Noname manuscript No.
(will be inserted by the editor)

An efficient approximate residual evaluation in the adaptive tensor product wavelet method

Sebastian Kestler · **Rob Stevenson**

Received: date / Accepted: date

Abstract A wide class of well-posed operator equations can be solved in optimal computational complexity by adaptive wavelet methods. A quantitative bottleneck is the approximate evaluation of the arising residuals that steer the adaptive refinements. In this paper, we consider multi-tree approximations from tensor product wavelet bases for solving linear PDE's. In this setting, we develop a new efficient approximate residual evaluation. Other than the commonly applied method, that uses the so-called `APPLY` routine, our approximate residual depends affinely on the current approximation of the solution. Our findings are illustrated by numerical results that show a considerable speed-up.

Keywords Adaptive wavelet method · multi-tree tensor product approximation · optimal computational complexity

Mathematics Subject Classification (2000) 41A30 · 41A63 · 65N30 · 65Y20

1 Introduction: Adaptive wavelet Galerkin methods

For some countable set ∇, $\mathbf{f} \in \ell_2(\nabla)$, and an $\mathbf{A} \in \mathscr{L}(\ell_2(\nabla), \ell_2(\nabla))$ that is symmetric and coercive, i.e., $\langle \mathbf{Av}, \mathbf{v} \rangle_{\ell_2(\nabla)} \gtrsim \|\mathbf{v}\|^2_{\ell_2(\nabla)}$ ($\mathbf{v} \in \ell_2(\nabla)$), consider the problem of finding the solution $\mathbf{u} \in \ell_2(\nabla)$ of $\mathbf{Au} = \mathbf{f}$.

S.K. was supported by the Deutsche Forschungsgemeinschaft within the Research Training Group (Graduiertenkolleg) GrK 1100 *Modellierung, Analyse und Simulation in der Wirtschaftsmathematik* at the University of Ulm.

Sebastian Kestler
Institute for Numerical Mathematics, University of Ulm, Helmholtzstrasse20, D-89069 Ulm, Germany
Tel.: +49-731-50-23935
Fax: +49-731-50-23548
E-mail: sebastian.kestler@uni-ulm.de

Rob Stevenson
Korteweg-de Vries Institute for Mathematics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands
Tel.: +31-20-525 5805
Fax: +31-20-525 7820
E-mail: r.p.stevenson@uva.nl

Such a problem arises from the equivalent formulation of an elliptic operator equation as a well-posed bi-infinite matrix-vector problem. Indeed, for some real Hilbert space H, let $A \in \mathscr{L}(H,H')$ be such that $(Av)(v) \gtrsim \|v\|_H^2$ $(v \in H)$, let $f \in H'$, and let $\Psi = \{\psi_\lambda : \lambda \in \nabla\}$ be a *Riesz basis* for H, where we have in mind Ψ to be a wavelet basis. Then $Au = f$ is equivalently formulated as $\mathbf{Au} = \mathbf{f}$, where $\mathbf{A} = (A\Psi)(\Psi)$, $u = \mathbf{u}^\top \Psi$ and $\mathbf{f} = f(\Psi)$.

Recall that Ψ being a Riesz basis for H means that $H' \to \ell_2(\nabla) : g \mapsto [g(\psi_\lambda)]_{\lambda\in\nabla}$, or equivalently, its adjoint $\ell_2(\nabla) \to H : \mathbf{v} \mapsto \mathbf{v}^\top \Psi$, is boundedly invertible.

The (idealized) adaptive wavelet-Galerkin method for solving $\mathbf{Au} = \mathbf{f}$ reads as follows. For $\Lambda \subset \nabla$, let $\mathbf{I}_\Lambda \in \mathscr{L}(\ell_2(\Lambda), \ell_2(\nabla))$ be the extension with zeros, $\mathbf{R}_\Lambda := \mathbf{I}_\Lambda^\top$, $\mathbf{f}_\Lambda := \mathbf{R}_\Lambda \mathbf{f}$, and $\mathbf{A}_\Lambda := \mathbf{R}_\Lambda \mathbf{A} \mathbf{I}_\Lambda$. Let $\mu \in (0,1)$ be some constant, and $\Lambda_0 \subset \nabla$.

```
for k = 0, 1, ... do
    solve A_{Λ_k} u_{Λ_k} = f_{Λ_k}
    find the smallest Λ_{k+1} ⊃ Λ_k such that
                ‖R_{Λ_{k+1}}(f − A I_{Λ_k} u_{Λ_k})‖_{ℓ2(Λ_{k+1})} ≥ μ ‖f − A I_{Λ_k} u_{Λ_k}‖_{ℓ2(∇)}
enddo
```

This algorithm, with an additional recurrent coarsening step, was introduced in [9], where in [18] it was demonstrated that this coarsening can be omitted. For μ being sufficiently small, the algorithm was proven to converge optimally in the following sense: If, for whatever $s > 0$, $\mathbf{u}$ belongs to the (unconstrained) *non-linear approximation class*,

$$\{\mathbf{v} \in \ell_2(\nabla) : |\mathbf{v}|_{\mathscr{A}^s} := \sup_{N\in\mathbb{N}_0} N^s \inf_{\{\Lambda\subset\nabla : \#\Lambda = N\}} \|\mathbf{v} - \mathbf{I}_\Lambda \mathbf{R}_\Lambda \mathbf{v}\|_{\ell_2(\nabla)} < \infty\},$$

then the sequence of Galerkin solutions satisfies

$$\sup_k (\#\Lambda_k)^s \|\mathbf{u} - \mathbf{I}_{\Lambda_k} \mathbf{u}_{\Lambda_k}\|_{\ell_2(\nabla)} < \infty. \tag{1}$$

The same statement is true when inside the algorithm, as well as in the definition of the approximation classes, only Λ from some subset of $\mathscr{P}(\nabla)$ are considered, which subset of "admissible" Λ should be closed under taking unions. The idea is to take this subset such that on the one hand it allows a more efficient implementation of the adaptive algorithm, whereas on the other hand the approximation classes become only "slightly" smaller. Examples of such a constrained approximation are given by tree or multi-tree approximation, where the latter applies with tensor product approximation as will be discussed later. This tree or multi-tree structure allows to switch in optimal computational complexity from the wavelet representation to a locally finite single scale representation, or to do this in each coordinate separately.

The above optimally converging adaptive wavelet Galerkin method is not an implementable algorithm, which is why we called it an idealized method. Indeed, already the residuals that are used to steer the adaptivity are generally infinitely supported, and therefore not computable.

It was, however, shown that convergence with the best possible rate is maintained under the following relaxations. The sequences $(\mathbf{u}_{\Lambda_k})_k$ and $(\mathbf{f} - \mathbf{A}\mathbf{I}_{\Lambda_k}\mathbf{u}_{\Lambda_k})_k$ of exact Galerkin solutions and residuals can be read as $(\mathbf{w}_{\Lambda_k})_k$ and $(\mathbf{r}^{(k)})_k$ when, for some sufficiently small constant $\eta > 0$,

$$\|\mathbf{f} - \mathbf{A}\mathbf{I}_{\Lambda_k}\mathbf{w}_{\Lambda_k} - \mathbf{r}^{(k)}\|_{\ell_2(\nabla)} \le \eta \|\mathbf{f} - \mathbf{A}\mathbf{I}_{\Lambda_k}\mathbf{w}_{\Lambda_k}\|_{\ell_2(\nabla)}, \tag{2}$$

$$\|\mathbf{u}_{\Lambda_k} - \mathbf{w}_{\Lambda_k}\|_{\ell_2(\Lambda_k)} \le \eta \|\mathbf{r}^{(k-1)}\|_{\ell_2(\nabla)}. \tag{3}$$

Moreover, instead of finding a truly smallest $\Lambda_{k+1} \supset \Lambda_k$ such that $\|\mathbf{R}_{\Lambda_{k+1}}\mathbf{r}^{(k)}\|_{\ell_2(\Lambda_{k+1})} \geq \mu\|\mathbf{r}^{(k)}\|_{\ell_2(\nabla)}$, it is sufficient when such a $\Lambda_{k+1} \supset \Lambda_k$ is found with $\#(\Lambda_{k+1}\setminus\Lambda_k)$ being minimal up to some constant factor (actually, this condition can be relaxed further, see [27, Algorithm 3.7]).

Condition (2) can be realized by an inner loop in which $\mathbf{f}-\mathbf{A}\mathbf{I}_{\Lambda_k}\mathbf{w}_{\Lambda_k}$ is approximated within some absolute tolerance ε, that is halved in each iteration until $\varepsilon \leq \frac{\eta}{1+\eta}\|\mathbf{r}^{(k)}\|_{\ell_2(\nabla)}$, which implies (2).

Concerning (3), since for any Λ, the matrix $\mathbf{A}_\Lambda$ is symmetric, positive definite, and well-conditioned, uniformly in Λ, the arising Galerkin problems can be efficiently approximately solved by the application of an iterative method, as the simple Richardson method. Such a method requires the approximate evaluation of residuals restricted to the current set Λ, i.e., residuals of type $\mathbf{R}_\Lambda(\mathbf{f}-\mathbf{A}\mathbf{I}_\Lambda\mathbf{w}_\Lambda) = \mathbf{f}_\Lambda - \mathbf{A}_\Lambda\mathbf{w}_\Lambda$, which task is easier than the approximation of an infinite residual $\mathbf{f}-\mathbf{A}\mathbf{I}_\Lambda\mathbf{w}_\Lambda$ required for (2).

For the resulting practical adaptive wavelet Galerkin method (`AWGM`), it was shown that *if* for fixed, sufficiently small $\eta > 0$, and for s for which $\mathbf{u} \in \mathscr{A}^s$, for any $\mathbf{w}_\Lambda \in \ell_2(\Lambda)$,

$$\begin{cases} \mathbf{f}-\mathbf{A}\mathbf{I}_\Lambda\mathbf{w}_\Lambda \text{ can be approximated within any absolute tolerance} \\ \varepsilon \geq \eta\|\mathbf{f}-\mathbf{A}\mathbf{I}_\Lambda\mathbf{w}_\Lambda\|_{\ell_2(\nabla)} \text{ at cost } \mathscr{O}(\varepsilon^{-1/s}+\#\Lambda), \end{cases} \tag{4}$$

only dependent on the exact solution $\mathbf{u}$ or, equivalently, on the right-hand side $\mathbf{f}$, *then* the `AWGM` is of *optimal computational complexity* in the sense that

$$\sup_k (\#\mathrm{ops}_k)^s \|\mathbf{u}-\mathbf{I}_{\Lambda_k}\mathbf{w}_{\Lambda_k}\|_{\ell_2(\nabla)} < \infty,$$

where $\#\mathrm{ops}_k$ is the number of arithmetic operations used to compute $\mathbf{w}_{\Lambda_k}$. The task (4) is the topic of the current paper.

Least-squares problems

Before focussing on task (4), we briefly discuss a more general setting. For real Hilbert spaces H and K, let $A \in \mathscr{L}(H,K')$ be a homeomorphism onto its range, i.e., $\|Av\|_{K'} \eqsim \|v\|_H$ ($v \in H$), meaning that in case this range is K', A is boundedly invertible. Let $\Psi = \{\psi_\lambda : \lambda \in \nabla_\Psi\}$ and $\Xi = \{\xi_\lambda : \lambda \in \nabla_\Xi\}$ be Riesz bases (of wavelet type) for H and K. Equipping K' with the (equivalent) norm $\|f\|_{K'} := \sup_{0\neq w\in K} \frac{|f(w)|}{\|\mathbf{w}\|_{\ell_2(\nabla_\Xi)}}$, where $w = \mathbf{w}^\top\Xi$, for given $f \in K'$ the problem of finding $u = \mathrm{argmin}_{v\in H}\|f-Av\|_{K'}$ (i.e., $Au = f$ when A is invertible) is equivalent to $\mathbf{A}^\top\mathbf{A}\mathbf{u} = \mathbf{A}^\top\mathbf{f}$, where $\mathbf{A} = (A\Psi)(\Xi)$, $u = \mathbf{u}^\top\Psi$, and $\mathbf{f} = f(\Xi)$.

The matrix $\mathbf{A}^\top\mathbf{A}$ is symmetric and coercive, so the adaptive wavelet Galerkin method can be applied to solve these normal equations. The residual from (4) now reads as $\mathbf{A}^\top(\mathbf{f}-\mathbf{A}\mathbf{I}_\Lambda\mathbf{w}_\Lambda)$. The task of its approximate evaluation can be viewed as the concatenation of two subtasks; first the approximate evaluation of the primal residual $\mathbf{f}-\mathbf{A}\mathbf{I}_\Lambda\mathbf{w}_\Lambda$, yielding $\mathbf{r}_{\tilde\Lambda} \in \ell_2(\tilde\Lambda)$ for some $\tilde\Lambda \subset \nabla_\Xi$, and then the approximate evaluation of the dual residual $\mathbf{g}-\mathbf{A}^\top\mathbf{I}_{\tilde\Lambda}\mathbf{r}_{\tilde\Lambda}$, where $\mathbf{g}$ happens to be zero.

Although both subtasks look similar, the second one is easier. Recalling that inside the adaptive wavelet Galerkin method, it always suffices to evaluate residuals up to some, sufficiently small, *relative* tolerance (cf. the lower bound on ε in (4)), and that $\|\mathbf{r}_{\tilde\Lambda}\|_{\ell_2(\tilde\Lambda)} \lesssim \|\mathbf{u}-\mathbf{I}_\Lambda\mathbf{w}_\Lambda\|_{\ell_2(\nabla_\Psi)} \lesssim \|\mathbf{A}^\top(\mathbf{f}-\mathbf{A}\mathbf{I}_\Lambda\mathbf{w}_\Lambda)\|_{\ell_2(\nabla_\Psi)}$, we infer that the second task is performed by replacing the multiplication with $\mathbf{A}^\top$ by that with $\mathbf{A}_\eta^\top$, when, for some sufficiently small $\eta > 0$, $\|\mathbf{A}-\mathbf{A}_\eta\|_{\mathscr{L}(\ell_2(\nabla_\Psi),\ell_2(\nabla_\Xi))} \leq \eta$. Under usual assumptions on the operator A and on

the collections Ψ and Ξ, which already will be needed to perform the first subtask, such an $\mathbf{A}_\eta$ exists which is sparse, dependent on η. We infer that the second task can be performed in $\mathcal{O}(\#\tilde{\Lambda})$ operations. So we conclude that also in this more general setting of least squares problems, the main task is the approximate evaluation of the primal residual satisfying (4).

To end this excursion into more general settings, we note that the adaptive wavelet Galerkin method can also be applied to solve non-linear (least-squares) problems, see [27].

Residual evaluation using the APPLY-routine

Returning to the task (4), for the unconstrained approximation setting already in [10] an approximate matrix-vector multiplication routine APPLY was developed for approximating $\mathbf{A}\mathbf{I}_{\Lambda_k}\mathbf{w}_{\Lambda_k}$. It consists of an approximation scheme for the individual columns of $\mathbf{A}$ with accuracies that are increasing as function of the modulus of the corresponding entry in the input vector. The resulting approximate multiplication of $\mathbf{A}$ with the vector $\mathbf{I}_{\Lambda_k}\mathbf{w}_{\Lambda_k}$ is therefore a *non-linear* mapping on $\ell_2(\Lambda_k)$.

Generally, each column of $\mathbf{A}$ has infinitely many non-zero entries. Under standard conditions on the wavelet bases and on the operator A, however, $\mathbf{A}$ is close to being a sparse matrix in the sense that these entries, re-ordered by decreasing modulus, form a sequence that rapidly tends to zero. Assuming a sufficient near-sparsity of $\mathbf{A}$, depending on s from (4), and that of $\mathbf{w}_{\Lambda_k}$, the latter in the sense that $\sup_k |\mathbf{I}_{\Lambda_k}\mathbf{w}_{\Lambda_k}|_{\mathscr{A}^s} < \infty$, the cost of this APPLY-routine with prescribed absolute tolerance $\varepsilon/2$ is $\mathcal{O}(\varepsilon^{-1/s} + \#\Lambda_k)$.

Given a Hilbert space H and a Riesz basis Ψ, there is a maximum $s_{\max}$ on the value of s for which $\mathbf{u} \in \mathscr{A}^s$ can be generally expected. For large classes of singular integral operators, and partial differential operators with smooth coefficients, for $s \leq s_{\max}$ the sufficient near-sparsity of $\mathbf{A}$ has been verified for bases of wavelets that are sufficiently smooth, and have sufficiently many vanishing moments. The condition $\sup_k |\mathbf{I}_{\Lambda_k}\mathbf{w}_{\Lambda_k}|_{\mathscr{A}^s} < \infty$ is a consequence of $\mathbf{u} \in \mathscr{A}^s$ and the convergence inside the adaptive wavelet Galerkin method of $(\mathbf{I}_{\Lambda_k}\mathbf{w}_{\Lambda_k})_k$ towards $\mathbf{u}$ with rate s.

Usually, approximating $\mathbf{f}$ does not pose any problem. From the aforementioned near-sparsity of $\mathbf{A}$, it follows that if, for $s \leq s_{\max}$, $\mathbf{u} \in \mathscr{A}^s$, then $\mathbf{f} = \mathbf{A}\mathbf{u} \in \mathscr{A}^s$, where in many cases even $\mathbf{f} \in \mathscr{A}^{\tilde{s}}$ for some $\tilde{s} > s$ can be demonstrated. Membership of $\mathbf{f} \in \mathscr{A}^s$ means that there exists an $\mathbf{f}_\varepsilon$ with $\|\mathbf{f} - \mathbf{f}_\varepsilon\|_{\ell(\nabla_\Xi)} \leq \varepsilon/2$, where $\operatorname{supp}\mathbf{f}_\varepsilon \subset \Lambda$ with $\#\Lambda \lesssim \varepsilon^{-1/s}$, and in the constrained approximation case, with Λ being admissible. Assuming that this $\mathbf{f}_\varepsilon$ can also be *constructed* in $\mathcal{O}(\varepsilon^{-1/s})$ operations, which involves a quadrature issue, we conclude that (4) is realized.

Alternative approaches

Although qualitatively fully satisfactory, quantitatively the approximate evaluation of the residual by means of the application of the APPLY-routine turns out to be quite demanding as has been observed by many researchers. Various attempts were made to improve the efficiency of the APPLY-routine, or to avoid its use.

In [7,8,14], we considered the case of Ψ (and Ξ when $K \neq H$) to be a tensor product of bases of univariate functions. The advantage of this tensor product setting is that dimension independent convergence rates are obtained. On the other hand, the application of such bases requires that the operator equation is posed on a product domain. Recently, in [6], it was demonstrated that also non-product domains can be considered by equipping them by *piecewise* tensor product bases constructed via a domain decomposition technique. The present work, however, will be restricted to product domains.

In the aforementioned papers, tensor product bases were constructed such that for A corresponding to a partial differential operator with constant coefficients, $\mathbf{A} = (A\Psi)(\Psi)$ (or $\mathbf{A} = (A\Psi)(\Xi)$) is *truly* sparse, and so can be applied *exactly* in linear complexity, yielding an adaptive wavelet Galerkin method that is also quantitatively efficient.

This approach, however, is restricted to the use of specific wavelet bases. So far, for operators of second order, we were able to construct them as globally C^1, piecewise polynomials of degree 4 (or higher), with relatively large supports. The bases are *bi*orthogonal, i.e., not L_2-orthogonal. The use of L_2-orthogonal univariate wavelets has the important advantage that the condition number of the resulting tensor product basis is bounded, uniformly in the number of factors, i.e., in the space dimension.

Residual evaluation in this paper

In the current paper, we consider constrained tensor product approximation, with the constraint being that only approximations from spans of collections of tensor product wavelets are considered with index sets that are multi-trees. The natural index set of a tensor product basis is the Cartesian product of the index sets of the univariate bases that are the factors in the product. A subset of this index set is a multi-tree when "frozen" in all but one of its coordinates, it is a tree in the remaining "free" coordinate.

For $\Psi = \{\psi_{\vec{\lambda}} : \vec{\lambda} \in \nabla_\Psi\}$, $\Xi = \{\xi_{\vec{\lambda}} : \vec{\lambda} \in \nabla_\Xi\}$ being tensor product bases for H and K, an $A \in \mathscr{L}(H,K')$ that corresponds to a linear partial differential operator with polynomial coefficients, and $f = Au$, let $\mathbf{A} := (A\Psi)(\Xi)$, $u = \mathbf{u}^\top\Psi$, and $\mathbf{f} := f(\Xi)$. Under the assumption of the functions in Ψ to be sufficiently smooth (C^1 for a second order operator), we will prove that for $\mathbf{u} \in \mathscr{A}^s$, $\Lambda_\Psi \subset \nabla_\Psi$ being a multi-tree, $\mathbf{w}_{\Lambda_\Psi} \in \ell_2(\Lambda_\Psi)$, and $\varepsilon \geq \eta\|\mathbf{f} - \mathbf{A}\mathbf{I}_{\Lambda_\Psi}\mathbf{w}_{\Lambda_\Psi}\|_{\ell_2(\nabla_\Xi)}$, there exists a multi-tree $\Lambda_\Xi \subset \nabla_\Xi$, with $\#\Lambda_\Xi \lesssim \#\Lambda_\Psi + \varepsilon^{-1/s}$, such that $(\mathbf{f} - \mathbf{A}\mathbf{I}_{\Lambda_\Psi}\mathbf{w}_{\Lambda_\Psi})|_{\Lambda_\Xi}$ approximates $\mathbf{f} - \mathbf{A}\mathbf{I}_{\Lambda_\Psi}\mathbf{w}_{\Lambda_\Psi}$ in $\ell_2(\nabla_\Xi)$ within tolerance ε.

Apart from the computation, or sufficiently accurate approximation of $\mathbf{f}|_{\Lambda_\Xi}$, the *evaluation* of this approximate residual requires the multiplication with $\mathbf{A}|_{\Lambda_\Xi \times \Lambda_\Psi}$. Thanks to the multi-tree structure of both Λ_Ξ and Λ_Ψ, the application of this *linear* mapping can be performed in $\mathscr{O}(\#\Lambda_\Xi + \#\Lambda_\Psi)$ operations by using a generalization of an algorithm for matrix-vector multiplication on sparse grids (e.g., see [3–5, 30, 31]). Because of the way in which multi-trees are traversed, the idea of this algorithm is known as the *unidirectional principle*. Generalizations to adaptive sparse grids can be found in e.g. [1–4, 19, 17, 24]. A generalization to multi-tree wavelet index sets as needed here, together with a formal proof of its linear complexity can be found in [21].

The idea behind the *construction* of the approximate residual evaluation is as follows. Firstly, for A and Ψ as above, where additionally Ψ consists of piecewise polynomials, there exists an *auxiliary* tensor product basis Θ for K, with dual $\tilde{\Theta} = \{\tilde{\theta}_{\vec{\lambda}} : \vec{\lambda} \in \nabla_{\tilde{\Theta}}\}$, such that for any given multi-tree $\Lambda_\Psi \subset \nabla_\Psi$, there exists a multi-tree $\Lambda_{\tilde{\Theta}} \subset \nabla_{\tilde{\Theta}}$ with $\#\Lambda_{\tilde{\Theta}} \lesssim \#\Lambda_\Psi$ and

$$\mathfrak{I}A|_{\operatorname{span}\{\psi_{\vec{\lambda}} : \vec{\lambda} \in \Lambda_\Psi\}} \subset \operatorname{span}\{\tilde{\theta}_{\vec{\lambda}} : \vec{\lambda} \in \Lambda_{\tilde{\Theta}}\}. \tag{5}$$

In particular, we have $A((\mathbf{I}_{\Lambda_\Psi}\mathbf{w}_{\Lambda_\Psi})^\top\Psi) \in \operatorname{span}\{\tilde{\theta}_{\vec{\lambda}} : \vec{\lambda} \in \Lambda_{\tilde{\Theta}}\}$.

Another consequence of (5) is that $\mathbf{u} \in \mathscr{A}^s$ implies that $f(\Theta) \in \mathscr{A}^s$. So there exists a multi-tree in $\nabla_{\tilde{\Theta}}$, with cardinality $\mathscr{O}(\varepsilon^{-1/s})$, such that f can be approximated within tolerance ε in $\|\cdot\|_{K'}$ by a linear combination f_ε of $\tilde{\theta}_{\vec{\lambda}}$ with $\vec{\lambda}$ from this multi-tree. By redefining $\Lambda_{\tilde{\Theta}}$ as the union of the latter multi-tree and the previous $\Lambda_{\tilde{\Theta}}$, we conclude that

$f_\varepsilon - A((\mathbf{I}_{\Lambda_\Psi}\mathbf{w}_{\Lambda_\Psi})^\top \Psi) \in \operatorname{span}\{\tilde\theta_{\vec\lambda} : \vec\lambda \in \Lambda_{\tilde\Theta}\}$ approximates the residual within tolerance ε, where $\#\Lambda_{\tilde\Theta} \lesssim \#\Lambda_\Psi + \varepsilon^{-1/s}$.

The second ingredient behind the construction is the proof of $\tilde\Theta(\Xi)$ being close to a sparse matrix, in the sense that for arbitrary, fixed $\eta > 0$, for any multi-tree $\Lambda_{\tilde\Theta} \subset \nabla_{\tilde\Theta}$, there exists a multi-tree $\Lambda_\Xi \subset \nabla_\Xi$ with $\#\Lambda_\Xi \lesssim \#\Lambda_{\tilde\Theta}$ and $\|\tilde\Theta(\Xi)|_{\nabla_\Xi\setminus\Lambda_\Xi\times\Lambda_{\tilde\Theta}}\|_{\mathscr{L}(\ell_2(\nabla_{\tilde\Theta}),\ell_2(\nabla_\Xi))} \le \eta$. So with $\mathbf{f}_\varepsilon := f_\varepsilon(\Xi)$, the approximation $(f_\varepsilon - A((\mathbf{I}_{\Lambda_\Psi}\mathbf{w}_{\Lambda_\Psi})^\top\Psi))(\Xi)|_{\Lambda_\Xi} = (\mathbf{f}_\varepsilon - \mathbf{A}\mathbf{I}_{\Lambda_\Psi}\mathbf{w}_{\Lambda_\Psi})|_{\Lambda_\Xi}$ to $\mathbf{f}_\varepsilon - \mathbf{A}\mathbf{I}_{\Lambda_\Psi}\mathbf{w}_{\Lambda_\Psi}$ has a relative error of order η. By $\|f - f_\varepsilon\|_{K'} \eqsim \|\mathbf{f} - \mathbf{f}_\varepsilon\|_{\ell_2(\nabla_\Xi)}$, we arrive at the estimate $\|(\mathbf{f} - \mathbf{A}\mathbf{I}_{\Lambda_\Psi}\mathbf{w}_{\Lambda_\Psi})|_{\nabla_\Xi\setminus\Lambda_\Xi}\|_{\ell_2(\nabla_\Xi)} \lesssim \varepsilon + \eta\|(\mathbf{f} - \mathbf{A}\mathbf{I}_{\Lambda_\Psi}\mathbf{w}_{\Lambda_\Psi})\|_{\ell_2(\nabla_\Xi)}$ as required.

Remark 1 In view of the use of tensor product bases and the unidirectional principle to evaluate the matrix-vector multiplication, our adaptive wavelet algorithm has similarities with adaptive sparse grid methods. There are, however, also important differences. In general, within adaptive sparse grid algorithms new degrees of freedom are added in neighborhoods of those current degrees of freedom that correspond to the largest coefficients, e.g., see [1–5, 19, 17, 20, 24, 25, 31]. This is due to the fact that in (most cases) adaptive sparse grids are based on (infinite) tensor product collections that are not Riesz bases for the closures of their spans. Consequently, the norm of the residual vector is not equivalent to the norm of the error, and so the residual cannot be used as a reliable and efficient a posteriori error estimator to steer the adaptive process. Nevertheless, as shown in the above mentioned references, this adaptive approach improves upon non-adaptive sparse grid methods in various practical examples. However, other than with the adaptive wavelet algorithm, in general, convergence and so in particular optimal convergence rates cannot be guaranteed.

Organization

The remainder of this paper is organized as follows: In Section 2, a general criterion is derived for our approximate residual to realize (4). In Section 3, the conditions of this criterion are verified in the setting of tensor product bases and an operator that corresponds to a partial differential operator with polynomial coefficients. The results of various experiments illustrating the performance of the new approximate residual evaluation are presented in Section 4.

2 Approximate evaluation of the residual

We take up the issue outlined in the introduction, and present a method that realizes task (4).

Theorem 1 *For some (real) Hilbert spaces H and K, let $A \in \mathscr{L}(H, K')$. Let*

(A). $\Xi = \{\xi_\lambda : \lambda \in \nabla_\Xi\} \subset K$ *be such that* $\|[g(\xi_\lambda)]_{\lambda\in\nabla_\Xi}\|_{\ell_2(\nabla_\Xi)} \eqsim \|g\|_{K'}$ $(g \in K')$*, i.e., Ξ is a* frame *for Ξ, e.g., a Riesz basis,*

(B). $\tilde\Theta = \{\tilde\theta_\lambda : \lambda \in \nabla_{\tilde\Theta}\} \subset K'$ *be such that* $\|\tilde{\mathbf{z}}\|_{\ell_2(\nabla_{\tilde\Theta})} \lesssim \|\tilde{\mathbf{z}}^\top\tilde\Theta\|_{K'}$ $(\tilde{\mathbf{z}} \in \ell_2(\nabla_{\tilde\Theta}))$*,*

(C). $\Psi = \{\psi_\lambda : \lambda \in \nabla_\Psi\} \subset H$.

We consider classes of "admissible" finite subsets of ∇_Ξ, $\nabla_{\tilde\Theta}$, and ∇_Ψ, respectively, which classes should be closed under taking unions, such that

(D). *for any admissible $\Lambda_\Psi \subset \nabla_\Psi$, there exists an admissible $\Lambda_{\tilde\Theta} \subset \nabla_{\tilde\Theta}$ such that* $A(\operatorname{span}\{\psi_\lambda : \lambda \in \Lambda_\Psi\}) \subset \operatorname{span}\{\tilde\theta_\lambda : \lambda \in \Lambda_{\tilde\Theta}\}$*, and* $\#\Lambda_{\tilde\Theta} \lesssim \#\Lambda_\Psi$*,*

(E). *for any $\eta > 0$, and any admissible $\Lambda_{\tilde\Theta} \subset \nabla_{\tilde\Theta}$, there exists an admissible $\Lambda_\Xi \subset \nabla_\Xi$ with $\#\Lambda_\Xi \lesssim \#\Lambda_{\tilde\Theta}$, generally dependent on η, such that*

$$\|[\tilde\theta_\mu(\xi_\lambda)]_{\lambda\in\nabla_\Xi\setminus\Lambda_\Xi,\mu\in\Lambda_{\tilde\Theta}}\|_{\mathscr{L}(\ell_2(\nabla_{\tilde\Theta}),\ell_2(\nabla_\Xi))} \le \eta.$$

Now let $u \in H$ be such that for some $s > 0$, for any $\varepsilon > 0$, there exists an admissible $\Lambda_\Psi \subset \nabla_\Psi$ with $\#\Lambda_\Psi \lesssim \varepsilon^{-1/s}$ and $\inf_{v\in\operatorname{span}\{\psi_\lambda:\lambda\in\Lambda_\Psi\}} \|u-v\|_H \le \varepsilon$.

Then for $f = Au$, an arbitrary, fixed $\eta > 0$, and any admissible $\Lambda_\Psi \subset \nabla_\Psi$, $w_{\Lambda_\Psi} \in \operatorname{span}\{\psi_\lambda : \lambda \in \Lambda_\Psi\}$, and $\varepsilon > 0$, there exists an admissible $\Lambda_\Xi \subset \nabla_\Xi$ with

$$\begin{aligned}\|[(f-Aw_{\Lambda_\Psi})(\xi_\lambda)]_{\lambda\in\nabla_\Xi} - [(f-Aw_{\Lambda_\Psi})(\xi_\lambda)]_{\lambda\in\Lambda_\Xi}\|_{\ell_2(\nabla_\Xi)} \\ \lesssim \eta\|[(f-Aw_{\Lambda_\Psi})(\xi_\lambda)]_{\lambda\in\nabla_\Xi}\|_{\ell_2(\nabla_\Xi)} + (1+\eta)\varepsilon,\end{aligned} \tag{6}$$

and $\#\Lambda_\Xi \lesssim \#\Lambda_\Psi + \varepsilon^{-1/s}$ (generally dependent on η), i.e., task (4) is realized.

Proof By the assumption on u, and $f = Au$ with $A \in \mathscr{L}(H,K')$, Condition (D) implies that there exists an admissible $\Lambda^{(1)}_{\tilde\Theta} \subset \nabla_{\tilde\Theta}$ with $\#\Lambda^{(1)}_{\tilde\Theta} \lesssim \varepsilon^{-1/s}$, and an $f_\varepsilon \in \operatorname{span}\{\tilde\theta_\lambda : \lambda \in \Lambda^{(1)}_{\tilde\Theta}\}$ with $\|f - f_\varepsilon\|_{K'} \le \varepsilon$.

With $\Lambda^{(2)}_{\tilde\Theta}$ denoting the set from (D) corresponding to Λ_Ψ, let $\Lambda_{\tilde\Theta} = \Lambda^{(1)}_{\tilde\Theta} \cup \Lambda^{(2)}_{\tilde\Theta}$, and let Λ_Ξ correspond to $\Lambda_{\tilde\Theta}$ according to (E). Then we have $\#\Lambda_\Xi \lesssim \#\Lambda_\Psi + \varepsilon^{-1/s}$, and

$$\|[(f_\varepsilon - Aw_{\Lambda_\Psi})(\xi_\lambda)]_{\lambda\in\nabla_\Xi\setminus\Lambda_\Xi}\|_{\ell_2(\nabla_\Xi)} \lesssim \eta\|f_\varepsilon - Aw_{\Lambda_\Psi}\|_{K'} \le \eta\|f - Aw_{\Lambda_\Psi}\|_{K'} + \eta\varepsilon$$

by (E) and (B). The proof is completed by $\|[(f_\varepsilon - f)(\xi_\lambda)]_{\lambda\in\nabla_\Xi\setminus\Lambda_\Xi}\|_{\ell_2(\nabla_\Xi)} \lesssim \|f_\varepsilon - f\|_{K'} \le \varepsilon$ by (A), and $\|f_\varepsilon - Aw_{\Lambda_\Psi}\|_{K'} \lesssim \|[(f_\varepsilon - Aw_{\Lambda_\Psi})(\xi_\lambda)]_{\lambda\in\nabla_\Xi}\|_{\ell_2(\nabla_\Xi)}$ again by (A).

Remark 2 For completeness, to show that with Theorem 1 indeed task (4) is realized, note that (6) means that for some constant $C > 0$,

$$\begin{aligned}\|[(f-Aw_{\Lambda_\Psi})(\xi_\lambda)]_{\lambda\in\nabla_\Xi} - [(f-Aw_{\Lambda_\Psi})(\xi_\lambda)]_{\lambda\in\Lambda_\Xi}\|_{\ell_2(\nabla_\Xi)} \\ \le C[\eta\|[(f-Aw_{\Lambda_\Psi})(\xi_\lambda)]_{\lambda\in\nabla_\Xi}\|_{\ell_2(\nabla_\Xi)} + (1+\eta)\varepsilon].\end{aligned} \tag{7}$$

Now given $\bar\eta, \bar\varepsilon > 0$, set $\eta = \frac{\bar\eta}{1+C}$, $\varepsilon = \frac{\bar\varepsilon}{C(C+1)}/(1+\eta)$. By substituting these η and ε into the upper bound (7), for $\bar\varepsilon \ge \bar\eta\|[(f-Aw_{\Lambda_\Psi})(\xi_\lambda)]_{\lambda\in\nabla_\Xi}\|_{\ell_2(\nabla_\Xi)}$ we find that this upper bound is less or equal to $\bar\varepsilon$ as required.

Remark 3 The verification of Conditions (D) and (E) in the next section gives information, in any case qualitatively, how to *construct* $\Lambda^{(2)}_{\tilde\Theta}$ from Λ_Ψ, and Λ_Ξ from $\Lambda_{\tilde\Theta} = \Lambda^{(1)}_{\tilde\Theta} \cup \Lambda^{(2)}_{\tilde\Theta}$. The construction of $\Lambda^{(1)}_{\tilde\Theta}$ depends on the right-hand side at hand, which is, however, known to the user. Fortunately, in many applications it suffices to take $\Lambda_{\tilde\Theta} = \Lambda^{(2)}_{\tilde\Theta}$.

Remark 4 Condition (B) is satisfied when there exists a $\Theta = \{\theta_\lambda : \lambda \in \nabla_{\tilde\Theta}\} \subset K$ with $\tilde\Theta(\Theta) = \mathrm{Id}$ and $\|\mathbf{z}^\top\Theta\|_K \lesssim \|\mathbf{z}\|_{\ell_2(\nabla_{\tilde\Theta})}$ ($\mathbf{z} \in \ell_2(\nabla_{\tilde\Theta})$) (so, in particular when Θ is a Riesz basis for K). Indeed, then $\|\tilde{\mathbf{z}}\|_{\ell_2(\nabla_{\tilde\Theta})} = \sup_{0\ne\mathbf{z}\in\ell_2(\nabla_{\tilde\Theta})} \frac{\langle\tilde{\mathbf{z}},\mathbf{z}\rangle_{\ell_2(\nabla_{\tilde\Theta})}}{\|\mathbf{z}\|_{\ell_2(\nabla_{\tilde\Theta})}} = \sup_{0\ne\mathbf{z}\in\ell_2(\nabla_{\tilde\Theta})} \frac{(\tilde{\mathbf{z}}^\top\tilde\Theta)(\mathbf{z}^\top\Theta)}{\|\mathbf{z}\|_{\ell_2(\nabla_{\tilde\Theta})}} \lesssim \|\tilde{\mathbf{z}}^\top\tilde\Theta\|_{K'}$. Both $\tilde\Theta$ and Θ are auxiliary collections that will not enter the computation.

Remark 5 In the proof of Theorem 1, in order to conclude that if u can be approximated in H from $\operatorname{span}\Psi$ at rate s, then so can f in K' from $\operatorname{span}\tilde\Theta$, we assumed that $f = Au$. In the setting of least squares problems discussed in the introduction, the assumption that $f = Au$ requires that the problem is consistent in the sense that $f \in \mathfrak{I}A$. When this does not hold true, an approximation rate s for the right-hand side has to be added as a separate condition.

3 Application in the tensor product setting

We verify the conditions of Theorem 1 for A being a linear partial differential operator having polynomial coefficients on a rectangular domain, and collections of tensor product functions.

3.1 Assumptions and verification of Condition (D)

For some $n \in \mathbb{N}$, and $\square := \prod_{i=1}^n (a_i, b_i)$ for some $a_i < b_i \in \mathbb{Z} \cup \{-\infty, \infty\}$, let $A \in \mathscr{L}(H, K')$, where

$$H, K \hookrightarrow L_2(\square) \simeq L_2(\square)' \hookrightarrow K'.$$

For some invertible diagonal matrices $D^\Psi = (d^\Psi_{\vec{\lambda}})_{\vec{\lambda}}$, $D^{\tilde{\Theta}} = (d^{\tilde{\Theta}}_{\vec{\lambda}})_{\vec{\lambda}}$, and $D^\Xi = (d^\Xi_{\vec{\lambda}})_{\vec{\lambda}}$, we assume that the collections $\Psi \subset H$, $\tilde{\Theta}$, and Ξ are of type

$$D^\Psi \Psi = \otimes_{i=1}^n \Psi^{(i)}, \Psi^{(i)} = \{\psi^{(i)}_\lambda : \lambda \in \nabla_{\Psi^{(i)}}\}, d^\Psi_{\vec{\lambda}} \psi_{\vec{\lambda}} := \bigotimes_{i=1}^n \psi^{(i)}_{\lambda_i} \ (\vec{\lambda} \in \nabla_\Psi := \prod_{i=1}^n \nabla_{\Psi^{(i)}}),$$

$$D^{\tilde{\Theta}} \tilde{\Theta} = \otimes_{i=1}^n \tilde{\Theta}^{(i)}, \tilde{\Theta}^{(i)} = \{\tilde{\theta}^{(i)}_\lambda : \lambda \in \nabla_{\tilde{\Theta}^{(i)}}\}, \ d^{\tilde{\Theta}}_{\vec{\lambda}} \tilde{\theta}_{\vec{\lambda}} := \bigotimes_{i=1}^n \tilde{\theta}^{(i)}_{\lambda_i} \ (\vec{\lambda} \in \nabla_{\tilde{\Theta}} := \prod_{i=1}^n \nabla_{\tilde{\Theta}^{(i)}}),$$

$$D^\Xi \Xi = \otimes_{i=1}^n \Xi^{(i)}, \ \Xi^{(i)} = \{\xi^{(i)}_\lambda : \lambda \in \nabla_{\Xi^{(i)}}\}, \ d^\Xi_{\vec{\lambda}} \ \xi_{\vec{\lambda}} := \bigotimes_{i=1}^n \xi^{(i)}_{\lambda_i} \ (\vec{\lambda} \in \nabla_\Xi := \prod_{i=1}^n \nabla_{\Xi^{(i)}}),$$

where $\Psi^{(i)}, \tilde{\Theta}^{(i)}, \Xi^{(i)} \subset L_2(a_i, b_i)$ are collections of functions *normalized in* $L_2(a_i, b_i)$.

We assume that $\Psi^{(i)}$, $\tilde{\Theta}^{(i)}$, and $\Xi^{(i)}$ are *local* in the sense that

$$\sup_{\lambda \in \nabla_{\Psi^{(i)}}} 2^{|\lambda|} \operatorname{diam} \operatorname{supp} \psi^{(i)}_\lambda < \infty, \tag{8}$$

$$\sup_{\ell \in \mathbb{N}_0} \sup_{x \in (a_i, b_i)} \#\{\lambda \in \nabla_{\Psi^{(i)}} : |\lambda| = \ell \wedge \operatorname{supp} \psi^{(i)}_\lambda \cap B(x; 2^{-\ell}) \neq \emptyset\} < \infty, \tag{9}$$

where $|\lambda| \in \mathbb{N}_0$ denotes the *level* of λ, and similarly for $\tilde{\Theta}^{(i)}$ and $\Xi^{(i)}$.

We assume that for $1 \leq i \leq n$, there exists another *local* collection $\Theta^{(i)} \subset L_2(a_i, b_i)$ that is *dual to* $\tilde{\Theta}^{(i)}$. Consequently,

$$\Theta := D^{\tilde{\Theta}} \otimes_{i=1}^n \Theta^{(i)}$$

is dual to $\tilde{\Theta}$.

We add the harmless assumptions that for all $\ell \in \mathbb{N}_0$

$$\bigcup_{\{\lambda \in \nabla_{\Psi^{(i)}} : |\lambda| = \ell\}} \operatorname{supp} \psi^{(i)}_\lambda = (a_i, b_i), \tag{10}$$

$$\bigcup_{\{\lambda \in \nabla_{\Xi^{(i)}} : |\lambda| = \ell\}} \operatorname{supp} \xi^{(i)}_\lambda = (a_i, b_i), \tag{11}$$

and, with $S^{(i)}_\lambda := \operatorname{supp} \theta^{(i)}_\lambda \cup \operatorname{supp} \tilde{\theta}^{(i)}_\lambda \quad (\lambda \in \nabla^{(i)}_{\tilde{\Theta}})$,

$$\bigcup_{\{\lambda \in \nabla_{\tilde{\Theta}^{(i)}} : |\lambda| = \ell\}} S^{(i)}_\lambda = (a_i, b_i). \tag{12}$$

Finally, we make the assumptions that for some $p \in \mathbb{N}_0$,

$$\begin{aligned}
&\operatorname{span}\{\psi_\lambda^{(i)} : \lambda \in \nabla_{\Psi^{(i)}}, |\lambda| = \ell\} \\
&\subset \big\{g : (a_i, b_i) \to \mathbb{R} : g|_{(j2^{-\ell},(j+1)2^{-\ell})} \in P_p,\ j \in [a_i 2^\ell, b_i 2^\ell - 1] \cap \mathbb{Z}\big\} \qquad (13)\\
&\subset \operatorname{span}\{\tilde\theta_\lambda^{(i)} : \lambda \in \nabla_{\tilde\Theta^{(i)}}, |\lambda| \le \ell\}, \qquad (14)
\end{aligned}$$

with the assumption (14) concerning the inclusion of all (*discontinuous*) piecewise polynomials being crucial.

We call a collection $\Lambda \subset \nabla_{\Psi^{(i)}}$ a *tree*, if whenever $\lambda \in \nabla_{\Psi^{(i)}}$ with $|\lambda| > 0$ is in Λ, then the support of $\psi_\lambda^{(i)}$ is covered by the supports of $\psi_\mu^{(i)}$ for some $\mu \in \Lambda$ with $|\mu| = |\lambda| - 1$. The definition of a tree $\Lambda \subset \nabla_{\Xi^{(i)}}$ is similar, whereas for the definition of a tree $\Lambda \subset \nabla_{\tilde\Theta^{(i)}}$, $\operatorname{supp}\tilde\theta_\mu^{(i)}$ and $\operatorname{supp}\tilde\theta_\lambda^{(i)}$ should be read as $S_\mu^{(i)}$ and $S_\lambda^{(i)}$, respectively.

A collection $\Lambda \subset \nabla_\Psi$ is called a *multi-tree*, and similar for $\Lambda \subset \nabla_{\tilde\Theta}$ or $\Lambda \subset \nabla_\Xi$, if when "frozen" in any $n-1$ coordinates it is a tree in the remaining coordinate. In the current setting, as admissible subsets of ∇_Ψ, $\nabla_{\tilde\Theta}$, or ∇_Ξ mentioned in Theorem 1, we consider multi-trees.

Remark 6 The restriction to index sets that are multi-trees is harmless in the sense that it still allows to approximate a wide class of functions, that includes solutions of PDEs, at optimal, n-independent rates (cf. [6, Thm. 3.2]).

Proposition 1 *In addition to the assumptions of $A \in \mathscr{L}(H, K')$, $\Psi \subset H$, the collections $\Psi^{(i)}$, $\tilde\Theta^{(i)}$ and its dual $\Theta^{(i)}$ being local, and* (10), (13)*, and* (14)*, let, for some densely embedded $\tilde H \subset H$, $A(\tilde H) \subset L_2(\square)$, and $A|_{\tilde H}$ be a* linear partial differential operator with constant coefficients*. Let* $\operatorname{span}\Psi \subset \tilde H$*. Then for any multi-tree $\Lambda_\Psi \subset \nabla_\Psi$, there exists a multi-tree $\Lambda_{\tilde\Theta} \subset \nabla_{\tilde\Theta}$ with $\#\Lambda_{\tilde\Theta} \lesssim \#\Lambda_\Psi$ and*

$$A(\operatorname{span}\{\psi_{\vec\lambda} : \vec\lambda \in \Lambda_\Psi\}) \subset \operatorname{span}\{\tilde\theta_{\vec\mu} : \vec\mu \in \Lambda_{\tilde\Theta}\}, \qquad (15)$$

i.e., condition (D) from Theorem 1 is valid.

Proof With $C_i := \sup_{\mu \in \nabla_{\tilde\Theta^{(i)}}} 2^{|\mu|} \operatorname{diam} S_\mu^{(i)}$, given a multi-tree $\Lambda_\Psi \subset \nabla_\Psi$, we take

$$\Lambda_{\tilde\Theta} := \{\vec\mu \in \nabla_{\tilde\Theta} : \exists \vec\lambda \in \Lambda_\Psi \text{ s.t. } |\vec\mu| = |\vec\lambda| \wedge \mathrm{d}(\operatorname{supp}\psi_{\lambda_i}^{(i)}, S_{\mu_i}^{(i)}) \le C_i 2^{-|\mu_i|}\,(\forall i)\},$$

where $|\vec\gamma| := (|\gamma_1|, \dots, |\gamma_n|)$, and for sets $X, Y \subset \mathbb{R}^n$, $\mathrm{d}(X,Y) := \inf_{x\in X, y\in Y} |x-y|$. Note that in view of (10), and since Λ_Ψ is a multi-tree, $\Lambda_{\tilde\Theta}$ does not change when the condition $|\vec\mu| = |\vec\lambda|$ reads as $|\vec\mu| \le |\vec\lambda|$.

The assumptions of $A|_{\tilde H}$ being a linear partial differential operator with constant coefficients, $A(\operatorname{span}\Psi) \subset L_2(\square)$, and (13)-(14) show that for $\vec\lambda \in \nabla_\Psi$, one has $A\psi_{\vec\lambda} \in \operatorname{span}\{\tilde\theta_{\vec\mu} : \vec\mu \in \nabla_{\tilde\Theta}, |\vec\mu| \le |\vec\lambda|\}$ and $\operatorname{supp} A\psi_{\vec\lambda} \subseteq \operatorname{supp}\psi_{\vec\lambda}$.

Now let $v \in \operatorname{span}\{\psi_{\vec\lambda} : \vec\lambda \in \Lambda_\Psi\}$, i.e., $v = \sum_{\vec\lambda\in\Lambda_\Psi} v_{\vec\lambda}\psi_{\vec\lambda}$. From A being linear, and Θ being dual to $\tilde\Theta$, we have

$$Av = \sum_{\{\vec\mu\in\nabla_{\tilde\Theta} : \|\vec\mu\|_\infty \le \max_{\vec\lambda\in\Lambda_\Psi}\|\vec\lambda\|_\infty\}} \sum_{\vec\lambda\in\Lambda_\Psi} v_{\vec\lambda}\,\langle A\psi_{\vec\lambda}, \theta_{\vec\mu}\rangle_{L_2(\square)}\,\tilde\theta_{\vec\mu}.$$

Now let $\vec{\mu} \notin \Lambda_{\tilde{\Theta}}$. Then for all $\vec{\lambda} \in \Lambda_\Psi$ there exists an i with either $|\mu_i| > |\lambda_i|$ or $|\operatorname{supp} \psi^{(i)}_{\lambda_i} \cap S^{(i)}_{\mu_i}| = 0$, and so $\langle A\psi_{\vec{\lambda}}, \theta_{\vec{\mu}}\rangle_{L_2(\square)} = 0$ from $\operatorname{supp} \theta^{(i)}_{\mu_i} \subseteq S^{(i)}_{\mu_i}$. We conclude that (15) is valid.

To show that $\Lambda_{\tilde{\Theta}}$ is a multi-tree, let $\vec{\mu} \in \Lambda_{\tilde{\Theta}}$, $\vec{\gamma} \in \nabla_{\tilde{\Theta}}$ with, for some i, $\gamma_j = \mu_j$ for $j \neq i$, and $|\gamma_i| = |\mu_i| - 1$ and $|S^{(i)}_{\mu_i} \cap S^{(i)}_{\gamma_i}| > 0$. There exists a $\vec{\lambda} \in \Lambda_\Psi$ with $|\vec{\mu}| \leq |\vec{\lambda}|$ and $\mathrm{d}(\operatorname{supp} \psi^{(i)}_{\lambda_i}, S^{(i)}_{\mu_i}) \leq C_i 2^{-|\mu_i|}$ forall i. We infer that $\mathrm{d}(\operatorname{supp} \psi^{(i)}_{\lambda_i}, S^{(i)}_{\gamma_i}) \leq \mathrm{d}(\operatorname{supp} \psi^{(i)}_{\lambda_i}, S^{(i)}_{\mu_i}) + \operatorname{diam} S^{(i)}_{\mu_i} \leq C_i 2^{-|\mu_i|} + C_i 2^{-|\mu_i|} = C_i 2^{-|\gamma_i|}$, and so $\vec{\gamma} \in \Lambda_{\tilde{\Theta}}$. Now apply (12).

By the locality of $\tilde{\Theta}^{(i)}$, $\Theta^{(i)}$, and $\Psi^{(i)}$, the definition of $\Lambda_{\tilde{\Theta}}$ shows that $\#\Lambda_{\tilde{\Theta}} \lesssim \#\Lambda_\Psi$, which completes the proof.

Example 1 With $H = K = H^1_0(\square)$, a model example of an A as in Proposition 1 is given by $(Au)(v) = \int_\Omega \nabla u \cdot \nabla v + \omega uv$ for some $\omega \geq 0$, and $\omega \gtrsim 1$ when $\square$ is unbounded (see [22] for the application of adaptive wavelet schemes for such domains). With $\tilde{H} := H^2(\square) \cap H^1_0(\square)$, integration by parts shows that $A(\tilde{H}) \subset L_2(\square)$ and $A|_{\tilde{H}} = -\Delta + \omega I$. Note that for the latter two properties it is essential that the boundary conditions on $\partial\square$ are of *homogeneous Dirichlet type*.

Remark 7 If $A|_{\tilde{H}}$ is a linear partial differential operator with *polynomial* coefficients, say of maximal degree $\tilde{p}$, then the statement of Proposition 1 is still valid, assuming that (14) holds for p reading as $p + \tilde{p}$. We expect that for any fixed $p + \tilde{p}$ suitable biorthogonal collections $(\Theta^{(i)}, \tilde{\Theta}^{(i)})$ can be found such that this second inclusion is valid, cf. Remark 8.

If, for some *fixed* $k \in \mathbb{N}$, (14) is valid with $\operatorname{span}\{\tilde{\theta}^{(i)}_\lambda : |\lambda| \leq \ell\}$ reading as $\operatorname{span}\{\tilde{\theta}^{(i)}_\lambda : |\lambda| \leq \ell - k\}$, i.e., if the coarsest scale in the $(\Theta^{(i)}, \tilde{\Theta}^{(i)})$ system is increased with k levels, then the statement of Proposition 1 is still valid when the coefficients of the partial differential operator are *piecewise* polynomial w.r.t. the partition $\bar{\square} = \cup_{\{\alpha \in \mathbb{Z}^n : 2^{-k}(\alpha + \bar{\square}) \subset \bar{\square}\}} 2^{-k}(\alpha + \bar{\square})$.

It is tempting to combine both above generalizations, and to approximate general smooth coefficients of a partial differential operator by piecewise polynomials. Yet, since for the multi-trees Λ^Ψ and $\Lambda^{\tilde{\Theta}}$ from Proposition 1 it holds that $\#\Lambda^{\tilde{\Theta}}/\#\Lambda^\Psi \to \infty$ when $\tilde{p}$ or k tend to infinity, it is not clear whether this yields some useful results.

3.2 Further assumptions and verification of Conditions (A), (B), and (E) of Theorem 1

We will consider a space K of the following form. For $1 \leq i \leq n$, for some $m_i \in \mathbb{N}_0$ and $\vec{\sigma}_i = (\sigma_{i,l}, \sigma_{i,r}) \in \{0, \ldots, m_i\}^2$, let

$$H^{m_i}_{\vec{\sigma}_i}(a_i, b_i) := \{v \in H^{m_i}(a_i, b_i) : v(a_i) = \cdots = v^{(\sigma_{i,l}-1)}(a_i) = 0, \\ v(b_i) = \cdots = v^{(\sigma_{i,r}-1)}(b_i) = 0\},$$

$$\mathscr{H}^s_i := [L_2(a_i, b_i), H^{m_i}_{\vec{\sigma}_i}(a_i, b_i)]_{s/m_i} \quad (s \in [0, m_i]).$$

For some finite $I \subset \prod_{i=1}^n [0, m_i]$, let

$$K = \bigcap_{\vec{q} \in I} \bigotimes_{i=1}^n \mathscr{H}^{q_i}_i, \tag{16}$$

equipped with the usual norm on an intersection space. Examples of applications will be given in Subsection 3.3.

We assume that

$$\|\sum_{\mu\in\nabla_{\tilde\Theta^{(i)}}} z_\mu 2^{-|\mu|s}\theta_\mu^{(i)}\|^2_{\mathscr{H}_i^s} \lesssim \sum_{\mu\in\nabla_{\tilde\Theta^{(i)}}} |z_\mu|^2 \quad (1\le i\le n, s\in\{0,m_i\}). \tag{17}$$

Then, taking

$$(d_{\vec\mu}^{\tilde\Theta})^{-2} \eqsim \sum_{q\in I}\prod_{i=1}^n 4^{|\mu_i|q_i}, \quad (\vec\mu\in\nabla_{\tilde\Theta}), \tag{18}$$

we infer that $\|\mathbf{z}^\top\Theta\|_K \lesssim \|\mathbf{z}\|_{\ell_2(\nabla_{\tilde\Theta})}$, and so (B) is valid by Remark 4. (For (B), it suffices when (18) holds with "$\eqsim$" reading as "$\gtrsim$", whereas for the proof of (E) the inequality "$\lesssim$" will be used.)

Additionally, we impose the (mild) piecewise smoothness conditions

$$\sup_{i,\mu\in\nabla_{\tilde\Theta^{(i)}}} 2^{-|\mu|/2}\|\tilde\theta_\mu^{(i)}\|_{L_\infty(a_i,b_i)} < \infty, \tag{19}$$

$$\sup_{i,\mu\in\nabla_{\tilde\Theta^{(i)}}} 2^{-3|\mu|/2} \sup_{\{j\in\mathbb{Z}:(j2^{-|\mu|},(j+1)2^{-|\mu|})\in(a_i,b_i)\}} \|\tilde\theta_\mu^{(i)}\|_{W^1_\infty(j2^{-|\mu|},(j+1)2^{-|\mu|})} < \infty, \tag{20}$$

at the "dual" side, the Jackson estimate

$$\inf_{w\in\mathrm{span}\{\theta_\mu^{(i)}:|\mu|\le\ell\}} \|u-w\|_{L_2(a_i,b_i)} \lesssim 2^{-\ell m_i}\|u\|_{H^{m_i}(a_i,b_i)} \quad (u\in H^{m_i}_{\vec\sigma_i}(a_i,b_i)) \tag{21}$$

at the "primal" side, and

$$\|\sum_{\{\mu\in\nabla_{\tilde\Theta^{(i)}}:|\mu|=\ell\}} \tilde z_\mu \tilde\theta_\mu{}^{(i)}\|^2_{L_2(a_i,b_i)} \lesssim \sum_{\{\mu\in\nabla_{\tilde\Theta^{(i)}}:|\mu|=\ell\}} |\tilde z_\mu|^2, \tag{22}$$

uniformly for all $\ell\in\mathbb{N}_0$.

Remark 8 Auxiliary biorthogonal collections $(\Theta^{(i)},\tilde\Theta^{(i)})$ that are both local, satisfy (12), (14), (17), (19), (20), (21), and (22) have been constructed in [7] for $m_i=2$, $\vec\sigma_i\in\{0,1\}^2$, and $p=4$, and so obviously for $m_i\le 2$, $\vec\sigma_i\in\{0,\min(1,m_i)\}^2$, and $p\le 4$. (The case $\vec\sigma_i=(1,1)$ considered in [7] is extended to $\vec\sigma_i\in\{0,1\}^2$ in [6, §7]). Instead of (17) and (22), for $s\in\{0,m_i\}$, $\{2^{-|\mu|s}\theta_\mu^{(i)}:\mu\in\nabla_{\tilde\Theta^{(i)}}\}$ is even a *Riesz basis* for $\mathscr{H}_i^s$. We expect that such $(\Theta^{(i)},\tilde\Theta^{(i)})$ exist for any $m_i\in\mathbb{N}_0$, $\vec\sigma_i\in\{0,\dots,m_i\}^2$ and $p\in\mathbb{N}_0$.

Concerning Ξ, we assume that

$$\Xi^{(i)}\subset H^{m_i}_{\vec\sigma_i}(a_i,b_i), \tag{23}$$

so that $\Xi\subset K$, the Bernstein inequality

$$\|\cdot\|_{H^{m_i}(a_i,b_i)} \lesssim 2^{\ell m_i}\|\cdot\|_{L_2(a_i,b_i)} \quad \text{on } \mathrm{span}\{\xi_\lambda^{(i)}:|\lambda|=\ell\}, \tag{24}$$

$$\|\sum_{\{\lambda\in\nabla_{\Xi^{(i)}}:|\lambda|=\ell\}} z_\lambda\xi_\lambda^{(i)}\|^2_{L_2(a_i,b_i)} \lesssim \sum_{\{\lambda\in\nabla_{\Xi^{(i)}}:|\lambda|=\ell\}} |z_\lambda|^2, \tag{25}$$

uniformly for all $\ell\in\mathbb{N}_0$,

$$(d_{\vec\lambda}^{\Xi})^2 \gtrsim \sum_{\vec q\in I}\prod_{i=1}^n 4^{|\lambda_i|q_i} \quad (\vec\lambda\in\nabla_\Xi), \tag{26}$$

and

$$\int_{a_i}^{b_i} \xi_\lambda^{(i)} dx = 0 \quad (i, \lambda \in \nabla_{\Xi^{(i)}} \text{ with } |\lambda| > 0), \tag{27}$$

the latter possibly with the exception of those λ for which $\mathrm{d}(\{a_i, b_i\}, \operatorname{supp} \xi_\lambda^{(i)}) \lesssim 2^{-|\lambda|}$.

Remark 9 If, for $s \in \{0, m_i\}$, $\{2^{-|\lambda|s} \xi_\lambda^{(i)} : \lambda \in \nabla_{\Xi^{(i)}}\}$ is a Riesz basis for $\mathscr{H}_i^s$, then (24) and (25) are valid (even with "$\eqsim$"). Taking $(d_\lambda^\Xi)^2 \eqsim \sum_{\vec{q} \in I} \prod_{i=1}^n 4^{|\lambda_i| q_i}$, (26) is valid and Ξ is a Riesz basis for K, so that in particular Condition (A) is valid. The condition (27) of the functions to have at least one vanishing moment, with the exception of functions on the coarsest level and possibly those with supports near the boundary, is a standard condition on a wavelet basis.

Proposition 2 *Assume* (12)*, the locality of* $\Xi^{(i)}, \Theta^{(i)}$*, and* $\tilde{\Theta}^{(i)}$*, and* (17)–(27)*. Then, for any* $\eta > 0$*, and any multi-tree* $\Lambda_{\tilde{\Theta}} \subset \nabla_{\tilde{\Theta}}$*, there exists a multi-tree* $\Lambda_\Xi \subset \nabla_\Xi$ *with* $\#\Lambda_\Xi \lesssim \#\Lambda_{\tilde{\Theta}}$*, generally dependent on* η*, such that*

$$\| [\langle \xi_{\vec{\lambda}}, \tilde{\theta}_{\vec{\mu}} \rangle_{L_2(\square)}]_{\vec{\lambda} \in \nabla_\Xi \setminus \Lambda_\Xi, \vec{\mu} \in \Lambda_{\tilde{\Theta}}} \|_{\mathscr{L}(\ell_2(\nabla_{\tilde{\Theta}}), \ell_2(\nabla_\Xi))} \leq \eta,$$

i.e., condition (E) of Theorem 1 is valid.

Proof With $D_i := \sup_{\lambda \in \nabla_{\Xi^{(i)}}} 2^{|\lambda|} \operatorname{diam} \operatorname{supp} \xi_\lambda^{(i)}$, for a multi-tree $\Lambda_{\tilde{\Theta}} \subset \nabla_{\tilde{\Theta}}$ and $\ell \in \mathbb{N}_0$, let

$$\Lambda_\Xi := \{\vec{\lambda} \in \nabla_\Xi : \exists \vec{\mu} \in \Lambda_{\tilde{\Theta}} \text{ s.t. } \forall i, |\lambda_i| \leq |\mu_i| + \ell \wedge \mathrm{d}(\operatorname{supp} \xi_{\lambda_i}^{(i)}, S_{\mu_i}^{(i)}) \leq D_i 2^{-|\lambda_i|}\}. \tag{28}$$

Let $\vec{\lambda} \in \Lambda_\Xi$ with for some i, $|\lambda_i| > 0$, and let $\vec{\mu} \in \Lambda_{\tilde{\Theta}}$ correspond to $\vec{\lambda}$ as in the definition of Λ_Ξ. From $\Lambda_{\tilde{\Theta}}$ being a multi-tree, we infer that there exists a $\vec{\nu} \in \nabla_{\tilde{\Theta}}$ with $\nu_j = \mu_j$ for $j \neq i$, $|\nu_i| = |\mu_i| - 1$, $\mathrm{d}(\operatorname{supp} \xi_{\lambda_i}^{(i)}, S_{\nu_i}^{(i)}) \leq D_i 2^{-|\lambda_i|}$, and $|S_{\nu_i}^{(i)} \cap S_{\mu_i}^{(i)}| > 0$. Now for any $\gamma_i \in \nabla_{\Xi^{(i)}}$ with $|\gamma_i| = |\lambda_i| - 1$ and $|\operatorname{supp} \xi_{\gamma_i}^{(i)} \cap \operatorname{supp} \xi_{\lambda_i}^{(i)}| > 0$, we have $\mathrm{d}(\operatorname{supp} \xi_{\gamma_i}^{(i)}, S_{\nu_i}^{(i)}) \leq \operatorname{diam} \operatorname{supp} \xi_{\lambda_i}^{(i)} + \mathrm{d}(\operatorname{supp} \xi_{\lambda_i}^{(i)}, S_{\nu_i}^{(i)}) \leq D_i 2^{-|\gamma_i|}$, showing that $\vec{\gamma} := (\gamma_1, \ldots, \gamma_n)$ with $\gamma_j = \lambda_j$ for $j \neq i$ is in Λ_Ξ. From (11), we conclude that Λ_Ξ is a multi-tree.

The locality of $\Xi^{(i)}$, $\Theta^{(i)}$, and $\tilde{\Theta}^{(i)}$ shows that $\#\Lambda_\Xi \lesssim \#\Lambda_{\tilde{\Theta}}$, dependent on ℓ.

In order to prove the last statement, with $I_{\vec{k},\vec{k}'} := [\langle \xi_{\vec{\lambda}}, \tilde{\theta}_{\vec{\mu}} \rangle_{L_2(\square)}]_{|\vec{\lambda}|=\vec{k}, |\vec{\mu}|=\vec{k}'}$, it is sufficient to show that for sufficiently large ℓ,

$$\sup_{\vec{k}} \sum_{\{\vec{k}' : \exists i \text{ s.t. } k_i > k_i' + \ell\}} \|I_{\vec{k},\vec{k}'}\| \times \sup_{\vec{k}'} \sum_{\{\vec{k} : \exists i \text{ s.t. } k_i > k_i' + \ell\}} \|I_{\vec{k},\vec{k}'}\| \leq \eta. \tag{29}$$

Indeed, for the corresponding block partitioning of $[\langle \xi_{\vec{\lambda}}, \tilde{\theta}_{\vec{\mu}} \rangle_{L_2(\square)}]_{\vec{\lambda} \in \nabla_\Xi \setminus \Lambda_\Xi, \vec{\mu} \in \Lambda_{\tilde{\Theta}}}$, denoted by $[\hat{I}_{\vec{k},\vec{k}'}]_{\vec{k},\vec{k}'}$, one has that $\hat{I}_{\vec{k},\vec{k}'}$ is equal to $I_{\vec{k},\vec{k}'}$ possibly with some columns or rows being removed. In particular, from $\operatorname{supp} \tilde{\theta}_{\mu_i}^{(i)} \subseteq S_{\mu_i}^{(i)}$ one has that $\hat{I}_{\vec{k},\vec{k}'} = 0$ when $k_i \leq k_i' + \ell$ for all i. Now an application of the Schur lemma shows that $\| [\langle \xi_{\vec{\lambda}}, \tilde{\theta}_{\vec{\mu}} \rangle_{L_2(\square)}]_{\vec{\lambda} \in \nabla_\Xi \setminus \Lambda_\Xi, \vec{\mu} \in \Lambda_{\tilde{\Theta}}} \|_{\mathscr{L}(\ell_2(\nabla_{\tilde{\Theta}}), \ell_2(\nabla_\Xi))}$ is less or equal to the left hand side from (29).

With $I_{k,k'}^{(i)} := [\langle \xi_\lambda^{(i)}, \tilde{\theta}_\mu^{(i)} \rangle_{L_2(a_i,b_i)}]_{|\lambda|=k, |\mu|=k'}$, we have

$$I_{\vec{k},\vec{k}'} = [1/d_{\vec{\lambda}}^\Xi]_{|\vec{\lambda}|=\vec{k}} \circ \bigotimes_{i=1}^n I_{k_i,k_i'}^{(i)} \circ [1/d_{\vec{\mu}}^{\tilde{\Theta}}]_{|\vec{\mu}|=\vec{k}'},$$

and so by (18) and (26),

$$\|I_{\vec{k},\vec{k}'}\| \lesssim \frac{\sqrt{\sum_{\vec{q}\in I}\prod_{i=1}^n 4^{k_i' q_i}}}{\sqrt{\sum_{\vec{q}\in I}\prod_{i=1}^n 4^{k_i q_i}}}\prod_{i=1}^n \|I^{(i)}_{k_i,k_i'}\|. \tag{30}$$

For $k \geq k'$, we split $I^{(i)}_{k,k'}$ into $I^{(i,s)}_{k,k'}$ and $I^{(i,r)}_{k,k'}$, where the first matrix contains the "singular" entries $(I^{(i)}_{k,k'})_{\lambda,\mu}$ of $I^{(i)}_{k,k'}$ with $(\operatorname{supp}\xi^{(i)}_\lambda)^\circ \cap 2^{-k'}\mathbb{Z} \neq \emptyset$ or $\operatorname{d}(\{a_i,b_i\},\operatorname{supp}\xi^{(i)}_\lambda) \lesssim 2^{-|\lambda|}$ (cf. line following (27)), and the second one contains the remaining "regular" entries.

The matrix $I^{(i,s)}_{k,k'}$ is sparse, uniformly in k,k', and from (19) and the locality of $\Xi^{(i)}$, for $|\lambda| = k$, $|\mu| = k'$, we have

$$|\langle \xi^{(i)}_\lambda, \tilde{\theta}^{(i)}_\mu\rangle_{L_2(a_i,b_i)}| \leq \|\xi^{(i)}_\lambda\|_{L_1(a_i,b_i)}\|\tilde{\theta}^{(i)}_\mu\|_{L_\infty(a_i,b_i)} \lesssim 2^{\frac{1}{2}(k'-k)}. \tag{31}$$

For the regular entries with $|\lambda| = k$, $|\mu| = k'$, from (27), the locality of $\Xi^{(i)}$, and (20), we have

$$\begin{aligned}|\langle \xi^{(i)}_\lambda, \tilde{\theta}^{(i)}_\mu\rangle_{L_2(a_i,b_i)}| &\lesssim \|\xi^{(i)}_\lambda\|_{L_1(a_i,b_i)} 2^{-k} \sup_{j\in[a_i2^{k'},b_i2^{k'}-1]\cap\mathbb{Z}} \|\tilde{\theta}^{(i)}_\mu\|_{W^1_\infty(j2^{-k'},(j+1)2^{-k'})} \\ &\lesssim 2^{\frac{3}{2}(k'-k)}.\end{aligned} \tag{32}$$

Using, by the locality of $\tilde{\Theta}^{(i)}$ and $\Xi^{(i)}$, that the number of non-zeros per row and column of $I^{(i,r)}_{k,k'}$ is $\mathscr{O}(1)$ or $\mathscr{O}(2^{k-k'})$, from the Schur lemma we infer that $\|I^{(i,r)}_{k,k'}\| \lesssim 2^{(k'-k)}$. We conclude that

$$\|I^{(i)}_{k,k'}\| \lesssim 2^{\frac{1}{2}(k'-k)} \quad (k \geq k'). \tag{33}$$

For $v \in \operatorname{span}\{\tilde{\theta}^{(i)}_\mu : \mu \in \nabla_{\tilde{\Theta}^{(i)}}, |\mu| = k'\}$, by the biorthogonality of $\Theta^{(i)}$, $\tilde{\Theta}^{(i)}$ and (21), we have

$$\|v\|_{(H^{m_i}_{\vec{\sigma}_i}(a_i,b_i))'} = \sup_{0\neq z\in H^{m_i}_{\vec{\sigma}_i}(a_i,b_i)} \frac{\langle v,z\rangle_{L_2(a_i,b_i)}}{\|z\|_{H^{m_i}(a_i,b_i)}} =$$

$$\sup_{0\neq z\in H^{m_i}_{\vec{\sigma}_i}(a_i,b_i)} \inf_{w\in\operatorname{span}\{\theta^{(i)}_\mu : \mu\in\nabla_{\tilde{\Theta}^{(i)}},|\mu|<k'\}} \frac{\langle v, z-w\rangle_{L_2(a_i,b_i)}}{\|z\|_{H^{m_i}(a_i,b_i)}} \lesssim 2^{-k'm_i}\|v\|_{L_2(a_i,b_i)}.$$

Now from $\Xi^{(i)} \subset H^{m_i}_{\vec{\sigma}_i}(a_i,b_i)$, (24), (25), and (22), for $k \leq k'$ we estimate

$$\begin{aligned}|\langle I^{(i)}_{k,k'}\tilde{\mathbf{z}},\mathbf{d}\rangle| &= |\langle \sum_{|\lambda|=k} d_\lambda \xi^{(i)}_\lambda, \sum_{|\mu|=k'} \tilde{z}_\mu \tilde{\theta}^{(i)}_\mu\rangle_{L_2(a_i,b_i)}| \\ &\lesssim \|\sum_{|\lambda|=k} d_\lambda \xi^{(i)}_\lambda\|_{H^{m_i}(a_i,b_i)} 2^{-k'm_i} \|\sum_{|\mu|=k'} \tilde{z}_\mu \tilde{\theta}^{(i)}_\mu\|_{L_2(a_i,b_i)} \\ &\lesssim 2^{km_i}\|\sum_{|\lambda|=k} d_\lambda \xi^{(i)}_\lambda\|_{L_2(a_i,b_i)} 2^{-k'm_i} \|\sum_{|\mu|=k'} \tilde{z}_\mu \tilde{\theta}^{(i)}_\mu\|_{L_2(a_i,b_i)} \\ &\lesssim 2^{(k-k')m_i}\|\mathbf{d}\|\|\tilde{\mathbf{z}}\|,\end{aligned}$$

or

$$\|I^{(i)}_{k,k'}\| \lesssim 2^{(k-k')m_i} \quad (k \leq k'). \tag{34}$$

From (30), $\frac{\sqrt{\sum_{\vec{q}\in I}\prod_{i=1}^n 4^{k_i' q_i}}}{\sqrt{\sum_{\vec{q}\in I}\prod_{i=1}^n 4^{k_i q_i}}} \leq \sum_{\vec{q}\in I}\prod_{i=1}^n 2^{(k_i'-k_i)q_i}$, (33), and (34), we infer that

$$\|I_{\vec{k},\vec{k}'}\| \lesssim \prod_{\{i:k_i\geq k_i'\}} 2^{\frac{1}{2}(k_i'-k_i)}.$$

We conclude that the left-hand side from (29) is $\mathscr{O}(2^{-\frac{1}{2}\ell})$, and so that (29) is valid when ℓ is sufficiently large, with which the proof is completed.

Remark 10 We assumed K to be the intersection of tensor products of Sobolev spaces of univariate functions of *nonnegative* orders. Now consider the case that in (16) some q_i *is* negative. Then with $(d_{\vec{\mu}}^{\tilde{\Theta}})^{-2}$ as in (18), $\vec{k},\vec{k}'$ can be selected with $k_i > k_i'$ and $k_j \leq k_j'$ for $j \neq i$ such that $\frac{\sqrt{\sum_{\vec{q}\in I}\prod_{i=1}^n 4^{k_i' q_i}}}{\sqrt{\sum_{\vec{q}\in I}\prod_{i=1}^n 4^{k_i q_i}}}$ is arbitrarily close to $2^{(k_i'-k_i)q_i}$. Thus, in order to bound $\|I_{\vec{k},\vec{k}'}\|$ in (30), for $q_i \leq -\frac{1}{2}$ the growth $2^{(k_i'-k_i)q_i}$ as function of $k_i - k_i'$ is not compensated by the bound $\|I_{k_i,k_i'}^{(i)}\| \lesssim 2^{\frac{1}{2}(k_i'-k_i)}$ from (33). Since for discontinuous $\tilde{\theta}_\lambda^{(i)}$, as needed for (14), generally the estimate (33) is sharp, we conclude that our assumptions on K seem hard to avoid.

Actually, the above reasoning shows that for $q_i \leq -\frac{1}{2}$ it can*not* be expected that $[\langle \xi_{\vec{\lambda}}, \tilde{\theta}_{\vec{\mu}}\rangle_{L_2(\square)}]_{\vec{\lambda}\in\nabla_\Xi,\vec{\mu}\in\nabla_{\tilde{\Theta}}} \in \mathscr{L}(\ell_2(\nabla_{\tilde{\Theta}}),\ell_2(\nabla_\Xi))$. Assuming that $\|\mathbf{z}^\top \Xi\|_K \lesssim \|\mathbf{z}\|_{\ell_2(\nabla_\Xi)}$, this means that it is not possible that $\|\tilde{\mathbf{z}}^\top \tilde{\Theta}\|_{K'} \lesssim \|\tilde{\mathbf{z}}\|_{\ell_2(\nabla_{\tilde{\Theta}})}$.

3.3 Examples

We give two examples of settings where K is the intersection of tensor products of univariate Sobolev spaces of nonnegative orders.

Example 2 When the operator A results from a variational formulation of a second order *elliptic* boundary value problem, e.g. Poisson's equation, on $\square$ with homogeneous Dirichlet boundary conditions, the natural spaces are

$$\begin{aligned} H = K = H_0^1(\square) =& H_{(1,1)}^1(a_1,b_1)\otimes L_2(a_2,b_2)\otimes\cdots\otimes L_2(a_n,b_n)\cap \\ &\vdots \\ &\cap L_2(a_1,b_1)\otimes\cdots\otimes L_2(a_{n-1},b_{n-1})\otimes H_{(1,1)}^1(a_n,b_n). \end{aligned}$$

Example 3 With "spatial domain" $\tilde{\square} = \prod_{i=2}^n (a_i,b_i)$, and $-\infty < a_1, b_1 < \infty$, for $t \in (a_1,b_1)$ a.e. let $b(t;\cdot,\cdot)$ be a bilinear form on $H_0^1(\tilde{\square})\times H_0^1(\tilde{\square})$ such that for some constant λ_0 and for a.e. $t \in (a_1,b_1)$,

$$\begin{aligned} |b(t;\omega,\zeta)| &\lesssim \|\omega\|_{H^1(\tilde{\square})}\|\zeta\|_{H^1(\tilde{\square})} \quad (\omega,\zeta\in H_0^1(\tilde{\square})), \\ b(t;\omega,\omega)+\lambda_0\|\omega\|_{L_2(\tilde{\square})}^2 &\gtrsim \|\omega\|_{H^1(\tilde{\square})}^2 \quad (\omega\in H_0^1(\tilde{\square})). \end{aligned}$$

With $B(t)\in\mathscr{L}(H_0^1(\tilde{\square}),H^{-1}(\tilde{\square}))$ defined by $(B(t)\omega)(\zeta) = b(t;\omega,\zeta)$, given a right-hand side $g(t)$ and an initial datum h, we consider the *parabolic problem* of solving

$$\tfrac{du}{dt}(t) + B(t)u(t) = g(t) \quad \text{in } H^{-1}(\tilde{\square}), \quad u(a_1) = h \text{ in } L_2(\tilde{\square}).$$

A well-posed simultaneous space-time variational formulation reads as finding $u \in H := L_2(a_1,b_1)\otimes H^1_0(\tilde{\square}) \cap H^1(a_1,b_1)\otimes H^{-1}(\tilde{\square})$ such that for all $v=(v_1,v_2)\in K := L_2(a_1,b_1)\otimes H^1_0(\tilde{\square})\times L_2(\tilde{\square})$,

$$\begin{aligned}(Au)(v) &:= \int_{a_1}^{b_1}\int_{\tilde{\square}} \frac{\partial u}{\partial t}(t,\vec{x})v_1(t,\vec{x}) + b(t;u(t,\vec{x}),v_1(t,\vec{x}))dtd\vec{x} + \int_{\tilde{\square}} u(a_1,\vec{x})v_2(\vec{x})d\vec{x}\\ &= \int_{a_1}^{b_1}\int_{\tilde{\square}} g(t,\vec{x})v_1(t,\vec{x})dtd\vec{x} + \int_{\tilde{\square}} h(\vec{x})v_2(\vec{x})d\vec{x} =: f(v).\end{aligned}$$

This problem needs a little massage to fit in the setting that has been considered. The space K is the Cartesian product of $K_1 := L_2(a_1,b_1)\otimes H^1_0(\tilde{\square})$ and $K_2 = L_2(\tilde{\square})$, $A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$, and $f = \begin{bmatrix} g \\ h \end{bmatrix}$. The task of the approximate evaluation of the (primal) (see §1) residual splits into corresponding tasks for both equations. The first equation fits in the framework presented so far, and so this task can be performed assuming suitable tensor product collections on H and K_1, and assuming that the spatial differential operator B has coefficients that are polynomials in t and $\vec{x}$.

Concerning the second equation, H and K_2 are function spaces on the different domains $\square = (a_1,b_1)\times\tilde{\square}$ and $\tilde{\square}$, which does not fit into the setting we have assumed from the very beginning of this section. The operator A_2, however, reads as $\mathrm{Id}\circ R$, with $\mathrm{Id}: L_2(\tilde{\square}) \to L_2(\tilde{\square})$ and $R: H \to L_2(\tilde{\square})$ defined by $(Rv)(t,x) = v(a_1,x)$. It is known that R is bounded (e.g. see [12, Ch.XVIII, §1, Th.1]), and its representation w.r.t. the collections $\Psi = (D^\Psi)^{-1}\otimes_{i=1}^n \Psi^{(i)} \subset H$ and $\otimes_{i=2}^n \Psi^{(i)} \subset L_2(\tilde{\square})$, i.e., $(R\Psi)(\otimes_{i=2}^n \tilde{\Psi}^{(i)}) = \big[(d^\Psi_{\vec{\lambda}})^{-1}\psi^{(1)}_{\lambda_1}(a_1)\prod_{i=2}^n \delta_{\lambda_i\mu_i}\big]_{\vec{\lambda},\vec{\mu}\in\nabla_\Psi}$, has at most one non-zero in each column. Results concerning the sparse evaluation of $(\mathrm{Id}\otimes_{i=2}^n \Psi^{(i)})(\Xi)$, with Ξ being some tensor product collection on $K_2 = L_2(\tilde{\square})$, which obviously can be derived in the framework that has been developed, now directly imply corresponding results for the evaluation of $(A_2\Psi)(\Xi)$.

3.4 Cost of the approximate residual evaluation

In the main Theorem 1, for $w_{\Lambda_\Psi} \in \mathrm{span}\{\psi_{\vec{\mu}} : \vec{\mu}\in\Lambda_\Psi\}$, we established an upper bound $\gtrsim \varepsilon$ on the error in the approximate residual

$$[(f - Aw_{\Lambda_\Psi})(\xi_{\vec{\lambda}})]_{\vec{\lambda}\in\Lambda_\Xi}, \tag{35}$$

as well as the bound $\#\Lambda_\Xi \lesssim \#\Lambda_\Psi + \varepsilon^{-1/s}$ on the cardinality of Λ_Ξ. Below we show the same bound for the number of arithmetic operations to *compute* this approximate residual, or an approximation of it that keeps its error on the same level.

Recall the setting of A being a linear partial differential operator with constant or polynomial coefficients, tensor product collections $D^\Psi\Psi = \otimes_{i=1}^n \Psi^{(i)}$, $D^\Xi\Xi = \otimes_{i=1}^n \Xi^{(i)}$, for some local $\Psi^{(i)}$ and $\Xi^{(i)}$. Then the bi-infinite system or stiffness matrix reads as

$$[(A\psi_{\vec{\mu}})(\xi_{\vec{\lambda}})]_{\vec{\lambda}\in\nabla_\Xi,\vec{\mu}\in\nabla_\Psi} = (D^\Xi)^{-1}\big[\sum_{j\in J}\bigotimes_{i=1}^n s_{j_i}(\xi^{(i)}_{\lambda_i},\psi^{(i)}_{\mu_i})\big](D^\Psi)^{-1} \tag{36}$$

for some finite index set J, and for bilinear forms $s_{j_i}(\cdot,\cdot)$ that are local, i.e., give zero when the supports of both arguments have empty intersection.

As shown in [21], building on the earlier work in the (adaptive) sparse grid setting (e.g., see [1–5, 17, 24, 30, 31]) for multi-trees $\Lambda_\Psi \subset \nabla_\Psi$ and $\Lambda_\Xi \subset \nabla_\Xi$, the multiplication with each of the matrices $[\bigotimes_{i=1}^n s_{j_i}(\xi_{\lambda_i}^{(i)}, \psi_{\mu_i}^{(i)})]_{\vec\lambda\in\Lambda_\Xi,\vec\mu\in\Lambda_\Psi}$, and so with the matrix $[(A\psi_{\vec\mu})(\xi_{\vec\lambda})]_{\vec\lambda\in\Lambda_\Xi,\vec\mu\in\Lambda_\Psi}$, can be performed in $\mathcal{O}(\#\Lambda_\Xi + \#\Lambda_\Psi)$ operations, although the matrix is generally not sparse.

Let us briefly sketch the two principal ideas behind the realization of the matrix-vector multiplication in linear complexity. Firstly, in the univariate case, say with $\Psi = \Xi$ and $\Lambda_\Xi = \Lambda_\Psi$, the idea is first to apply a transformation from the multilevel representation to a locally finite single-scale representation, then to do the matrix-vector multiplication in single-scale coordinates, and finally to return to a multi-scale representation by applying the transposed transformation. Thanks to the tree constraint on $\Lambda_\Xi = \Lambda_\Psi$, the basis transformations can be performed in linear complexity, and so can the matrix-vector multiplication, since in single-scale coordinates the matrix is sparse.

Secondly, to explain the generalization to the bivariate, and with that to the multivariate case, we consider a simplified setting that essentially corresponds to pretending that for each univariate wavelet basis, on each level there is only one wavelet. For $i = 1, 2$, let $S^{(i)} \in \mathbb{R}^{m\times m}$. For $k \le m$, let $I_k : \mathbb{R}^k \to \mathbb{R}^m$ denote the extension with zeros, and $P_k = I_k^\top$. Motivated by our observations for the univariate case, just explained, assume that multiplication with $P_{k'} S^{(i)} I_k$ can be performed in $\mathcal{O}(\max(k,k'))$ operations.

Now let $\Lambda \subseteq \{1,\dots,m\}^2$ be such that when frozen in either the first or second coordinate, at any value, it is a set of the form $\{1,\dots,k\}$ for some $k \le m$, i.e., Λ is a "multi-tree". With $I_\Lambda : \mathbb{R}^{k\times k} \to \mathbb{R}^{m\times m}$ denoting the extension with zeros, and $P_\Lambda = I_\Lambda^\top$, we are interested in the matrix-vector multiplication with the matrix $P_\Lambda(S^{(1)}\otimes S^{(2)})I_\Lambda$.

A straightforward matrix-vector multiplication, taking advantage of the tensor product structure ($S^{(1)}\otimes S^{(2)}$ "as a whole" is a $m^2\times m^2$ matrix), requires $\gtrsim m^2$ operations, assuming that $\max_{\lambda\in\Lambda}\lambda_1 = \max_{\lambda\in\Lambda}\lambda_2 = m$. This is even true in the extremal case that $\Lambda = \{\lambda \in \{1,\dots,m\}^2 : \min(\lambda_1,\lambda_2) = 1\}$. To solve this problem, the idea is to split $S^{(1)}$ as $L^{(1)} + U^{(1)}$, where $L^{(1)}_{ij} = S^{(1)}_{ij}$ when $i > j$, and zero elsewhere. By definition of $U^{(1)}$, $L^{(1)}$, and Λ being a "multi-tree", it holds that

$$(U^{(1)}\otimes\mathrm{Id})I_\Lambda = I_\Lambda P_\Lambda (U^{(1)}\otimes\mathrm{Id})I_\Lambda, \quad P_\Lambda(L^{(1)}\otimes\mathrm{Id}) = P_\Lambda(L^{(1)}\otimes\mathrm{Id})I_\Lambda P_\Lambda,$$

from which we infer that

$$\begin{aligned}
P_\Lambda(S^{(1)}\otimes S^{(2)})I_\Lambda &= P_\Lambda((U^{(1)}+L^{(1)})\otimes S^{(2)})I_\Lambda \\
&= P_\Lambda(L^{(1)}\otimes\mathrm{Id})(\mathrm{Id}\otimes S^{(2)})I_\Lambda + P_\Lambda(U^{(1)}\otimes\mathrm{Id})(\mathrm{Id}\otimes S^{(2)})I_\Lambda \\
&= P_\Lambda(L^{(1)}\otimes\mathrm{Id})(\mathrm{Id}\otimes S^{(2)})I_\Lambda + P_\Lambda(\mathrm{Id}\otimes S^{(2)})(U^{(1)}\otimes\mathrm{Id})I_\Lambda \\
&= P_\Lambda(L^{(1)}\otimes\mathrm{Id})I_\Lambda P_\Lambda(\mathrm{Id}\otimes S^{(2)})I_\Lambda + P_\Lambda(\mathrm{Id}\otimes S^{(2)})I_\Lambda P_\Lambda(U^{(1)}\otimes\mathrm{Id})I_\Lambda.
\end{aligned}$$

The matrix-vector multiplication with each of the three matrices $P_\Lambda(L^{(1)}\otimes\mathrm{Id})I_\Lambda$, $P_\Lambda(\mathrm{Id}\otimes S^{(2)})I_\Lambda$, and $P_\Lambda(U^{(1)}\otimes\mathrm{Id})I_\Lambda$ requires $\mathcal{O}(\#\Lambda)$ operations, and we conclude that so does this way of multiplying $P_\Lambda(S^{(1)}\otimes S^{(2)})I_\Lambda$ with a vector.

Concerning the first term of the approximate residual (35), only in cases where an analytic expression of $f(\psi_{\vec\lambda})$ can be derived, the exact evaluation of $[f(\xi_{\vec\lambda})]_{\vec\lambda\in\Lambda_\Xi}$ can be computed in $\mathcal{O}(\#\Lambda_\Xi)$ operations. In general quadrature is needed, which is a delicate issue with tensor product approximation. With Φ being a tensor product of univariate single scale bases such that $\mathrm{span}\{\xi_{\vec\lambda} : \vec\lambda \in \Lambda_\Xi\} \subset \mathrm{span}\,\Phi$, it is not recommendable to compute $[f(\xi_{\vec\lambda})]_{\vec\lambda\in\Lambda_\Xi}$

from $[f(\phi)]_{\phi\in\Phi}$, or from an approximation of the latter by the application of numerical quadrature. The reason is that for the smallest of such Φ, $\#\Phi$ can be of the order of $(\#\Lambda_\Xi)^n$.

Assuming a continuous f, the following procedure can be applied instead. Approximate f from the span of a collection $\Pi = \{\pi_{\vec{\mu}} : \vec{\mu} \in \nabla_\Pi\}$ of tensor product *interpolets*, i.e. $D^\Pi \Pi = \otimes_{i=1}^n \Pi^{(i)}$, for some invertible diagonal matrix D^Π, and collections of univariate interpolets $\Pi^{(i)}$. A well-known example of a collection of univariate interpolets is given by the continuous piecewise linear hierarchical basis. The duals of the univariate interpolets are (uniformly) finite linear combinations of Dirac functionals, and so are the duals $\tilde{\pi}_{\vec{\mu}}$ of the tensor product interpolets. So any coefficient in the expansion $f = \sum_{\vec{\mu}\in\nabla_\Pi} \tilde{\pi}_{\vec{\mu}}(f)\pi_{\vec{\mu}}$ can be computed in $\mathscr{O}(1)$ operations. Assuming a sufficient (mixed) smoothness of f, and taking interpolets of a sufficiently high order, for any $\varepsilon > 0$ there exists a multi-tree $\Lambda_\Pi \subset \nabla_\Pi$ such that $\|f - \sum_{\vec{\mu}\in\Lambda_\Pi} \tilde{\pi}_{\vec{\mu}}(f)\pi_{\vec{\mu}}\|_{K'} \le \varepsilon$, where $\#\Lambda_\Pi \lesssim \varepsilon^{-1/\tilde{s}}$ for some $\tilde{s} \ge s_{\max}$, e.g. see [5]. Here recall that $s_{\max}$ is defined as the best possible rate that can be expected for approximation in H from $\operatorname{span}\Psi$, so being an upper bound for s.

Now by replacing $[f(\xi_{\vec{\lambda}})]_{\vec{\lambda}\in\Lambda_\Xi}$ by $\big[\sum_{\vec{\mu}\in\Lambda_\Pi} \tilde{\pi}_{\vec{\mu}}(f)\langle\xi_{\vec{\lambda}}, \pi_{\vec{\mu}}\rangle_{L_2(\square)}\big]_{\vec{\lambda}\in\Lambda_\Xi}$ in the approximate residual computation, we make an additional error that, by Condition (A) of Theorem 1, is of order ε. Its computation requires the multiplication with $[\bigotimes_{i=1}^n \langle\xi_{\lambda_i}^{(i)}, \pi_{\mu_i}^{(i)}\rangle_{L_2(a_i,b_i)}]_{\vec{\lambda}\in\Lambda_\Xi,\vec{\mu}\in\Lambda_\Pi}$, which can be performed in $\mathscr{O}(\#\Lambda_\Xi + \#\Lambda_\Pi)$ operations by the application of the algorithm from [21].

4 Numerical experiments

In this section, we illustrate our theoretical results by selected numerical experiments. A crucial point is the construction of Λ_Ξ for the approximation $(\mathbf{f} - \mathbf{A}\mathbf{I}_{\Lambda_\Psi}\mathbf{w}_{\Lambda_\Psi})|_{\Lambda_\Xi}$ to the exact residual $\mathbf{f} - \mathbf{A}\mathbf{I}_{\Lambda_\Psi}\mathbf{w}_{\Lambda_\Psi} = (f - A(\mathbf{w}_{\Lambda_\Psi}^\top\Psi))(\Xi)$. In particular, we investigate quantitative properties of this multi-tree based residual approximation, and compare it to the standard approach for residual approximation being based on an `APPLY` routine. Moreover, we present advantages of the usage of L_2-orthonormal (multi-) wavelet bases in this context.

We consider Poisson's equation with homogeneous Dirichlet boundary conditions as discussed in Examples 1 and 2. With $H = K = H_0^1(\square)$, $\square = (0,1)^n$, for given $f \in H'$, we seek $u \in H$ such that

$$(Au)(v) := \int_\square \nabla u \cdot \nabla v = \int_\square fv \quad (v \in H). \tag{37}$$

For our convenience, we shall choose $f \in H'$ later on as a *constant* function and the spatial dimension $n \in \{2,3\}$. Since f does *not* vanish at the boundary of $\square$, we point out that an adaptive (wavelet) method is required in order to obtain the best possible approximation rate $s_{\max} = p$ for piecewise polynomial wavelets of degree p (see [11]).

4.1 Representation of (37) in terms of L_2-orthonormal multi-wavelet bases

To derive the equivalent ℓ_2-problem $\mathbf{A}\mathbf{u} = \mathbf{f}$ associated to (37), we use $L_2(0,1)$-orthonormal multi-wavelet bases $\Psi^{(i)} = \{\psi_\lambda^{(i)} : \lambda \in \nabla_{\Psi^{(i)}}\} \subset H^2(\square) \cap H_0^1(\square)$ for $i \in \{1,\dots,n\}$ so that

$$\Psi := (D^\Psi)^{-1} \bigotimes_{i=1}^n \Psi^{(i)} = \{\psi_{\vec{\lambda}} := (d_{\vec{\lambda}}^\Psi)^{-1} \otimes_{i=1}^n \psi_{\lambda_i}^{(i)} : \vec{\lambda} \in \nabla_\Psi\},$$

where $d_{\vec{\lambda}}^{\Psi} := |\psi_{\lambda_1}^{(1)} \otimes \cdots \otimes \psi_{\lambda_n}^{(n)}|_{H^1(\square)}$, is a Riesz basis for $H_0^1(\square)$. Moreover, we set $\Xi := \Psi$.

The construction of an $L_2(0,1)$-orthonormal wavelet basis is based on a multiresolution analysis (MRA) $(V_j)_{j\in\mathbb{N}_0}$ where V_j, for integer $p \geq 2$, is given by

$$V_j := \prod_{k\in[0,2^j-1]\cap\mathbb{Z}} P_p((2^{-j}k, 2^{-j}(k+1)) \cap C^1(0,1) \cap H_0^1(0,1).$$

Following the guideline given in [15,16], a so called *intertwining* MRA $(\check{V}_j)_{j\in\mathbb{N}_0}$ with $V_j \subset \check{V}_j \subset V_{j+m}$ for some $m \in \mathbb{N}$ can be constructed such that the single scale spaces $\check{V}_j$ are spanned by uniformly local, globally C^1, *orthonormal* scaling functions which are moreover piecewise polynomials of degree p. The corresponding wavelets $\psi_\lambda^{(i)} \in \Psi^{(i)}$ for $i \in \{1,\ldots,n\}$, being finite linear combinations of these scaling functions, inherit these properties. Moreover, if $\psi_\lambda^{(i)}$ does not correspond to a boundary adapted wavelet, it has $p+1$ vanishing moments. A detailed description will be given in [29].

The choice of L_2-orthonormal wavelet bases $\Psi^{(i)}$ ($i \in \{1,\ldots,n\}$) yields certain advantages concerning quantitative properties of the AWGM which we detail next.

Efficient realization of matrix-vector products

As explained in Subsection 3.4, with A being a second-order differential operator with constant, or polynomial coefficients, the application of the system matrix $\mathbf{R}_{\tilde{\Lambda}}\mathbf{A}\mathbf{I}_\Lambda$ can be realized within $\mathcal{O}(\#\tilde{\Lambda} + \#\Lambda)$ operations, when $\tilde{\Lambda}, \Lambda \subset \nabla_\Psi$ are finite multi-trees, and the $\Psi^{(i)}$ ($i \in \{1,\ldots,n\}$) are *any* univariate Riesz wavelet bases satisfying assumptions (8), (9), (10), (13) (see [21]).

However, we observe that due to the $L_2(0,1)$-orthonormality of the univariate wavelet basis, the bi-infinite system matrix $\mathbf{A}$ from (37) has the form

$$(D^\Psi)^{-1}\big[A^{(1)}\otimes \mathrm{Id}^{(2)}\otimes\cdots\otimes \mathrm{Id}^{(n)} + \cdots + \mathrm{Id}^{(1)}\otimes\cdots\otimes\mathrm{Id}^{(n-1)}\otimes A^{(n)}\big](D^\Psi)^{-1}, \qquad (38)$$

where, for $i \in \{1,\ldots,n\}$, $\mathrm{Id}^{(i)}$ denotes a (bi-infinite) *identity* matrix indexed by $\nabla_{\Psi^{(i)}}$ and $A^{(i)} := [\int_0^1 \partial\psi_\lambda^{(i)}\partial\psi_\mu^{(i)}]_{\lambda,\mu\in\nabla_{\Psi^{(i)}}}$ a univariate stiffness matrix.

Having an $\mathbf{A}$ of this special form is very advantageous when it comes to the application of $\mathbf{R}_{\tilde{\Lambda}}\mathbf{A}\mathbf{I}_\Lambda$ for $\tilde{\Lambda}, \Lambda \subset \nabla_\Psi$ being two finite multi-trees. Indeed, computing e.g. $\mathbf{R}_{\tilde{\Lambda}}[A^{(1)}\otimes \mathrm{Id}^{(2)}\otimes\cdots\otimes\mathrm{Id}^{(n)}]\mathbf{I}_\Lambda$ requires only the application of $A^{(1)}$ w.r.t. the first coordinate which can be implemented efficiently (see [21]). We refer to such an operation as *unidirectional*. Thus, the number of unidirectional operations for the multiplication with $\mathbf{R}_{\tilde{\Lambda}}\mathbf{A}\mathbf{I}_\Lambda$ is *exactly* n.

Conversely, choosing univariate bases $\Psi^{(i)}$ that are only biorthogonal, we need to replace $\mathrm{Id}^{(i)}$ in (38) by univariate mass matrices $M^{(i)} := [\int_0^1 \psi_\lambda^{(i)}\psi_\mu^{(i)}]_{\lambda,\mu\in\nabla_{\Psi^{(i)}}}$ for $i \in \{1,\ldots,n\}$. In this case, the application of $\mathbf{R}_{\tilde{\Lambda}}\mathbf{A}\mathbf{I}_\Lambda$ by means of unidirectional operations is more complicated and relies on the splitting of $A^{(i)}$ and $M^{(i)}$ into upper and strict lower triangular block matrices for $i \in \{1,\ldots,n\}$ (see Subsection 3.4): for $S^{(i)} \in \{A^{(i)}, M^{(i)}\}$ for $i \in \{1,2\}$, we split $S^{(1)}$ into an upper and strict lower triangular matrix, $S^{(1)} = U^{(1)} + L^{(1)}$, so that we obtain:

$$\mathbf{R}_{\tilde{\Lambda}}[S^{(1)}\otimes S^{(2)}]\mathbf{I}_\Lambda = \mathbf{R}_{\tilde{\Lambda}}[(L^{(1)}\otimes\mathrm{Id}^{(2)})(\mathrm{Id}^{(1)}\otimes S^{(2)})]\mathbf{I}_\Lambda + \mathbf{R}_{\tilde{\Lambda}}[(\mathrm{Id}^{(1)}\otimes S^{(2)})(U^{(1)}\otimes\mathrm{Id}^{(2)})]\mathbf{I}_\Lambda.$$

Observe that the application of $\mathbf{R}_{\tilde{\Lambda}}[S^{(1)}\otimes S^{(2)}]\mathbf{I}_\Lambda$ requires $2^n = 2^2 = 4$ unidirectional operations. For higher dimensions $n > 2$, the application of $\mathbf{R}_{\tilde{\Lambda}}\mathbf{A}\mathbf{I}_\Lambda$ is then realized by a recursion requiring, however, $\mathcal{O}(n\cdot 2^n)$ unidirectional operations. Thus, compared to the case of $L_2(0,1)$-orthonormal bases, we have an additional factor of 2^n.

Remark 11 The usage of *prewavelets* as described in [17,23] can reduce the number of unidirectional operations to $\mathscr{O}(n)$ as in this case, the strict lower triangular block matrix associated to $M^{(i)}$ for $i \in \{1,\dots,n\}$ is zero.

Condition numbers

For the adaptive wavelet Galerkin method (AWGM), the condition number $\kappa(\mathbf{A})$ of $\mathbf{A}$ is of great importance. Firstly, it constitutes an upper bound for the condition number of $\mathbf{A}_\Lambda$ arising in Galerkin systems $\mathbf{A}_\Lambda \mathbf{u}_\Lambda = \mathbf{f}_\Lambda$ to be approximately solved by an iterative linear system solver. Secondly, mainly $\kappa(\mathbf{A})$ determines how large we may choose the bulk chasing parameter μ (cf. line after (3)) in order to guarantee optimal complexity.

In our case, $\kappa(\mathbf{A})$ can be bounded *independently* of the spatial dimension n by means of the Riesz constants $c^{(i)}_{\min}$, $c^{(i)}_{\max}$ of the $|\cdot|_{H^1(0,1)}$ normalized univariate Riesz bases $\Psi^{(i)}$ ($i \in \{1,\dots,n\}$). More precisely, with

$$c^{(i)}_{\min} \le \| \sum_{\lambda \in \nabla_{\Psi^{(i)}}} v_\lambda \partial \psi^{(i)}_\lambda \|^2_{L_2} \Big/ \sum_{\lambda \in \nabla_{\Psi^{(i)}}} |v_\lambda|^2 \|\partial \psi^{(i)}_\lambda\|^2_{L_2} \le c^{(i)}_{\max},$$

for all $(2^{|\lambda|} v_\lambda)_{\lambda \in \nabla_{\Psi^{(i)}}} \in \ell_2(\nabla_{\Psi^{(i)}})$, it can be shown that $\kappa(\mathbf{A}) \le \max_i c^{(i)}_{\max}/c^{(i)}_{\min}$ (cf., e.g., [7,13]). For the constructions of $\Psi^{(i)}$ for $i \in \{1,\dots,n\}$ used for the numerical experiments below, we estimated these constants numerically and used the bounds $c^{(i)}_{\min} \ge 0.3$, $c^{(i)}_{\max} \le 2.1$ for $p = 2$ and $c^{(i)}_{\min} \ge 0.3$, $c^{(i)}_{\max} \le 2.0$ for $p = 3$.

4.2 Approximate residuals

Let us now come to the approximation of the exact residual $\mathbf{f}_{\Lambda_k} - \mathbf{A}\mathbf{I}_{\Lambda_k}\mathbf{w}_{\Lambda_k}$ where, as above, $\mathbf{w}_{\Lambda_k} \in \ell_2(\Lambda_k)$ denotes an approximate Galerkin solution of $\mathbf{A}_{\Lambda_k}\mathbf{u}_{\Lambda_k} = \mathbf{f}_{\Lambda_k}$ for $k = 1,2,\dots$ inside the adaptive wavelet Galerkin iteration.

The standard approach to approximate the exact (infinite) residual $\mathbf{f} - \mathbf{A}\mathbf{I}_{\Lambda_k}\mathbf{w}_{\Lambda_k}$ is by using a RHS and a APPLY routine that approximate $\mathbf{f}$ and $\mathbf{A}\mathbf{I}_{\Lambda_k}\mathbf{w}_{\Lambda_k}$ within prescribed absolute tolerances. An APPLY based approximate residual evaluation then takes the form

$$\mathbf{r}^{(k)}_{\mathtt{APPLY}}(\varepsilon) := \mathtt{RHS}[\varepsilon/2] - \mathtt{APPLY}[\mathbf{w}_{\Lambda_k}, \varepsilon/2]. \tag{39}$$

As shown in [13], $\mathbf{A}$ is very close to a sparse matrix. Indeed, since $\psi^{(i)}_{\lambda_i} \in C^1(0,1)$ for all $\lambda_i \in \nabla_{\Psi^{(i)}}$ and $i \in \{1,\dots,n\}$, it holds that $\|\mathbf{A} - \mathbf{A}_j\| \lesssim 2^{-\frac{3}{2}j}$ for all $j \in \mathbb{N}_0$ where $\mathbf{A}_j$ are *sparse compressions* of $\mathbf{A}$ having at most $\mathscr{O}(nj)$ non-zero entries per row and column (cf. [13, Remark 3.4 & Theorem 3.5]), which can be computed exactly at unit cost.

For our simple choice of f in (37) being a constant function, an entry $\mathbf{f}_{\vec{\lambda}} = f(\psi_{\vec{\lambda}})$ of $\mathbf{f}$ is only non-zero if all the indices $\lambda_1,\dots,\lambda_n$ correspond to either scaling functions or to so called boundary wavelets which both do not have vanishing moments. Moreover, any entry $\mathbf{f}_{\vec{\lambda}}$ can be computed exactly at unit cost.

From both observations it can be inferred, see e.g. [26, Sect. 3], that the cost of evaluating $\mathbf{r}^{(k)}_{\mathtt{APPLY}}(\varepsilon)$ is $\mathscr{O}(\varepsilon^{-1/s}|\mathbf{I}_{\Lambda_k}\mathbf{w}_{\Lambda_k}|^{-1/s}_{\mathscr{A}^s} + \#\Lambda_k)$, even without any restriction on $\varepsilon > 0$. Knowing that for $s > 0$ for which $\mathbf{u} \in \mathscr{A}^s$, one has $\sup_k |\mathbf{I}_{\Lambda_k}\mathbf{w}_{\Lambda_k}|_{\mathscr{A}^s} \lesssim |\mathbf{u}|_{\mathscr{A}^s} < \infty$, we conclude that the condition (4) for being a valid approximate residual evaluation is satisfied.

Restricting to sets Λ_k that are multi-trees, our focus is on the approximate residual evaluation of the form

$$\mathbf{r}^{(k)} := (\mathbf{f} - \mathbf{A}\mathbf{I}_{\Lambda_k}\mathbf{w}_{\Lambda_k})|_{\tilde{\Lambda}_k}, \tag{40}$$

where $\tilde{\Lambda}_k \subset \nabla_\Psi$ (recall that $\Xi = \Psi$) with $\#\tilde{\Lambda}_k \lesssim \#\Lambda_k + \varepsilon^{-1/s}$ is a sufficiently large multi-tree as in Theorem 1. Since it is difficult to extract from the proof of Theorem 1 reasonably sharp quantitative information for the construction of $\tilde{\Lambda}_k$, we do some experiments.

First of all, we take $\tilde{\Lambda}_k$ to be only dependent on Λ_k, which is justified since f has a finite representation in the auxiliary collection $\tilde{\Theta}$ (cf. Remark 3). For general sufficiently smooth f, its representation w.r.t. $\tilde{\Theta}$ will be that close to being sparse that also then it can be expected that it is sufficient to select $\tilde{\Lambda}_k$ only dependent on Λ_k.

Our experiments are based on the construction $\tilde{\Lambda}_k := \cup_{j=1}^n \tilde{\Lambda}_k^{(j)}$ where for $j \in \{1,\dots,n\}$

$$\begin{aligned}\tilde{\Lambda}_k^{(j)} := \{\vec{\lambda} \in \nabla_\Psi : \exists \vec{\mu} \in \Lambda_k \text{ s.t. } \forall i \in \{1,\dots,n\}, |\lambda_i| \le |\mu_i| + \delta_{i,j}\cdot\ell \\ \wedge\, \mathrm{d}(\operatorname{supp}\psi_{\lambda_i}^{(i)}, \operatorname{supp}\psi_{\mu_i}^{(i)}) \le D_i 2^{-|\lambda_i|}\},\end{aligned} \tag{41}$$

where $\ell \in \mathbb{N}$ and $D_i := \sup_{\lambda\in\nabla_{\Psi^{(i)}}} 2^{|\lambda|}\operatorname{diam}\operatorname{supp}\psi_\lambda^{(i)}$ for $i \in \{1,\dots,n\}$. So, $\tilde{\Lambda}_k^{(j)}$ is a (multi-tree) extension of Λ_k by ℓ levels in the j-th coordinate direction. It can be shown that for all j, $\tilde{\Lambda}_k^{(j)}$ is a multi-tree and so is $\tilde{\Lambda}_k$, being a union of multi-trees. Moreover, for the cardinality of $\tilde{\Lambda}_k$, we find that

$$\#\tilde{\Lambda}_k \lesssim n\cdot 2^\ell \cdot \#\Lambda_k. \tag{42}$$

Although not exactly of the same type, the construction of the extension $\tilde{\Lambda}_k$ of Λ_k is similar to that of the extension Λ_Ξ of Λ_Ψ via $\Lambda_{\tilde{\Theta}}$ from the proofs of Propositions 1 and 2. The extension used here turns out to give quantitatively the best results.

4.3 Selected numerical results

We start with some remarks concerning the implementation and parameters of the `AWGM`.

Concerning the right-hand side f in (37), we choose $f \equiv 20$ for $n = 2$ and $f \equiv 100$ for $n = 3$. We emphasize that the symmetry of the solution is not exploited here within the `AWGM`.

In view of (3), we compute an approximate Galerkin solution $\mathbf{w}_{\Lambda_k}$ such that $\|\mathbf{f}_{\Lambda_k} - \mathbf{A}_{\Lambda_k}\mathbf{w}_{\Lambda_k}\|_{\ell_2(\nabla_\Psi)} \le \gamma\|\mathbf{r}^{(k-1)}\|_{\ell_2(\nabla_\Psi)}$ by using the Conjugate Gradients (CG) method, where we took $\gamma = 0.1$.

For computing $\Lambda_{k+1} \supset \Lambda_k$ such that $\|\mathbf{R}_{\Lambda_{k+1}}\mathbf{r}^{(k)}\|_{\ell_2(\Lambda_{k+1})} \ge \mu\|\mathbf{r}^{(k)}\|_{\ell_2(\nabla_\Psi)}$ and $\#(\Lambda_{k+1}\setminus\Lambda_k)$ is minimal up to some constant factor, we choose $\mu = 0.7$. This value of μ is *not* within the range for which (quasi-) optimality of the `AWGM` can be shown theoretically (cf., e.g., [26, Theorem 4.1]), but it yields better quantitative results than with admissible values of μ.

Finally, in order to keep the length of the multi-tree based residual as small as possible (see (42)), we use $\ell = 1$ for the residuals $\mathbf{r}^{(k)}$ defined in (40) in conjunction with (41).

The implementation of `AWGM` is obtained by using the C++-library `LAWA` ([28]). Numerical results reported below were obtained on a PC with 2.83 GHz (Intel Core 2 Duo) and 8 GiB memory. Note that no parallelization techniques were used here and 4 GiB were sufficient for the numerical results in 2d and 3d. Only for high accurate approximation of the exact residual, 8 GiB of memory were required.

Convergence

We start with investigating the approximate residual $\mathbf{r}^{(k)}$ defined in (40) quantitatively. In Figure 1, the convergence history of the approximate residual $\|\mathbf{r}^{(k)}\|_{\ell_2(\nabla_\Psi)}$ for all $k = 1,2,\ldots$ encountered in the course of the AWGM is shown where we set $N_k := \#\Lambda_k$. Observe that the best possible (asymptotic) convergence rate being $s_{\max} = p$ is attained by the AWGM quite fast for $n = 2$. In the three-dimensional case, more degrees of freedom are required to see this asymptotic rate. The fact that the asymptotic regime starts quite late for $n = 3$ is caused by

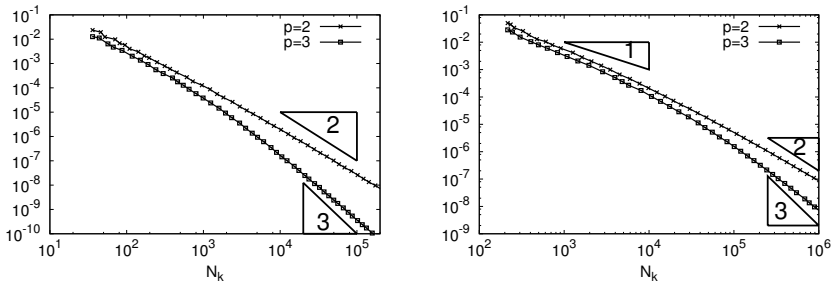


Fig. 1 Convergence history of $\|\mathbf{r}^{(k)}\|_{\ell_2(\nabla_\Psi)}/\|\mathbf{f}\|_{\ell_2(\nabla_\Psi)}$ within the AWGM for polynomial degrees $p \in \{2,3\}$ for $n = 2$ (left) and $n = 3$ (right).

$|\mathbf{u}|_{\mathscr{A}^p}$ being large. Note that the underlying problem has singularities in all 8 corners and also along all 12 edges of the cube $(0,1)^3$, see also the right part of Figure 2. In the left part of Figure 2, we also show results of a (nearly) best N-term approximation. To this end, we first computed a highly accurate approximation $\mathbf{u}_{\bar\varepsilon}$ to $\mathbf{u}$ (with $\|\mathbf{u}-\mathbf{u}_{\bar\varepsilon}\|_{\ell_2(\nabla_\Psi)} \le \bar\varepsilon$ for $\bar\varepsilon$ sufficiently small). Secondly, we compared $\|\mathbf{u}_{\bar\varepsilon} - \mathbf{w}_{\Lambda_k}\|_{\ell_2(\nabla_\Psi)}$ and $\|\mathbf{u}_{\bar\varepsilon} - \mathbf{u}_{\bar\varepsilon,N_k}\|_{\ell_2(\nabla_\Psi)}$, ($N_k := \#\Lambda_k$) and $\mathbf{u}_{\bar\varepsilon,N_k}$ is a best N_k-term approximation to $\mathbf{u}_{\bar\varepsilon}$ (here, the N_k largest coefficients in modulus). We observe that the approximate (Galerkin) solutions $\mathbf{w}_{\Lambda_k}$ are very close to corresponding best N-term approximations. This illustrates both that the AWGM produces approximations that are nearly as good as best N-term multi-tree approximations, and that, for this problem, best N-term multi-tree approximations are nearly as good as *unconstrained* best N-term approximations.

Next, we investigate numerically for which values of η the relative approximation error estimate (2) is valid. For this purpose we need a highly accurate approximation to the exact residual which we realize by computing $\mathbf{r}^{(k)}_{\mathtt{ex}} := \mathbf{r}^{(k)}_{\mathtt{APPLY}}(\bar\varepsilon)$. Here, $\bar\varepsilon$ is (once again) a sufficiently small tolerance. In Figure 3 we show the quantity

$$\|\mathbf{r}^{(k)}_{\mathtt{ex}} - \mathbf{r}^{(k)}\|_{\ell_2(\nabla_\Psi)}/\|\mathbf{r}^{(k)}_{\mathtt{ex}}\|_{\ell_2(\nabla_\Psi)}, \tag{43}$$

as far as $\mathbf{r}^{(k)}_{\mathtt{ex}}$ (requiring a very large amount of storage) was storable. Our choice $\mathbf{r}^{(k)}$ is actually a good approximation of the exact residual $\mathbf{f} - \mathbf{A}\mathbf{I}_{\Lambda_k}\mathbf{w}_{\Lambda_k}$. In view of (2), asymptotically, a value of $\eta \approx 0.175$ is estimated numerically for $p = 2$ and $\eta \approx 0.125$ for $p = 3$.

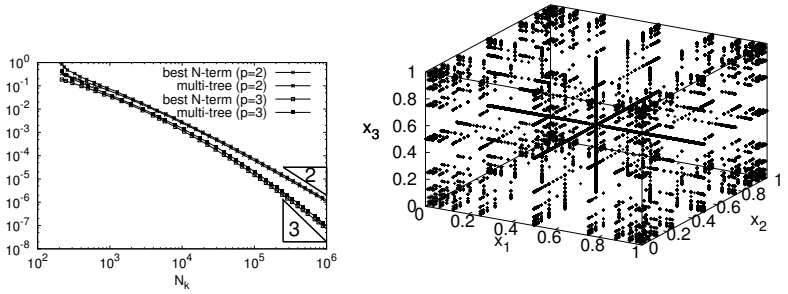


Fig. 2 Comparison of best-N_k-term and multi-tree approximation for $n = 3$ (left) and visualization of a multi-tree Λ_k with $\#\Lambda_k \approx 2 \times 10^5$ obtained by the AWGM with $p = 2$ (right). Here, the barycenters ('+') of $\operatorname{supp} \psi_{\vec{\lambda}}$ for $\vec{\lambda} \in \Lambda_k$ are plotted.

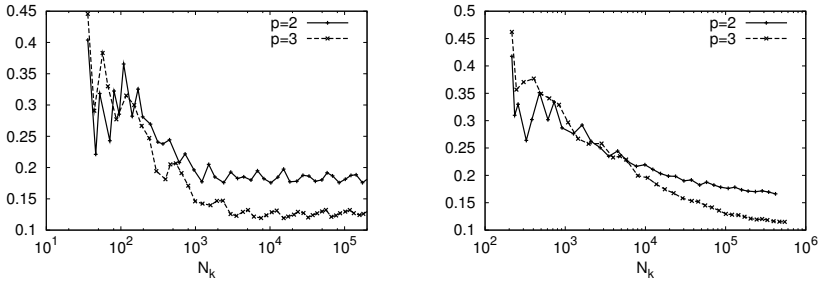


Fig. 3 Relative error in $\mathbf{r}^{(k)}$ (see (43)) for $n = 2$ (left) and $n = 3$ (right).

Efficiency of the new approximate residual

We also compared the multi-tree based residual $\mathbf{r}^{(k)}$ with the APPLY based residual as defined in (39). To this end, we selected ε_k such that $\|\mathbf{r}^{(k)}_{\mathtt{ex}} - \mathbf{r}^{(k)}_{\mathtt{APPLY}}(\varepsilon_k)\|_{\ell_2(\nabla_\Psi)}$ is close to $\|\mathbf{r}^{(k)}_{\mathtt{ex}} - \mathbf{r}^{(k)}\|_{\ell_2(\nabla_\Psi)}$. To optimize the computation of $\mathbf{r}^{(k)}_{\mathtt{APPLY}}(\varepsilon_k)$, we used the APPLY routine stated in [26, Section 3] which, by our experiences, yields quantitatively the best results. Moreover, the special form of $\mathbf{A}$ stated in (38) is exploited to keep the length as well as computation time for $\mathbf{r}^{(k)}_{\mathtt{APPLY}}(\varepsilon_k)$ as short as possible (see also [13]).

It is clear that for an efficient implementation of the AWGM, residuals of short length are mandatory. So, first we compare the lengths of the supports for the two different type of approximate residuals. The corresponding results are shown in Figure 4. For spatial dimension $n = 2$, we found for the relative length $\#\operatorname{supp}\mathbf{r}^{(k)}/N_k \approx 5$ ($N_k = \#\Lambda_k$), independent of the polynomial degree p. For $n = 3$, we estimated $\#\operatorname{supp}\mathbf{r}^{(k)}/N_k \approx 10$. The corresponding support sizes of the APPLY based residual approximation are much larger. As it can be seen from Figure 4, for $n = 2$, the support length of the APPLY based approximate residual is about a factor of 10 ($p = 3$) to 11 ($p = 2$) times larger than the corresponding support length of the multi-tree based one. For $n = 3$, we find corresponding factors $\#\operatorname{supp}\mathbf{r}^{(k)}_{\mathtt{APPLY}}/\#\operatorname{supp}\mathbf{r}^{(k)}$ of approximately 8 ($p = 3$) and 9 ($p = 2$).

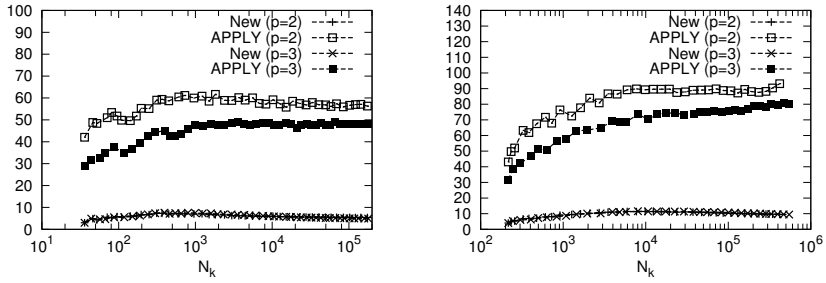
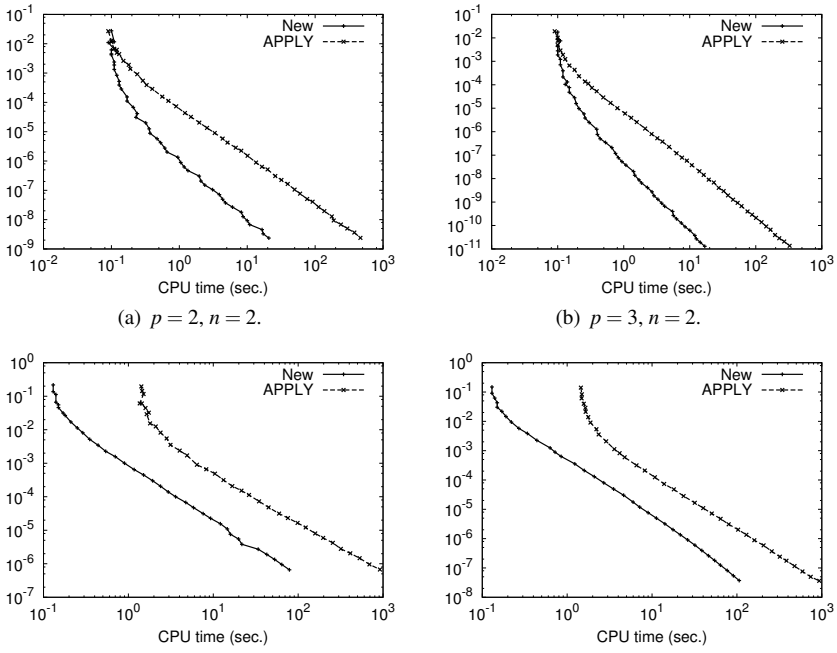


Fig. 4 Relative length of residuals, i.e., $\#\mathrm{supp}\,\mathbf{r}^{(k)}/N_k$ and $\#\mathrm{supp}\,\mathbf{r}^{(k)}_{\mathtt{APPLY}}(\varepsilon_k)/N_k$ where $N_k := \#\Lambda_k$ for $p \in \{2,3\}$ and $n = 2$ (left) and $n = 3$ (right).

Beyond support lengths of the approximate residuals, we also investigated the corresponding computation times. Here, except for a multi-tree Λ_k of relatively small cardinality, we observed for $n = 2$ that the computation of $\mathbf{r}^{(k)}$ is a factor of about 14 ($p = 3$) to 16 ($p = 2$) times faster than the computation of $\mathbf{r}^{(k)}_{\mathtt{APPLY}}(\varepsilon_k)$. Correspondingly, for spatial dimensional $n = 3$, the computation of the multi-tree based approximate residual is approximately 9 ($p = 3$) to 10 ($p = 2$) times faster than the `APPLY` based one.



(a) $p = 2, n = 2$. (b) $p = 3, n = 2$.

(c) $p = 2, n = 3$. (d) $p = 3, n = 3$.

Fig. 5 The residual approximation errors $\|\mathbf{r}^{(k)}_{\mathtt{ex}} - \mathbf{r}^{(k)}\|_{\ell_2(\nabla_\Psi)}$ ('New') and $\|\mathbf{r}^{(k)}_{\mathtt{ex}} - \mathbf{r}^{(k)}_{\mathtt{APPLY}}(\varepsilon_k)\|_{\ell_2(\nabla_\Psi)}$ ('APPLY') are plotted against the corresponding computation times.

Remark 12 It is possible to use larger values of ℓ (i.e., $\ell > 1$) in (41) or alternative constructions of $\tilde{\Lambda}_k$, e.g.,

$$\tilde{\Lambda}_k = \{\vec{\lambda} \in \nabla_\Psi : \exists \vec{\mu} \in \Lambda_k \text{ s.t. } \forall i, |\lambda_i| \leq |\mu_i| + \ell \wedge \mathrm{d}(\operatorname{supp} \psi^{(i)}_{\lambda_i}, \operatorname{supp} \psi^{(i)}_{\mu_i}) \leq D_i 2^{-|\lambda_i|}\},$$

However, we observed that the corresponding residuals have much larger supports and, correspondingly, higher computations times than $\mathbf{r}^{(k)}$. At the same time, this effort did not pay off in form of faster convergence of the `AWGM`.

Computation times

In Figure 6, we show the computation times for the application of $\mathbf{R}_{\Lambda_k}\mathbf{A}\mathbf{I}_{\Lambda_k}$ (required for solving associated Galerkin systems) and for computing the residual $\mathbf{r}^{(k)}$ for $n = 3$. These show that also for higher number of degrees of freedoms ($\#\Lambda_k > 10^5$), an efficient implementation of the `AWGM` is possible. Here, we benefit from the structure of $\mathbf{A}$ being due to the discretization of (37) by L_2-orthonormal wavelet bases as described in Section 4.1.

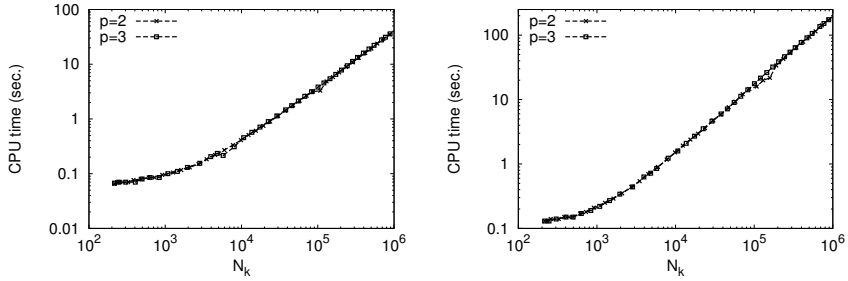


Fig. 6 Computation times for the application of $\mathbf{R}_{\Lambda_k}\mathbf{A}\mathbf{I}_{\Lambda_k}$ (left) and for computing $\mathbf{r}^{(k)}$ (right) for $n = 3$ ($N_k = \#\Lambda_k$).

Acknowledgements The authors are grateful to Andreas Rupp from the Institute for Numerical Mathematics at the University of Ulm for providing implementations of L_2-orthonormal multi-wavelets on the interval.

References

1. S. Achatz. *Adaptive finite Dünngitter-Elemente höherer Ordnung für elliptische partielle Differentialgleichungen mit variablen Koeffizienten*. PhD thesis, Technische Universität München, 2003.
2. R. Balder. *Adaptive Verfahren für elliptische und parabolische Differentialgleichungen auf dünnen Gittern*. PhD thesis, Technische Universität München, 1994.
3. R. Balder and Ch. Zenger. The solution of multidimensional real Helmholtz equations on sparse grids. *SIAM J. Sci. Comput.*, 17(3):631–646, 1996.
4. H.J. Bungartz. *Dünne Gitter und deren Anwendung bei der adaptiven Lösung der dreidimensionalen Poisson-Gleichung*. PhD thesis, Technische Universität München, 1992.
5. H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numer.*, 13:147–269, 2004.
6. N.G. Chegini, S. Dahlke, U. Friedrich, and R.P. Stevenson. Piecewise tensor product wavelet bases by extensions and approximation rates. Technical report, KdV Institute for Mathematics, University of Amsterdam, 2011. To appear in *Math. Comp.*

7. N.G. Chegini and R.P. Stevenson. The adaptive tensor product wavelet scheme: Sparse matrices and the application to singularly perturbed problems. *IMA J. Numer. Anal.*, 32(1):75–104, 2011.
8. N.G. Chegini and R.P. Stevenson. Adaptive wavelets schemes for parabolic problems: Sparse matrices and numerical results. *SIAM J. Numer. Anal.*, 49(1):182–212, 2011.
9. A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore. Tree approximation and optimal encoding. *Appl. Comput. Harmon. Anal.*, 11(2):192–226, 2001.
10. A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods for elliptic operator equations – Convergence rates. *Math. Comp*, 70:27–75, 2001.
11. M. Dauge and R.P. Stevenson. Sparse tensor product wavelet approximation of singular functions. *SIAM J. Math. Anal.*, 42(5):2203-2228, 2010.
12. R. Dautray and J.-L. Lions. *Mathematical analysis and numerical methods for science and technology. Vol. 5*. Springer-Verlag, Berlin, 1992. Evolution problems I.
13. T.-J. Dijkema, C. Schwab, and R.P. Stevenson. An adaptive wavelet method for solving high-dimensional elliptic PDEs. *Constructive Approximation*, 30(3):423–455, 2009.
14. T.-J. Dijkema and R.P. Stevenson. A sparse Laplacian in tensor product wavelet coordinates. *Numer. Math.*, 115(3):433–449, 2010.
15. G. C. Donovan, J. S. Geronimo and D. P. Hardin. Intertwining multiresolution analyses and the construction of piecewise-polynomial wavelets. *SIAM J. Math. Anal.*, 27(6):1791–1815, 1996.
16. G. C. Donovan, J. S. Geronimo and D. P. Hardin. Orthogonal polynomials and the construction of piecewise polynomial smooth wavelets. *SIAM J. Math. Anal.*, 30(5):1029–1056, 1999.
17. C. Feuersänger. Dünngitterverfahren für hochdimensionale elliptische partielle Differentialgleichungen. Master's thesis, Institut für Numerische Simulation, Universität Bonn, 2005.
18. T. Gantumur, H. Harbrecht, and R.P. Stevenson. An optimal adaptive wavelet method without coarsening of the iterands. *Math. Comp.*, 76:615–629, 2007.
19. M. Griebel. Adaptive sparse grid multilevel methods for elliptic PDEs based on finite differences. *Computing*, 61(2):151–179, 1998.
20. M. Hegland. Adaptive sparse grids. *ANZIAM Journal*, 44:C335–C353, 2003.
21. S. Kestler and R.P. Stevenson. Fast evaluation of system matrices w.r.t. multi-tree collections of tensor product refinable basis functions. Technical report, 2012. Submitted.
22. S. Kestler and K. Urban. Adaptive wavelet methods on unbounded domains. *J. Sci. Comput.*, 53(2):342–376, 2012.
23. A. Niedermeier and S. Zimmer. Implementational aspects of prewavelet sparse grid methods. In *Eleventh International Conference on Domain Decomposition Methods (London, 1998)*, pages 314–321. 1999.
24. D. Pflüger. *Spatially Adaptive Sparse Grids for High-Dimensional Problems*. PhD thesis, Technische Universität München, 2010.
25. K. Schneider, M. Farge, F. Koster, and M. Griebel. Adaptive wavelet methods for the Navier-Stokes equations. In *Numerical flow simulation, II*, volume 75 of *Notes Numer. Fluid Mech.*, pages 303–318. Springer, Berlin, 2001.
26. R.P. Stevenson. Adaptive methods for solving operator equations: An overview. In R. DeVore and A. Kunoth, editors, *Multiscale, Nonlinear and Adaptive Approximation: Dedicated to Wolfgang Dahmen on the Occasion of his 60th Birthday*, pages 543–598. Springer (Berlin), 2009.
27. R.P. Stevenson. Adaptive wavelet methods for linear and non-linear least squares problems. Technical report, KdVI, UvA Amsterdam, November 2011. Submitted.
28. A. Stippler. LAWA - Library for Adaptive Wavelet Applications. *http://lawa.sourceforge.net*, 2009. (last accessed February 16, 2013).
29. A. Rupp. PhD thesis, University of Ulm, to appear 2013.
30. A. Zeiser. Fast matrix-vector multiplication in the sparse-grid Galerkin method. *J. Sci. Comput.*, 47(3):328–346, 2011.
31. C. Zenger. Sparse grids. In W. Hackbusch, editor, *Parallel algorithms for partial differential equations*, volume 31 of *Notes on numerical fluid mechanics*, pages 241–251. Vieweg, 1991.