_0|^2_{\mathbb{H}^1(\mathrm{D})} \\
&\quad + (3\delta^2 + 36)\|u_0' - \tilde{v}_1\|^2_{\mathbb{L}^2(\mathrm{D})}\|w_1\|^2_{\mathbb{L}^\infty(\mathrm{Y})} + (3\delta^2 + 36)\|\tilde{v}_1\|^2_{\mathbb{L}^\infty(\mathrm{D})}\|w_1 - \tilde{w}_1\|^2_{\mathbb{L}^2(\mathrm{Y})} \\
&\quad + 9\delta^2|u_0' - \tilde{v}_1|^2_{\mathbb{H}^1(\mathrm{D})}\|w_1\|^2_{\mathbb{L}^\infty(\mathrm{Y})} + 9\|u_0' - \tilde{v}_1\|^2_{\mathbb{L}^2(\mathrm{D})}\|w_1'\|^2_{\mathbb{L}^\infty(\mathrm{Y})} \\
&\quad + 9\delta^2\|\tilde{v}_1'\|^2_{\mathbb{L}^\infty(\mathrm{D})}\|w_1 - \tilde{w}_1\|^2_{\mathbb{L}^2(\mathrm{Y})} + 9\|\tilde{v}_1\|^2_{\mathbb{L}^\infty(\mathrm{D})}|w_1 - \tilde{w}_1|^2_{\mathbb{H}^1(\mathrm{Y})} \\
&\le 3\|u_0 - \tilde{u}_0\|^2_{\mathbb{H}^1(\mathrm{D})} \\
&\quad + 39\delta^2\|u_0' - \tilde{v}_1\|^2_{\mathbb{H}^1(\mathrm{D})}\|w_1\|^2_{\mathbb{L}^\infty(\mathrm{Y})} + 39\|\tilde{v}_1\|^2_{\mathbb{L}^\infty(\mathrm{D})}\|w_1 - \tilde{w}_1\|^2_{\mathbb{H}^1(\mathrm{Y})} \\
&\quad + 39\delta^2\|\tilde{v}_1'\|^2_{\mathbb{L}^\infty(\mathrm{D})}\|w_1 - \tilde{w}_1\|^2_{\mathbb{H}^1(\mathrm{Y})} + 39\|u_0' - \tilde{v}_1\|^2_{\mathbb{H}^1(\mathrm{D})}\|w_1'\|^2_{\mathbb{L}^\infty(\mathrm{Y})} \\
&\le 3\|u_0 - \tilde{u}_0\|^2_{\mathbb{H}^1(\mathrm{D})} + 39\|u_0' - \tilde{v}_1\|^2_{\mathbb{H}^1(\mathrm{D})}\big\{\|w_1\|^2_{\mathbb{L}^\infty(\mathrm{Y})} + \|w_1'\|^2_{\mathbb{L}^\infty(\mathrm{Y})}\big\} \\
&\quad + 39\big\{\|\tilde{v}_1\|^2_{\mathbb{L}^\infty(\mathrm{D})} + \|\tilde{v}_1'\|^2_{\mathbb{L}^\infty(\mathrm{D})}\big\}\|w_1 - \tilde{w}_1\|^2_{\mathbb{H}^1(\mathrm{Y})}
\end{aligned}$$

Using the Nikolskii inequality (see proposition 4.8) and the estimates (4.20)–(4.21), we obtain

$$\begin{aligned}
\big\|\tilde{U}_1^\delta - U_1^\delta\big\|^2_{\mathbb{H}^1(\mathrm{D})} &\le 3C^2p^2\exp(-2bp) \\
&\quad + 39C^2p^2\exp(-2bp)\big\{\|w_1\|^2_{\mathbb{L}^\infty(\mathrm{Y})} + \|w_1'\|^2_{\mathbb{L}^\infty(\mathrm{Y})}\big\} \\
&\quad + 39\big\{\|u_0'\|^2_{\mathbb{L}^2(\mathrm{D})} + \|u_0''\|^2_{\mathbb{L}^2(\mathrm{D})}\big\}C^2p^2(p+1)^2\exp(-2bp) \\
&\le \hat{C}^2p^4\exp(-2bp)
\end{aligned}$$

with a positive constant $\hat{C}$ independent of p and δ.

Furthermore, we estimate the stability of the approximation as follows: since $|\psi(x)| \le 1$, $|\psi'(x)| \le 2/\delta$ and $|\psi''(x)| \le 2/\delta^2$ for all $x \in \mathrm{D}$:

$$\begin{aligned}
\big|\tilde{U}_1^\delta\big|^2_{\mathbb{H}^2(\mathrm{D})} &\le 2|\tilde{u}_0|^2_{\mathbb{H}^2(\mathrm{D})} + 12\frac{4}{\delta^2}\|\tilde{v}_1\|^2_{\mathbb{L}^2(\mathrm{D})}\|\tilde{w}_1\|^2_{\mathbb{L}^\infty(\mathrm{Y})} + 12\delta^2|\tilde{v}_1|^2_{\mathbb{H}^2(\mathrm{D})}\|\tilde{w}_1\|^2_{\mathbb{L}^\infty(\mathrm{Y})} \\
&\quad + 12\frac{1}{\delta^2}\|\tilde{v}_1\|^2_{\mathbb{L}^2(\mathrm{D})}\|\tilde{w}_1''\|^2_{\mathbb{L}^\infty(\mathrm{Y})} + 12|\tilde{v}_1|^2_{\mathbb{H}^1(\mathrm{D})}\|\tilde{w}_1'\|^2_{\mathbb{L}^\infty(\mathrm{Y})}
\end{aligned}$$

$$+\,12\cdot\frac{4}{\delta^2}\,\|\tilde{v}_1\|^2_{\mathbb{L}^2(\mathrm{D})}\|\tilde{w}_1'\|^2_{\mathbb{L}^\infty(\mathrm{Y})}+12\cdot 4|\tilde{v}_1|^2_{\mathbb{H}^1(\mathrm{D})}\|\tilde{w}_1\|^2_{\mathbb{L}^\infty(\mathrm{Y})}.$$

By the Jackson–Nikolskii inequality (see proposition 4.12) and the estimates (4.20)–(4.21), we therefore obtain

$$\begin{aligned}
\left|\tilde{U}_1^\delta\right|^2_{\mathbb{H}^2(\mathrm{D})} &\le 2C_2^2p^2+\frac{48}{\delta^2}C_0^2\,2p\,\|\tilde{w}_1\|^2_{\mathbb{L}^2(\mathrm{Y})}+12\delta^2C_2^2p^2\,2p\,\|\tilde{w}_1\|^2_{\mathbb{L}^2(\mathrm{Y})}\\
&\quad+\frac{12}{\delta^2}\,C_0^2\,2p\,\|\tilde{w}_1''\|^2_{\mathbb{L}^2(\mathrm{Y})}+12\,C_1^2\,2p\,\|\tilde{w}_1'\|^2_{\mathbb{L}^2(\mathrm{Y})}\\
&\quad+\frac{48}{\delta^2}C_0^2\,2p\,\|\tilde{w}_1'\|^2_{\mathbb{L}^2(\mathrm{Y})}+48\,C_1^2\,2p\,\|\tilde{w}_1\|^2_{\mathbb{L}^2(\mathrm{Y})}\\
&\le 2C_2^2p^2+\frac{48}{\delta^2}C_0^2\,2p\,\|w_1\|^2_{\mathbb{L}^2(\mathrm{Y})}+12\delta^2C_2^2p^2\,2p\,\|w_1\|^2_{\mathbb{L}^2(\mathrm{Y})}\\
&\quad+\frac{12}{\delta^2}\,C_0^2\,2p\,p^4\,\|w_1''\|^2_{\mathbb{L}^2(\mathrm{Y})}+12\,C_1^2\,2p\,p^2\,\|w_1'\|^2_{\mathbb{L}^2(\mathrm{Y})}\\
&\quad+\frac{48}{\delta^2}C_0^2\,2p\,p^2\,\|w_1'\|^2_{\mathbb{L}^2(\mathrm{Y})}+48\,C_1^2\,2p\,\|w_1\|^2_{\mathbb{L}^2(\mathrm{Y})}\le\hat{C}_2^2\,\frac{1}{\delta^2}\,p^5
\end{aligned}$$

with a positive constant $\hat{C}_2$ independent of p and δ. □

Theorem 4.14. *Let f, c and a_1 satisfy assumption 2.1. Then there exist positive constants $\hat{c}$, λ and R such that the following holds.*

For all $\delta>0$ such that $\log_2(\delta^{-1}-1)\in\mathbb{N}$, the solution u^δ of the problem (2.1) admits an approximation $u^{\delta,l}\in\mathscr{F}^l_{u^\delta}\subset \mathrm{S}^1(\mathrm{D},\mathcal{T}^l)$ with

$$l=\left\lceil\lambda\log_2\delta^{-1}\right\rceil, \tag{4.23a}$$

where $\mathrm{S}^1(\mathrm{D},\mathcal{T}^l)$ and $\mathscr{F}^l_{u^\delta}$ are given by (3.2) and (3.3), such that the error bound

$$\left\|u^\delta-u^{\delta,l}\right\|_{\mathbb{H}^1(\mathrm{D})}\le\hat{c}\sqrt{\delta} \tag{4.23b}$$

holds and the coefficient vector $\boldsymbol{u}^l=\mathscr{A}^l u^{\delta,l}$ admits an exact QTT representation with ranks bounded from above by

$$R\,\log_2^2\delta^{-1}. \tag{4.23c}$$

Proof. First, we observe that since a_1 and f are analytic in $\overline{\mathrm{D}}$, the solution w_1 of the unit-cell problem (4.1) and the solution u_0 of (4.2) together with its first derivative u_0', all of which are independent of the scale parameter δ, are likewise analytic in $\overline{\mathrm{Y}}$ and $\overline{\mathrm{D}}$ respectively. Then, by lemma 4.13, the approximation $\tilde{U}_1^\delta$

given by (4.18) and (4.6) satisfies the bounds of (4.19) with positive constants b, $\hat{C}$ and $\hat{C}_2$ for all $p \in \mathbb{N}$. Let us, for some $\alpha > 0$, set

$$\varkappa_\alpha = \sup_{t>1} \left(\frac{1}{2} + \alpha\right) \frac{\log t}{t^\alpha} \quad \text{and} \quad p = \lfloor b^{-1} \log \delta^{-1/2-\alpha} \rfloor. \tag{4.24}$$

Then we have $\varkappa_\alpha < \infty$, $\delta^\alpha p \le b^{-1} \varkappa_\alpha$ and $\exp(-bp) \le e\, \delta^{1/2+\alpha}$.

Consider $\lambda = 3/2 + 5\alpha/2$ and the corresponding l given by (4.23a). Let $u^{\delta,l} \in \mathscr{F}^l_{u^\delta}$ be the nodal interpolant of $\tilde{U}^\delta_1$ given by (4.4), (4.6) at the nodes of $\mathcal{T}^l$ satisfying the boundary conditions, i.e. the element of $\mathscr{F}^l_{u^\delta}$ such that $\boldsymbol{u}^l = \mathscr{A}^l u^{\delta,l} = \mathscr{A}^l \tilde{U}^\delta_1$. From proposition 4.5, we obtain a bound on the error of the nodal interpolation:

$$\begin{aligned} \left\| u^{\delta,l} - \tilde{U}^\delta_1 \right\|^2_{\mathbb{H}^1(\mathrm{D})} &\le \frac{1}{8}\, h_l^2 \left\{ 1 + \frac{1}{8}\, h_l^2 \right\} |\tilde{U}^\delta_1|^2_{\mathbb{H}^2(\mathrm{D})} \\ &\le \frac{9}{64}\, \hat{C}_2^2\, 2^{-2l}\, \frac{p^5}{\delta^2} \le \frac{9}{64}\, \hat{C}_2^2\, (\delta^\alpha p)^5\, \delta \le \frac{9}{64}\, \hat{C}_2^2\, (b^{-1}\varkappa_\alpha)^5\, \delta, \end{aligned} \tag{4.25}$$

using the second bound of (4.19). The first bound of (4.19) results in

$$\left\| \tilde{U}^\delta_1 - U^\delta_1 \right\|^2_{\mathbb{H}^1(\mathrm{D})} \le \hat{C}^2 p^2 \exp(-2bp) \le \hat{C}^2 e^2\, (\delta^\alpha p)^2 \delta \le \hat{C}^2 e^2\, (b^{-1}\varkappa_\alpha)^2\, \delta \tag{4.26}$$

By lemma 4.2, the inequality (4.5) holds with a positive constant c. Combining it with (4.25) and (4.26) and applying the triangle inequality, we arrive at

$$\left\| u^\delta - u^{\delta,l} \right\|_{\mathbb{H}^1(\mathrm{D})} = \left\| \left(u^\delta - U^\delta_1\right) + \left(U^\delta_1 - \tilde{U}^\delta_1\right) + \left(\tilde{U}^\delta_1 - u^{\delta,l}\right) \right\|_{\mathbb{H}^1(\mathrm{D})} \le \hat{c}\, \sqrt{\delta}$$

with $\hat{c} = c + \hat{C} e\, b^{-1} \varkappa_\alpha + (3/8)\, \hat{C}_2\, (b^{-1}\varkappa_\alpha)^{5/2}$, which proves (4.23b).

To prove the rank bound (4.23c), let us introduce 2^l-vectors $\boldsymbol{u}^l_0 = \mathscr{A}^l \pi_p u_0$, $\boldsymbol{v}^l_1 = \mathscr{A}^l \pi_p u'_0$, $\boldsymbol{w}^l_1 = \mathscr{A}^l (\tau_p \tilde{w}_1)(\cdot/\delta)$ and $\boldsymbol{\sigma}^l = \mathscr{A}^l \sigma(\rho(\cdot)/\delta)$. Then for the vector $\boldsymbol{u}^l = \mathscr{A}^l u^{\delta,l} = \mathscr{A}^l \tilde{U}^\delta_1$, by (4.18), we have the decomposition

$$\boldsymbol{u}^l = \boldsymbol{u}^l_0 + \delta\, \boldsymbol{\sigma}^l \odot \boldsymbol{v}^l_1 \odot \boldsymbol{w}^l_1. \tag{4.27}$$

By proposition 3.5, there exist exact QTT representations of $\boldsymbol{u}^l_0$ and $\boldsymbol{v}^l_1$ with ranks bounded by $p+1$. Also, from proposition 3.6 we conclude that $\boldsymbol{w}^l_1$ admits an exact QTT representation of ranks bounded by $2p$. Also, by (4.6) and [29, lemma 3.7], $\boldsymbol{\sigma}^l$ admits an exact QTT representation of ranks bounded by 5. By proposition 3.2, these arguments and (4.27) result in the QTT rank bound $r = (p+1) + 5(p+1)2p = (p+1)(10p+1)$ for $\boldsymbol{u}^l$. Due to the choice of p, we have $p \le b^{-1}(1/2+\alpha)\, \log \delta^{-1}$, and $r \le R\, \log_2^2 \delta^{-1}$ holds with a positive constant R independent of δ, which proves (4.23c). □

Remark 4.15. *The preceding tensor-rank bounds for the first order finite-element solution are polylogarithmic in the scale parameter δ. They imply corresponding results for model-order-reduction and reduced-basis methods. Consider* parametric families of two-scale problems (2.1), (2.4), (2.6) *under assumption 2.1 corresponding to pairs of a_1^μ and f^μ parametrized by a parameter $\mu \in \mathcal{M}$. Here, $\mathcal{M}$ denotes a closed and bounded parameter set (more generally, a compact, metric space would suffice for the following). For every $\delta \in (0,1)$ such that $\log_2(\delta^{-1}-1) \in \mathbb{N}$, for all $\mu \in \mathcal{M}$ and for $l \sim \log \delta^{-1}$ (4.23a), denote by $\hat{u}^{\delta,\mu,l} \in \mathrm{S}^1(\mathrm{D},\mathcal{T}^l)$ and $u^{\delta,\mu,l} \in \mathrm{S}^1(\mathrm{D},\mathcal{T}^l)$ the exact Galerkin solution and its QTT-FE approximation given by theorem 4.14 respectively. Consider the sets*

$$\hat{\mathcal{U}}^{\delta,l} = \left\{\hat{u}^{\delta,\mu,l} \colon \mu \in \mathcal{M}\right\} \quad \textit{and} \quad \mathcal{U}^{\delta,l} = \left\{u^{\delta,\mu,l} \colon \mu \in \mathcal{M}\right\},$$

the exact solutions from the former being approximated by the corresponding approximate solutions from the latter.

Assume that, with respect to the parameter $\mu \in \mathcal{M}$, all a_1^μ admit holomorphic extensions to a strip S_θ (4.15) that are uniformly separated from zero and all f^μ uniformly satisfy the analyticity condition (4.8). Then, for $p = \mathcal{O}(\log \delta^{-1})$ (4.24), $\mathcal{W}^{\delta,l} \subset \mathrm{S}^1(\mathrm{D},\mathcal{T}^l)$, the linear space of continuous, piecewise-linear interpolants on $\mathcal{T}^l$ of products of algebraic and trigonometric polynomials of degrees at most $p+1$ and p respectively, satisfies $\dim \mathcal{W}^{\delta,l} = (p+2)(2p+1) = \mathcal{O}(\log^2 \delta^{-1})$ *and, due to (4.18), we have $\mathcal{U}^{\delta,l} \subset \mathcal{W}^{\delta,l}$. The subspace $\mathcal{W}^{\delta,l}$ thus realizes an upper bound on the* Kolmogorov n-width d_n *in* $\mathbb{H}^1(\mathrm{D})$ *(see, e.g. [46] and the references there for definitions of this terminology) of the set $\hat{\mathcal{U}}^{\delta,l}$ of approximate solutions: there holds*

$$d_n\Big(\hat{\mathcal{U}}^{\delta,l}, \mathbb{H}^1(\mathrm{D})\Big) \leq C\big\{\exp(-b\sqrt{n}\,) + \sqrt{\delta}\,\big\}.$$

Here, b and C are positive constants which are independent of δ and n.

5 QTT-FEM for the Helmholtz Equation

We now consider the model problem (2.1) under assumption 2.2. We establish QTT rank bounds of the solution which are explicit in the wavenumber k.

For the sake of brevity, we denote with $\|\cdot\|$ the $\mathbb{L}^2(\mathrm{D})$-norm and $|\cdot|_m = \|\cdot^{(m)}\|$, the $\mathbb{H}^m(\mathrm{D})$-seminorm. The error analysis is based on the *wavenumber-weighted* $\mathbb{H}^1$ norm $\|\cdot\|_{(k)}$ given by $\|v\|_{(k)}^2 = \|v\|^2 + k^2|v|_1^2$ for every $v \in V$. We work in the variational space $V = \mathbb{H}^1_{0,\{0\}}(\mathrm{D}) = \{u \in \mathbb{H}^1(\mathrm{D}) :\ u(0) = 0\}$.

5.1 Polynomial approximation of the solution

We start by collecting some a-priori estimates on the solution u^k of (2.9) for (real-valued) wavenumber $k > 0$

Proposition 5.1. *[20, theorem 4.4] Consider* (2.9)–(2.10) *under assumption 2.2. Then, there exists a unique solution $u^k \in V$ and there hold the wavenumber-explicit a-priori estimates*

$$\|u^k\| \leq k^{-1}\|f\| \ , \quad |u^k|_1 \leq \|f\| \ , \quad |u^k|_2 \leq (1+k)\|f\| \ . \tag{5.1}$$

As for the analysis in the homogenization case in Section 4, for the QTT rank analysis we assume that f in (2.1) is analytic in $\overline{\mathrm{D}}$. This implies that, for every polynomial degree $p \in \mathbb{N}$, the $\mathbb{L}^2(\mathrm{D})$-projection f_p of f onto the space of polynomials of degree at most p defined in D converge exponentially: there exist positive constants b and C such that for every $p \in \mathbb{N}$ the bound

$$\|f - f_p\| \leq C \exp(-bp) \tag{5.2}$$

holds. We remark that the $\mathbb{L}^2(\mathrm{D})$-projection f_p is realized by a p-term truncated Legendre expansion of f (for a convergence proof, see, e.g. [8, theorem 12.4.7]). The a-priori bounds (5.1) and the error bound (5.2) imply, by superposition, the following result.

Proposition 5.2. *Consider* (2.1) *under assumption 2.2. Denote by u_p^k the unique solution of* (2.1) *under assumption 2.2 with f_p in place of f. Then there exist positive constants b and C independent of k such that there holds the error bound*

$$\|u^k - u_p^k\|_{\mathbb{H}^1(\mathrm{D})} \leq C \exp(-bp). \tag{5.3}$$

We denote the Helmholtz operator (2.1), (2.8) by L_k. The inequality (5.3) implies a bound on the n-width of the solution set $\mathcal{X}_k = L_k^{-1}\mathcal{A}(\overline{\mathrm{D}})$, where $\mathcal{A}(\overline{\mathrm{D}})$ denotes the set of functions analytic on $\overline{\mathrm{D}}$.

To study the QTT rank of u_p^k with $p \in \mathbb{N}$, with particular attention to its dependence on the wavenumber k, we represent L_k^{-1} by its Green function:

$$u_p^k(x) = (L_k^{-1} f_p)(x) = \int_0^1 G^k(x,s) f_p(s)\,\mathrm{d}s, \tag{5.4}$$

where the Green's function $G^k(x,s)$ of (2.9)–(2.10) with $g = 0$ is given by

$$G^k(x,s) = \frac{1}{k}\begin{cases} \sin(kx)\exp(\mathrm{i}ks) & 0 \leq x \leq s \leq 1, \\ \sin(ks)\exp(\mathrm{i}kx) & 0 \leq s \leq x \leq 1. \end{cases} \tag{5.5}$$

5.2 QTT-FE approximation of the solution

For discrete approximations of the solution, we consider functions from $V^l = \mathrm{S}^1_0(\mathrm{D}, \mathcal{T}^l) = \mathrm{S}^1(\mathrm{D}, \mathcal{T}^l) \cap V$ (3.2), $l \in \mathbb{N}$, as introduced in section 3.1 with $\mu_1 = 0$ and $\mu_2 = 1$ due to the boundary conditions (2.10).

Theorem 5.3. *Let assumption 2.2 hold. Then there exist positive constants $\hat{C}$, l_0, R and r_0 independent of the wave number k such that the following holds.*

For every $k > 0$, the unique solution u^k of the problem (2.9)–(2.10) *in the weak form* (2.2)–(2.11) *admits a sequence of approximations $u^{k,l} \in \mathscr{F}^l_{u^k} \subset \mathrm{S}^1(\mathrm{D}, \mathcal{T}^l)$, $l > l_0$, where $\mathrm{S}^1(\mathrm{D}, \mathcal{T}^l)$ and $\mathscr{F}^l_{u^k}$ are given by* (3.2) *and* (3.3)*, satisfying the accuracy estimate*

$$\|u^k - u^{k,l}\|_{\mathbb{H}^1(\mathrm{D})} \leq \hat{C}\,(k+2)\,2^{-l} \tag{5.6a}$$

and such that every coefficient vector $\boldsymbol{u}^{k,l} = \mathscr{A}^l u^{k,l}$, $l > l_0$, admits an exact QTT representation with ranks bounded from above by

$$R\,l + r_0. \tag{5.6b}$$

Proof. Let us consider $p \in \mathbb{N}$ and the corresponding function u^k_p defined in proposition 5.2. Then we define a 2^l-component vector $\boldsymbol{u}^{k,l} = \mathscr{A}^l u^k_p$ and $u^{k,l} \in \mathscr{F}^l_{u^k} \subset \mathrm{S}^1(\mathrm{D}, \mathcal{T}^l)$ such that $\boldsymbol{u}^{k,l} = \mathscr{A}^l u^{k,l}$. In other words, $u^{k,l}$ is the piecewise-linear interpolant of u^k_p at the nodes of $\mathcal{T}^l$ except the origin and satisfying the boundary condition (2.10) at the origin, and $\boldsymbol{u}^{k,l}$ is the corresponding coefficient vector.

By the triangle inequality, we have

$$\|u^k - u^{k,l}\|_{\mathbb{H}^1(\mathrm{D})} \leq \|u^k - u^k_p\|_{\mathbb{H}^1(\mathrm{D})} + \|u^k_p - u^{k,l}\|_{\mathbb{H}^1(\mathrm{D})}. \tag{5.7}$$

By proposition 5.2, there exist positive constants C_1 and b_1 such that, for all $p \in \mathbb{N}$ and $k > 0$, the following bound holds:

$$\|u^k - u^k_p\|_{\mathbb{H}^1(\mathrm{D})} \leq C_1 \exp(-b_1 p). \tag{5.8}$$

By proposition 4.5, we bound the error of the nodal interpolation by

$$\begin{aligned} \|u^k_p - u^{k,l}\|^2_{\mathbb{H}^1(\mathrm{D})} &\leq \frac{1}{8}\,2^{-2l}\left\{1 + \frac{1}{8}\,2^{-2l}\right\} |u^k_p|^2_{\mathbb{H}^2(\mathrm{D})} \\ &\leq \frac{9}{64}\,2^{-2l}\,C^2\,(1+k)^2\,\|f_p\|^2_{\mathbb{L}^2(\mathrm{D})} \leq 2^{-2l}\,\hat{C}^2\,(1+k)^2, \end{aligned} \tag{5.9}$$

where we use the third bound of proposition 5.1 with a positive constant C independent of k and p, take into account the projection property of f_p and introduce $\hat{C} = 3C\|f\|_{\mathbb{L}^2(\mathrm{D})}/8$.

Let us set

$$p = \lceil b_1^{-1}(l \log 2 + \log C_1/\hat{C}) \rceil. \tag{5.10}$$

Then we have $p \in \mathbb{N}$ for large enough $l \in \mathbb{N}$ and, from (5.7)–(5.9), we obtain the inequality

$$\|u^k - u^{k,l}\|_{\mathbb{H}^1(\mathrm{D})} \leq \hat{C}\,(k+2)\,2^{-l}.$$

We rewrite (5.4)–(5.5) as follows:

$$\begin{aligned}(2\mathrm{i}k)\, u_p^k(x) = e^{\mathrm{i}kx} \int_0^1 e^{\mathrm{i}ks} f_p(s)\,\mathrm{d}s \\ - e^{\mathrm{i}kx} \int_0^x e^{-\mathrm{i}ks} f_p(s)\,\mathrm{d}s - e^{-\mathrm{i}kx} \int_x^1 e^{\mathrm{i}ks} f_p(s)\,\mathrm{d}s \quad \text{for all} \quad x \in \overline{\mathrm{D}}.\end{aligned} \tag{5.11}$$

For every $\alpha \in \mathbb{N}$ and for all $\xi, \eta \in \overline{\mathrm{D}}$ and $\kappa \in \mathbb{R} \setminus \{0\}$, the following holds:

$$\int_\xi^\eta e^{\mathrm{i}\kappa s} s^\alpha \,\mathrm{d}s = \sum_{\beta=0}^{\alpha} \frac{(-1)^{\alpha-\beta}}{(\mathrm{i}\kappa)^{\alpha-\beta+1}}\, \frac{\alpha!}{\beta!} \left(e^{\mathrm{i}\kappa s} s^\beta\right)\Big|_{s=\xi}^{s=\eta}. \tag{5.12}$$

Since f_p is a polynomial of degree p at most, there exist $c_\alpha \in \mathbb{C}$, $\alpha = 0, 1, \ldots, p$, such that $f_p(x) = \sum_{\alpha=0}^{p} c_\alpha\, x^\alpha$ for all $x \in \mathrm{D}$. Then, combining (5.11) and (5.12), we obtain

$$\begin{aligned} u_p^k(x) &= \frac{1}{2\mathrm{i}k} \sum_{\alpha=0}^{p} c_\alpha \sum_{\beta=0}^{\alpha} \frac{(-1)^{\alpha-\beta}}{(\mathrm{i}k)^{\alpha-\beta+1}}\, \frac{\alpha!}{\beta!} \left\{ e^{\mathrm{i}k(1+x)} + \left(1 + (-1)^{\alpha-\beta}\right) x^\beta - e^{\mathrm{i}k(1-x)} \right\} \\ &= e^{\mathrm{i}k} \sin(kx) \sum_{\alpha=0}^{p} c_\alpha \sum_{\beta=0}^{\alpha} \frac{(-1)^{\alpha-\beta}}{(\mathrm{i}k)^{\alpha-\beta+1}}\, \frac{\alpha!}{\beta!} + \sum_{\alpha=0}^{p} c_\alpha \sum_{\substack{\beta=0,1,\ldots,\alpha: \\ \alpha-\beta \in 2\mathbb{Z}}} \frac{1}{(\mathrm{i}k)^{\alpha-\beta+2}}\, \frac{\alpha!}{\beta!}\, x^\beta \end{aligned} \tag{5.13}$$

for all $x \in \mathrm{D}$. Applying propositions 3.6 and 3.5, we obtain that the vector $\boldsymbol{u}^{k,l} = \mathscr{A}^l u_p^k$ of 2^l nodal values of u_p^k admits an exact QTT representation with ranks bounded from above by

$$r \leq 2 + (p+1) = p + 3. \tag{5.14}$$

Together with (5.10), that leads to the claimed rank bound (5.6b).

□

Similarly to what was outlined in remark 4.15, our analysis admits the following extension.

Remark 5.4. *The preceding rank bounds imply corresponding results for model-order-reduction and reduced-basis methods. Consider* parametric families of Helmholtz problems (2.9)–(2.10) *under assumption 2.2 corresponding to right-hand sides f^μ parametrized by a parameter $\mu \in \mathcal{M}$. Here, $\mathcal{M}$ denotes a closed and bounded parameter set (more generally, a compact, metric space would suffice for the following). For all $k \in \mathbb{R}$ and $l \in \mathbb{N}$ such that $k > 0$ and $l > l_0$, where l_0 is the constant given by theorem 5.3, which is independent of k, we denote by $\hat{u}^{k,\mu,l} \in \mathrm{S}^1(\mathrm{D}, \mathcal{T}^l)$ and $u^{k,\mu,l} \in \mathrm{S}^1(\mathrm{D}, \mathcal{T}^l)$ the exact Galerkin solution and its QTT-FE approximation given by theorem 5.3 respectively. Consider the sets*

$$\hat{\mathcal{U}}^{k,l} = \left\{\hat{u}^{k,\mu,l} \colon \mu \in \mathcal{M}\right\} \quad \textit{and} \quad \mathcal{U}^{k,l} = \left\{u^{k,\mu,l} \colon \mu \in \mathcal{M}\right\},$$

the exact solutions from the former being approximated by the corresponding approximate solutions from the latter.

Assume that, with respect to the parameter $\mu \in \mathcal{M}$, all f^μ uniformly satisfy the analyticity condition (4.8)*. Then, for $p = \mathcal{O}(l)$* (5.10)*, $\mathcal{W}^{k,l} \subset \mathrm{S}^1(\mathrm{D}, \mathcal{T}^l)$, the linear space of continuous, piecewise-linear interpolants on $\mathcal{T}^l$ of all functions of the form* (5.13)*, satisfies* $\dim \mathcal{W}^{k,l} = \lceil (p+1)/2 \rceil = \mathcal{O}(l)$ *and, due to* (4.18)*, we have $\mathcal{U}^{k,l} \subset \mathcal{W}^{k,l}$. The subspace $\mathcal{W}^{k,l}$ thus realizes an upper bound on the* Kolmogorov n-width d_n *in $\mathbb{H}^1(\mathrm{D})$ of the set $\hat{\mathcal{U}}^{k,l}$ of approximate solutions: there holds*

$$d_n\Big(\hat{\mathcal{U}}^{k,l}, \mathbb{H}^1(\mathrm{D})\Big) \leq C\big\{\exp(-bn) + (k+2)\, 2^{-l}\big\}.$$

Here, b and C are positive constants independent of k and n.

6 QTT Compressibility of the System Matrix

For $l \in \mathbb{N}$, we consider the solution of (2.2) in the form of a FE approximation $u^l \in \mathscr{F}_u^l$, where the admissible approximation set $\mathscr{F}_u^l$ is given by (3.3). For every $l \in \mathbb{N}$, we represent the solution by a vector $\boldsymbol{u}^l \in \mathbb{R}^{2^l}$ of its values at the active nodes of $\mathcal{T}^l$:

$$\boldsymbol{u}^l = \mathscr{A}^l u^l$$

in the sense of (3.1)–(3.5). The discrete problem reads as follows: find $\boldsymbol{u} \in \mathbb{R}^n$ such that

$$(\boldsymbol{A}^l + \boldsymbol{C}^l)\boldsymbol{u}^l = \boldsymbol{f}^l. \tag{6.1}$$

Here, $\boldsymbol{f}^l$ is the load-vector corresponding to the right-hand side of (2.1) and $\boldsymbol{A}^l$ and $\boldsymbol{C}^l$ are matrices corresponding to the two terms in the left-hand side of (2.1): the first is a tridiagonal stiffness matrix, and the second is a diagonal matrix.

To demonstrate the QTT compressibility of the matrix of (6.1), we consider the case of homogeneous Dirichlet boundary conditions. Then, using the midpoint quadrature rule, we construct the matrix $\boldsymbol{A}^l$ from the values of the diffusion coefficient a in the following way: $2^{-2l}\boldsymbol{A}^l_{ii} = a(\tilde{x}_i)+a(\tilde{x}_{i+1})$ for $0 \le i < n$ and $2^{-2l}\boldsymbol{A}^l_{i,i+1} = 2^{-2l}\boldsymbol{A}^l_{i+1,i} = -a(\tilde{x}_{i+1})$ for $0 \le i < n-1$. The second matrix, $\boldsymbol{C}^l$, is approximated as follows: $\boldsymbol{C}^l_{ii} = c(z_i)$ for $0 \le i < n$. We define vectors $\boldsymbol{a}_+ = \mathcal{I}^l_+ a, \boldsymbol{a}_- = \mathcal{I}^l_- a, \boldsymbol{c} = \mathscr{A}^l c$. Then

$$\boldsymbol{A}^l = (\boldsymbol{I}^l - \boldsymbol{S}^l)\operatorname{diag}\boldsymbol{a}^l_+ + (\boldsymbol{I}^l - {\boldsymbol{S}^l}^\top)\operatorname{diag}\boldsymbol{a}^l_- \quad \text{and} \quad \boldsymbol{C}^l = \operatorname{diag}\boldsymbol{c}^l, \tag{6.2}$$

where a square matrix $\boldsymbol{S}^l$ denotes the matrix of one-position downward shift:

$$\boldsymbol{S}^l = \begin{pmatrix} 0 & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix}.$$

Let us assume that the coefficients a and c are approximated with $\tilde{a}$ and $\tilde{c}$. This approximation induces the corresponding approximations $\tilde{\boldsymbol{a}}^l_+$, $\tilde{\boldsymbol{a}}^l_-$ and $\tilde{\boldsymbol{c}}^l$ of $\boldsymbol{a}^l_+$, $\boldsymbol{a}^l_-$ and $\boldsymbol{c}$. Then approximations $\tilde{\boldsymbol{A}}^l$, $\tilde{\boldsymbol{C}}^l$ and $\tilde{\boldsymbol{A}}^l + \tilde{\boldsymbol{C}}^l$ of the matrices $\boldsymbol{A}^l$, $\boldsymbol{C}^l$ and $\boldsymbol{A}^l + \boldsymbol{C}^l$ arise analogously to (6.2).

Lemma 6.1. *Assume that the vectors $\tilde{\boldsymbol{a}}^l_+$ and $\tilde{\boldsymbol{a}}^l_-$ have QTT representations with the ranks bounded by $R_{\tilde{a}}$ and $\tilde{\boldsymbol{c}}^l$ has one with the ranks bounded by $R_{\tilde{c}}$. Then the matrices $\tilde{\boldsymbol{A}}^l$, $\tilde{\boldsymbol{C}}^l$ and $\tilde{\boldsymbol{A}}^l + \tilde{\boldsymbol{C}}^l$ defined above can be represented in the QTT format with the ranks bounded by $6R_{\tilde{a}}$, $R_{\tilde{c}}$ and $6R_{\tilde{a}} + R_{\tilde{c}}$ respectively.*

Proof. According to proposition 3.3, the diagonalization of a QTT-structured vector results in a matrix with the corresponding format with the same ranks. On the other hand, as it follows from [30, lemma 3.1], the bidiagonal matrices $\boldsymbol{I}^l - \boldsymbol{S}^l$ and $\boldsymbol{I}^l - (\boldsymbol{S}^l)^\top$ can be represented in the QTT format with the ranks bounded by 3. By applying proposition 3.4 to the representation (6.2), we obtain the rank bounds for $\tilde{\boldsymbol{A}}^l$ and $\tilde{\boldsymbol{C}}^l$. Together with proposition 3.2, they lead to the rank bound for $\tilde{\boldsymbol{A}}^l + \tilde{\boldsymbol{C}}^l$. □

A representation of the form (6.2) was suggested in [10, p. 10]. In that work, under the same assumption on $\tilde{\boldsymbol{a}}_+$ and $\tilde{\boldsymbol{a}}_-$, a QTT representation of $\tilde{\boldsymbol{A}}$ with ranks bounded by $7R_{\tilde{a}}$ was considered. Here, in Lemma 6.1, we make a straightforward reduction of the constant from 7 to 6.

Assume that the coefficients a and c are analytic in $\overline{\mathrm{D}}$. Then the approximations $\tilde{a}$ and $\tilde{c}$ can be constructed as algebraic polynomials of degrees p_a and p_c, similarly to u_0 in theorem 4.14. If the coefficients satisfy assumption (2.1), then $\tilde{a}$ and $\tilde{c}$ can be constructed as rescaled trigonometric polynomials of degrees p_a and p_c, similarly to $w_1(\cdot/\delta)$ in theorem 4.14. In either case, we have the bounds

$$\frac{\|\tilde{a}-a\|_{L^\infty(\mathrm{D})}}{\|a\|_{L^\infty(\mathrm{D})}} \le C_a \exp(-b_a p_a), \quad \frac{\|\tilde{c}-c\|_{L^\infty(\mathrm{D})}}{\|c\|_{L^\infty(\mathrm{D})}} \le C_c \exp(-b_c p_c).$$

Here, the constants C_a, b_a and C_c, b_c depend only on the properties of the functions being approximated, namely on the size of the domains of possible holomorphic extensions of the functions into the complex plane and on the growth of the extensions in these domains. In either case, by propositions 3.5 and 3.6, we obtain the bounds on the accuracy estimate of the QTT-structured representation of the matrices, see Theorem 6.2 below. With $\|\cdot\|_{\max}$ we denote the maximum-entry matrix norm; for example, $\|\boldsymbol{A}^l\|_{\max} = \max_{0\le i,j<n}|\boldsymbol{A}^l_{ij}|$.

Theorem 6.2. *There exist constants $C_a, b_a > 0$ such that, for every $r \in \mathbb{N}$, the matrix $\boldsymbol{A}$ has an approximation $\tilde{\boldsymbol{A}}$ that admits QTT representation with ranks bounded by r and satisfies the accuracy estimate*

$$\frac{\|\tilde{\boldsymbol{A}}^l - \boldsymbol{A}^l\|_{\max}}{\|\boldsymbol{A}^l\|_{\max}} \le C_a \exp(-b_a r) .$$

Analogous statements hold, with different constants, for the matrices $\boldsymbol{C}^l$ and $\boldsymbol{A}^l + \boldsymbol{C}^l$.

The results in this section cover, in particular, the case of two-scale diffusion, i.e. (2.1) under assumption 2.1. In the Helmholtz case (2.9)–(2.10), the coefficients are constant and the matrices $\boldsymbol{A}^l$ and $\boldsymbol{C}^l$ admit, independently of the wavenumber k, *exact QTT-representations* with ranks bounded by 3 and 2 respectively, see [30, lemma 3.1] for the first term of $\mathfrak{a}$ in (2.11) and proposition 3.3 for the second and third terms.

7 Numerical Experiments

In this section we provide numerical experiments for two problems, one is the multiscale problem and another is the 1D dimensional Helmholtz problem with radiating boundary condition.

Our goal is to show that the solution can be well-approximated in the QTT-format with a small error in the $\mathbb{H}^1$-norm and compare the behavior of the error with respect to the number of degrees of freedom and the multiscale parameter δ. Supplementary materials with the code are available online[1].

Once we have confirmed the QTT-structure of the solution and of the matrix itself, we can make use of black-box QTT-solvers, for example the DMRG [41] or AMEn [12] solvers. We note in passing that special treatment for small (relative to the machine precision) virtual meshwidth is needed, since in this case the condition number of the matrix which arises from the virtual discretization could be prohibitive. Even if the original matrix is represented exactly within machine precision, the solution to the perturbed problem resulting from QTT-formatted rank truncation could be very far from the true one. Thus, and efficient numerical algorithm to determine the QTT-structured solution requires numerical algorithms and/or discretization schemes that are better suited for this purpose. We note that the system with better condition number can be dense and even larger than the initial one, but still useful for the QTT approach due to the logarithmic complexity.

In the numerical experiment below, we study, in particular, how certain estimates ε_l of the $\mathbb{H}^1$-norm errors of QTT-FE approximate solutions depend on the number l of levels and on the corresponding number of QTT parameters N_l used to represent the approximate solutions. In both the problems, we observe the following convergence with respect to l:

$$\varepsilon_l \leq C \exp(-b N_l^{1/\kappa}) \tag{7.1}$$

with positive constants C, b and κ independent of l. Both the left- and right-hand sides of (7.1) depend also on the respective scale parameter, δ or k.

7.1 Two-scale diffusion problem

7.1.1 Discrete problem

We consider (2.1), (2.4), i.e.

$$(au')' = 1 \quad \text{in} \quad \mathrm{D}, \qquad u(1) = u(0) = 0,$$

under assumption 2.1, but with the coefficient a is taken, differently from (2.5), as in [18]:

$$a(x) = a_0(x)\, a_1\!\left(\frac{x}{\delta}\right) \quad \text{for all} \quad x \in \mathrm{D}, \tag{7.2}$$

with $a_0(x) = 1 + x$ for all $x \in \mathrm{D}$ and $a_1(y) = 2/3\,(1 + \cos^2(y))$ for all $y \in \mathrm{Y}$.

[1] `https://github.com/rakhuba/homhelm_experiments/blob/master/data_preparation.ipynb`

To discretize the problem, we use finite-element approximations based on 2^l interior nodes, where $l \in \mathbb{N}$, see section 3.1 with $\mu_1 = \mu_2 = 0$. For every $l \in \mathbb{N}$, the Galerkin solution $u^{\delta,l}$ is parametrized by the 2^l-component vector $\boldsymbol{u}^{\delta,l} = \mathscr{A}^l u^{\delta,l}$ solving the Galerkin system of the form

$$\boldsymbol{A}^{\delta,l}\boldsymbol{u}^{\delta,l} = \boldsymbol{f}^l. \tag{7.3}$$

7.1.2 QTT approximation of the discrete solution

For various $l \in \mathbb{N}$, we approximate these solutions by $u_{\mathrm{qtt}}^{\delta,l}$ such that

$$\|u_{\mathrm{qtt}}^{\delta,l} - u^{\delta,l}\|_{\mathbb{H}^1(\mathrm{D})} \leq \tau_l(\delta) \tag{7.4}$$

with a certain accuracy $\tau_l(\delta)$ and that the corresponding 2^l-component vector $\boldsymbol{u}_{\mathrm{qtt}}^{\delta,l} = \mathscr{A}^l u_{\mathrm{qtt}}^{\delta,l}$ is represented in the QTT format with the minimal number of parameters N_l (3.11). For the error of the QTT-structured approximation with respect to the exact solution,

$$\hat{\varepsilon}_l(\delta) = \|u_{\mathrm{qtt}}^{\delta,l} - u^{\delta}\|_{\mathbb{H}^1(\mathrm{D})}, \tag{7.5}$$

the triangle inequality results in the bound

$$\hat{\varepsilon}_l(\delta) \leq \|u^{\delta,l} - u^{\delta}\|_{\mathbb{H}^1(\mathrm{D})} + \tau_l(\delta). \tag{7.6}$$

In particular, for $\tau_l = \|u^{\delta,l} - u^{\delta}\|_{\mathbb{H}^1(\mathrm{D})}$ for all $l \in \mathbb{N}$, using the quasi-optimality of the Galerkin FEM, we arrive at the quasi-optimality of the QTT-structured solutions $u_{\mathrm{qtt}}^{\delta,l}$, $l \in \mathbb{N}$, under (7.4).

Since the exact solution u^{δ} is not known analytically, $\|u^{\delta,l} - u^{\delta}\|_{\mathbb{H}^1(\mathrm{D})}$ has to be evaluated approximately. For each $l \in \mathbb{N}$, we use Aitken's extrapolation procedure with an adaptive selection of the order using solutions on three consecutive discretization levels ($l-2$, $l-1$ and l) to obtain an extrapolation $u_{\mathrm{ext}}^{\delta,l}$. Then, using (7.4)–(7.6) with $\tau_l(\delta) = \|u^{\delta,l} - u_{\mathrm{ext}}^{\delta,l}\|_{\mathbb{H}^1(\mathrm{D})}$ and the approximation $\|u^{\delta,l} - u^{\delta}\|_{\mathbb{H}^1(\mathrm{D})} \approx \|u_{\mathrm{ext}}^{\delta,l} - u^{\delta,l}\|_{\mathbb{H}^1(\mathrm{D})}$, we arrive at the error estimate $\varepsilon_l(\delta)$ given by

$$\varepsilon_l(\delta) = 2\,\|u_{\mathrm{ext}}^{\delta,l} - u^{\delta,l}\|_{\mathbb{H}^1(\mathrm{D})}, \tag{7.7}$$

which approximates $\hat{\varepsilon}_l(\delta)$ given by (7.5).

7.1.3 Results for various values of the microscale parameter

For every $m = 1, 2, 3, 4$, we consider $\delta = 10^{-m}$. The error estimate $\varepsilon_l(\delta)$, as expected, converges algebraically with respect to 2^l, which is the number of virtual degrees of freedom, see figure 1. On the other hand, with respect to N_l (3.11), which is the number of QTT parameters, the same error estimate converges exponentially. Namely, we observe (7.1) with $\kappa \approx 2$, see figure 2. This convergence is superior to the theoretical bound with $\kappa = 5$, which follows immediately from (3.11) and theorem 4.14 in the case of $a_0 = 1$ in (7.2).

We also study the dependence of the error on the microscale parameter δ. For each pair of $\varepsilon = 4 \cdot 10^{-m}$ with $m = 2, 3, 4$, and $\delta = 10^{-m}$ with $m = 1, 2, 3, 4$, we find the minimal value of l such that $\|u^{\delta,l} - u^{\delta}\mathrm{ext}\|_{\mathbb{H}^1(\mathrm{D})} \leq \varepsilon/2$. Then we obtain the corresponding QTT approximation $u^{\delta,l}_{\mathrm{qtt}}$ of $u^{\delta,l}$ such that $\|u^{\delta,l}_{\mathrm{qtt}} - u^{\delta,l}\|_{\mathbb{H}^1(\mathrm{D})} \leq \varepsilon/2$, and hence $\|u^{\delta,l}_{\mathrm{qtt}} - u^{\delta}\|_{\mathbb{H}^1(\mathrm{D})} \approx \|u^{\delta,l}_{\mathrm{qtt}} - u^{\delta,l}_{\mathrm{ext}}\|_{\mathbb{H}^1(\mathrm{D})} \leq \varepsilon$. The results are presented in figure 3, which shows the dependence $N_l \sim \log^{\vartheta} \delta^{-1}$ with $\vartheta < 3$.

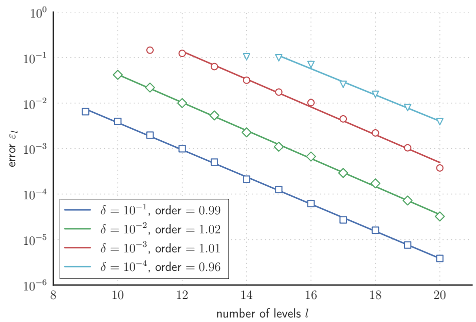

Figure 1: Error estimate $\varepsilon_l(\delta)$ (7.7) w.r.t the number l of levels, for various δ.

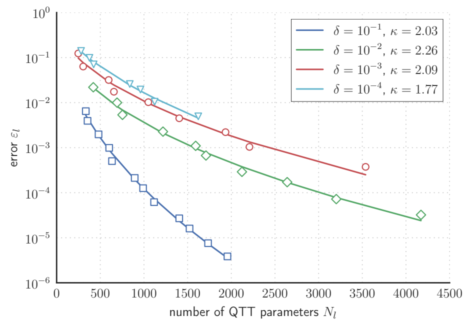

Figure 2: Error estimate $\varepsilon_l(\delta)$ (7.7) w.r.t the number l of levels, for various δ, plotted against the corresponding numbers N_l (3.11) of QTT parameters.

7.2 Helmholtz problem

As the second example, we consider the one-dimensional Helmholtz problem (2.1), (2.9) with radiating boundary conditions (2.10) and the right-hand side $f = 1$.

The $\mathbb{H}^1$-error of QTT-FE approximations is estimated as in section 7.1.1, by $\varepsilon_l(k)$ for all l and k.

As expected, the QTT-FE approximations we consider converge algebraically with respect to 2^l, which is the number of virtual degrees of freedom, see figure 4. With respect to the corresponding values of the number N_l (3.11)

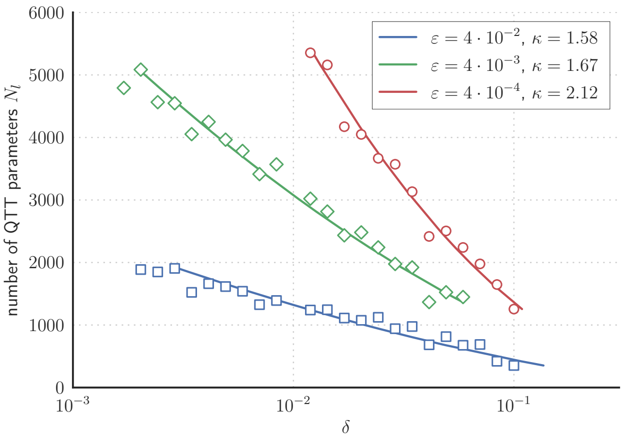

Figure 3: The numbers N_l (3.11) of QTT parameters w.r.t. the microscale parameter δ, for various values of the tolerance ε. For each data point, the number l of levels is set the minimal such that $\|u^{\delta,l} - u^{\delta}_{\mathrm{ext}}\|_{\mathbb{H}^1(\mathrm{D})} \leq \varepsilon/2$. The corresponding QTT approximation $u^{\delta,l}_{\mathrm{qtt}}$ of $u^{\delta,l}$ is obtained so that $\|u^{\delta,l}_{\mathrm{qtt}} - u^{\delta,l}\|_{\mathbb{H}^1(\mathrm{D})} \leq \varepsilon/2$, and hence $\|u^{\delta,l}_{\mathrm{qtt}} - u^{\delta}\|_{\mathbb{H}^1(\mathrm{D})} \approx \|u^{\delta,l}_{\mathrm{qtt}} - u^{\delta,l}_{\mathrm{ext}}\|_{\mathbb{H}^1(\mathrm{D})} \leq \varepsilon$. The reference lines indicate the dependence $N_l \sim \log^{\vartheta} \delta^{-1}$ with $\vartheta < 3$ independent of δ.

of QTT parameters, we observe the exponential convergence (7.1) with $\kappa < 3/2$, see figure (5).

We also study, in the same way as in section 7.1.3, the dependence of the error on the microscale parameter, k. The results are presented in figure 6, which illustrates the moderate dependence of the number of QTT parameters N_l (3.11) on the the wavenumber k: $N_l \sim \log^{\vartheta} k$ with $\vartheta < 1$, cf. the bound with $\vartheta = 3$ resulting from theorem 5.3 and the inequality of (3.11).

8 Conclusion

In the present paper, we proposed and analyzed a QTT-structured finite elements for two classes of model multiscale problems in one space dimension. For solutions with singularities, it was proven in [27, 23, 28] that the lowest-order FE approximation on uniform meshes (in the present setting, continuous and piecewise-linear FE) with classical, Lagrangian nodal basis functions and with QTT-formatted coefficient vectors achieves exponential convergence rates typ-

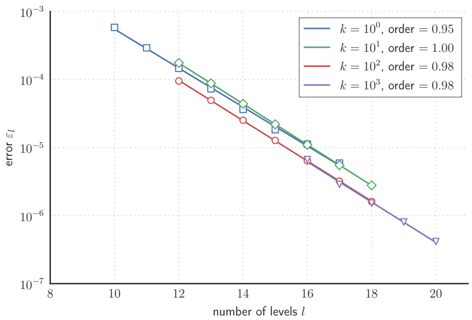

Figure 4: Error estimate $\varepsilon_l(k)$ (7.7) w.r.t the number l of levels, for various k.

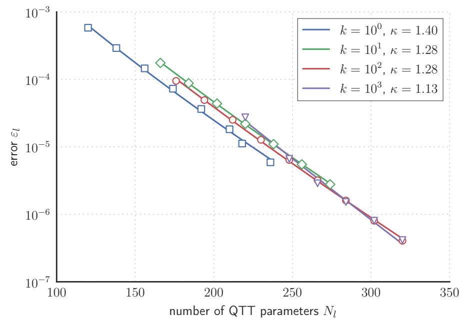

Figure 5: Error estimate $\varepsilon_l(k)$ (7.7) w.r.t the number l of levels, for various k, plotted against the corresponding numbers N_l (3.11) of QTT parameters.

ically afforded by *hp*-FEM Importantly, in QTT-formatted approximations of solutions, the mesh becomes *virtual* in the sense that it is never activated in full. Instead, nonlinear approximations extract, from this space and at runtime (i.e. "online"), low-parametric subspaces and manifolds which capture the solutions to the level of accuracy and resolution afforded by the finest FE mesh which could be virtually addressed in the adopted tensor formats.

For homogenization problems, we show that QTT-formatted low-order FE approximations *without additional features for microscale resolution* (such as numerical homogenization, HMM, oversampling, etc., which require, in some form, a-priori provision of information on the microstructure of problem and solution) achieve scale resolution and exponential convergence rates. In particular, a two-scale diffusion problem admits approximate QTT-FE solutions of accuracy $\mathcal{O}(\sqrt{\delta})$ in the energy norm and ranks $\mathcal{O}(\log^2 \delta^{-1})$, where δ is the scale parameter. We show similar results for the model Helmholtz problems, namely that they admit approximate QTT-FE solutions of accuracy $\mathcal{O}(k\,2^{-l})$ in the energy norm and ranks $\mathcal{O}(l)$, where k is the wavenumber and $l \in \mathbb{N}$ is the number of virtual levels underlying quantization.

In remarks 4.15 and 5.4, we noted the exponential smallness of the Kolmogorov n-widths of the solution sets, uniformly with respect to the scale parameter. By the arguments of [38, 47], these bounds imply in particular exponential convergence rates of reduced basis and model order reduction methods with respect to the subspace dimension n, *with constants which are bounded independently of the scale parameter.*

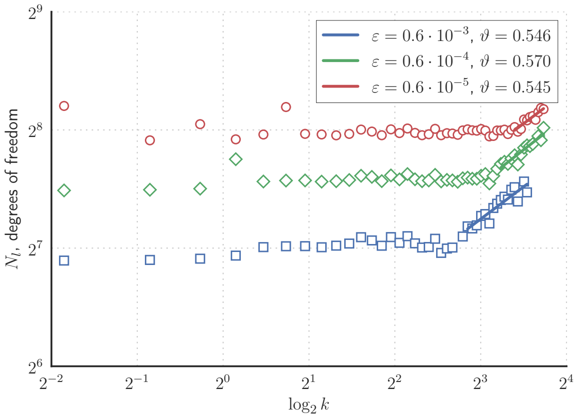

Figure 6: The numbers N_l (3.11) of QTT parameters w.r.t. the microscale parameter k, for various values of the tolerance ε. For each data point, the number l of levels is set the minimal such that $\|u^{k,l} - u^k_{\text{ext}}\|_{\mathbb{H}^1(\mathrm{D})} \leq \varepsilon/2$. The corresponding QTT approximation $u^{k,l}_{\text{qtt}}$ of $u^{k,l}$ is obtained so that $\|u^{k,l}_{\text{qtt}} - u^{k,l}\|_{\mathbb{H}^1(\mathrm{D})} \leq \varepsilon/2$, and hence $\|u^{k,l}_{\text{qtt}} - u^k\|_{\mathbb{H}^1(\mathrm{D})} \approx \|u^{k,l}_{\text{qtt}} - u^{k,l}_{\text{ext}}\|_{\mathbb{H}^1(\mathrm{D})} \leq \varepsilon$. The reference lines indicate the dependence $N_l \sim \log^\vartheta k$ with $\vartheta < 1$ independent of k.

The present results were developed in one space dimension. Two- and three-dimensional problems show, in numerical experiments, completely analogous behavior; the corresponding analysis will be presented elsewhere. All proofs given in the present paper generalize to the corresponding error bounds for space dimensions $d = 2, 3$ using tensorized meshes, where the techniques of [48, 6, 7] can be applied for singular problems to achieve similar, polylogarithmic in accuracy rank bounds for elliptic problems with corner singularities [23, 27, 28].

Since algorithmic realizations of QTT-structured low-order FEM can be based to a large extent on standard FEM with uniform meshes, the observations in the present paper offer a perspective of reproducing the performance of special methods (such as HMM and scale-resolving gFEM) in solving boundary-value problems for diffusion equations with multiple scales and for Helmholtz equations at large wavenumbers *without code redevelopment* for a range of problems of engineering interest.

Acknowledgment CS was supported by the European Research Council through the FP7 Advanced Grant AdG247277. The research of VK was performed at the Seminar for Applied Mathematics, ETH Zurich. IO and MR were supported by Russian Science Foundation grant 14-11-00659.

References

[1] Roman Andreev and Christine Tobler. Multilevel preconditioning and low-rank tensor iteration for space–time simultaneous discretizations of parabolic pdes. *Numerical Linear Algebra with Applications*, 22(2):317–337, 2015. 9

[2] Ivo Babuška. Error-bounds for finite element method. *Numerische Mathematik*, 16(4):322–333, 1971. 6

[3] Markus Bachmayr and Wolfgang Dahmen. Adaptive low-rank methods for problems on Sobolev spaces with error control in L_2. arXiv preprint 1412.3951, 2014. 9

[4] N. Bakhvalov and G. Panasenko. *Homogenisation: Averaging Processes in Periodic Media*, volume 36 of *Mathematics and its Applications*. Springer Netherlands. 13

[5] Jonas Ballani and Lars Grasedyck. A projection method to solve linear systems in tensor format. *Numerical Linear Algebra with Applications*, 2012. 9

[6] L. Beirão da Veiga, A. Buffa, J. Rivas, and G. Sangalli. Some estimates for h–p–k-refinement in isogeometric analysis. *Numerische Mathematik*, 118(2):271–305, 2011. 14, 34

[7] Annalisa Buffa, G. Sangalli, and Christoph Schwab. Exponential convergence of the hp version of isogeometric analysis in 1D, 2014. Proceedings. 34

[8] Philip J. Davis. *Interpolation and approximation*. Dover books on advanced mathematics. Dover Publications, 1975. 23

[9] S. Dolgov, B. Khoromskij, and I. Oseledets. Fast solution of parabolic problems in the tensor train/quantized tensor train format with initial application to the fokker–planck equation. *SIAM Journal on Scientific Computing*, 34(6):A3016–A3038, 2012. 9

[10] Sergey V. Dolgov, Vladimir A. Kazeev, and Boris N. Khoromskij. The tensor-structured solution of one-dimensional elliptic differential equations with high-dimensional parameters. Preprint 51, Max-Planck-Institut für Mathematik in den Naturwissenschaften, 2012. 2, 27

[11] Sergey V. Dolgov and Boris N. Khoromskij. Tensor-product approach to global time-space-parametric discretization of chemical master equation. Preprint 68, Max-Planck-Institut für Mathematik in den Naturwissenschaften, November 2012. 9

[12] Sergey V. Dolgov and Dmitry V. Savostyanov. Alternating minimal energy methods for linear systems in higher dimensions. *SIAM Journal on Scientific Computing*, 36(5):A2248–A2271, 2014. 29

[13] Lars Grasedyck. Hierarchical Singular Value Decomposition of tensors. *SIAM Journal on Matrix Analysis and Applications*, 31(4):2029–2054, 2010. 11

[14] Lars Grasedyck. Polynomial approximation in Hierarchical Tucker Format by vector-tensorization. Preprint 308, Institut für Geometrie und Praktische Mathematik, RWTH Aachen, April 2010. 11

[15] Lars Grasedyck, Daniel Kressner, and Christine Tobler. A literature survey of low-rank tensor approximation techniques. *GAMM-Mitteilungen*, 36(1):53–78, 2013. 2, 9

[16] W. Hackbusch and S. Kühn. A new scheme for the tensor representation. *Journal of Fourier Analysis and Applications*, 15(5):706–722, 2009. 10.1007/s00041-009-9094-9. 11

[17] Wolfgang Hackbusch. *Tensor Spaces and Numerical Tensor Calculus*, volume 42 of *Springer Series in Computational Mathematics*. Springer, 2012. 11

[18] Viet Ha Hoang and Christoph Schwab. High-dimensional finite elements for elliptic problems with multiple scales. *Multiscale Modeling & Simulation*, 3(1):168–194, 2005. 29

[19] 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(2):A683–A713, 2012. 9

[20] Frank Ihlenburg. *Finite element analysis of acoustic scattering*, volume 132 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1998. 5, 23

[21] V. V. Jikov, S. M. Kozlov, and O. A. Oleinik. *Homogenization of Differential Operators and Integral Functionals*. Springer-Verlag, 1994. 13

[22] Ho Tho Kau. An inequality for algebraic polynomials, and the dependence between the best polynomial approximations $e(f)_{L_p}$ and $e(f)_{L_q}$ of functions $f(x) \in L_p$ (in Russian). *Acta Mathematica Academiae Scientiarum Hungarica*, 27(1-2):141–147, 1976. 15

[23] Vladimir Kazeev. *Quantized tensor-structured finite elements for second-order elliptic PDEs in two dimensions*. PhD thesis, SAM, ETH Zurich, ETH Dissertation No. 23002. 2, 32, 34

[24] Vladimir Kazeev, Mustafa Khammash, Michael Nip, and Christoph Schwab. Direct solution of the Chemical Master Equation using Quantized Tensor Trains. *PLOS Computational Biology*, 10(3), March 2014. 9

[25] Vladimir Kazeev, Oleg Reichmann, and Christoph Schwab. *hp*-DG-QTT solution of high-dimensional degenerate diffusion equations. Research Report 11, Seminar for Applied Mathematics, ETH Zürich, 2012. 2

[26] Vladimir Kazeev, Oleg Reichmann, and Christoph Schwab. Low-rank tensor structure of linear diffusion operators in the TT and QTT formats. *Linear Algebra and its Applications*, 2013. 11

[27] Vladimir Kazeev and Christoph Schwab. Approximation of singularities by quantized-tensor FEM. *PAMM*, 15(1):743–746, 2015. 2, 32, 34

[28] Vladimir Kazeev and Christoph Schwab. Quantized tensor-structured finite elements for second-order elliptic PDEs in two dimensions. Research Report 24, Seminar for Applied Mathematics, ETH Zürich, 2015. 2, 14, 15, 32, 34

[29] Vladimir Kazeev and Christoph Schwab. Tensor approximation of stationary distributions of chemical reaction networks. *SIAM Journal on Matrix Analysis and Applications*, 36(3):1221–1247, 2015. 9, 21

[30] Vladimir A. Kazeev and Boris N. Khoromskij. Low-rank explicit QTT representation of the Laplace operator and its inverse. *SIAM Journal on Matrix Analysis and Applications*, 33(3):742–758, 2012. 11, 27, 28

[31] Vladimir A. Kazeev, Boris N. Khoromskij, and Eugene E. Tyrtyshnikov. Multilevel Toeplitz matrices generated by tensor-structured vectors and convolution with logarithmic complexity. *SIAM Journal on Scientific Computing*, 2013. 11

[32] B. N. Khoromskij. $\mathcal{O}(d \log n)$-quantics approximation of n-d tensors in high-dimensional numerical modeling. *Constructive Approximation*, 34(2):257–280, 2011. 10.1007/s00365-011-9131-1. 2, 9, 11, 12

[33] B. N. Khoromskij, V. Khoromskaia, and H.-J. Flad. Numerical solution of the Hartre-Fock equation in multilevel tensor-structured format. *SIAM Journal on Scientific Computing*, 33(1):45–65, 2011. 2

[34] B. N. Khoromskij and Ch. Schwab. Tensor-structured Galerkin approximation of parametric and stochastic elliptic PDEs. *SIAM Journal on Scientific Computing*, 33(1):364–385, 2011. 2

[35] Boris N. Khoromskij and Sergey I. Repin. A fast iteration method for solving elliptic problems with quasiperiodic coefficients. arXiv preprint 1510.00284, October 2015. 2

[36] Daniel Kressner, Michael Steinlechner, and André Uschmajew. Low-rank tensor methods with subspace correction for symmetric eigenvalue problems. Technical report 40, MATHICSE EPFL, 2013. 9

[37] Daniel Kressner, Michael Steinlechner, and Bart Vandereycken. A fast iteration method for solving elliptic problems with quasiperiodic coefficients. arXiv preprint 1508.02988, August 2015. 9

[38] Y. Maday, O. Mula, and G. Turinici. Convergence analysis of the generalized empirical interpolation method. Technical report, HAL-UPMC. 33

[39] R. J. Nessel and G. Wilmes. Nikolskii-type inequalities for trigonometric polynomials and entire functions of exponential type. *Journal of the Australian Mathematical Society (Series A)*, 25:7–18, 2 1978. 17

[40] I. Oseledets. Approximation of matrices with logarithmic number of parameters. *Doklady Mathematics*, 80:653–654, 2009. 2, 9

[41] I. Oseledets and S. Dolgov. Solution of linear systems and matrix inversion in the TT-format. *SIAM Journal on Scientific Computing*, 34(5):A2718–A2739, 2012. 29

[42] I. V. Oseledets. Approximation of $2^d \times 2^d$ matrices using tensor decomposition. *SIAM Journal on Matrix Analysis and Applications*, 31(4):2130–2145, 2010. 9

[43] I. V. Oseledets. Tensor Train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011. 7, 8, 9, 11

[44] 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(5):3744–3759, October 2009. 7

[45] I.V. Oseledets. Constructive representation of functions in low-rank tensor formats. *Constructive Approximation*, 37(1):1–18, 2013. 11, 12

[46] Allan Pinkus. *n-widths in approximation theory*, volume 7 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1985. 22

[47] Alfio Quarteroni, Andrea Manzoni, and Federico Negri. *Reduced Basis Methods for Partial Differential Equations*, volume 92 of *UNITEXT*. Springer, 2016. 33

[48] Ch. Schwab. *p- and hp-FEM: Theory and Application to Solid and Fluid Mechanics*. Oxford University Press, Oxford, 1998. 14, 34

[49] E. E. Tyrtyshnikov. Tensor approximations of matrices generated by asymptotically smooth functions. *Sbornik: Mathematics*, 194(5):941–954, 2003. 10

[50] F. Verstraete, D. Porras, and J. I. Cirac. Density matrix renormalization group and periodic boundary conditions: A quantum information perspective. *Phys. Rev. Lett.*, 93(22):227205, November 2004. 11

[51] Guifré Vidal. Efficient classical simulation of slightly entangled quantum computations. *Phys. Rev. Lett.*, 91(14):147902, October 2003. 11

[52] Steven R. White. Density-matrix algorithms for quantum renormalization groups. *Phys. Rev. B*, 48(14):10345–10356, October 1993. 11

Recent Research Reports

Nr.	Authors/Title
2015-43	Ch. Schwab and R. Stevenson Fractional space-time variational formulations of (Navier-) Stokes equations
2015-44	S. May and M. Berger An Explicit Implicit Scheme for Cut Cells in Embedded Boundary Meshes
2015-45	L. Jacobe de Naurois and A. Jentzen and T. Welti Weak convergence rates for spatial spectral Galerkin approximations of semilinear stochastic wave equations with multiplicative noise
2015-46	A. Jentzen and R. Kurniawan Weak convergence rates for Euler-type approximations of semilinear stochastic evolution equations with nonlinear diffusion coefficients
2015-47	A. Andersson and A. Jentzen and R. Kurniawan Existence, uniqueness, and regularity for stochastic evolution equations with irregular initial values
2016-01	A. Jentzen and Th. Müller-Gronbach and L. Yaroslavtseva On stochastic differential equations with arbitrary slow convergence rates for strong approximation
2016-02	L. Herrmann and A. Lang and Ch. Schwab Numerical Analysis of Lognormal Diffusions on the Sphere
2016-03	W. Dahmen and G. Kutyniok and W. Lim and Ch. Schwab and G. Welper Adaptive Anisotropic Petrov-Galerkin Methods for First Order Transport Equations
2016-04	A. Obermeier and P. Grohs On the Approximation of Functions with Line Singularities by Ridgelets
2016-05	D. Schoetzau and Ch. Schwab Exponential Convergence of hp-FEM for Elliptic Problems in Polyhedra: