Electronic Transactions on Numerical Analysis.
Volume 62, pp. 188–207, 2024.

ISSN 1068–9613.
DOI: 10.1553/etna_vol62s188

ETNA
Kent State University and
Johann Radon Institute (RICAM)

OPERATOR-DEPENDENT PROLONGATION AND RESTRICTION FOR THE PARAMETER-DEPENDENT MULTIGRID METHOD USING LOW-RANK TENSOR FORMATS*

LARS GRASEDYCK† AND TIM A. WERTHMANN†

Abstract. Iterative solution methods, such as the parameter-dependent multigrid method, solve linear systems arising from partial differential equations that depend on parameters. When parameters introduce non-smooth dependencies, such as jumping coefficients, the convergence of the parameter-dependent multigrid method declines or even results in divergence. The goal is to enhance robustness of this multigrid method to enable effective solutions of problems involving jumping coefficients.

An operator-dependent prolongation and restriction, inspired by block Gaussian elimination, is derived that fulfills the approximation property under exact arithmetic, thereby enhancing the method's robustness. Using an approximation of this prolongation and restriction directly in a low-rank tensor format is a trade-off between computational cost and guaranteed convergence. Numerical experiments provide empirical support for the effectiveness of the method, even when using a lower accuracy to compute the approximation within the low-rank format. The proposed operator-dependent prolongation and restriction improves the convergence of the parameter-dependent multigrid method in the presence of jumping coefficients.

Key words. multigrid, transfer operators, partial differential equations, iterative solvers, low-rank tensor format, jumping coefficients

AMS subject classifications. 65N55, 15A69

1. Introduction.

Modern scientific problems are frequently addressed through the formulation of partial differential equations (PDEs), where it becomes mandatory to incorporate dependencies on parameters or uncertainties. In many instances, the consideration of such dependencies enhances the model's ability to reflect real-world complexities. This motivates us to focus on solving parameter-dependent linear systems, exemplified through the exploration of the following well-established model PDE:

$$\begin{aligned} -\nabla \cdot (\sigma(x,p)\nabla u(x,p)) &= f(x), \quad x \in \Omega, \\ u(x,p) &= 0, \qquad x \in \partial\Omega. \end{aligned} \tag{1.1}$$

In (1.1), the diffusion $\sigma(x,p)$ depends on some parameters $p \coloneqq (p^{(1)}, \dots, p^{(d)})$, where each $p^{(\nu)}$ takes only n_ν discrete values, i.e., $p^{(\nu)} \in \{p^{(\nu)}(1), p^{(\nu)}(2), \dots, p^{(\nu)}(n_\nu)\} =: \mathcal{P}^{(\nu)}$ for all $\nu = 1, \dots, d$. After employing a suitable discretization (e.g., a finite-element method (FEM) discretization, even though our arguments are invariant to the discretization method chosen), we obtain a linear system

$$A(p)u(p) = f \tag{1.2}$$

for each combination of parameters $p \in \bigtimes_{\nu=1}^{d} \mathcal{P}^{(\nu)} =: \mathcal{P}$. As there are n_ν different, discrete values for every $p^{(\nu)}$, applying classical methods means that we need to solve $\prod_{\nu=1}^{d} n_\nu \approx n^d$ linear systems with $n \coloneqq \max n_\nu$. The corresponding exponential scaling in the arithmetic and storage costs renders classical methods infeasible.

Low-rank tensor formats are one way to represent the operator and the right-hand side of (1.2) in a data-sparse way. This allows us to represent all n^d linear systems simultaneously

*Received January 30, 2024. Accepted January 23, 2025. Published online on February 13, 2025. Recommended by Peter Benner. The authors have been supported by the German Research Foundation (DFG) within the SFB 1481 (INST 222/1473-1 subproject B02).

†Institut für Geometrie und Praktische Mathematik, RWTH Aachen University, Templergraben 55, 52056 Aachen, Germany ({lgr,werthmann}@igpm.rwth-aachen.de). Corresponding author: T. A. Werthmann.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

with a complexity linear in d. Further, in our manuscript [16], we used the multigrid method to solve linear systems inside low-rank tensor formats and derived an approximate Jacobi smoother. For continuous $\sigma(x,p)$, the standard linear prolongation and restriction was used to achieve fast multigrid convergence for parameter-dependent problems.

We focus on jumping parameters, such as random or high-contrast variations. For instance, the function $\sigma(x,p)$ is discontinuous at internal boundaries within the domain Ω. Following [11], as a consequence, $\nabla u(x,p)$ is discontinuous, too. Using standard linear interpolation for the prolongation and restriction is inaccurate and the convergence rate of the multigrid method declines. Therefore, the question arises how to deal with these jumping parameters in the parameter-dependent multigrid method.

To answer this question, in Section 2, we summarize the parameter-dependent multigrid theory, including parameter-dependent smoothers and linear interpolation. Then, in Section 3, we address data-sparse representation in low-rank tensor formats and introduce exponential sums crucial for approximating the inverse of a low-rank tensor operator. This sets the stage for our main problem in Section 4: Addressing jumping coefficients in (1.1) and explaining why linear interpolation falls short in such cases. In Section 5, we motivate the derivation of an operator-dependent prolongation and restriction based on block Gaussian elimination.

In Section 6, the first part of our main contribution is the derivation of an operator-dependent prolongation and restriction independent of the chosen data-sparse representation, for which we prove the approximation property. The second part of our main contribution is a low-rank tensor format representation of this operator-dependent prolongation and restriction, which we derive using an approximation based on exponential sums.

Using our operator-dependent prolongation and restriction yields a fast convergence of the low-rank tensor format multigrid method for discontinuous parameters. Numerical experiments, in Section 7, for one- and two-dimensional geometries with three parameters, support our theoretical results. Section 8 places our results within the broader context of related work, and the conclusion is presented in Section 9.

2. The parameter-dependent multigrid method. The subsequent section draws, for ease of presentation, upon our previous discussion of multigrid methods in [16]. The multigrid method is a solver for linear systems, with a complexity that often scales linearly or quasi-linearly in the grid size. This makes it particularly effective for handling large linear systems of equations. The fundamental concept is to obtain a smooth approximation for the error of a given estimate for the fine grid solution on a coarser grid. By solving the linear system on coarser grids with fewer computations, this recursive approach to error correction defines the essence of the multigrid method. Consequently, the method solves the system on a sequence of grids, progressing from a coarse grid to progressively finer grids. At each level of the hierarchy, a combination of relaxation techniques (smoothers) and grid transfers (prolongation and restriction) is employed to iteratively refine a solution. Thus, the method smoothes errors on a fine grid and corrects them on coarse grids, enabling rapid convergence to an accurate solution.

In our manuscript [16], we transferred the parameter-independent case of the convergence theory of this iterative method to the case of parameter-dependent problems. Our results are based on the matrix theory established by Hackbusch and others; for a comprehensive overview of the matrix results, see [19]. To summarize the main results, we define the following notation.

Let X_ℓ denote the grid of level $\ell \in \mathbb{N}$ with grid size $h_\ell > 0$, and let $A_\ell(p)$ denote the parameter-dependent matrix corresponding to this grid. Analogously, let us denote by $X_{\ell-1}$ the next coarser grid with corresponding matrix $A_{\ell-1}(p)$. Further, let $P_\ell(p) : X_{\ell-1} \to X_\ell$ denote a parameter-dependent prolongation and $R_\ell(p) : X_\ell \to X_{\ell-1}$ a parameter-dependent restriction, as well as $S_\ell(p)$ a parameter-dependent iteration matrix of the smoother corre-

ETNA
Kent State University and
Johann Radon Institute (RICAM)

sponding to the grid of level ℓ. Then the parameter-dependent iteration matrix of the two-grid method (TGM) including $\nu_1 \in \mathbb{N}$ pre-smoothing steps and $\nu_2 \in \mathbb{N}_0$ post-smoothing steps is given by

$$M_\ell^{\mathrm{TGM}(\nu_1,\nu_2)}(p) = S_\ell^{\nu_2}(p)\big[\mathrm{Id} - P_\ell(p)(A_{\ell-1}(p))^{-1}R_\ell(p)A_\ell(p)\big]S_\ell^{\nu_1}(p). \tag{2.1}$$

For convenience, we will choose $\nu_1 = \nu$ and $\nu_2 = 0$ in the following theoretical analysis of the method. We also assume that $A_\ell(p)$ is symmetric positive definite for all $p \in \mathcal{P}$.

The convergence analysis of Hackbusch is based on the splitting of the iteration matrix from (2.1) into two factors:

$$\|M_\ell^{\mathrm{TGM}(\nu,0)}(p)\|_2 \le \|(A_\ell(p))^{-1} - P_\ell(p)(A_{\ell-1}(p))^{-1}R_\ell(p)\|_2 \, \|A_\ell(p)S_\ell^{\nu}(p)\|_2. \tag{2.2}$$

The first factor on the right-hand side of (2.2) reflects the grid transfer process. Consequently, our objective is to ensure that the error on the coarse grid serves as a reliable approximation of the error on the fine grid. This property is referred to as the *approximation property*.

DEFINITION 2.1 (Approximation property). *The* approximation property *is given by*

$$\|(A_\ell(p))^{-1} - P_\ell(p)(A_{\ell-1}(p))^{-1}R_\ell(p)\|_2 \le \frac{C_A}{\|A_\ell(p)\|_2}$$

for all $\ell \in \mathbb{N}$ and $p \in \mathcal{P}$, with a constant $C_A > 0$ independent of ℓ and uniform for all $p \in \mathcal{P}$.

In the case of continuous coefficients, one can use a linear interpolation as grid transfer, e.g., for a one-dimensional geometry $\Omega \subseteq \mathbb{R}$ such an interpolation is given by the stencil $\big[0.5 \quad 1 \quad 0.5\big]$. This has the advantage that it can, in the case of continuous coefficients, be used for all parameters and is therefore parameter-independent. Our main contribution will be to derive a parameter-dependent prolongation that fulfills the approximation property even for more complex problems, like discontinuous coefficients.

The second factor on the right-hand side of (2.2) corresponds to the relaxation techniques. Therefore, our aim is to guarantee that the approximation of the fine grid error is sufficiently smooth to be approximated on the coarser grid. This property is denoted as the *smoothing property*.

DEFINITION 2.2 (Smoothing property). *For an iteration with parameter-dependent iteration matrix $S_\ell(p)$, the* smoothing property *is defined as*

$$\|A_\ell(p)S_\ell^{\nu}(p)\|_2 \le \eta(\nu)\|A_\ell(p)\|_2$$

for all $0 \le \nu \le \overline{\nu}(h_\ell)$, $p \in \mathcal{P}$, and $\ell \in \mathbb{N}_0$, with functions η and $\overline{\nu}$ satisfying

$$\lim_{\nu\to\infty} \eta(\nu) = 0, \quad \lim_{h\to 0} \overline{\nu}(h) = \infty, \quad \text{or} \quad \overline{\nu}(h) = \infty$$

independent of the level ℓ and uniform for all $p \in \mathcal{P}$.

In our manuscript [16], we transferred the smoothing property of the parameter-independent method to the damped parameter-dependent Jacobi method, defined as $S_\ell(p) \coloneqq \mathrm{Id} - \omega(p)(D_\ell(p))^{-1}A_\ell(p)$, and presented an approximation of this smoother using exponential sums within low-rank tensor formats to approximate the parameter-dependent inverse diagonal $(D_\ell(p))^{-1} \coloneqq (\mathrm{diag}(A_\ell(p)))^{-1}$. Section 3 provides a brief introduction to exponential sums and their underlying concept, as we rely on their ability to approximate a tensor inverse within the low-rank format. As the error of the exponential sums approximation decays exponentially fast, for a sufficient accuracy the smoothing property therefore holds when using an approximation by few summands.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

By combining the smoothing and approximation properties, one directly obtains the convergence of the two-grid method.

THEOREM 2.3. *Assume that the smoothing property of Definition 2.2 with* $\overline{\nu}(h) = \infty$ *and the approximation property of Definition 2.1 are fulfilled. For a given* $0 < \zeta < 1$*, there exists a lower bound* $\underline{\nu} \in \mathbb{N}_0$*, such that*

$$\|M_\ell^{\mathrm{TGM}(\nu,0)}(p)\|_2 \leq C_A \eta(\nu) \leq \zeta$$

holds for all $\nu \geq \underline{\nu}$*,* $\ell \in \mathbb{N}$*, and* $p \in \mathcal{P}$*.*

This theory then leads to Algorithm 1, where the parameter-dependent multigrid pseudo-code is given for ν_1 pre-, ν_2 post-smoothing steps, and γ coarse grid corrections. We call the multigrid method with $\gamma = 1$ the V-cycle, and for $\gamma = 2$ the W-cycle.

Algorithm 1 $u_\ell(p) \leftarrow \mathrm{multigrid}(u_\ell(p), f_\ell(p), \ell)$

if $\ell = 0$ **then**
  $u_\ell(p) \leftarrow$ solve $A_\ell(p)u_\ell(p) = f_\ell(p)$
else
  $u_\ell(p) \leftarrow S_\ell^{\nu_1}(p)(u_\ell(p), f_\ell(p))$
  $d_{\ell-1}(p) \leftarrow R_\ell(p)(f_\ell(p) - A_\ell(p)u_\ell(p))$
  $e_{\ell-1}(p) \leftarrow 0$
  for 1 **to** γ **do**
    $e_{\ell-1}(p) \leftarrow \mathrm{multigrid}(e_{\ell-1}(p), d_{\ell-1}(p), \ell - 1)$
  end for
  $u_\ell(p) \leftarrow u_\ell(p) + P_\ell(p)e_{\ell-1}(p)$
  $u_\ell(p) \leftarrow S_\ell^{\nu_2}(p)(u_\ell(p), f_\ell(p))$
end if

Algorithm 1 outlines the process. We start by smoothing the current solution approximation $u_\ell(p)$. Next, we restrict the current residual to a coarser grid, recursively applying the method. Afterward, we prolongate the information from the coarse grid back to the current fine grid and perform smoothing again. On the coarsest level $\ell = 0$, a method is needed to solve the linear system, for which we use the chosen smoother $S_\ell(p)$.

This underlines that we need to represent the smoother, the operator, the right-hand side, the prolongation, and the restriction in a parameter-dependent way. To achieve this, we use low-rank tensor formats.

3. Low-rank tensor formats. We revisit low-rank formats, based on [16, 31], as they are instrumental in capturing the parameter dependency.

We consider that the operator, as expressed in (1.2), has an affine-linear form, i.e., $A(p) \coloneqq A^{(0)} + \sum_{k=1}^{d} p^{(k)} A^{(k)}$. Utilizing the Kronecker product to represent this operator in a data-sparse way, we obtain

$$\begin{aligned}\mathcal{A} &= \mathrm{Id} \otimes \cdots \otimes \mathrm{Id} \otimes A^{(0)} + \mathrm{Id} \otimes \cdots \otimes \mathrm{Id} \otimes \mathrm{diag}(p_1^{(1)}, \ldots, p_{n_1}^{(1)}) \otimes A^{(1)} \\ &\quad + \cdots + \mathrm{diag}(p_1^{(d)}, \ldots, p_{n_d}^{(d)}) \otimes \cdots \otimes \mathrm{Id} \otimes \mathrm{Id} \otimes A^{(d)} \\ &= \sum_{k=0}^{d} \bigotimes_{\nu=0}^{d} A_\nu^k \quad \text{with } A_\nu^k = \begin{cases} A^{(k)}, & \text{if } \nu = d, \\ \mathrm{diag}(p_1^{(\nu)}, \ldots, p_{n_\nu}^{(\nu)}), & \text{if } \nu + k = d \text{ and } \nu \neq 0, \\ \mathrm{Id}, & \text{otherwise.} \end{cases}\end{aligned} \tag{3.1}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

This data-sparse representation reduces the storage cost from $\mathcal{O}(n^d)$ to $\mathcal{O}(ndr)$, employing the so-called CANDECOMP/PARAFAC (CP) decomposition, introduced in [8, 25].

DEFINITION 3.1 (CP decomposition). *A* CP representation *of a tensor* $\mathcal{B} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$, *with* representation rank $r \in \mathbb{N}_0$, *is defined as*

$$\mathcal{B} = \sum_{k=1}^{r} \bigotimes_{\ell=1}^{d} b_k^{(\ell)} \quad \text{with } b_k^{(\ell)} \in \mathbb{R}^{n_\ell}, \tag{3.2}$$

where each $\ell \in \mathcal{D} := \{1, \dots, d\}$ *is termed a* mode, *and* d *is the* dimension. *The minimal* r, *making* (3.2) *hold, is referred to as the* CP rank *of* $\mathcal{B}$. *In this case,* (3.2) *is called the* CP decomposition *of* $\mathcal{B}$. *We call a tensor of the form* (3.2) *a* CP vector. *Tensors of the form* $\bigotimes_{\ell=1}^{d} b^{(\ell)} \neq 0$, *i.e., of rank* 1, *are called* elementary tensors. *A CP representation of a linear tensor operator* $\mathcal{A}$ *from* $\mathbb{R}^{n_1 \times \cdots \times n_d}$ *to* $\mathbb{R}^{n_1 \times \cdots \times n_d}$, *with representation rank* r *and dimension* d, *is defined as*

$$\mathcal{A} = \sum_{k=1}^{r} \bigotimes_{\ell=1}^{d} A_k^{(\ell)} \quad \text{with } A_k^{(\ell)} \in \mathbb{R}^{n_\ell \times n_\ell}. \tag{3.3}$$

We call a tensor of the form (3.3) *a* CP operator.

It is important to note that $b_k^{(\ell)}$ in (3.2) are vectors and the $A_k^{(\ell)}$ in (3.3) are matrices.

Using the CP format, one can derive a data-sparse representation of the operator and the right-hand side of (1.2). Therefore, we are capable of representing the operator and right-hand side in a parameter-dependent way.

To employ the parameter-dependent multigrid method for solving, it is essential to express the smoother in a data-sparse form. However, directly finding an inverse operator within the CP format poses a challenging problem, and deriving a smoother for the multigrid method becomes a complex task.

An effective approach for achieving this involves employing exponential sums. Given appropriate assumptions, we can express $1/x = \int_0^\infty \exp(-tx)\,\mathrm{d}t$, and through a suitable numerical integration formula, we can derive $1/x \approx \sum_{k=1}^{r} \alpha_k \exp(-t_k x)$. Employing this method allows for variable separation, resulting in an approximate direct inverse in the CP format:

$$\frac{1}{x+y} \approx \sum_{k=1}^{r} \alpha_k \exp(-t_k(x+y)) = \sum_{k=1}^{r} \alpha_k \exp(-t_k x) \cdot \exp(-t_k y),$$

and for commuting operators A and B:

$$(A+B)^{-1} \approx \sum_{k=1}^{r} \alpha_k \exp(-t_k(A+B)) = \sum_{k=1}^{r} \alpha_k \exp(-t_k A) \otimes \exp(-t_k B).$$

It is important to note that the error of the exponential sums approximation can be bounded based on r; see [22]. However, the stringent assumptions required for applying exponential sums make their use impractical for arbitrary operators in the CP format. In [16], we developed a method to apply exponential sums to the diagonal of affine-linear operators in the form of (1.2). This allowed us to derive an approximate Jacobi smoother directly in the CP format, with a proven smoothing property.

The task of finding a prolongation and restriction that fulfills the approximation property and can be represented within a low-rank format is challenging. In [16], we opted for the

ETNA
Kent State University and
Johann Radon Institute (RICAM)

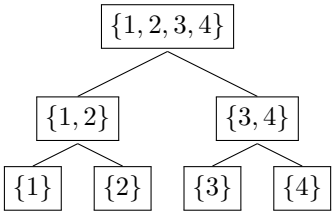

FIG. 3.1. *A dimension tree for dimension $d = 4$.*

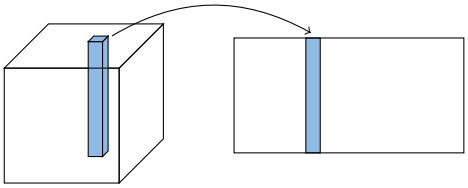

FIG. 3.2. *Visual representation of a matricization for dimension $d = 3$.*

standard linear prolongation and restriction, represented in the CP format by $\mathrm{Id} \otimes \cdots \otimes \mathrm{Id} \otimes P$. For problems with smooth σ in (1.2), we observed for this parameter-independent prolongation and restriction a rapid multigrid convergence. The derivation of an operator-dependent and parameter-dependent prolongation and restriction is our main contribution.

Since we have to perform arithmetic operations in the CP format, an issue arises due to the resulting rank growth. For example, the addition of CP tensors results in a CP tensor whose rank is the sum of the ranks of the tensors being added. Given computational constraints that limit the acceptable approximation rank, we need a method to compute an approximation with lower rank. Further, to ensure the convergence of the multigrid method, or any other iterative solver, this approximation needs to be error-controlled; cf. [23]. The complication arises from the fact that the set of the CP format with rank r is not closed [10]. Consequently, obtaining an error-controlled approximation with lower rank is an ill-posed problem in the CP format; cf. [10]. To address this, we will briefly introduce the *hierarchical Tucker* format; proposed in [24] and further analyzed in [14].

The idea of the hierarchical Tucker format is to establish a hierarchy among the modes $\mathcal{D} = \{1, \ldots, d\}$. This is achieved by defining a *dimension tree*; analogously to [14, Definition 3.3]. An example for $d = 4$ is shown in Figure 3.1.

DEFINITION 3.2 (Dimension tree). *A* dimension tree $\mathcal{T}$ *for dimension $d \in \mathbb{N}$ is represented as a binary tree. Nodes within this tree are labeled with non-empty subsets of $\mathcal{D}$. The root node is labeled with $\mathcal{D}$, each leaf node is labeled with a single-element subset $z = \{\ell\} \subseteq \mathcal{D}$, and each inner node is labeled with the disjoint union of its two children. We will identify a node with its label z and hence write $z \in \mathcal{T}$.*

The labels within dimension trees give rise to the associated *matricization*, a concept that we define as in [14, Definition 3.3].

DEFINITION 3.3 (Matricization and vectorization). *Let $\mathcal{B} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$, $z \subseteq \mathcal{D}$ with $z \neq \emptyset$, and $g \coloneqq \mathcal{D} \setminus z$. The* matricization *of $\mathcal{B}$ corresponding to z is defined as $\mathcal{B}^{(z)} \in \mathbb{R}^{n_z \times n_g}$, where $n_z \coloneqq \prod_{\ell \in z} n_\ell$ and $n_g \coloneqq \prod_{\ell \in g} n_\ell$, with $\mathcal{B}^{(z)}[(i_j)_{j \in z}, (i_j)_{j \in g}] \coloneqq \mathcal{B}[i_1, \ldots, i_d]$ for all $i = (i_j)_{j \in \mathcal{D}}$. In particular, $\mathcal{B}^{(\mathcal{D})} \in \mathbb{R}^{n_1 \cdots n_d}$ holds, which can also be interpreted as the* vectorization *of $\mathcal{B}$.*

A matricization can be visualized as an unfolding of the tensor as illustrated in Figure 3.2.

Drawing upon the notion of matricizations, the *hierarchical Tucker rank* is defined in accordance with [14, Definition 3.4].

DEFINITION 3.4 (Hierarchical Tucker rank). *Let $\mathcal{B} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ and $\mathcal{T}$ be a dimension tree. The* hierarchical Tucker rank *of $\mathcal{B}$ is defined as $\mathrm{rank}_{\mathcal{T}}(\mathcal{B}) \coloneqq (r_z)_{z \in \mathcal{T}}$, where $r_z \coloneqq \mathrm{rank}(\mathcal{B}^{(z)})$ denotes the matrix rank of the matricization $\mathcal{B}^{(z)}$ for all $z \in \mathcal{T}$.*

The set of tensors with hierarchical Tucker rank node-wise bounded by $(r_z)_{z \in \mathcal{T}}$ is defined as $\mathcal{H}$-Tucker$(\mathcal{T}, (r_z)_{z \in \mathcal{T}}) \coloneqq \{\mathcal{F} \in \mathbb{R}^{n_1 \times \cdots \times n_d} \mid \mathrm{rank}(\mathcal{F}^{(z)}) \leq r_z \text{ for all } z \in \mathcal{T}\}$.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

TABLE 3.1
Operations and their costs in the hierarchical Tucker format, where d is the dimension, n is the mode size, and r is the rank.

Operation	Cost	Reference
Storage	$\mathcal{O}(dr^3 + dnr)$	[14, Lemma 3.7]
Orthonormalization	$\mathcal{O}(dnr^2 + dr^4)$	[20, (13.16b)]
Addition	$\mathcal{O}(dnr^2 + dr^4)$	[20, 13.1.4]
Evaluation	$\mathcal{O}(dr^3)$	[20, 13.2.3]
Inner product/Norm	$\mathcal{O}(dnr^2 + dr^4)$	[20, Lemma 13.7]
Operator application	$\mathcal{O}(dn^2r)$	[20, 13.9.1]
Truncation	$\mathcal{O}(dr^2n + dr^4)$	[20, (11.46c)]
Maximum	$\mathcal{O}(d^2 \log_2(n) nr^2 + d^2 \log_2(n) r^6)$	[15, (7.4)]

Using the dimension tree, the concept of matricization, and the hierarchical Tucker rank, one can define the representation of a tensor within the hierarchical Tucker format; cf. [14, Definition 3.6]. The memory required for a hierarchical Tucker representation, with dimension tree $\mathcal{T}$ and representation rank $(r_z)_{z\in\mathcal{T}}$, of a tensor $\mathcal{B} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ for $n = \max_{\ell\in\mathcal{D}} n_\ell$ and $r = \max_{z\in\mathcal{T}} r_z$ is given by $\mathcal{O}(rdn + r^3d)$; cf. [14, Lemma 3.7]. The existence of a truncation method of a low-rank tensor $\mathcal{B} \in \mathcal{H}\text{-Tucker}(\mathcal{T}, (r_z)_{z\in\mathcal{T}})$ down to lower rank $(\tilde{r}_z)_{z\in\mathcal{T}}$ with an arithmetic cost in $\mathcal{O}(r^2dn+r^4d)$ was proven in [14]. The resulting approximation $\tilde{\mathcal{B}} := \operatorname{truncate}(\mathcal{B}) \in \mathcal{H}\text{-Tucker}(\mathcal{T}, (\tilde{r}_z)_{z\in\mathcal{T}})$ fulfills the quasi-optimal error estimation

$$\|\mathcal{B} - \tilde{\mathcal{B}}\| \le \sqrt{2d-3} \inf_{\mathcal{F}\in\mathcal{H}\text{-Tucker}(\mathcal{T},(\tilde{r}_z)_{z\in\mathcal{T}})} \|\mathcal{B} - \mathcal{F}\|.$$

Moreover, we can transfer a CP representation of a tensor with CP rank r into a hierarchical Tucker representation with rank node-wise bounded by r; cf. [20, Theorem 11.17]. Subsequently, we can express both the operator and the right-hand side in the hierarchical Tucker format. A summary of several arithmetic operations feasible within the hierarchical Tucker format is presented in Table 3.1.

Utilizing these arithmetic operators we can now numerically approximate the solution of a parameter-dependent linear system within the hierarchical Tucker format. This is applicable when the tensor operator $\mathcal{A}$ is positive definite and symmetric, employing the multigrid method outlined in Algorithm 1. An observation from Table 3.1 reveals that, for small ranks r, the majority of operations required for the multigrid method exhibit linear scaling with both the dimension d and mode size n. Consequently, this presents an efficient approach for solving parameter-dependent linear systems using low-rank tensor formats.

4. Motivation for operator-dependent prolongation and restriction. In this section, we motivate the challenges posed by discontinuous coefficients, even in the context of parameter-independent problems. To illustrate this, we use an insightful example from the literature [9, 11]. Let u_ℓ be the approximate solution of $A_\ell u_\ell = f_\ell$ and u be the continuous solution. We consider the error equation $r_\ell = A_\ell e_\ell$, where $r_\ell = f_\ell - A_\ell u_\ell$ is the residual and $e_\ell = u - u_\ell$ is the corresponding error.

After applying the smoothing step from the multigrid method, the residual r_ℓ is also smoothed. The key difference when dealing with a smooth σ versus a non-smooth σ now occurs:

- If σ is continuous, the error e_ℓ is smoothed and thus well approximated by standard linear interpolation of a coarse grid function, as visualized in Figure 4.1. So in this case, the multigrid method works as expected also using low-rank formats.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

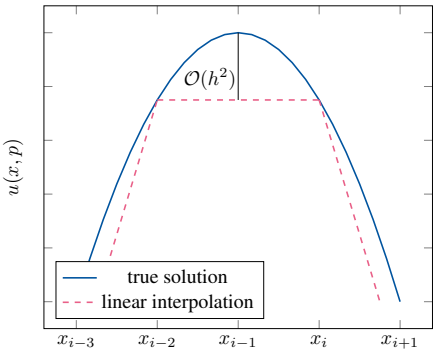

FIG. 4.1. *Using linear interpolation on a function with continuous gradient yields an error of* $\mathcal{O}(h^2)$.

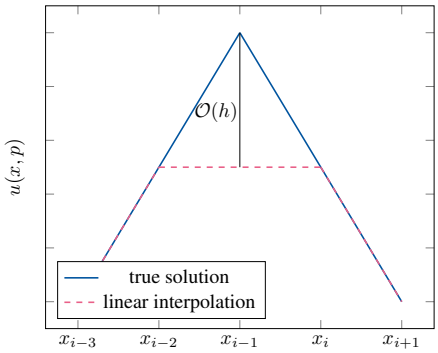

FIG. 4.2. *Using linear interpolation on a function with discontinuous gradient can yield an error of* $\mathcal{O}(h)$ *instead of* $\mathcal{O}(h^2)$.

- However, near a discontinuity point of the diffusion coefficient, a smooth residual r_ℓ corresponds to an error e_ℓ with discontinuous gradient. For this reason, the canonical linear interpolation of a coarse grid function no longer adequately captures the error, resulting in a reduction in approximation accuracy. As a result, the error induced by linear interpolation is of $\mathcal{O}(h)$ instead of $\mathcal{O}(h^2)$, as depicted in Figure 4.2.

Consequently, in the case of discontinuous coefficients, linear interpolation fails to provide a sufficiently accurate approximation of the function on the fine grid. This leads to a decrease in the convergence rate of the multigrid method or even divergence; cf. [1]. So we need alternative prolongations and restrictions.

To address this problem in the classical parameter-independent multigrid method, a vast number of operator-dependent prolongation and restrictions were employed as seen, e.g., in [1, 12, 26]. However, these approaches often rely on fixed parameter combinations and the local entry-wise combination of operator entries. Implementing such a strategy for all parameter combinations independently of each other is infeasible because of the exponential scaling in the arithmetic and storage cost.

Therefore, as a first problem, we need an ansatz capable of computing an operator-dependent prolongation and restriction simultaneously for all parameters. Additionally, these prolongations and restrictions should fulfill the approximation property for all $p \in \mathcal{P}$. As a second problem, we also need a method to compute and represent these prolongations and restrictions in a parameter-dependent way using low-rank formats.

5. A block Gaussian idea. In this section, we discuss the derivation of a general concept of an operator-dependent prolongation and restriction that fulfills the approximation property. We shall transition from a geometric perspective of the multigrid method to an algebraic point of view. This shift in perspective will result in several modifications, including a different derivation of the coarse grid operator as a Schur complement. This derivation will be conducted within the tensor format, which facilitates the computation of an approximation. The objective is to derive a prolongation and restriction from the algebraic approach for arbitrary parameter-dependent problems. To facilitate the computations, it is necessary to approximate them within the low-rank tensor format.

We adapt, for ease of presentation, the motivation of our idea from the matrix case, presented in [28], to the parameter-dependent case. Consider the parameter-dependent linear system at level $\ell \in \mathbb{N}$,

$$A_\ell(p)u_\ell(p) = f_\ell. \tag{5.1}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

This system is given for all p in the parameter space $\mathcal{P}$.

Let $u_\ell^{\mathrm{F}}(p)$ represent the unknowns on $\Omega_\ell \setminus \Omega_{\ell-1}$, i.e., the unknowns that *only* live on the fine grid. Further, let $u_\ell^{\mathrm{C}}(p)$ denote the unknowns on the next coarser grid $\Omega_{\ell-1}$. Furthermore, we partition both the operator and the right-hand side based on the distinction between fine and coarse grids. Here, this partitioning relies entirely on the geometry. Further, this article does not address or require the consideration of algebraic partitioning, as encountered in algebraic multigrid methods. In our context, all the necessary geometric information is readily available as the parameters do not change the geometry. We define $K_\ell(p) \coloneqq \mathrm{Id} - P_\ell(p) A_{\ell-1}^{-1}(p) R_\ell(p) A_\ell(p)$ as the coarse-level correction operator.

The idea is to choose the prolongation and restriction in such a way that the multigrid method degenerates into a direct solver, e.g., it holds that $K_\ell(p) S_\ell^\nu(p) = 0$ or that $S_\ell^\nu(p) K_\ell(p) = 0$. To achieve this, we aim to choose the prolongation and restriction such that we separate the equations for $u_\ell^{\mathrm{F}}(p)$ and $u_\ell^{\mathrm{C}}(p)$ using block Gaussian elimination. To do so, we rewrite (5.1) in the following block form:

$$\begin{bmatrix} A_\ell^{\mathrm{FF}}(p) & A_\ell^{\mathrm{FC}}(p) \\ A_\ell^{\mathrm{CF}}(p) & A_\ell^{\mathrm{CC}}(p) \end{bmatrix} \begin{bmatrix} u_\ell^{\mathrm{F}}(p) \\ u_\ell^{\mathrm{C}}(p) \end{bmatrix} = \begin{bmatrix} f_\ell^{\mathrm{F}} \\ f_\ell^{\mathrm{C}} \end{bmatrix}. \tag{5.2}$$

Then, using exact block Gaussian elimination, i.e.,

$$\begin{aligned} &\begin{bmatrix} \mathrm{Id} & -(A_\ell^{\mathrm{FF}}(p))^{-1} A_\ell^{\mathrm{FC}}(p) \\ 0 & \mathrm{Id} \end{bmatrix}^\top \begin{bmatrix} A_\ell^{\mathrm{FF}}(p) & A_\ell^{\mathrm{FC}}(p) \\ A_\ell^{\mathrm{CF}}(p) & A_\ell^{\mathrm{CC}}(p) \end{bmatrix} \begin{bmatrix} \mathrm{Id} & -(A_\ell^{\mathrm{FF}}(p))^{-1} A_\ell^{\mathrm{FC}}(p) \\ 0 & \mathrm{Id} \end{bmatrix} \\ &\quad = \begin{bmatrix} A_\ell^{\mathrm{FF}}(p) & 0 \\ 0 & A_\ell^{\mathrm{CC}}(p) - A_\ell^{\mathrm{CF}}(p)(A_\ell^{\mathrm{FF}}(p))^{-1} A_\ell^{\mathrm{FC}}(p) \end{bmatrix}, \end{aligned}$$

we get a separated representation of the operator. Therefore, our operator-dependent prolongation, which provides a separation of the fine grid and coarse grid, is given by

$$P_\ell(p) = \begin{bmatrix} -(A_\ell^{\mathrm{FF}}(p))^{-1} A_\ell^{\mathrm{FC}}(p) \\ \mathrm{Id} \end{bmatrix}. \tag{5.3}$$

One benefit of (5.3) is that it enables a highly accurate representation of the parameter-dependent prolongation in the low-rank format using exponential sums. We remark that the error using those exponential sums decays with an exponential rate for $\varepsilon \to 0$. Consequently, this algebraic perspective yields a low-rank prolongation that is well suited for discontinuous diffusion coefficients.

REMARK 5.1. It should be noted that utilizing this highly accurate representation has the potential to result in high ranks in the tensor format. To address this, a rougher approximation employing some truncation would be preferable. However, such an approach requires some additional modification of the operator, e.g., a scaling, to prove multigrid convergence; see, e.g., [30]. As this task is non-trivial within the tensor format, it is thus beyond the scope of this article.

If we envision a (maybe hypothetical) smoother designed to *exclusively* solve the fine grid equations for $u_\ell^{\mathrm{F}}(p)$, our above choice of the prolongation and restriction becomes optimal. This is because the remaining unknowns $u_\ell^{\mathrm{C}}(p)$ on the coarse grid are solved by $A_{\ell-1}^{-1}(p)$, and, due to the separation of the problem, there is no interference between these components. If $A_\ell^{-1}(p)$ and $A_{\ell-1}^{-1}(p)$ exist and it holds that

$$S_\ell^\nu(p) = \begin{bmatrix} 0 & P_\ell(p) \end{bmatrix} = \begin{bmatrix} 0 & P_\ell^{\mathrm{FC}}(p) \\ 0 & P_\ell^{\mathrm{CC}}(p) \end{bmatrix},$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

then $K_\ell(p)S_\ell^\nu(p) = 0$ as $K_\ell(p)P_\ell(p) = 0$, i.e., the resulting multigrid method is a direct solver.

If we then define $P_\ell^{\mathrm{FC}}(p) = -A_\ell^{\mathrm{FF}}(p)^{-1}A_\ell^{\mathrm{FC}}(p)$, i.e., define the prolongation by solving $A_\ell^{\mathrm{FF}}(p)e_\ell^{\mathrm{F}} + A_\ell^{\mathrm{FC}}(p)e_\ell^{\mathrm{C}} = 0$ for e_ℓ^{F}, then it holds that

$$S_\ell^\nu(p) = \begin{bmatrix} 0 & -(A_\ell^{\mathrm{FF}}(p))^{-1}A_\ell^{\mathrm{FC}}(p) \\ 0 & \mathrm{Id} \end{bmatrix}. \tag{5.4}$$

Using this $S_\ell^\nu(p)$ corresponds to solving the F-equations for $u_\ell^{\mathrm{F}}(p)$, i.e., $A_\ell^{\mathrm{FF}}(p)u_\ell^{\mathrm{F}}(p) + A_\ell^{\mathrm{FC}}(p)u_\ell^{\mathrm{C}}(p) = f_\ell^{\mathrm{F}}$, or in terms of the error,

$$e_\ell^{\mathrm{F}}(p) = -(A_\ell^{\mathrm{FF}}(p))^{-1}A_\ell^{\mathrm{FC}}(p)e_\ell^{\mathrm{C}}. \tag{5.5}$$

For this choice of $P_\ell^{\mathrm{FC}}(p)$ the Galerkin operator is the Schur complement, i.e.,

$$A_{\ell-1}(p) = P^\top(p)A_\ell(p)P(p) \coloneqq A_\ell^{\mathrm{CC}}(p) - A_\ell^{\mathrm{CF}}(p)(A_\ell^{\mathrm{FF}}(p))^{-1}A_\ell^{\mathrm{FC}}(p),$$

and is regular if $A_\ell(p)$ and $A_\ell^{\mathrm{FF}}(p)$ are regular.

LEMMA 5.2. *Let $A_\ell(p)$ and $A_\ell^{\mathrm{FF}}(p)$ be regular and let $S_\ell^\nu(p)$ be defined by (5.4). Then the following are true:*

- *If $P_\ell^{\mathrm{FC}}(p) = -(A_\ell^{\mathrm{FF}}(p))^{-1}A_\ell^{\mathrm{FC}}(p)$ then $A_{\ell-1}^{-1}(p)$ exists and $K_\ell(p)S_\ell^\nu(p) = 0$.*
- *The Galerkin operator is just the Schur complement corresponding to (5.2), i.e., $A_{\ell-1}(p) = A_\ell^{\mathrm{CC}}(p) - A_\ell^{\mathrm{CF}}(p)(A_\ell^{\mathrm{FF}}(p))^{-1}A_\ell^{\mathrm{FC}}(p)$.*

We observe that $R_\ell(p)A_\ell(p)e_\ell = A_{\ell-1}(p)e_\ell^{\mathrm{C}}$ holds for all $e_\ell = (e_\ell^{\mathrm{F}}, e_\ell^{\mathrm{C}})$. From this and $R_\ell(p)A_\ell(p)K_\ell(p) = 0$ it follows that the application of the coarse-level correction operator $K_\ell(p)$ yields error vectors e_ℓ with $e_\ell^{\mathrm{C}} = 0$ and therefore, due to (5.5), post-smoothing reduces the total error to 0. If we specify the prolongation and restriction following these ideas, we obtain a direct method.

As a next step, using the choice of the operator-dependent prolongation and restriction as outlined above, we prove that this fulfills the approximation property in case of exact arithmetic. Additionally, we must address our second problem: How to compute or approximate these operator-dependent prolongations and restrictions within low-rank formats.

6. Our operator-dependent prolongation and restriction. Following our motivation from Section 5, we define the following operator-dependent prolongation and restriction.

DEFINITION 6.1. *Let $A_\ell(p)u_\ell(p) = f_\ell$ be a parameter-dependent linear system such that*

$$\begin{bmatrix} A_\ell^{\mathrm{FF}}(p) & A_\ell^{\mathrm{FC}}(p) \\ A_\ell^{\mathrm{CF}}(p) & A_\ell^{\mathrm{CC}}(p) \end{bmatrix} \begin{bmatrix} u_\ell^{\mathrm{F}}(p) \\ u_\ell^{\mathrm{C}}(p) \end{bmatrix} = \begin{bmatrix} f_\ell^{\mathrm{F}} \\ f_\ell^{\mathrm{C}} \end{bmatrix}$$

holds, where $u_\ell^{\mathrm{F}}(p)$ represent the unknowns on $\Omega_\ell \setminus \Omega_{\ell-1}$ and $u_\ell^{\mathrm{C}}(p)$ denote the unknowns on the next coarser grid $\Omega_{\ell-1}$.

Then, we define the operator-dependent prolongation

$$P_\ell(p) = \begin{bmatrix} -(A_\ell^{\mathrm{FF}}(p))^{-1}A_\ell^{\mathrm{FC}}(p) \\ \mathrm{Id} \end{bmatrix}$$

and the operator-dependent restriction

$$R_\ell(p) = \begin{bmatrix} -A_\ell^{\mathrm{CF}}(p)(A_\ell^{\mathrm{FF}}(p))^{-\top} & \mathrm{Id} \end{bmatrix}.$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

We prove that the selection of the prolongation and restriction, as in Definition 6.1, fulfills the approximation property under the constraint of exact arithmetic. In the parameter-independent matrix case, this proof is known, but for ease of presentation we present a parameter-dependent version.

THEOREM 6.2. *Let $\sigma(x,p)$ in (1.1) be uniformly bounded from above and below for all $p \in \mathcal{P}$. Further, we assume $C_A \lambda_{\min}(A_\ell^{\mathrm{FF}}(p)) \geq \lambda_{\max}(A_\ell(p))$ holds for each, fixed $p \in \mathcal{P}$.*

Then, for the parameter-dependent and operator-dependent prolongation and restriction in Definition 6.1, the approximation property from Definition 2.1 holds.

Proof. In the proof, we refrain from explicitly stating the dependency on p.

We calculate for fixed, but arbitrary, p and ℓ:

$$
\begin{aligned}
&A_{\ell-1} \\
&\quad = R_\ell A_\ell P_\ell \\
&\quad = \begin{bmatrix} -(A_\ell^{\mathrm{FC}})^\top (A_\ell^{\mathrm{FF}})^{-\top} & \mathrm{Id} \end{bmatrix} \begin{bmatrix} A_\ell^{\mathrm{FF}} & A_\ell^{\mathrm{FC}} \\ A_\ell^{\mathrm{CF}} & A_\ell^{\mathrm{CC}} \end{bmatrix} \begin{bmatrix} -(A_\ell^{\mathrm{FF}})^{-1} A_\ell^{\mathrm{FC}} \\ \mathrm{Id} \end{bmatrix} \\
&\quad = \begin{bmatrix} -(A_\ell^{\mathrm{FC}})^\top (A_\ell^{\mathrm{FF}})^{-\top} A_\ell^{\mathrm{FF}} + A_\ell^{\mathrm{CF}} & -(A_\ell^{\mathrm{FC}})^\top (A_\ell^{\mathrm{FF}})^{-\top} A_\ell^{\mathrm{FC}} + A_\ell^{\mathrm{CC}} \end{bmatrix} \begin{bmatrix} -(A_\ell^{\mathrm{FF}})^{-1} A_\ell^{\mathrm{FC}} \\ \mathrm{Id} \end{bmatrix} \\
&\quad = ((A_\ell^{\mathrm{FC}})^\top (A_\ell^{\mathrm{FF}})^{-\top} A_\ell^{\mathrm{FF}} - A_\ell^{\mathrm{CF}})(A_\ell^{\mathrm{FF}})^{-1} A_\ell^{\mathrm{FC}} - (A_\ell^{\mathrm{FC}})^\top (A_\ell^{\mathrm{FF}})^{-\top} A_\ell^{\mathrm{FC}} + A_\ell^{\mathrm{CC}} \\
&\quad = A_\ell^{\mathrm{CC}} - A_\ell^{\mathrm{CF}} (A_\ell^{\mathrm{FF}})^{-1} A_\ell^{\mathrm{FC}} =: M.
\end{aligned}
$$

Using the Schur complement we calculate

$$
\begin{aligned}
A_\ell^{-1} &= \begin{bmatrix} A_\ell^{\mathrm{FF}} & A_\ell^{\mathrm{FC}} \\ A_\ell^{\mathrm{CF}} & A_\ell^{\mathrm{CC}} \end{bmatrix}^{-1} \\
&= \begin{bmatrix} (A_\ell^{\mathrm{FF}})^{-1} + (A_\ell^{\mathrm{FF}})^{-1} A_\ell^{\mathrm{FC}} M^{-1} A_\ell^{\mathrm{CF}} (A_\ell^{\mathrm{FF}})^{-1} & -(A_\ell^{\mathrm{FF}})^{-1} A_\ell^{\mathrm{FC}} M^{-1} \\ -M^{-1} A_\ell^{\mathrm{CF}} (A_\ell^{\mathrm{FF}})^{-1} & M^{-1} \end{bmatrix}
\end{aligned}
$$

Using $(A_\ell^{\mathrm{FF}})^{-1} = (A_\ell^{\mathrm{FF}})^{-\top}$ and $(A_\ell^{\mathrm{FC}})^\top = A_\ell^{\mathrm{CF}}$ we calculate

$$
\begin{aligned}
&\|A_\ell^{-1} - P_\ell A_{\ell-1}^{-1} R_\ell\|_2 \\
&\quad = \|A_\ell^{-1} - P_\ell (R_\ell A_\ell P_\ell)^{-1} R_\ell\|_2 \\
&\quad = \left\| A_\ell^{-1} - \begin{bmatrix} -(A_\ell^{\mathrm{FF}})^{-1} A_\ell^{\mathrm{FC}} \\ \mathrm{Id} \end{bmatrix} M^{-1} \begin{bmatrix} -(A_\ell^{\mathrm{FC}})^\top (A_\ell^{\mathrm{FF}})^{-\top} & \mathrm{Id} \end{bmatrix} \right\|_2 \\
&\quad = \left\| A_\ell^{-1} - \begin{bmatrix} (A_\ell^{\mathrm{FF}})^{-1} A_\ell^{\mathrm{FC}} M^{-1} (A_\ell^{\mathrm{FC}})^\top (A_\ell^{\mathrm{FF}})^{-\top} & -(A_\ell^{\mathrm{FF}})^{-1} A_\ell^{\mathrm{FC}} M^{-1} \\ -M^{-1} (A_\ell^{\mathrm{FC}})^\top (A_\ell^{\mathrm{FF}})^{-\top} & M^{-1} \end{bmatrix} \right\|_2 \\
&\quad = \left\| A_\ell^{-1} - \begin{bmatrix} (A_\ell^{\mathrm{FF}})^{-1} A_\ell^{\mathrm{FC}} M^{-1} A_\ell^{\mathrm{CF}} (A_\ell^{\mathrm{FF}})^{-1} & -(A_\ell^{\mathrm{FF}})^{-1} A_\ell^{\mathrm{FC}} M^{-1} \\ -M^{-1} A_\ell^{\mathrm{CF}} (A_\ell^{\mathrm{FF}})^{-1} & M^{-1} \end{bmatrix} \right\|_2 \\
&\quad = \left\| \begin{bmatrix} (A_\ell^{\mathrm{FF}})^{-1} & 0 \\ 0 & 0 \end{bmatrix} \right\|_2 \\
&\quad \leq \frac{C_A}{\|A_\ell\|_2}
\end{aligned}
$$

for a constant $C_A > 0$ independent of ℓ. As we choose p and ℓ to be arbitrary, the statement follows. ∎

ETNA
Kent State University and
Johann Radon Institute (RICAM)

For this reason, $P_\ell(p)$ as in (5.3) is an optimal prolongation that fulfills the approximation property.

The assumption $C_A \lambda_{\min}(A_\ell^{\mathrm{FF}}(p)) \geq \lambda_{\max}(A_\ell(p)) < \infty$ can be proven for Laplace-like problems as in our guiding problem (1.1) using, e.g., a suitable discretization such that the constant C_A depends only on the bounds of the diffusion coefficients, see [2].

REMARK 6.3. The choice to employ the prolongation and restriction as in Definition 6.1, coupled with the use of the Schur complement as the coarse grid matrix, introduces an inconsistency between the defined smoothing property and the employed smoothers. While our employed smoothers do indeed smooth the error, they do not align with Definition 2.2. Changing the definition for our smoother in the parameter-dependent way is beyond the scope of this work and merits further detailed investigation in future research.

This leads to our second problem: How to invert $A_\ell^{\mathrm{FF}}(p)$ or how to approximate the operator-dependent prolongation and restriction from Definition 6.1 within the low-rank format. To delve into this, we talk about different structured forms of $A_\ell^{\mathrm{FF}}(p)$ and $(A_\ell^{\mathrm{FF}}(p))^{-1}$.

Prolongation and restriction using low-rank formats. For a one-dimensional geometry $\Omega \subseteq \mathbb{R}$ in (1.1), the $A_\ell^{\mathrm{FF}}(p)$ part is a diagonal operator; see [28]. In the case of parameter-independent problems based on classical methods, inverting it is straightforward since only a diagonal matrix requires inversion. However, for parameter-dependent problems using low-rank formats, this becomes a non-trivial task. As demonstrated in [16], one can employ exponential sums to approximate the inverse. As exploiting the diagonal structure allows for the application of exponential sums within the CP format, this results in an error-controlled approximation of this part of the operator.

For more complex geometries Ω, such as the two-dimensional case in (1.1), the inversion of $A_\ell^{\mathrm{FF}}(p)$ becomes non-trivial even in the parameter-independent scenario. In the two-dimensional case $\Omega \subseteq \mathbb{R}^2$, the $A_\ell^{\mathrm{FF}}(p)$ part is no longer a diagonal operator, and its inversion could result in a full matrix. Consequently, various approximations of this optimal choice of the prolongation have been proposed even for the matrix case. For example in [27, 32], they suggested to use an approximation

$$\begin{bmatrix} \tilde{A}_\ell^{\mathrm{FF}}(p) & \tilde{A}_\ell^{\mathrm{FC}}(p) \\ \tilde{A}_\ell^{\mathrm{CF}}(p) & A_\ell^{\mathrm{CC}}(p) \end{bmatrix} \quad \text{for} \quad \begin{bmatrix} A_\ell^{\mathrm{FF}}(p) & A_\ell^{\mathrm{FC}}(p) \\ A_\ell^{\mathrm{CF}}(p) & A_\ell^{\mathrm{CC}}(p) \end{bmatrix}$$

such that:

- $\tilde{A}_\ell^{\mathrm{FF}}(p)$ is easy to invert,
- $-(\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} \tilde{A}_\ell^{\mathrm{FC}}(p)$ is an approximation of $-(A_\ell^{\mathrm{FF}}(p))^{-1} A_\ell^{\mathrm{FC}}(p)$, and
- $\begin{bmatrix} -(\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} \tilde{A}_\ell^{\mathrm{FC}}(p) & \mathrm{Id} \end{bmatrix}$ is a local, i.e., sparse, operator.

In the parameter-dependent scenario, the sparsity of the operator becomes less crucial for us, as we work with a data-sparse format. Instead, the emphasis is on the ease of approximating the inversion of $\tilde{A}_\ell^{\mathrm{FF}}(p)$ without resorting to additional linear solvers. To do so, our focus narrows down to approximating the upper-left block operator. Therefore, we opt not to approximate the other two proposed parts of the operator at this stage, deferring this aspect to future work. We further defer many of the other proposed approximations from [27, 32] to further work, which boils down to an entry-wise normalization that is unknown within low-rank tensor formats.

Instead, we focus on two approaches: Our first straightforward approach to solve this problem is to invert only the diagonal of $\tilde{A}_\ell^{\mathrm{FF}}(p)$ as an approximation. As we are then back in the case of approximating a parameter-dependent *diagonal* operator within the low-rank format, we achieve this using our previous work in [16]. Such an approximation is also proposed for the matrix case in [27, Section 2.2].

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Our second approach is based on our work in [18], where we directly applied exponential sums for the approximated inverse of an operator in the hierarchical Tucker format. This enables us to compute an approximation of the inverse in a direct manner. Despite the potential for a higher rank growth due to numerous operator applications, the use of a truncation allows us to consistently compute a rough approximation of the inverse. This approximation aligns with the recommendations from the literature for prolongation and restriction shown above.

In this context, employing a rough accuracy during the inversion in the low-rank format ensures that $\tilde{A}_\ell^{\mathrm{FF}}(p)$ is computationally inexpensive to invert. Additionally, by controlling the truncation error, we can guarantee a bound on the error such that $-(\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1}\tilde{A}_\ell^{\mathrm{FC}}(p)$ provides a reliable approximation of $-(A_\ell^{\mathrm{FF}}(p))^{-1}A_\ell^{\mathrm{FC}}(p)$. Utilizing only a rough truncation accuracy ensures that the operator is data-sparse. However, a too inaccurate approximation could lead to a decrease in the convergence, leading to a trade-off between arithmetic effort and guaranteed convergence.

As the next step, we talk about the representation of the operator-dependent prolongation using low-rank formats in the case of a parameter-dependent problem. Here, the operator $\mathcal{A}_\ell$ is represented as a tensor operator using the hierarchical Tucker format; cf. Section 3. Additionally, both the solution $\mathcal{U}_\ell$ and the right-hand side $\mathcal{F}_\ell$ are represented as tensor vectors using the hierarchical Tucker format, as detailed in Section 3.

To precompute the operator-dependent prolongation and restriction for the parameter-dependent problem within low-rank tensor formats, our initial step involves splitting the operator into the desired block form. Given that the operator is of the form $\mathcal{A}_\ell(x_1, x_2, p_1, \dots, p_d)$ and represents a multilinear mapping, it directly inherits the fine and coarse grid splitting from the splitting of the parameter-independent $A_\ell^{(\nu)}$ for all $\nu \in \{0, \dots, d\}$, i.e.,

$$A_\ell^{(\nu)} = \begin{bmatrix} (A_\ell^{(\nu)})^{\mathrm{FF}} & (A_\ell^{(\nu)})^{\mathrm{FC}} \\ (A_\ell^{(\nu)})^{\mathrm{CF}} & (A_\ell^{(\nu)})^{\mathrm{CC}} \end{bmatrix}.$$

THEOREM 6.4. *Assume $A_\ell(p)$ is given as an affine-linear operator with a CP representation $\mathcal{A}_\ell$ as in (3.1) or a hierarchical Tucker representation.*

Then we can split $\mathcal{A}_\ell$ into

$$\mathcal{A}_\ell = \begin{bmatrix} \mathcal{A}_\ell^{\mathrm{FF}} & \mathcal{A}_\ell^{\mathrm{FC}} \\ \mathcal{A}_\ell^{\mathrm{CF}} & \mathcal{A}_\ell^{\mathrm{CC}} \end{bmatrix},$$

where $\mathcal{A}_\ell^{\mathrm{FF}}$, $\mathcal{A}_\ell^{\mathrm{FC}}$, $\mathcal{A}_\ell^{\mathrm{CF}}$, and $\mathcal{A}_\ell^{\mathrm{CC}}$ have the same format as $\mathcal{A}_\ell$, i.e., for a CP representation

$$\begin{aligned} \mathcal{A}_\ell^{\mathrm{FF}} &= \mathrm{Id} \otimes \cdots \otimes \mathrm{Id} \otimes ({A_\ell}^{(0)})^{\mathrm{FF}} + \mathrm{Id} \otimes \cdots \otimes \mathrm{Id} \otimes \mathrm{diag}(p_1^{(1)}, \dots, p_{n_1}^{(1)}) \otimes ({A_\ell}^{(1)})^{\mathrm{FF}} \\ &\quad + \cdots + \mathrm{diag}(p_1^{(d)}, \dots, p_{n_d}^{(d)}) \otimes \cdots \otimes \mathrm{Id} \otimes \mathrm{Id} \otimes ({A_\ell}^{(d)})^{\mathrm{FF}} \\ &= \sum_{k=0}^{d} \bigotimes_{\nu=0}^{d} A_\nu^k \quad \text{with } A_\nu^k = \begin{cases} ({A_\ell}^{(k)})^{\mathrm{FF}}, & \text{if } \nu = d, \\ \mathrm{diag}(p_1^{(\nu)}, \dots, p_{n_\nu}^{(\nu)}), & \text{if } \nu + k = d \text{ and } \nu \neq 0, \\ \mathrm{Id}, & \text{otherwise,} \end{cases} \end{aligned}$$

and likewise for $\mathcal{A}_\ell^{\mathrm{FC}}$, $\mathcal{A}_\ell^{\mathrm{CF}}$, and $\mathcal{A}_\ell^{\mathrm{CC}}$.

After the splitting in Theorem 6.4, we next approximate the inverse of $A_\ell^{\mathrm{FF}}(p)$ using exponential sums. Using this approximation, we get the following result that motivates why a rough approximation of the optimal prolongation could still lead to an operator-dependent prolongation and restriction that fulfills the approximation property.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

THEOREM 6.5. *Let us define an operator-dependent prolongation through*

$$\tilde{P}_\ell(p) = \begin{bmatrix} -(\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} A_\ell^{\mathrm{FC}}(p) \\ \mathrm{Id} \end{bmatrix},$$

with $(\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1}$ *an approximation of* $(A_\ell^{\mathrm{FF}}(p))^{-1}$.

Then for

$$\begin{aligned}
&\begin{bmatrix} \mathrm{Id} & \tilde{P}_\ell(p) \\ 0 & \end{bmatrix}^\top \begin{bmatrix} A_\ell^{\mathrm{FF}}(p) & A_\ell^{\mathrm{FC}}(p) \\ A_\ell^{\mathrm{CF}}(p) & A_\ell^{\mathrm{CC}}(p) \end{bmatrix} \begin{bmatrix} \mathrm{Id} & \tilde{P}_\ell(p) \\ 0 & \end{bmatrix} \\
&\quad = \begin{bmatrix} \mathrm{Id} & -(\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} A_\ell^{\mathrm{FC}}(p) \\ 0 & \mathrm{Id} \end{bmatrix}^\top \begin{bmatrix} A_\ell^{\mathrm{FF}}(p) & A_\ell^{\mathrm{FC}}(p) \\ A_\ell^{\mathrm{CF}}(p) & A_\ell^{\mathrm{CC}}(p) \end{bmatrix} \begin{bmatrix} \mathrm{Id} & -(\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} A_\ell^{\mathrm{FC}}(p) \\ 0 & \mathrm{Id} \end{bmatrix} \\
&\quad = \begin{bmatrix} A_\ell^{\mathrm{FF}}(p) & \tilde{A}_\ell^{\mathrm{FC}}(p) \\ \tilde{A}_\ell^{\mathrm{CF}}(p) & \tilde{A}_\ell^{\mathrm{CC}}(p) \end{bmatrix},
\end{aligned}$$

it holds that

$$\begin{aligned}
\tilde{A}_\ell^{\mathrm{FC}}(p) &= \left((\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} - (A_\ell^{\mathrm{FF}}(p))^{-1} \right) A_\ell^{\mathrm{FF}}(p) A_\ell^{\mathrm{FC}}(p), \\
\tilde{A}_\ell^{\mathrm{CF}}(p) &= A_\ell^{\mathrm{CF}}(p) A_\ell^{\mathrm{FF}}(p) \left((\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} - (A_\ell^{\mathrm{FF}}(p))^{-1} \right),
\end{aligned}$$

and

$$\begin{aligned}
\tilde{A}_\ell^{\mathrm{CC}}(p) = A_\ell^{\mathrm{CC}}(p) &- A_\ell^{\mathrm{CF}}(p) (\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} A_\ell^{\mathrm{FC}}(p) \\
&+ A_\ell^{\mathrm{CF}}(p) (\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} A_\ell^{\mathrm{FF}}(p) \left((\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} - (A_\ell^{\mathrm{FF}}(p))^{-1} \right) A_\ell^{\mathrm{FC}}(p).
\end{aligned}$$

Proof. We start with the claim and compute:

$$\begin{aligned}
&\begin{bmatrix} \mathrm{Id} & -(\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} A_\ell^{\mathrm{FC}}(p) \\ 0 & \mathrm{Id} \end{bmatrix}^\top \begin{bmatrix} A_\ell^{\mathrm{FF}}(p) & A_\ell^{\mathrm{FC}}(p) \\ A_\ell^{\mathrm{CF}}(p) & A_\ell^{\mathrm{CC}}(p) \end{bmatrix} \begin{bmatrix} \mathrm{Id} & -(\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} A_\ell^{\mathrm{FC}}(p) \\ 0 & \mathrm{Id} \end{bmatrix} \\
&\quad = \begin{bmatrix} \mathrm{Id} & 0 \\ -A_\ell^{\mathrm{CF}}(p) (\tilde{A}_\ell^{\mathrm{FF}}(p))^{-\top} & \mathrm{Id} \end{bmatrix} \begin{bmatrix} A_\ell^{\mathrm{FF}}(p) & -A_\ell^{\mathrm{FF}}(p) (\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} A_\ell^{\mathrm{FC}}(p) + A_\ell^{\mathrm{FC}}(p) \\ A_\ell^{\mathrm{CF}}(p) & -A_\ell^{\mathrm{CF}}(p) (\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} A_\ell^{\mathrm{FC}}(p) + A_\ell^{\mathrm{CC}}(p) \end{bmatrix} \\
&\quad = \begin{bmatrix} A_\ell^{\mathrm{FF}}(p) & \tilde{A}_\ell^{\mathrm{FC}}(p) \\ \tilde{A}_\ell^{\mathrm{CF}}(p) & \tilde{A}_\ell^{\mathrm{CC}}(p) \end{bmatrix},
\end{aligned}$$

with

$$\begin{aligned}
\tilde{A}_\ell^{\mathrm{FC}}(p) &\coloneqq A_\ell^{\mathrm{FC}}(p) - A_\ell^{\mathrm{FF}}(p) (\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} A_\ell^{\mathrm{FC}}(p), \\
\tilde{A}_\ell^{\mathrm{CF}}(p) &\coloneqq A_\ell^{\mathrm{CF}}(p) - A_\ell^{\mathrm{CF}}(p) (\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} A_\ell^{\mathrm{FF}}(p),
\end{aligned}$$

and

$$\begin{aligned}
\tilde{A}_\ell^{\mathrm{CC}}(p) \coloneqq A_\ell^{\mathrm{CC}}(p) &- 2 A_\ell^{\mathrm{CF}}(p) (\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} A_\ell^{\mathrm{FC}}(p) \\
&+ A_\ell^{\mathrm{CF}}(p) (\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} A_\ell^{\mathrm{FF}}(p) (\tilde{A}_\ell^{\mathrm{FF}}(p))^{-\top} A_\ell^{\mathrm{FC}}(p).
\end{aligned}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

As it now holds that

$$\begin{aligned}(\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} A_\ell^{\mathrm{FF}}(p) &= (\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} A_\ell^{\mathrm{FF}}(p) - A_\ell^{\mathrm{FF}}(p)^{-1} A_\ell^{\mathrm{FF}}(p) + A_\ell^{\mathrm{FF}}(p)^{-1} A_\ell^{\mathrm{FF}}(p) \\ &= \left((\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} - A_\ell^{\mathrm{FF}}(p)^{-1}\right) A_\ell^{\mathrm{FF}}(p) + \mathrm{Id}\end{aligned}$$

and also

$$A_\ell^{\mathrm{FF}}(p)(\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} = A_\ell^{\mathrm{FF}}(p)\left((\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1} - A_\ell^{\mathrm{FF}}(p)^{-1}\right) + \mathrm{Id},$$

the claim follows. □

As an approximation in Theorem 6.5 using exponential sums can be made arbitrarily accurate, it holds that

$$\|(A_\ell^{\mathrm{FF}}(p))^{-1} - (\tilde{A}_\ell^{\mathrm{FF}}(p))^{-1}\| \leq \varepsilon. \tag{6.1}$$

Now for $\varepsilon \to 0$ in (6.1), Theorem 6.5 gives us a prolongation and restriction that separate the fine and coarse grids according to our motivation from Section 5. So if we were to compute in exact arithmetic, this approximation would fulfill, according to Theorem 6.2, the approximation property. Using such a high accuracy in the non-diagonal case could lead to a possible rapid rank growth, therefore, Theorem 6.5 should be understood as a motivation why the rough approximation works so well in practice. Further, it is evident that utilizing a rougher approximation would lead to the necessity for some changes to the operator to guarantee the convergence, e.g., a scaling of the operator. Consequently, this work can be regarded as a preliminary investigation, as such a scaling is non-trivial in the low-rank tensor case.

7. Numerical experiments. We now present numerical experiments that support the findings from the preceding sections. As a general setting, we compute to a relative residual of 1×10^{-3} and truncate to a relative accuracy of 1×10^{-4}. The smoother employed is an approximated Jacobi method with $\nu_1 = \nu_2 = 2$ pre- and post-smoothing iterations.

In our first experiment, we compare the relative residual of a one-dimensional Laplace operator with $d = 3$ parameters. The geometric mode size is 16 with three multigrid levels in Figure 7.1 and 32 with four multigrid levels in Figure 7.2. The parameter mode size is 101 with a step size of 1×10^{-2}. Figures 7.1 and 7.2 display log–linear plots of the relative residual as a function of the number of multigrid steps.

In Figure 7.1, one observes that the multigrid method utilizing the operator-dependent prolongation and restriction exhibits a fast convergence rate, while using standard linear interpolation results in decreased convergence.

Figure 7.2 reinforces this observation, showing that the proposed method maintains a fast convergence rate, whereas the standard linear interpolation diverges. In this one-dimensional case, A_ℓ^{FF} is a diagonal matrix and therefore the optimal operator-dependent prolongation and restriction can be approximated within the CP format with an exponentially fast decaying error, ensuring the approximation property for $\varepsilon \to 0$. Conversely, with a relatively large grid size and jumping (discontinuous) parameters, the standard linear interpolation fails to meet the approximation property, leading to divergence. This underscores the necessity of robust operator-dependent prolongation and restriction for parameter-dependent multigrid methods using low-rank tensor methods.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

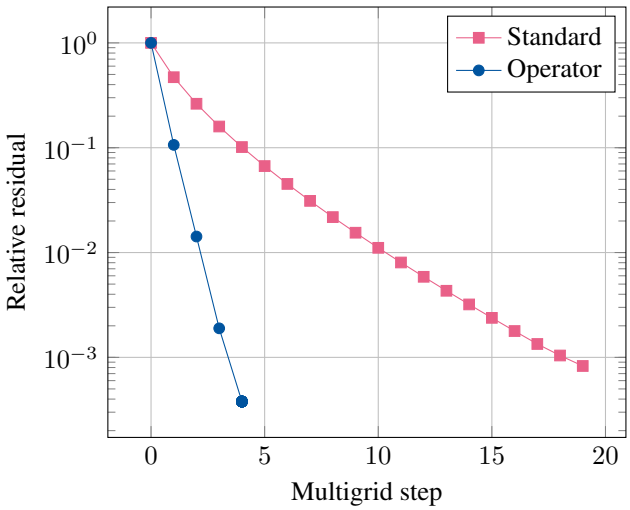

FIG. 7.1. *Comparison of a standard linear prolongation and restriction with our operator-dependent prolongation and restriction for the parameter-dependent multigrid method for a one-dimensional Laplace operator with* $d = 3$ *parameters with a geometric mode size of* 16 *and a parameter mode size of* 101.

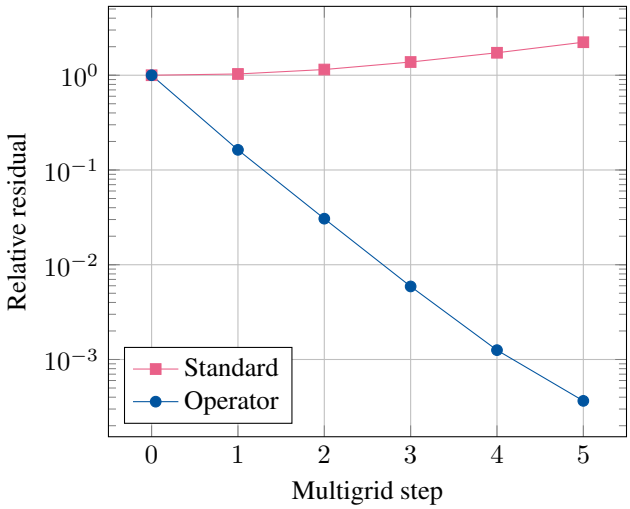

FIG. 7.2. *Comparison of a standard linear prolongation and restriction with our operator-dependent prolongation and restriction for the parameter-dependent multigrid method for a one-dimensional Laplace operator with* $d = 3$ *parameters with a geometric mode size of* 32 *and a parameter mode size of* 101.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

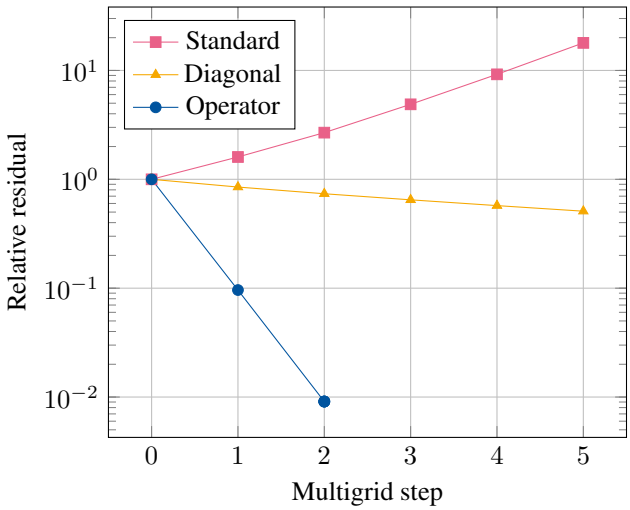

FIG. 7.3. *Comparison of a standard linear prolongation and restriction, with a diagonal approximation, and our operator-dependent prolongation and restriction for the parameter-dependent multigrid method for a two-dimensional Laplace operator with* $d = 3$ *parameters with a geometric mode size of* 32×32 *and a parameter mode size of* 101.

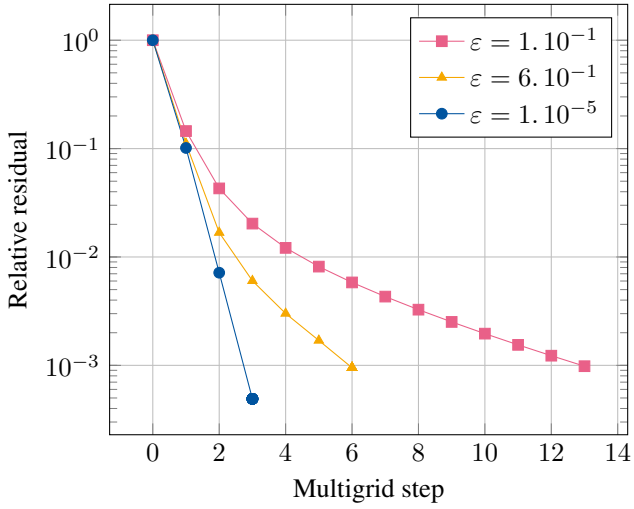

FIG. 7.4. *Comparison of different accuracy value* ε *in* (6.1) *for our operator-dependent prolongation and restriction for the parameter-dependent multigrid method for a one-dimensional Laplace operator with* $d = 3$ *parameters with a geometric mode size of* 32 *and a parameter mode size of* 101.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Moving to more complex geometries, our next experiment involves a two-dimensional Laplace operator with $d = 3$ parameters. We select a geometric mode size of 32×32 with four multigrid levels and a parameter mode size of 101 with a step size of 1×10^{-2}. In the log–linear plot of Figure 7.3, the relative residual is shown as a function of the number of multigrid steps, evaluating the standard linear prolongation and restriction against a directly computed approximation of the operator-dependent prolongation and restriction within the hierarchical Tucker format, as well as a diagonal approximation of the upper-left block.

Once again, Figure 7.3 illustrates the fast multigrid convergence of the method employing the approximated operator-dependent prolongation and restriction. The linear interpolation method diverges, and the diagonal approximation, while convergent, exhibits slower convergence. This suggests that relying solely on a diagonal approximation of the upper-left block may not be sufficient for achieving fast convergence.

To delve deeper into the insufficiency of a diagonal approximation, we conduct a further experiment comparing different truncation accuracies when computing the operator-dependent prolongation and restriction within the low-rank format. Here, we use a one-dimensional Laplace operator with three parameters, opting for a geometric mode size of 32 with three multigrid levels. The parameter mode size remains at 101 with a step size of 1×10^{-1}. In the log–linear plot of Figure 7.4, the relative residual is given as a function of the number of multigrid steps for various truncation accuracies.

Figure 7.4 indicates that, up to a relatively rough accuracy, the approximated operator-dependent prolongation and restriction maintain fast multigrid convergence. The experiment implies that the error allowed in practice could be rather high, so that it results in a fast method, as such a rough approximation is relatively cheap to compute.

8. Related and future work. The combination of the multigrid method with low-rank tensor formats was discussed by Hackbusch in [21]. There he emphasized the use of the Richardson method as smoother in combination with the standard linear interpolation for prolongation and restriction, coupled with the low-rank tensor arithmetic. Further, Hackbusch addressed the challenge of selecting an efficient smoother within the low-rank format, as the inversion of even the diagonal within the format is a complex task.

In the context of computing the stationary distribution of Markov chains [6, 7], the multigrid method was employed together with an alternating least-squares-based approximation of the smoother. Notably, this approach needs the solution of a linear system as part of the smoothing step, a distinction from our method, wherein the inverse part can be precomputed.

In [16], we presented the approximation of the Jacobi method in the CP format as smoother. Furthermore, we presented numerical experiments illustrating that our method leads to the standard multigrid convergence rates. Building on this, in [18] we extended the method to compute the approximation of the inverse directly within the hierarchical Tucker format.

Exploring low-rank representations for linear systems, such as those expressed in (1.1), and their solution was the focus of several articles; see [3, 17] for an overview. In [4], an iterative method based on soft thresholding was employed, presenting an alternative to truncation methods. The theory derived there could potentially improve the error and rank guarantees of our proposed method and is therefore a promising avenue for future work.

Krylov-based methods [5, 13, 29] have been used in solving similar systems. A potential application of our proposed multigrid method would be as a preconditioner in those methods, as a good precondition is often needed to ensure fast convergence. Building on the well-known combination of the conjugate gradient method with the multigrid method as smoother in the matrix case, a future direction could involve the investigation of the efficiency of the parameter-dependent multigrid method using a coarse truncation accuracy.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

9. Conclusion. Expanding upon our prior research in [16], where we derived an approximation of the parameter-dependent Jacobi method as smoother, this article expands the robustness of the multigrid method for parameter-dependent problems. Our investigation, however, unveils a critical challenge when dealing with jumping coefficients in the parameter-dependent multigrid method. The use of the standard linear interpolation in this context results in diminishing convergence rates or outright divergence.

To address this issue, we derive an operator-dependent prolongation and restriction based on the concept of block Gaussian elimination. Notably, building on our previous work, we compute this operator-dependent prolongation and restriction directly within the low-rank format. In the case of exact arithmetic, we prove that it fulfills the approximation property.

Furthermore, in numerical experiments involving truncation, we present the effectiveness of our approach.

Acknowledgments. We thank the anonymous reviewers for their careful reading of our manuscript and their many insightful comments and suggestions. We would also like to thank Maren Klever for her critical reading of and suggestions for this article.

REFERENCES

[1] R. E. ALCOUFFE, A. BRANDT, J. E. DENDY, JR., AND J. W. PAINTER, *The multigrid method for the diffusion equation with strongly discontinuous coefficients*, SIAM J. Sci. Statist. Comput., 2 (1981), pp. 430–454.
[2] O. AXELSSON, *A survey of algebraic multilevel iteration (AMLI) methods*, BIT, 43 (2003), pp. 863–879.
[3] M. BACHMAYR, *Low-rank tensor methods for partial differential equations*, Acta Numer., 32 (2023), pp. 1–121.
[4] M. BACHMAYR AND R. SCHNEIDER, *Iterative methods based on soft thresholding of hierarchical tensors*, Found. Comput. Math., 17 (2017), pp. 1037–1083.
[5] J. BALLANI AND L. GRASEDYCK, *A projection method to solve linear systems in tensor format*, Numer. Linear Algebra Appl., 20 (2013), pp. 27–43.
[6] M. BOLTEN, K. KAHL, D. KRESSNER, F. MACEDO, AND S. SOKOLOVIĆ, *Multigrid methods combined with low-rank approximation for tensor-structured Markov chains*, Electron. Trans. Numer. Anal., 48 (2018), pp. 348–361. https://etna.ricam.oeaw.ac.at/vol.48.2018/pp348-361.dir/pp348-361.pdf
[7] M. BOLTEN, K. KAHL, AND S. SOKOLOVIĆ, *Multigrid methods for tensor structured Markov chains with low rank approximation*, SIAM J. Sci. Comput., 38 (2016), pp. A649–A667.
[8] J. D. CARROLL AND J.-J. CHANG, *Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart–Young" decomposition*, Psychometrika, 35 (1970), pp. 283–319.
[9] T. F. CHAN AND W. L. WAN, *Robust multigrid methods for nonsmooth coefficient elliptic linear systems*, J. Comput. Appl. Math., 123 (2000), pp. 323–352.
[10] V. DE SILVA AND L.-H. LIM, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1084–1127.
[11] P. M. DE ZEEUW, *Matrix-dependent prolongations and restrictions in a blackbox multigrid solver*, J. Comput. Appl. Math., 33 (1990), pp. 1–27.
[12] J. E. DENDY, JR., *Black box multigrid*, J. Comput. Phys., 48 (1982), pp. 366–386.
[13] S. V. DOLGOV, *TT-GMRES: solution to a linear system in the structured tensor format*, Russian J. Numer. Anal. Math. Modelling, 28 (2013), pp. 149–172.
[14] L. GRASEDYCK, *Hierarchical singular value decomposition of tensors*, SIAM J. Matrix Anal. Appl., 31 (2009/10), pp. 2029–2054.
[15] L. GRASEDYCK, L. JUSCHKA, AND C. LÖBBERT, *Finding entries of maximum absolute value in low-rank tensors*, Preprint on arXiv, 2019. https://arxiv.org/abs/1912.02072
[16] L. GRASEDYCK, M. KLEVER, C. LÖBBERT, AND T. A. WERTHMANN, *A parameter-dependent smoother for the multigrid method*, Preprint on arXiv, 2019. https://arxiv.org/abs/2008.00927
[17] L. GRASEDYCK, D. KRESSNER, AND C. TOBLER, *A literature survey of low-rank tensor approximation techniques*, GAMM-Mitt., 36 (2013), pp. 53–78.
[18] L. GRASEDYCK AND T. A. WERTHMANN, *Computing tensor operator exponentials within low-rank tensor formats with application to the parameter-dependent multigrid method*, PAMM-Proc. Appl. Math. Mech., 22 (2023), Paper No. e202200093, 6 pages.
[19] W. HACKBUSCH, *Multigrid Methods and Applications*, Springer, Berlin, 1985.
[20] ———, *Tensor Spaces and Numerical Tensor Calculus*, Springer, Heidelberg, 2012.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

[21] ———, *Solution of linear systems in high spatial dimensions*, Comput. Vis. Sci., 17 (2015), pp. 111–118.
[22] ———, *Computation of best l^∞ exponential sums for $1/x$ by Remez' algorithm*, Comput. Vis. Sci., 20 (2019), pp. 1–11.
[23] W. HACKBUSCH, B. N. KHOROMSKIJ, AND E. E. TYRTYSHNIKOV, *Approximate iterations for structured matrices*, Numer. Math., 109 (2008), pp. 365–383.
[24] W. HACKBUSCH AND S. KÜHN, *A new scheme for the tensor representation*, J. Fourier Anal. Appl., 15 (2009), pp. 706–722.
[25] R. A. HARSHMAN, *Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multi-code factor analysis,* UCLA Working Papers in Phonetics 16, UCLA Linguistics Department, UCLA, Los Angeles, 1970, 84 pages.
[26] R. KETTLER, *Analysis and comparison of relaxation schemes in robust multigrid and preconditioned conjugate gradient methods*, in Multigrid Methods (Cologne, 1981), W. Hackbusch and U. Trottenberg, eds., Lecture Notes Math., 960, Springer, Berlin, 1982, pp. 502–534.
[27] S. KNAPEK, *Matrix-dependent multigrid homogeneization for diffusion problems*, SIAM J. Sci. Comput., 20 (1998), pp. 515–533.
[28] A. KRECHEL AND K. STÜBEN, *Operator dependent interpolation in algebraic multigrid*, in Multigrid Methods V (Stuttgart, 1996), W. Hackbusch and G. Wittum, eds., Lecture Notes Comput. Sci. Eng., Vol. 3, Springer, Berlin, 1998, pp. 189–211.
[29] D. KRESSNER AND C. TOBLER, *Low-rank tensor Krylov subspace methods for parametrized linear systems*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 1288–1316.
[30] Y. NOTAY, *Using approximate inverses in algebraic multilevel methods*, Numer. Math., 80 (1998), pp. 397–417.
[31] A. RÖRICH, T. A. WERTHMANN, D. GÖDDEKE, AND L. GRASEDYCK, *Bayesian inversion for electromyography using low-rank tensor formats*, Inverse Problems, 37 (2021), Paper No. 055003, 22 pages.
[32] J. XU AND L. ZIKATANOV, *Algebraic multigrid methods*, Acta Numer., 26 (2017), pp. 591–721.