Foundations of Computational Mathematics (2020) 20:1175–1236
https://doi.org/10.1007/s10208-020-09446-z

FOUNDATIONS OF COMPUTATIONAL MATHEMATICS
The Journal of the Society for the Foundations of Computational Mathematics

Stability of Low-Rank Tensor Representations and Structured Multilevel Preconditioning for Elliptic PDEs

Markus Bachmayr[1] · **Vladimir Kazeev**[2]

Received: 5 March 2018 / Revised: 18 June 2019 / Accepted: 5 November 2019 /
Published online: 23 January 2020

Abstract

Folding grid value vectors of size 2^L into Lth-order tensors of mode size $2 \times \cdots \times 2$, combined with low-rank representation in the tensor train format, has been shown to result in highly efficient approximations for various classes of functions. These include solutions of elliptic PDEs on nonsmooth domains or with oscillatory data. This tensor-structured approach is attractive because it leads to highly compressed, adaptive approximations based on simple discretizations. Standard choices of the underlying bases, such as piecewise multilinear finite elements on uniform tensor product grids, entail the well-known *matrix ill-conditioning* of discrete operators. We demonstrate that, for low-rank representations, the use of tensor structure itself additionally introduces *representation ill-conditioning*, a new effect specific to computations in tensor networks. We analyze the tensor structure of a BPX preconditioner for a second-order linear elliptic operator and construct an explicit tensor-structured representation of the preconditioner, with ranks independent of the number L of discretization levels. The straightforward application of the preconditioner yields discrete operators whose matrix conditioning is uniform with respect to the discretization parameter, but in decompositions that suffer from representation ill-conditioning. By additionally eliminating certain redundancies in the representations of the preconditioned discrete operators, we obtain reduced-rank decompositions that are free of both matrix and representation ill-conditioning. For an iterative solver based on soft thresholding of low-rank tensors, we obtain convergence and complexity estimates and demonstrate its reliability and efficiency for discretizations with up to 2^{50} nodes in each dimension.

Keywords Elliptic boundary value problems · Multilevel preconditioning · Tensor decompositions · Representation condition number · Solver complexity

Communicated by Endre Süli.

M.B. acknowledges support by the Hausdorff Center of Mathematics, University of Bonn. M.B. was partially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Projektnummer 211504053 - SFB 1060.

Extended author information available on the last page of the article

Mathematics Subject Classification 15A69 · 35J25 · 65N12 · 65N30 · 65N55 · 65F08 · 65F35 · 65Y20

1 Introduction

The direct textbook treatment of elliptic PDEs by low-order discretizations on uniform grids becomes unaffordable for many important problem classes. The high computational costs are due to the prohibitively large number of degrees of freedom required to resolve specific features of solutions, such as singularities and high-frequency oscillations, that arise in problems with nonsmooth or oscillatory data. More efficient discretizations can be obtained with basis functions that are adapted to the given problem and require fewer degrees of freedom. However, the construction and analysis of such methods (for instance, of hp-adaptive solvers) generally depends on specific features of the considered problem classes and accordingly specialized analytical tools.

By the approach considered in this work, efficiency is achieved in a different way: extremely large arrays of coefficients parametrizing simple, uniformly refined low-order discretizations are themselves parametrized as nonlinear functions of relatively few effective degrees of freedom. The latter parametrization is based on representing the coefficient arrays, reshaped into high-order tensors, in the *tensor train* decomposition with low ranks. This representation exploits low-rank structure with respect to a hierarchy of dyadic scales, providing, at each scale, a problem-adapted basis that can be computed using standard techniques of numerical linear algebra. In other words, for the identification of suitable degrees of freedom, this approach avoids relying on problem-specific *a priori* information; instead, suitable degrees of freedom are found by the low-rank tensor compression of generic, conceptually straightforward discretizations.

In numerical solvers for PDE problems that operate on such highly compressed, nonlinear representations of basis coefficients, new difficulties arise compared to a standard entrywise representation. As we demonstrate in this contribution, specific types of ill-conditioning in such tensor representations can dramatically affect the numerical stability of solvers. We show how a special low-rank representation of a BPX preconditioner allows to overcome these difficulties and obtain estimates for the total computational complexity of computing solutions with low-rank tensor train structure.

1.1 Low-Rank Tensor Approximations

The development of low-rank tensor representations [18,25,45,47,50], such as the *tensor train* format, has originally been motivated by applications to high-dimensional PDEs. As observed in [19,37,43,44], the artificial treatment of coefficient vectors in lower-dimensional problems as high-dimensional quantities, known in the literature as *quantized tensor train* (QTT) decomposition or *tensorization*, leads to highly efficient approximations in many problems of interest. See [38] for a general overview and, for instance, [29,36] for further applications.

To briefly illustrate this concept, let us suppose that a function u has an accurate approximation $u \approx \sum_{j=1}^{N} \boldsymbol{u}_j \phi_j$ in terms of the basis functions $\{\phi_j\}_{j=1,\dots,N}$ with the coefficient vector $\boldsymbol{u} = (\boldsymbol{u}_j)_{j=1,\dots,N} \in \mathbb{R}^N$. The basic idea is to reinterpret $\boldsymbol{u}$ as a higher-order tensor of mode sizes $n_1 \times \cdots \times n_L$ with $\prod_{\ell=1}^{L} n_\ell = N$ via the identification

$$j \quad \leftrightarrow \quad (i_1, \dots, i_L) \in \{0, \dots, n_1 - 1\} \times \cdots \times \{0, \dots, n_L - 1\}$$

provided by the unique decomposition

$$j - 1 = \sum_{\ell=1}^{L} i_\ell \prod_{k=\ell+1}^{L} n_k \quad \text{with} \quad i_\ell \in \{0, \dots, n_\ell - 1\} \quad \text{for all} \quad \ell = 1, \dots, L\,.$$

We assume a simple choice of basis functions, such as low-order splines, combined with a compressed, nonlinearly parametrized approximation of the corresponding coefficients $\boldsymbol{u}$ in the tensor train format,

$$\boldsymbol{u}_{i_1,\dots,i_L} \approx \sum_{\alpha_1=1}^{r_1} \cdots \sum_{\alpha_{L-1}=1}^{r_{L-1}} U_1(1, i_1, \alpha_1)\, U_2(\alpha_1, i_2, \alpha_2) \cdots U_L(\alpha_{L-1}, i_L, 1). \quad (1)$$

The actual degrees of freedom are now the entries of the third-order tensors $U_\ell \in \mathbb{R}^{r_{\ell-1} \times n_\ell \times r_\ell}$ with $\ell \in \{1, \dots, L\}$, which are referred to as *cores* (where $r_0 = r_L = 1$ for notational convenience). In the case of $n_\ell = n \in \mathbb{N}$ for all ℓ, which we consider in this work, the total number of parameters defining this approximation equals $\sum_{\ell=1}^{L} n_\ell\, r_{\ell-1}\, r_\ell \lesssim (\log N) \max\{r_1^2, \dots, r_{L-1}^2\}$.

For certain representative approximation problems (such as functions with isolated singularities or high-frequency oscillations), as shown in [19,31,34,37], one obtains approximations where the rank parameters $r_1, \dots, r_{L-1}$ grow at most polylogarithmically in the corresponding error. This suggests the possibility of constructing numerical methods with complexity scaling as $(\log N)^\alpha$ for a fixed α.

1.2 Multilevel Low-Rank Approximations for Elliptic Boundary Value Problems

In this work, we focus on the application of low-rank tensor techniques for solving second-order elliptic boundary value problems on domains $\Omega \subset \mathbb{R}^D$, where we are mainly interested in the cases of $D \in \{1, 2, 3\}$. First, consider the exact solution u and finite element solutions u_h, where $h > 0$ is a mesh-size parameter, that are simple, low-order finite element functions with coefficient vectors $\boldsymbol{u}_h$. These are given by suitable linear systems of the form $\boldsymbol{A}_h \boldsymbol{u}_h = \boldsymbol{f}_h$. For each mesh size h, one can seek instead u_h^{LR} from the same finite element space whose coefficient vector $\boldsymbol{u}_h^{\mathrm{LR}}$ is a low-rank approximation in form (1) of $\boldsymbol{u}_h$. In order to benefit from the complexity reduction afforded by representation (1), the vector $\boldsymbol{u}_h^{\mathrm{LR}}$ needs to be computed directly in this low-rank representation. Using corresponding representations of $\boldsymbol{A}_h$ and $\boldsymbol{f}_h$, this can be achieved by iteratively solving the nonlinear problem in terms of the cores

$U_1, \dots, U_L$ of $\boldsymbol{u}_h^{\mathrm{LR}}$ in (1). In our setting, the binary indexing $(i_1, \dots, i_L)$ used in the interpretation of $\boldsymbol{u}_h$ as a tensor of order L corresponds to uniform grid refinement with L levels, and thus $h \sim 2^{-L}$. The separation of variables expressed by (1) therefore applies not to the spatial dimensions but rather to the dyadic scales of u_h^{LR}.

In our model problem, the underlying discretization uses piecewise D-linear finite elements. Using the triangle inequality, we can decompose the error $u - u_h^{\mathrm{LR}}$ into a discretization error $u - u_h$, for which on uniform meshes one obtains bounds of the form

$$\|u - u_h\|_{\mathrm{H}^1} \leq C_u h^s, \tag{2}$$

with $C_u > 0$ depending only on u and $0 < s \leq 1$, and the computation error $u_h - u_h^{\mathrm{LR}}$ including the error of low-rank approximation. In problems where u exhibits, for instance, singularities or high-frequency oscillations, one may be dealing with C_u extremely large or with $s \ll 1$. Thus, achieving reasonable total errors may require values of h that are so small that the entrywise representations of coefficient vectors and matrices are computationally infeasible.

Under natural assumptions on the data and on the underlying mesh, the problem of finding u_h remains well-conditioned with respect to the problem data independently of h. However, for very small h as considered here, it becomes a nontrivial issue to ensure numerical stability of algorithms, since these are affected by the condition numbers $\mathcal{O}(h^{-2})$ of $\boldsymbol{A}_h$. Regardless of the type of solver that is employed, preconditioning $\boldsymbol{A}_h$ becomes a necessity for avoiding numerical instabilities even for moderately small h. As a first step, we therefore construct a preconditioner for $\boldsymbol{A}_h$ that can be applied directly in low-rank form, where both the resulting matrix condition numbers after preconditioning and the tensor representation ranks are uniformly bounded with respect to the discretization level L.

However, we also find that when such a preconditioner is applied as usual by the standard matrix-vector multiplication in the tensor format, numerical solvers *still* stagnate at an error $\|u_h - u_h^{\mathrm{LR}}\|_{\mathrm{H}^1}$ of order $\mathcal{O}(h^{-2}\varepsilon)$, where ε is the machine precision. This shows that ensuring uniformly bounded matrix condition numbers by preconditioning is not sufficient for low-rank tensor methods to remain numerically stable for very small h. It turns out that tensor representations of vectors in the form of (1) generated by the action of $\boldsymbol{A}_h$ can be extremely sensitive to perturbations of each single core. This new type of ill-conditioning cannot be eradicated by simply multiplying by the preconditioner, and any further numerical manipulations of the resulting tensor representations are prone to large round-off errors. To quantify this effect, we introduce the notion of *representation condition numbers*.

Without addressing the issue of representation ill-conditioning, one can therefore only expect $\|u - u_h\|_{\mathrm{H}^1} = \mathcal{O}(C_u h^s + h^{-2}\varepsilon)$. With the optimal choice of h, this yields a total error of order $\mathcal{O}(C_u^{2/(2+s)} \varepsilon^{s/(2+s)})$; even in the ideal case $s = 1$, one thus has a limitation to $\mathcal{O}(\varepsilon^{1/3})$. In the present paper, by analytically combining the low-rank representations of the preconditioner and of the stiffness matrix, we obtain a tensor representation that retains favorable representation condition numbers also for large L and leads to solvers that remain numerically stable even for h on the order of the

machine precision ε. For the problems preconditioned in this manner, we can apply results from [4,6] to obtain bounds for the number of operations required for computing $\boldsymbol{u}_h^{\mathrm{LR}}$, in terms of the ranks of low-rank *best* approximations of $\boldsymbol{u}_h$ with the same error. Since the costs depend only weakly on the discretization level L, one may then in fact simply choose L so large that $h \approx \varepsilon$. This ensures that the discretization error $\|u - u_h\|_{\mathrm{H}^1}$ is negligible in all practical situations and only the explicitly controllable low-rank approximation error $\|u_h - u_h^{\mathrm{LR}}\|_{\mathrm{H}^1}$ remains.

1.3 Conditioning of Tensor Train Representations

Let us now briefly outline the source of numerical instability that we need to mitigate here. Subspace-based tensor decompositions such as the Tucker format, hierarchical tensors [25] or the presently considered tensor train format [45] share the basic stability property that the existence of low-rank best approximations with fixed-rank parameters is guaranteed. In contrast, such best approximation problems for canonical tensors are in general ill-posed [14], and one has the well-known border rank phenomena where given tensors can be approximated arbitrarily well by tensors of lower canonical ranks. In subspace-based formats, such pathologies of the canonical rank are avoided by working only with matrix ranks of certain tensor matricizations. This leads to natural higher-order generalizations of the singular value decomposition (SVD), in particular the TT-SVD algorithm for tensor trains.

However, when performing computations in such tensor formats, tensors in general do not remain in orthogonalized standard representations, such as those given by the TT-SVD. For instance, the action of low-rank representations of finite element stiffness matrices in iterative solvers may create tensor train representations with substantial redundancies that are far from their respective SVD forms. A return to the rank-reduced SVD form can then in principle be accomplished by applying standard linear algebra operations (such as QR decomposition and SVD) to the representation components.

As we demonstrate in what follows, in relevant cases, tensor train representations can become so ill-conditioned that performing this rank reduction with machine precision no longer produces useful results. To our knowledge, this particular point has not received attention in the literature so far. As we consider in further detail in Sect. 4, a particular instance where this effect occurs is multilevel low-rank representations of discretization matrices of differential operators.

In order to illustrate these issues, let us consider a low-rank matrix $M = AB^{\mathsf{T}}$ with $A \in \mathbb{R}^{m\times r}$ and $B \in \mathbb{R}^{n\times r}$. Performing numerical manipulations of A, for instance a QR factorization with machine precision, amounts to replacing M by $\tilde{M} = \tilde{A}B^{\mathsf{T}}$ with $\|A - \tilde{A}\|_F \leq \delta\|A\|_F$, where δ will ideally be close to the relative machine precision. Similarly to standard perturbation estimates for matrix products (see, e.g., [27, Sec. 3]), one obtains the generally sharp worst-case bound

$$\|M - \tilde{M}\|_F \leq \delta\|A\|_F\|B\|_{2\to 2}.$$

In the case of high-order tensor train representations, one may think of B as composed of many individual cores. Even when each of these cores looks completely innocent,

their cumulative effect can lead to very large $\|B\|_{2\to 2}$. In cases where cancellations occur in the product with A, the size of $\|M\|_F$, however, can be small compared to $\|B\|_{2\to 2}$, and perturbations to A are strongly amplified. This means that any numerical manipulation of such representations (such as orthogonalization, which is also the first step in performing a TT-SVD, see Sect. 3.6) can introduce extremely large errors in the represented tensor.

We define the representation condition number of an operator in low-rank representation as the factor by which its action may deteriorate the conditioning of tensor train representations. In the case of the finite element stiffness matrices $\boldsymbol{A}_h$, we find that this condition number scales (matching the standard matrix condition number) as $\mathcal{O}(h^{-2})$, which agrees with the numerically observed loss of precision. One may regard this as a tensor-decomposition analogue of the classical amplification of relative errors by ill-conditioned matrices. However, this error amplification manifests itself not in the action of the tensor representation of $\boldsymbol{A}_h$ on any single tensor core, which by itself is harmless, but rather in the *cumulative* effect that emerges when further operations are performed on the resulting output cores.

1.4 Novelty and Relation to Previous Work

As a main contribution of this work, we introduce basic notions and auxiliary results for studying the representation conditioning of tensor train representations. In particular, our finding that the stiffness matrix represented in low-rank format has a representation condition number of order 2^{2L} explains numerical instabilities in its direct application for large L as observed in tests in [11]. We prove a new result on a BPX preconditioner for second-order elliptic problems that is tailored to our purposes, and we construct a low-rank decomposition of the preconditioned stiffness matrix with the following properties: it is well-conditioned uniformly in discretization level L as a matrix; its ranks are independent of L; and its representation condition numbers remain moderate for large L. Based on these properties, we establish an estimate for the total computational complexity of finding approximate solutions in low-rank form. These complexity bounds are shown for an iterative solver based on the soft thresholding of tensors [6], for which the ranks of approximate solutions can be estimated in terms of the ranks of the exact Galerkin solution. We identify appropriate approximability assumptions on solutions in the present context, which are slightly different from those proved in [34].

Difficulties with the numerical stability of solvers for large L have also been noted previously in [34]. In [11,46], a reformulation as a constrained minimization problem with Volterra integral operators is proposed. It is demonstrated numerically in [11] up to $L \approx 20$ to lead to improved numerical stability, compared to a direct finite difference discretization, for Poisson-type problems with $D = 2$ dimensions. However, in this reformulation, which so far has been studied only experimentally, the matrix condition number still grows exponentially with respect to L, and numerical stability is still observed to be lacking for larger values of L.

A different class of preconditioners based on approximate matrix exponentials has been proposed for QTT decompositions in [39]. In the different context of separation

of spatial coordinates in high-dimensional problems, tensor representations have been combined with multilevel preconditioners based on multigrid methods [8,23], BPX preconditioners [1] and wavelet Riesz bases [5]. There the required representation ranks of preconditioners have been observed to increase with discretization levels, in contrast to the uniformly bounded ranks that we obtain in our present setting of tensor separation between scales.

1.5 Outline

In Sect. 2, we consider the structure of discretization matrices in detail and establish a result on symmetric BPX preconditioning. In Sect. 3, we recapitulate basic notation and operations for the tensor train format. In Sect. 4, we introduce notions of representation condition numbers of tensor decompositions and investigate some of their basic properties. Building on these concepts, in Sect. 5 we construct well-conditioned multilevel low-rank representations of preconditioned discretization matrices. In Sect. 6, we discuss the implications of our findings on the complexity of finding approximate solutions and illustrate the performance of numerical solvers in Sect. 7.

We use the following general notational conventions: $A \lesssim B$ denotes $A \leq CB$ with C independent of any parameters explicitly appearing in the expressions A and B, and $A \sim B$ denotes $A \lesssim B \wedge A \gtrsim B$. We use $\|\cdot\|_2$ to denote the ℓ^2-norm both of vectors and of higher-order tensors, and $\|\cdot\|_{2\to 2}$ to denote the associated operator norm. In addition, $\|\cdot\|_F$ denotes the Frobenius norm of matrices. By $\langle\cdot,\cdot\rangle$, we denote the ℓ^2-inner product of vectors and tensors or the L^2-inner product of functions, as well as the corresponding duality product.

2 Discretization and Preconditioning

The model problem that we focus on in what follows is posed on the product domain $\Omega = \hat{\Omega}^D \subset \mathbb{R}^D$ with $\hat{\Omega} = (0,1)$. With $\Gamma = \{x \in \partial\Omega : x_1 \cdots x_D = 0\}$, we consider the corresponding Sobolev space of functions defined on Ω and vanishing on Γ,

$$V = \{v \in \mathrm{H}^1(\Omega) : v|_\Gamma = 0\}, \tag{3}$$

with norm $\|v\|_V = \|v\|_{\mathrm{H}^1_0(\Omega)} \sim \|v\|_{\mathrm{H}^1(\Omega)}$. On this space, we consider the variational problem

$$\text{find } u \in V \text{ such that } \quad a(u,v) = f(v) \quad \text{for all} \quad v \in V, \tag{4}$$

where $a : V \times V \to \mathbb{R}$ is the bilinear form given by

$$a(w,v) = \int_\Omega (\nabla v)^{\mathsf{T}} A \nabla w + \int_\Omega c\, v\, w \quad \text{for all} \quad w, v \in V, \tag{5}$$

and $f \in V'$ is a given linear form. We assume the diffusion and reaction coefficients $A \in \mathrm{L}^\infty\big(\Omega, \mathbb{R}^{D\times D}\big)$ and $c \in \mathrm{L}^\infty(\Omega)$ to be strongly elliptic and nonnegative,

respectively:

$$\underline{A} = \operatorname*{ess\,inf}_{\Omega} \inf_{\xi \in \mathbb{R}^D \setminus \{0\}} \frac{\xi^{\mathsf{T}} A \xi}{\xi^{\mathsf{T}} \xi} > 0 \quad \text{and} \quad c \geq 0 \quad \text{a.e. on } \Omega\,.$$

Problem (4) is a variational formulation of a boundary value problem for a reaction-diffusion equation with homogeneous mixed boundary conditions: of Dirichlet type on Γ and of Neumann type on $\partial\Omega \setminus \Gamma$.

Under the assumptions on the data made so far, the bilinear form a is continuous and coercive and the linear form f is continuous. By the Lax–Milgram theorem, (4) has a unique solution satisfying

$$\|u\|_V \leq \underline{A}^{-1} \|f\|_{V'}\,. \tag{6}$$

Additional assumptions on the data of problem (4), essential for its tensor-structured preconditioning and solution, are stated in Sects. 2 and 5.

In what follows, we consider a hierarchy of discretizations based on piecewise D-linear nodal basis functions on a sequence of uniform grids with cell sizes $2^{-\ell} \times \cdots \times 2^{-\ell}$, $\ell = 0, 1, 2, \ldots$; the basis functions can be written as tensor products of standard univariate hat functions.

In this section, we describe V_ℓ with $\ell \in \mathbb{N}_0$, nested finite-dimensional subspaces of V introduced in (3). We will use these subspaces to approximate the solution of the variational problem stated in (4).

2.1 Finite Element Spaces for $\hat{\Omega} = (0, 1)$

Throughout this section, we assume that an arbitrary number $\ell \in \mathbb{N}_0$ of refinement levels is fixed. We consider a uniform partition of $\hat{\Omega}$ into 2^ℓ subintervals and corresponding 2^ℓ continuous piecewise linear functions defined on $\hat{\Omega}$. Then, by tensorization, we introduce basis functions defined on Ω.

First, we consider the uniform partition of $\hat{\Omega}$ that consists of the 2^ℓ intervals

$$\hat{\Omega}_{\ell,i} = (\hat{\tau}_{\ell,i-1}, \hat{\tau}_{\ell,i}) \quad \text{with} \quad i \in \hat{\mathscr{I}}_\ell = \{1, \ldots, 2^\ell\} \tag{7}$$

given by the $2^\ell + 1$ nodes

$$\hat{\tau}_{\ell,j} = 2^{-\ell} j \quad \text{with} \quad j = 0, \ldots, 2^\ell\,. \tag{8}$$

For each $i \in \hat{\mathscr{I}}_\ell$, we introduce an affine mapping $\hat{\phi}_{\ell,i}$ from $(-1, 1)$ onto $\hat{\Omega}_{\ell,i}$:

$$\hat{\phi}_{\ell,i}(t) = \frac{1}{2}(\hat{\tau}_{\ell,i} + \hat{\tau}_{\ell,i-1}) + \frac{t}{2}(\hat{\tau}_{\ell,i} - \hat{\tau}_{\ell,i-1}) = 2^{-\ell} i + 2^{-\ell-1}(t - 1) \tag{9}$$

for all $t \in (-1, 1)$.

Further, we consider nodal functions defined on $\hat{\Omega}$ and associated with these nodes: for each $j \in \hat{\mathscr{J}}_\ell$, by $\hat{\varphi}_{\ell,j}$ we denote the function that is linear on each $\hat{\Omega}_{\ell,i}$ with $i \in \hat{\mathscr{J}}_\ell$, continuous on $\hat{\Omega}$ and such that

$$\hat{\varphi}_{\ell,j}(\hat{\tau}_{\ell,j'}) = 2^{\frac{\ell}{2}} \delta_{jj'} \quad \text{for all} \quad j' = 0, \dots, 2^\ell . \tag{10}$$

The ℓ-dependent normalization factor in the right-hand side of (10) results in the uniform normalization

$$\|\hat{\varphi}_{\ell,j}\|_{\mathrm{L}^2(\hat{\Omega})} \sim 1 . \tag{11}$$

By the above construction of basis functions, each $\hat{\varphi}_{\ell,j}$ with $j \in \hat{\mathscr{J}}_\ell$ is a degree-one polynomial on every $\hat{\Omega}_{\ell,i}$ with $i \in \hat{\mathscr{J}}_\ell$. This implies that, for $\alpha = 0, 1$, there exist matrices $\hat{\boldsymbol{M}}_{\ell,\alpha}$ with rows and columns indexed by $\hat{\mathscr{J}}_\ell \times \{\alpha, 1\}$ and $\hat{\mathscr{J}}_\ell$, respectively, such that

$$\partial^\alpha \hat{\varphi}_{\ell,j} \circ \hat{\phi}_{\ell,i} = \sum_{\beta=\alpha,1} (\hat{\boldsymbol{M}}_{\ell,\alpha})_{i\beta\, j}\ \hat{\psi}_\beta \quad \text{on} \quad (-1, 1) \tag{12a}$$

for all $i, j \in \hat{\mathscr{J}}_\ell$, where $\hat{\psi}_0$ and $\hat{\psi}_1$ are the standard monomials of degree zero and one,

$$\hat{\psi}_0(t) = 1 \quad \text{and} \quad \hat{\psi}_1(t) = t \quad \text{for all} \quad t \in (-1, 1) . \tag{12b}$$

We note that the matrix $\hat{\boldsymbol{M}}_{\ell,0}$ is rectangular of size $2^{\ell+1} \times 2^\ell$ and the matrix $\hat{\boldsymbol{M}}_{\ell,1}$ is a square matrix of order 2^ℓ.

For the basis functions defined in (12b), since $\hat{\psi}_1' = \hat{\psi}_0$, the odd rows of $\hat{\boldsymbol{M}}_{\ell,0}$ form a multiple of $\hat{\boldsymbol{M}}_{\ell,1}$: for $\beta = 1$ and all $i, j \in \hat{\mathscr{J}}_\ell$, we have

$$(\hat{\boldsymbol{M}}_{\ell,1})_{i\beta\, j} = 2^{\ell+1} (\hat{\boldsymbol{M}}_{\ell,0})_{i\beta\, j} . \tag{12c}$$

Furthermore, the matrices $\hat{\boldsymbol{M}}_{\ell,0}$ and $\hat{\boldsymbol{M}}_{\ell,1}$ have the following explicit form, which will be used below:

$$\begin{aligned} \hat{\boldsymbol{M}}_{\ell,0} &= 2^{\frac{1}{2}\ell-1} \left\{ (\hat{\boldsymbol{I}}_\ell + \hat{\boldsymbol{S}}_\ell) \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix} + (\hat{\boldsymbol{I}}_\ell - \hat{\boldsymbol{S}}_\ell) \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} , \\ \hat{\boldsymbol{M}}_{\ell,1} &= 2^{\frac{1}{2}\ell-1+(\ell+1)} (\hat{\boldsymbol{I}}_\ell - \hat{\boldsymbol{S}}_\ell) , \end{aligned} \tag{12d}$$

where

$$\hat{\boldsymbol{I}}_\ell = \begin{pmatrix} 1 & & \\ 0 & \ddots & \\ & \ddots & \ddots \\ & & 0 \quad 1 \end{pmatrix} \quad \text{and} \quad \hat{\boldsymbol{S}}_\ell = \begin{pmatrix} 0 & & \\ 1 & \ddots & \\ & \ddots & \ddots \\ & & 1 \quad 0 \end{pmatrix} \tag{12e}$$

are square matrices of order 2^ℓ.

The finite element spaces $\operatorname{span}\{\hat\varphi_{\ell,j}\}_{j\in\hat{\mathscr{J}}_\ell}$ with $\ell\in\mathbb{N}_0$ are nested: for all $L,\ell\in\mathbb{N}_0$ such that $\ell\le L$, we have

$$\hat\varphi_{\ell,j}=\sum_{j'\in\hat{\mathscr{J}}_L}(\hat{\boldsymbol{P}}_{\ell,L})_{j'\,j}\;\hat\varphi_{L,j'}\quad\text{for all}\quad j\in\hat{\mathscr{J}}_\ell\,,\tag{13}$$

where $\hat{\boldsymbol{P}}_{\ell,L}$ is the matrix of the identity operator from $\operatorname{span}\{\hat\varphi_{\ell,j}\}_{j\in\hat{\mathscr{J}}_\ell}$ to $\operatorname{span}\{\hat\varphi_{L,j'}\}_{j'\in\hat{\mathscr{J}}_L}$ with respect to the bases defined in (10):

$$\hat{\boldsymbol{P}}_{\ell,L}=2^{(\ell-L)/2}\big(\hat{\boldsymbol{I}}_\ell\otimes\hat{\boldsymbol{\eta}}_{L-\ell}+\hat{\boldsymbol{S}}_\ell\otimes(\hat{\boldsymbol{\xi}}_{L-\ell}-\hat{\boldsymbol{\eta}}_{L-\ell})\big)\tag{14}$$

where

$$\hat{\boldsymbol{\xi}}_k=\begin{pmatrix}1\\1\\\vdots\\1\\1\end{pmatrix}\quad\text{and}\quad\hat{\boldsymbol{\eta}}_k=2^{-k}\begin{pmatrix}1\\2\\\vdots\\2^k-1\\2^k\end{pmatrix}\tag{15}$$

are 2^k-component vectors for each $k\in\mathbb{N}_0$.

2.2 Finite Element Spaces for $\Omega=(0,1)^D$

Partition (7) induces a uniform tensor product partition of Ω that consists of the $2^{D\ell}$ elements

$$\Omega_{\ell,i}=\bigtimes_{d=1}^{D}\hat\Omega_{\ell,i_d}\quad\text{with}\quad i=(i_1,\dots,i_D)\in\mathscr{J}_\ell=\hat{\mathscr{J}}_\ell^{\,D}=\{1,\dots,2^\ell\}^D\,.\tag{16}$$

Tensorizing (10), we obtain the $2^{D\ell}$ functions

$$\varphi_{\ell,j}=\bigotimes_{d=1}^{D}\hat\varphi_{\ell,j_d}\quad\text{with}\quad j=(j_1,\dots,j_D)\in\mathscr{J}_\ell\,,\tag{17}$$

which are continuous on $\overline{\Omega}$ and D-linear on each of the partition elements given by (16). We will use these functions as a basis of a finite-dimensional subspace of V,

$$V_\ell=\operatorname{span}\{\varphi_{\ell,j}\}_{j\in\mathscr{J}_\ell}\subset V\,.\tag{18}$$

The normalization of univariate factors in (11) implies

$$\|\varphi_{\ell,j}\|_{\mathrm{L}^2(\Omega)}\sim 1\,,\tag{19}$$

and hence

$$\Big\| \sum_{i \in \mathscr{J}_\ell} \boldsymbol{v}_i \, \varphi_{\ell,i} \Big\|_{\mathrm{L}^2(\Omega)} \sim \|\boldsymbol{v}\|_{\ell^2} \quad \text{for all} \quad \boldsymbol{v} \in \mathbb{R}^{\mathscr{J}_\ell},$$

with equivalence constants independent of $\ell \in \mathbb{N}$. Also, relationship (12a) results in

$$\partial^\alpha \varphi_{\ell,j} \circ \phi_{\ell,i} = \sum_{\beta \in \{\alpha_1,1\} \times \cdots \times \{\alpha_D,1\}} (\boldsymbol{M}_{\ell,\alpha})_{i\beta\, j} \, \psi_\beta \quad \text{on} \quad (-1,1)^D \tag{20a}$$

for all $\alpha = (\alpha_1, \ldots, \alpha_D) \in \{0,1\}^D$ and $i, j \in \mathscr{J}_\ell$ with

$$\phi_{\ell,i} = \bigotimes_{d=1}^{D} \hat{\phi}_{\ell,i_d} \quad \text{and} \quad \psi_\beta = \bigotimes_{d=1}^{D} \hat{\psi}_{\beta_d} \tag{20b}$$

for all $i = (i_1, \ldots, i_D) \in \mathscr{J}_\ell$ and $\beta = (\beta_1, \ldots, \beta_D) \in \{0,1\}^D$ and with $\boldsymbol{M}_{\ell,\alpha}$ given by

$$(\boldsymbol{M}_{\ell,\alpha})_{i\beta\, j} = \prod_{k=1}^{D} (\hat{\boldsymbol{M}}_{\ell,\alpha_k})_{i_k \beta_k\, j_k} \tag{20c}$$

for all $i = (i_1, \ldots, i_D) \in \mathscr{J}_\ell$, $j = (j_1, \ldots, j_D) \in \mathscr{J}_\ell$ and $\beta = (\beta_1, \ldots, \beta_D) \in \{0,1\}^D$. Note that, for each $\alpha \in \{0,1\}^D$, the rows and columns of $\boldsymbol{M}_{L,\alpha}$ are indexed by $\mathscr{J}_L \times \{\alpha_1, 1\} \times \cdots \times \{\alpha_D, 1\}$ and $\mathscr{J}_L$, respectively. The embedding (12c) implies

$$(\boldsymbol{M}_{\ell,\alpha'})_{i\beta\, j} = 2^{|\alpha'-\alpha|(\ell+1)} (\boldsymbol{M}_{\ell,\alpha})_{i\beta\, j} \tag{20d}$$

for all $i, j \in \mathscr{J}_\ell$ and $\alpha, \alpha', \beta \in \{0,1\}^D$ such that $\alpha_k \le \alpha'_k \le \beta_k$ for each $k = 1, \ldots, D$.

The finite element spaces V_ℓ with $\ell \in \mathbb{N}_0$ are also nested: for all $L, \ell \in \mathbb{N}_0$ such that $\ell \le L$, we have $V_\ell \subset V_L$. In particular, the basis functions of V_ℓ and V_L introduced in (17) satisfy the refinement relation

$$\varphi_{\ell,j} = \sum_{j' \in \mathscr{J}_L} (\boldsymbol{P}_{\ell,L})_{j'\, j} \; \varphi_{L,j'} \quad \text{for all} \quad j \in \mathscr{J}_\ell, \tag{21}$$

where

$$\boldsymbol{P}_{\ell,L} = \bigotimes_{k=1}^{D} \hat{\boldsymbol{P}}_{\ell,L} \tag{22}$$

with $\hat{\boldsymbol{P}}_{\ell,L}$ given by (14).

The stiffness matrix for the bilinear form a and discretization level ℓ is given by

$$\boldsymbol{A}_\ell = \big(a(\varphi_{\ell,i}, \varphi_{\ell,j})\big)_{j,i\in\mathscr{J}_\ell}. \tag{23}$$

Note that due to (19),

$$\langle \boldsymbol{A}_\ell \boldsymbol{v}, \boldsymbol{v}\rangle \sim \Big\| \sum_{i\in\mathscr{J}_\ell} \boldsymbol{v}_i\, \varphi_{\ell,i} \Big\|_V^2 \quad \text{for all} \quad \boldsymbol{v} \in \mathbb{R}^{\mathscr{J}_\ell}.$$

For the right-hand side, we set $\boldsymbol{f}_\ell = \big(f(\varphi_{\ell,i})\big)_{i\in\mathscr{J}_\ell}$.

2.3 Representation of Differential Operators

The bilinear form $a\colon V\times V\to\mathbb{R}$ in (5) can be rewritten in the form

$$a(u,v) = \sum_{(\alpha,\alpha')\in\mathscr{D}} \int_\Omega c_{\alpha\alpha'} (\partial^\alpha v)(\partial^{\alpha'} u) \quad \text{for all} \quad u, v \in V \tag{24a}$$

with a $\mathscr{D}\subset\{0,1\}^D\times\{0,1\}^D$. We assume that each coefficient function $c_{\alpha\alpha'}\in \mathrm{L}^\infty(\Omega)$ with $(\alpha,\alpha')\in\mathscr{D}$ is given by

$$c_{\alpha\alpha'}\circ\phi_{L,i} = \sum_{\gamma\in\Gamma_{\alpha\alpha'}} (\boldsymbol{c}_{L,\alpha,\alpha'})_{i\gamma}\, \chi_{\alpha\alpha'\gamma} \quad \text{on} \quad (-1,1)^D \quad \text{for all} \quad i\in\mathscr{J}_L \tag{24b}$$

in terms of the affine transformations $\phi_{L,i}$ with $i\in\mathscr{J}_L$ defined by (9) and (20b), a finite index set $\Gamma_{\alpha\alpha'}$ of cardinality $R_{\alpha\alpha'} = |\Gamma_{\alpha\alpha'}|$, functions $\chi_{\alpha\alpha'\gamma}\in \mathrm{L}^\infty((-1,1)^D)$ with $\gamma\in\Gamma_{\alpha\alpha'}$ and a coefficient vector $\boldsymbol{c}_{L,\alpha,\alpha'}\in\mathbb{R}^{\mathscr{J}_L\times\Gamma_{\alpha\alpha'}}\simeq\mathbb{R}^{2^{DL}R_{\alpha\alpha'}}$.

In this section, we analyze the dimension structure of the matrix $\boldsymbol{A}_L$ of a restricted to $V_L\times V_L$ with respect to the basis of $\varphi_{L,j}$ with $j\in\mathscr{J}_L$, whose entries are

$$(\boldsymbol{A}_L)_{j\,j'} = a(\varphi_{L,j},\varphi_{L,j'}) = \sum_{(\alpha,\alpha')\in\mathscr{D}} \int_\Omega c_{\alpha\alpha'} (\partial^\alpha\varphi_{L,j})(\partial^{\alpha'}\varphi_{L,j'}) \quad \text{with} \quad j,j'\in\mathscr{J}_L, \tag{25a}$$

induced by the tensor product dimension structure of the basis. Splitting integration over the elements $\Omega_{\ell,i}$ with $i\in\mathscr{J}_L$, given by (16), and applying (20a), we obtain

$$\begin{aligned}(\boldsymbol{A}_L)_{j\,j'} &= \sum_{(\alpha,\alpha')\in\mathscr{D}} \sum_{i\in\mathscr{J}_L} \int_{\Omega_{L,i}} c_{\alpha\alpha'} (\partial^\alpha\varphi_{L,j})(\partial^{\alpha'}\varphi_{L,j'}) \\ &= \sum_{(\alpha,\alpha')\in\mathscr{D}} \sum_{i\in\mathscr{J}_L} \sum_{\gamma\in\Gamma_{\alpha\alpha'}} 2^{-D(L+1)}\, (\boldsymbol{c}_{L,\alpha\alpha'})_{i\gamma} \int_{(-1,1)^D} \chi_{\alpha\alpha'\gamma}\end{aligned}$$

$$\sum_{\beta\in\{\alpha_1,1\}\times\cdots\times\{\alpha_D,1\}} (\boldsymbol{M}_{L,\alpha})_{i\beta\,j}\ \partial^{\alpha}\psi_{\beta} \sum_{\beta'\in\{\alpha'_1,1\}\times\cdots\times\{\alpha'_D,1\}} (\boldsymbol{M}_{L,\alpha'})_{i\beta'\,j'}\ \partial^{\alpha'}\psi_{\beta'}\,. \tag{25b}$$

Let us now, for all $\alpha,\alpha'\in\mathscr{D}$, introduce a matrix $\boldsymbol{\Lambda}_{L,\alpha,\alpha'}$ of size $2^{D(L+1)-|\alpha|}\times 2^{D(L+1)-|\alpha'|}$:

$$(\boldsymbol{\Lambda}_{L,\alpha,\alpha'})_{i\beta\,i'\beta'} = \delta_{ii'}\,2^{-D(L+1)} \sum_{\gamma\in\Gamma_{\alpha\alpha'}} (\boldsymbol{c}_{L,\alpha,\alpha'})_{i\gamma} \int_{(-1,1)^D} \chi_{\alpha\alpha'\gamma}\,(\partial^{\alpha}\psi_{\beta})\,(\partial^{\alpha'}\psi_{\beta'}) \tag{26a}$$

for all $i,i'\in\mathscr{J}_L$, $\beta\in\{\alpha_1,1\}\times\cdots\times\{\alpha_D,1\}$ and $\beta'\in\{\alpha'_1,1\}\times\cdots\times\{\alpha'_D,1\}$. Using these matrices, we can rewrite (25b) as

$$\boldsymbol{A}_L = \sum_{(\alpha,\alpha')\in\mathscr{D}} \boldsymbol{M}_{L,\alpha}^{\mathsf{T}}\,\boldsymbol{\Lambda}_{L,\alpha,\alpha'}\,\boldsymbol{M}_{L,\alpha'}\,. \tag{26b}$$

Example 1 In the case of the negative Laplacian, we deal with a bilinear form given by (24a) with $\mathscr{D}=\big((\delta_{k1},\dots,\delta_{kD}),(\delta_{k1},\dots,\delta_{kD})\big)_{k=1}^{D}$ and $c_{\alpha\alpha'}=1$ for all $(\alpha,\alpha')\in\mathscr{D}$. For each (α,α), the corresponding coefficient is of form (24b) with $\Gamma_{\alpha\alpha'}=\{0\}$, $\chi_{\alpha\alpha'0}=1$ and $(\boldsymbol{c}_{L,\alpha\alpha'})_{i0}=1$ for all $i\in\mathscr{J}_L$. The corresponding matrix $\boldsymbol{\Lambda}_{L,\alpha,\alpha}$ given by (26a) takes the Kronecker product form

$$\boldsymbol{\Lambda}_{L,\alpha,\alpha} = \bigotimes_{k=1}^{D} \hat{\boldsymbol{\Lambda}}_{L,\alpha_k,\alpha_k}\,, \tag{27a}$$

where the factors $\hat{\boldsymbol{\Lambda}}_{L,0,0}$ and $\hat{\boldsymbol{\Lambda}}_{L,1,1}$ are diagonal matrices independent of $(\alpha,\alpha')\in\mathscr{D}$ whose rows and columns are indexed by $\mathscr{J}_L\otimes\{0,1\}$ and $\mathscr{J}_L\otimes\{1\}$, respectively. Specifically, their nonzero entries are

$$(\hat{\boldsymbol{\Lambda}}_{L,0,0})_{i,0\,i,0} = (\hat{\boldsymbol{\Lambda}}_{L,1,1})_{i,1\,i,1} = 2^{-L} \quad\text{and}\quad (\hat{\boldsymbol{\Lambda}}_{L,0,0})_{i,1\,i,1} = \frac{1}{3}\,2^{-L},\quad i\in\mathscr{J}_L\,. \tag{27b}$$

The multilevel tensor structure of the factorization (26b) and, in particular, of $\boldsymbol{\Lambda}_{L,\alpha,\alpha'}$ with $(\alpha,\alpha')\in\mathscr{D}$ is investigated in Sect. 5. This analysis applies to the case of general nonconstant coefficients $c_{\alpha\alpha'}$ with $(\alpha,\alpha')\in\mathscr{D}$ under the assumption that each of them exhibits the multilevel low-rank structure in the sense of Sect. 3. Specifically, in Sect. 5, we analyze the low-rank structure of every factor matrix $\boldsymbol{M}_{L,\alpha}$ with $\alpha\in\{0,1\}^D$ and also show how the low-rank structure of $c_{\alpha\alpha'}$ with $(\alpha,\alpha')\in\mathscr{D}$ translates into that of $\boldsymbol{\Lambda}_{L,\alpha,\alpha'}$. First, however, in the remainder of Sect. 2 we turn to the multilevel preconditioning of $\boldsymbol{A}_L$. This gives rise to the preconditioned operator $\boldsymbol{B}_L$ and matrices $\boldsymbol{Q}_{L,\alpha}$ with $\alpha\in\{0,1\}^D$, defined in (33c), which relate to $\boldsymbol{B}_L$ as $\boldsymbol{M}_{L,\alpha}$ with $\alpha\in\{0,1\}^D$ to $\boldsymbol{A}_L$. The low-rank multilevel structure of $\boldsymbol{B}_L$ and $\boldsymbol{Q}_{L,\alpha}$ with $\alpha\in\{0,1\}^D$ is the main topic of Sect. 5.

Remark 1 In the case of one dimension ($D = 1$), let us consider a diffusion operator with a coefficient c that is piecewise constant: $c \circ \hat{\phi}_{L,i} = (\hat{\boldsymbol{c}}_L)_i$ on $(-1, 1)$ for all $i \in \hat{\mathscr{J}}_L$, cf. (24b). Such coefficients appear, for example, as approximations in the midpoint quadrature rule. Then, representation (25b) takes the form

$$\begin{aligned} \boldsymbol{A}_L = 2^{-L}\, \hat{\boldsymbol{M}}_{L,1}^{\mathsf{T}}\, (\operatorname{diag} \hat{\boldsymbol{c}}_L)\, \hat{\boldsymbol{M}}_{L,1} = 2^{2L} \big[\operatorname{diag} \big((\hat{\boldsymbol{I}}_L + \hat{\boldsymbol{S}}_L^{\mathsf{T}})\, \hat{\boldsymbol{c}}_L \big) \\ - \hat{\boldsymbol{S}}_L^{\mathsf{T}} (\operatorname{diag} \hat{\boldsymbol{c}}_L) - (\operatorname{diag} \hat{\boldsymbol{c}})\, \hat{\boldsymbol{S}}_L \big], \end{aligned} \tag{28}$$

where $\hat{\boldsymbol{M}}_{L,1} = 2^{\frac{3}{2}L} (\hat{\boldsymbol{I}}_\ell - \hat{\boldsymbol{S}}_\ell)$ is defined by (12a) and is given explicitly by (12d). The representation (28) has been used for this one-dimensional case in [15,16,31]; representation (26b) provides a generalization to higher dimensions and general coefficients.

2.4 Multilevel Preconditioning

Among the various existing methods for preconditioning discretization matrices of second-order elliptic problems, we are especially interested in approaches that provide optimal preconditioning and at the same time lead to favorable multilevel low-rank structures. A choice that meets these criteria is based on the classical BPX preconditioner [10]. For our particular purposes, in what follows we also obtain a new result on symmetric preconditioning by this method.

The BPX preconditioner requires a hierarchy of nested finite element spaces $V_0 \subset V_1 \subset \cdots \subset V_L \subset V$, which in the present case are the uniformly refined spaces defined in (18). The standard implementable form of the preconditioner (cf. [10,53]) is then given by

$$C_{2,L}\, v = \sum_{\ell=0}^{L} 2^{-2\ell} \sum_{j \in \mathscr{J}_L} \langle v, \varphi_{\ell,j} \rangle \varphi_{\ell,j}, \quad v \in V_L.$$

Interpreting $C_{2,L}$ as a mapping of coefficient sequences $(\langle v, \varphi_{L,j} \rangle)_{j \in \mathscr{J}_L}$ to nodal values of finite element functions, one obtains the corresponding matrix representation

$$\boldsymbol{C}_{2,L} = \sum_{\ell=0}^{L} 2^{-2\ell}\, \boldsymbol{P}_{\ell,L}\, \boldsymbol{P}_{\ell,L}^{\mathsf{T}}, \tag{29}$$

where $\boldsymbol{P}_{\ell,L}$ is as in (21), (22). The following result on the BPX preconditioner (29) was established in [12,48], see also [9,54].

Theorem 1 *Let $\boldsymbol{A}_L$ and $\boldsymbol{C}_{2,L}$ be as in (23) and (29). Then, there exist $c, C > 0$ independent of L such that*

$$c\, \langle \boldsymbol{C}_{2,L}^{-1} \boldsymbol{v}, \boldsymbol{v} \rangle \le \langle \boldsymbol{A}_L \boldsymbol{v}, \boldsymbol{v} \rangle \le C\, \langle \boldsymbol{C}_{2,L}^{-1} \boldsymbol{v}, \boldsymbol{v} \rangle, \quad \boldsymbol{v} \in \mathbb{R}^{\mathscr{J}_L}.$$

This preconditioner is therefore optimal; that is, the condition numbers of preconditioned systems remain bounded uniformly in the discretization level. It is usually applied in the form of a left-sided preconditioning: it implies in particular that $\mathrm{cond}(\boldsymbol{C}_{2,L}^{1/2}\boldsymbol{A}_L\,\boldsymbol{C}_{2,L}^{1/2})$ is uniformly bounded with respect to L and that there exists $\omega > 0$ such that the iteration $\boldsymbol{u}^{k+1} = \boldsymbol{u}^k - \omega\,\boldsymbol{C}_{2,L}\big(\boldsymbol{A}_L\boldsymbol{u}^k - \boldsymbol{f}_L\big)$ converges at an L-independent rate. Also standard implementations of the preconditioned conjugate gradient method use only the action of $\boldsymbol{C}_{2,L}$.

For our purposes, for several reasons explained in further detail in what follows, we require *symmetric* preconditioning, that is, an implementable operator $\boldsymbol{C}_L$ such that $\boldsymbol{C}_L\boldsymbol{A}_L\boldsymbol{C}_L$ is well-conditioned. Although $\boldsymbol{C}_{2,L}^{1/2}$ provides optimal symmetric preconditioning by Theorem 1, this is not directly numerically realizable.

We thus instead consider two-sided preconditioning by the implementable operator

$$\boldsymbol{C}_L = \sum_{\ell=0}^{L} 2^{-\ell}\,\boldsymbol{P}_{\ell,L}\,\boldsymbol{P}_{\ell,L}^{\mathsf{T}}\,. \tag{30}$$

For bounding the condition number of the symmetrically preconditioned operator $\boldsymbol{C}_L\boldsymbol{A}_L\boldsymbol{C}_L$, we need to establish spectral equivalence of $\boldsymbol{A}_L$ and $\boldsymbol{C}_L^{-2}$. This is not a direct consequence of Theorem 1. Although relying mainly on adaptations of established techniques as in [26,51,54], the following result appears to be new. The proof is given in "Appendix A."

Theorem 2 *With $\boldsymbol{A}_L$ as in (23) and $\boldsymbol{C}_L$ as in (30), there exist $c, C > 0$ independent of L such that*

$$c\|\boldsymbol{v}\|_2^2 \le \langle \boldsymbol{C}_L\boldsymbol{A}_L\boldsymbol{C}_L\boldsymbol{v}, \boldsymbol{v}\rangle \le C\|\boldsymbol{v}\|_2^2, \quad \boldsymbol{v} \in \mathbb{R}^{\mathscr{J}_L}. \tag{31}$$

Remark 2 As an immediate consequence of Theorem 2,

$$\|v\|_{\mathrm{H}^1} \sim \|\boldsymbol{C}_L^{-1}\boldsymbol{v}\|_2 \qquad \text{for } v = \sum_{j\in\mathscr{J}_L} \boldsymbol{v}_j\,\varphi_{L,j}, \quad \boldsymbol{v} \in \mathbb{R}^{\mathscr{J}_L}, \tag{32}$$

which means that the functions $\sum_{i\in\mathscr{J}_L}(\boldsymbol{C}_L)_{ij}\varphi_{L,i}$, $j \in \mathscr{J}_L$, form a Riesz basis of the subspace $V_L \subset \mathrm{H}^1(\Omega)$ with bounds independent of L.

In what follows, we consider the symmetrically preconditioned problem of finding $\boldsymbol{u}_L$ such that

$$\boldsymbol{B}_L\boldsymbol{u}_L = \boldsymbol{g}_L \quad \text{where} \quad \boldsymbol{B}_L = \boldsymbol{C}_L\boldsymbol{A}_L\boldsymbol{C}_L \text{ and } \boldsymbol{g}_L = \boldsymbol{C}_L\boldsymbol{f}_L. \tag{33a}$$

Then, $\bar{\boldsymbol{u}}_L = \boldsymbol{C}_L\boldsymbol{u}_L$ satisfies $\boldsymbol{A}_L\bar{\boldsymbol{u}}_L = \boldsymbol{f}_L$; that is, $\bar{\boldsymbol{u}}_L$ are the (rescaled) nodal values of the Galerkin solution at level L. Using (26b), we obtain

$$\boldsymbol{B}_L = \sum_{(\alpha,\alpha')\in\mathscr{D}} \boldsymbol{Q}_{L,\alpha}^{\mathsf{T}}\,\boldsymbol{\Lambda}_{L,\alpha\alpha'}\,\boldsymbol{Q}_{L,\alpha'}\,, \tag{33b}$$

where

$$\boldsymbol{Q}_{L,\alpha} = \boldsymbol{M}_{L,\alpha}\, \boldsymbol{C}_L \tag{33c}$$

for all $\alpha \in \{0, 1\}^D$.

For our purposes, the symmetrically preconditioned operator is preferable mainly for two reasons. On the one hand, an important advantage of the symmetric preconditioning (33b) consists in the norm equivalence (32), since ultimately we are interested in numerical schemes with guaranteed convergence in the H^1 norm. With low-rank methods using SVD-based rank truncations, as considered in further detail in Sect. 6, for any $\varepsilon > 0$ we can find $\boldsymbol{v}$ such that $\|\boldsymbol{u}_L - \boldsymbol{v}\|_2 \leq \varepsilon$ with $\boldsymbol{u}_L$ as in (33a). With the nodal basis coefficients $\bar{\boldsymbol{v}} = \boldsymbol{C}_L \boldsymbol{v}$, for the corresponding finite element functions $v = \sum_{j \in \mathscr{J}_L} \bar{\boldsymbol{v}}_j\, \varphi_{L,j}$ and $u_L = \sum_{j \in \mathscr{J}_L} \bar{\boldsymbol{u}}_{L,j}\, \varphi_{L,j}$ we have $\|u_L - v\|_{\mathrm{H}^1} \lesssim \|\boldsymbol{C}_L^{-1}(\bar{\boldsymbol{u}}_L - \bar{\boldsymbol{v}})\|_2 = \|\boldsymbol{u}_L - \boldsymbol{v}\|_2 \leq \varepsilon$ by (32). On the other hand, the symmetric preconditioning (33b) allows for the explicit assembly of the preconditioned operator $\boldsymbol{B}_L$ directly in the low-rank form, as considered in detail in Sect. 5.

3 Tensor Train Decomposition

In this section, we recapitulate the definition of the tensor train (TT) decomposition of multidimensional arrays and present the notation that we need for the following sections.

3.1 Tensor Train Decomposition of Multidimensional Arrays

Throughout this section, we assume that $L \in \mathbb{N}$. Let $n_1, \ldots, n_L \in \mathbb{N}$ and $\boldsymbol{u}$ be a multidimensional vector of dimension $n_1 \cdots n_L$. Let $r_1, \ldots, r_{L-1} \in \mathbb{N}$ and, for $\ell = 1, \ldots, L$, let U_ℓ be arrays of size $r_{\ell-1} \times n_\ell \times r_\ell$, where $r_0 = 1$ and $r_L = 1$. The vector $\boldsymbol{u}$ is said to be represented in the *tensor train (TT) decomposition* [45,47] with *ranks* $r_1, \ldots, r_{L-1}$ and *cores* $U_1, \ldots, U_L$ if

$$\boldsymbol{u}_{j_1,\ldots,j_L} = \sum_{\alpha_1=1}^{r_1} \cdots \sum_{\alpha_{L-1}=1}^{r_{L-1}} U_1(\alpha_0, j_1, \alpha_1) \cdots U_L(\alpha_{L-1}, j_L, \alpha_L) \tag{34a}$$

for all $j_\ell = 1, \ldots, n_\ell$ with $\ell = 1, \ldots, L$, where $\alpha_0 \equiv 1$ and $\alpha_L \equiv 1$ are dummy indices.

The TT decomposition for matrices is defined analogously. Assume that $m_1, n_1, \ldots, m_L, n_L \in \mathbb{N}$ and that $\boldsymbol{A}$ is a matrix of size $(m_1 \cdots m_L) \times (n_1 \cdots n_L)$. Let $p_1, \ldots, p_{L-1} \in \mathbb{N}$ and, for each $\ell = 1, \ldots, L$, let A_ℓ be an array of size $p_{\ell-1} \times m_\ell \times n_\ell \times p_\ell$, where $p_0 = 1$ and $p_L = 1$. Then, the representation

$$\boldsymbol{A}_{i_1,\ldots,i_L\ \ j_1,\ldots,j_L} = \sum_{\beta_1=1}^{p_1} \cdots \sum_{\beta_{L-1}=1}^{p_{L-1}} A_1(\beta_0, i_1, j_1, \beta_1) \cdots A_L(\beta_{L-1}, i_L, j_L, \beta_L) \tag{34b}$$

for all $i_\ell = 1, \ldots, n_\ell$ with $\ell = 1, \ldots, L$, where $\beta_0 \equiv 1$ and $\beta_L \equiv 1$ are dummy indices, is called a tensor train decomposition of the matrix $\boldsymbol{A}$ with *ranks* $p_1, \ldots, p_{L-1}$ and *cores* $A_1, \ldots, A_L$.

The TT decomposition uses one of many possible ways to separate variables in multidimensional arrays; see, e.g., the survey [40] and the monograph [22]. The TT decomposition is a particular case of the more general *hierarchical tensor representation*, also known as the *hierarchical Tucker representation* [18,25]. Both the TT and hierarchical tensor representations can be interpreted as successive subspace approximation or low-rank matrix factorization, and this relation allows for the quasi-optimal low-rank approximation of tensors built upon standard matrix algorithms.

The number of parameters of the representation, formally linear in L, is mainly governed by the ranks, such as $r_1, \ldots r_{L-1}$ in (34a) and $p_1, \ldots, p_{L-1}$ in (34b). In many applications, the complexity is observed, theoretically as well as numerically, to depend moderately on L (see, e.g., [20]), which allows to lift or completely avoid the so-called *curse of dimensionality* associated with the entrywise storage of high-dimensional arrays.

The use of L for the dimensionality of tensors in this section is not accidental: in the present paper, the "dimension" index $\ell \in \{1, \ldots, L\}$ enumerates the levels of discretization, and each of the mode indices (i_ℓ and j_ℓ with $\ell \in \{1, \ldots, L\}$ above) represents the corresponding D bits of the D "physical" dimensions. In this case, the TT format separates not "physical" dimensions of tensors but rather the levels of the "physical" dimensions and adaptive low-rank approximation allows to resolve this multilevel structure in vectors and matrices. In this setting, the TT decomposition is known as the quantized tensor train (QTT) decomposition [21,37,43,44]. This idea is further explained in Sect. 3.7.

3.2 Core Notation

In this section, we present the notation developed in [30,32,35], which we extensively use to work with TT representations. For the sake of brevity, several definitions and properties will be stated for cores with two mode indices, which naturally arise in TT representations of matrices. The setting with a single mode index per core can be considered a particular case in the same way as vectors can be considered one-column matrices.

If $U^{[\alpha,\beta]}$ with $\alpha = 1, \ldots, p$ and $\beta = 1, \ldots, q$ are tensors of size $m \times n$, we call the array U of size $p \times m \times n \times q$ given by

$$U(\alpha, i, j, \beta) = U_{ij}^{[\alpha,\beta]} \tag{35}$$

for all $\alpha = 1, \ldots, p$, $i = 1, \ldots, m$, $j = 1, \ldots, n$ and $\beta = 1, \ldots, q$ a *core* of *rank* $p \times q$ and *mode size* $m \times n$. Conversely, for any core U of rank $p \times q$ and mode size $m \times n$, we refer to each tensor $U^{[\alpha,\beta]}$ with $\alpha = 1, \ldots, p$ and $\beta = 1, \ldots, q$ as *block* (α, β) of the core U.

For explicitly defining a core U, as a tensor of order four as in (35), in terms of its blocks (which in turn can be matrices or vectors), we use the notation

$$U = \begin{bmatrix} U^{[1,1]} & \ldots & U^{[1,q]} \\ \vdots & \ddots & \vdots \\ U^{[p,1]} & \ldots & U^{[p,q]} \end{bmatrix}, \tag{36}$$

where square brackets are used for distinction from matrices. The following matrices are examples of blocks that we frequently use in this paper:

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad J = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad I_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad I_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{37}$$

To apply the usual matrix transposition to TT decompositions of matrices, we will use the transposition of mode indices of cores:

$$U^{\mathsf{T}}(\alpha, i, j, \beta) = U(\alpha, j, i, \beta), \quad \text{i.e.,} \quad \left(U^{\mathsf{T}}\right)^{[\alpha,\beta]} = \left(U^{[\alpha,\beta]}\right)^{\mathsf{T}} \tag{38}$$

in terms of matrix transposition, for all values of the indices.

Similarly to (35), for any core U of rank $p \times q$ and mode size $m \times n$, we refer to each matrix $U^{\{ij\}}$ with $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$ given by

$$U(\alpha, i, j, \beta) = U_{\alpha\beta}^{\{i,j\}} \tag{39}$$

for all $\alpha = 1, \ldots, p$ and $\beta = 1, \ldots, q$ as *slice* (i, j) of the core U.

3.3 Strong Kronecker Product

We are interested in cores as factors of TT decompositions, and now we present how decompositions of forms (34a)–(34b) can be expressed in terms of cores. For that purpose, we use the *strong Kronecker product*, introduced for two-level matrices in [13]. In order to avoid confusion with the Hadamard and tensor products, we denote this operation by $\bowtie$, as in [35, Definition 2.1], where it was introduced specifically for connecting cores into TT representations.

Definition 1 (Strong Kronecker product of cores) Let $p, q, r \in \mathbb{N}$ and $m_1, m_2, n_1, n_2 \in \mathbb{N}$. Consider cores U and V of ranks $p \times r$ and $r \times q$ and of mode size $m_1 \times m_2$ and $n_1 \times n_2$, respectively. The *strong Kronecker product* $U \bowtie V$ of U and V is the core of rank $p \times q$ and mode size $m_1 m_2 \times n_1 n_2$ given, in terms of the matrix multiplication of slices (of size $p \times r$ and $r \times q$), by

$$(U \bowtie V)^{\{i_1 i_2,\, j_1 j_2\}} = U^{\{i_1, j_1\}}\, V^{\{i_2, j_2\}}$$

for all combinations of $i_k \in \{1, \ldots, m_k\}$ and $j_k \in \{1, \ldots, n_k\}$ with $k = 1, 2$.

In other words, we define $U \bowtie V$ as the usual matrix product of the corresponding core matrices, their entries (blocks) being multiplied by means of the Kronecker product. For example, we have

$$\begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \bowtie \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix} = \begin{bmatrix} V_{11} \otimes W_{11} + V_{12} \otimes W_{21} & V_{11} \otimes W_{12} + V_{12} \otimes W_{22} \\ V_{21} \otimes W_{11} + V_{22} \otimes W_{21} & V_{21} \otimes W_{12} + V_{22} \otimes W_{22} \end{bmatrix} \tag{40}$$

for two cores of rank 2×2. Using the strong Kronecker product, we can rewrite (34a) and (34b) as follows:

$$\boldsymbol{u} = [\boldsymbol{u}] = U_1 \bowtie \cdots \bowtie U_L \quad \text{and} \quad \boldsymbol{A} = [\boldsymbol{A}] = A_1 \bowtie \cdots \bowtie A_L\,, \tag{41}$$

where the first equalities indicate that any tensor of dimension $m \times n$ can be identified with a core of rank 1×1 and mode size $m \times n$.

3.4 Representation Map

Since many different tuples of cores may represent (or approximate) the same tensor, we need to distinguish representations as tuples of cores. We denote such tuples by sans-serif letters; for example,

$$\mathsf{U} = (U_1, \ldots, U_L) \quad \text{and} \quad \mathsf{A} = (A_1, \ldots, A_L) \tag{42a}$$

for the decompositions given by (34a) and (34b). Further, we denote by τ the function mapping *tuples of cores* into *cores* (in particular, into tensors when the rank of the resulting core is 1×1):

$$\tau(U_1, \ldots, U_L) = U_1 \bowtie \cdots \bowtie U_L \tag{42b}$$

for any cores $U_1, \ldots, U_L$ such that the right-hand side exists in the sense of Definition 1. Under (42a), this allows to rewrite (34a)–(34b) and (41) as

$$\boldsymbol{u} = [\boldsymbol{u}] = \tau(\mathsf{U}) \quad \text{and} \quad \boldsymbol{A} = [\boldsymbol{A}] = \tau(\mathsf{A})\,. \tag{42c}$$

For the sets of all tuples of $L \in \mathbb{N}$ cores with compatible ranks, we write $\mathrm{TT}_L = \mathrm{TT}^1_L$ in the case of blocks with one mode index, and TT^2_L in the case of two mode indices as in (35).

Furthermore, let us assume that $\mathsf{U} = (U_1, \dots, U_L) \in \mathrm{TT}_L$, i.e., that $U_1, \dots, U_L$ are cores such that $\tau(U_1, \dots, U_L)$ is a core of rank $r_0 \times r_L$ and mode size n, where $r_0, r_L, n \in \mathbb{N}$. Then, by τ^- and τ^+, we denote the matrices of size $r_0 n \times r_L$ and $r_0 \times n r_L$, respectively, given as follows:

$$\big(\tau^-(U_1, \dots, U_L)\big)_{\beta_0 i \;\; \beta_L} = \big(\tau(U_1, \dots, U_L)\big)(\beta_0, i, \beta_L) \tag{43a}$$

and

$$\big(\tau^+(U_1, \dots, U_L)\big)_{\beta_0 \;\; i \beta_L} = \big(\tau(U_1, \dots, U_L)\big)(\beta_0, i, \beta_L) \tag{43b}$$

for all $\beta_0 = 1, \dots, r_0, i = 1, \dots, n$ and $\beta_L = 1, \dots, r_L$. These matrices may be called matricizations of the core $\tau(U_1, \dots, U_L)$: they are obtained by interpreting the rank indices as row and column indices, which is consistent with (36), and by interpreting all mode indices as either row or column indices. For notational convenience, we set $\tau^-(\varnothing) = 1$ and $\tau^+(\varnothing) = 1$ for empty lists of cores.

Moreover, for each $\ell = 1, \dots, L$, we define

$$\tau^-_\ell(\mathsf{U}) = \tau^-(U_1, \dots, U_{\ell-1}) \quad \text{for each} \quad \ell = 1, \dots, L+1 \tag{43c}$$

and

$$\tau^+_\ell(\mathsf{U}) = \tau^+(U_{\ell+1}, \dots, U_L) \quad \text{for each} \quad \ell = 0, \dots, L\,. \tag{43d}$$

In particular, we have $\tau^-_1(U_1, \dots, U_L) = 1$, $\tau^-_{L+1}(U_1, \dots, U_L) = \tau^-(U_1, \dots, U_L)$ and $\tau^+_L(U_1, \dots, U_L) = 1$, $\tau^+_0(U_1, \dots, U_L) = \tau^+(U_1, \dots, U_L)$.

3.5 Unfolding Matrices, Ranks and Orthogonality

Let us consider a vector $\boldsymbol{u}$ of size $n_1 \cdots n_L$ and a matrix $\boldsymbol{A}$ of size $m_1 \cdots m_L \times n_1 \cdots n_L$. For every $\ell = 1, \dots, L-1$, we denote by $\mathrm{U}_\ell(\boldsymbol{u})$ and $\mathrm{U}_\ell(\boldsymbol{A})$ the ℓth *unfolding matrices* of $\boldsymbol{u}$ and $\boldsymbol{A}$, which are the matrices of size $n_1 \cdots n_\ell \times n_{\ell+1} \cdots n_L$ and $m_1 n_1 \cdots m_\ell n_\ell \times m_{\ell+1} n_{\ell+1} \cdots m_L n_L$ given by

$$\big(\mathrm{U}_\ell(\boldsymbol{u})\big)_{j_1, \dots, j_\ell \;\; j_{\ell+1}, \dots, j_L} = \boldsymbol{u}_{j_1, \dots, j_\ell, j_{\ell+1}, \dots, j_L}\,, \tag{44a}$$

$$\big(\mathrm{U}_\ell(\boldsymbol{A})\big)_{i_1 j_1, \dots, i_\ell j_\ell \;\; i_{\ell+1} j_{\ell+1}, \dots, i_L j_L} = \boldsymbol{A}_{i_1, \dots, i_\ell, i_{\ell+1}, \dots, i_L \;\; j_1, \dots, j_\ell, j_{\ell+1}, \dots, j_L} \tag{44b}$$

for all $i_k = 1, \dots, m_k$ and $j_k = 1, \dots, n_k$ with $k = 1, \dots, L$. For the ranks of the unfolding matrices, we use the notation

$$\mathrm{rank}_\ell(\boldsymbol{u}) = \mathrm{rank}\, \mathrm{U}_\ell(\boldsymbol{u}) \quad \text{and} \quad \mathrm{rank}_\ell(\boldsymbol{A}) = \mathrm{rank}\, \mathrm{U}_\ell(\boldsymbol{A}) \tag{44c}$$

for each $\ell = 1, \dots, L-1$.

The decompositions given by (34a)–(34b) or, equivalently, by (42c) imply $\operatorname{rank}_\ell(\boldsymbol{u}) \le r_\ell$ and $\operatorname{rank}_\ell(\boldsymbol{A}) \le p_\ell$ for each $\ell = 1, \dots, L-1$; furthermore, the decompositions provide low-rank factorizations of the unfolding matrices with the respective numbers of rank-one terms. For example, in the case of a vector, using the notation introduced in (43c)–(43d), we can write $\mathrm{U}_\ell(\boldsymbol{u}) = \tau^-_{\ell+1}(\mathsf{U})\,\tau^+_\ell(\mathsf{U})$.

Conversely, if $\boldsymbol{u}$ and $\boldsymbol{A}$ are such that, for every $\ell = 1, \dots, L-1$, the unfolding matrices $\mathrm{U}_\ell(\boldsymbol{u})$ and $\mathrm{U}_\ell(\boldsymbol{A})$ have approximations of ranks r_ℓ and p_ℓ, respectively, and of accuracy ε_ℓ in the Frobenius norm, then representations $\mathsf{U} = (U_1, \dots, U_L)$ and $\mathsf{A} = (A_1, \dots, A_L)$ of ranks $r_1, \dots, r_{L-1}$ and $p_1, \dots, p_{L-1}$ such that

$$\|\tau(\mathsf{U}) - \boldsymbol{u}\|_2^2 \le \varepsilon^2 \quad \text{and} \quad \|\tau(\mathsf{A}) - \boldsymbol{A}\|_{\mathrm{F}}^2 \le \varepsilon^2$$

with $\varepsilon^2 = \varepsilon_1^2 + \cdots + \varepsilon_{\ell-1}^2$ exist [45, Theorem 2.2] and can be constructed by the TT-SVD algorithm [45, Algorithm 1].

Next, we recapitulate the notion of orthogonality of decompositions in terms of the matricization operators defined in (43a)–(43d). If a core U is such that the matrix $\tau^-(U)$ has orthonormal columns, then the core is called *left-orthogonal*. Similarly, if the matrix $\tau^+(U)$ has orthonormal rows, then the core is called *right-orthogonal*. Further, if $\mathsf{U} \in \mathrm{TT}_L$ is such that the columns of each matrix $\tau^-_\ell(\mathsf{U})$ with $\ell = 2, \dots, L+1$ are orthonormal, then the decomposition is called *left-orthogonal*. Analogously, if the rows of each matrix $\tau^+_\ell(\mathsf{U})$ with $\ell = 0, \dots, L-1$ are orthonormal, then the decomposition is called *right-orthogonal*. It is easy to see that any core U of the form $U = U_1 \bowtie U_2$ is left- or right-orthogonal if both U_1 and U_2 are left- or right-orthogonal, respectively. As a result, any decomposition $\mathsf{U} = (U_1, \dots, U_L)$ is left- or right-orthogonal if each of the cores $U_1, \dots, U_L$ is left- or right-orthogonal.

Moreover, we say that U is in *left-orthogonal TT-SVD form* if $\tau^-_{\ell+1}(\mathsf{U})$ has orthonormal columns and $\tau^+_\ell(\mathsf{U})$ has orthogonal rows for each $\ell = 1, \dots, L-1$; in other words, these matrices provide the SVD of $\mathrm{U}_\ell(\boldsymbol{u})$ for each ℓ, where the norms of the rows of $\tau^+_\ell(\mathsf{U})$ are the corresponding singular values, and $\|\boldsymbol{u}\|_2 = \|U_L\|_2$. Analogously, U is in *right-orthogonal TT-SVD form* if $\tau^-_{\ell+1}(\mathsf{U})$ has orthogonal columns and $\tau^+_\ell(\mathsf{U})$ has orthonormal rows. These TT-SVD forms can be obtained numerically for any given U by the procedure [45, Algorithm 1] without rank truncation.

3.6 Operations on Cores

We require several further operations, which are explained in this section. We start with the mode product of cores, which was introduced in [30, Definition 2.2] and which generalizes matrix multiplication to the case of cores.

Definition 2 (Mode product of cores) Let $p, p', r, r' \in \mathbb{N}$ and $m, n, k \in \mathbb{N}$. Consider cores A and B of ranks $p \times p'$ and $r \times r'$ and of mode size $m \times k$ and $k \times n$, respectively. The *mode core product* $A \bullet B$ of A and B is the core of rank $pq \times p'q'$ and mode size $m \times n$ given, in terms of the matrix multiplication of blocks (of sizes $m \times k$ and

$k \times n$), by

$$(A \bullet B)^{[\alpha\beta,\,\alpha'\beta']} = A^{[\alpha,\alpha']}\, B^{[\beta,\beta']}$$

for all combinations of $\alpha = 1, \ldots, p$, $\alpha' = 1, \ldots, p'$, $\beta = 1, \ldots, q$ and $\beta = 1, \ldots, q'$. If B has only one mode index, we apply the above definition, introducing a dummy mode size $n = 1$ in B and discarding it in $A \bullet B$.

For example, for a core A with two mode indices and a core B with one or two mode indices, each core being of rank 2×2, we have

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \bullet \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} & A_{11}B_{12} & A_{12}B_{11} & A_{12}B_{12} \\ A_{11}B_{21} & A_{11}B_{22} & A_{12}B_{21} & A_{12}B_{22} \\ A_{21}B_{11} & A_{21}B_{12} & A_{22}B_{11} & A_{22}B_{12} \\ A_{21}B_{21} & A_{21}B_{22} & A_{22}B_{21} & A_{22}B_{22} \end{bmatrix} \tag{45}$$

if the first mode size of B equals the second of A.

The mode product and the strong Kronecker product inherit distributivity from the usual matrix product and from the Kronecker product: for $\mathsf{A} = (A_1, \ldots, A_L)$ and $\mathsf{U} = (U_1, \ldots, U_L)$ such that the products $A_\ell \bullet U_\ell$ with $\ell = 1, \ldots, L$ are all defined, we have that the product $\tau(\mathsf{A}) \bullet \tau(\mathsf{U})$ is defined and is given by

$$\begin{aligned} \tau(\mathsf{A}) \bullet \tau(\mathsf{U}) &\equiv (A_1 \bowtie \cdots \bowtie A_L) \bullet (U_1 \bowtie \cdots \bowtie U_L) \\ &= (A_1 \bullet U_1) \bowtie \cdots \bowtie (A_L \bullet U_L) \equiv \tau(A_1 \bullet U_1, \ldots, A_L \bullet U_L)\,. \end{aligned} \tag{46}$$

When $\tau(\mathsf{A})$ and $\tau(\mathsf{U})$ are both of rank 1×1 and can therefore be identified with matrices, $\tau(\mathsf{A}) \bullet \tau(\mathsf{U})$ is the core of rank 1×1 identified with the matrix–matrix product of these matrices, and (46) gives a representation for the product of a matrix $\boldsymbol{A} = \tau(\mathsf{A})$ and a vector $\boldsymbol{u} = \tau(\mathsf{U})$ given by (34b) and (34a).

Finally, our derivations involve Kronecker products of cores, which are defined as the Kronecker product of the corresponding arrays. For any $p, p', q, q' \in \mathbb{N}$ and $m, n, m', n' \in \mathbb{N}$, let A be a core of rank $p \times p'$ and mode size $m \times n$ and let B be a core of rank $q \times q'$ and mode size $m' \times n'$. Then, the Kronecker product $A \otimes B$ of A and B is the core of rank $pq \times p'q'$ and mode size $mm' \times nn'$ given by

$$(U \otimes V)^{[\alpha\beta,\,\alpha'\beta']} = U^{[\alpha,\alpha']} \otimes V^{[\beta,\beta']} \tag{47a}$$

in terms of the Kronecker products of all pairs of block tensors or, equivalently, by

$$(U \otimes V)^{\{ii',\,jj'\}} = U^{\{i,j\}} \otimes V^{\{i',j'\}} \tag{47b}$$

in terms of the Kronecker products of all pairs of slice matrices. Similarly to (46), we have

$$\begin{aligned}\tau(\mathsf{A}) \otimes \tau(\mathsf{B}) &\equiv (A_1 \bowtie \cdots \bowtie A_L) \otimes (B_1 \bowtie \cdots \bowtie B_L) \\ &= (A_1 \otimes B_1) \bowtie \cdots \bowtie (A_L \otimes B_L) \equiv \tau(A_1 \otimes B_1, \ldots, A_L \otimes B_L)\end{aligned} \tag{48}$$

for any representation $\mathsf{A} = (A_1, \ldots, A_L)$ and $\mathsf{B} = (B_1, \ldots, B_L)$. Relations (46) and (48) indicate the well-known fact that the matrix and Kronecker products can be recast core-wise; see, e.g., [22,40,45].

One of the most important properties of the TT decomposition of tensors is that any representation can be made left- or right-orthogonal in the sense of Sect. 3.5 by the successive application of the QR decomposition [18,22,25,41,45]. We now briefly present an algorithm for the left-orthogonalization of a decomposition, which we use as an example in the discussion of representation conditioning. This scheme is also the first step in the computation of the TT-SVD form of a TT representation, as in [45, Algorithm 2].

Algorithm 3.1 left-orthogonalization orth^- of a TT representation (right-orthogonalization orth^+ can be performed analogously)

1: **function** $\mathsf{V} = \mathrm{orth}^-(\mathsf{U})$
input: a representation $\mathsf{U} = (U_1, \ldots, U_L) \in \mathrm{TT}_L^S$ with $L, S \in \mathbb{N}$
output: a left-orthogonal representation $\mathsf{V} = (V_1, \ldots, V_L) \in \mathrm{TT}_L^S$ such that $\tau(\mathsf{V}) = \tau(\mathsf{U})$
2: set $W_1 = U_1$ ▷ $U_1 \bowtie U_2 \bowtie \cdots \bowtie U_L = W_1 \bowtie U_2 \bowtie \cdots \bowtie U_L$
3: **for** $\ell = 1, \ldots, L-1$ ▷ sweep through the representation from left to right
4: compute a matrix QR decomposition: $\tau^-(W_\ell) = Q_\ell R_\ell$
5: define V_ℓ, of same dimensions as U_ℓ, so that $\tau^-(V_\ell) = Q_\ell$
6: define $W_{\ell+1}$, of same dimensions as $U_{\ell+1}$, so that $\tau^+(W_{\ell+1}) = R_\ell \tau^+(U_{\ell+1})$
▷ $V_1 \bowtie \cdots \bowtie V_{\ell-1} \bowtie W_\ell \bowtie U_{\ell+1} \bowtie \cdots \bowtie U_L = V_1 \bowtie \cdots \bowtie V_\ell \bowtie W_{\ell+1} \bowtie U_{\ell+2} \bowtie \cdots \bowtie U_L$
7: **end for**
8: set $V_L = W_L$ ▷ $V_1 \bowtie \cdots \bowtie V_{L-1} \bowtie W_L = V_1 \bowtie \cdots \bowtie V_{L-1} \bowtie V_L$
9: **end function**

In exact arithmetic, we have $\tau(\mathsf{V}) = \tau(\mathsf{U})$ for any $\mathsf{U} \in \mathrm{TT}_L^S$ with $L, S \in \mathbb{N}$ and $\mathsf{V} = \mathrm{orth}^-(\mathsf{U})$, and this is the view adhered to in the references cited above. However, the situation is drastically different when errors are introduced (e.g., due to round-off) in the course of orthogonalization, namely in lines 4 and 6 of Algorithm 3.1.

3.7 Low-Rank Multilevel Decomposition of Vectors and Matrices

Here, we discuss how we use the tensor train decomposition for the resolution of low-rank multilevel structure in vectors and matrices involved in the solution of (4).

To reorder the entries of Kronecker products, we use particular permutation matrices defined as follows. First, for every $L \in \mathbb{N}$, we define $\boldsymbol{\Pi}_L$ as the permutation matrix of order 2^{DL} such that

$$(\boldsymbol{\Pi}_L)_{i_{1,1},\ldots,i_{D,1},\ldots\ldots,i_{1,L},\ldots,i_{D,L}\;\; i_{1,1},\ldots,i_{1,L},\ldots\ldots,i_{D,1},\ldots,i_{D,L}} = 1 \tag{49}$$

for all $i_{k,\ell} = 1, 2$ with $k = 1, \ldots, D$ and $\ell = 1, \ldots, L$. In our present setting, we are interested in functions

$$u_L = \sum_{j \in \mathscr{J}_L} (\bar{\boldsymbol{u}}_L)_j \, \varphi_{L,j} \in V_L \tag{50a}$$

whose coefficients admit low-rank TT representations in the following sense:

$$\boldsymbol{\Pi}_L \, \bar{\boldsymbol{u}}_L = \tau(\mathsf{U}) = U_1 \bowtie \cdots \bowtie U_L \tag{50b}$$

with some $\mathsf{U} = (U_1, \ldots, U_L)$.

The set $\mathscr{J}_L$, which is defined by (16), is isomorphic to $\{1, 2\}^{DL}$. The matrix $\boldsymbol{\Pi}_L$, when applied to a vector whose components are indexed by $\mathscr{J}_L$, folds the vector into a DL-dimensional array, transposes the DL indices according to the transformation of ordering in the product $\{1, \ldots, D\} \times \{1, \ldots, L\}$ from big-endian to little-endian and unfolds the resulting array back into a vector.

In other words, the matrix $\boldsymbol{\Pi}_L$, acting on a vector whose entries are enumerated so that the indices corresponding to each dimension and all of the levels occur one after another, rearranges the entries in such a way that the indices corresponding to each level and all of the dimensions occur one after another. In the present paper, we will use $\boldsymbol{\Pi}_L$ to permute the rows and columns of matrices, as the following example illustrates.

Example 2 In the case of $D = 2$ and $L = 3$, the following relation holds:

$$\boldsymbol{\Pi}_L \Big(\underbrace{I \otimes J \otimes J^{\mathsf{T}}}_{\text{dimension 1}} \otimes \underbrace{I \otimes I_1 \otimes I_2}_{\text{dimension 2}} \Big) \boldsymbol{\Pi}_L^{\mathsf{T}} = \underbrace{I \otimes I}_{\text{level 1}} \otimes \underbrace{J \otimes I_1}_{\text{level 2}} \otimes \underbrace{J^{\mathsf{T}} \otimes I_2}_{\text{level 3}},$$

where we use the matrices that we defined in (37) above.

Similarly, for every $L \in \mathbb{N}$ and $\alpha \in \{0, 1\}^D$, we introduce $\widetilde{\boldsymbol{\Pi}}_{L,\alpha}$ as a permutation matrix of order $2^{D(L+1)-|\alpha|}$ with rows and columns indexed by $\mathscr{J}_L \times \{\alpha_1, 1\} \times \cdots \times \{\alpha_D, 1\}$, where $\mathscr{J}_L$ is given by (16). Specifically, we define $\widetilde{\boldsymbol{\Pi}}_{L,\alpha}$ by

$$(\widetilde{\boldsymbol{\Pi}}_{L,\alpha})_{i_{1,1},\ldots,i_{1,D},\ldots\ldots,i_{1,L},\ldots,i_{D,L},\beta_1,\ldots,\beta_D \;\; i_{1,1},\ldots,i_{1,L},\beta_1,\ldots\ldots,i_{D,1},\ldots,i_{D,L},\beta_D} = 1 \tag{51}$$

for all $i_{k,\ell} = 1, 2$ with $k = 1, \ldots, D$ and $\ell = 1, \ldots, L$ and for all $\beta_k \in \{\alpha_k, 1\}$ with $k = 1, \ldots, D$.

4 Representation Conditioning

Since the TT decomposition is based on low-rank matrix factorization, redundancy (linear dependence) in explicit TT representations can be eliminated analytically. This is illustrated in "Appendix B": see (114a)–(114c) and, for more practical examples, the proof of Lemma 5. On the other hand, in the course of computations, this reduction

has to be done numerically. In exact arithmetic, it can always be achieved by the TT rounding algorithm [45, Algorithm 2] using the TT-SVD. In practice, however, it may fail due to round-off errors: a small perturbation of a single core in a TT decomposition may, through catastrophic cancellations, introduce a large perturbation in the represented tensor. This can occur even in the course of orthogonalization (Algorithm 3.1), which is essential for ensuring the stability of the TT rounding algorithm. We now turn to an analysis of the potential for such error amplification, which we refer to as *representation conditioning*.

4.1 Examples of Ill-Conditioned Tensor Representations

We first consider a simple example of a tensor where relative perturbations on the order of the machine precision can lead to large changes in the represented tensors.

Example 3 Take $D = 1$ (so that $\mathscr{I} = \{0, 1\}$) and let $\boldsymbol{x}$ be the tensor with all entries equal to one, $\boldsymbol{x}_{i_1,\dots,i_L} = 1$ for $i_1, \dots, i_L \in \mathscr{I}$. Clearly, $\boldsymbol{x}$ can be represented by $\mathsf{X} = (X_\ell)_{\ell=1,\dots,L}$ with $\operatorname{ranks}(\mathsf{X}) = (1, \dots, 1)$, where $X_\ell = [(1, 1)^\mathsf{T}]$ for each ℓ. However, we also have an alternative representation Y with $\operatorname{ranks}(\mathsf{Y}) = (2, \dots, 2)$: for any fixed $R > 0$ and $y_0 = (0, 0)^\mathsf{T}$, $y_R = (R, R)^\mathsf{T}$, we instead set

$$Y_1 = \begin{bmatrix}(1 + R^{-L})y_R & -y_R\end{bmatrix}, \quad Y_2 = \dots = Y_{L-1} = \begin{bmatrix} y_R & y_0 \\ y_0 & y_R \end{bmatrix}, \quad Y_L = \begin{bmatrix} y_R \\ y_R \end{bmatrix}. \tag{52}$$

For $\varepsilon > 0$, consider a perturbation of Y by replacing Y_ℓ for some fixed $1 < \ell < L$ by

$$\tilde{Y}_\ell = \begin{bmatrix} (1 + \varepsilon)y_R & y_0 \\ y_0 & y_R \end{bmatrix}.$$

This corresponds to a relative error of order ε with respect to $\|Y_\ell\|_2$. The resulting perturbed tensor $\boldsymbol{x}_\varepsilon$ is again constant with entries $1+(R^L+1)\varepsilon$, and therefore satisfies

$$\frac{\|\boldsymbol{x} - \boldsymbol{x}_\varepsilon\|_2}{\|\boldsymbol{x}\|_2} = (R^L + 1)\varepsilon. \tag{53}$$

For instance, with $R = 4$ and $L \geq 25$, we obtain $R^L > 10^{15}$. Consequently, any numerical manipulation of the representations can then lead to very large round-off errors that leave no significant digits in the output; in particular, an automatic rank reduction of the representation by SVD will in general not produce any useful result.

To illustrate this numerically, we consider the left-orthogonalization $\operatorname{orth}^-(\mathsf{Y})$ with $R = 4$ and machine precision $\varepsilon \approx 2 \times 10^{-16}$, which is also the first step in computing the TT-SVD. In exact arithmetic, the tensor $\tau(\operatorname{orth}^-(\mathsf{Y}))$ is identical to $\tau(\mathsf{Y})$; however, in inexact arithmetic, this can be far from true. The associated relative numerical errors are compared to bound (53) in Table 1. We consider two ways of evaluating the difference in ℓ^2-norm: by extracting all tensor entries and computing the norm of their differences directly, or by assembling the difference in TT format and computing its norm using another orthogonalization. Due to numerical effects, the resulting values

Table 1 Relative errors $\|\tau(\mathsf{Y}) - \tau(\mathrm{orth}^{-}(\mathsf{Y}))\|_2 / \|\tau(\mathsf{Y})\|_2$ for Y as in Example 3 with $R = 4$, with difference computed using two different methods: (a) entrywise, (b) in TT format; compared to $(R^L + 1)\varepsilon$

	$L = 5$	$L = 10$	$L = 15$	$L = 20$	$L = 25$
diff. (a)	4.17×10^{-13}	6.06×10^{-10}	6.95×10^{-07}	9.64×10^{-04}	9.48×10^{-01}
diff. (b)	3.51×10^{-13}	3.82×10^{-10}	7.10×10^{-07}	7.02×10^{-04}	$1.07 \times 10^{+00}$
$(R^L + 1)\varepsilon$	2.28×10^{-13}	2.33×10^{-10}	2.38×10^{-07}	2.44×10^{-04}	2.50×10^{-01}

are not identical, but agree in their order of magnitude, which is also the same as predicted for a particular perturbation by (53).

The type of instability observed in Example 3 occurs in a similar way in other operations, for instance in the computation of inner products, or even in the extraction of a single entry of the tensor. Due to its fundamental importance in many algorithms, we use orthogonalization as an illustrative example in what follows.

Example 3 may seem artificial, since in the explicit construction of tensor representations one will usually try to avoid such redundant representations that can cause cancellations. However, redundancies of this kind may also be generated when matrix-vector products are performed. We next consider an example of practical relevance where a matrix and a vector are each given in a multilevel tensor representation of minimal ranks, but the resulting representation of their product has a similar ill-conditioning as the previous example.

Example 4 We consider the negative Laplacian with homogeneous Dirichlet boundary conditions on $(0, 1)$, discretized by piecewise linear finite elements on a uniform grid with 2^L interior nodes. The resulting stiffness matrix $\boldsymbol{A}_L^{\mathrm{DD}} \in \mathbb{R}^{2^L \times 2^L}$ satisfies $\boldsymbol{A}_L^{\mathrm{DD}} = A_1 \bowtie \cdots \bowtie A_L$ with $A_1 = 4\begin{bmatrix} I & J^{\mathsf{T}} & J \end{bmatrix}$,

$$A_2 = \cdots = A_{L-1} = 4\begin{bmatrix} I & J^{\mathsf{T}} & J \\ & J & \\ & & J^{\mathsf{T}} \end{bmatrix} \quad \text{and} \quad A_L = 4H^2\begin{bmatrix} 2I - J - J^{\mathsf{T}} \\ -J \\ -J^{\mathsf{T}} \end{bmatrix}, \tag{54}$$

as derived in [35, Cor. 3.2], where $H = 1 + 2^{-L}$ and the elementary blocks are as defined in (37). The first eigenvector of $\boldsymbol{A}_L^{\mathrm{DD}}$, corresponding to the lowest eigenvalue $\lambda_{\min,L} \approx \pi^2$, is $\boldsymbol{x}_{\min,L} = \big(\sin(\pi i 2^{-L}/H)\big)_{i=1,\ldots,2^L} = X_1 \bowtie \cdots \bowtie X_L$, where

$$X_1 = \begin{bmatrix} x_{\mathrm{c}}^1 & x_{\mathrm{s}}^1 \end{bmatrix}, \qquad X_\ell = \begin{bmatrix} x_{\mathrm{c}}^\ell & x_{\mathrm{s}}^\ell \\ -x_{\mathrm{s}}^\ell & x_{\mathrm{c}}^\ell \end{bmatrix} \quad \text{for} \quad \ell = 2, \ldots, L-1, \qquad X_L = \begin{bmatrix} \hat{x}_{\mathrm{s}}^\ell \\ \hat{x}_{\mathrm{c}}^\ell \end{bmatrix}, \tag{55}$$

with $t_\ell = \pi 2^{-\ell}/H$,

$$x_{\mathrm{c}}^\ell = \begin{pmatrix} 1 \\ \cos(t_\ell) \end{pmatrix}, \quad x_{\mathrm{s}}^\ell = \begin{pmatrix} 0 \\ \sin(t_\ell) \end{pmatrix}, \quad \hat{x}_{\mathrm{c}}^\ell = \begin{pmatrix} \cos(t_L) \\ \cos(2t_L) \end{pmatrix}, \quad \hat{x}_{\mathrm{s}}^\ell = \begin{pmatrix} \sin(t_L) \\ \sin(2t_L)) \end{pmatrix}.$$

Table 2 Relative errors $e_{\mathsf{A}\bullet\mathsf{V}} = \|\tau(\mathsf{A}\bullet\mathsf{V}) - \tau(\mathrm{orth}^-(\mathsf{A}\bullet\mathsf{V}))\|_2/\|\tau(\mathsf{A}\bullet\mathsf{V})\|_2$ compared to $e_{\mathsf{V}} = \|\tau(\mathsf{V}) - \tau(\mathrm{orth}^-(\mathsf{V}))\|_2/\|\tau(\mathsf{V})\|_2$, for A, V as in Example 4

	$L = 20$	$L = 25$	$L = 30$	$L = 35$	$L = 40$
e_{V}	1.70×10^{-15}	1.42×10^{-15}	1.92×10^{-15}	3.15×10^{-15}	2.73×10^{-15}
$e_{\mathsf{A}\bullet\mathsf{V}}$	2.97×10^{-05}	4.50×10^{-02}	$4.21 \times 10^{+01}$	$3.46 \times 10^{+04}$	$4.05 \times 10^{+07}$
$2^{2L}\varepsilon$	2.44×10^{-04}	2.50×10^{-01}	$2.56 \times 10^{+02}$	$2.62 \times 10^{+05}$	$2.68 \times 10^{+08}$

Then, the representation $\mathsf{A}\bullet\mathsf{X}$ of the matrix-vector product $\boldsymbol{A}_L^{\mathrm{DD}}\boldsymbol{x}_{\min,L}$ in exact arithmetic satisfies $\tau(\mathsf{A}\bullet\mathsf{X}) = \boldsymbol{A}_L^{\mathrm{DD}}\boldsymbol{x}_{\min,L} = \lambda_{\min,L}\boldsymbol{x}_{\min,L} = \lambda_{\min,L}\tau(\mathsf{X})$.

We consider a similar numerical test as in Example 3, comparing the relative error in $\mathrm{orth}^-(\mathsf{A}\bullet\mathsf{X})$ to that of $\mathrm{orth}^-(\mathsf{X})$. The results are given in Table 2, where differences are computed in the TT format. Whereas the numerical manipulation of X leads to errors close to the machine precision ε, in $\mathrm{orth}^-(\mathsf{A}\bullet\mathsf{X})$ we observe large relative errors of order $2^{2L}\varepsilon$. Note that the representation (54) of A_L^{DD} has a similar structure as the redundant representation (52) in the previous example: the cores $A_1, \ldots, A_{L-1}$ have only positive entries, whereas A_L can introduce cancellations, in particular when the matrix is applied to low-frequency grid functions as above.

4.2 Representation Amplification Factors and Condition Numbers

We now introduce a quantitative measure for the stability of TT representations under numerical manipulations. To first order in the size of the perturbation, it is determined by the relative condition numbers of the multilinear mapping τ with respect to the component tensors in its argument. Here, we use the appropriate metric on the components that corresponds to the above-considered perturbations arising in linear algebra operations.

Definition 3 We define the *representation amplification factors* of $\mathsf{X} \in \mathrm{TT}_L$, for $\ell = 1, \ldots, L$, by

$$\mathrm{ramp}_\ell(\mathsf{X}) = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\sup\Big\{\|\tau(\tilde{\mathsf{X}}) - \tau(\mathsf{X})\|_2 : \tilde{\mathsf{X}} \in \mathrm{TT}_L, \\ \|\tilde{X}_\ell - X_\ell\|_2 \le \varepsilon\|X_\ell\|_2 \text{ and } \tilde{X}_k = X_k \text{ for } k \ne \ell\Big\}, \tag{56}$$

and the *representation condition numbers* by

$$\mathrm{rcond}_\ell(\mathsf{X}) = \frac{\mathrm{ramp}_\ell(\mathsf{X})}{\|\tau(\mathsf{X})\|_2}. \tag{57}$$

By the multilinearity of τ, if $\mathsf{X}, \tilde{\mathsf{X}} \in \mathrm{TT}_L$ with $\boldsymbol{x} = \tau(\mathsf{X})$, $\tilde{\boldsymbol{x}} = \tau(\tilde{\mathsf{X}})$ are such that $\|\tilde{X}_\ell - X_\ell\|_2 \le \varepsilon\|X_\ell\|_2$ for each ℓ, then for such relative perturbations of size ε of

cores we have the bounds

$$\|\boldsymbol{x}-\tilde{\boldsymbol{x}}\|_2 \leq \sum_{\ell=1}^{L} \mathrm{ramp}_\ell(\mathsf{X})\,\varepsilon + \mathscr{O}(\varepsilon^2), \quad \frac{\|\boldsymbol{x}-\tilde{\boldsymbol{x}}\|_2}{\|\boldsymbol{x}\|_2} \leq \sum_{\ell=1}^{L} \mathrm{rcond}_\ell(\mathsf{X})\,\varepsilon + \mathscr{O}(\varepsilon^2).$$

In the following characterization, we use the notation τ_ℓ^- and τ_ℓ^+ for left and right partial matricizations as introduced in (43c)–(43d).

Proposition 1 *For any $X \in \mathrm{TT}_L$ and $\ell = 1, \dots, L$,*

$$\mathrm{ramp}_\ell(X) = \|\tau_\ell^-(X)\|_{2\to 2}\|X_\ell\|_2\|\tau_\ell^+(X)\|_{2\to 2}.$$

Proof For fixed ℓ in (56), let $\mathsf{X}, \tilde{\mathsf{X}}$ satisfy the conditions in the supremum. Then,

$$\begin{aligned}\|\tau(\tilde{\mathsf{X}}) - \tau(\mathsf{X})\|_2^2 &= \sum_{i_1,\dots,i_L} \left[X_1^{\{i_1\}} \cdots \left(X_\ell^{\{i_\ell\}} - \tilde{X}_\ell^{\{i_\ell\}}\right) \cdots X_L^{\{i_L\}}\right]^2 \\ &= \sum_{i_\ell} \left\|\tau_\ell^-(\mathsf{X})(X_\ell^{\{i_\ell\}} - \tilde{X}_\ell^{\{i_\ell\}})\tau_\ell^+(\mathsf{X})\right\|_2^2.\end{aligned}$$

The claim thus follows by taking the supremum over $\tilde{X}_\ell$ such that $\sum_{i\in\mathscr{I}} \|\tilde{X}_\ell^{\{i\}} - X_\ell^{\{i\}}\|_F^2 \leq \varepsilon^2\|X_\ell\|_2^2$, which is in fact attained. □

Remark 3 The quantities in Definition 3 measuring the amplification of perturbations can be defined in an analogous way for more general tensor networks by considering perturbations in the respective components; see [7,22,42,49] for an overview on such more general tensor formats.

We have the following general observations concerning possible representation condition numbers, where in certain special cases, we can also give bounds that depend only on the TT ranks. Here, we use the notion of TT-SVD forms introduced in Sect. 3.5.

Proposition 2 *Let $X \in \mathrm{TT}_L$, then the following hold for $\ell = 1, \dots, L$.*

(i) *One has* $\mathrm{rcond}_\ell(X) \geq 1$.
(ii) *If* $\mathrm{rank}_{\ell-1}(X) = \mathrm{rank}_\ell(X) = 1$, *then* $\mathrm{rcond}_\ell(X) = 1$.
(iii) *If X is in right-orthogonal TT-SVD form, then* $\mathrm{rcond}_\ell(X) \leq \sqrt{\mathrm{rank}_{\ell-1}(X)}$; *if it is in left-orthogonal TT-SVD form, then* $\mathrm{rcond}_\ell(X) \leq \sqrt{\mathrm{rank}_\ell(X)}$.

Proof Statement (i) follows by estimating $\|\tau(\mathsf{X})\|_2$ as in the proof of Proposition 1; (ii) follows directly from properties of the Kronecker product. To show (iii), it suffices to consider the right-orthogonal case. With $\boldsymbol{x} = \tau(\mathsf{X})$ and $r_\ell = \mathrm{rank}_\ell(\mathsf{X})$ for each ℓ, we need to show that $\mathrm{ramp}_\ell(\mathsf{X}) \leq \sqrt{r_{\ell-1}}\|\boldsymbol{x}\|_2$ for each ℓ. Since $\tau_\ell^+(\mathsf{X})$ has orthonormal rows, $\|\tau_\ell^+(\mathsf{X})\|_{2\to 2} = 1$ for each ℓ. For $\ell = 1$, we also have $\|\tau_\ell^-(\mathsf{X})\|_{2\to 2} = 1$ by definition and $\|X_\ell\|_2 = \|\boldsymbol{x}\|_2$. For $\ell > 1$, by right-orthogonality of X_ℓ we have $\|X_\ell\|_2 = \sqrt{r_{\ell-1}}$. In this case, since the representation is in TT-SVD form, $\tau_\ell^-(\mathsf{X})$ has orthogonal columns whose ℓ^2-norms are the singular values of $\mathrm{U}_\ell(\boldsymbol{x})$, and thus $\|\tau_\ell^-(\mathsf{X})\|_{2\to 2} \leq \|\boldsymbol{x}\|_2$. □

Modifications to the components of a TT representation that leave the represented tensor unchanged can still lead to a change in the representation condition numbers. This change can be bounded from above as follows.

Proposition 3 *For given $X \in \mathrm{TT}_L$, $1 \le \ell < L$, and invertible $R \in \mathbb{R}^{r_\ell \times r_\ell}$, where $r_\ell = \mathrm{rank}_\ell(X)$, let $\tilde{X}$ be identical to X except for $\tilde{X}_\ell^{\{i\}} = X_\ell^{\{i\}} R$, $\tilde{X}_{\ell+1}^{\{i\}} = R^{-1} X_{\ell+1}^{\{i\}}$ for $i \in \mathscr{I}$. Then, $\tau(X) = \tau(\tilde{X})$ and*

$$\mathrm{ramp}_\ell(\tilde{X}) \le \mathrm{cond}(R)\,\mathrm{ramp}_\ell(X), \quad \mathrm{ramp}_{\ell+1}(\tilde{X}) \le \mathrm{cond}(R)\,\mathrm{ramp}_{\ell+1}(X). \tag{58}$$

In the particular case when the matrix $\tau_\ell^+(\tilde{X})$ has orthonormal rows, one has the stronger bound

$$\mathrm{ramp}_\ell(\tilde{X}) \le \mathrm{ramp}_\ell(X). \tag{59}$$

If $\tilde{X}_{\ell+1}$ is right-orthogonal, then

$$\mathrm{ramp}_{\ell+1}(\tilde{X}) \le \sqrt{r_\ell}\,\mathrm{ramp}_{\ell+1}(X). \tag{60}$$

Proof The estimates (58) follow from

$$\|\tilde{X}_\ell\|_2 \le \|X_\ell\|_2 \|R\|_{2\to 2}, \quad \|\tau_\ell^+(\tilde{\mathsf{X}})\|_2 \le \|R^{-1}\|_{2\to 2} \|\tau_\ell^+(\mathsf{X})\|_2$$

for the first, and analogous estimates for the second inequality. To see (59), observe that $R\tau_\ell^+(\tilde{\mathsf{X}}) = \tau_\ell^+(\mathsf{X})$ and that under the given additional assumption, $\|\tau_\ell^+(\tilde{\mathsf{X}})\|_{2\to 2} = 1$ and $\|\tau_\ell^+(\mathsf{X})\|_{2\to 2} = \|R\|_{2\to 2}$. Under the further assumption for (60), we have $\|X_{\ell+1}\|_2 = \|R\|_F$, and thus

$$\begin{aligned}\mathrm{ramp}_{\ell+1}(\tilde{\mathsf{X}}) &= \|\tau_{\ell+1}^-(\mathsf{X})R\|_{2\to 2}\|\tilde{X}_{\ell+1}\|_2\|\tau_{\ell+1}^+(\mathsf{X})\|_{2\to 2}\\ &\le \|\tau_{\ell+1}^-(\mathsf{X})\|_{2\to 2}\|R\|_F\sqrt{r_\ell}\|\tau_{\ell+1}^+(\mathsf{X})\|_{2\to 2}\\ &\le \sqrt{r_\ell}\ \mathrm{ramp}_{\ell+1}(\mathsf{X}).\end{aligned}$$

□

Note that the improved bounds (59) and (60), which do not depend on the particular transformation R, correspond to the transformations made in algorithms for right-orthogonalizing $\mathsf{X} \in \mathrm{TT}_L$. When the roles of $\tilde{X}_\ell$, $\tilde{X}_{\ell+1}$ and the corresponding orthogonality requirements are reversed, (59) and (60) are replaced by $\mathrm{ramp}_{\ell+1}(\tilde{\mathsf{X}}) \le \mathrm{ramp}_{\ell+1}(\mathsf{X})$ and $\mathrm{ramp}_\ell(\tilde{\mathsf{X}}) \le \sqrt{r_{\ell+1}}\ \mathrm{ramp}_\ell(\mathsf{X})$.

4.3 Orthogonalization as an Example of a Numerical Operation

Orthogonalization of tensor train representations is usually done via QR decompositions of matricized cores. When performed at machine precision ε, these decompositions are affected by round-off errors: applied to $M \in \mathbb{R}^{m\times n}$, where $mn\varepsilon$

is sufficiently small, as shown in [27, §19] the standard Householder algorithm yields $\tilde{Q}, \tilde{R}$ such that

$$\|M - \tilde{Q}\tilde{R}\|_F \leq C_{\mathrm{QR}} mn^{3/2} \varepsilon \|M\|_F. \tag{61}$$

As a consequence of Proposition 3, we obtain a statement on the numerical errors incurred by orthogonalization of TT representations. As a simplifying assumption, let us suppose that the QR factorizations in $\mathrm{orth}^-(\mathsf{X})$, $\mathrm{orth}^+(\mathsf{X})$ of $\mathsf{X} \in \mathrm{TT}_L$ are computed with machine precision ε up to the error bound (61), but that matrix–matrix multiplications are performed exactly (and hence the computed Householder reflectors act as exactly orthogonal matrices). Then, recursively using (59), (60), we obtain

$$\|\tau(\mathrm{orth}^+(\mathsf{X})) - \tau(\mathsf{X})\|_2 \leq C_{\mathrm{QR}} \sum_{\ell=2}^{L} (2^D r_{\ell-1} r_\ell)^{3/2} \, \mathrm{ramp}_\ell(\mathsf{X}) \, \varepsilon + \mathscr{O}(\varepsilon^2), \tag{62}$$

$$\|\tau(\mathrm{orth}^-(\mathsf{X})) - \tau(\mathsf{X})\|_2 \leq C_{\mathrm{QR}} \sum_{\ell=1}^{L-1} (2^D r_{\ell-1} r_\ell)^{3/2} \, \mathrm{ramp}_\ell(\mathsf{X}) \varepsilon + \mathscr{O}(\varepsilon^2), \tag{63}$$

where $r_\ell = \mathrm{rank}_\ell(\mathsf{X})$ for $\ell = 1, \dots, L$. The analogous statements for the relative errors $\|\tau(\mathrm{orth}^+(\mathsf{X})) - \tau(\mathsf{X})\|_2 / \|\tau(\mathsf{X})\|_2$ and $\|\tau(\mathrm{orth}^-(\mathsf{X})) - \tau(\mathsf{X})\|_2 / \|\tau(\mathsf{X})\|_2$ hold with ramp replaced by rcond.

Taking into account further numerical effects due to inexact matrix–matrix multiplications leads to substantially more complicated bounds involving additional prefactors depending more strongly on intermediate steps in the algorithms. As our numerical illustrations in Sect. 4.1 demonstrate, however, the order of magnitude of the resulting errors is typically already very well predicted by the bounds (62), (63).

4.4 Representations of Operators

Definition 4 For $\ell = 1, \dots, L$, we define the representation amplification factor and representation condition number of the matrix representation $\mathsf{A} \in \mathrm{TT}_L^2$ by

$$\mathrm{mramp}_\ell(\mathsf{A}) = \sup_{\mathsf{X} \in \mathrm{TT}_L} \frac{\mathrm{ramp}_\ell(\mathsf{A} \bullet \mathsf{X})}{\mathrm{ramp}_\ell(\mathsf{X})}, \quad \mathrm{mrcond}_\ell(\mathsf{A}) = \sup_{\mathsf{X} \in \mathrm{TT}_L} \frac{\mathrm{rcond}_\ell(\mathsf{A} \bullet \mathsf{X})}{\mathrm{rcond}_\ell(\mathsf{X})}. \tag{64}$$

In other words, these are the largest factors by which the action of the matrix representation A can possibly change the representation amplification factors and the condition numbers of a vector representation. By definition, these functions are submultiplicative:

$$\mathrm{mramp}_\ell(\mathsf{A} \bullet \mathsf{B}) \leq \mathrm{mramp}_\ell(\mathsf{A}) \, \mathrm{mramp}_\ell(\mathsf{B}),$$
$$\mathrm{mrcond}_\ell(\mathsf{A} \bullet \mathsf{B}) \leq \mathrm{mrcond}_\ell(\mathsf{A}) \, \mathrm{mrcond}_\ell(\mathsf{B}).$$

We do not have an explicit representation of these quantities as in Proposition 1, but we obtain the following upper bound in terms of the components of representations.

Proposition 4 *For $A \in \mathrm{TT}_L^2$, we define the matrices*

$$\boldsymbol{A}_{\ell,k}^{-} = \big((A_1 \bowtie \cdots \bowtie A_{\ell-1})(1, i, j, k)\big)_{i\in\mathscr{I}^{\ell-1}, j\in\mathscr{I}^{\ell-1}}, \qquad k = 1, \dots, R_{\ell-1},$$
$$\boldsymbol{A}_{\ell,k}^{+} = \big((A_{\ell+1} \bowtie \cdots \bowtie A_L)(k, i, j, 1)\big)_{i\in\mathscr{I}^{L-\ell}, j\in\mathscr{I}^{L-\ell}}, \qquad k = 1, \dots, R_\ell.$$

Then, $\operatorname{mramp}_\ell(A) \le \beta_\ell(A)$ *for* $\ell = 1, \dots, L$, *where we define*

$$\beta_\ell(A) = \Big(\sum_{k^-=1}^{R_{\ell-1}} \|\boldsymbol{A}_{\ell,k^-}^{-}\|_{2\to2}^2 \sum_{k^+=1}^{R_\ell} \|\boldsymbol{A}_{\ell,k^+}^{+}\|_{2\to2}^2 \sum_{k^-=1}^{R_{\ell-1}} \sum_{k^+=1}^{R_\ell} \big\|A_\ell^{[k^-,k^+]}\big\|_{2\to2}^2\Big)^{\frac{1}{2}}, \quad (65)$$

and if $\tau(A)$ *is invertible,*

$$\operatorname{mrcond}_\ell(A) \le \|\tau(A)^{-1}\|_{2\to2} \operatorname{mramp}_\ell(A). \quad (66)$$

Proof By Proposition 1, with $\mathsf{Y} = \mathsf{A} \bullet \mathsf{X}$,

$$\operatorname{mramp}_\ell(\mathsf{A}) = \sup_{\mathsf{X}\in\mathrm{TT}_L} \frac{\|\tau_\ell^-(\mathsf{Y})\|_{2\to2}\|\tau_\ell^+(\mathsf{Y})\|_{2\to2}\|Y_\ell\|_2}{\|\tau_\ell^-(\mathsf{X})\|_{2\to2}\|\tau_\ell^+(\mathsf{X})\|_{2\to2}\|X_\ell\|_2}.$$

The first statement follows with the estimates

$$\begin{aligned}\|Y_\ell\|_2^2 &= \sum_{K^-=1}^{R_{\ell-1}} \sum_{K^+=1}^{R_\ell} \sum_{k^-=1}^{r_{\ell-1}} \sum_{k^+=1}^{r_\ell} \|A_\ell^{[K^-,K^+]}\, X_\ell^{[k^-,k^+]}\|_2^2 \\ &\le \sum_{K^-=1}^{R_{\ell-1}} \sum_{K^+=1}^{R_\ell} \big\|A_\ell^{[K^-,K^+]}\big\|_{2\to2}^2\, \|X_\ell\|_2^2\end{aligned}$$

and

$$\|\tau_\ell^-(\mathsf{Y})\|_{2\to2}^2 \le \sup_{\|y\|_2=1} \sum_{k=1}^{R_{\ell-1}} \big\|\boldsymbol{A}_{\ell,k}^{-}\, \tau_\ell^-(\mathsf{X})\, y\big\|_2^2 \le \sum_{k=1}^{R_{\ell-1}} \|\boldsymbol{A}_{\ell,k}^{-}\|_{2\to2}^2 \|\tau_\ell^-(\mathsf{X})\|_{2\to2}^2,$$

as well as the analogous bound for $\tau_\ell^+(\mathsf{Y})$. For (66), note that if $\tau(\mathsf{A})$ is invertible, then

$$\operatorname{mrcond}_\ell(\mathsf{A}) \le \left(\sup_{\mathsf{X}\in\mathrm{TT}_L} \frac{\|\tau(\mathsf{X})\|_2}{\|\tau(\mathsf{Y})\|_2}\right) \operatorname{mramp}_\ell(\mathsf{A}) = \|\tau(\mathsf{A})^{-1}\|_{2\to2} \operatorname{mramp}_\ell(\mathsf{A}).$$

□

In certain situations, Proposition 4 provides qualitatively sharp bounds. We now demonstrate this in the simple example of the stiffness matrix for the Dirichlet Laplacian on $(0, 1)$. Similar results are observed numerically for direct representations of more general stiffness matrices of second-order elliptic problems.

Proposition 5 *Let $\boldsymbol{A}_L^{\mathrm{DD}}$ be as in Example 4, and let A with $\tau(\mathsf{A}) = \boldsymbol{A}_L^{\mathrm{DD}}$ be as in (54). Then, for $\ell = 1, \dots, L$, one has $\operatorname{mramp}_\ell(\mathsf{A}) \sim 2^{2L}$ and $2^{(3L+\ell)/2} \lesssim \operatorname{mrcond}_\ell(\mathsf{A}) \lesssim 2^{2L}$.*

Proof The upper bounds follow by direct computation from Proposition 4 via evaluation of the auxiliary matrices in (65). For the lower bound on $\operatorname{mramp}_\ell(\mathsf{A})$, we estimate the supremum from below using the representation $\mathsf{X}_{\max}$ analogous to (55) of the eigenvector $\boldsymbol{x}_{\max,L} = \big(\sin(\pi i/H)\big)_{i=1,\dots,2^L}$ corresponding to the largest eigenvalue $\lambda_{\max,L} \sim 2^{2L}$. To this end, it suffices to evaluate $\operatorname{ramp}_\ell(\mathsf{A} \bullet \mathsf{X}_{\max})/\operatorname{ramp}_\ell(\mathsf{X}_{\max})$ via Proposition 1 in a direct but tedious calculation. For the lower bound on $\operatorname{mrcond}_\ell(\mathsf{A})$, we instead use $\boldsymbol{x}_{\min,L} = \big(\sin(\pi i 2^{-L}/H)\big)_{i=1,\dots,2^L}$ in the representation (55). □

Thus, applying the matrix representation A to the tensor decomposition X of a vector may in general increase its representation condition number by a factor proportional to 2^{2L}. For instance, if X is given in TT-SVD form with representation condition number close to one, the further numerical manipulation of $\mathsf{A} \bullet \mathsf{X}$ can cause errors of order $\mathcal{O}(2^{2L}\varepsilon\|\tau(\mathsf{X})\|_2)$. This effect is observed also in the numerical tests in Sect. 7.1.

5 Multilevel Low-Rank Tensor Structure of the Operator

In this section, we analyze the low-rank structure of the preconditioner $\boldsymbol{C}_L$, given by (30), and of the preconditioned discrete differential operator $\boldsymbol{B}_L$ in the form of (33b). The resulting low-rank representations are designed specifically to have small representation condition numbers in the sense of Definition 4, which is not generally the case for low-rank decompositions of $\boldsymbol{B}_L$.

The central idea for obtaining well-conditioned representations is to directly combine the representations of differential operators $\hat{\boldsymbol{M}}_{L,\alpha}$ as in (12d) with those of the averaging matrices $\hat{\boldsymbol{P}}_{\ell,L}$ in the preconditioner. This leads to a natural rank-reduced representation of the products $\hat{\boldsymbol{M}}_{L,\alpha}\hat{\boldsymbol{P}}_{\ell,L}$, where the cancellations causing representation ill-conditioning that are present in the tensor decomposition of $\hat{\boldsymbol{M}}_{L,\alpha}$ are explicitly absorbed by the preconditioner and thus removed from the final representation.

5.1 Auxiliary Results

In this section, for $\ell \in \mathbb{N}$, we present explicit joint representations of the identity matrix $\hat{\boldsymbol{I}}_\ell$ and of the shift matrix $\hat{\boldsymbol{S}}_\ell$, given by (12e), and of the linear vectors $\hat{\boldsymbol{\xi}}_\ell$ and $\hat{\boldsymbol{\eta}}_\ell$, defined in (15). These representations will be presented here in terms of the following cores:

$$\hat{U} = \begin{bmatrix} I & J^{\mathsf{T}} \\ & J \end{bmatrix}, \qquad \hat{X} = \frac{1}{2}\begin{bmatrix} \begin{pmatrix}1\\2\end{pmatrix} & \begin{pmatrix}0\\1\end{pmatrix} \\ \begin{pmatrix}1\\0\end{pmatrix} & \begin{pmatrix}2\\1\end{pmatrix} \end{bmatrix} \quad \text{and} \quad \hat{P} = \begin{bmatrix}1\\0\end{bmatrix}. \tag{67}$$

Our derivations will also involve the square Kronecker product matrices

$$\hat{\boldsymbol{J}}_\ell = J^{\otimes \ell} = \begin{pmatrix} 0 & & 1 \\ & \ddots & \\ 1 & & 0 \end{pmatrix} \tag{68}$$

with $\ell \in \mathbb{N}$ and iterated strong Kronecker products, such as $\hat{U}^{\bowtie \ell} = U \bowtie \cdots \bowtie U$ with $\ell \in \mathbb{N}$ factors.

We start with the following auxiliary result, which appeared in slightly different forms in [30,35]. The brief derivation, in the form given here, provides an illustration and simplifies further proofs given below.

Lemma 1 *For every $\ell \in \mathbb{N}$, the matrices $\hat{\boldsymbol{I}}_\ell$, $\hat{\boldsymbol{S}}_\ell$ and $\hat{\boldsymbol{J}}_\ell$, given by (12e) and (68), satisfy*

$$\begin{bmatrix} \hat{\boldsymbol{I}}_\ell & \hat{\boldsymbol{S}}_\ell \\ & \hat{\boldsymbol{J}}_\ell \end{bmatrix} = \hat{U}^{\bowtie \ell} \equiv \begin{bmatrix} I & J^T \\ & J \end{bmatrix}^{\bowtie \ell}, \tag{69}$$

where the blocks I and J are given by (37) and the core $\hat{U}$ is as defined in (67).

Proof For $\ell = 1$, the claim is trivial. Let us assume that $\ell > 1$. Then, splitting each of the matrices $\hat{\boldsymbol{S}}_\ell$, $\hat{\boldsymbol{I}}_\ell$ and $\hat{\boldsymbol{J}}_\ell$ into four blocks, we obtain the following recurrence relations:

$$\begin{aligned} \hat{\boldsymbol{I}}_\ell &= I \otimes \hat{\boldsymbol{I}}_{\ell-1} = \begin{bmatrix} I \end{bmatrix} \bowtie \begin{bmatrix} \hat{\boldsymbol{I}}_{\ell-1} \end{bmatrix}, \\ \hat{\boldsymbol{S}}_\ell &= I \otimes \hat{\boldsymbol{S}}_{\ell-1} + J^{\mathsf{T}} \otimes \hat{\boldsymbol{J}}_{\ell-1} = \begin{bmatrix} I & J^{\mathsf{T}} \end{bmatrix} \bowtie \begin{bmatrix} \hat{\boldsymbol{S}}_{\ell-1} \\ \hat{\boldsymbol{J}}_{\ell-1} \end{bmatrix}, \\ \hat{\boldsymbol{J}}_\ell &= J \otimes \hat{\boldsymbol{J}}_{\ell-1} = \begin{bmatrix} J \end{bmatrix} \bowtie \begin{bmatrix} \hat{\boldsymbol{J}}_{\ell-1} \end{bmatrix}. \end{aligned} \tag{70}$$

Using the core product, these relations can be recast as

$$\begin{bmatrix} \hat{\boldsymbol{I}}_\ell & \hat{\boldsymbol{S}}_\ell \\ & \hat{\boldsymbol{J}}_\ell \end{bmatrix} = \hat{U} \bowtie \begin{bmatrix} \hat{\boldsymbol{I}}_{\ell-1} & \hat{\boldsymbol{S}}_{\ell-1} \\ & \hat{\boldsymbol{J}}_{\ell-1} \end{bmatrix}. \tag{71}$$

Applying (71) recursively, we obtain (69). □

As the following auxiliary result shows, a similar technique applies to cores whose blocks are vectors.

Lemma 2 *For every $\ell \in \mathbb{N}_0$, the vectors $\hat{\boldsymbol{\xi}}_\ell$ and $\hat{\boldsymbol{\eta}}_\ell$, given by (15), satisfy*

$$\begin{bmatrix} & \hat{\boldsymbol{\eta}}_\ell \\ \hat{\boldsymbol{\xi}}_\ell & -\hat{\boldsymbol{\eta}}_\ell \end{bmatrix} = \hat{X}^{\bowtie \ell} \bowtie \hat{P}, \tag{72}$$

where $\hat{X}$ is given by (67).

Proof For $\ell = 0, 1$, the claim is trivial. Let us assume that $\ell > 1$. Splitting each of the vectors $\hat{\boldsymbol{\xi}}_\ell$ and $\hat{\boldsymbol{\eta}}_\ell$ into two blocks, we arrive at the recursion

$$\hat{\boldsymbol{\xi}}_\ell = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \otimes \hat{\boldsymbol{\xi}}_{\ell-1}, \qquad \hat{\boldsymbol{\eta}}_\ell = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} \otimes \hat{\boldsymbol{\eta}}_{\ell-1} + \begin{pmatrix} 0 \\ 1/2 \end{pmatrix} \otimes \hat{\boldsymbol{\xi}}_{\ell-1}, \tag{73a}$$

from which it is easy to see that

$$\begin{aligned} \hat{\boldsymbol{\eta}}_\ell &= \begin{pmatrix} 1/2 \\ 1 \end{pmatrix} \otimes \hat{\boldsymbol{\eta}}_{\ell-1} + \begin{pmatrix} 0 \\ 1/2 \end{pmatrix} \otimes (\hat{\boldsymbol{\xi}}_{\ell-1} - \hat{\boldsymbol{\eta}}_{\ell-1}) \,, \\ \hat{\boldsymbol{\xi}}_\ell - \hat{\boldsymbol{\eta}}_\ell &= \begin{pmatrix} 1/2 \\ 0 \end{pmatrix} \otimes \hat{\boldsymbol{\eta}}_{\ell-1} + \begin{pmatrix} 1 \\ 1/2 \end{pmatrix} \otimes (\hat{\boldsymbol{\xi}}_{\ell-1} - \hat{\boldsymbol{\eta}}_{\ell-1}) \,. \end{aligned} \tag{73b}$$

Using the core product, relations (73a) and (73b) can be recast as

$$\begin{bmatrix} \hat{\boldsymbol{\eta}}_\ell \\ \hat{\boldsymbol{\xi}}_\ell - \hat{\boldsymbol{\eta}}_\ell \end{bmatrix} = \hat{X} \bowtie \begin{bmatrix} \hat{\boldsymbol{\eta}}_{\ell-1} \\ \hat{\boldsymbol{\xi}}_{\ell-1} - \hat{\boldsymbol{\eta}}_{\ell-1} \end{bmatrix}. \tag{74}$$

Applying (74) recursively and comparing $\hat{\boldsymbol{\xi}}_1$ and $\hat{\boldsymbol{\eta}}_1$ with the first column of the core $\hat{X}$, which is given by $\hat{X} \bowtie \hat{P}$, we obtain (72). □

5.2 Explicit Analysis of Univariate Factors

In this section, we show how the auxiliary results of Sect. 5.1 translate into low-rank decompositions of the univariate factors $\hat{\boldsymbol{M}}_{L,\alpha}$ with $\alpha \in \{0, 1\}$ and $\hat{\boldsymbol{P}}_{\ell,L}$ with $\ell = 0, \dots, L$, where $L \in \mathbb{N}$. These matrices are introduced in (12d) and (14). Let

$$\begin{aligned} &\hat{A} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad \hat{T}_0 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \\ &\hat{V} = \frac{1}{2}\, \hat{T}_0 \bowtie \hat{U} \bowtie \hat{T}_0 = \frac{1}{2} \begin{bmatrix} I + J^{\mathsf{T}} + J & I - J^{\mathsf{T}} - J \\ I + J^{\mathsf{T}} - J & I - J^{\mathsf{T}} + J \end{bmatrix}, \quad \hat{M}_0 = \frac{1}{2} \begin{bmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{bmatrix}, \quad \hat{M}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \end{aligned} \tag{75}$$

Lemma 3 *For every $L \in \mathbb{N}$ and for $\alpha = 0, 1$, the matrix $\hat{\boldsymbol{M}}_{L,\alpha}$, given by (12d), satisfies*

$$\hat{\boldsymbol{M}}_{L,\alpha} = 2^{(\alpha+\frac{1}{2})L}\ \hat{A} \bowtie \hat{U}^{\bowtie \ell} \bowtie \hat{T}_0 \bowtie \hat{V}^{\bowtie (L-\ell)} \bowtie \hat{M}_\alpha \tag{76}$$

for every $\ell = 0, \dots, L$, where the cores $\hat{A}$, $\hat{U}$, $\hat{V}$, $\hat{T}_0$ and $\hat{M}_\alpha$ with $\alpha = 0, 1$ are given by (67) and (75).

Proof Consider $L \in \mathbb{N}$ and $\alpha \in \{0, 1\}$. Immediately from (12d), we obtain the representation

$$\hat{\boldsymbol{M}}_{L,\alpha} = 2^{(\alpha+\frac{1}{2})L}\ \hat{A} \bowtie \begin{bmatrix} \hat{\boldsymbol{I}}_L & \hat{\boldsymbol{S}}_L \\ & \hat{\boldsymbol{J}}_L \end{bmatrix} \bowtie \hat{T}_0 \bowtie \hat{M}_\alpha\,.$$

Applying Lemma 1, we arrive at the claimed decomposition in the case of $\ell = L$,

$$\hat{\boldsymbol{M}}_{L,\alpha} = 2^{(\alpha+\frac{1}{2})L}\ \hat{A} \bowtie \hat{U}^{\bowtie L} \bowtie \hat{T}_0 \bowtie \hat{M}_\alpha\,.$$

Using that $\hat{T}_0 \bowtie \hat{T}_0 = 2\hat{I}$, we obtain

$$\hat{\boldsymbol{M}}_{L,\alpha} = 2^{(\alpha+\frac{1}{2})L}\ \hat{A} \bowtie \hat{U}^{\bowtie \ell} \bowtie \hat{T}_0 \bowtie \left(\frac{1}{2}\hat{T}_0 \bowtie \hat{U} \bowtie \hat{T}_0\right)^{\bowtie(L-\ell)} \bowtie \hat{M}_\alpha$$

for every $\ell = 0, \ldots, L-1$, which completes the proof due to (75). □

Lemma 4 *For all $L \in \mathbb{N}_0$ and $\ell = 0, \ldots, L$, the matrix $\hat{\boldsymbol{P}}_{\ell,L}$, given by (14), has the representation*

$$\hat{\boldsymbol{P}}_{\ell,L} = 2^{-\frac{1}{2}(L-\ell)}\ \hat{A} \bowtie \hat{U}^{\bowtie \ell} \bowtie \hat{X}^{\bowtie(L-\ell)} \bowtie \hat{P}\,, \tag{77}$$

where $\hat{A}$, $\hat{U}$, $\hat{X}$ and $\hat{P}$ are the cores given by (67) and (75).

Proof We start with rewriting (14) in terms of the core product as

$$\hat{\boldsymbol{P}}_{\ell,L} = 2^{-\frac{1}{2}(L-\ell)}\ \hat{A} \bowtie \begin{bmatrix} \hat{\boldsymbol{I}}_\ell & \hat{\boldsymbol{S}}_\ell \\ & \hat{\boldsymbol{J}}_\ell \end{bmatrix} \bowtie \begin{bmatrix} \hat{\boldsymbol{\eta}}_{L-\ell} \\ \hat{\boldsymbol{\xi}}_{L-\ell} - \hat{\boldsymbol{\eta}}_{L-\ell} \end{bmatrix},$$

where the middle core should be omitted when $\ell = 0$. Applying Lemma 1 (for $\ell > 0$) and Lemma 2 to expand the middle and the last cores, we prove the claim. □

5.3 Explicit Analysis of Univariate Factors Under Preconditioning

Here, obtain an optimal-rank representation of the product $\boldsymbol{M}_{L,\alpha}\,\boldsymbol{P}_{\ell,L}$ and note how the products $\hat{\boldsymbol{M}}_{L,\alpha}\,\hat{\boldsymbol{P}}_{\ell,L}\,\hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}}$ and $\hat{\boldsymbol{P}}_{\ell,L}\,\hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}}$ can be represented, all for $L \in \mathbb{N}$, $\alpha \in \{0,1\}^D$ and $\ell = 0, \ldots, L$.

The optimal-rank representation of the product $\boldsymbol{M}_{L,\alpha}\,\boldsymbol{P}_{\ell,L}$ is obtained in terms of the following cores:

$$\hat{T}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \hat{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \hat{Y}_0 = \frac{1}{2}\begin{bmatrix} \binom{2}{2} \\ \binom{-1}{1} \quad \binom{1}{1} \end{bmatrix},$$

$$\hat{Y}_1 = \frac{1}{2}\left[\binom{1}{1}\right], \quad \hat{N}_1 = [1] \quad \text{and} \quad \hat{N}_0 = \frac{1}{2}\begin{bmatrix} \binom{1}{0} \\ \binom{0}{1} \end{bmatrix}. \tag{78}$$

The proof of the following lemma is rather technical and is therefore given in "Appendix B."

Lemma 5 *For all $L, \ell \in \mathbb{N}_0$ such that $\ell \leq L$, the matrices $\hat{\boldsymbol{M}}_{L,\alpha}\, \hat{\boldsymbol{P}}_{\ell,L}$ with $\alpha = 0, 1$, where the factors are given by (12d) and (14), admit the representation*

$$\hat{\boldsymbol{M}}_{L,\alpha}\, \hat{\boldsymbol{P}}_{\ell,L} = 2^{(\alpha+\frac{1}{2})\ell}\ \hat{A} \bowtie \hat{U}^{\bowtie \ell} \bowtie \hat{T}_\alpha \bowtie \hat{Y}_\alpha^{\bowtie (L-\ell)} \bowtie \hat{N}_\alpha\,, \tag{79}$$

where the cores $\hat{A}$, $\hat{U}$ and $\hat{T}_\alpha$, $\hat{Y}_\alpha$, $\hat{N}_\alpha$ with $\alpha = 0, 1$ are as in (67) and (75).

Combining decomposition (77) and its transpose, we can rewrite the product $\hat{\boldsymbol{P}}_{\ell,L}\, \hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}}$ core-wise:

$$\hat{\boldsymbol{P}}_{\ell,L}\, \hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}} = 2^{-(L-\ell)}\ \hat{A}_\flat \bowtie \hat{U}_\flat^{\bowtie \ell} \bowtie \hat{X}_\flat^{\bowtie (L-\ell)} \bowtie \hat{P}_\flat\,, \tag{80}$$

where the factors are

$$\hat{A}_\flat = \hat{A} \bullet \hat{A}\,, \quad \hat{U}_\flat = \hat{U} \bullet \hat{U}^{\mathsf{T}}, \quad \hat{X}_\flat = \hat{X} \bullet \hat{X}^{\mathsf{T}}, \quad \hat{P}_\flat = \hat{P} \bullet \hat{P}\,. \tag{81}$$

We remark that the ranks of the decomposition (80) are $4, \ldots, 4$.

Applying the same argument to the product $\hat{\boldsymbol{M}}_{L,\alpha}\, (\hat{\boldsymbol{P}}_{\ell,L}\, \hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}})$, the factors $\hat{\boldsymbol{M}}_{L,\alpha}$ and $\hat{\boldsymbol{P}}_{\ell,L}\, \hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}}$ being taken in the form of (76) and (80), we could obtain its explicit decomposition with ranks $2^3, \ldots, 2^3$. Instead, we multiply $\hat{\boldsymbol{M}}_{L,\alpha}\, \hat{\boldsymbol{P}}_{\ell,L}$ and $\hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}}$ using the representations (79) and (77) to form a representation of the same product $\hat{\boldsymbol{M}}_{L,\alpha}\, \hat{\boldsymbol{P}}_{\ell,L}\, \hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}}$. This representation has the ranks $2^2, \ldots, 2^2, 2^{2-\alpha}, \ldots, 2^{2-\alpha}$, which means that the ranks of unfolding matrices $1, \ldots, \ell - 1$ and $\ell, \ldots, L - \alpha$ are bounded by 4 and $2^{2-\alpha}$, respectively. As we discuss in Sect. 5.4, this reduction is substantial in the case of multiple dimensions, when the exponents (2 or $2 - \alpha$ instead of 3) that correspond to the dimensions are summed.

Specifically, combining (79) and (77), we arrive at

$$\hat{\boldsymbol{M}}_{L,\alpha}\, \hat{\boldsymbol{P}}_{\ell,L}\, \hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}} = 2^{(\alpha+\frac{1}{2})L-(L-\ell)}\ \hat{A}_\flat \bowtie \hat{U}_\flat^{\bowtie \ell} \bowtie \hat{W}_\alpha \bowtie \hat{Z}_\alpha^{\bowtie (L-\ell)} \bowtie \hat{K}_\alpha\,, \tag{82}$$

where

$$\hat{W}_\alpha = \hat{T}_\alpha \bullet \hat{I}\,, \quad \hat{Z}_\alpha = \hat{Y}_\alpha \bullet \hat{X}^{\mathsf{T}}, \quad \hat{K}_\alpha = \hat{N}_\alpha \bullet \hat{P} \quad \text{with} \quad \alpha = 0, 1\,. \tag{83}$$

Decomposition (82) is exact and explicit, the latter meaning that all the cores involved are provided in closed form. Since $\hat{U}_\flat$ and $\hat{Y}_\alpha$ are of ranks $2^2 \times 2^2$ and $2^{2-\alpha} \times 2^{2-\alpha}$, respectively, the ranks of decomposition (82) are $2^2, \ldots, 2^2, 2^{2-\alpha}, \ldots, 2^{2-\alpha}$.

Direct calculation with expressions given in (67)–(75) leads to $\hat{A}_\flat = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$,

$$\hat{P}_\flat = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \hat{W}_0 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \quad \text{and} \quad \hat{W}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{84}$$

Explicit expression for $\hat{U}_\flat$, $\hat{X}_\flat$ and $\hat{Z}_\alpha$, $\hat{K}_\alpha$ with $\alpha = 0, 1$ can be likewise calculated based on (67) and (75), from which we refrain to keep exposition concise.

5.4 Analysis in *D* Dimensions by Tensorization

In this section, we generalize the results of Sect. 5.3 to the case of multiple dimensions and analyze the low-rank tensor structure of the preconditioner $\boldsymbol{C}_L$, given by (30), and of the preconditioned discrete differential operator $\boldsymbol{B}_L$ in the form of (33b). For the latter, we first derive a representation of the matrices $\boldsymbol{Q}_{L,\alpha}$ with $L \in \mathbb{N}$ and $\alpha \in \{0, 1\}^D$, which are defined in (33c).

The representations derived below are composed from the following cores:

$$\begin{aligned} A_\flat = \hat{A}_\flat^{\otimes D}, \quad U_\flat = \hat{U}_\flat^{\otimes D}, \quad X_\flat = \hat{X}_\flat^{\otimes D}, \quad P_\flat = \hat{P}_\flat^{\otimes D}, \\ W_\alpha = \otimes_{k=1}^D \hat{W}_{\alpha_k}, \quad Z_\alpha = \otimes_{k=1}^D \hat{Z}_{\alpha_k}, \quad K_\alpha = \otimes_{k=1}^D \hat{K}_{\alpha_k} \end{aligned} \tag{85}$$

for all $\alpha \in \{0, 1\}^D$, where the factors are given by (81) and (83).

Tensorizing (80) core-wise and distributing the scaling factor over the cores, we obtain the decompositions

$$\begin{aligned} 2^{-\ell} \boldsymbol{\Pi}_L \boldsymbol{P}_{\ell,L} \boldsymbol{P}_{\ell,L}^\mathsf{T} \boldsymbol{\Pi}_L^\mathsf{T} &= 2^{-\ell - D(L-\ell)} A_\flat \bowtie U_\flat^{\bowtie \ell} \bowtie X_\flat^{\bowtie (L-\ell)} \bowtie P_\flat \\ &= 2^{-\ell} A_\flat \bowtie U_\flat^{\bowtie \ell} \bowtie (2^{-D} X_\flat)^{\bowtie (L-\ell)} \bowtie P_\flat \end{aligned} \tag{86}$$

of ranks $2^{2D}, \ldots, 2^{2D}$, where the cores are given by (85) and the permutation matrix $\boldsymbol{\Pi}_L$ is as defined in (49). Applying [35, Lemma 5.5] to the sum of such matrices with $\ell = 1, \ldots, L$ and adding the term corresponding to $\ell = 0$, we obtain the following result.

Theorem 3 *For any $L \in \mathbb{N}$, the matrix $\boldsymbol{C}_L$, defined by (30), admits the decomposition*

$$\boldsymbol{\Pi}_L \boldsymbol{C}_L \boldsymbol{\Pi}_L^\mathsf{T} = \begin{bmatrix} A_\flat & A_\flat \end{bmatrix} \bowtie C_1 \bowtie \cdots \bowtie C_L \bowtie \begin{bmatrix} \\ P_\flat \end{bmatrix} \tag{87}$$

of ranks $2^{2D} + 2^{2D}, \ldots, 2^{2D} + 2^{2D}$, all equal to 2^{2D+1}, where the middle cores are

$$C_\ell = \begin{bmatrix} U_\flat & 2^{-\ell} U_\flat \\ & 2^{-D} X_\flat \end{bmatrix} \quad \textit{with} \quad \ell = 1, \ldots, L,$$

the subcores being as in (85).

For any $L \in \mathbb{N}$, $\ell = 0, 1 \ldots, L$ and $\alpha \in \{0,1\}^D$, tensorizing (82) core-wise and distributing the scaling factor over the cores result in the decompositions

$$\begin{aligned} &2^{-\ell}\, \widetilde{\boldsymbol{\Pi}}_{L,\alpha}\, \boldsymbol{M}_{L,\alpha}\, \boldsymbol{P}_{\ell,L}\, \boldsymbol{P}_{\ell,L}^{\mathsf{T}} \boldsymbol{\Pi}_L^{\mathsf{T}} \\ &\quad = 2^{-\ell+(|\alpha|+\frac{1}{2}D)L - D(L-\ell)}\; A_\flat \bowtie U_\flat^{\bowtie \ell} \bowtie W_\alpha \bowtie Z_\alpha^{\bowtie (L-\ell)} \bowtie K_\alpha \\ &\quad = 2^{-(1-|\alpha|)\ell}\; A_\flat \bowtie (2^{\frac{1}{2}D}\, U_\flat)^{\bowtie \ell} \bowtie W_\alpha \bowtie (2^{|\alpha|-\frac{1}{2}D}\, Z_\alpha)^{\bowtie (L-\ell)} \bowtie K_\alpha \end{aligned} \tag{88}$$

of ranks $2^{2D}, \ldots, 2^{2D}, 2^{2D-|\alpha|}, \ldots, 2^{2D-|\alpha|}$, where $\boldsymbol{M}_{L,\alpha}$ and $\boldsymbol{P}_{\ell,L}$ are given by (20c) and (22), the cores are given by (85) and the permutation matrices $\boldsymbol{\Pi}_L$ and $\widetilde{\boldsymbol{\Pi}}_{L,\alpha}$ are as defined in (49) and (51).

Similarly as for $\boldsymbol{C}_L$ above, we can apply [35, Lemma 5.5] to the sum of the matrices given by (88) with $\ell = 1, \ldots, L$ and add the term corresponding to $\ell = 0$. This leads to the following result, which is analogous to Theorem 3.

Theorem 4 *For any $L \in \mathbb{N}$ and $\alpha \in \{0,1\}^D$, the matrix $\boldsymbol{Q}_{L,\alpha}$, given by (33c), admits the decomposition*

$$\widetilde{\boldsymbol{\Pi}}_{L,\alpha}\, \boldsymbol{Q}_{\ell,L,\alpha} \boldsymbol{\Pi}_L^{T} = \begin{bmatrix} A_\flat & A_\flat \bowtie W_\alpha \end{bmatrix} \bowtie Q_1 \bowtie \cdots \bowtie Q_L \bowtie \begin{bmatrix} \\ K_\alpha \end{bmatrix} \tag{89}$$

of ranks $2^{2D} + 2^{2D-|\alpha|}, \ldots, 2^{2D} + 2^{2D-|\alpha|}$, all bounded from above by 2^{2D+1}, where the middle cores are

$$Q_\ell = \begin{bmatrix} U_\flat & 2^{-(1-|\alpha|)\ell}\, U_\flat \bowtie W_\alpha \\ & 2^{|\alpha|-\frac{1}{2}D}\, Z_\alpha \end{bmatrix} \quad \textit{with} \quad \ell = 1, \ldots, L\,,$$

the subcores being defined by (85).

In Example 1, the case of the Laplace operator was considered and the factors $\boldsymbol{\Lambda}_{L,\alpha\alpha'}$ with $(\alpha,\alpha') \in \mathscr{D}$ for the suitable $\mathscr{D}$ were explicitly given in the Kronecker product form (27a). That form immediately leads to a multilevel TT decomposition of ranks $1, \ldots, 1$ for each $\boldsymbol{\Lambda}_{L,\alpha\alpha'}$. Here, we analyze the structure of $\boldsymbol{\Lambda}_{L,\alpha\alpha'}$ with $(\alpha,\alpha') \in \mathscr{D}$ in the general setting of Sect. 2.3, for an arbitrary $\mathscr{D} \subset \{0,1\}^D \times \{0,1\}^D$ of differential indices, under the additional assumption that the coefficient functions (24b) exhibit low-rank structure.

Specifically, for each $(\alpha,\alpha') \in \mathscr{D}$, we assume that the coefficient vector $\boldsymbol{c}_{L,\alpha\alpha'} \in \mathbb{R}^{\mathscr{I}_L \times \Gamma_{\alpha\alpha'}} \simeq \mathbb{R}^{2^{DL} R_{\alpha\alpha'}}$ parametrizing the coefficient function $c_{\alpha\alpha'}$ through (24b) is given in a multilevel TT representation of ranks $r_{0,\alpha\alpha'}, \ldots, r_{L,\alpha\alpha'}$:

$$\widetilde{\boldsymbol{\Pi}}_{L,\alpha}\, \boldsymbol{c}_{L,\alpha\alpha'} = C_{L,0,\alpha,\alpha'} \bowtie C_{L,1,\alpha,\alpha'} \bowtie \cdots \bowtie C_{L,L,\alpha,\alpha'} \bowtie C_{L,L+1,\alpha,\alpha'}\,, \tag{90a}$$

where each of $C_{L,1,\alpha,\alpha'}, \ldots, C_{L,L,\alpha,\alpha'}$ is of mode size 2^D, whereas $C_{L,0,\alpha,\alpha'}$ is of mode size 1 and $C_{L,L+1,\alpha,\alpha'}$ is of mode size $R_{\alpha\alpha'} = |\Gamma_{\alpha\alpha'}|$. Then, the corresponding

factor $\boldsymbol{\Lambda}_{L,\alpha\alpha'}$, given by (26a), can as well be represented with ranks $r_{0,\alpha\alpha'}, \dots, r_{L,\alpha\alpha'}$:

$$\widetilde{\boldsymbol{\Pi}}_{L,\alpha}\, \boldsymbol{\Lambda}_{L,\alpha,\alpha'} \widetilde{\boldsymbol{\Pi}}_{L,\alpha}^{\mathsf{T}} = \Lambda_{L,0,\alpha,\alpha'} \bowtie \Lambda_{L,1,\alpha,\alpha'} \bowtie \cdots \bowtie \Lambda_{L,L,\alpha,\alpha'} \bowtie \Lambda_{L,L+1,\alpha,\alpha'}\,, \tag{90b}$$

where the cores are defined in terms of those appearing in (90a) as follows. First, one sets $\Lambda_{L,0,\alpha,\alpha'} = C_{L,0,\alpha,\alpha'}$ and defines each core $\Lambda_{L,\ell,\alpha,\alpha'}$ with $\ell = 1, \dots, L$ by

$$(\Lambda_{L,\ell,\alpha,\alpha'})_{\gamma_{\ell-1}\, i_\ell\, i'_\ell\, \gamma_\ell} = 2^{-D}\, \delta_{i_\ell i'_\ell}\, (C_{L,\ell,\alpha,\alpha'})_{\gamma_{\ell-1}\, i_\ell\, \gamma_\ell} \tag{90c}$$

for all $\gamma_{\ell-1} = 1, \dots, r_{\ell-1,\alpha\alpha'}$, $\gamma_\ell = 1, \dots, r_{\ell,\alpha\alpha'}$ and $i_\ell, i'_\ell = 1, 2$. Then, the last core should be defined by

$$(\Lambda_{L,L+1,\alpha,\alpha'})_{\gamma_L\, \beta\, \beta'} = 2^{-D} \sum_{\gamma \in \Gamma_{\alpha\alpha'}} (C_{L,L+1,\alpha,\alpha'})_{\gamma_L\, \gamma} \int_{(-1,1)^D} \chi_{\alpha\alpha'\gamma}\, (\partial^\alpha \psi_\beta)\, (\partial^{\alpha'} \psi_{\beta'}) \tag{90d}$$

for all $\gamma_L = 1, \dots, r_{L,\alpha\alpha'}$, $\beta \in \{\alpha_1, 1\} \times \cdots \times \{\alpha_D, 1\}$ and $\beta' \in \{\alpha'_1, 1\} \times \cdots \times \{\alpha'_D, 1\}$, cf. (26a).

Using the fact that the ranks add under addition and multiply under multiplication [45], we obtain the following result.

Theorem 5 *For $\mathscr{D} \subset \{0,1\}^D \times \{0,1\}^D$ and $L \in \mathbb{N}$, consider a bilinear form of the type (24a)–(24b), where each coefficient vector $\boldsymbol{c}_{L,\alpha\alpha'}$ with $(\alpha, \alpha') \in \mathscr{D}$ admits a multilevel TT decomposition of the form (90a) with ranks $r_{0,\alpha\alpha'}, \dots, r_{L,\alpha\alpha'}$ not exceeding $r \in \mathbb{N}$. Then, the preconditioned matrix $\boldsymbol{B}_L$ of a, defined by (25a), (30) and (33a), admits a multilevel TT decomposition*

$$\boldsymbol{\Pi}_L\, \boldsymbol{B}_L\, \boldsymbol{\Pi}_L^T = B_{L,0} \bowtie B_{L,1} \bowtie \cdots \bowtie B_{L,L} \bowtie B_{L,L+1}$$

of ranks $R_0, \dots, R_L$, where

$$R_\ell = 2^{4D} \sum_{(\alpha,\alpha') \in \mathscr{D}} \left(1 + 2^{-|\alpha|}\right)^2 r_{\ell,\alpha\alpha'} \le 2^{4D+2} \sum_{(\alpha,\alpha') \in \mathscr{D}} r_{\ell,\alpha\alpha'} \le 12\, D^2\, 2^{4D}\, r \tag{91}$$

for $\ell = 0, \dots, L$.

Remark 4 (Sharper bounds in specific cases) The last inequality of (91) is given for a general case with D^2 second-order terms (no symmetry is assumed), D first-order terms and a zero-order term. However, for the Laplacian in the case $D = 2$, the first equality given in (91) results in $R_\ell = 1152$, which is a marked reduction from the bound $R_\ell \le 12288$ obtained for a general second-order bilinear form with constant coefficients.

Remark 5 (Inexact application) In computations, algorithms using products of $\boldsymbol{B}_L$ with vectors rather than explicit representations of $\boldsymbol{B}_L$ may be expected to be more efficient. Indeed, such products can be formed by adding the products of the terms in the sum (33b), and for each term the product can be computed by three multiplications. On the intermediate results obtained between these multiplications and additions, low-rank re-approximation can be performed, as explained further in the example of the discretized Laplacian in Sect. 5.5. The given bounds for TT ranks appear to be highly pessimistic for such inexact schemes.

Remark 6 The analysis in D dimensions is given here for the most generic discretization obtained by tensorization. The approach can be applied to discretizations that are not of tensor product form in order to mitigate the growth of the rank bounds with respect to D.

5.5 Numerical Illustrations

In summary, we obtain a combined tensor representation B_L with $\tau(\mathsf{B}_L) = \boldsymbol{\Pi}_L \boldsymbol{B}_L \boldsymbol{\Pi}_L^{\mathsf{T}} = \boldsymbol{\Pi}_L (\boldsymbol{C}_L \boldsymbol{A}_L \boldsymbol{C}_L) \boldsymbol{\Pi}_L^{\mathsf{T}}$. Similarly, from Theorem 3, we also have C_L with $\tau(\mathsf{C}_L) = \boldsymbol{\Pi}_L \boldsymbol{C}_L \boldsymbol{\Pi}_L^{\mathsf{T}}$. With a representation A_L of the stiffness matrix $\boldsymbol{A}_L$, such that $\tau(\mathsf{A}_L) = \boldsymbol{\Pi}_L \boldsymbol{A}_L \boldsymbol{\Pi}_L^{\mathsf{T}}$, one can alternatively consider the simple product representation $\mathsf{C}_L \bullet \mathsf{A}_L \bullet \mathsf{C}_L$, which corresponds to performing the action of the preconditioner $\boldsymbol{C}_L$ separately from that of $\boldsymbol{A}_L$.

Note that, in Sect. 4.4, we have assumed decompositions consisting of L cores. The decompositions in Theorems 3, 4 and 5 comprise $L+2$ cores, with first and last playing special roles since they can be merged with the respective adjacent cores. The cores in these extended decompositions are thus indexed by $\ell = 0, \ldots, L+1$ in what follows, so that again the bounds for $\ell = 1, \ldots, L$ are relevant.

One benefit of the combined representation B_L is the rank reduction compared to $\mathsf{C}_L \bullet \mathsf{A}_L \bullet \mathsf{C}_L$. More importantly, however, the decomposition B_L is constructed so that the representation condition numbers $\mathrm{mrcond}_\ell(\mathsf{B}_L)$, $\ell = 1, \ldots, L$, remain moderate even for large L. In contrast, the representation condition numbers of $\mathsf{C}_L \bullet \mathsf{A}_L \bullet \mathsf{C}_L$ are in general of the same order of magnitude as those of A_L – in other words, whereas the *matrix* condition number of $\boldsymbol{C}_L \boldsymbol{A}_L \boldsymbol{C}_L$ is uniformly bounded, for improving also the *representation* condition number, applying the preconditioner $\boldsymbol{C}_L$ separately is insufficient and one instead needs a carefully constructed *combined* representation B_L.

We now present numerical observations that illustrate how different the decompositions A_L, $\mathsf{C}_L \bullet \mathsf{A}_L \bullet \mathsf{C}_L$ and B_L are in terms of representation conditioning and demonstrate the improvement afforded by our findings presented in Sects. 4, 5.2 and 5.4. As in Example 1, we consider the case of the Laplacian: $\boldsymbol{A}_L = \boldsymbol{D}_L$ with $\boldsymbol{D}_L$ as in (103). Using (37), for $D = 1$ we have $\boldsymbol{A}_L = A_1 \bowtie \cdots \bowtie A_L$ with $A_1 = 4\,[\,I \;\; J^{\mathsf{T}} \;\; J \;\; I_2\,]$,

$$A_2 = \cdots = A_{L-1} = 4\begin{bmatrix} I & J^{\mathsf{T}} & J & \\ & J & & \\ & & J^{\mathsf{T}} & \\ & & & I_2 \end{bmatrix}, \quad A_L = 4\begin{bmatrix} 2I - J - J^{\mathsf{T}} \\ -J \\ -J^{\mathsf{T}} \\ -I_2 \end{bmatrix},$$

as derived in [35]; similar representations can be obtained for $D > 1$ by tensorization.

We first consider the upper bounds β_ℓ, defined in (65), for mramp_ℓ from Proposition 4. Since both $\|\boldsymbol{A}_L^{-1}\|$ and $\|\boldsymbol{B}_L^{-1}\|$ are bounded independently of L, by (66), up to fixed constants the respective β_ℓ are also upper bounds of the corresponding representation condition numbers mrcond_ℓ.

For B_L, instead of directly computing the estimates for $\mathrm{mramp}_\ell(\mathsf{B}_L)$ with $\ell = 1, \ldots, L$ given by Proposition 4, we will do this for the factors of a decomposition that is equivalent to B_L and is also based on (33b). Let us note that the equality

$$\mathsf{B}_L = \sum_{k=1}^{D} \Theta_{L,k}^{\mathsf{T}} \bullet \Theta_{L,k} \tag{92}$$

of decompositions holds in terms of the factors $\Theta_{L,k}$ with $k = 1, \ldots, D$ given as follows: for every k, we set $\Theta_{L,k} = \Lambda_{L,k}^{1/2} \bullet \mathsf{Q}_{L,\alpha}$ with $\alpha = (\delta_{k1}, \ldots, \delta_{kD})$, where $\Lambda_{L,\alpha,\alpha}^{1/2}$ is the decomposition of $\boldsymbol{\Lambda}_{L,\alpha,\alpha}^{1/2}$, which is diagonal and of Kronecker product form (27a); thus, its decomposition with ranks $1, \ldots, 1$ is obtained by element-wise application of the square root to each core. Equality (92) results in the second of the following inequalities:

$$\max_{\ell=1,\ldots,L} \mathrm{mrcond}_\ell(\mathsf{B}_L) \lesssim \max_{\ell=1,\ldots,L} \mathrm{mramp}_\ell(\mathsf{B}_L) \lesssim \max_{\ell=1,\ldots,L} [\beta_\ell(\Theta_{L,1})]^2, \tag{93}$$

where the equivalence is uniform with respect to $L \in \mathbb{N}$ and, for each $L \in \mathbb{N}$, β_ℓ with $\ell = 1, \ldots, L$ are as defined in (65). As well as in (93), the alternate form (92) of B_L is used to improve the efficiency of residual approximation in the numerical tests of Sect. 7.

Figure 1a shows the computed values of $\max_\ell \beta_\ell(\Theta_{L,1})$ for different values of L and $D = 1, 2$, where we observe $\max_\ell \beta_\ell(\Theta_{L,1}) = \mathscr{O}(L)$ in both cases, corresponding to

$$\max_{\ell=1,\ldots,L} \mathrm{mrcond}_\ell(\mathsf{B}_L) \lesssim \max_{\ell=1,\ldots,L} \mathrm{mramp}_\ell(\mathsf{B}_L) \lesssim \max_{\ell=1,\ldots,L} \beta_\ell(\mathsf{B}_L) \lesssim L^2.$$

In contrast, as shown in Fig. 1b, both $\max_\ell \beta_\ell(\mathsf{A}_L)$ and $\max_\ell \beta_\ell(\mathsf{C}_L \bullet \mathsf{A}_L \bullet \mathsf{C}_L)$ increase exponentially with respect to L.

Although Proposition 5 shows that they can lead to useful qualitative statements, the upper bounds provided by β_ℓ cannot be expected to be quantitatively sharp. The direct evaluation of the suprema in the definitions (64) is in general infeasible, but testing with concrete $\mathsf{V} \in \mathrm{TT}_L$ can provide some further insight. For $D = 1$, we use

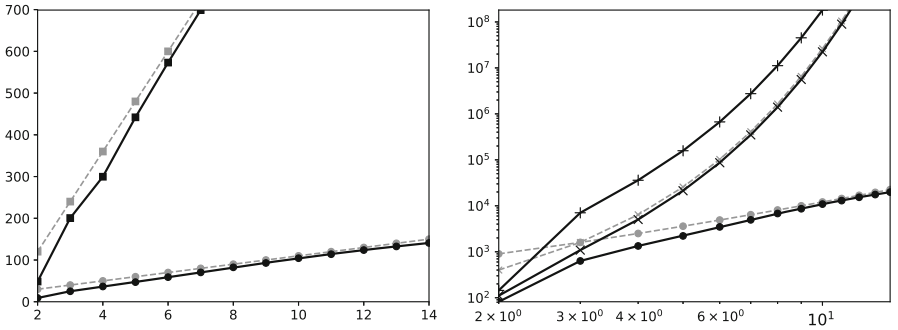


Fig. 1 Upper bounds $\max_{\ell=1,\dots,L} \beta_\ell$ as in (65) for $\max_{\ell=1,\dots,L} \mathrm{mramp}_\ell$ from Proposition 4, in dependence of L: **a** $\max_\ell \beta_\ell(\Theta_{L,1})$ for $D = 1$ (circles) and $D = 2$ (squares), with dashed lines representing $10(L+1)$ and $120(L-1)$, respectively; **b** $\max_\ell[\beta_\ell(\Theta_{L,1})]^2$ (circles), $\max_\ell \beta_\ell(\mathsf{A}_L)$ (crosses), and $\max_\ell \beta_\ell(\mathsf{C}_L \bullet \mathsf{A}_L \bullet \mathsf{C}_L)$ (plusses), for $D = 1$, with dashed lines representing $(11L)^2$ and 25×2^{2L}, respectively. The quantities $\max_\ell[\beta_\ell(\Theta_{L,1})]^2$ bound $\max_\ell[\beta_\ell(\mathsf{B}_L)]$ up to a constant independent of L, see (93)

TT-SVD representations $\mathsf{V}_1, \mathsf{V}_{\min}, \mathsf{V}_{\max}$ (of maximum ranks 1, 2, and 2, respectively) of the vectors

$$\boldsymbol{v}_1 = \big(c_1\big)_{k=1,\dots,2^L}, \quad \boldsymbol{v}_{\min} = \big(c_{\min} \sin(\tfrac{\pi}{2} x_i)\big)_{i=1,\dots,2^L},$$
$$\boldsymbol{v}_{\max} = \big(c_{\max} \sin(\tfrac{\pi}{2}(1 + 2^{L+1}) x_i)\big)_{i=1,\dots,2^L},$$

with $x_i = 2^{-L} i$ and with constants c_1, $c_{\min}$, $c_{\max}$ chosen so that $\|\boldsymbol{v}_1\|_2 = \|\boldsymbol{v}_{\min}\|_2 = \|\boldsymbol{v}_{\max}\|_2 = 1$. By Proposition 2(iii), $\mathrm{rcond}_\ell(\mathsf{V}_1) = 1$ and $1 \le \mathrm{rcond}_\ell(\mathsf{V}_{\min}) \le \sqrt{2}$, $1 \le \mathrm{rcond}_\ell(\mathsf{V}_{\max}) \le \sqrt{2}$. Consequently, as in the examples of Sect. 4.1, for each such a choice of V and any representation of a matrix M, the absolute and relative errors incurred by the orthogonalization of $\mathsf{M} \bullet \mathsf{V}$ give an indication of the order of magnitude of $\mathrm{ramp}_\ell(\mathsf{M} \bullet \mathsf{V})$ and $\mathrm{rcond}_\ell(\mathsf{M} \bullet \mathsf{V})$.

The results are summarized in Tables 3 and 4. We see that in all cases, the absolute and relative errors for B_L are close to machine precision $\varepsilon \approx 2.2 \times 10^{-16}$, which is quantitatively better than indicated by the upper bounds in Fig. 1. For A_L and $\mathsf{C}_L \bullet \mathsf{A}_L \bullet \mathsf{C}_L$, we observe an amplification of relative errors that is exponential in L (and in fact slightly worse for $\mathsf{C}_L \bullet \mathsf{A}_L \bullet \mathsf{C}_L$). The absolute errors for C_L are close to ε, which is important for the evaluation of preconditioned right-hand sides; the corresponding relative errors increase with L in the case of $\mathsf{V}_{\max}$, which is to be expected since $\boldsymbol{C}_L$ damps high-frequency oscillations.

6 Complexity of Solvers

We now consider the numerical computation of $\boldsymbol{u}_L$ solving $\boldsymbol{B}_L \boldsymbol{u}_L = \boldsymbol{f}_L$ with $\boldsymbol{B}_L = \boldsymbol{C}_L \boldsymbol{A}_L \boldsymbol{C}_L$ and $\boldsymbol{g}_L = \boldsymbol{C}_L \boldsymbol{f}_L$ as in (33a). Here, the objective is to find $u_\varepsilon \in V_{L(\varepsilon)}$ such that $\|u - u_\varepsilon\|_{\mathrm{H}^1} \lesssim \varepsilon$, and we obtain an estimate for the computational complexity of

Table 3 Absolute errors $\|\tau(\mathsf{M}\bullet\mathsf{V})-\tau(\mathrm{orth}^{-}(\mathsf{M}\bullet\mathsf{V}))\|_2$ with $\mathsf{M}=\mathsf{B}_L,\ \mathsf{C}_L,\ \mathsf{C}_L\bullet\mathsf{A}_L\bullet\mathsf{C}_L,\ \mathsf{A}_L$ and $\mathsf{V}=\mathsf{V}_1,\mathsf{V}_{\min},\mathsf{V}_{\max}$, as given in Sect. 5.5

V	M	$L=20$	$L=30$	$L=40$
V_1	B_L	1.47×10^{-14}	2.08×10^{-14}	3.30×10^{-14}
	C_L	1.16×10^{-15}	2.05×10^{-15}	5.70×10^{-15}
	$\mathsf{C}_L\bullet\mathsf{A}_L\bullet\mathsf{C}_L$	3.06×10^{-04}	$2.65\times10^{+02}$	$3.27\times10^{+08}$
	A_L	2.66×10^{-04}	$2.08\times10^{+02}$	$2.13\times10^{+08}$
$\mathsf{V}_{\min}$	B_L	1.89×10^{-14}	3.78×10^{-14}	2.96×10^{-14}
	C_L	2.69×10^{-15}	1.70×10^{-15}	2.20×10^{-15}
	$\mathsf{C}_L\bullet\mathsf{A}_L\bullet\mathsf{C}_L$	4.58×10^{-04}	$3.60\times10^{+02}$	$5.23\times10^{+08}$
	A_L	4.99×10^{-04}	$5.96\times10^{+02}$	$4.27\times10^{+08}$
$\mathsf{V}_{\max}$	B_L	1.31×10^{-14}	1.20×10^{-14}	9.29×10^{-15}
	C_L	9.82×10^{-17}	1.20×10^{-16}	1.07×10^{-16}
	$\mathsf{C}_L\bullet\mathsf{A}_L\bullet\mathsf{C}_L$	1.08×10^{-04}	$1.80\times10^{+02}$	$1.26\times10^{+08}$
	A_L	6.62×10^{-03}	$1.43\times10^{+04}$	$1.12\times10^{+10}$

Table 4 Relative errors $\|\tau(\mathsf{M}\bullet\mathsf{V})-\tau(\mathrm{orth}^{-}(\mathsf{M}\bullet\mathsf{V}))\|_2/\|\tau(\mathsf{M}\bullet\mathsf{V})\|_2$ with M and V as in Table 3

V	M	$L=20$	$L=30$	$L=40$
V_1	B_L	2.87×10^{-15}	4.06×10^{-15}	6.44×10^{-15}
	C_L	1.11×10^{-15}	1.95×10^{-15}	5.44×10^{-15}
	$\mathsf{C}_L\bullet\mathsf{A}_L\bullet\mathsf{C}_L$	5.97×10^{-05}	$1.89\times10^{+00}$	$2.18\times10^{+00}$
	A_L	2.48×10^{-13}	5.92×10^{-12}	1.85×10^{-10}
$\mathsf{V}_{\min}$	B_L	4.17×10^{-15}	8.32×10^{-15}	6.52×10^{-15}
	C_L	2.40×10^{-15}	1.52×10^{-15}	1.97×10^{-15}
	$\mathsf{C}_L\bullet\mathsf{A}_L\bullet\mathsf{C}_L$	1.01×10^{-04}	$1.50\times10^{+00}$	$5.41\times10^{+00}$
	A_L	2.02×10^{-04}	7.22×10^{-01}	6.59×10^{-01}
$\mathsf{V}_{\max}$	B_L	3.28×10^{-15}	3.00×10^{-15}	2.32×10^{-15}
	C_L	6.91×10^{-11}	8.61×10^{-08}	7.91×10^{-05}
	$\mathsf{C}_L\bullet\mathsf{A}_L\bullet\mathsf{C}_L$	2.70×10^{-05}	$6.32\times10^{+00}$	$2.78\times10^{+00}$
	A_L	1.51×10^{-15}	3.10×10^{-15}	2.31×10^{-15}

achieving this goal. Assuming that $L(\varepsilon)\sim|\log\varepsilon|$ is suitably chosen a priori and that the TT singular values of $\boldsymbol{u}_L$ satisfy a natural decay estimate, we show that the number of arithmetic operations for computing a tensor train representation of u_ε is of order $\mathcal{O}(|\log\varepsilon|^{\theta})$, where $\theta>0$ depends only on the low-rank approximability of the $\boldsymbol{u}_L$.

Remark 7 The methods we consider rely on the accurate evaluation of residuals $\boldsymbol{B}_L\boldsymbol{v}-\boldsymbol{C}_L\boldsymbol{f}_L$. As we have seen in Sect. 5.5, for the representations B_L and C_L of $\boldsymbol{B}_L$ and $\boldsymbol{C}_L$ that we have constructed, the quantities $\mathrm{mramp}_\ell(\mathsf{B}_L)$ and $\mathrm{mramp}_\ell(\mathsf{C}_L)$ grow only moderately with respect to L. Indeed, the results of Table 3 indicate that provided that $\boldsymbol{v}$

and $\boldsymbol{f}_L$ are given in well-conditioned representations, the corresponding residuals can be evaluated with an absolute error close to machine precision, which is corroborated also by our further numerical tests in Sect. 7. For the convergence analysis of this section, we assume exact arithmetic.

6.1 Estimates of Ranks and Computational Costs

To estimate the computational complexity of finding approximate solutions, we use the quasi-optimality properties of an iterative method using soft thresholding of hierarchical tensors introduced in [6]. This construction directly carries over to the special case of the TT format, leading to a soft thresholding operation $\mathscr{S}_\alpha$ that is non-expansive with respect to the ℓ^2-norm. It can be realized numerically for TT representations, described in [6, Sec. 3], at essentially the same cost as the TT-SVD.

Note that since $\boldsymbol{B}_L$ is well-conditioned uniformly with respect to L, as a consequence of Theorem 1 we can choose $\omega > 0$ such that $\xi = \sup_{L>0} \|I - \omega \boldsymbol{B}_L\|$ satisfies $\xi < 1$. The basic iterative method applied to the present problem has the form

$$\boldsymbol{u}_L^{n+1} = \mathscr{S}_{\alpha_n}\big(\boldsymbol{u}_L^n - \omega(\boldsymbol{B}_L \boldsymbol{u}_L^n - \boldsymbol{g}_L)\big), \quad n \geq 0, \tag{94}$$

with $\boldsymbol{u}_L^0 = 0$ and $\alpha_n \to 0$ determined (according to [6, Alg. 2]) as follows: set $\alpha_0 = \omega \|\boldsymbol{g}_L\|_2/(d-1)$, and for a fixed $\bar{B} > \|\boldsymbol{B}_L\|_{2\to 2}$, take

$$\alpha_{n+1} = \begin{cases} \frac{1}{2}\alpha_n, & \text{if } \|\boldsymbol{u}_L^{n+1} - \boldsymbol{u}_L^n\|_2 \leq \frac{1-\xi}{\xi \bar{B}} \|\boldsymbol{B}_L \boldsymbol{u}_L^{n+1} - \boldsymbol{g}_L\|_2, \\ \alpha_n, & \text{else.} \end{cases} \tag{95}$$

In what follows, we refer to the algorithm given by (94), (95) as STSOLVE.

Recall that $u_L = \sum_{j \in \mathscr{J}_L} (\boldsymbol{C}_L \boldsymbol{u}_L)_j \varphi_{L,j}$, with analogous notation for the iterates, where $\|u_L\|_V \sim \|\boldsymbol{u}_L\|_2$. Our convergence analysis is based on the following assumption on uniform decay of singular values, which is discussed further in Sect. 6.2.

Assumption 1 For all $L \in \mathbb{N}$ and $\ell = 1, \dots, L$, let the singular values $\sigma_{\ell,j}(\boldsymbol{u}_L)$ with $j = 1, \dots, 2^{D \max(\ell, L-\ell)}$ of the ℓth unfolding matrix $\mathrm{U}_\ell(\boldsymbol{u}_L)$, defined as in (44a), satisfy the bound

$$\sigma_{\ell,j}(\boldsymbol{u}_L) \leq C e^{-cj^\beta} \quad \text{for all} \quad j = 1, \dots, 2^{D \max(\ell, L-\ell)} \tag{96}$$

with $C, c, \beta > 0$ independent of ℓ and L.

Theorem 6 *Let $\varepsilon > 0$. Then,* STSOLVE *stops with $\boldsymbol{u}_{L,\varepsilon}$ such that*

$$\|u_L - u_{L,\varepsilon}\|_{\mathrm{H}^1} \lesssim \|\boldsymbol{u}_L - \boldsymbol{u}_{L,\varepsilon}\|_2 \leq \varepsilon$$

after finitely many steps. In addition, let Assumption 1 hold. Then, there exist $c_1, c_2 > 0$ and $\rho \in (0,1)$ independent of L and n such that with $\varepsilon_n = \rho^{n/\log L}$,

$$\|u_L - u_L^n\|_{\mathrm{H}^1} \leq c_1 L \varepsilon_n, \quad \max_{\ell=1,\dots,L-1} \mathrm{rank}_\ell(\boldsymbol{u}_L^n) \leq c_2 L^2 \big(1 + |\log \varepsilon_n|\big)^{\frac{1}{\beta}}.$$

Proof This is the statement of [6, Thm. 5.1(ii)] applied to our setting, combined with [6, Rem. 5.6] concerning the dependence of ε_n on L. □

The above statement makes assumptions on the low-rank approximability of the approximations u_L. We next relate this, by an appropriate choice of L, to the approximability of the exact solution $u \in V$ of (4).

Corollary 1 *Assume that there exist $C_1 > 0$ and $s > 0$ such that $\|u - u_L\|_{\mathrm{H}^1} \le C_1 2^{-sL}$. Then, for given $\varepsilon \in (0,1)$, taking $L = \frac{1}{s}(1+|\log\varepsilon|)$, with $c_1, c_2 > 0$ and $\varepsilon_n = \rho^{n/\log L}$ as in Theorem 6, for $n > 0$ we have*

$$\|u_L - u_L^n\|_{\mathrm{H}^1} \le c_1 s^{-1}(1+|\log\varepsilon|)\varepsilon_n,$$
$$\max_{\ell=1,\dots,L-1} \operatorname{rank}_\ell(\boldsymbol{u}_L^n) \le c_2 s^{-2}(1+|\log\varepsilon|)^2\big(1+|\log\varepsilon_n|\big)^{\frac{1}{\beta}},$$

and for $N = (|\log\varepsilon| + \log L)\log L \lesssim (1+|\log\varepsilon|)\log(1+|\log\varepsilon|)$, we obtain

$$\|u - u_L^N\|_{\mathrm{H}^1} \le C_2\varepsilon, \qquad \max_{\ell=1,\dots,L-1} \operatorname{rank}_\ell(\boldsymbol{u}_L^N) \le C_3(1+|\log\varepsilon|)^{2+\frac{1}{\beta}},$$

where $C_2, C_3 > 0$ depend on c_1, c_2, ρ, C_1, and s.

Remark 8 (Complexity bounds) If $\boldsymbol{B}_L$ has fixed representation ranks, as in the case of the Laplacian, the costs of each step are dominated by those of applying $\mathscr{S}_{\alpha_n}$, which are of order $\mathscr{O}(L(\max_\ell \operatorname{rank}_\ell(\boldsymbol{u}_L^n))^3)$. By Corollary 1, the total number of operations for N steps to guarantee an H^1-error of order ε is thus bounded by

$$C(1+|\log\varepsilon|)^{8+\frac{3}{\beta}}\log(1+|\log\varepsilon|) \tag{97}$$

with a uniform constant $C > 0$.

In cases with variable coefficients such that $\boldsymbol{B}_L$ does not have an exact low-rank form, but needs to be applied approximately, the iteration given in (94) and (95) can be adapted to residual approximations with prescribed tolerance as given in [6, Alg. 3], which preserves the statement of Theorem 6 as shown in [6, Prop. 5.9]. Depending on the L- and ε-dependent rank bounds for $\boldsymbol{B}_L$, one may then obtain additional factors in estimate (97).

Remark 9 Complexity estimates are also given in [4] for a similar iterative method based on hierarchical SVD truncation (which in the present setting translates to a direct TT-SVD truncation). A simplified version of this method operating on fixed discretizations is given in [6, Alg. 4]. Based on the theory for this method, one can also derive rank and complexity bounds similar to (97), but with a less favorable exponent: for this method, one arrives at a number of operations bounded by $C(1+|\log\varepsilon|)^{t+\frac{3}{\beta}}$ for some $C > 0$, where $t > 0$ now depends on the representation ranks and condition number of $\boldsymbol{B}_L$, and the bound can be substantially worse than (97). The practical performance of the scheme from [4], however, tends to be comparable to the one of STSOLVE considered above.

Remark 10 Alternatively, the linear systems $\boldsymbol{B}_L \boldsymbol{u}_L = \boldsymbol{g}_L$ can be solved by the AMEn methods introduced in [17]. The basic version analyzed in [17, Sec. 5] relies on residual approximations of a certain quality and increases approximation ranks in each iteration. However, the available convergence results only lead to a complexity bound that increases faster than exponentially in L. In the practical implementation that we also consider for comparison in Sect. 7, the basic method is combined with a faster heuristic residual approximation scheme based on the alternating least squares (ALS) method and with additional rank reduction steps. Although no convergence analysis is available for this version, the method performs well in our tests with well-conditioned $\boldsymbol{B}_L$.

6.2 Low-Rank Approximability Assumptions

For the case of one or two dimensions, a low-rank approximation analysis for the solution of problem (4) under certain analyticity assumptions on the coefficients and right-hand side, following from the regularity analysis developed in [2,3], is available in [28,33,34]. The following result can be obtained as an immediate consequence of [34, Theorem 5.16].

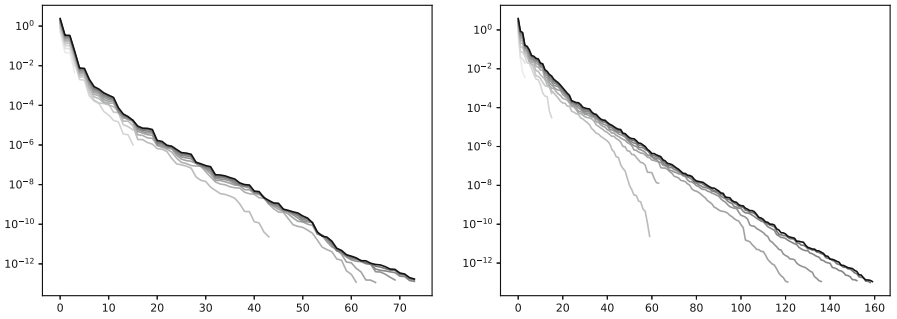
Theorem 7 *Consider problem* (4) *with* $D = 2$ *dimensions under the ellipticity and regularity assumptions made in Sect. 2. Assume additionally that the data (the diffusion coefficient and the right-hand side) are analytic on* $\overline{\Omega}$*. Then, the following holds with positive constants* C, C', b, b'*. For all* $L, R \in \mathbb{N}$*, the exact solution* u *admits an approximation* $u_{L,R} \in V_L$ *that can be exactly represented in the multilevel TT decomposition in the sense of* (50a)–(50b)*, with ranks not exceeding* R *and such that*

$$\|u - u_{L,R}\|_{\mathrm{H}^1(\Omega)} \leq C e^{-bL} + C' e^{-b'\sqrt{R}} . \tag{98}$$

Theorem 7 and analogous results for highly oscillatory solutions [31] cover the tensor approximation of exact solutions in the nodal basis, described in Sect. 2.2. The requirements of Assumption 1 are somewhat different: they refer to the solution of the Galerkin discretization (uniformly in the discretization level L), and the application of $\boldsymbol{C}_L^{-1}$ to the corresponding coefficient $\bar{\boldsymbol{u}}_L$ (which is with respect to the nodal basis) yields the coefficient $\boldsymbol{u}_L = \boldsymbol{C}_L^{-1} \bar{\boldsymbol{u}}_L$ with respect to the preconditioned basis. Nevertheless, the H^1-errors bounded implicitly by the decay of singular values in Assumption 1 and explicitly by the second term in the right-hand side of (98) both correspond to low-rank tensor approximation within the underlying finite element space V_L.

The verification of the low-rank approximability of $\boldsymbol{u}_L$, $L \in \mathbb{N}$, stipulated in Assumption 1 requires the result of Theorem 7 to be complemented by two further ingredients: bounds on the ranks of Galerkin discretizations (as opposed to interpolants of the exact solution); and suitable low-rank approximations of $\boldsymbol{C}_L^{-1}$, (which, unlike $\boldsymbol{C}_L$, does not have an explicit low-rank form).

In the present work, we restrict ourselves to studying the resulting approximability of $\boldsymbol{u}_L$ numerically. We are not aware of existing analysis that would allow to arrive at conclusions on Galerkin solution ranks, covering also the convergence behavior for accuracies below the size of the Galerkin discretization error; this appears to be a



(a) $\max_{\ell=1,\dots,L-1} \sigma_{\ell,j}(\boldsymbol{C}_L \boldsymbol{u}_L)$ vs. j **(b)** $\max_{\ell=1,\dots,L-1} \sigma_{\ell,j}(\boldsymbol{u}_L)$ vs. j

Fig. 2 Singular values of unfolding matrices (see Assumption 1) for u solving $-\Delta u = 1$ on $(0,1)^2$ with boundary conditions according to (3), for $L = 2, \dots, 12$

question of independent interest. In certain special cases, such as Poisson problems in $D = 1$, the Galerkin solution can in fact be shown to be the nodal interpolant of the exact solution. For more general problems and for $D > 1$, however, this is in general not the case.

The numerically observed decay of matricization singular values of the preconditioned solution coefficients $\boldsymbol{u}_L$ (with $\|\boldsymbol{u}_L\|_2 \sim \|u_L\|_{\mathrm{H}^1}$) and of the vector of scaled nodal values $\bar{\boldsymbol{u}}_L = \boldsymbol{C}_L \boldsymbol{u}_L$ (with $\|\bar{\boldsymbol{u}}_L\|_2 \sim \|u_L\|_{\mathrm{L}^2}$) for a Poisson problem in spatial dimension $D = 2$ is illustrated in Fig. 2. We find that the action of $\boldsymbol{C}_L^{-1}$ on the vector of nodal values preserves the exponential decay of singular values, but at a slightly modified rate. This is consistent with the further numerical tests for this problem in Sect. 7.4. Similar results are also observed in further experiments presented in Sect. 7.

7 Numerical Experiments

In our numerical tests, we apply the preconditioned discretization matrices in well-conditioned tensor representations obtained in Sect. 5 to different problems of type (4), both with constant and with highly oscillatory diffusion coefficients A in (5).

For solving the resulting systems of equations, on the one hand we use STSOLVE analyzed in Sect. 6, implemented in the Julia programming language; on the other hand, we compare to results obtained using a Fortran implementation of the AMEN solver [17] wrapped by the Python version of the TT Toolbox by I. Oseledets.

These two solvers have quite distinct characteristics. The parameters for STSOLVE are chosen such that the convergence and complexity estimates of Theorem 6 are guaranteed, which leads to a very conservative control of the iteration. Since residuals are approximated with guaranteed accuracy, this method yields rigorous error bounds. In contrast, the considered version of AMEN uses several heuristic extensions, as described in [17, Sec. 6]. In particular, it uses a simplified ALS-type residual approximation that has strongly reduced complexity, but does not give any error guarantees.

Moreover, in the given results, iteration numbers for AMEn need to be interpreted differently, where each iteration in the convergence plots comprises several substeps with local residual evaluations for each core.

7.1 Results Without Preconditioning

We first illustrate the results obtained by a direct application of multilevel tensor representations of stiffness matrices $\boldsymbol{A}_L$ without preconditioning. Such representations have been derived, for instance, in [35]. In the present case of mixed Dirichlet and Neumann boundary conditions, this leads to representations similar to the pure Dirichlet case in (54). Here, we consider the case $D = 1$, where for simplicity we take reaction coefficient $c = 0$ and right-hand side $f = 1$, that is, we solve the weak formulation of

$$-u'' = 1, \quad u(0) = 0, \ u'(1) = 0. \tag{99}$$

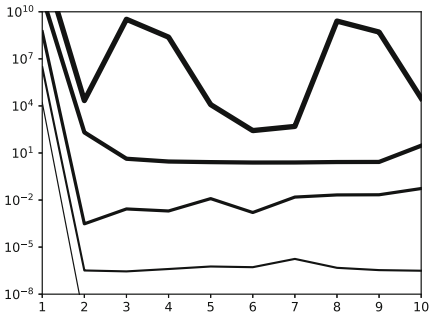
Using AMEn directly with system matrix $\boldsymbol{A}_L$ and right-hand side $\boldsymbol{f}_L$, we observe that the resulting residual indicators stagnate at values above $2^{2L}\varepsilon$, where $\varepsilon \approx 2.2 \times 10^{-16}$ is the relative machine precision. This is to be expected in view of the matrix and representation ill-conditioning of $\boldsymbol{A}_L$.

If we instead implement the preconditioned matrix $\boldsymbol{C}_L\boldsymbol{A}_L\boldsymbol{C}_L$ by pre- and post-multiplying with a separate tensor representation $\boldsymbol{C}_L$ of the preconditioner, we still obtain essentially the same type of stagnation at approximately $2^{2L}\varepsilon$. Since the represented matrix $\boldsymbol{C}_L\boldsymbol{A}_L\boldsymbol{C}_L$ is now well-conditioned, these remaining catastrophic round-off errors and the resulting stagnation are entirely due to *representation* ill-conditioning, which is not removed by simply multiplying by the preconditioner. This effect is observed both with AMEn and with STSolve. The results are shown in Fig. 3, with the residual values with respect to the system matrices $\boldsymbol{A}_L$ and $\boldsymbol{C}_L\boldsymbol{A}_L\boldsymbol{C}_L$, respectively.

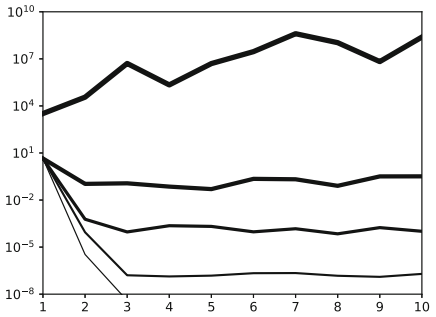
7.2 Constant Coefficient Diffusion, $D = 1$

We now consider the same basic test case (99), but with $\boldsymbol{B}_L = \boldsymbol{C}_L\boldsymbol{A}_L\boldsymbol{C}_L$ in the combined tensor representation constructed in Sect. 5. In this and the following tests, residual values always refer to the preconditioned residuals $\|\boldsymbol{B}_L \cdot -\boldsymbol{g}_L\|_2$, which is proportional to the H^1-errors in the corresponding grid functions. With a target residual of 10^{-12}, both AMEn and STSolve converge unaffected by any round-off errors for very large values of L (Fig. 4). Indeed, this remains true for values L that are substantially larger than in the case $L = 50$ shown here, but since the corresponding mesh widths are then smaller than machine precision, the results are more difficult to interpret.

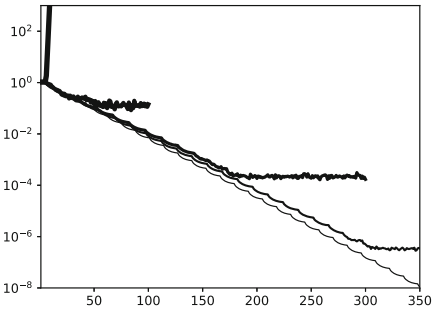
For the AMEn solver, we assemble the complete representation of $\boldsymbol{B}_L$. In exact arithmetic, this would in fact be equivalent to applying representations A_L and C_L separately, and differences are entirely due to the different tensor decomposition in the previous case. With STSolve, we have the additional option of using error-controlled



(a) AMEN with $\boldsymbol{A_L}$



(b) AMEN with $\boldsymbol{C_L A_L C_L}$



(c) STSOLVE with $\boldsymbol{C_L A_L C_L}$

Fig. 3 Results for Sect. 7.1, computed residual bounds in dependence on iteration count: **a** AMEN applied directly to $\boldsymbol{A}_L$, **b** AMEN with directly multiplied $\boldsymbol{C}_L\boldsymbol{A}_L\boldsymbol{C}_L$, (c) STSOLVE with directly multiplied $\boldsymbol{C}_L\boldsymbol{A}_L\boldsymbol{C}_L$; each for $L = 10, 15, 20, 25, 30$ (by increasing line thickness)

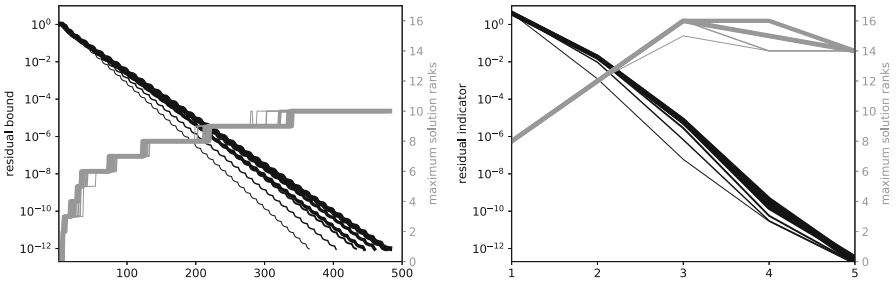
inexact residual evaluations as in [6, Alg. 3] to reduce the arising ranks of intermediate results; as shown in [6, Prop. 5.9], the statement of Theorem 6 still applies to this modification. To this end, we use that the tensor representation can be directly rewritten in the form $\boldsymbol{B}_L = \boldsymbol{\Theta}_{L,1}^{\mathsf{T}} \boldsymbol{\Theta}_{L,1}$ as in (92), where $\|\boldsymbol{\Theta}_{L,1}\|$ is uniformly bounded with respect to L, and apply an additional recompression by TT-SVD after applying $\boldsymbol{\Theta}_{L,1}$.

7.3 Highly Oscillatory Diffusion Coefficients, $D = 1$

We next consider the family of problems with oscillatory diffusion coefficients on $\Omega = (0, 1)$ given by

$$-(a_K u')' = 1, \quad u(0) = 0,\ u'(1) = 0, \quad a_K(x) = \big(2 + \cos(K\pi x)\big)^{-1} \tag{100}$$

for large values of K. The exact solution reads



(a) STSOLVE **(b)** AMEN

Fig. 4 Results for Sect. 7.2: residual bounds (black) and maximum approximation ranks (gray), with well-conditioned combined representation of $\boldsymbol{B}_L = \boldsymbol{C}_L \boldsymbol{A}_L \boldsymbol{C}_L$ for $L = 10, 15, 20, 25, 30, 35, 40, 45, 50$ (by increasing line thickness)

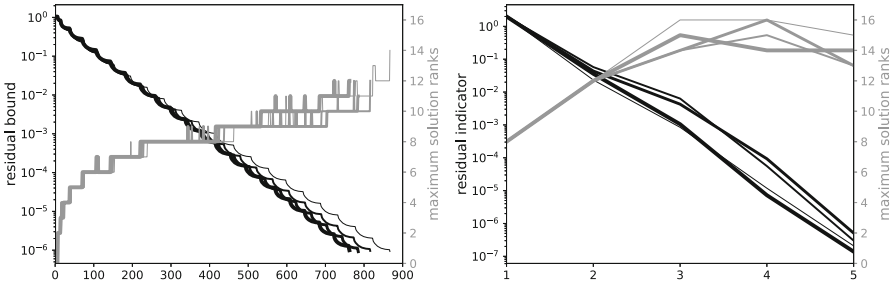
$$u(x) = x(2-x) + (K\pi)^{-1}\big[(1-x)\sin(K\pi x) + (K\pi)^{-1}\big(1-\cos(K\pi x)\big)\big]. \tag{101}$$

For $K \in 4\mathbb{N}$, we represent the vectors $\boldsymbol{u}_{\mathrm{ex}}$ and $\boldsymbol{v}_{\mathrm{ex}}$ of nodal values of u and u' in the multiscale TT format with ranks bounded by seven and six, respectively.

The coefficient a_K does not have an explicit low-rank form, and we compute approximations as follows: using the explicit rank-three representation of $c(x) = 2 + \cos(K\pi x)$, using STSOLVE we solve the equation $c(x_i)\, a_K(x_i) = 1$ in the points $x_i = 2^{-L}(i - \frac{1}{2})$, $i = 1, \ldots, 2^L$, as an elliptic problem on $\ell^2(\{1, \ldots, 2^L\})$ for a_K; the tolerance is chosen to ensure a sufficient uniform error bound.

We compare the results for the values $K = 2^{10}, 2^{20}, 2^{30}, 2^{40}$ with $L = 50$ in Fig. 5. The observed convergence patterns of both methods show hardly any influence of the value of K. Note that the computed preconditioned coefficients $\boldsymbol{u}_L$ do not satisfy the same rank bound as (101) (which holds for $\boldsymbol{C}_L \boldsymbol{u}_L$, the corresponding vector of scaled nodal values). In each case, comparison with the explicit low-rank form of $\boldsymbol{u}_{\mathrm{ex}}$, $\boldsymbol{v}_{\mathrm{ex}}$ shows that the expected total error bounds are achieved.

More specifically, approximations of the H^1-error in the solutions can be obtained in a numerically stable way by evaluating $\|\boldsymbol{u}_{\mathrm{ex}} - \boldsymbol{C}_L \boldsymbol{u}_L\|_2$ and $\|\boldsymbol{v}_{\mathrm{ex}} - \boldsymbol{\Theta}_{L,1} \boldsymbol{u}_L\|_2$, where $\boldsymbol{\Theta}_{L,1}$ is the factor of the preconditioned Laplacian stiffness matrix as in Sect. 7.2. In Table 5, we summarize the obtained approximations of H^1-errors for different solver tolerances and parameters L. We observe an effect that is particular to the present one-dimensional setting, where the accuracy in the nodal values is limited only by the solver tolerance as soon as L is sufficiently large for resolving the oscillations in the solution.



(a) STSOLVE **(b)** AMEN

Fig. 5 Results for Sect. 7.3: residual bounds (black) and maximum approximation ranks (gray), with well-conditioned representation of $\boldsymbol{B}_L$ for oscillatory coefficient a_K with $K = 2^{10}, 2^{20}, 2^{30}, 2^{40}$ (by increasing line thickness) and $L = 50$

Table 5 H^1-errors in approximations computed by AMEN with $K = 2^{30}$, solver tolerances 10^{-4}, 10^{-6}, 10^{-8} and discretization parameters L

Tol.	$L = 10$	$L = 20$	$L = 30$	$L = 40$
10^{-4}	3.65×10^{-01}	3.65×10^{-01}	3.21×10^{-05}	3.45×10^{-05}
10^{-6}	3.65×10^{-01}	3.65×10^{-01}	2.89×10^{-07}	2.88×10^{-07}
10^{-8}	3.65×10^{-01}	3.65×10^{-01}	3.73×10^{-08}	2.71×10^{-08}

7.4 Constant Coefficient Diffusion, $D = 2$

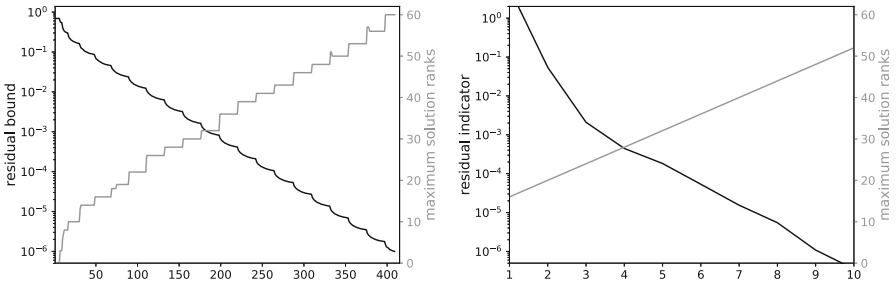
On $\Omega = (0, 1)^2$, we consider (4) with $A = 1$, $c = 0$ and $f = 1$, that is, the weak form of

$$-\Delta u = 1, \quad u|_{\Gamma} = 0, \quad \partial_n u|_{\partial\Omega\setminus\Gamma} = 0, \tag{102}$$

with Γ as in (3). Both STSOLVE and AMEN show the expected convergence for $L = 50$ (Fig. 6), with ranks that are consistent with the singular value decay of discretized solutions of Fig. 2b.

Similarly to Sect. 7.2, STSOLVE is used with inexact residual evaluation, now using that the tensor representation of $\boldsymbol{B}_L$ can be written in the form $\boldsymbol{B}_L = \boldsymbol{\Theta}_{L,1}^{\mathsf{T}} \boldsymbol{\Theta}_{L,1} + \boldsymbol{\Theta}_{L,2}^{\mathsf{T}} \boldsymbol{\Theta}_{L,2}$ as in (92). Here, $\boldsymbol{\Theta}_{L,1}$ and $\boldsymbol{\Theta}_{L,2}$ are uniformly bounded, and each has maximum representation rank 24. Although these ranks remain independent of L, additional rank reductions in this decomposition are important from a quantitative point of view: since $\boldsymbol{B}_L$ has maximum representation rank 1152, applying it directly would lead to very large ranks. In the available version of AMEN, the decomposition of $\boldsymbol{B}_L$ needs to be used directly, but the impact of large residual ranks is limited due to the ALS-type residual approximation. In this case, the main downside of the direct assembly of $\boldsymbol{B}_L$ is in the higher memory requirements for large L.

In terms of computational costs, the error-controlled full residual approximation used by STSOLVE is substantially more expensive in all considered tests than the heuristic ALS-based residual approximation used by AMEN. The precise CPU timings are of limited significance due to the different implementations, but we observe running



(a) STSOLVE **(b)** AMEN

Fig. 6 Results for Sect. 7.4: residual bounds (black) and maximum approximation ranks (gray), for well-conditioned representation of $\boldsymbol{B}_L$, $L = 50$

times on the order of several minutes with STSOLVE and of seconds with AMEN in the tests with $D = 1$, and of several hours with STSOLVE and several minutes with AMEN in the case of $D = 2$. Although no convergence analysis is available for this AMEN implementation, especially for the present well-conditioned representations it is thus an interesting practical choice.

8 Conclusion and Outlook

We have identified notions of condition numbers of tensor representations that determine the propagation of errors in numerical algorithms. In the application to multilevel tensor-structured discretizations of second-order elliptic PDEs, the careful construction of tensor representations of preconditioned system matrices guided by these notions leads to solvers that remain numerically stable also for very large discretization levels. For one such method based on soft thresholding of tensors, we have shown that the total number of arithmetic operations scales like a fixed power of the logarithm of the prescribed bound on the total solution error.

The new variant of BPX preconditioning that we have analyzed leads to a very natural low-rank structure of the symmetrically preconditioned stiffness matrix. Remarkably, unlike the rank increase with discretization levels observed in the case of separation of *spatial coordinates* [5], in the present case of tensor separation of *scales*, we obtain preconditioner representation ranks that remain uniformly bounded with respect to the discretization level. Similar results can be obtained for related preconditioners based on wavelet transforms, which are the subject of ongoing work.

For the preconditioned solvers, the relevant approximability properties of solutions we have identified are slightly different from the ones for nodal basis coefficients studied, e.g., in [34]. The numerically observed favorable decay of TT singular values of preconditioned quantities thus requires further investigation; it also depends on the particular choice of preconditioner.

The practical application to more general problems was not considered here to avoid further technicalities, but one can similarly treat different boundary conditions, more general coefficients (such as highly oscillatory diffusion coefficients in $D > 1$) or more general domains by techniques developed in [28]. We also expect that our basic

considerations concerning the combined low-rank representations of preconditioners and discretization matrices of differential operators can be applied, with potentially more technical effort, to other types of basis expansions and to different classes of PDE problems.

Although the representation ranks of preconditioned matrices that we obtain are bounded independently of the discretization level, they are fairly large for $D > 1$. This suggests the further investigation of solvers with improved quantitative performance, in particular the combination of AMEn-type methods with efficient residual approximation strategies for preconditioned operator representations.

We expect that the framework we have proposed here for studying the conditioning of tensor representations can be developed further to provide more detailed information, as well as sharper computable bounds for representations of matrices.

Acknowledgements Open Access funding provided by Projekt DEAL.

A Preconditioner Optimality

In preparation of the proof of Theorem 2, we define the square matrix $\boldsymbol{D}_L$ of order 2^{DL} by

$$(\boldsymbol{D}_L)_{j\,j'} = \big\langle \nabla\varphi_{L,j}, \nabla\varphi_{L,j'}\big\rangle_{\mathrm{L}^2(\Omega)} \quad \text{for all} \quad j, j' \in \mathscr{J}_L. \tag{103}$$

Since the bilinear form a is elliptic on V with $\|\cdot\|_V = \|\nabla\cdot\|_{\mathrm{L}^2(\Omega)^d}$, we obtain

$$\begin{aligned}\langle \boldsymbol{C}_L\boldsymbol{A}_L\boldsymbol{C}_L\boldsymbol{v}, \boldsymbol{v}\rangle &= a\Big(\sum_{\ell=0}^{L} 2^{-\ell} \sum_{j\in\mathscr{J}_\ell} \varphi_{\ell,j}(\boldsymbol{P}_{\ell,L}^{\mathsf{T}}\boldsymbol{v})_j, \sum_{\ell=0}^{L} 2^{-\ell} \sum_{j\in\mathscr{J}_\ell} \varphi_{\ell,j}(\boldsymbol{P}_{\ell,L}^{\mathsf{T}}\boldsymbol{v})_j\Big)\\ &\sim \Big\|\nabla\Big(\sum_{\ell=0}^{L} 2^{-\ell} \sum_{j\in\mathscr{J}_\ell} \varphi_{\ell,j}(\boldsymbol{P}_{\ell,L}^{\mathsf{T}}\boldsymbol{v})_j\Big)\Big\|_{\mathrm{L}^2}^2 = \langle \boldsymbol{C}_L\boldsymbol{D}_L\boldsymbol{C}_L\boldsymbol{v}, \boldsymbol{v}\rangle,\end{aligned}$$

and it thus suffices to show (31) for $\boldsymbol{D}_L$ in place of $\boldsymbol{A}_L$.

For $\ell = 0, \dots, L$, we introduce the nested subspaces $\mathscr{V}_\ell = \operatorname{ran}\boldsymbol{P}_{\ell,L} \subseteq \mathbb{R}^{\mathscr{J}_L}$; that is, the spaces $\mathscr{V}_\ell$ are spanned by vectors of finest-grid nodal values of the functions $\varphi_{\ell,j}$, $j \in \mathscr{J}_\ell$. In particular, $\mathscr{V}_L = \mathbb{R}^{\mathscr{J}_L}$.

Lemma 6 *For $\ell, \ell' \in \{0, \dots, L\}$, let*

$$\boldsymbol{L}_{\ell,\ell'} = 2^{-\ell-\ell'}\, \boldsymbol{P}_{\ell,L}^{\mathsf{T}}\, \boldsymbol{D}_L\, \boldsymbol{P}_{\ell',L}. \tag{104}$$

Then, for $0 \le k \le \ell$, $0 \le k' \le \ell'$,

$$\left|\left\langle \boldsymbol{L}_{\ell,\ell'} \boldsymbol{P}_{\ell',L}^{\mathsf{T}} \boldsymbol{w}_{k'},\, \boldsymbol{P}_{\ell,L}^{\mathsf{T}} \boldsymbol{w}_k \right\rangle\right| \lesssim 2^{-\frac{1}{2}|\ell'-\ell|}\, 2^{\frac{1}{2}(k'-\ell')}\, \|\boldsymbol{w}_{k'}\|_2\, 2^{\frac{1}{2}(k-\ell)}\, \|\boldsymbol{w}_k\|_2 \tag{105}$$

for all $\boldsymbol{w}_k \in \mathscr{V}_k$ and $\boldsymbol{w}_{k'} \in \mathscr{V}_{k'}$.

Proof The matrices defined in 104 can also be expressed in terms of

$$\begin{aligned} \hat{\boldsymbol{L}}_{\ell,\ell'} &= \big(2^{-\ell-\ell'} \langle \hat{\varphi}'_{\ell,j}, \hat{\varphi}'_{\ell',j'} \rangle_{\mathrm{L}^2(0,1)}\big)_{j \in \hat{\mathscr{J}}_\ell,\, j' \in \hat{\mathscr{J}}_{\ell'}}, \\ \hat{\boldsymbol{E}}_{\ell,\ell'} &= \big(\langle \hat{\varphi}_{\ell,j}, \hat{\varphi}_{\ell',j'} \rangle_{\mathrm{L}^2(0,1)}\big)_{j \in \hat{\mathscr{J}}_\ell,\, j' \in \hat{\mathscr{J}}_{\ell'}} \end{aligned} \tag{106}$$

as

$$\boldsymbol{L}_{\ell,\ell'} = \sum_{d=1}^{D} \Big(\bigotimes_{i=1}^{d-1} \hat{\boldsymbol{E}}_{\ell,\ell'} \Big) \otimes \hat{\boldsymbol{L}}_{\ell,\ell'} \otimes \Big(\bigotimes_{i=d+1}^{D} \hat{\boldsymbol{E}}_{\ell,\ell'} \Big). \tag{107}$$

The matrices $\hat{\boldsymbol{L}}_{\ell,\ell'}$ for $\ell > \ell'$ can be written in terms of $\hat{\boldsymbol{L}}_{\ell',\ell'}$ as follows: since $\hat{\varphi}'_{\ell,j}$ for $j \in \hat{\mathscr{J}}_\ell$ with $j < 2^\ell$ are L^2-orthogonal to constants, the inner products of these functions with $\hat{\varphi}'_{\ell',j'}$, $j' \in \hat{\mathscr{J}}_{\ell'}$, can be nonzero only when $j = 2^{\ell-\ell'} j'$. For $\ell \ge \ell'$, we thus define $\hat{\boldsymbol{\Xi}} \in \mathbb{R}^{\hat{\mathscr{J}}_\ell \times \hat{\mathscr{J}}_{\ell'}}$ by

$$\big(\hat{\boldsymbol{\Xi}}_{\ell,\ell'}\big)_{j\,j'} = \delta_{j,\,2^{\ell-\ell'} j'} \quad \text{for all} \quad j \in \hat{\mathscr{J}}_\ell,\ j' \in \hat{\mathscr{J}}_{\ell'}.$$

Additionally taking into account the difference in L^2-normalization factors between levels ℓ and ℓ', we obtain

$$\hat{\boldsymbol{L}}_{\ell,\ell'} = 2^{-\frac{1}{2}|\ell'-\ell|}\, \hat{\boldsymbol{\Xi}}_{\ell,\ell'} \hat{\boldsymbol{L}}_{\ell',\ell'}\,.$$

Let $\hat{\mathscr{V}}_{k,\ell} = \operatorname{ran} \hat{\boldsymbol{P}}_{k,\ell} \subseteq \mathbb{R}^{\hat{\mathscr{J}}_\ell}$ and $\hat{\mathscr{V}}_k = \operatorname{ran} \hat{\boldsymbol{P}}_{k,L}$. For $k \le \ell$ and $\boldsymbol{w} \in \hat{\mathscr{V}}_{k,\ell}$, let $w \in V_k$ be the function represented by $\boldsymbol{w}$. Then, by (106) and the standard inverse estimate for V_k (see, e.g., [24, Sec. 8.8.3]), we have $\langle \hat{\boldsymbol{L}}_{\ell,\ell} \boldsymbol{w}, \boldsymbol{w} \rangle = 2^{-2\ell} |w|^2_{\mathrm{H}^1_0} \lesssim 2^{2(k-\ell)} \|w\|^2_{\mathrm{L}^2}$, and thus

$$\langle \hat{\boldsymbol{L}}_{\ell,\ell} \boldsymbol{w}, \boldsymbol{w} \rangle \le 2^{2(k-\ell)} \|\boldsymbol{w}\|_2^2, \qquad k \le \ell,\ \boldsymbol{w} \in \hat{\mathscr{V}}_{k,\ell}; \tag{108}$$

in particular, we also have $\|\hat{\boldsymbol{L}}_{\ell,\ell}\|_{2\to 2} \lesssim 1$. Moreover, one has

$$\langle \hat{\boldsymbol{\Xi}}_{\ell,\ell'} \hat{\boldsymbol{L}}_{\ell',\ell'} \hat{\boldsymbol{\Xi}}_{\ell,\ell'}^{\mathsf{T}} \boldsymbol{w}, \boldsymbol{w} \rangle \le 2^{k-\ell+\min\{k-\ell',0\}} \|\boldsymbol{w}\|_2^2, \qquad k, \ell' \le \ell,\ \boldsymbol{w} \in \hat{\mathscr{V}}_{k,\ell}. \tag{109}$$

To see this, denote again by $w \in V_k$ the function represented by $\boldsymbol{w}$, and consider first $\ell' \le k \le \ell$. Then, $\tilde{\boldsymbol{w}} := \hat{\boldsymbol{\Xi}}_{\ell,\ell'}^{\mathsf{T}} \boldsymbol{w}$ corresponds to evaluations of w on the grid of level ℓ', which is coarser than the one on which it is piecewise linear, and consequently $2^{\ell-k} \sum_{j' \in \hat{\mathscr{J}}_{\ell'}} |\tilde{\boldsymbol{w}}_{j'}|^2 \lesssim \sum_{j \in \hat{\mathscr{J}}_\ell} |\boldsymbol{w}_j|^2$. Thus, $\|\tilde{\boldsymbol{w}}\|_2 = \|\hat{\boldsymbol{\Xi}}_{\ell,\ell'}^{\mathsf{T}} \boldsymbol{w}\|_2 \lesssim 2^{\frac{1}{2}(k-\ell)} \|\boldsymbol{w}\|_2$, and (109) follows in this case. If $k < \ell' \le \ell$, $\tilde{\boldsymbol{w}} \in \hat{\mathscr{V}}_{k,\ell'}$ corresponds to a reinterpolation of w that is still on a finer level than k, and thus $\|\tilde{\boldsymbol{w}}\|_2 \le 2^{\frac{1}{2}(\ell'-\ell)} \|\boldsymbol{w}\|_2$. Using (108), we thus obtain $\langle \hat{\boldsymbol{L}}_{\ell',\ell'} \tilde{\boldsymbol{w}}, \tilde{\boldsymbol{w}} \rangle \lesssim 2^{2(k-\ell')} \|\tilde{\boldsymbol{w}}\|_2^2 \lesssim 2^{2k-2\ell'+\ell'-\ell} \|\boldsymbol{w}\|_2^2$, which gives (109).

We next show that

$$\|\hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}} \hat{\boldsymbol{P}}_{k,L} - \hat{\boldsymbol{P}}_{k,\ell}\|_{2\to 2} \lesssim 2^{\frac{1}{2}(k-\ell)}, \quad k \le \ell. \tag{110}$$

Let $s_{ji} := (\hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}} \hat{\boldsymbol{P}}_{k,L})_{ji}$, $v_{ji} := (\hat{\boldsymbol{P}}_{k,\ell})_{ji}$, $j \in \hat{\mathscr{J}}_\ell$, $i \in \hat{\mathscr{J}}_k$. Recalling (10), and taking into account that $\operatorname{supp} \hat{\varphi}_{\ell,j} = [2^{-\ell}(j-1), 2^{-\ell}(j+1)] \cap [0,1]$,

$$s_{ji} = 2^{-L} \sum_{n=2^{L-\ell}(j-1)}^{\min\{2^{L-\ell}(j+1),2^L\}} \hat{\varphi}_{\ell,j}(2^{-L}n)\, \hat{\varphi}_{k,i}(2^{-L}n), \qquad v_{ji} = 2^{-\frac{1}{2}\ell} \hat{\varphi}_{k,i}(2^{-\ell}j).$$

Whenever $\hat{\varphi}_{k,i}$ is linear on $\operatorname{supp} \hat{\varphi}_{\ell,j}$, one has $s_{ji} = v_{ji}$ by the symmetries in the summation in s_{ji}. This fails to hold only when $j = 2^{\ell-k} i$. In these cases, one easily verifies that $|s_{ji} - v_{ji}| \lesssim 2^{\frac{3}{2}(k-\ell)}$ when $j < 2^\ell$ and $|s_{ji} - v_{ji}| \lesssim 2^{\frac{1}{2}(k-\ell)}$ for $i = 2^k$, $j = 2^\ell$, with L-independent constants. Using interpolation to bound $\|\hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}} \hat{\boldsymbol{P}}_{k,L} - \hat{\boldsymbol{P}}_{k,\ell}\|_{2\to 2}$ by the corresponding row- and column-sum norms, where the number of nonzero entries in each row and column is uniformly bounded, we obtain (110).

Note that for any $\boldsymbol{w} \in \hat{\mathscr{V}}_k$ there exists a unique $\boldsymbol{z} \in \mathbb{R}^{\hat{\mathscr{J}}_k}$ such that $\boldsymbol{w} = \hat{\boldsymbol{P}}_{k,L} \boldsymbol{z}$, where $\|\boldsymbol{w}\|_2 \sim \|\boldsymbol{z}\|_2$ with constants independent of k, L. As a consequence, using this with (110), we obtain $\|\hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}} \boldsymbol{w} - \hat{\boldsymbol{P}}_{k,\ell} \boldsymbol{z}\|_2 \lesssim 2^{\frac{1}{2}(k-\ell)} \|\boldsymbol{w}\|_2$ for such $\boldsymbol{w}$ and $\boldsymbol{z}$. Since

$$\begin{aligned}
\langle \hat{\boldsymbol{L}}_{\ell,\ell} \hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}} \boldsymbol{w}, \hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}} \boldsymbol{w} \rangle &= \langle \hat{\boldsymbol{L}}_{\ell,\ell} \hat{\boldsymbol{P}}_{k,\ell} \boldsymbol{z}, \hat{\boldsymbol{P}}_{k,\ell} \boldsymbol{z} \rangle \\
&\quad + 2 \langle \hat{\boldsymbol{L}}_{\ell,\ell} (\hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}} \boldsymbol{w} - \hat{\boldsymbol{P}}_{k,\ell} \boldsymbol{z}), \hat{\boldsymbol{P}}_{k,\ell} \boldsymbol{z} \rangle \\
&\quad + \langle \hat{\boldsymbol{L}}_{\ell,\ell} (\hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}} \boldsymbol{w} - \hat{\boldsymbol{P}}_{k,\ell} \boldsymbol{z}), (\hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}} \boldsymbol{w} - \hat{\boldsymbol{P}}_{k,\ell} \boldsymbol{z}) \rangle,
\end{aligned}$$

using (108) for $\hat{\boldsymbol{P}}_{k,\ell} \boldsymbol{z} \in \hat{\mathscr{V}}_{k,\ell}$, $\|\hat{\boldsymbol{L}}_{\ell,\ell}\|_{2\to 2} \lesssim 1$, and the Cauchy–Schwarz inequality for the middle term on the right, we obtain

$$\langle \hat{\boldsymbol{L}}_{\ell,\ell} \hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}} \boldsymbol{w}, \hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}} \boldsymbol{w} \rangle \lesssim (2^{2(k-\ell)} + 2^{\frac{3}{2}(k-\ell)} + 2^{k-\ell}) \|\boldsymbol{w}\|_2^2 \lesssim 2^{k-\ell} \|\boldsymbol{w}\|_2^2,$$

and similarly, using (109) in the same manner,

$$\langle \hat{\boldsymbol{\Xi}}_{\ell,\ell'} \hat{\boldsymbol{L}}_{\ell',\ell'} \hat{\boldsymbol{\Xi}}_{\ell,\ell'}^{\mathsf{T}} \hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}} \boldsymbol{w}, \hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}} \boldsymbol{w} \rangle \lesssim 2^{k-\ell} \|\boldsymbol{w}\|_2^2,$$

for any $\boldsymbol{w} \in \hat{\mathscr{V}}_k$, $k \le \ell$.

Consequently, with $0 \le k \le \ell$, $0 \le k' \le \ell'$, $\ell \le \ell'$, for all $\boldsymbol{w}_k \in \hat{\mathscr{V}}_k$ and $\boldsymbol{w}_{k'} \in \hat{\mathscr{V}}_{k'}$,

$$\begin{aligned}\big|\langle \hat{\boldsymbol{L}}_{\ell,\ell'} \hat{\boldsymbol{P}}_{\ell',L}^{\mathsf{T}} \boldsymbol{w}_{k'}, \hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}} \boldsymbol{w}_k \rangle\big| &= 2^{-\frac{1}{2}|\ell'-\ell|} \big|\langle \hat{\boldsymbol{L}}_{\ell,\ell'} \hat{\boldsymbol{P}}_{\ell',L}^{\mathsf{T}} \boldsymbol{w}_{k'}, \hat{\boldsymbol{\Xi}}_{\ell,\ell'}^{\mathsf{T}} \hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}} \boldsymbol{w}_k \rangle\big| \\ &\le 2^{-\frac{1}{2}|\ell'-\ell|} \langle \hat{\boldsymbol{L}}_{\ell,\ell'} \hat{\boldsymbol{P}}_{\ell',L}^{\mathsf{T}} \boldsymbol{w}_{k'}, \hat{\boldsymbol{P}}_{\ell',L}^{\mathsf{T}} \boldsymbol{w}_{k'} \rangle^{\frac{1}{2}} \\ &\quad \times \langle \hat{\boldsymbol{L}}_{\ell,\ell'} \hat{\boldsymbol{\Xi}}_{\ell,\ell'}^{\mathsf{T}} \hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}} \boldsymbol{w}_k, \hat{\boldsymbol{\Xi}}_{\ell,\ell'}^{\mathsf{T}} \hat{\boldsymbol{P}}_{\ell,L}^{\mathsf{T}} \boldsymbol{w}_k \rangle^{\frac{1}{2}} \\ &\le 2^{-\frac{1}{2}|\ell'-\ell|}\, 2^{\frac{1}{2}(k'-\ell')} \|\boldsymbol{w}_{k'}\|_2\, 2^{\frac{1}{2}(k-\ell)} \|\boldsymbol{w}_k\|_2 \,. \end{aligned} \tag{111}$$

By (107), since $\|\hat{\boldsymbol{E}}_{\ell,\ell'}\|_{2\to 2} \le 1$, this implies (105). □

Proof (Theorem 2) Theorem 1 implies in particular that $\langle \boldsymbol{C}_{2,L} \boldsymbol{v}, \boldsymbol{v} \rangle \sim \langle \boldsymbol{D}_L^{-1} \boldsymbol{v}, \boldsymbol{v} \rangle$ for all $\boldsymbol{v}$, that is,

$$\langle \boldsymbol{D}_L^{-1} \boldsymbol{v}, \boldsymbol{v} \rangle \sim \sum_{\ell=0}^{L} \big\| 2^{-\ell} \boldsymbol{P}_{\ell,L}^{\mathsf{T}} \boldsymbol{v} \big\|_2^2. \tag{112}$$

We use this in the following proof of the lower bound in (31), which is inspired by arguments using frame theory from [26]. Let $\bar{\mathscr{V}}_L = \times_{\ell=0}^{L} \mathbb{R}^{\mathscr{J}_\ell}$. We consider the mappings $\boldsymbol{F} : \mathscr{V}_L \to \bar{\mathscr{V}}_L$ and $\boldsymbol{F}^{\mathsf{T}} : \bar{\mathscr{V}}_L \to \mathscr{V}_L$ given by

$$\boldsymbol{F} : \boldsymbol{v} \mapsto \big(2^{-\ell} \boldsymbol{P}_{\ell,L}^{\mathsf{T}} \boldsymbol{v}\big)_{\ell=0,\dots,L}, \qquad \boldsymbol{F}^{\mathsf{T}} : (\boldsymbol{v}_\ell)_{\ell=0,\dots,L} \mapsto \sum_{\ell=0}^{L} 2^{-\ell} \boldsymbol{P}_{\ell,L} \boldsymbol{v}_\ell.$$

For any $\boldsymbol{w} = (\boldsymbol{w}_\ell)_{\ell=0,\dots,L} \in \operatorname{ran} \boldsymbol{F}$, where $\boldsymbol{w} = \boldsymbol{F}\boldsymbol{v}$ for $\boldsymbol{v} \in \mathscr{V}_L$, we obtain

$$\|\boldsymbol{F}^{\mathsf{T}} \boldsymbol{w}\|_{\boldsymbol{D}_L} = \sup_{z \ne 0} \frac{\langle \boldsymbol{F}^{\mathsf{T}} \boldsymbol{w}, z \rangle}{\|z\|_{\boldsymbol{D}_L^{-1}}} = \sup_{z \ne 0} \frac{\langle \boldsymbol{F}\boldsymbol{v}, \boldsymbol{F}z \rangle}{\sqrt{\langle \boldsymbol{D}_L^{-1} z, z \rangle}} \sim \|\boldsymbol{F}\boldsymbol{v}\|_2 = \|\boldsymbol{w}\|_2$$

by (112). Now let $\boldsymbol{G} : \mathscr{V}_L \to \bar{\mathscr{V}}_L$, $\boldsymbol{v} \mapsto \big(\boldsymbol{P}_{\ell,L}^{\mathsf{T}} \boldsymbol{v}\big)_{\ell=0,\dots,L}$. Then, $\operatorname{ran} \boldsymbol{G} \subset \operatorname{ran} \boldsymbol{F}$, and thus

$$\langle \boldsymbol{C}_L \boldsymbol{D}_L \boldsymbol{C}_L \boldsymbol{v}, \boldsymbol{v} \rangle = \|\boldsymbol{F}^{\mathsf{T}} \boldsymbol{G} \boldsymbol{v}\|_{\boldsymbol{D}_L}^2 \sim \|\boldsymbol{G}\boldsymbol{v}\|_2^2 \gtrsim \|\boldsymbol{v}\|_2^2,$$

which shows the lower bound in (31).

Arguing along similar lines to obtain the upper bound in (31) would lead to a constant depending linearly on L, and we thus now turn to a different approach using Lemma 6. Let $\boldsymbol{R}_\ell = \boldsymbol{P}_{\ell,L} (\boldsymbol{P}_{\ell,L}^{\mathsf{T}} \boldsymbol{P}_{\ell,L})^{-1} \boldsymbol{P}_{\ell,L}^{\mathsf{T}}$ be the discrete orthogonal projector onto $\mathscr{V}_\ell$. For any $\boldsymbol{w} \in \mathscr{V}_L$, setting $\boldsymbol{w}_0 = \boldsymbol{R}_0 \boldsymbol{w}$ and $\boldsymbol{w}_\ell = (\boldsymbol{R}_\ell - \boldsymbol{R}_{\ell-1})\, \boldsymbol{w}$ for $\ell = 1, \dots, L$, we obtain the decomposition

$$\boldsymbol{w} = \sum_{\ell=0}^{L} \boldsymbol{w}_\ell \quad \text{with} \quad \|\boldsymbol{w}\|_2^2 = \sum_{\ell=0}^{L} \|\boldsymbol{w}_\ell\|_2^2 \,, \tag{113}$$

which yields

$$\langle \boldsymbol{C}_L \boldsymbol{D}_L \boldsymbol{C}_L \boldsymbol{w}, \boldsymbol{w} \rangle = \sum_{\ell,\ell'=0}^{L} \Big\langle \boldsymbol{L}_{\ell,\ell'} \boldsymbol{P}_{\ell',L}^{\mathsf{T}} \sum_{k'=0}^{\ell'} \boldsymbol{w}_{k'},\ \boldsymbol{P}_{\ell,L}^{\mathsf{T}} \sum_{k=0}^{\ell} \boldsymbol{w}_k \Big\rangle.$$

For $n = 0, 1, \ldots, L$, by Lemma 6,

$$\begin{aligned}
&\sum_{\ell=0}^{L-n} \Big\langle \boldsymbol{L}_{\ell,\ell+n} \boldsymbol{P}_{\ell+n,L}^{\mathsf{T}} \sum_{k'=0}^{\ell+n} \boldsymbol{w}_{k'},\ \boldsymbol{P}_{\ell,L}^{\mathsf{T}} \sum_{k=0}^{\ell} \boldsymbol{w}_k \Big\rangle \\
&\quad \lesssim 2^{-\frac{1}{2}n} \sum_{\ell=0}^{L-n} \sum_{k'=0}^{\ell+n} \sum_{k=0}^{\ell} 2^{\frac{1}{2}(k'-\ell-n)}\, 2^{\frac{1}{2}(k-\ell)}\, \|\boldsymbol{w}_k\|_2\, \|\boldsymbol{w}_{k'}\|_2 \\
&\quad \le 2^{-\frac{1}{2}n} \sum_{\ell=0}^{L-n} \Big\{ \sum_{k'=0}^{\ell+n} 2^{\frac{1}{2}(k'-\ell-n)}\, \|\boldsymbol{w}_{k'}\|_2^2 + \sum_{k=0}^{\ell} 2^{\frac{1}{2}(k-\ell)}\, \|\boldsymbol{w}_k\|_2^2 \Big\}.
\end{aligned}$$

We thus arrive at

$$\langle \boldsymbol{C}_L \boldsymbol{D}_L \boldsymbol{C}_L \boldsymbol{w}, \boldsymbol{w} \rangle \lesssim \sum_{n=0}^{L} 2^{-\frac{1}{2}n} \sum_{\ell=0}^{L} \|\boldsymbol{w}_\ell\|_2^2 \lesssim \|\boldsymbol{w}\|_2^2,$$

completing the proof of the upper bound in (31) and hence of Theorem 2. □

Remark 11 Although we have used some simplifications due to the tensor structure in our particular setting, the proof of Theorem 2 carries over to more general hierarchies of finite element spaces, provided that one can establish a corresponding strengthened Cauchy–Schwarz inequality as in (105), see, e.g., [9,52,54].

B Rank-Reduced Decomposition

The following proof of Lemma 5 relies on properties of the strong Kronecker product inherited from the matrix and Kronecker products: linearity, associativity and distributivity. In particular, products of cores can be transformed into products of smaller cores by eliminating linear dependence from the decomposition, as the following example illustrates.

For any scalar coefficients α, β and blocks *or subcores* V_{11}, V_{12}, V_{21}, V_{22}, W_{11}, W_{12} of suitable rank and mode size, we have

$$\begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \ltimes \begin{bmatrix} \alpha W_{11} & \alpha W_{12} \\ \beta W_{11} & \beta W_{12} \end{bmatrix} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \ltimes \left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \ltimes \begin{bmatrix} W_{11} & W_{12} \end{bmatrix} \right) \tag{114a}$$

$$= \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \ltimes \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \ltimes \begin{bmatrix} W_{11} & W_{12} \end{bmatrix} \tag{114b}$$

$$= \left(\begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \bowtie \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \right) \bowtie \begin{bmatrix} W_{11} & W_{12} \end{bmatrix} = \begin{bmatrix} \alpha V_{11} + \beta V_{12} \\ \alpha V_{21} + \beta V_{22} \end{bmatrix} \bowtie \begin{bmatrix} W_{11} & W_{12} \end{bmatrix}. \tag{114c}$$

When the partitioning shown in (114a)–(114c) is in terms of blocks (which, by our identification convention, are subcores of rank 1×1), the rank of the product is 2×2. The left-hand side of (114a) and the right-hand side of (114c) represent this core "in the TT format," which has only one rank parameter and happens to be nothing else than low-rank matrix factorization in these two cases. The "ranks" of the first decomposition, equal to 2, are larger than the "ranks" of the last decomposition, equal to 1.

The TT representation (114b) consists of three cores and has ranks 2, 1. However, all mode indices of its middle core are dummy indices (the mode size of the middle core is 1×1), so the middle core can be merged with either of the neighboring cores *without changing the decomposition scheme* (by the latter we mean the set and the ordering of the variables separated by the TT format).

Proof (Lemma 5) Let $\hat{\boldsymbol{N}}_{\ell,L,\alpha} = \hat{\boldsymbol{M}}_{L,\alpha} \hat{\boldsymbol{P}}_{\ell,L}$ and $c_{\ell,L} = 2^{(\alpha+\frac{1}{2})L - \frac{1}{2}(L-\ell)}$. Applying Lemma 3 with the same ℓ as fixed here, we obtain

$$\hat{\boldsymbol{M}}_{L,\alpha} = 2^{(\alpha+\frac{1}{2})L} \; \hat{A} \bowtie \hat{U}^{\bowtie \ell} \bowtie \hat{T}_0 \bowtie \hat{V}^{\bowtie (L-\ell)} \bowtie \hat{I} \bowtie \hat{M}_\alpha, \tag{115a}$$

where $\hat{I}$ is as defined in (78). On the other hand, Lemma 4 gives the decomposition

$$\hat{\boldsymbol{P}}_{\ell,L} = 2^{-\frac{1}{2}(L-\ell)} \; \hat{A} \bowtie \hat{U}^{\bowtie \ell} \bowtie \hat{I} \bowtie \hat{X}^{\bowtie (L-\ell)} \bowtie \hat{P} \bowtie \begin{bmatrix} 1 \end{bmatrix}. \tag{115b}$$

Rewriting matrix multiplication core-wise, we combine the rank 2 decompositions given by (115a)–(115b) into a rank 4 decomposition for the product:

$$\hat{\boldsymbol{N}}_{\ell,L,\alpha} = c_{\ell,L} \; \hat{A}_\flat \bowtie \hat{U}_\sharp^{\bowtie \ell} \bowtie \hat{W}_0 \bowtie \hat{Y}_\sharp^{\bowtie (L-\ell)} \bowtie E \bowtie \hat{M}_\alpha, \tag{115c}$$

where $\hat{A}_\flat$ and $\hat{W}_\alpha$ with $\alpha = 0, 1$ are as in (81) and (83) and $E = \hat{I} \bullet \hat{P}$, $\hat{U}_\sharp = \hat{U} \bullet \hat{U}$ and $\hat{Y}_\sharp = \hat{V} \bullet \hat{X}$ are newly introduced cores. Direct calculation with expressions given in (67), (75) and (78) yields

$$E = \begin{bmatrix} 1 & \\ 0 & \\ & 1 \\ & 0 \end{bmatrix}, \quad \hat{U}_\sharp = \begin{bmatrix} I & J^\mathsf{T} & J^\mathsf{T} & \\ & J & & I_2 \\ & & J & I_1 \end{bmatrix}, \quad \hat{Y}_\sharp = \frac{1}{4} \begin{bmatrix} \binom{3}{3} & \binom{1}{1} & \binom{-1}{1} & \binom{-1}{1} \\ \binom{1}{1} & \binom{3}{3} & \binom{1}{-1} & \binom{1}{-1} \\ \binom{-1}{3} & \binom{-1}{1} & \binom{3}{1} & \binom{1}{1} \\ \binom{1}{1} & \binom{1}{3} & \binom{1}{-1} & \binom{3}{-1} \end{bmatrix}$$

in terms of the blocks I, I_1, I_2 and J defined in (37).

Sweeping from level L to level 1. Let us define the following cores:

$$C = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \\ & 1 & & 0 \end{bmatrix} \quad \text{and} \quad G = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \\ & & & 0 \end{bmatrix}.$$

First, we note that the second and fourth rows in each of the cores E and $\hat{Y}_\sharp \bowtie C$ are equal. This implies that $E = C \bowtie E$ and $\hat{Y}_\sharp \bowtie C = C \bowtie \hat{Y}_\sharp \bowtie C$. Further, in each of the cores $\hat{W}_0 \bowtie C$ and $\hat{U}_\sharp$, the last row is zero, so that $\hat{W}_0 \bowtie C = G \bowtie \hat{W}_0 \bowtie C$ and $\hat{U}_\sharp = G \bowtie \hat{U}_\sharp$. These equalities allow to sweep the cores C and G through the last $L - \ell$ and first ℓ levels, respectively: starting from (115c), we obtain

$$\begin{aligned} \hat{N}_{\ell,L,\alpha} &= c_{\ell,L}\ \hat{A}_\flat \bowtie \hat{U}_\sharp^{\bowtie \ell} \bowtie \hat{W}_0 \bowtie \hat{Y}_\sharp^{\bowtie(L-\ell)} \bowtie C \bowtie E \bowtie \hat{M}_\alpha \\ &= c_{\ell,L}\ \hat{A}_\flat \bowtie \hat{U}_\sharp^{\bowtie \ell} \bowtie \hat{W}_0 \bowtie C \bowtie (\hat{Y}_\sharp \bowtie C)^{\bowtie(L-\ell)} \bowtie E \bowtie \hat{M}_\alpha \\ &= c_{\ell,L}\ \hat{A}_\flat \bowtie \hat{U}_\sharp^{\bowtie \ell} \bowtie G \bowtie \hat{W}_0 \bowtie C \bowtie (\hat{Y}_\sharp \bowtie C)^{\bowtie(L-\ell)} \bowtie E \bowtie \hat{M}_\alpha \\ &= c_{\ell,L}\ \hat{A}_\flat \bowtie (\hat{U}_\sharp \bowtie G)^{\bowtie \ell} \bowtie \hat{W}_0 \bowtie C \bowtie (\hat{Y}_\sharp \bowtie C)^{\bowtie(L-\ell)} \bowtie E \bowtie \hat{M}_\alpha\,. \end{aligned} \tag{115d}$$

Sweeping from level 1 *to level L*. Further, we notice that the cores

$$F = \begin{bmatrix} 1 & & & \\ & 1 & 1 & 0 \end{bmatrix} \quad \text{and} \quad H = \begin{bmatrix} 1 & 1 & \\ & -1 & 1 & 0 \end{bmatrix}$$

satisfy the relations $\hat{A}_\flat = \hat{A} \bowtie F$, $F \bowtie \hat{U}_\sharp \bowtie G = \hat{U} \bowtie F$, $F \bowtie \hat{W}_0 \bowtie C = \hat{T}_0 \bowtie H$, $H \bowtie \hat{Y}_\sharp \bowtie C = \hat{Y}_0 \bowtie H$ and $H \bowtie E = \hat{I}$. These relations allow to sweep the cores F and H through the first ℓ and last $L - \ell$ levels respectively: continuing (115c), we derive

$$\begin{aligned} \hat{N}_{\ell,L,\alpha} &= c_{\ell,L}\ \hat{A} \bowtie F \bowtie (\hat{U}_\sharp \bowtie G)^{\bowtie \ell} \bowtie \hat{W}_0 \bowtie C \bowtie (\hat{Y}_\sharp \bowtie C)^{\bowtie(L-\ell)} \bowtie E \bowtie \hat{M}_\alpha \\ &= c_{\ell,L}\ \hat{A} \bowtie \hat{U}^{\bowtie \ell} \bowtie F \bowtie \hat{W}_0 \bowtie C \bowtie (\hat{Y}_\sharp \bowtie C)^{\bowtie(L-\ell)} \bowtie E \bowtie \hat{M}_\alpha \\ &= c_{\ell,L}\ \hat{A} \bowtie \hat{U}^{\bowtie \ell} \bowtie \hat{T}_0 \bowtie H \bowtie (\hat{Y}_\sharp \bowtie C)^{\bowtie(L-\ell)} \bowtie E \bowtie \hat{M}_\alpha \\ &= c_{\ell,L}\ \hat{A} \bowtie \hat{U}^{\bowtie \ell} \bowtie \hat{T}_0 \bowtie \hat{Y}_0^{\bowtie(L-\ell)} \bowtie H \bowtie E \bowtie \hat{M}_\alpha \\ &= c_{\ell,L}\ \hat{A} \bowtie \hat{U}^{\bowtie \ell} \bowtie \hat{T}_0 \bowtie \hat{Y}_0^{\bowtie(L-\ell)} \bowtie \hat{M}_\alpha\,. \end{aligned} \tag{115e}$$

This proves the claim in the case of $\alpha = 0$ since $\hat{M}_0 = \hat{N}_0$ by (75) and (78).

Sweeping from level L to level ℓ. In the decomposition (115e), the ranks involved in the core products to the right of $\hat{T}_0$ (in particular, those bounding the ranks of unfolding matrices $\ell, \ldots, L-1+\alpha$) are all equal to two. To prove the claim, it remains to consider the case of $\alpha = 1$ and obtain a reduced decomposition in which those ranks are all equal to one instead of two. To this end, we note that $\hat{Y}_0 \bowtie \hat{M}_1 = \hat{M}_1 \bowtie \hat{Y}_1 = \hat{M}_1 \bowtie \hat{Y}_1 \bowtie \hat{N}_1$

and $\hat{T}_0 \bowtie \hat{M}_1 = \hat{T}_1$. Applying these relations to (115e), we obtain the claim in the case of $\alpha = 1$. □

References

1. Andreev, R., Tobler, C.: 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)
2. Babuška, I., Guo, B.: The *h*-*p* version of the finite element method for domains with curved boundaries. SIAM Journal on Numerical Analysis **25**(4), 837–861 (1988)
3. Babuška, I., Guo, B.: Regularity of the solution of elliptic problems with piecewise analytic data. Part I. boundary value problems for linear elliptic equation of second order. SIAM Journal on Mathematical Analysis **19**(1), 172–203 (1988)
4. Bachmayr, M., Dahmen, W.: Adaptive near-optimal rank tensor approximation for high-dimensional operator equations. Found. Comput. Math. **15**(4), 839–898 (2015)
5. Bachmayr, M., Dahmen, W.: Adaptive low-rank methods: problems on Sobolev spaces. SIAM J. Numer. Anal. **54**(2), 744–796 (2016)
6. Bachmayr, M., Schneider, R.: Iterative methods based on soft thresholding of hierarchical tensors. Found. Comput. Math. **17**, 1037–1083 (2017)
7. Bachmayr, M., Schneider, R., Uschmajew, A.: Tensor networks and hierarchical tensors for the solution of high-dimensional partial differential equations. Found. Comput. Math. **16**(6), 1423–1472 (2016)
8. Ballani, J., Grasedyck, L.: A projection method to solve linear systems in tensor format. Numerical Linear Algebra with Applications **20**(1), 27–43 (2013)
9. Bornemann, F., Yserentant, H.: A basic norm equivalence for the theory of multilevel methods. Numer. Math. **64**(4), 455–476 (1993)
10. Bramble, J.H., Pasciak, J.E., Xu, J.: Parallel multilevel preconditioners. Math. Comp. **55**(191), 1–22 (1990)
11. Chertkov, A.V., Oseledets, I.V., Rakhuba, M.V.: Robust discretization in quantized tensor train format for elliptic problems in two dimensions. arXiv:1612.01166 (2016)
12. Dahmen, W., Kunoth, A.: Multilevel preconditioning. Numer. Math. **63**(3), 315–344 (1992)
13. De Launey, W., Seberry, J.: The strong Kronecker product. Journal of Combinatorial Theory, Series A **66**(2), 192–213 (1994)
14. de Silva, V., Lim, L.H.: Tensor rank and the ill-posedness of the best low-rank approximation problem. SIAM Journal on Matrix Analysis and Applications **30**(3), 1084–1127 (2008)
15. Dolgov, S.V., Kazeev, V.A., Khoromskij, B.N.: Direct tensor-product solution of one-dimensional elliptic equations with parameter-dependent coefficients. Mathematics and Computers in Simulation **145**(Supplement C), 136–155 (2018)
16. Dolgov, S.V., Khoromskij, B.N., Oseledets, I.V., Tyrtyshnikov, E.E.: Tensor structured iterative solution of elliptic problems with jumping coefficients. Preprint 55, Max Planck Institute for Mathematics in the Sciences (2010)
17. Dolgov, S.V., Savostyanov, D.V.: Alternating minimal energy methods for linear systems in higher dimensions. SIAM J. Sci. Comput. **36**(5), A2248–A2271 (2014)
18. Grasedyck, L.: Hierarchical singular value decomposition of tensors. SIAM Journal on Matrix Analysis and Applications **31**(4), 2029–2054 (2010)
19. Grasedyck, L.: Polynomial approximation in hierarchical Tucker format by vector-tensorization. Preprint 308, Institut für Geometrie und Praktische Mathematik, RWTH Aachen (2010)
20. Grasedyck, L., Kressner, D., Tobler, C.: A literature survey of low-rank tensor approximation techniques. GAMM-Mitteilungen **36**(1), 53–78 (2013)
21. Hackbusch, W.: Tensorisation of vectors and their efficient convolution. Numerische Mathematik **119**(3), 465 (2011)
22. Hackbusch, W.: Tensor Spaces and Numerical Tensor Calculus, *Springer Series in Computational Mathematics*, vol. 42. Springer (2012)
23. Hackbusch, W.: Solution of linear systems in high spatial dimensions. Computing and Visualization in Science **17**(3), 111–118 (2015)

24. Hackbusch, W.: Elliptic Differential Equations: Theory and Numerical Treatment, *Springer Series in Computational Mathematics*, vol. 18, second edn. Springer (2017)
25. Hackbusch, W., Kühn, S.: A new scheme for the tensor representation. J. Fourier Anal. Appl. **15**(5), 706–722 (2009)
26. Harbrecht, H., Schneider, R., Schwab, C.: Multilevel frames for sparse tensor product spaces. Numer. Math. **110**(2), 199–220 (2008)
27. Higham, N.J.: Accuracy and stability of numerical algorithms, second edn. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2002)
28. Kazeev, V.: Quantized tensor-structured finite elements for second-order elliptic PDEs in two dimensions. Ph.D. thesis, ETH Zürich (2015). https://doi.org/10.3929/ethz-a-010554062
29. Kazeev, V., Khammash, M., Nip, M., Schwab, C.: Direct solution of the chemical master equation using quantized tensor trains. PLOS Computational Biology **10**(3), 742–758 (2014)
30. Kazeev, V., Khoromskij, B., Tyrtyshnikov, E.: Multilevel Toeplitz matrices generated by tensor-structured vectors and convolution with logarithmic complexity. SIAM Journal on Scientific Computing **35**(3), A1511–A1536 (2013)
31. Kazeev, V., Oseledets, I., Rakhuba, M., Schwab, C.: QTT-finite-element approximation for multiscale problems I: model problems in one dimension. Adv. Comput. Math. **43**(2), 411–442 (2017)
32. Kazeev, V., Reichmann, O., Schwab, C.: Low-rank tensor structure of linear diffusion operators in the TT and QTT formats. Linear Algebra and its Applications **438**(11), 4204–4221 (2013)
33. Kazeev, V., Schwab, C.: Approximation of singularities by quantized-tensor FEM. Proceedings in Applied Mathematics and Mechanics **15**(1), 743–746 (2015)
34. Kazeev, V., Schwab, C.: Quantized tensor-structured finite elements for second-order elliptic PDEs in two dimensions. Numerische Mathematik **138**, 133–190 (2018)
35. Kazeev, V.A., Khoromskij, B.N.: Low-rank explicit QTT representation of the Laplace operator and its inverse. SIAM Journal on Matrix Analysis and Applications **33**(3), 742–758 (2012)
36. Khoromskaia, V., Khoromskij, B.N.: Grid-based lattice summation of electrostatic potentials by assembled rank-structured tensor approximation. Comp. Phys. Communications **185**(12), 3162–3174 (2014)
37. Khoromskij, B.N.: $\mathcal{O}(d \log n)$-quantics approximation of n-d tensors in high-dimensional numerical modeling. Constructive Approximation **34**(2), 257–280 (2011)
38. Khoromskij, B.N.: Tensor Numerical Methods in Scientific Computing. De Gruyter Verlag (2018)
39. Khoromskij, B.N., Oseledets, I.V.: QTT approximation of elliptic solution operators in higher dimensions. Russ. J. Numer. Anal. Math. Modelling **26**(3), 303–322 (2011)
40. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Review **51**(3), 455–500 (2009)
41. Kressner, D., Tobler, C.: Algorithm 941: Htucker—a matlab toolbox for tensors in hierarchical Tucker format. ACM Transactions on Mathematical Software **40**(3), 22:1–22:22 (2014)
42. Orús, R.: A practical introduction to tensor networks: Matrix product states and projected entangled pair states. Annals of Physics **349**(Supplement C), 117–158 (2014)
43. Oseledets, I.: Approximation of matrices with logarithmic number of parameters. Doklady Mathematics **80**(2), 653–654 (2009)
44. Oseledets, I.V.: Approximation of $2^d \times 2^d$ matrices using tensor decomposition. SIAM Journal on Matrix Analysis and Applications **31**(4), 2130–2145 (2010)
45. Oseledets, I.V.: Tensor Train decomposition. SIAM Journal on Scientific Computing **33**(5), 2295–2317 (2011)
46. Oseledets, I.V., Rakhuba, M.V., Chertkov, A.V.: Black-box solver for multiscale modelling using the QTT format. In: Proc. ECCOMAS. Crete Island, Greece (2016)
47. Oseledets, I.V., Tyrtyshnikov, E.E.: Breaking the curse of dimensionality, or how to use SVD in many dimensions. SIAM Journal on Scientific Computing **31**(5), 3744–3759 (2009)
48. Oswald, P.: On discrete norm estimates related to multilevel preconditioners in the finite element method. In: Constructive Theory of Functions, Proc. Int. Conf. Varna, 1991, pp. 203–214. Bulg. Acad. Sci., Sofia (1992)
49. Schollwöck, U.: The density-matrix renormalization group in the age of matrix product states. Annals of Physics **326**(1), 96–192 (2011). January 2011 Special Issue
50. Uschmajew, A., Vandereycken, B.: The geometry of algorithms using hierarchical tensors. Linear Algebra and its Applications **439**(1), 133–166 (2013)
51. Vassilevski, P.S., Wang, J.: Stabilizing the hierarchical basis by approximate wavelets. I. Theory. Numer. Linear Algebra Appl. **4**(2), 103–126 (1997)

52. Yserentant, H.: On the multilevel splitting of finite element spaces. Numer. Math. **49**(4), 379–412 (1986)
53. Yserentant, H.: Two preconditioners based on the multi-level splitting of finite element spaces. Numer. Math. **58**(2), 163–184 (1990)
54. Zhang, X.: Multilevel Schwarz methods. Numer. Math. **63**(4), 521–539 (1992)

Affiliations

Markus Bachmayr[1] · Vladimir Kazeev[2]

✉ Markus Bachmayr
bachmayr@uni-mainz.de

Vladimir Kazeev
vladimir.kazeev@univie.ac.at

1 Institut für Mathematik, Johannes Gutenberg-Universität Mainz, Staudingerweg 9, 55128 Mainz, Germany

2 Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria