**CONSTRUCTIVE
APPROXIMATION**

An Adaptive Wavelet Method for Solving High-Dimensional Elliptic PDEs

**Tammo Jan Dijkema · Christoph Schwab ·
Rob Stevenson**

Abstract Adaptive tensor product wavelet methods are applied for solving Poisson's equation, as well as anisotropic generalizations, in high space dimensions. It will be demonstrated that the resulting approximations converge in energy norm with the same rate as the best approximations from the span of the best N tensor product wavelets, where moreover the constant factor that we may lose is independent of the space dimension n. The cost of producing these approximations will be proportional to their length with a constant factor that may grow with n, but only linearly.

Communicated by Ian Sloan.

T.J. Dijkema
Department of Mathematics, Utrecht University, P.O. Box 80.010, 3508 TA Utrecht, The Netherlands
e-mail: t.j.dijkema@uu.nl

C. Schwab
Seminar for Applied Mathematics, ETHZ HG G58.1, ETH Zürich, 8092 Zürich, Switzerland
e-mail: schwab@sam.math.ethz.ch

R. Stevenson (✉)
Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Plantage Muidergracht 24,
1018 TV Amsterdam, The Netherlands
e-mail: R.P.Stevenson@uva.nl

1 Introduction

On $\Omega = (0, 1)^n$, and with Γ_D the union of one or more $(n - 1)$-dimensional faces of $\partial\Omega$, for given $f \in (H^1_{0,\Gamma_D}(\Omega))'$, we study the numerical solution of the problem of finding $u \in H^1_{0,\Gamma_D}(\Omega)$ such that

$$a(u, v) := \int_\Omega c_0 uv + \sum_{m=1}^n c_m \partial_m u \partial_m v = f(v) \quad \left(v \in H^1_{0,\Gamma_D}(\Omega)\right), \qquad (1.1)$$

where $c_0 \geq 0$ and $c_m > 0$ $(m = 1, \ldots, n)$ are constants.

PDEs such as (1.1) in spatial domains of high dimension arise in numerous areas. We only mention mathematical finance (valuation of derivative contracts on large baskets), elliptic homogenization problems with multiple separated length scales (e.g., [23]), deterministic methods for stochastic PDEs [28], the N-electron Schrödinger equation in molecular dynamics [29] and problems from molecular biology.

Since the work of R. Feynman [13], (strong) solutions $u(x)$ at a point $x \in \Omega$ of the Dirichlet Problem (1.1) are well known to be characterized by the first exit-time of the n-dimensional Wiener process W_t^x started at $t = 0$ in $x \in \Omega$ (see, e.g., [14] and the references therein for a detailed account). For large dimension n, this probabilistic characterization of $u(x)$ can be used, in conjunction with the Monte Carlo Simulation of sample paths of W_t^x, to obtain numerical approximations to $u(x)$ *in a single point* $x \in \Omega$ of accuracy $O(1/\sqrt{m})$, where m denotes the number of sample paths; the work for generating m paths in $\mathbb{R}^n$ scales linearly with n.

Here, our aim is to develop deterministic approximations of $u(x)$ *at all points* $x \in (0, 1)^n$ that are efficient for large n. With a standard, piecewise polynomial approximation procedure in $(0, 1)^n$, the error in energy norm

$$\|\|\cdot\|\| := a(\cdot, \cdot)^{\frac{1}{2}} \qquad (1.2)$$

on $H^1_0(\Omega)$ behaves at best as $\sim N^{-(d-1)/n}$, where $d \geq 2$ is the polynomial order and N is the number of degrees of freedom. The rate $(d - 1)/n$ being inversely proportional to the space dimension n is known as *curse of dimensionality*.

Using that $\Omega = (0, 1)^n$ is a tensor product domain, the curse of dimensionality can be circumvented by applying hyperbolic cross approximation or sparse grids [2, 30]. With this approach, for any fixed n the error behaves as $\sim (\log N)^{(n-1)(d-1)} \times N^{-(d-1)}$, or, with some small modification (see [20]), even as $\sim N^{-(d-1)}$, *assuming* that $u \in \bigcap_{k=1}^n \bigotimes_{m=1}^n H^{d-1+\delta_{km}}(0, 1)$ or, for some $\varepsilon > 0$, $u \in \bigcap_{k=1}^n \bigotimes_{m=1}^n \times H^{d-1+\delta_{km}+(1-\delta_{km})\varepsilon}(0, 1)$, respectively. It can be shown that these regularity conditions are satisfied for smooth f that satisfy homogeneous Dirichlet boundary conditions of a sufficiently high order (for $n \geq 2$, an order larger than $d - \frac{5}{2} - \frac{1}{n}$ suffices). In a forthcoming work [10], we will show that for $n \geq 2$, $\Gamma_D = \partial\Omega$ and general smooth f that do not vanish at the boundary, the solution $u \in \bigcap_{k=1}^n \bigotimes_{m=1}^n H^{s+\delta_{km}}(0, 1)$ if and only if $s < \frac{1}{2} + \frac{1}{n}$, essentially thus being the rate of these sparse approximations regardless of d.

The key to overcoming regularity restrictions is to apply *nonlinear approximation*. Let us write the Dirichlet boundary as

$$\Gamma_D := \bigcup_{m=1}^{n} [0,1]^{m-1} \times z_m \times [0,1]^{n-m},$$

where for $1 \le m \le n$, $z_m \subseteq \{0,1\}$ with $\bigcup_{m=1}^{n} z_m \neq \emptyset$. For $z \subseteq \{0,1\}$, let $\{\psi_\lambda^{(z)} : \lambda \in \nabla_z\}$ be a Riesz basis for $L_2(0,1)$ consisting of *wavelets* of order d, such that $\{2^{-|\lambda|} \psi_\lambda^{(z)} : \lambda \in \nabla_z\}$ is a Riesz basis for $H_{0,z}^1(0,1)$. As usual, here $|\lambda| \in \mathbb{N}_0$ denotes the *level* of $\psi_\lambda^{(z)}$, and for $\ell \in \mathbb{N}_0$, $\#\{\lambda \in \nabla_z : |\lambda| = \ell\}$ is of the order of 2^ℓ. Now for any n, the normalized *tensor product basis*

$$\Psi := \left\{ \psi_\lambda := \bigotimes_{m=1}^{n} \psi_{\lambda_m}^{(z_m)} \Big/ \left\| \bigotimes_{m=1}^{n} \psi_{\lambda_m}^{(z_m)} \right\| : \lambda \in \nabla := \prod_{m=1}^{n} \nabla_{z_m} \right\} \qquad (1.3)$$

is a Riesz basis for $H_{0,\Gamma_D}^1(0,1)^n$, even uniformly in $c_0 \ge 0$ and $c_m > 0$ $(m = 1, \dots, n)$ when this space is equipped with norm $\|\|\cdot\|\|$. This means that *the condition number of Ψ with respect to* $\|\|\cdot\|\|$, being the quotient

$$\sup_{0 \neq \mathbf{v} \in \ell_2(\nabla)} \frac{\|\|\sum_{\lambda \in \nabla} v_\lambda \psi_\lambda\|\|^2}{\|\mathbf{v}\|_{\ell_2(\nabla)}^2} \Big/ \inf_{0 \neq \mathbf{v} \in \ell_2(\nabla)} \frac{\|\|\sum_{\lambda \in \nabla} v_\lambda \psi_\lambda\|\|^2}{\|\mathbf{v}\|_{\ell_2(\nabla)}^2} \qquad (1.4)$$

is bounded, uniformly in $c_0 \ge 0$ and $c_m > 0$ $(m = 1, \dots, n)$. With $|\lambda| := (|\lambda_1|, \dots, |\lambda_n|)$, for any $\ell \in \mathbb{N}_0$, $\{\psi_\lambda : \lambda \in \nabla, \||\lambda|\|_1 \le \ell\}$ spans the corresponding sparse grid space of dimension N being of the order of $2^\ell \ell^{n-1}$.

Instead of restricting ourselves to sparse grid approximation, we consider approximations to u from the span of $\{\psi_\lambda : \lambda \in \Lambda_N\}$, where $\Lambda_N \subset \nabla$ is *any* subset with $\#\Lambda_N = N$. Because of the boundedness of the condition number, approximating u by $\sum_{\lambda \in \Lambda_N} v_\lambda \psi_\lambda$ in $\|\|\cdot\|\|$ is equivalent to approximating its representation $\mathbf{u}$ with respect to Ψ by $(v_\lambda)_{\lambda \in \Lambda_N}$ in $\|\cdot\|_{\ell_2(\nabla)}$. We have $\inf_{\{\mathbf{v} \in \ell_2(\nabla) : \text{supp } \mathbf{v} \subset \Lambda_N\}} \|\mathbf{u} - \mathbf{v}\|_{\ell_2(\nabla)} = \|\mathbf{u} - \mathbf{u}|_{\Lambda_N}\|_{\ell_2(\nabla)}$, where $\mathbf{u}|_{\Lambda_N}$ is the vector in $\ell_2(\nabla)$ that coincides with $\mathbf{u}$ on its support being Λ_N. For approximating $\mathbf{u}$ in $\|\cdot\|_{\ell_2(\nabla)}$, a best choice for Λ_N is one such that $\mathbf{u}|_{\Lambda_N}$ is a *best N-term approximation* to $\mathbf{u}$, denoted as $\mathbf{u}_N$, meaning that it contains the N largest coefficients in modulus of $\mathbf{u}$.

The class

$$\mathcal{A}_\infty^s := \left\{ \mathbf{v} \in \ell_2(\nabla) : \|\mathbf{v}\|_{\mathcal{A}_\infty^s} := \sup_{\varepsilon > 0} \varepsilon \times \left[\min\{N \in \mathbb{N}_0 : \|\mathbf{v} - \mathbf{v}_N\|_{\ell_2(\nabla)} \le \varepsilon\} \right]^s < \infty \right\}^1$$

gathers under one roof all $\mathbf{v}$ whose best N-term approximations converge to $\mathbf{v}$ with rate $s > 0$. Since for $\varepsilon < \|\mathbf{v}\|$, the value of N in the definition of $\|\mathbf{v}\|_{\mathcal{A}_\infty^s}$ is positive, note that $\|\mathbf{v}\|_{\mathcal{A}_\infty^s} \ge \sup_{\varepsilon < \|\mathbf{v}\|_{\ell_2(\nabla)}} \varepsilon = \|\mathbf{v}\|_{\ell_2(\nabla)}$. Although $\mathcal{A}_\infty^s$ is non-empty for any

---

[1] We have $2^{-s}\|\|\cdot\|\|_{\mathcal{A}_\infty^s} \le \|\cdot\|_{\mathcal{A}_\infty^s} \le \|\|\cdot\|\|_{\mathcal{A}_\infty^s}$, where $\|\|\mathbf{v}\|\|_{\mathcal{A}_\infty^s} := \sup_{N \in \mathbb{N}_0} (N+1)^s \|\mathbf{v} - \mathbf{v}_N\|_{\ell_2(\nabla)}$ is the common definition of the (quasi) norm on $\mathcal{A}_\infty^s$.

s, as it contains any finitely supported vector, in view of the order d of polynomial reproduction being applied, the representation $\mathbf{v}$ of an arbitrarily smooth v cannot be expected to be in $\mathcal{A}_\infty^s$ for $s > d - 1$. On the other hand, in [26], Nitsche showed that for sufficiently smooth wavelets, e.g., spline wavelets, for $0 < s < d - 1$ and with $\tau = (s + \frac{1}{2})^{-1}$,

$$\mathbf{v} \in \mathcal{A}_\tau^s \quad \Longleftrightarrow \quad v \in \bigcap_{k=1}^n \bigotimes_{m=1}^n {}_\tau B_\tau^{s+\delta_{mk}}\big(L_\tau(0,1)\big),$$

where $\mathcal{A}_\tau^s := \{\mathbf{v} \in \ell_2(\nabla) : \sum_{N \in \mathbb{N}} (N^s \|\mathbf{v} - \mathbf{v}_N\|_{\ell_2(\nabla)})^\tau N^{-1} < \infty\}$, $B_p^t(L_p(0,1))$ is a Besov space measuring "t orders of smoothness in L_p", and $\bigotimes_\tau$ denotes the so-called "τ tensor product". Note that $\mathcal{A}_\tau^s$ is even (slightly) smaller than $\mathcal{A}_\infty^s$.

The upshot of this result is that $\bigcap_{k=1}^n \bigotimes_{\tau m=1}^n B_\tau^{s+\delta_{mk}}(L_\tau(0,1))$ is much larger, with an increasing difference when s and n get larger, than $\bigcap_{k=1}^n \bigotimes_{m=1}^n H^{s+\delta_{km}}(0,1)$, membership of which is needed to guarantee the same rate with sparse grid approximation.

In particular, for two and three space dimensions, in [25] it was shown that for general, sufficiently smooth f, the solution u of (1.1) is in $\mathcal{A}_\infty^s$ for any $s < d - 1$, and actually even that the best N-term approximations converge with rate $d - 1$ up to some log-factors (in a forthcoming paper, we will even show that $u \in \mathcal{A}_\infty^{d-1}$). The proof in [25] makes use of the splitting of u into known singular functions and a smooth remainder, which is also available for (1.1) in more than three dimensions in [8]. It indicates the potential of best N-term approximation in tensor product bases. For background on nonlinear approximation, we refer for example to [3, 9].

The above considerations concern best N-term approximations that, however, are not feasible in practice, already because $\mathbf{u}$ is not given explicitly. It can be found as the solution of the infinite matrix-vector problem

$$\mathbf{Au} = \mathbf{f}, \tag{1.5}$$

with "stiffness matrix" $\mathbf{A} = [a(\boldsymbol{\psi}_\mu, \boldsymbol{\psi}_\lambda)]_{\lambda,\mu \in \nabla}$ and $\mathbf{f} = [f(\boldsymbol{\psi}_\lambda)]_{\lambda \in \nabla}$. This infinite matrix problem is equivalent to (1.1).

In [4, 5], optimal *adaptive* algorithms were introduced for solving (1.5). It was shown that whenever for some $s > 0$, $\mathbf{u}$ happens to belong to $\mathcal{A}_\infty^s$, then, under two assumptions discussed below, the sequence of approximations produced by these algorithms converge to $\mathbf{u}$ with *this rate s*, requiring a number of operations equivalent to their length. The first assumption is that one has available a routine **RHS** that given an $\varepsilon > 0$ produces an approximation $\mathbf{f}_\varepsilon := \mathbf{RHS}[\varepsilon]$ with $\|\mathbf{f} - \mathbf{f}_\varepsilon\| \le \varepsilon$ and

$$\mathbf{rhs}_s := \sup_{0 < \varepsilon < \|\mathbf{f}\|_{\ell_2(\nabla)}} \varepsilon \times \big[\# \text{ operations required by the call } \mathbf{RHS}[\varepsilon]\big]^s < \infty. \tag{1.6}$$

The second assumption is that for some $s^* > s$, $\mathbf{A}$ is s^*-computable, which is a quantitative measure of how well $\mathbf{A}$ can be approximated by computable sparse matrices. This assumption implies that $\mathbf{A} : \mathcal{A}_\infty^s \to \mathcal{A}_\infty^s$ is bounded [4], and thus that $\|\mathbf{f}\|_{\mathcal{A}_\infty^s} \lesssim \|\mathbf{u}\|_{\mathcal{A}_\infty^s}$, where the $\lesssim$ symbol has its usual definition recalled at the end of

this introduction. So in any case if we could realize (quasi-) best N-term approximations for $\mathbf{f}$ in $\mathcal{O}(N)$ operations, then the resulting routine **RHS** would satisfy $\mathbf{rhs}_s < \infty$, in particular with $\mathbf{rhs}_s \lesssim \|\mathbf{u}\|_{\mathcal{A}_\infty^s}$.

In the tensor product basis setting, applying as a building block univariate biorthogonal spline wavelets, in [27] we proved that $\mathbf{A}$ is s^*-computable for some $s^* > d - 1$, which s^* is thus larger than any s for which $\mathbf{u} \in \mathcal{A}_\infty^s$ might be expected. We showed this result in a more general setting than we consider here. Most importantly, we allowed general, i.e., non-separable, smooth variable coefficients in the differential operator, meaning that we had to design and analyze a quadrature scheme. Our results imply that for any n, and $s \in [0, d - 1]$, the adaptive wavelet schemes produce an approximation to $\mathbf{u} \in \mathcal{A}_\infty^s$ within any given *tolerance* $0 < \varepsilon \lesssim \|\mathbf{f}\|_{\ell_2(\nabla)}$ in $\| \cdot \|_{\ell_2(\nabla)}$ with a *support length* not exceeding

$$D_1 \varepsilon^{-1/s} |\mathbf{u}|_{\mathcal{A}_\infty^s}^{1/s},$$

taking a *number of operations* not exceeding

$$D_2 \varepsilon^{-1/s} |\mathbf{u}|_{\mathcal{A}_\infty^s}^{1/s} + D_3 \varepsilon^{-1/s} \mathbf{rhs}_s^{1/s},$$

where D_1, D_2 and D_3 are some constants, independent of $c_0 \geq 0$ and $c_m > 0$ $(m = 1, \ldots, n)$. Only knowing that $\mathbf{u} \in \mathcal{A}_\infty^s$, up to the factors D_1 or $D_2 + D_3[\mathbf{rhs}_s/|\mathbf{u}|_{\mathcal{A}_\infty^s}]^{1/s}$, this length or number of operations are indeed the best that generally can be expected.

What we did not analyze, however, is the

dependence of the constants D_1, D_2, D_3 and the condition number of $\boldsymbol{\Psi}$ with respect to $\|\|\cdot\|\|$ on the space dimension n.

Concerning the latter, note that instead of approximating $\mathbf{u}$, our ultimate goal is to approximate u in $\|\|\cdot\|\|$ within some given tolerance with, up to some constant factor, the smallest linear combination of wavelets.

The condition number of $\boldsymbol{\Psi}$ with respect to $\|\|\cdot\|\|$ is equal to the spectral condition number $\kappa(\mathbf{A})$ of $\mathbf{A}$, which for biorthogonal wavelets can be expected to grow exponentially with n. Since in any case also D_2 is an increasing function of $\kappa(\mathbf{A})$, it may be that, although optimal for any fixed n, the method has only practical value for relatively small values of n.

In view of this, in the current paper we apply univariate $L_2(0, 1)$-*orthonormal, piecewise polynomial* wavelets as introduced in [11, 12, 19]. In this case, $\kappa(\mathbf{A})$ is bounded *uniformly in n*, and $c_0 \geq 0$ and $c_m > 0$ $(m = 1, \ldots, n)$. Thanks to both the L_2-orthonormality and the fact that the wavelets are piecewise polynomial, the stiffness matrix $\mathbf{A}$ appears to be very close to a sparse matrix. We give a detailed description of an adaptive wavelet algorithm for which aforementioned statements are valid with

$$D_1, \frac{D_2}{n}, D_3 \text{ constants, } \textit{independent of } n \text{ and } c_0 \geq 0, c_m > 0 \ (m = 1, \ldots, n).$$

So only the constant involved in the operation count may grow with the space dimension, but only linearly.

The algorithm we use is a modification developed in [16] of the adaptive wavelet method from [4]. With this modification, the recurrent coarsening of the iterands is avoided, which yields a quantitatively better performing algorithm. In the current paper, an additional quantitative improvement will be obtained by the use of a modified approximate matrix vector routine **APPLY** that will use a posteriori information to optimize the accuracies with which the columns of the infinite stiffness matrix are approximated.

The remainder of this paper is organized as follows. In Sect. 2, from [11] we recall the construction of $L_2(0, 1)$-orthonormal, piecewise polynomial wavelets. We show that, as a consequence of the $L_2(0, 1)$-orthonormality, the stiffness matrix **A** of the n-dimensional Laplacian in tensor product wavelet coordinates is well-conditioned uniformly in n.

In Sect. 3, we construct sparse approximations to **A**. Using that the wavelets are $L_2(0, 1)$-orthonormal and piecewise polynomial, we construct approximations that with a multiple of jn non-zeros per column have an error of order $2^{-j/2}$.

Sect. 4 deals with adaptive wavelet schemes for solving general bi-infinite matrix-vector equations with symmetric positive definite system matrices. The construction and the theory behind these schemes from [4, 16] is summarized. The most time-consuming ingredient of such schemes is the approximate matrix-vector routine **APPLY**. We develop a new version of this routine that performs an optimization based on a posteriori optimization, and which has strongly improved quantitative properties.

In Sect. 5, we specialize the results about adaptive wavelet schemes to the application of solving problem (1.1) in high space dimensions.

Finally, in Sect. 6 we discuss the implementation of the scheme and present numerical results.

In this paper, unless explicitly stated otherwise, by $C \lesssim D$ we will mean that C can be bounded by a multiple of D, independently of parameters which C and D may depend on, in particular the space dimension n or the coefficients c_m ($0 \leq m \leq n$). Obviously, $C \gtrsim D$ is defined as $D \lesssim C$, and $C \eqsim D$ as $C \lesssim D$ and $C \gtrsim D$.

We use notations like $\vec{v}$ or $\vec{B}$ to indicate vectors in $\ell_2(\nabla_z)$ or linear mappings $\ell_2(\nabla_z) \to \ell_2(\nabla_z)$, respectively, and **v** or **B** to indicate vectors in $\ell_2(\nabla)$ or linear mappings $\ell_2(\nabla) \to \ell_2(\nabla)$. We use the notation $\| \cdot \|$ to abbreviate all of $\| \cdot \|_{L_2(0,1)}$, $\| \cdot \|_{L_2(0,1)^n}$, $\| \cdot \|_{\ell_2(\nabla_z)}$, $\| \cdot \|_{\ell_2(\nabla_z) \to \ell_2(\nabla_z)}$, $\| \cdot \|_{\ell_2(\nabla)}$ and $\| \cdot \|_{\ell_2(\nabla) \to \ell_2(\nabla)}$. We write $\langle \cdot, \cdot \rangle$ instead of $\langle \cdot, \cdot \rangle_{\ell_2(\nabla_z)}$ or $\langle \cdot, \cdot \rangle_{\ell_2(\nabla)}$. We use the notation $\|| \cdot \||$, earlier defined as $a(\cdot, \cdot)^{\frac{1}{2}}$, also to denote $\langle \mathbf{A} \cdot, \cdot \rangle^{\frac{1}{2}}$. Note that for any $\mathbf{v} \in \ell_2(\nabla)$, $\|| \mathbf{v} \|| = \|| \sum_{\lambda \in \nabla} \mathbf{v}_\lambda \boldsymbol{\psi}_\lambda \||$.

2 Univariate $L_2(0, 1)$-orthonormal, Piecewise Polynomial Wavelets

For $z \subseteq \{0, 1\}$, let $\{\psi_\lambda^{(z)} : \lambda \in \nabla_z\}$ be a Riesz basis for $L_2(0, 1)$, normalized with respect to $\| \cdot \|$, such that $\{2^{-|\lambda|} \psi_\lambda^{(z)} : \lambda \in \nabla_z\}$ is a Riesz basis for $H_{0,z}^1(0, 1)$. Let $\lambda_{\max}^{(0,z)}, \lambda_{\max}^{(1,z)} (\geq 1)$ and $\lambda_{\min}^{(0,z)}, \lambda_{\min}^{(1,z)} (\leq 1)$ be the smallest or largest constants such that

$$\lambda_{\min}^{(0,z)} \leq \left\| \sum_{\lambda \in \nabla_z} \vec{v}_\lambda \psi_\lambda^{(z)} \right\|^2 \bigg/ \|\vec{v}\|^2 \leq \lambda_{\max}^{(0,z)}, \quad \left(\vec{v} \in \ell_2(\nabla_z) \right),$$

$$\lambda_{\min}^{(1,z)} \leq \left\| \sum_{\lambda \in \nabla_z} \vec{v}_\lambda \dot{\psi}_\lambda^{(z)} \right\|^2 \bigg/ \sum_{\lambda \in \nabla_z} |\vec{v}_\lambda|^2 \|\dot{\psi}_\lambda^{(z)}\|^2 \leq \lambda_{\max}^{(1,z)}, \quad \left((2^{|\lambda|} \vec{v}_\lambda)_\lambda \in \ell_2(\nabla_z) \right).$$

$$(2.1)$$

(Throughout this paper, a "dot" on top of a univariate function denotes its derivative). Then, with $\nabla := \prod_{m=1}^{n} \nabla_{z_m}$, by a tensor product argument we have

$$\prod_{m=1}^{n} \lambda_{\min}^{(0,z_m)} \leq \left\| \sum_{\boldsymbol{\lambda} \in \nabla} \mathbf{v}_{\boldsymbol{\lambda}} \bigotimes_{k=1}^{n} \psi_{\lambda_k}^{(z_k)} \right\|^2 \bigg/ \|\mathbf{v}\|^2 \leq \prod_{m=1}^{n} \lambda_{\max}^{(0,z_m)}, \quad \left(\mathbf{v} \in \ell_2(\nabla) \right),$$

$$\lambda_{\min}^{(1,z_m)} \prod_{k \neq m} \lambda_{\min}^{(0,z_k)} \leq \left\| \sum_{\boldsymbol{\lambda} \in \nabla} \mathbf{v}_{\boldsymbol{\lambda}} \partial_m \bigotimes_{k=1}^{n} \psi_{\lambda_k}^{(z_k)} \right\|^2 \bigg/ \sum_{\boldsymbol{\lambda} \in \nabla} |\mathbf{v}_{\boldsymbol{\lambda}}|^2 \|\dot{\psi}_{\lambda_m}^{(z_m)}\|^2 \leq \lambda_{\max}^{(1,z_m)} \prod_{k \neq m} \lambda_{\max}^{(0,z_k)}$$

$(1 \leq m \leq n, (2^{|\lambda_m|} \mathbf{v}_{\boldsymbol{\lambda}})_{\boldsymbol{\lambda} \in \nabla} \in \ell_2(\nabla))$, with bounds that cannot be improved. By summing over these inequalities, multiplied by c_0 or c_m $(1 \leq m \leq n)$, we infer that

$$\min_m \min\left(\lambda_{\min}^{(0,z_m)}, \lambda_{\min}^{(1,z_m)} \right) \prod_{k \neq m} \lambda_{\min}^{(0,z_k)}$$

$$\leq \frac{\|| \sum_{\boldsymbol{\lambda} \in \nabla} \mathbf{v}_{\boldsymbol{\lambda}} \bigotimes_{k=1}^{n} \psi_{\lambda_k}^{(z_k)} \||^2}{\sum_{\boldsymbol{\lambda} \in \nabla}[c_0 + \sum_{m=1}^{n} c_m \|\dot{\psi}_{\lambda_m}^{(z_m)}\|^2]|\mathbf{v}_{\boldsymbol{\lambda}}|^2} \leq \max_m \max\left(\lambda_{\max}^{(0,z_m)}, \lambda_{\max}^{(1,z_m)} \right) \prod_{k \neq m} \lambda_{\max}^{(0,z_k)}$$

$(([c_0 + \sum_{m=1}^{n} c_m 4^{|\lambda_m|}]^{\frac{1}{2}} \mathbf{v}_{\boldsymbol{\lambda}})_{\boldsymbol{\lambda}} \in \ell_2(\nabla))$. In view of the definition of $\boldsymbol{\Psi}$ in (1.3) and that of the stiffness matrix $\mathbf{A}$, using that $\|| \bigotimes_{m=1}^{n} \psi_{\lambda_m}^{(z_m)} \||^2 = c_0 + \sum_{m=1}^{n} c_m \|\dot{\psi}_{\lambda_m}^{(z_m)}\|^2$, we arrive at

$$\min_m \min\left(\lambda_{\min}^{(0,z_m)}, \lambda_{\min}^{(1,z_m)} \right) \prod_{k \neq m} \lambda_{\min}^{(0,z_k)}$$

$$\leq \frac{\langle \mathbf{A}\mathbf{v}, \mathbf{v} \rangle}{\|\mathbf{v}\|^2} = \frac{\|| \sum_{\boldsymbol{\lambda} \in \nabla} \mathbf{v}_{\boldsymbol{\lambda}} \boldsymbol{\psi}_{\boldsymbol{\lambda}} \||^2}{\|\mathbf{v}\|^2} \leq \max_m \max\left(\lambda_{\max}^{(0,z_m)}, \lambda_{\max}^{(1,z_m)} \right) \prod_{k \neq m} \lambda_{\max}^{(0,z_k)},$$

where moreover $\boldsymbol{\Psi}$ is a basis for $H_{0,\Gamma_D}^1(0,1)^n$ (cf. [21]).

So *if for all* $z \subseteq \bigcup_{m=1}^{n} z_m$, $\lambda_{\max}^{(0,z)} = \lambda_{\min}^{(0,z)}$, i.e., $\{\psi_\lambda^{(z)} : \lambda \in \nabla_z\}$ is an orthonormal basis for $L_2(0,1)$, *then*

$$\|\mathbf{A}\| \leq \max_z \lambda_{\max}^{(1,z)}, \qquad \|\mathbf{A}^{-1}\| \leq \max_z 1/\lambda_{\min}^{(1,z)}, \qquad (2.2)$$

which bounds are thus independent of n, $c_0 \geq 0$ and $c_m > 0$ $(1 \leq m \leq n)$. On the other hand, if for some $z \subseteq \bigcup_{m=1}^{n} z_m$, $\lambda_{\max}^{(0,z)} > \lambda_{\min}^{(0,z)}$, then one may verify that $\kappa(\mathbf{A}) := \|\mathbf{A}\| \|\mathbf{A}^{-1}\|$ grows exponentially with the number of coordinate directions that have
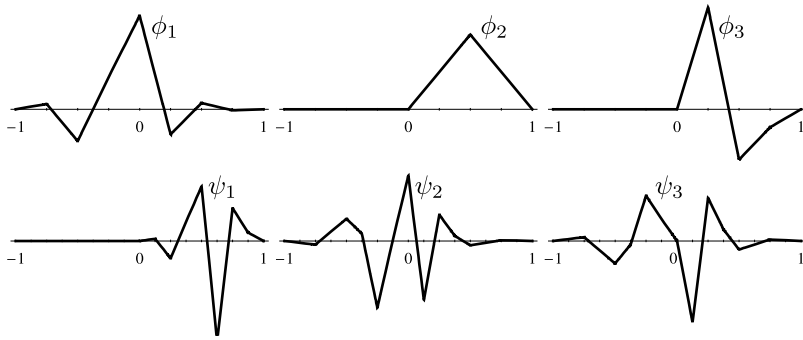
Fig. 1 $L_2(\mathbb{R})$-orthonormal, continuous piecewise linear scaling functions and wavelets. Their values in $\frac{1}{4}\mathbb{Z}$ or $\frac{1}{8}\mathbb{Z}$ can be found in [11]

these Dirichlet boundaries. So in order to be able to obtain satisfactory results for large n, there is no option other than to start with *univariate $L_2(0,1)$-orthonormal wavelet bases*.

In view of obtaining a stiffness matrix that can be sufficiently well approximated by sparse matrices, in addition we need *compactly supported wavelets*. Examples of orthonormal compactly supported wavelet bases are given by those from the Daubechies family [7]. With those wavelets, $\int_0^1 \dot{\psi}_\mu^{(z)} \dot{\psi}_\lambda^{(z)}$ can only be expected to be zero when the supports have empty intersection. In contrast, for compactly supported, orthonormal *piecewise polynomial* wavelets, say of order d, $\int_0^1 \dot{\psi}_\mu^{(z)} \dot{\psi}_\lambda^{(z)}$ is already zero when the singular support of one wavelet, say $\psi_\lambda^{(z)}$, has empty intersection with the interior of the convex hull of the support of the other. Indeed, in that case $\int_0^1 \dot{\psi}_\mu^{(z)} \dot{\psi}_\lambda^{(z)} = -\int_0^1 \psi_\mu^{(z)} \ddot{\psi}_\lambda^{(z)} = 0$, since on the convex hull of supp $\psi_\mu^{(z)}$, $\ddot{\psi}_\lambda^{(z)} \in P_{d-3}$, and $\psi_\mu^{(z)}$ has $d \geq d-2$ *vanishing moments*, i.e., $\psi_\mu^{(z)} \perp P_{d-1}$ (when supp $\psi_\mu^{(z)}$ has non-empty intersection with the Dirichlet boundary z, the latter, and thus $\int \dot{\psi}_\mu^{(z)} \dot{\psi}_\lambda^{(z)} = 0$, is not necessarily valid).

Compactly supported, orthonormal, piecewise polynomial (multi-) wavelets were constructed in [11, 12, 19]. In [11], it was shown that if $\cdots \subset V_{-1} \subset V_0 \subset V_1 \subset \cdots$ is a multiresolution analysis with finitely many generators (scaling functions), then for some $m, q \in \mathbb{N}_0$, there exists a multiresolution analysis $\cdots \subset \check{V}_{-1} \subset \check{V}_0 \subset \check{V}_1 \subset \cdots$ with $V_q \subset \check{V}_0 \subset V_{q+m}$ that has finitely many *orthogonal* generators. From those, finitely many orthogonal generators (wavelets) for the orthogonal complements $\check{V}_\ell \ominus^{\perp L_2(\mathbb{R})} \check{V}_{\ell-1}$ can be constructed. In [12], explicit constructions were given for V_ℓ being the spline space

$$\left\{ u \in L_2(\mathbb{R}) \cap C^r(\mathbb{R}) : u|_{[i2^{-\ell},(i+1)2^{-\ell}]} \in P_{d-1}, i \in \mathbb{Z} \right\}$$

for $r \in \{0,1\}$ and $d \geq r+2$, as well as some examples for $r = 2$.

For $r = 0$ and $d = 2$, i.e., *continuous piecewise linears*, (q,m) can be taken to be $(1,1)$, i.e., $V_1 \subset \check{V}_0 \subset V_2$, with generating orthonormal scaling functions and wavelets illustrated in Fig. 1. For $\ell \in \mathbb{Z}$, the collections $\{\phi_{j,\ell,i} := 2^{\ell/2}\phi_j(2^\ell \cdot -i):$
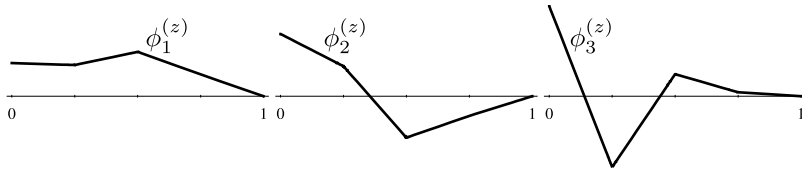
Fig. 2 The scaling functions on the lowest level, for the case $z = \{1\}$

$j \in \{1, 2, 3\}$, $i \in \mathbb{Z}$ and $\{\psi_{j,\ell,i} := 2^{(\ell-1)/2}\psi_j(2^{\ell-1} \cdot -i) : j \in \{1, 2, 3\}, i \in \mathbb{Z}\}$ are orthonormal bases for $\check{V}_\ell$ and $\check{V}_\ell \ominus^{\perp L_2(\mathbb{R})} \check{V}_{\ell-1}$, respectively.

These scaling functions and wavelets on the line can be used to construct such functions on the interval. With $\check{V}_\ell^{(z)} := \check{V}_\ell|_{[0,1]} \cap H^1_{0,z}(0, 1)$, for the choice $z = \{1\}$, i.e., homogeneous Dirichlet boundary conditions at the right boundary, $L_2(0, 1)$-orthonormal bases for $\check{V}_0^{(z)}$ and $\check{V}_\ell^{(z)} \ominus^{\perp L_2(0,1)} \check{V}_{\ell-1}^{(z)}$ are given by $\{\phi_1^{(z)}, \phi_2^{(z)}, \phi_3^{(z)}\}$, illustrated in Fig. 2, and

$$\left\{\psi_{1,\ell,i} : 0 \leq i \leq 2^{\ell-1} - 1\right\} \cup \left\{\psi_{j,\ell,i} : 1 \leq i \leq 2^{\ell-1} - 1, j = 2, 3\right\}$$
$$\cup \left\{\frac{\psi_{2,\ell,0}|_{[0,1]}}{\|\psi_2|_{[0,1]}\|_{L_2(0,1)}}\right\} \cup \left\{\frac{\psi_{3,\ell,2^{\ell-1}}|_{[0,1]}}{\|\psi_3|_{[-1,0]}\|_{L_2(-1,0)}}\right\},$$

respectively.

The collection $\{\phi_1^{(z)}, \phi_2^{(z)}, \phi_3^{(z)}\}$ was constructed as follows: As initial choice for a basis for $\check{V}_0^{(z)}$, we took the orthonormal collection $\{\frac{\phi_1|_{[0,1]}}{\|\phi_1|_{[0,1]}\|_{L_2(0,1)}}, \phi_2, \phi_3\}$. By applying suitable rotations to the pairs of the first and the second and the first and the third basis functions, the resulting modified second and third basis functions were given one vanishing moment. By again applying a suitable rotation to these modified second and third basis functions, the resulting final third basis function got two vanishing moments.

Our experiments, reported in Sect. 6, are carried out with these wavelets. That is, we took $\Psi^{(z)} = \{\psi_\lambda^{(z)} : \lambda \in \Lambda_z\}$ to be the union of the basis for $\check{V}_0^{(z)}$ and that for $\check{V}_\ell^{(z)} \ominus^{\perp L_2(0,1)} \check{V}_{\ell-1}^{(z)}$ for all $\ell \in \mathbb{N}$, where, as usual, we define the level $|\lambda|$ of $\psi_\lambda^{(z)}$ to be 0 or ℓ, respectively. Since these wavelets satisfy appropriate Jackson and Bernstein estimates, $\Psi^{(z)}$ and $\{2^{-|\lambda|}\psi_\lambda^{(z)} : \lambda \in \nabla_z\}$ are Riesz bases for $L_2(0, 1)$ and $H^1_{0,z}(0, 1)$, respectively. Only one scaling function and, on each level $\ell > 0$, the rightmost wavelet of type 3 have no vanishing moments.

For other choices of $z \subseteq \{0, 1\}$, the adaptation of the bases from [12] on the line to the interval can follow similar lines.

3 Near Sparsity of A

For a class of univariate wavelet bases that contains those discussed above, in this section, we investigate how well the stiffness matrix **A** with respect to the resulting tensor product basis can be approximated by sparse matrices. As outlined in the introduction, near sparsity of **A** is an essential ingredient of an adaptive wavelet method.

We consider $\Psi^{(z)} = \{\psi_\lambda^{(z)} : \lambda \in \nabla_z\}$ to be an orthonormal basis for $L_2(0, 1)$, such that $\{2^{-|\lambda|}\psi_\lambda^{(z)} : \lambda \in \nabla_z\}$ is a Riesz basis for $H_{0,z}^1(0, 1)$. Apart from that, we assume that

   (i) $\operatorname{diam}(\operatorname{supp} \psi_\lambda^{(z)}) \eqsim 2^{-|\lambda|}$
  (ii) For some $d \in \mathbb{N}$, each wavelet $\psi_\lambda^{(z)}$ is piecewise polynomial of order d, that
 (iii) Has a singular support that consists of a bounded number of points, uniformly in λ
 (iv) Each point in $[0, 1]$ is contained in the interior of the convex hull of the support of at most a bounded number of $\psi_\lambda^{(z)}$ with $|\lambda| = \ell$, uniformly in $\ell \in \mathbb{N}_0$
  (v) $\|\psi_\lambda^{(z)}\|_{L_\infty} \lesssim 2^{\frac{1}{2}|\lambda|}$, $\|\dot{\psi}_\lambda^{(z)}\|_{L_\infty} \lesssim 2^{\frac{3}{2}|\lambda|}$
 (vi) Wavelets on levels larger than 0 whose supports have empty intersection with z have $d - 2$ vanishing moments

Remark 3.1 By the above assumptions we have $\#\{\lambda \in \nabla_z : |\lambda| = \ell\} \lesssim 2^\ell$. Actually, for Ψ being a basis for $L_2(0, 1)$, it is also needed that $\#\{\lambda \in \nabla_z : |\lambda| = \ell\} \gtrsim 2^\ell$.

Remark 3.2 Thinking of a wavelet basis of order d, i.e., a basis that locally reproduces any polynomial of order d that vanishes at the Dirichlet boundary, the orthogonality of the basis implies assumption (vi) with $d - 2$ reading as d.

Because of the form of the bilinear form (1.1), and the use of tensor product wavelets, the stiffness matrix **A** will be a sum of tensor product operators involving factors that are stiffness matrices of one-dimensional Laplacians in wavelet coordinates. The following result concerns near-sparsity of these matrices:

Proposition 3.3 *Let* $\vec{A}_z = [\frac{\int_0^1 \dot{\psi}_\mu^{(z)} \dot{\psi}_\lambda^{(z)}}{\|\dot{\psi}_\mu^{(z)}\|\|\dot{\psi}_\lambda^{(z)}\|}]_{\lambda,\mu \in \nabla_z}$, *and for* $j \in \mathbb{N}$, *let* $\vec{A}_z^{(j)}$ *be constructed from* $\vec{A}_z$ *by dropping all entries* $(\vec{A}_z)_{\lambda,\mu}$ *with* $||\lambda| - |\mu|| \geq j$. *Then the number of nonzero entries in each row or column of* $\vec{A}_z^{(j)}$ *is* $\mathcal{O}(j)$, *and*

$$\left\| \vec{A}_z - \vec{A}_z^{(j)} \right\| \leq C_{z,\mathrm{sp}} 2^{-j/2},$$

for some constant $C_{z,\mathrm{sp}}$.

Proof If $\operatorname{supp} \psi_\mu^{(z)} \subset (0, 1)$, and the interior of the convex hull of $\operatorname{supp} \psi_\mu^{(z)}$ has empty intersection with $\operatorname{sing\,supp} \psi_\lambda^{(z)}$, then $\int_0^1 \dot{\psi}_\mu^{(z)} \dot{\psi}_\lambda^{(z)} = -\int_0^1 \psi_\mu^{(z)} \ddot{\psi}_\lambda^{(z)} = 0$ by (ii) and (vi). By (iii) and (iv), we conclude that the number of non-zero entries $(\vec{A}_z)_{\lambda,\mu}$ for $|\mu| = \ell$ is bounded uniformly in λ and ℓ. By symmetry, the first statement follows.
From

$$\left| \int_0^1 \dot{\psi}_\mu^{(z)} \dot{\psi}_\lambda^{(z)} \right| \leq \operatorname{diam}\left(\operatorname{supp} \psi_\mu^{(z)} \cap \operatorname{supp} \psi_\lambda^{(z)}\right) \|\dot{\psi}_\mu^{(z)}\|_{L_\infty} \|\dot{\psi}_\lambda^{(z)}\|_{L_\infty}$$

$$\lesssim 2^{-\max(|\mu|,|\lambda|)} 2^{\frac{3}{2}|\mu|} 2^{\frac{3}{2}|\lambda|} \eqsim 2^{-\frac{1}{2}||\lambda|-|\mu||} \|\dot{\psi}_\mu^{(z)}\| \|\dot{\psi}_\lambda^{(z)}\|$$

by (v) and (vi), together with the previous observation, the second statement follows by an application of the Schur lemma. □

Remark 3.4 If $z = \{0, 1\}$ and the wavelets satisfy the additional smoothness condition $\|\ddot{\psi}_\lambda^{(z)}\|_{L_\infty} \lesssim 2^{|\lambda|\frac{5}{2}}$, then from $|\int_0^1 \dot{\psi}_\mu^{(z)} \dot{\psi}_\lambda^{(z)}| = |\int_0^1 \psi_\mu^{(z)} \ddot{\psi}_\lambda^{(z)}| \le \mathrm{diam}(\mathrm{supp}\,\psi_\mu^{(z)}) \times \|\psi_\mu^{(z)}\|_{L_\infty} \|\ddot{\psi}_\lambda^{(z)}\|_{L_\infty} \lesssim 2^{\frac{3}{2}(|\lambda|-|\mu|)} 2^{|\lambda|+|\mu|}$, one infers that even $\|\vec{A} - \vec{A}^{(j)}\| \lesssim 2^{-3j/2}$. Even $\|\vec{A} - \vec{A}^{(j)}\| \lesssim 2^{-5j/2}$ could be obtained when additionally $\|\dddot{\psi}_\lambda^{(z)}\|_{L_\infty} \lesssim 2^{|\lambda|\frac{7}{2}}$ and all wavelets on positive levels have at least one vanishing moment. For orthogonal wavelets, however, the last requirement is in conflict with $z = \{0, 1\}$ that was used for the integration by parts.

Using the above univariate wavelets, let $\boldsymbol{\Psi} := \{\boldsymbol{\psi}_\lambda : \lambda \in \nabla\}$ be defined as in (1.3).

Theorem 3.5 *For $j \in \mathbb{N}$, let $\mathbf{A}^{(j)}$ be the matrix created from $\mathbf{A} = a(\boldsymbol{\psi}_\mu, \boldsymbol{\psi}_\lambda)_{\lambda,\mu \in \nabla}$ with $a(\cdot, \cdot)$ from (1.1) by dropping all entries $\mathbf{A}_{\lambda,\mu}$ for which $\||\lambda| - |\mu|\|_\infty \ge j$. Then the number of non-zero entries in each row or column of $\mathbf{A}^{(j)}$ is bounded by some absolute multiple of jn, and*

$$\|\mathbf{A} - \mathbf{A}^{(j)}\| \le \max_z \|\vec{A}_z - \vec{A}_z^{(j)}\| \quad \left(\le \max_z C_{z,\mathrm{sp}} 2^{-j/2}\right).$$

Proof In view of Proposition 3.3, we only have to prove the first inequality. Let $\vec{B}_z := [\int_0^1 \dot{\psi}_\mu^{(z)} \dot{\psi}_\lambda^{(z)}]_{\lambda,\mu \in \nabla_z}$, $\vec{D}_z := \mathrm{diag}(\vec{B}_z)$, and let $\vec{B}_z^{(j)}$ be constructed from $\vec{B}_z$ by dropping all entries $(\vec{B}_z)_{\lambda,\mu}$ with $\||\lambda| - |\mu|\| \ge j$. From $\vec{A}_z := \vec{D}_z^{-\frac{1}{2}} \vec{B}_z \vec{D}_z^{-\frac{1}{2}}$, $\vec{A}_z^{(j)} := \vec{D}_z^{-\frac{1}{2}} \vec{B}_z^{(j)} \vec{D}_z^{-\frac{1}{2}}$, and since $\vec{A}_z - \vec{A}_z^{(j)}$ is symmetric, we have

$$-\|\vec{A}_z - \vec{A}_z^{(j)}\| \vec{D}_z \le \vec{B}_z - \vec{B}_z^{(j)} \le \|\vec{A}_z - \vec{A}_z^{(j)}\| \vec{D}_z. \tag{3.1}$$

(Since these inequalities concern unbounded matrices, they should be interpreted as inequalities between scalar products involving any finitely supported vector.)

Let

$$\mathbf{B} := \left[a\left(\bigotimes_{m=1}^n \psi_{\boldsymbol{\mu}_m}^{(z_m)}, \bigotimes_{m=1}^n \psi_{\boldsymbol{\lambda}_m}^{(z_m)}\right)\right]_{\lambda,\mu \in \nabla}$$

$$= c_0 + c_1 \vec{B}_{z_1} \otimes I \otimes \cdots \otimes I + \cdots + c_n I \otimes \cdots \otimes I \otimes \vec{B}_{z_n},$$

$$\mathbf{D} := c_0 + c_1 \vec{D}_{z_1} \otimes I \otimes \cdots \otimes I + \cdots + c_n I \otimes \cdots \otimes I \otimes \vec{D}_{z_n},$$

$$\mathbf{B}^{(j)} := c_0 + c_1 \vec{B}_{z_1}^{(j)} \otimes I \otimes \cdots \otimes I + \cdots + c_n I \otimes \cdots \otimes I \otimes \vec{B}_{z_n}^{(j)}.$$

Then $\mathbf{A} = \mathbf{D}^{-\frac{1}{2}} \mathbf{B} \mathbf{D}^{-\frac{1}{2}}$ and $\mathbf{A}^{(j)} = \mathbf{D}^{-\frac{1}{2}} \mathbf{B}^{(j)} \mathbf{D}^{-\frac{1}{2}}$. By applying (3.1) n times, we infer that

$$-\max_z \|\vec{A}_z - \vec{A}_z^{(j)}\| (\mathbf{D} - c_0) \le \mathbf{B}, \qquad -\mathbf{B}^{(j)} \le \max_z \|\vec{A}_z - \vec{A}_z^{(j)}\| (\mathbf{D} - c_0),$$

and so by $\mathbf{D} - c_0 \le \mathbf{D}$, that $\|\mathbf{A} - \mathbf{A}^{(j)}\| \le \max_z \|\vec{A}_z - \vec{A}_z^{(j)}\|$. □

Remark 3.6 For $n > 1$, the matrix $\mathbf{A}^{(j)}$ still has many entries that can be dropped without significantly increasing the error. Indeed, $\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\mu}}$ and so $\mathbf{A}^{(j)}_{\boldsymbol{\lambda},\boldsymbol{\mu}}$ can only be non-zero when $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ differ in at most one coordinate, say the ith one. Then $\mathbf{A}^{(j)}_{\boldsymbol{\lambda},\boldsymbol{\mu}} = \frac{\|\dot{\psi}_{\lambda_i}^{(z_i)}\|}{\|\boldsymbol{\psi}_{\boldsymbol{\lambda}}\|}\vec{A}^{(j)}_{\lambda_i,\mu_i}\frac{\|\dot{\psi}_{\mu_i}^{(z_i)}\|}{\|\boldsymbol{\psi}_{\boldsymbol{\mu}}\|}$, where the pre- and post factor are of the order $\frac{2^{|\lambda_i|}}{\sqrt{\sum_{m=1}^{n}4^{|\lambda_m|}}}$ and $\frac{2^{|\mu_i|}}{\sqrt{\sum_{m=1}^{n}4^{|\mu_m|}}}$, respectively, which can be very small. Using the Schur lemma, one can show that in any case elements with modulus less than some multiple of $\frac{2^{-j/2}}{(j+1)n}$ can be dropped from $\mathbf{A}^{(j)}$ while keeping the error on the same level.

We recall the following definition from [15]:

A matrix $\mathbf{E}$ *is said to be* s^*-computable *when for some constant* C *and any* $j \in \mathbb{N}_0$*, there exists a matrix* $\mathbf{E}^{(j)}$ *having in each column not more than* Cj *non-zero entries, whose joint computation takes not more than* Cj *operations, such that for any* $\bar{s} < s^*$*, there exists a constant* $D_{\bar{s}}$ *such that* $\|\mathbf{E} - \mathbf{E}^{(j)}\| \le D_{\bar{s}}j^{-\bar{s}}$.

Theorem 3.5 shows that $\mathbf{A}$ is s^*-computable for $s^* = \infty$ with $C \approx n$ and, for any $\bar{s}$, with $D_{\bar{s}}$ being an absolute constant (the spectral argument applied in the proof of Theorem 3.5 prevents that also $D_{\bar{s}}$ grows linearly with n, which would be the case if simply a repeated triangle inequality was applied). In the next sections, it will not be essential that $s^* = \infty$, but it suffices that it is larger than $d - 1$, being the maximum rate of best N-term approximation that can be expected.

Remark 3.7 We discuss some *generalizations* of Theorem 3.5. Instead of (1.1), let

$$a(u, v) = \int_{\Omega} c_0 u v + \sum_{i,j=1}^{n} d_{ij}\partial_j u \partial_i v,$$

where the d_{ij} are constants such that for some constants $Q \ge q > 0$,

$$q \le \sum_{i,j=1}^{n} d_{ij}\xi_i\xi_j \Big/ \sum_{m=1}^{n} c_m\xi_m^2 \le Q, \quad (\xi \in \mathbb{R}^n).$$

From

$$\frac{|\int_0^1 \psi_\mu^{(z)}\dot{\psi}_\lambda^{(z)}|}{\sqrt{\|\dot{\psi}_\mu^{(z)}\|\|\dot{\psi}_\lambda^{(z)}\|}} \lesssim 2^{||\lambda|-|\mu||},$$

and, with $\triangle_k^{(i,j)} := \vec{D}_{z_k}^{\frac{1}{2}}$ for $k \in \{i, j\}$ and $\triangle_k^{(i,j)} := I$ otherwise, from $\sum_{i,j=1}^{n} d_{ij} \otimes_{k=1}^{n} \triangle_k^{(i,j)} \le Q(\mathbf{D} - c_0)$, one may verify that with $m(n) := \#\{d_{ij} \ne 0\}$, there exist $\mathbf{A}^{(j)}$ having $\mathcal{O}(jm(n))$ non-zeros in each column with $\|\mathbf{A} - \mathbf{A}_j\| \le 2^{-j/2}$.

By additionally using results from [27], for variable, sufficiently smooth coefficients d_{ij} that each depend on at most a uniformly bounded number of space variables x_k, one can show that $\mathbf{A}$ is s^*-computable for some $s^* > d - 1$ with $C \approx m(n)$

and, for any $\bar{s} < s^*$, with $D_{\bar{s}}$ being an absolute constant. Concluding, we can say that in these generalized settings the results of this paper are still valid with the factor n showing up in the operation count of the adaptive wavelet algorithm now reading as $m(n) \in [n, n^2]$.

4 The Adaptive Wavelet Scheme

4.1 Underlying Principles

We recall the main ideas behind the adaptive wavelet method introduced in [16]. It is a modification of the scheme by Cohen, Dahmen, and DeVore from [4], in which the recurrent coarsening of the iterands is avoided. This modification results in a quantitatively better performing method, which was confirmed in numerical experiments [6, 16].

Throughout this Sect. 4, $\mathbf{A} : \ell_2(\boldsymbol{\Lambda}) \to \ell_2(\boldsymbol{\Lambda})$ is some general bounded, symmetric and positive definite operator that is s^*-computable, where $\boldsymbol{\Lambda}$ is some countable index set. Not until the next section do we return to our particular matrix $\mathbf{A}$ and investigate the issue of the dependence of constants on the space dimension.

For any $\boldsymbol{\Lambda} \subset \boldsymbol{\nabla}$, with $\ell_2(\boldsymbol{\Lambda})$ we will mean the subspace of $\mathbf{v} \in \ell_2(\boldsymbol{\nabla})$ with supports in $\boldsymbol{\Lambda}$. The trivial embedding of $\ell_2(\boldsymbol{\Lambda})$ into $\ell_2(\boldsymbol{\nabla})$ will be denoted by $\mathbf{I}_{\boldsymbol{\Lambda}}$, and its adjoint with respect to $\langle \cdot, \cdot \rangle$, i.e., the operator that replaces coefficients outside $\boldsymbol{\Lambda}$ by zeros, will be denoted by $\mathbf{P}_{\boldsymbol{\Lambda}}$. We set $\mathbf{A}_{\boldsymbol{\Lambda}} = \mathbf{P}_{\boldsymbol{\Lambda}} \mathbf{A} \mathbf{I}_{\boldsymbol{\Lambda}}$. Using that $\mathbf{A}$ is symmetric and positive definite, one verifies that for $\mathbf{v} \in \ell_2(\boldsymbol{\nabla})$ and $\mathbf{v}_{\boldsymbol{\Lambda}} \in \ell_2(\boldsymbol{\Lambda})$,

$$\left\| \mathbf{A}^{-1} \right\|^{-\frac{1}{2}} \|\mathbf{v}\| \leq \||\mathbf{v}\|| \leq \|\mathbf{A}\|^{\frac{1}{2}} \|\mathbf{v}\|, \qquad \||\mathbf{A}\mathbf{v}\|| \leq \|\mathbf{A}\|^{\frac{1}{2}} \||\mathbf{v}\||,$$

$$\left\| \mathbf{A}^{-1} \right\|^{-\frac{1}{2}} \||\mathbf{v}_{\boldsymbol{\Lambda}}\|| \leq \||\mathbf{A}_{\boldsymbol{\Lambda}} \mathbf{v}_{\boldsymbol{\Lambda}}\||,$$

which properties will be used often in the following. The solution $\mathbf{u}_{\boldsymbol{\Lambda}} \in \ell_2(\boldsymbol{\Lambda})$ of $\mathbf{A}_{\boldsymbol{\Lambda}} \mathbf{u}_{\boldsymbol{\Lambda}} = \mathbf{P}_{\boldsymbol{\Lambda}} \mathbf{f}$ is the *Galerkin approximation* to $\mathbf{u}$ from $\ell_2(\boldsymbol{\Lambda})$. With respect to $\||\cdot\||$, it is the best approximation to $\mathbf{u}$ from this subspace.

The methods from [4, 16] generate a sequence of (approximate) Galerkin solutions with respect to a sequence of nested subspaces of $\ell_2(\boldsymbol{\nabla})$. In Part (a) of the following lemma, which is [4, Lemma 4.1], it is shown how a given subspace can be enlarged such that the error in the Galerkin solution is reduced by a constant factor. In Part (b), from [16], and in the discussion following this lemma, it is shown that if one is not too greedy, i.e., if one is content with a moderate error reduction, then besides having convergence, one keeps quasi-optimal control over the dimensions of the subspaces. Although $\mathbf{w}$ in this lemma can be any element of $\ell_2(\boldsymbol{\Lambda}_0)$, one may think of it as being (an approximation to) the Galerkin solution from $\ell_2(\boldsymbol{\Lambda}_0)$. For convenience we recall the short proof.

Lemma 4.1 *Let $\mu \in (0, 1]$, $\boldsymbol{\Lambda}_0 \subset \boldsymbol{\Lambda}_1 \subset \boldsymbol{\nabla}$, $\mathbf{w} \in \ell_2(\boldsymbol{\Lambda}_0)$ such that*

$$\left\| \mathbf{P}_{\boldsymbol{\Lambda}_1} (\mathbf{f} - \mathbf{A}\mathbf{w}) \right\| \geq \mu \|\mathbf{f} - \mathbf{A}\mathbf{w}\|. \tag{4.1}$$

(a) *Then, for* $\mathbf{u}_{\mathbf{\Lambda}_1} \in \ell_2(\mathbf{\Lambda}_1)$ *being the solution of* $\mathbf{A}_{\mathbf{\Lambda}_1}\mathbf{u}_{\mathbf{\Lambda}_1} = \mathbf{P}_{\mathbf{\Lambda}_1}\mathbf{f}$, *we have*

$$\||\mathbf{u} - \mathbf{u}_{\mathbf{\Lambda}_1}|\| \leq \rho \||\mathbf{u} - \mathbf{w}|\|,$$

where $\rho = \rho(\mathbf{A}, \mu) := [1 - \kappa(\mathbf{A})^{-1}\mu^2]^{\frac{1}{2}} < 1$.

(b) *If* $\mu < \kappa(\mathbf{A})^{-\frac{1}{2}}$ *and* $\mathbf{\Lambda}_1$ *is the* smallest *set satisfying* (4.1), *then*

$$\#(\mathbf{\Lambda}_1\backslash\mathbf{\Lambda}_0) \leq \min\{N : \||\mathbf{u} - \mathbf{u}_N|\| \leq \sigma \||\mathbf{u} - \mathbf{w}|\|\}, \tag{4.2}$$

where $\sigma = \sigma(\mathbf{A}, \mu) := [1 - \mu^2\kappa(\mathbf{A})]^{\frac{1}{2}} > 0$.

Proof (a) We have

$$\||\mathbf{u}_{\mathbf{\Lambda}_1} - \mathbf{w}|\| \geq \|\mathbf{A}\|^{-\frac{1}{2}}\|\mathbf{A}(\mathbf{u}_{\mathbf{\Lambda}_1} - \mathbf{w})\| \geq \|\mathbf{A}\|^{-\frac{1}{2}}\|\mathbf{P}_{\mathbf{\Lambda}_1}(\mathbf{f} - \mathbf{A}\mathbf{w})\|$$

$$\geq \|\mathbf{A}\|^{-\frac{1}{2}}\mu\|\mathbf{f} - \mathbf{A}\mathbf{w}\| \geq \kappa(\mathbf{A})^{-\frac{1}{2}}\mu\||\mathbf{u} - \mathbf{w}|\|.$$

The proof of (a) is completed by using the Galerkin orthogonality

$$\||\mathbf{u} - \mathbf{w}|\|^2 = \||\mathbf{u} - \mathbf{u}_{\mathbf{\Lambda}_1}|\|^2 + \||\mathbf{u}_{\mathbf{\Lambda}_1} - \mathbf{w}|\|^2.$$

(b) For an N as in the right-hand side of (4.2), let $\check{\mathbf{\Lambda}} := \mathbf{\Lambda}_0 \bigcup \operatorname{supp} \mathbf{u}_N$. Then, for the solution of $\mathbf{A}_{\check{\mathbf{\Lambda}}}\mathbf{u}_{\check{\mathbf{\Lambda}}} = \mathbf{P}_{\check{\mathbf{\Lambda}}}\mathbf{f}$, we have $\||\mathbf{u} - \mathbf{u}_{\check{\mathbf{\Lambda}}}|\| \leq \||\mathbf{u} - \mathbf{u}_N|\|$, and so by Galerkin orthogonality

$$\||\mathbf{u}_{\check{\mathbf{\Lambda}}} - \mathbf{w}|\| \geq \mu\kappa(\mathbf{A})^{\frac{1}{2}}\||\mathbf{u} - \mathbf{w}|\|,$$

giving

$$\|\mathbf{P}_{\check{\mathbf{\Lambda}}}(\mathbf{f} - \mathbf{A}\mathbf{w})\| = \|\mathbf{A}_{\check{\mathbf{\Lambda}}}(\mathbf{u}_{\check{\mathbf{\Lambda}}} - \mathbf{w})\| \geq \|\mathbf{A}^{-1}\|^{-\frac{1}{2}}\||\mathbf{u}_{\check{\mathbf{\Lambda}}} - \mathbf{w}|\|$$

$$\geq \|\mathbf{A}^{-1}\|^{-\frac{1}{2}}\mu\kappa(\mathbf{A})^{\frac{1}{2}}\||\mathbf{u} - \mathbf{w}|\| \geq \mu\|\mathbf{f} - \mathbf{A}\mathbf{w}\|.$$

By our assumption on $\mathbf{\Lambda}_1$, we conclude that $\#(\mathbf{\Lambda}_1\backslash\mathbf{\Lambda}_0) \leq \#(\check{\mathbf{\Lambda}}\backslash\mathbf{\Lambda}_0) \leq N$. $\qquad\square$

Let us think of $\mathbf{w}$ in Lemma 4.1 as the exact Galerkin solution on $\mathbf{\Lambda}_0$. Then starting with $\mathbf{\Lambda}_0 = \emptyset$, a repeated application of Part (a) gives a sequence $\mathbf{\Lambda}_0 \subset \mathbf{\Lambda}_1 \subset \cdots$, and corresponding Galerkin solutions $\mathbf{u}_{\mathbf{\Lambda}_0}, \mathbf{u}_{\mathbf{\Lambda}_1}, \ldots$ with $\||\mathbf{u} - \mathbf{u}_{\mathbf{\Lambda}_k}|\| \leq \rho\||\mathbf{u} - \mathbf{u}_{\mathbf{\Lambda}_{k-1}}|\|$. This is the idealized version of the adaptive wavelet method we are going to apply. Note that generally the residuals $\mathbf{f} - \mathbf{A}\mathbf{u}_{\mathbf{\Lambda}_k}$ are infinitely supported vectors, so that this algorithm cannot be performed in practice. For the moment ignoring this fact, let us take $\mu < \kappa(\mathbf{A})^{-\frac{1}{2}}$. Then, assuming that $\mathbf{u} \in \mathcal{A}_\infty^s$ for some $s > 0$, Part (b) shows that

$$\#(\mathbf{\Lambda}_k\backslash\mathbf{\Lambda}_{k-1}) \leq \min\{N : \||\mathbf{u} - \mathbf{u}_N|\| \leq \sigma\||\mathbf{u} - \mathbf{u}_{\mathbf{\Lambda}_{k-1}}|\|\}$$

$$\leq \left[\|\mathbf{A}\|^{-\frac{1}{2}}\sigma\||\mathbf{u} - \mathbf{u}_{\mathbf{\Lambda}_{k-1}}|\|\right]^{-1/s}\|\mathbf{u}\|_{\mathcal{A}_\infty^s}^{1/s},$$

by $\||| \cdot \||| \leq \|\mathbf{A}\|^{\frac{1}{2}} \| \cdot \|$ and the definition of $\| \cdot \|_{\mathcal{A}_\infty^s}$. By combining both estimates, for $\ell \in \mathbb{N}$ we have

$$\#\boldsymbol{\Lambda}_\ell = \sum_{k=1}^{\ell} \#(\boldsymbol{\Lambda}_k \backslash \boldsymbol{\Lambda}_{k-1}) \leq \|\mathbf{A}\|^{1/2s} \sigma^{-1/s} \|\mathbf{u}\|_{\mathcal{A}_\infty^s}^{1/s} \sum_{k=1}^{\ell} \||| \mathbf{u} - \mathbf{u}_{\boldsymbol{\Lambda}_{k-1}} \|||^{-1/s}$$

$$\leq \|\mathbf{A}\|^{1/2s} \sigma^{-1/s} \|\mathbf{u}\|_{\mathcal{A}_\infty^s}^{1/s} \||| \mathbf{u} - \mathbf{u}_{\boldsymbol{\Lambda}_\ell} \|||^{-1/s} \sum_{k=1}^{\ell} \left(\rho^{\ell-k+1} \right)^{1/s}$$

$$\leq \|\mathbf{A}\|^{1/s} \frac{(\rho/\sigma)^{1/s}}{1 - \rho^{1/s}} \|\mathbf{f} - \mathbf{A}\mathbf{u}_{\boldsymbol{\Lambda}_\ell} \|^{-1/s} \|\mathbf{u}\|_{\mathcal{A}_\infty^s}^{1/s}. \tag{4.3}$$

In view of $\mathbf{u} \in \mathcal{A}_\infty^s$, generally the support of any approximation to $\mathbf{u}$ with an $\| \cdot \|$-error equal to that of $\mathbf{u}_{\boldsymbol{\Lambda}_\ell}$ cannot be expected to be smaller than $\| \mathbf{u} - \mathbf{u}_{\boldsymbol{\Lambda}_\ell} \|^{-1/s} \|\mathbf{u}\|_{\mathcal{A}_\infty^s}^{1/s} \geq \|\mathbf{A}^{-1}\|^{-1/s} \|\mathbf{f} - \mathbf{A}\mathbf{u}_{\boldsymbol{\Lambda}_\ell}\|^{-1/s} \|\mathbf{u}\|_{\mathcal{A}_\infty^s}^{1/s}$. Comparing this lower bound with the upper bound from (4.3), we conclude that up to the factor $\frac{(\rho/\sigma)^{1/s}}{1-\rho^{1/s}} \kappa(\mathbf{A})^{1/s}$, generally the cardinality of the set $\boldsymbol{\Lambda}_\ell$ produced by the algorithm is the smallest that can be expected.

Remark 4.2 Thinking of $\mu = \beta \kappa(\mathbf{A})^{-\frac{1}{2}}$ for some fixed $\beta < 1$, $\frac{(\rho/\sigma)^{1/s}}{1-\rho^{1/s}} \kappa(\mathbf{A})^{1/s}$ is a bounded function of $(s, \kappa(\mathbf{A}))$ on any compact subset of $(0, \infty) \times [1, \infty)$. For any s, $\frac{(\rho/\sigma)^{1/s}}{1-\rho^{1/s}} \kappa(\mathbf{A})^{1/s}$ tends to infinity when $\kappa(\mathbf{A})$ does, underlining the importance of using $L_2(0, 1)$-orthonormal univariate wavelets in our application of solving elliptic PDEs in high space dimensions.

4.2 Practical Scheme and Rates

As we said, the algorithm from the previous subsection is not feasible in practice because it uses the generally infinitely supported residuals. Apart from stopping the iteration, which we did not discuss so far, these residuals are only used to indicate in which way the current subspace has to be enlarged. For a proper enlargement it is, however, sufficient to know where the largest entries of the residual are located, and for that goal it suffices to use an approximate residual with a sufficiently small relative error.

We will assume the availability of the following routines. Realizations will be discussed in the forthcoming subsections.

APPLY$[\boldsymbol{\Lambda}, \mathbf{w}, \varepsilon] \rightarrow \mathbf{z}$: *For $\varepsilon > 0$, $\boldsymbol{\Lambda} \subset \nabla$, and a finitely supported $\mathbf{w} \in \ell_2(\boldsymbol{\Lambda})$, it yields a finitely supported $\mathbf{z} \in \ell_2(\boldsymbol{\Lambda})$ with $\|\mathbf{A}_{\boldsymbol{\Lambda}} \mathbf{w} - \mathbf{z}\| \leq \varepsilon$.*

RHS$[\boldsymbol{\Lambda}, \varepsilon] \rightarrow \mathbf{g}$: *For $\varepsilon > 0$, $\boldsymbol{\Lambda} \subset \nabla$, it yields a finitely supported $\mathbf{g} \in \ell_2(\boldsymbol{\Lambda})$ with $\|\mathbf{P}_{\boldsymbol{\Lambda}} \mathbf{f} - \mathbf{g}\| \leq \varepsilon$.*

Furthermore, in view of minimizing the computational complexity, we will not solve the Galerkin systems $\mathbf{A}_{\boldsymbol{\Lambda}_k} \mathbf{u}_{\boldsymbol{\Lambda}_k} = \mathbf{P}_{\boldsymbol{\Lambda}_k} \mathbf{f}$ exactly. Instead, we solve them using

an iterative method starting with the previous iterand, and with a tolerance being a sufficiently small multiple of the error in that iterand. Moreover, we will allow that instead of the exact right-hand side $\mathbf{P}_{\Lambda_k}\mathbf{f}$, we only have a sufficiently accurate approximation to it available. We assume the availability of the following routine, whose realization will be discussed in the next subsection.

GALSOLVE$[\Lambda, \mathbf{g}_\Lambda, \mathbf{w}_\Lambda^{(0)}, \delta, \varepsilon] \to \mathbf{w}_\Lambda$: *For $\delta \geq \varepsilon > 0$, $\Lambda \subset \nabla$, $\mathbf{g}_\Lambda, \mathbf{w}_\Lambda^{(0)} \in \ell_2(\Lambda)$ with $\|\mathbf{g}_\Lambda - \mathbf{A}_\Lambda \mathbf{w}_\Lambda^{(0)}\| \leq \delta$, it yields a $\mathbf{w}_\Lambda \in \ell_2(\Lambda)$ with $\|\mathbf{g}_\Lambda - \mathbf{A}_\Lambda \mathbf{w}_\Lambda\| \leq \varepsilon$.*

Now suppose that we have available an approximation $\mathbf{w}$ to $\mathbf{u}$, with support contained in some finite $\Lambda_0 \subset \nabla$, as well as an approximation $\mathbf{r}$ to $\mathbf{f} - \mathbf{A}\mathbf{w}$ with $\|(\mathbf{f} - \mathbf{A}\mathbf{w}) - \mathbf{r}\| \leq \omega\|\mathbf{r}\|$ for some constant $\omega > 0$. Then our task will be to enlarge Λ_0 to some $\Lambda_1 \subset \nabla$ such that for some constant $\alpha > \omega$,

$$\|\mathbf{P}_{\Lambda_1}\mathbf{r}\| \geq \alpha\|\mathbf{r}\|. \tag{4.4}$$

Indeed, then $\|\mathbf{P}_{\Lambda_1}(\mathbf{f} - \mathbf{A}\mathbf{w})\| \geq \|\mathbf{P}_{\Lambda_1}\mathbf{r}\| - \omega\|\mathbf{r}\| \geq (\alpha - \omega)\|\mathbf{r}\| \geq \frac{\alpha-\omega}{1+\omega}\|\mathbf{f} - \mathbf{A}\mathbf{w}\|$, and so an application of Lemma 4.1(a) shows that the error in the Galerkin solution from $\ell_2(\Lambda_1)$ is strictly less than that in $\mathbf{w}$.

Moreover, with $\mu := \frac{\alpha+\omega}{1-\omega}$, and $\nabla \supset \hat{\Lambda} \supset \Lambda_0$ being *any* set with $\|\mathbf{P}_{\hat{\Lambda}}(\mathbf{f} - \mathbf{A}\mathbf{w})\| \geq \mu\|\mathbf{f} - \mathbf{A}\mathbf{w}\|$, we have $\mu\|\mathbf{r}\| \leq \mu\|\mathbf{f} - \mathbf{A}\mathbf{w}\| + \mu\omega\|\mathbf{r}\| \leq \|\mathbf{P}_{\hat{\Lambda}}(\mathbf{f} - \mathbf{A}\mathbf{w})\| + \mu\omega\|\mathbf{r}\| \leq \|\mathbf{P}_{\hat{\Lambda}}\mathbf{r}\| + (1+\mu)\omega\|\mathbf{r}\|$, or $\|\mathbf{P}_{\hat{\Lambda}}\mathbf{r}\| \geq \alpha\|\mathbf{r}\|$. So if we would select $\Lambda_1 \supset \Lambda_0$ satisfying (4.4) with *minimal* cardinality, then $\#(\Lambda_1 \backslash \Lambda_0) \leq \#(\hat{\Lambda} \backslash \Lambda_0)$, and if additionally $\mu < \kappa(\mathbf{A})^{-\frac{1}{2}}$ then using Lemma 4.1(b) we would have quasi optimal control over the cardinality of this set.

Remark 4.3 The cost of our adaptive wavelet algorithm will be dominated by that of the computation of the approximate residuals. Assuming $\mathbf{u} \in \mathcal{A}_\infty^s$, (the upper bound for) the latter will scale as the prescribed tolerance to the power $-1/s$. So in particular for relatively small values of s, it will be beneficial if we could reduce the prescribed tolerances by some factor. Above, we estimated $\|\mathbf{P}_{\hat{\Lambda}}(\mathbf{f} - \mathbf{A}\mathbf{w})\| \leq \|\mathbf{P}_{\hat{\Lambda}}\mathbf{r}\| + \omega\|\mathbf{r}\|$ for $\hat{\Lambda} \supset \Lambda_0$ being the smallest set with $\|\mathbf{P}_{\hat{\Lambda}}(\mathbf{f} - \mathbf{A}\mathbf{w})\| \geq \mu\|\mathbf{f} - \mathbf{A}\mathbf{w}\|$, and we applied similar steps with interchanged roles of $\mathbf{r}$ and $\mathbf{f} - \mathbf{A}\mathbf{w}$. The first estimate can only be sharp when the error $(\mathbf{f} - \mathbf{A}\mathbf{w}) - \mathbf{r}$ is fully supported in $\hat{\Lambda}$, which is unlikely. Making the assumption that the error is more equally distributed in the sense that $\|\mathbf{P}_{\hat{\Lambda}}(\mathbf{f} - \mathbf{A}\mathbf{w})\| \leq \|\mathbf{P}_{\hat{\Lambda}}\mathbf{r}\| + \omega\mu\|\mathbf{r}\|$, (and similarly $\|\mathbf{P}_{\Lambda_1}\mathbf{r}\| \leq \|\mathbf{r}\| + \alpha\omega\|\mathbf{r}\|$), instead of $\omega < \alpha$ and $\mu := \frac{\alpha+\omega}{1-\omega} < \kappa(\mathbf{A})^{-\frac{1}{2}}$, we end up with the relaxed conditions $\omega < \frac{1}{2}$, and $\mu := \frac{\alpha}{1-2\omega} < \kappa(\mathbf{A})^{-\frac{1}{2}}$.

Also the upper bound $\kappa(\mathbf{A})^{-\frac{1}{2}}$ for μ derived in Lemma 4.1(b) is the result of a succession of two worst case analyses that unlikely or even impossibly could occur at the same time. So although we have no rigorous proof for this, we expect that in practice the conditions on the parameters α, ω, μ can be violated to some unfortunately unknown extent, by that improving the quantitative behavior of the algorithm. The same comment applies to the forthcoming parameter γ for the calls of **GALSOLVE**.

Finding the *smallest* $\mathbf{\Lambda}_1 \supset \mathbf{\Lambda}_0$ for which (4.4), or equivalently,

$$\left\| \mathbf{r}|_{\nabla \setminus \mathbf{\Lambda}_0} - \mathbf{r}|_{\mathbf{\Lambda}_1 \setminus \mathbf{\Lambda}_0} \right\| \leq \sqrt{1 - \alpha^2}\, \|\mathbf{r}\|$$

is valid requires the sorting of all coefficients of $\mathbf{r}_{\nabla \setminus \mathbf{\Lambda}_0}$ by their modulus, which, with $L := \# \operatorname{supp} \mathbf{r}|_{\nabla \setminus \mathbf{\Lambda}_0}$, already takes the order of $L \log L$ operations. In order to avoid the log-factor, we realize that it is not essential that $\#(\mathbf{\Lambda}_1 \setminus \mathbf{\Lambda}_0)$ be truly minimal. It is sufficient that it is minimal up to some constant factor that for convenience will be fixed to 2. It requires an approximate sorting of the coefficients of $\mathbf{r}_{\nabla \setminus \mathbf{\Lambda}_0}$ by their modulus, which will be performed by a call of the following routine. This routine will also be used for the implementation of the **APPLY** routine, in which setting the idea of performing an approximate ordering was introduced in [1, 24].

BUCKETSORT$[\mathbf{v}, \varepsilon] \to [(\mathbf{v}_{[p]})_{1 \leq p \leq P}, \Pi]$: *The input* $\mathbf{v} \in \ell_2(\mathbf{\Lambda})$ *should be finitely supported, and* $\varepsilon > 0$. *Let* P *be the smallest integer with* $2^{-P/2} \|\mathbf{v}\|_\infty \sqrt{\# \operatorname{supp} \mathbf{v}} \leq \varepsilon$. *Store the indices of* $\mathbf{v}$ *in* P *buckets, depending on the modulus of the corresponding coefficient to be in one of the* P *intervals* $(\frac{1}{\sqrt{2}} \|\mathbf{v}\|_\infty, \|\mathbf{v}\|_\infty]$, $(\frac{1}{2} \|\mathbf{v}\|_\infty, \frac{1}{\sqrt{2}} \|\mathbf{v}\|_\infty], \dots$, *or* $(2^{-P/2} \|\mathbf{v}\|_\infty, 2^{-(P-1)/2} \|\mathbf{v}\|_\infty]$, *and discard them otherwise. For* $1 \leq p \leq P$, *define* $\mathbf{v}_{[p]}$ *as the restriction of* $\mathbf{v}$ *to the indices from the pth bucket. Define* $\Pi : \mathbb{N} \to \nabla$ *by numbering the indices in the buckets from the first to the last, and within each bucket in arbitrary order. The definition of* Π *beyond the last numbered index is irrelevant.*

Note that the number of buckets P is $\max(0, \lceil 2 \log_2(\|\mathbf{v}\|_\infty \sqrt{\# \operatorname{supp} \mathbf{v}}/\varepsilon) \rceil)$. This P is taken such that the squared sum of the coefficients corresponding to the indices not stored in any bucket is not larger than ε^2. This means that for the task of finding a (quasi) minimal Λ such that $\|\mathbf{v} - \mathbf{v}|_\Lambda\| \leq \varepsilon$ these coefficients can be discarded anyway. By furthermore using that the squared coefficients corresponding to indices within one bucket differ at most by a factor 2, one arrives at the following result (cf. [16, Remark 2.3]):

Proposition 4.4 *The number of operations required by a call of the routine* $[\cdot, \Pi] :=$ **BUCKETSORT**$[\mathbf{v}, \varepsilon]$ *is bounded by some absolute multiple of* $\# \operatorname{supp} \mathbf{v} + \max(1, \lceil \log(\|\mathbf{v}\|_\infty \sqrt{\# \operatorname{supp} \mathbf{v}}/\varepsilon) \rceil)$. *For any* $\delta \geq \varepsilon$, *the smallest* $K \in \mathbb{N}_0$ *with* $\|\mathbf{v} - \mathbf{v}|_{\Pi(\{1, \dots, K\})}\| \leq \delta$ *satisfies* $K \leq 2 \min\{N \in \mathbb{N}_0 : \|\mathbf{v} - \mathbf{v}_N\| \leq \delta\}$.

The above considerations lead to the following practical variant of the scheme from the previous subsection:

SOLVE$[\nu_{-1}, \varepsilon] \to \mathbf{w}_k$:
% *Let* $\alpha, \omega, \gamma, \theta$ *be constants with* $\omega \in (0, \alpha)$, $\frac{\alpha + \omega}{1 - \omega} < \kappa(\mathbf{A})^{-\frac{1}{2}}$, $\theta > 0$ *and*
% $\gamma \in (0, \frac{1}{6} \kappa(\mathbf{A})^{-\frac{1}{2}} \frac{\alpha - \omega}{1 + \omega})$. *The parameter* ν_{-1} *is an estimate for the norm of the*
% *initial residual* $\mathbf{f}$.
$k := 0$, $\mathbf{w}_k := 0$, $\mathbf{\Lambda}_k := \emptyset$
do $\zeta := \theta \nu_{k-1}$
    do $\zeta := \zeta/2$, $\mathbf{r}_k := $ **RHS**$[\nabla, \zeta/2] - $ **APPLY**$[\nabla, \mathbf{w}_k, \zeta/2]$

```
        if νk := ‖rk‖ + ζ ≤ ε then stop endif
    until ζ ≤ ω‖rk‖
```
$[\cdot, \Pi] := \textbf{BUCKETSORT}[\mathbf{r}_k|_{\nabla \setminus \Lambda_k}, \sqrt{1 - \alpha^2}\, \|\mathbf{r}_k\|]$
determine the smallest $K \in \mathbb{N}_0$ *with* $\|\mathbf{r}_k|_{\nabla \setminus \Lambda_k} - \mathbf{r}_k|_{\Pi(\{1,\ldots,K\})}\| \leq \sqrt{1 - \alpha^2}\, \|\mathbf{r}_k\|$
$\mathbf{\Lambda}_{k+1} := \mathbf{\Lambda}_k \cup \Pi(\{1, \ldots, K\})$
$\mathbf{g}_{k+1} := \textbf{RHS}[\mathbf{\Lambda}_{k+1}, \gamma\, \nu_k]$
$\mathbf{w}_{k+1} := \textbf{GALSOLVE}[\mathbf{\Lambda}_{k+1}, \mathbf{g}_{k+1}, \mathbf{w}_k, (1 + \gamma)\nu_k, \gamma\, \nu_k]$
$k := k + 1$
```
enddo
```

In the inner loop of this adaptive wavelet algorithm **SOLVE**, an approximate residual $\mathbf{r}_k$ of the current iterand $\mathbf{w}_k$ is determined with $\|\mathbf{r}_k^* - \mathbf{r}_k\| \leq \omega\|\mathbf{r}_k\|$, with $\mathbf{r}_k^*$ denoting the exact residual. This is done by computing approximate residuals with geometrically decreasing tolerances ζ until either this condition is met, or $\|\mathbf{r}_k^*\| \leq \|\mathbf{r}_k\| + \|\mathbf{r}_k^* - \mathbf{r}_k\| \leq \|\mathbf{r}_k\| + \zeta =: \nu_k \leq \varepsilon$, in which case the algorithm stops.

Because of the geometrical decrease of the tolerances, the cost of this inner loop is always dominated by the cost of its last iteration. From a quantitative point of view, however, it is useful to tune the constant θ to be as large as possible such that the inner loop "usually" terminates in the first iteration. When the $\|\mathbf{r}_k\|$ exhibit a linear decrease as a function of k, which is generally the best that can be expected, a selection of θ to be (slightly smaller than) $\frac{2\omega}{1+\omega}$ times the rate seems to be a sensible choice. Indeed, if $\nu_{k-1} = (1+\omega)\|\mathbf{r}_{k-1}\|$, then $\frac{1}{2}\frac{2\omega}{1+\omega}\frac{\|\mathbf{r}_k\|}{\|\mathbf{r}_{k-1}\|}\nu_{k-1} = \omega\|\mathbf{r}_k\|$.

By adding perturbation arguments to the analysis from the previous subsection, we end up with the following result. For details we refer to [16]. Inspection of the proof shows that a constant that is unspecified there is actually of the form $D_4\|\mathbf{A}\|^{1/s}$, with D_4 as given in the current theorem.

Theorem 4.5 *For any* ε, $\nu_{-1} > 0$, $\mathbf{w} := \textbf{SOLVE}[\nu_{-1}, \varepsilon]$ *terminates with*

$$\|\mathbf{f} - \mathbf{Aw}\| \leq \varepsilon.$$

If, for some $s > 0$, $\mathbf{u} \in \mathcal{A}_\infty^s$, *then*

$$\#\operatorname{supp}\mathbf{w} \leq D_4\|\mathbf{A}\|^{1/s}\varepsilon^{-1/s}\|\mathbf{u}\|_{\mathcal{A}_\infty^s}^{1/s}$$

for some constant $D_4 > 0$, *only dependent on* $\kappa(\mathbf{A})$ *when it tends to* ∞, s *when it tends to* 0 *or* ∞, *and* ω, α *and* γ *when they approach the boundaries of their domains of definition.*

4.3 Cost of **SOLVE**

Using that $\max_{\lambda \in \nabla \setminus \Lambda_k} |(\mathbf{r}_k)_\lambda| \leq \|\mathbf{r}_k\|$, Proposition 4.4 shows that the cost of a call $[\cdot, \Pi] := \textbf{BUCKETSORT}[\mathbf{r}_k|_{\nabla \setminus \Lambda_k}, \sqrt{1 - \alpha^2}\, \|\mathbf{r}_k\|]$ in **SOLVE** can be bounded by some absolute multiple of $\#\operatorname{supp}\mathbf{r}_k$. The same holds for the cost of evaluating the next line in the algorithm. Since the preceding computation of $\mathbf{r}_k$ requires in any case the order of $\#\operatorname{supp}\mathbf{r}_k$ operations, the cost of the call of **SOLVE** will be determined by that of the calls of **RHS**, **APPLY** and **GALSOLVE**.

The realization and cost of the routines **APPLY** and **GALSOLVE** will be discussed in this subsection. A realization of a routine **RHS** will be specific for the problem at hand, and will be discussed in Sect. 5.

The proof of Theorem 4.5 from [16] shows that whenever $\mathbf{u} \in \mathcal{A}_\infty^s$ for some $s > 0$, then any $\mathbf{w}_k$ computed inside **SOLVE** satisfies

$$\# \operatorname{supp} \mathbf{w}_k \leq D_5 \|\mathbf{u} - \mathbf{w}_k\|^{-1/s} \|\mathbf{u}\|_{\mathcal{A}_\infty^s}^{1/s}$$

for some constant $D_5 > 0$, only dependent on $\kappa(\mathbf{A})$, s, ω, α and γ in the way indicated in Theorem 4.5. Putting $d := D_5^{-1} + 1$, and, for tolerances $\varepsilon \geq d \|\mathbf{u} - \mathbf{w}_k\|$, by approximating $\mathbf{w}_k$ by $\mathbf{u}_N$ with $\|\mathbf{u} - \mathbf{u}_N\| \leq (1 - 1/d)\varepsilon$, so that indeed $\|\mathbf{w}_k - \mathbf{u}_N\| \leq \varepsilon$, or by $\mathbf{w}_k$ itself otherwise, one easily verifies that

$$\|\mathbf{w}_k\|_{\mathcal{A}_\infty^s} = \sup_{\varepsilon > 0} \varepsilon \min_{\{N \in \mathbb{N}_0 : \|\mathbf{w}_k - (\mathbf{w}_k)_N\| \leq \varepsilon\}} N^s \leq (1 + D_5) \|\mathbf{u}\|_{\mathcal{A}_\infty^s}. \qquad (4.5)$$

This uniform boundedness of the iterands in $\|\cdot\|_{\mathcal{A}_\infty^s}$ is the key to the design of suitable routines **APPLY** and **GALSOLVE**.

The idea behind the approximate application of $\mathbf{A}$ introduced in [4] is as follows: Chop the input vector $\mathbf{w} = \mathbf{w}_k$ into $j := \lceil \log_2(\# \operatorname{supp} \mathbf{w} + 1) - 1 \rceil$ pieces

$$\mathbf{w} = \mathbf{w}_1 + [\mathbf{w}_3 - \mathbf{w}_1] + \cdots + [\mathbf{w}_{2^{j+1}-1} - \mathbf{w}_{2^j-1}],$$

where $\mathbf{w}_{2^{p+1}-1}$ is the best $2^{p+1} - 1$ approximation to $\mathbf{w}$, and approximate the application of $\mathbf{A}$ to each of the pieces separately. We have $\|\mathbf{w}_{2^{p+1}-1} - \mathbf{w}_{2^p-1}\| \leq \|\mathbf{w} - \mathbf{w}_{2^p-1}\| \lesssim 2^{-sp} \|\mathbf{w}\|_{\mathcal{A}_\infty^s}$ (cf. footnote 1). So the larger is $\# \operatorname{supp}[\mathbf{w}_{2^{p+1}-1} - \mathbf{w}_{2^p-1}] = 2^p$, and so the more columns of the approximation of $\mathbf{A}$ are needed, the smaller is $\|\mathbf{w}_{2^{p+1}-1} - \mathbf{w}_{2^p-1}\|$ and so the less accurate this approximation can be. So this approximate matrix vector product uses both the near sparsity of the matrix and that of the vector.

The following routine **APPLY** is based on a variation of the above principle. There will be the following differences with the original routine from [4] aiming at increasing the efficiency:

- In order to avoid suboptimal log-factors due to sorting needed to determine best N-term approximations, the approximate sorting routine **BUCKETSORT** is applied.
- Since there is little to be gained by approximating columns of $\mathbf{A}$ corresponding to indices inside one bucket with different accuracies, we will chop $\mathbf{w}$ in pieces corresponding to the subdivision of the indices over the buckets. These pieces do not have a prescribed cardinality, but instead have a prescribed range of values. A similar idea was applied in [6], with ranges that depend on some upper bound for the value of s for which $\mathbf{u} \in \mathcal{A}^s$ can be expected. The **APPLY** routine below is universal in the sense that nothing about the value of s is needed.
- Instead of using a priori fixed accuracies of the approximate matrix vector product on each of the pieces, to minimize the cost for achieving a prescribed tolerance, we use an optimization strategy based on a posteriori information.

APPLY$[\Lambda, \mathbf{w}, \varepsilon] \to \mathbf{z}$:

% $\mathbf{w} \in \ell_2(\Lambda)$ *is finitely supported,* $\varepsilon > 0$.

% *For* $j \in \mathbb{N}_0$, *let* $\mathbf{A}^{(j)}$ *and* $\overline{e}_j$ *be matrices and constants such that* $\mathbf{A}^{(0)} := 0$,

% $\|\mathbf{A} - \mathbf{A}^{(j)}\| \leq \overline{e}_j$ *and* $\lim_{j \to \infty} \overline{e}_j = 0$.

% *For* $j \in \mathbb{N}_0$, *let* c_j *be an upperbound for both the number of non-zero entries*

% *in each column of* $\mathbf{A}^{(j)}$, *and for the number of arithmetic operations needed*

% *for their joint computation, where* $\sup_{j \in \mathbb{N}} \frac{c_{j+1}}{c_j} < \infty$ *and* $c_0 := 0$.

$[(\mathbf{w}_{[p]})_p, \cdot] := \mathbf{BUCKETSORT}[\mathbf{w}, \varepsilon/(2\overline{e}_0)]$

Compute the smallest $\ell \in \mathbb{N}_0$ *with*

$$\delta := \overline{e}_0 \left\| \mathbf{w} - \sum_{p=1}^{\ell} \mathbf{w}_{[p]} \right\| \leq \varepsilon/2.$$

Determine, in $\mathcal{O}(\#\operatorname{supp}\mathbf{w})$ *operations,* $\mathbf{j} \in \mathbb{N}_0^{\ell}$ *such that* $\sum_{p=1}^{\ell} \overline{e}_{\mathbf{j}_p} \|\mathbf{w}_{[p]}\| \leq \varepsilon - \delta$ *and* $c_{\mathbf{j}_p} \lesssim c_{\tilde{\mathbf{j}}_p}$ $(p = 1, \ldots, \ell)$, *where* $\tilde{\mathbf{j}} \in \mathbb{N}_0^{\ell}$ *is the solution of*

$$\sum_{p=1}^{\ell} c_{\tilde{\mathbf{j}}_p} \#\operatorname{supp}\mathbf{w}_{[p]} \to \min!, \qquad \sum_{p=1}^{\ell} \overline{e}_{\tilde{\mathbf{j}}_p} \|\mathbf{w}_{[p]}\| \leq \varepsilon - \delta. \tag{4.6}$$

Compute

$$z := \mathbf{P}_\Lambda \sum_{p=1}^{\ell} \mathbf{A}^{(\mathbf{j}_p)} \mathbf{w}_{[p]}.$$

Theorem 4.6 *For* $\mathbf{z} := \mathbf{APPLY}[\Lambda, \mathbf{w}, \varepsilon]$, *we have* $\|\mathbf{P}_\Lambda \mathbf{A}\mathbf{w} - \mathbf{z}\| \leq \varepsilon$. *If for some* $s^* > 0$, *for any* $\overline{s} < s^*$, $\sup_{j \in \mathbb{N}} e_j c_j^{\overline{s}} < \infty$, *i.e., if* $\mathbf{A}$ *is* s^*-*computable, then for any* $s \in (0, s^*)$,

$$\#\operatorname{supp}\mathbf{z} \lesssim \varepsilon^{-1/s} \|\mathbf{w}\|_{\mathcal{A}_\infty^s}^{1/s}, \tag{4.7}$$

and the number of operations required by the call can be bounded by some absolute multiple of

$$1 + \#\operatorname{supp}\mathbf{w} + \varepsilon^{-1/s} \|\mathbf{w}\|_{\mathcal{A}_\infty^s}^{1/s}, \tag{4.8}$$

both estimates only dependent on s, *when it tends to* 0 *or to* s^*, *and on* $\overline{e}_0$, *being an upper bound for* $\|\mathbf{A}\|$.

Before proving this theorem, we discuss how to determine valid values of $\mathbf{j}$ in the **APPLY** routine in two common situations. If for some constants C and D, $c_j = Cj$ and $\overline{e}_j = Dj^{-s^*}$, so that $\mathbf{A}$ is s^*-computable, then

$$\tilde{\mathbf{j}}_p = \left(\frac{\|\mathbf{w}_{[p]}\|}{\#\operatorname{supp}\mathbf{w}_{[p]}} \right)^{\frac{1}{s^*+1}} \left(\frac{\sum_{q=1}^{\ell} \|\mathbf{w}_{[q]}\|^{\frac{1}{s^*+1}} (\#\operatorname{supp}\mathbf{w}_{[q]})^{\frac{s^*}{1+s^*}}}{(\varepsilon - \delta)/D} \right)^{\frac{1}{s^*}}$$

is the solution of (4.6) when minimization is performed over $\mathbb{R}^\ell$. If for some constants C, D and $\varrho > 0$, $c_j = Cj$ and $\overline{e}_j = D2^{-\varrho j}$, so that $\mathbf{A}$ is even ∞-computable, then

$$\tilde{\mathbf{j}}_p = \log_2\left(\frac{\|\mathbf{w}_{[p]}\| \sum_{q=1}^{\ell} \# \operatorname{supp} \mathbf{w}_{[q]}}{\# \operatorname{supp} \mathbf{w}_{[p]}(\varepsilon - \delta)/D}\right)\Big/\varrho$$

is the solution of (4.6) when minimization is performed over $\mathbb{R}^\ell$. Assuming these $\tilde{\mathbf{j}}_p$ are non-negative, by rounding them up to the nearest value in $\mathbb{N}_0$ one obtains a valid $\mathbf{j}$.

Proof of Theorem 4.6 Let $s \in (0, s^*)$ and $s < \bar{s}_1 < \bar{s}_2 < s^*$. The first statement follows from $\|\mathbf{A}\|\|\mathbf{w} - \sum_{p=1}^{\ell} \mathbf{w}_{[p]}\| \leq \delta$ and $\sum_{p=1}^{\ell} \|\mathbf{A} - \mathbf{A}^{(\mathbf{j}_p)}\|\|\mathbf{w}_{[p]}\| \leq \varepsilon - \delta$.

By Proposition 4.4, the cost of a call $[(\mathbf{w}_{[p]})_p, \cdot] = \mathbf{BUCKETSORT}[\mathbf{w}, \varepsilon/(2\overline{e}_0)]$ is bounded by some absolute multiple of

$$\# \operatorname{supp} \mathbf{w} + \max\left(1, \lceil \log(\# \operatorname{supp} \mathbf{w}) + \log(\|\mathbf{w}\|_\infty/(\varepsilon/2\overline{e}_0)) \rceil\right)$$

$$\lesssim \# \operatorname{supp} \mathbf{w} + \max\left(1, \log(\varepsilon^{-1}\|\mathbf{w}\|)\right)$$

$$\lesssim \# \operatorname{supp} \mathbf{w} + 1 + \varepsilon^{-1/s}\|\mathbf{w}\|_{\mathcal{A}_\infty^s}^{1/s}.$$

With $\tau := (\frac{1}{2} + s)^{-1} > 2$, it is known (see, e.g., [9]) that

$$\#\{\lambda \in \nabla : |\mathbf{w}_\lambda| > \eta\} \lesssim \eta^{-\tau}\|\mathbf{w}\|_{\mathcal{A}_\infty^s}^\tau \quad (\eta > 0).$$

As a consequence, we have

$$\# \operatorname{supp} \mathbf{w}_{[p]} \lesssim 2^{p\tau/2}\|\mathbf{w}\|_\infty^{-\tau}\|\mathbf{w}\|_{\mathcal{A}_\infty^s}^\tau$$

and

$$\|\mathbf{w}_{[p]}\| \lesssim 2^{-p/2}\|\mathbf{w}\|_\infty \sqrt{\# \operatorname{supp} \mathbf{w}_{[p]}} \lesssim 2^{-ps\tau/2}\|\mathbf{w}\|_\infty^{1-\tau/2}\|\mathbf{w}\|_{\mathcal{A}_\infty^s}^{\tau/2}.$$

The proof will be completed once we have shown that there exists a $\mathbf{j} \in \mathbb{N}_0^\ell$ with $\sum_{p=1}^{\ell} \overline{e}_{\mathbf{j}_p}\|\mathbf{w}_{[p]}\| \leq \varepsilon - \delta$ and $\sum_{p=1}^{\ell} c_{\mathbf{j}_p} \# \operatorname{supp} \mathbf{w}_{[p]} \lesssim \varepsilon^{-1/s}\|\mathbf{w}\|_{\mathcal{A}_\infty^s}^{1/s}$. For $\ell = 0$ there is nothing to prove, so we assume that $\ell > 0$.

By definition of ℓ, we have

$$\varepsilon/2 < \overline{e}_0\left\|\mathbf{w} - \sum_{p=1}^{\ell-1} \mathbf{w}_{[p]}\right\| = \overline{e}_0\sqrt{\sum_{p=\ell}^{\infty} \|\mathbf{w}_{[p]}\|^2} \lesssim \overline{e}_0 2^{-\ell s\tau/2}\|\mathbf{w}\|_\infty^{1-\tau/2}\|\mathbf{w}\|_{\mathcal{A}_\infty^s}^{\tau/2},$$

or

$$2^{\ell\tau/2}\|\mathbf{w}\|_\infty^{-\tau}\|\mathbf{w}\|_{\mathcal{A}_\infty^s}^\tau \lesssim \varepsilon^{-1/s}\|\mathbf{w}\|_{\mathcal{A}_\infty^s}^{1/s}. \tag{4.9}$$

Note that here we used the notation $\mathbf{w}_{[p]}$ also to denote the restriction of $\mathbf{w}$ to indices from buckets that actually were not generated in **BUCKETSORT**.

Let $J \geq \ell$ be the smallest integer such that $\sum_{p=1}^{\ell} 2^{-(J-p)\bar{s}_1 \tau/2} \|\mathbf{w}_{[p]}\| \leq \varepsilon - \delta$. So when $J > \ell$, from $\bar{s}_1 > s$ we have

$$\varepsilon/2 \leq \varepsilon - \delta < \sum_{p=1}^{\ell} 2^{-(J-1-p)\bar{s}_1 \tau/2} \|\mathbf{w}_{[p]}\|$$

$$< \sum_{p=1}^{\ell} 2^{-(J-1-p)\bar{s}_1 \tau/2} 2^{-ps\tau/2} \|\mathbf{w}\|_\infty^{1-\tau/2} \|\mathbf{w}\|_{\mathcal{A}_\infty^s}^{\tau/2}$$

$$\lesssim 2^{-(J-1-\ell)\bar{s}_1 \tau/2} 2^{-\ell s \tau/2} \|\mathbf{w}\|_\infty^{1-\tau/2} \|\mathbf{w}\|_{\mathcal{A}_\infty^s}^{\tau/2}$$

$$\leq 2^{-(J-1)s\tau/2} \|\mathbf{w}\|_\infty^{1-\tau/2} \|\mathbf{w}\|_{\mathcal{A}_\infty^s}^{\tau/2}$$

or

$$2^{J\tau/2} \|\mathbf{w}\|_\infty^{-\tau} \|\mathbf{w}\|_{\mathcal{A}_\infty^s}^\tau \lesssim \varepsilon^{-1/s} \|\mathbf{w}\|_{\mathcal{A}_\infty^s}^{1/s}, \tag{4.10}$$

which by (4.9) is also valid when $J = \ell$.

Now select $\mathbf{j}_p$ to be the smallest integer such that $\bar{e}_{\mathbf{j}_p} \leq 2^{-(J-p)\bar{s}_1 \tau/2}$. Then indeed $\sum_{p=1}^{\ell} \bar{e}_{\mathbf{j}_p} \|\mathbf{w}_{[p]}\| \leq \varepsilon - \delta$, and by $\bar{e}_{j-1} c_{j-1}^{\bar{s}_2} \lesssim 1$ and $c_j \lesssim c_{j-1}$, we have $c_{\mathbf{j}_p} \lesssim 2^{(J-p)(\bar{s}_1/\bar{s}_2)\tau/2}$. From (4.10), we conclude that

$$\sum_{p=1}^{\ell} c_{\mathbf{j}_p} \# \operatorname{supp} \mathbf{w}_{[p]} \lesssim \sum_{p=1}^{\ell} 2^{(J-p)(\bar{s}_1/\bar{s}_2)\tau/2} 2^{p\tau/2} \|\mathbf{w}\|_\infty^{-\tau} \|\mathbf{w}\|_{\mathcal{A}_\infty^s}^\tau$$

$$\lesssim 2^{(J-\ell)(\bar{s}_1/\bar{s}_2)\tau/2} 2^{\ell\tau/2} \|\mathbf{w}\|_\infty^{-\tau} \|\mathbf{w}\|_{\mathcal{A}_\infty^s}^\tau$$

$$\lesssim 2^{J\tau/2} \|\mathbf{w}\|_\infty^{-\tau} \|\mathbf{w}\|_{\mathcal{A}_\infty^s}^\tau \lesssim \varepsilon^{-1/s} \|\mathbf{w}\|_{\mathcal{A}_\infty^s}^{1/s},$$

which completes the proof. $\square$

Next we present a realization of the routine **GALSOLVE**. Given $\boldsymbol{\Lambda} \subset \nabla$, $\mathbf{g}_{\boldsymbol{\Lambda}}, \mathbf{w}_{\boldsymbol{\Lambda}}^{(0)} \in \ell_2(\boldsymbol{\Lambda})$, and $\delta \geq \varepsilon > 0$ with $\|\mathbf{g}_{\boldsymbol{\Lambda}} - \mathbf{A}_{\boldsymbol{\Lambda}} \mathbf{w}_{\boldsymbol{\Lambda}}^{(0)}\| \leq \delta$, it should produce a $\mathbf{w}_{\boldsymbol{\Lambda}} \in \ell_2(\boldsymbol{\Lambda})$ with $\|\mathbf{g}_{\boldsymbol{\Lambda}} - \mathbf{A}_{\boldsymbol{\Lambda}} \mathbf{w}_{\boldsymbol{\Lambda}}\| \leq \varepsilon$. **GALSOLVE** will only be called for $\mathbf{w}_{\boldsymbol{\Lambda}}^{(0)}$ being the previous iterand produced inside **SOLVE**, and for δ/ε being uniformly bounded, the latter meaning that only a fixed reduction of the initial error has to be achieved.

An obvious approach is to apply a fixed number of iterations of some iterative method. A problem, however, is that generally $\mathbf{A}_{\boldsymbol{\Lambda}}$ is not truly sparse, in the sense that its number of non-zero entries is not of the order of $\#\boldsymbol{\Lambda}$. The approach we follow is based on the observation that if for some constant $\mu < 1$, a fixed, sufficiently large j is chosen such that with $\mathbf{A}_{\boldsymbol{\Lambda}}^{(j)} := \mathbf{P}_{\boldsymbol{\Lambda}} \mathbf{A}^{(j)} \mathbf{I}_{\boldsymbol{\Lambda}}$, $\|\mathbf{I} - \mathbf{A}_{\boldsymbol{\Lambda}} (\mathbf{A}_{\boldsymbol{\Lambda}}^{(j)})^{-1}\| \leq \frac{\mu\varepsilon}{\delta}$, then $\|\mathbf{g}_{\boldsymbol{\Lambda}} - \mathbf{A}_{\boldsymbol{\Lambda}} [\mathbf{w}_{\boldsymbol{\Lambda}}^{(0)} + (\mathbf{A}_{\boldsymbol{\Lambda}}^{(j)})^{-1} (\mathbf{g}_{\boldsymbol{\Lambda}} - \mathbf{A}_{\boldsymbol{\Lambda}} \mathbf{w}_{\boldsymbol{\Lambda}}^{(0)})]\| \leq \mu\varepsilon$. Having chosen μ being strictly less than one, room is left to approximate the application of $(\mathbf{A}_{\boldsymbol{\Lambda}}^{(j)})^{-1}$ by an iterative

method ($\mathbf{A}^{(j)}$ *is* sparse), and to replace the initial residual $\mathbf{g_\Lambda} - \mathbf{A_\Lambda}\mathbf{w}_\Lambda^{(0)}$ by an approximation. Concerning the latter, we approximate $\mathbf{A_\Lambda}\mathbf{w}_\Lambda^{(0)}$ by a call of **APPLY**. Its cost is controlled in terms of $\#\mathbf{\Lambda}$ and $\|\mathbf{w}_\Lambda^{(0)}\|_{\mathcal{A}_\infty^s}$, the latter known to be bounded (cf. (4.5)).

The *defect correction* procedure outlined above is an alternative for the approximate application of Richardson's iteration to $\mathbf{A_\Lambda}\mathbf{w_\Lambda} = \mathbf{g_\Lambda}$ starting with $\mathbf{w}_\Lambda^{(0)}$ as proposed in [4], where in *each* iteration the application of $\mathbf{A_\Lambda}$ is approximated by a call of **APPLY**. Since, due to the built-in adaptivity, calls of **APPLY** are quantitatively demanding, whereas furthermore within our defect correction method the iterative method can be chosen to be of optimal Krylov type, we expect the latter method to have better quantitative properties.

GALSOLVE$[\mathbf{\Lambda}, \mathbf{g_\Lambda}, \mathbf{w}_\Lambda^{(0)}, \delta, \varepsilon] \to \mathbf{w_\Lambda}$:
% The input satisfies $\delta \geq \varepsilon > 0$, $\mathbf{\Lambda} \subset \nabla$, $\mathbf{g_\Lambda}, \mathbf{w}_\Lambda^{(0)} \in \ell_2(\mathbf{\Lambda})$ with $\|\mathbf{g_\Lambda} - \mathbf{A_\Lambda}\mathbf{w}_\Lambda^{(0)}\| \leq \delta$.
*% Let $(\overline{e}_j)_j$, $(c_j)_j$ be defined as in the **APPLY** routine, and let $\underline{e}_0^{-1}$ be an upper*
% bound for $\|\mathbf{A}^{-1}\|$.
Compute the smallest $j \in \mathbb{N}_0$ with $\overline{e}_j(\underline{e}_0)^{-1} \leq \frac{\varepsilon}{3\varepsilon+3\delta}$
$\mathbf{r}_\Lambda^{(0)} := \mathbf{g_\Lambda} - \mathbf{APPLY}[\mathbf{\Lambda}, \mathbf{w}_\Lambda^{(0)}, \frac{\varepsilon}{3}]$
With $\mathbf{A}_\Lambda^{(j)} := \mathbf{P_\Lambda}\mathbf{A}^{(j)}\mathbf{I_\Lambda}$, approximate $(\mathbf{A}_\Lambda^{(j)})^{-1}\mathbf{r}_\Lambda^{(0)}$ using the Conjugate Residuals method starting with zero until the current iterand $\mathbf{e_\Lambda}$ satisfies $\|\mathbf{r}_\Lambda^{(0)} - \mathbf{A}_\Lambda^{(j)}\mathbf{e_\Lambda}\| \leq \frac{\varepsilon}{3}$
$\mathbf{w_\Lambda} := \mathbf{w}_\Lambda^{(0)} + \mathbf{e_\Lambda}$

Theorem 4.7 $\mathbf{w_\Lambda} := \mathbf{GALSOLVE}[\mathbf{\Lambda}, \mathbf{g_\Lambda}, \mathbf{w}_\Lambda^{(0)}, \delta, \varepsilon]$ *satisfies* $\|\mathbf{g_\Lambda} - \mathbf{A_\Lambda}\mathbf{w_\Lambda}\| \leq \varepsilon$. *If for any* $\overline{s} < s^*$, $\sup_{j \in \mathbb{N}} \overline{e}_j c_j^{\overline{s}} < \infty$, *i.e., if $\mathbf{A}$ is s^*-computable, then for any $s < s^*$, for some increasing function η, the number of operations required by the call can be bounded by some absolute multiple of*

$$\varepsilon^{-1/s}\|\mathbf{w}_\Lambda^{(0)}\|_{\mathcal{A}_\infty^s}^{1/s} + \eta(\delta/\varepsilon)\#\mathbf{\Lambda} + 1, \tag{4.11}$$

only dependent on s when it tends to 0 or to s^*, and on $\overline{e}_0$ and $\underline{e}_0$.

Proof The selection of j implies that $\|\mathbf{A_\Lambda} - \mathbf{A}_\Lambda^{(j)}\|\|\mathbf{A}^{-1}\| \leq \frac{\varepsilon}{3\varepsilon+3\delta}$. So writing $\langle\mathbf{A}_\Lambda^{(j)}\mathbf{v_\Lambda}, \mathbf{v_\Lambda}\rangle = \langle\mathbf{A_\Lambda}\mathbf{v_\Lambda}, \mathbf{v_\Lambda}\rangle - \langle(\mathbf{A_\Lambda} - \mathbf{A}_\Lambda^{(j)})\mathbf{v_\Lambda}, \mathbf{v_\Lambda}\rangle$, we find

$$\left(1 - \frac{\varepsilon}{3\varepsilon+3\delta}\right)\|\mathbf{A}^{-1}\|^{-1}\|\mathbf{v_\Lambda}\|^2 \leq \langle\mathbf{A}_\Lambda^{(j)}\mathbf{v_\Lambda}, \mathbf{v_\Lambda}\rangle$$

$$\leq \|\mathbf{A}\|\|\mathbf{v_\Lambda}\|^2 + \frac{\varepsilon}{3\varepsilon+3\delta}\|\mathbf{A}^{-1}\|^{-1}\|\mathbf{v_\Lambda}\|^2.$$

We infer that $\|\mathbf{A_\Lambda} - \mathbf{A}_\Lambda^{(j)}\|\|(\mathbf{A}_\Lambda^{(j)})^{-1}\| \leq \frac{\frac{\varepsilon}{3\varepsilon+3\delta}}{1 - \frac{\varepsilon}{3\varepsilon+3\delta}} = \frac{\varepsilon}{2\varepsilon+3\delta}$, and by $\frac{\varepsilon}{3\varepsilon+3\delta} \leq \frac{1}{6}$, that $\kappa(\mathbf{A}_\Lambda^{(j)}) \leq \frac{1}{5} + \frac{6}{5}\kappa(\mathbf{A})$. Writing

$$\mathbf{g}_\Lambda - \mathbf{A}_\Lambda \mathbf{w}_\Lambda = \big(\mathbf{g}_\Lambda - \mathbf{A}_\Lambda \mathbf{w}_\Lambda^{(0)} - \mathbf{r}_\Lambda^{(0)}\big) + \big(\mathbf{r}_\Lambda^{(0)} - \mathbf{A}_\Lambda^{(j)}\mathbf{e}_\Lambda\big)$$
$$+ \big(\mathbf{I} - \mathbf{A}_\Lambda \big(\mathbf{A}_\Lambda^{(j)}\big)^{-1}\big)\big(\mathbf{r}_\Lambda^{(0)} + \mathbf{A}_\Lambda^{(j)}\mathbf{e}_\Lambda - \mathbf{r}_\Lambda^{(0)}\big),$$

and using $\|\mathbf{r}_\Lambda^{(0)}\| \le \delta + \frac{\varepsilon}{3}$, we find $\|\mathbf{g}_\Lambda - \mathbf{A}_\Lambda \mathbf{w}_\Lambda\| \le \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{2\varepsilon+3\delta}(\delta + \frac{\varepsilon}{3} + \frac{\varepsilon}{3}) \le \varepsilon$.

By Theorem 4.6, for any $s \in (0, s^*)$ the number of operations needed for the call **APPLY**$[\Lambda, \mathbf{w}_\Lambda^{(0)}, \frac{\varepsilon}{3}]$ can be bounded by some absolute multiple of $1 + \#\Lambda + \varepsilon^{-1/s}\|\mathbf{w}_\Lambda^{(0)}\|_{\mathcal{A}_\infty^s}^{1/s}$, dependent on s when it tends to 0 or to s^*, and on $\overline{e}_0$. For each $\overline{s} < s^*$, the number of non-zero entries in each column of $\mathbf{A}_\Lambda^{(j)}$ as well as the number of operations needed for their joint computation is bounded by a multiple of $((3 + \frac{3\delta}{\varepsilon})/\underline{e}_0)^{1/\overline{s}}$. Each iteration of Conjugate Residuals therefore requires not more than a multiple of $(1 + ((3 + \frac{3\delta}{\varepsilon})/\underline{e}_0)^{1/\overline{s}})\#\Lambda$ operations. The number of iterations is bounded by the smallest $\ell \in \mathbb{N}$ for which $\frac{2c^\ell}{1+c^{2\ell}} \ge \frac{\varepsilon/3}{\delta+\varepsilon/3}$ where $c = \frac{\sqrt{\kappa(\mathbf{A}_\Lambda^{(j)})}-1}{\sqrt{\kappa(\mathbf{A}_\Lambda^{(j)})}+1}$ (see, e.g., [22]). Since $\frac{2c^\ell}{1+c^{2\ell}} \le 2c^\ell$, this number is not larger than $\lceil \log(2(3\delta/\varepsilon + 1)/\log(1/c)\rceil$. Using that $\log(1/c)^{-1} \lesssim \sqrt{\kappa(\mathbf{A})} \le \sqrt{\overline{e}_0/\underline{e}_0}$, the proof is completed. $\square$

Using that the tolerances for calls of **APPLY** and **RHS** inside **SOLVE** are geometrically decreasing, that for $\mathbf{u} \in \mathcal{A}_\infty^s$ the iterands $\mathbf{w}_k$ in **SOLVE** satisfy $\|\mathbf{w}_k\|_{\mathcal{A}_\infty^s} \lesssim \|\mathbf{u}\|_{\mathcal{A}_\infty^s}$ (cf. (4.5)), and finally that calls of **GALSOLVE**$[\cdot, \cdot, \cdot, \delta, \varepsilon]$ are only made for $\delta/\varepsilon \lesssim 1$, using Theorems 4.6 and 4.7 we end up with the following result. Details can be verified by substituting the bounds from Theorems 4.6 and 4.7 into the corresponding proof from [16].

Theorem 4.8 *If $\nu_{-1} \eqsim \|\mathbf{f}\| \gtrsim \varepsilon$, and for some $0 < s < s^*$, $\mathbf{u} \in \mathcal{A}_\infty^s$ and $\mathbf{rhs}_s < \infty$ (cf. (1.6)), then the number of operations required by the call **SOLVE**$[\nu_{-1}, \varepsilon]$ is bounded by a multiple of*

$$\varepsilon^{-1/s}\big[\|\mathbf{u}\|_{\mathcal{A}_\infty^s}^{1/s} + \mathbf{rhs}_s^{1/s}\big], \tag{4.12}$$

only dependent on s when it tends to 0 or to s^*, on $\overline{e}_0$ and $\underline{e}_0$, and on the parameters $\alpha, \omega, \gamma, \theta$ when they approach the boundaries of their domains of definition.

Remark 4.9 The expression (4.12) is also an upper bound for the sum of the lengths of all vectors that were generated by the call **SOLVE**$[\nu_{-1}, \varepsilon]$.

5 Application to High Dimensional Elliptic PDEs

We are going to apply **SOLVE** to systems $\mathbf{Au} = \mathbf{f}$ that result from the boundary value problems (1.1) written in the tensor product wavelet basis (1.3) with univariate wavelet bases $\Psi^{(z)}$ as in Sect. 3. With $\overline{e}_0, \underline{e}_0$ being available upper or lower bounds for $\max_z \lambda_{\max}^{(1,z)}$ or $\min_z \lambda_{\min}^{(1,z)}$ from (2.1), respectively, we have that $\|\mathbf{A}\| \le \overline{e}_0$ and $\|\mathbf{A}^{-1}\| \le \underline{e}_0^{-1}$ uniformly in n, $c_0 \ge 0$ and $c_m > 0$ ($1 \le m \le n$). With C_{sp} being some

available upper bound for $\max_z \sup_j 2^{j/2} \| \vec{A}_z - \vec{A}_z^{(j)} \|$ (cf. Proposition 3.3), we know how to construct a matrix $\mathbf{A}^{(j)}$ having a multiple of jn non-zeros in each row and column with $\| \mathbf{A} - \mathbf{A}^{(j)} \| \leq C_{\mathrm{sp}} 2^{-j/2}$ uniformly in n, $c_0 \geq 0$ and $c_m > 0$ ($1 \leq m \leq n$). So on the one hand, for any fixed n, $\mathbf{A}$ is ∞-computable, but on the other hand, the number of non-zeros in each row and column of $\mathbf{A}^{(j)}$ grows linearly with n.

Taking the latter into account within the preceding analysis, the expressions (4.7), (4.8), and (4.11) read as

$$\begin{aligned} & \# \mathrm{supp}\, \mathbf{z} \lesssim n \varepsilon^{-1/s} \| \mathbf{w} \|_{\mathcal{A}_\infty^s}^{1/s}, \\ & 1 + \# \mathrm{supp}\, \mathbf{w} + n \varepsilon^{-1/s} \| \mathbf{w} \|_{\mathcal{A}_\infty^s}^{1/s}, \end{aligned} \tag{5.1}$$

and

$$n \varepsilon^{-1/s} \big\| \mathbf{w}_{\mathbf{\Lambda}}^{(0)} \big\|_{\mathcal{A}_\infty^s}^{1/s} + \eta(\delta/\varepsilon) n \# \mathbf{\Lambda} + 1,$$

respectively, only dependent on s when it tends to 0 or ∞. This leads to the following version of Theorems 4.5 and 4.8 when specialized to these systems resulting from high dimensional elliptic equations.

Theorem 5.1 $\mathbf{w} := \mathbf{SOLVE}[\nu_{-1}, \varepsilon]$ *terminates with* $\| \mathbf{f} - \mathbf{Aw} \| \leq \varepsilon$, *so that* $\| u - \sum_{\lambda \in \nabla} w_\lambda \psi_\lambda \| = \| \mathbf{u} - \mathbf{w} \| \leq \underline{e}_0^{-\frac{1}{2}} \varepsilon$. *If, for some* $s > 0$, $\mathbf{u} \in \mathcal{A}_\infty^s$, *then*

$$\mathrm{supp}\, \mathbf{w} \lesssim \varepsilon^{-1/s} \| \mathbf{u} \|_{\mathcal{A}_\infty^s}^{1/s}.$$

If $\nu_{-1} \eqsim \| \mathbf{f} \| \gtrsim \varepsilon$, *and* $\mathbf{rhs}_s < \infty$ (*cf.* (1.6)), *then the number of operations required by the call is bounded by a multiple of*

$$\varepsilon^{-1/s} \big[n \| \mathbf{u} \|_{\mathcal{A}_\infty^s}^{1/s} + \mathbf{rhs}_s^{1/s} \big], \tag{5.2}$$

everything uniformly in n, $c_0 \geq 0$, *and* $c_m > 0$ ($1 \leq m \leq n$), *and only dependent on* s *when it tends to 0 or to* ∞, *and on the parameters* $\alpha, \omega, \gamma, \theta$ *when they approach the boundaries of their domains of definition.*

What is left is the discussion of *a valid routine* **RHS**. As a consequence of Theorem 4.6, for any $s > 0$, $\mathbf{A} : \mathcal{A}_\infty^s \to \mathcal{A}_\infty^s$ is bounded with $\| \mathbf{Av} \|_{\mathcal{A}_\infty^s} \lesssim n \| \mathbf{v} \|_{\mathcal{A}_\infty^s}$. Indeed, for $\mathbf{v} \in \mathcal{A}_\infty^s$ and $\varepsilon > 0$, let $N \in \mathbb{N}$ be such that $\| \mathbf{v} - \mathbf{v}_N \| \leq \varepsilon/(2 \| \mathbf{A} \|)$. Put $\mathbf{z} := \mathbf{APPLY}[\nabla, \mathbf{v}_N, \varepsilon/2]$. Then $\# \mathrm{supp}\, \mathbf{z} \lesssim n(\varepsilon/2)^{-1/s} \| \mathbf{v}_N \|_{\mathcal{A}_\infty^s}^{1/s}$ (here we applied (5.1) replacing (4.7)). Now the statement follows from $\| \mathbf{v}_N \|_{\mathcal{A}_\infty^s} \leq \| \mathbf{v} \|_{\mathcal{A}_\infty^s}$, $\| \mathbf{Av} - \mathbf{z} \| \leq \| \mathbf{A} \| \| \mathbf{v} - \mathbf{v}_N \| + \| \mathbf{Av}_N - \mathbf{z} \| \leq \varepsilon$ and the definition of $\mathcal{A}_\infty^s$.

As a consequence, if the right-hand side vector $\mathbf{f}$ is such that we can realize its (quasi) best N-term approximations in $\mathcal{O}(N)$ operations, uniformly in n, then assuming $\mathbf{u} \in \mathcal{A}_\infty^s$, the resulting routine **RHS** satisfies $\mathbf{rhs}_s \lesssim n \| \mathbf{u} \|_{\mathcal{A}_\infty^s}$. In this situation, the upper bound (5.2) for the cost of **SOLVE** is never dominated by the term $\varepsilon^{-1/s} \mathbf{rhs}_s^{1/s}$. Unfortunately, even for a family over n of smooth right-hand sides f, it seems hard to guarantee these conditions uniformly in n. On the other hand, the

estimate $\|\mathbf{f}\|_{\mathcal{A}_\infty^s} \leq \|\mathbf{A}\|_{\mathcal{A}_\infty^s \to \mathcal{A}_\infty^s} \|\mathbf{u}\|_{\mathcal{A}_\infty^s}$ might be very crude, so that quasi-optimal approximations, uniformly in n, are actually not needed to arrive at $\mathbf{rhs}_s \lesssim n\|\mathbf{u}\|_{\mathcal{A}_\infty^s}$.

In the following, we will focus on the situation that for some $\bar{s} > s$, $\mathbf{rhs}_{\bar{s}} < \infty$, however generally dependent on n and on $c_0 \geq 0$ and $c_m > 0$ ($1 \leq m \leq n$). In this situation, the bound (5.2) may be read as $\varepsilon^{-1/s} n \|\mathbf{u}\|_{\mathcal{A}_\infty^s}^{1/s} + \varepsilon^{-1/\bar{s}} \mathbf{rhs}_{\bar{s}}^{1/\bar{s}}$, meaning that in any case for any *fixed* n and c_m ($0 \leq m \leq n$), the cost of approximating the right-hand side is *asymptotically* negligible, i.e., when ε tends to zero.

Let us start by assuming that *all* univariate wavelets $\psi_\lambda^{(z)}$ on positive levels have d vanishing moments (cf. Remark 3.2), for the moment ignoring that this is not true for wavelets whose supports have non-empty intersection with a Dirichlet boundary. Then from $|\int_0^1 v\psi_\lambda^{(z)}| = \inf_{p \in P_{d-1}} |\int_0^1 (v-p)\psi_\lambda^{(z)}| \lesssim 2^{-(d+\frac{1}{2})|\lambda|} |v|_{W_\infty^d(0,1)}$, a tensor product argument, and taking into account the normalization of the product wavelets, for smooth f we infer that

$$\left| f(\boldsymbol{\psi}_\lambda) \right| = \left| \int_{(0,1)^n} f\boldsymbol{\psi}_\lambda \right| \lesssim \frac{2^{-(\frac{1}{2}+d)\|\lambda\|_1}}{\sqrt{c_0 + \sum_{m=1}^n c_m 4^{|\lambda_m|}}} \lesssim 2^{-((\frac{1}{2}+d)\|\lambda\|_1 + \|\lambda\|_\infty)}. \quad (5.3)$$

In view of this estimate, it is natural to consider approximations to $\mathbf{f}$ by, for some parameter ℓ, dropping all entries $f(\boldsymbol{\psi}_\lambda)$ with indices $\boldsymbol{\lambda}$ outside

$$\nabla_\ell^{(\text{optsp})} := \left\{ \lambda \in \nabla : \left(\frac{1}{2}+d\right)\|\lambda\|_1 + \|\lambda\|_\infty \leq \left(\frac{1}{2}+d+\frac{1}{n}\right)\ell \right\}. \quad (5.4)$$

Using that $\#\{\lambda \in \nabla_z : |\lambda| \leq k\} \approx 2^k$ (Remark 3.1), one may verify that $\#\nabla_\ell^{(\text{optsp})} \approx 2^\ell$ with a constant factor that depends on n. The error in this approximation can therefore be bounded by a multiple of $\sqrt{\sum_{k>\ell} 2^k 4^{-(\frac{1}{2}+d+\frac{1}{n})k}} \approx 2^{-(d+\frac{1}{n})\ell}$. So with an approximation having support length N, the error is not larger than $CN^{-(d+\frac{1}{n})}$, for some constant C that may depend on n (and on the coefficients c_m and f).

Remark 5.2 Following [20], we call $\text{span}\{\boldsymbol{\psi}_\lambda : \lambda \in \nabla_\ell^{(\text{optsp})}\}$ an *optimized sparse grid* space. With the standard sparse grid space $\text{span}\{\boldsymbol{\psi}_\lambda : \lambda \in \nabla, \|\lambda\|_1 \leq \ell\}$, one would need $N(\log N)^{n-1}$ unknowns to guarantee an error $\mathcal{O}(N^{-(d+\frac{1}{n})})$, or equivalently, with N unknowns the error is $\mathcal{O}(N^{-(d+\frac{1}{n})}(\log N)^{(n-1)(d+\frac{1}{n})})$. See Fig. 3 for an illustration.

Remark 5.3 Using $\|\mathbf{f} - \mathbf{g}\| \approx \|f - \sum_{\lambda \in \nabla} g_\lambda \boldsymbol{\psi}_\lambda\|_{H^{-1}(0,1)^n}$, one might have expected that with a $\mathbf{g}$ of length N an error $N^{-(d+1)}$ should be realizable. The analysis above shows, however, that generally the rate $d + \frac{1}{n}$ cannot be improved, and more generally, that when errors are measured in negative norms a mild deterioration of the rate as n increases cannot be avoided. Yet, in view of our aim to approximate the right-hand side with a rate better than that for the solution, note that for any n, $d + \frac{1}{n}$ is safely larger than the generally best possible rate $d - 1$ for approximating u.

The fact that for $z \neq \emptyset$, on all levels there are univariate wavelets having no vanishing moments is a price that has to be paid for the use of orthonormal wavelets. Indeed,

Fig. 3 Standard and optimized sparse grid space in two space dimensions for the approximation of the right-hand side
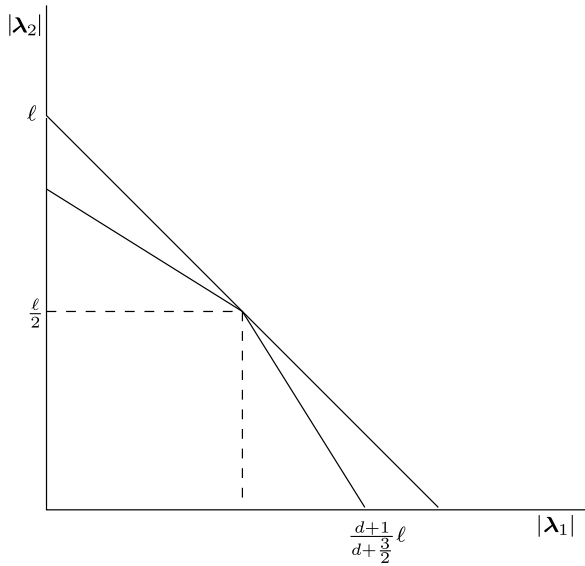
considering biorthogonal wavelets, one could select dual spaces that do not incorporate Dirichlet boundary conditions, so that all primal wavelets on positive levels have vanishing moments.

For $\lambda \in \nabla$, let $d(\lambda)$ denote the number of vanishing moments of $\psi_\lambda^{(z)}$, now being either d or 0. Then $|\int_0^1 v\psi_\lambda^{(z)}| \lesssim 2^{-(d(\lambda)+\frac{1}{2})|\lambda|}|v|_{W_\infty^{d(\lambda)}(0,1)}$, and instead of (5.3), we obtain

$$\left|f(\boldsymbol{\psi}_\lambda)\right| = \left|\int_{(0,1)^n} f\boldsymbol{\psi}_\lambda\right| \lesssim 2^{-(\|\|\boldsymbol{\lambda}\|\|_\infty + \sum_{m=1}^n (\frac{1}{2}+d(\lambda_m))|\lambda_m|)}.$$

In view of this estimate, we construct approximations to **f** by dropping all entries $\mathbf{f}(\psi_\lambda)$ with $\|\|\boldsymbol{\lambda}\|\|_\infty + \sum_{m=1}^n (\frac{1}{2} + d(\lambda_m))|\lambda_m| > (\frac{1}{2} + d + \frac{1}{n})\ell$ for some parameter ℓ. Using that for any $c > 0$,

$$\#\{\lambda \in \nabla : |\lambda| \leq k \text{ and } d(\lambda) = d, \text{ or } |\lambda| \leq ck \text{ and } d(\lambda) = 0\} \approx 2^k + c(k+1) \approx 2^k,$$

dependent on c but uniform in k, we obtain that

$$\#\left\{\boldsymbol{\lambda} \in \boldsymbol{\nabla} : \|\boldsymbol{\lambda}\|_\infty + \sum_{m=1}^n \left(\frac{1}{2} + d(\boldsymbol{\lambda}_m)\right)|\boldsymbol{\lambda}_m| \leq \left(\frac{1}{2} + d + \frac{1}{n}\right)k\right\} \approx 2^k.$$

As above, we conclude that with an approximation having support length N, the error is $\mathcal{O}(N^{-(d+\frac{1}{n})})$, for some constant C that may depend on n (and on the coefficients c_m and f).

In general, entries $f(\boldsymbol{\psi}_\lambda)$ have to be approximated by the application of a suitable *quadrature rule*. For simplicity, here we discuss this issue only in the situation that $z = \emptyset$ so that (5.3) is valid for all wavelets. In [27], we studied quadrature rules for approximating the entries of the stiffness matrix for high dimensional PDEs having

non-constant but smooth coefficients. From the analysis there, we deduce that, for sufficiently smooth f, the application of a product composite quadrature rule taking N function evaluations yields an approximation $\widetilde{f(\boldsymbol{\psi}_{\boldsymbol{\lambda}})}$ with

$$\left| f(\boldsymbol{\psi}_{\boldsymbol{\lambda}}) - \widetilde{f(\boldsymbol{\psi}_{\boldsymbol{\lambda}})} \right| \lesssim N^{-\alpha} 2^{-((\frac{1}{2}+d)\|\boldsymbol{\lambda}\|_1 + \|\boldsymbol{\lambda}\|_\infty)},$$

where α is proportional to the *order* of the one-dimensional composite rule, and inversely proportional to n. This dependence of α on n can be (nearly) removed by applying sparse product rules. This gives the advantage that the order of the rule can be chosen to be (nearly) independent of n, and with that so are the smoothness requirements on f.

Let us now select $\alpha > d + \frac{1}{n}$ and a parameter $\sigma \in (d + \frac{1}{n}, \alpha)$. Then given $\boldsymbol{\nabla}_\ell^{(\mathrm{optsp})}$, for approximating $f(\boldsymbol{\psi}_{\boldsymbol{\lambda}})$ we take the rule with $N = N_{\boldsymbol{\lambda}} \approx 2^{(\ell-k)\frac{\sigma}{\alpha}}$ when $\boldsymbol{\lambda} \in \boldsymbol{\nabla}_k^{(\mathrm{optsp})} \backslash \boldsymbol{\nabla}_{k-1}^{(\mathrm{optsp})}$ ($\boldsymbol{\nabla}_{-1}^{(\mathrm{optsp})} := \emptyset$). Assuming that each evaluation of f takes $\mathcal{O}(1)$ operations, the total work for approximating $[f(\boldsymbol{\psi}_{\boldsymbol{\lambda}})]_{\boldsymbol{\lambda} \in \boldsymbol{\nabla}_\ell^{(\mathrm{optsp})}}$ is then bounded by some multiple of $\sum_{k=0}^\ell 2^k 2^{(\ell-k)\frac{\sigma}{\alpha}} \approx 2^\ell$, whereas the $\ell_2(\boldsymbol{\nabla})$-norm of the quadrature error can be bounded by some multiple of $\sqrt{\sum_{k=0}^\ell 2^k 4^{(k-\ell)\sigma} 4^{-(\frac{1}{2}+d+\frac{1}{n})k}} \approx 2^{-(d+\frac{1}{n})\ell}$. We conclude that for sufficiently smooth f, $\mathbf{rhs}_{d+\frac{1}{n}} < \infty$ (generally depending on n, the coefficients c_m and on f).

6 Implementation and Numerical Results

For some general, fixed symmetric positive definite bi-infinite matrix $\mathbf{A}$, our considerations in Sect. 4 concerning the operation count of **SOLVE** apply under the assumption that any entry of any vector that is generated inside this routine can be stored in or fetched from memory in $\mathcal{O}(1)$ operations. This assumption is valid in the unrealistic situation that we have an unlimited amount of memory at our disposal, where each location can be accessed in $\mathcal{O}(1)$ operations. Indeed, in that case, using some ordering on the index set $\boldsymbol{\nabla}$ to store any vector from $\ell_2(\boldsymbol{\nabla})$ as an infinite array, we have direct access to each of its entries. If, instead of assuming an infinite memory, we restrict ourselves to wavelet indices up to a certain high level, and allocate arrays for corresponding vectors, then we see that adaptivity does not result in a reduction of memory requirements compared to non-adaptive schemes.

Therefore, let us consider the storage of any finitely supported $\mathbf{v} \in \ell_2(\boldsymbol{\nabla})$ as a *linked list* of its non-zero entries ordered by their indices, which requires an amount of memory that is proportional to its support length. Here we make the assumption that the amount of memory needed to store any index $\lambda \in \nabla$ that is encountered is fixed. Then in view of Remark 4.9, the total *amount of memory* needed for a call **SOLVE**$[\cdot, \varepsilon]$ is bounded by an absolute multiple of

$$\varepsilon^{-1/s} \big[\|\mathbf{u}\|_{\mathcal{A}_\infty^s}^{1/s} + \mathbf{rhs}_s^{1/s} \big].$$

With this datastructure, the addition of two vectors requires a number of operations that is proportional to the *sum* of their lengths. This has an unfortunate consequence

for the cost of a call **APPLY**[∇, **w**, ε]. In this routine the sum is computed of a number of vectors, each of them being an approximate column of **A** multiplied by the corresponding coefficient from the input vector (from which very small coefficients were removed beforehand). The number of vectors K and the sum of their lengths L, and with that the maximal possible length of the output vector, are all bounded by some absolute multiple of $\varepsilon^{-1/s}\|\mathbf{w}\|_{\mathcal{A}_\infty^s}^{1/s}$. If one would add the second until the Kth vector to the first one, then one could end up with a complexity that is quadratic in L. Instead, if one adds pairwise vectors 1 and 2, 3 and 4, etc., and proceeds similarly with the set of $\lceil K/2 \rceil$ resulting vectors, until finally one vector is left being thus the output vector, the total complexity is $(\log K)L$. Indeed, the cost of each step of summing all pairs of current vectors is proportional to the total sum of their lengths, which is never larger than L, whereas the number of these steps is $\lceil \log_2 K \rceil$. Realizing that the indices in the output vector are ordered, one infers that such a log-factor cannot be avoided. By substituting the bounds for K and L, we conclude that with the storage of a vector as a linked list, the cost of this call of **APPLY** is bounded by a multiple of $1 + \#\operatorname{supp}\mathbf{w} + \log(\varepsilon^{-1/s}\|\mathbf{w}\|_{\mathcal{A}_\infty^s}^{1/s})\varepsilon^{-1/s}\|\mathbf{w}\|_{\mathcal{A}_\infty^s}^{1/s}$. With that we find that the *number of operations* needed for a call **SOLVE**[$\cdot$, ε] is bounded by a constant multiple of

$$\log\big(\varepsilon^{-1/s}\|\mathbf{u}\|_{\mathcal{A}_\infty^s}^{1/s}\big)\varepsilon^{-1/s}\|\mathbf{u}\|_{\mathcal{A}_\infty^s}^{1/s} + \varepsilon^{-1/s}\mathbf{rhs}_s^{1/s},$$

so that we end up with suboptimal computational complexity. So far this phenomenon seems to be ignored in the literature on adaptive wavelet methods.

Specializing the discussion to the solution of elliptic boundary value problems in high space dimensions using tensor product wavelet bases, we have to take into account that an index is an n-tuple. In view of the bound (5.2), the storage requirements are therefore of the order

$$n\varepsilon^{-1/s}\big[n\|\mathbf{u}\|_{\mathcal{A}_\infty^s}^{1/s} + \mathbf{rhs}_s^{1/s}\big].$$

Since comparing two indices now requires $\mathcal{O}(n)$ operations, the cost of the addition of two vectors stored as ordered linked lists is now proportional to n times the sum of their lengths. We end up with a total computational complexity of the order

$$n\big[\log\big(n\varepsilon^{-1/s}\|\mathbf{u}\|_{\mathcal{A}_\infty^s}^{1/s}\big)n\varepsilon^{-1/s}\|\mathbf{u}\|_{\mathcal{A}_\infty^s}^{1/s} + \varepsilon^{-1/s}\mathbf{rhs}_s^{1/s}\big].$$

In our actual implementation, we have not used linked lists but *hash tables* to store the vectors. We chose this mainly because of the convenience of programming, but also since we expect that *usually* it results in faster code. Clearly, given any hash function $\nabla \to \{1, \ldots, M\}$, with M being the size of the hash table, one can always construct $\Lambda \subset \nabla$ whose elements are all mapped onto a single entry in the hash table (collision), in which case the hash-based implementation will be very slow. In our tests, the cost of storing vectors was proportional to their length.

For our *numerical tests*, we considered the Laplace operator, i.e., $c_0 = 0$ and $c_1 = \cdots = c_n = 1$. We took $\Gamma_D = \overline{\partial\Omega \cap \mathbb{R}_{>0}^n}$, i.e, homogeneous Dirichlet boundary conditions at the right boundary point in each coordinate direction ($z_m \equiv z = \{1\}$). We used the collection of univariate $L_2(0,1)$-orthonormal continuous piecewise linear (multi-) wavelets discussed in Sect. 2 as building block for the tensor product

wavelet basis for $H_{0,\Gamma_D}^1(0,1)^n$. Numerically, we estimated the extremal eigenvalues $\lambda_{\min}^{(1,z)} \approx 0.19$ and $\lambda_{\max}^{(1,z)} \approx 2.8$ of the one-dimensional Laplace operator in wavelet coordinates $\vec{A}_z$, being also lower or upper bounds for the extremal eigenvalues of the n-dimensional Laplace operator in tensor product wavelet coordinates $\mathbf{A}$ (see (2.2)). These bounds are used in the routines **APPLY** and **GALSOLVE**.

Sparse approximations to $\mathbf{A}$ are constructed from sparse approximations to $\vec{A}$. In Sect. 3, we considered $\vec{A}_z^{(j)}$ by dropping all entries $(\vec{A}_z)_{\lambda,\mu}$ with $||\lambda| - |\mu|| > j$. It turns out that this $\vec{A}_z^{(j)}$ contains entries that are much smaller in modulus than some of those that were dropped. Therefore, in order to improve quantitative properties of the algorithm, in our experiments we used approximations to $\vec{A}$ by dropping all entries with modulus below some given tolerance. For the resulting approximation with a maximal number of j non-zero entries per row and column, which we again denote as $\vec{A}_z^{(j)}$, we found numerically that $\|\vec{A}_z - \vec{A}_z^{(j)}\| \lesssim 1.45 \times 2^{-0.05*j}$. Using notations as in the proof of Theorem 3.5, with $\vec{B}_z^{(j)} := \vec{D}_z^{\frac{1}{2}} \vec{A}_z^{(j)} \vec{D}_z^{\frac{1}{2}}$, the maximal number of non-zeros per row and column of $\mathbf{A}^{(j)} := \mathbf{D}^{\frac{1}{2}} (\vec{B}_z^{(j)} \otimes I \otimes \cdots \otimes I + \cdots + I \otimes \cdots \otimes I \otimes \vec{B}_z^{(j)})\mathbf{D}^{\frac{1}{2}}$ is nj (or actually $n(j-1)+1$), where $\|\mathbf{A} - \mathbf{A}^{(j)}\| \lesssim 1.45 \times 2^{-0.05*j}$. This estimate was used to perform the optimization in the **APPLY** routine as discussed following Theorem 4.6. The use of these modified sparse approximations reduced the cost of a call of **SOLVE** by approximately 30%.

Additionally, we applied an extra dropping in the spirit of Remark 3.6. When $\vec{A}_z^{(j)}$ was constructed by dropping all elements in $\vec{A}_z$ with modulus less than ε, we dropped all elements from the resulting $\mathbf{A}^{(j)}$ with modulus less than ε. We observed that this additional dropping hardly influences the accuracies of the approximate matrix vector multiplications, whereas, for $n > 1$, the cost of calls of **SOLVE** were reduced by approximately a factor 2.

For our convenience, we took as right-hand side $f = 1$. In this case, the solution u is the restriction to $(0,1)^n$ of $\tilde{u}$ solving $-\triangle \tilde{u} = 1$ on $\tilde{\Omega} = (0,2)^n$, $\tilde{u} = 0$ on $\partial\tilde{\Omega}$. For any $\ell \in \mathbb{N}_0$, there exists one $\lambda \in \nabla_z$ with $|\lambda| = \ell$ and $\int_0^1 \psi_\lambda^{(z)} \neq 0$, and so for any $\boldsymbol{\ell} \in \mathbb{N}_0^n$, there exists one $\boldsymbol{\lambda} \in \nabla$ with $|\boldsymbol{\lambda}| = \boldsymbol{\ell}$ and $\mathbf{f}_{\boldsymbol{\lambda}} = \int_{(0,1)^n} \boldsymbol{\psi}_{\boldsymbol{\lambda}} \neq 0$. Based on the estimate $|\mathbf{f}_{\boldsymbol{\lambda}}| \lesssim 2^{-(\|\boldsymbol{\lambda}\|_\infty + \frac{1}{2}\|\boldsymbol{\lambda}\|_1)}$, we considered optimized sparse grid approximations $[\mathbf{f}_{\boldsymbol{\lambda}}]_{\boldsymbol{\lambda} \in \nabla_\ell^{(\text{optsp})}}$ with $\nabla_\ell^{(\text{optsp})}$ from (5.4) with $d = 0$. We simply estimated the error in this approximation by the $\ell_2(\nabla)$-norm of $[\mathbf{f}_{\boldsymbol{\lambda}}]_{\boldsymbol{\lambda} \in \nabla_{\ell+1}^{(\text{optsp})} \setminus \nabla_\ell^{(\text{optsp})}}$, and took the smallest ℓ for which this norm is less than the given tolerance. Using that $\mathbf{f}$ is exceptionally sparse, one easily verifies that $\mathbf{rhs}_s < \infty$ for any s (dependent on n).

Concerning the parameters in **SOLVE**, some testing indicated that the best results are obtained with approximately $\alpha = 0.6$, $\omega = 0.4$, $\gamma = 0.2$, and $\theta = 2\omega/(1+\omega)$. Not very surprisingly (cf. Remark 4.3), these values are actually outside the ranges for which the algorithm was proven to be optimal. For 'admissible' parameter values, however, we obtained similar approximate solutions although at higher cost.

Numerical results obtained with **SOLVE** are given in Fig. 4. For any n, we know that $\mathbf{u} \in \mathcal{A}_\infty^1$ (see Sect. 1). As shown by Theorem 5.1, as a consequence, the approximations $\widetilde{\mathbf{u}}_N$ produced by **SOLVE** also convergence with asymptotic rate 1. With $\tilde{\mathbf{f}} - \tilde{\mathbf{A}}\widetilde{\mathbf{u}}_N$ denoting the approximate residual of $\widetilde{\mathbf{u}}_N$ computed inside **SOLVE** with
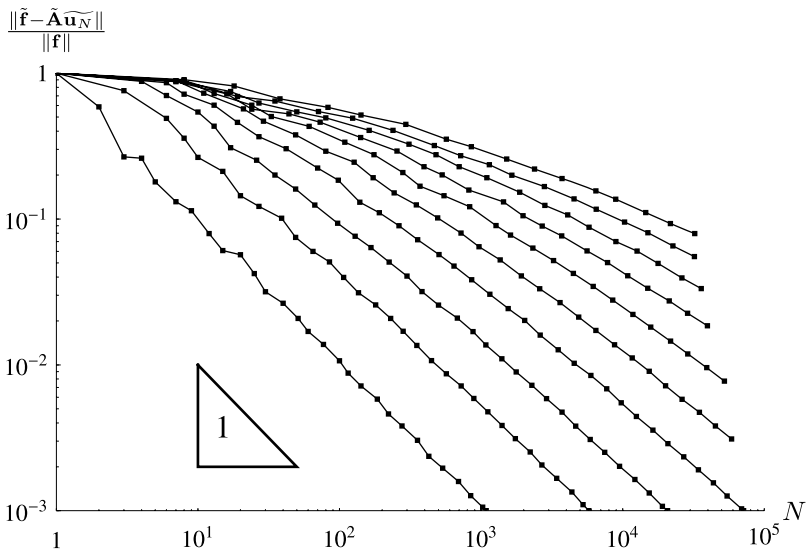
Fig. 4 $\frac{\|\tilde{\mathbf{f}} - \tilde{\mathbf{A}}\widetilde{\mathbf{u}_N}\|}{\|\mathbf{f}\|}$ for $n \in \{1, 2, \ldots, 10\}$, where $\tilde{\mathbf{f}} - \tilde{\mathbf{A}}\widetilde{\mathbf{u}_N}$ denotes an approximate residual, with a small relative error, of the approximate solution $\widetilde{\mathbf{u}_N}$ produced by **SOLVE** with support length N

small relative error, moreover we have $\sup_N N\|\tilde{\mathbf{f}} - \tilde{\mathbf{A}}\widetilde{\mathbf{u}_N}\| \eqsim \sup_N N\|\mathbf{u} - \widetilde{\mathbf{u}_N}\| \eqsim \sup_N N\|\mathbf{u} - \mathbf{u}_N\| \eqsim \sup_N N \inf_{\mathbf{v} \in \ell_2(\nabla)} \|u - \sum_{\lambda \in \nabla} v_\lambda \psi_\lambda\|$ $(\eqsim \|\mathbf{u}\|_{\mathcal{A}^1_\infty})$ with constant(s) that are *independent of n*. In view of Fig. 4, we conclude that for this family of solutions with right-hand sides $f = 1$, apparently $\|\mathbf{u}\|_{\mathcal{A}^1_\infty}/\|\mathbf{u}\|$ is increasing as a function of n, so that, as a consequence, the larger n is, the later the asymptotic rate shows up. Better results as a function of n will be obtained with coordinate directions with two Neumann boundaries, since for $f = 1$, the solution in those directions will be constant. For one Dirichlet and one Neumann boundary condition in each coordinate direction as we considered, for $n \geq 2$ the rate with (optimized) sparse grid approximation will be $\frac{1}{2} + \frac{1}{n}$ (see Sect. 1).

Finally, in Fig. 5, for $n \in \{1, 2, \ldots, 10\}$, we show the support size of the approximate residual computed inside **SOLVE** divided by n times the support size of the corresponding approximate solution. As predicted by Theorem 4.8, this quotient is bounded uniformly in n. The fact that it behaves like a constant shows that the cost of a call of **SOLVE** that produces an approximate solution of length N, being up to a constant multiple as good as a best N-term approximation, is proportional to n times N.

Remark 6.1 Our solution method did not make use of the symmetry in the right-hand function $f = 1$ and thus in the solution. It also did not use the fact that this f is a separable function so that its representation is a rank 1 tensor. In this sense we expect that the results we obtained are representative for general moderately smooth f that do not vanish at the Dirichlet boundary. Methods that do exploit the property of the right-hand side vector being a low rank tensor search for an approximate solution in
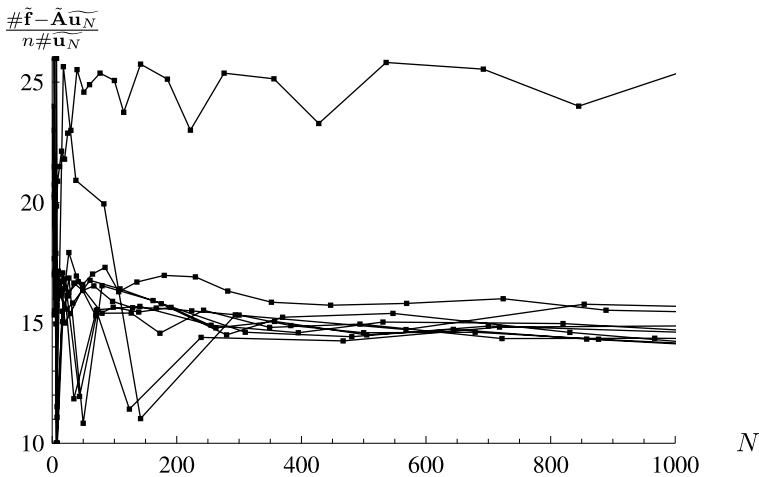
Fig. 5 Support size of the approximate residual computed inside **SOLVE** divided by n times the support size of the corresponding approximate solution for $n \in \{1, 2, \ldots, 10\}$. The *line* on *top* corresponds to $n = 1$

data sparse, low rank format, see [17, 18]. It would be interesting to see whether such methods can be combined with adaptivity.

References

1. Barinka, A.: Fast evaluation tools for adaptive wavelet schemes. PhD thesis, RTWH Aachen, March 2005
2. Bungartz, H.J., Griebel, M.: Sparse grids. Acta Numer. **13**, 147–269 (2004)
3. Cohen, A.: Numerical Analysis of Wavelet Methods. Elsevier, Amsterdam (2003)
4. Cohen, A., Dahmen, W., DeVore, R.: Adaptive wavelet methods for elliptic operator equations—convergence rates. Math. Comput. **70**, 27–75 (2001)
5. Cohen, A., Dahmen, W., DeVore, R.: Adaptive wavelet methods II—Beyond the elliptic case. Found. Comput. Math. **2**(3), 203–245 (2002)
6. Dahmen, W., Harbrecht, H., Schneider, R.: Compression techniques for boundary integral equations—optimal complexity estimates. SIAM J. Numer. Anal. **43**(6), 2251–2271 (2006)
7. Daubechies, I.: Orthonormal bases of compactly supported wavelets. Commun. Pure Appl. Math. **41**, 909–996 (1988)
8. Dauge, M.: Elliptic Boundary Value Problems on Corner Domain, Smoothness and Asymptotics of Solutions. Lecture Notes in Mathematics, vol. 1341. Springer, Berlin (1988)
9. DeVore, R.: Nonlinear approximation. Acta Numer. **7**, 51–150 (1998)
10. Dijkema, T.J.: Adaptive tensor product wavelet methods for solving PDEs. PhD thesis, Utrecht University (June 2009)
11. Donovan, G.C., Geronimo, J.S., Hardin, D.P.: Intertwining multiresolution analyses and the construction of piecewise-polynomial wavelets. SIAM J. Math. Anal. **27**(6), 1791–1815 (1996)
12. Donovan, G.C., Geronimo, J.S., Hardin, D.P.: Orthogonal polynomials and the construction of piecewise polynomial smooth wavelets. SIAM J. Math. Anal. **30**(5), 1029–1056 (1999)
13. Feynman, A.R., Hibbs, A.R.: Quantum Mechanics and Path Integrals. McGraw-Hill, New York (1965)

14. Freidlin, M.: Functional Integration and Partial Differential Equations. Annals of Mathematics Studies, vol. 109. Princeton University Press, Princeton (1985)
15. Gantumur, T., Stevenson, R.P.: Computation of differential operators in wavelet coordinates. Math. Comput. **75**, 697–709 (2006)
16. Gantumur, T., Harbrecht, H., Stevenson, R.P.: An optimal adaptive wavelet method without coarsening of the iterands. Math. Comput. **76**, 615–629 (2007)
17. Gavrilyuk, I.P., Hackbusch, W., Khoromskij, B.N.: Hierarchical tensor-product approximation to the inverse and related operators for high-dimensional elliptic problems. Computing **74**(2), 131–157 (2005)
18. Grasedyck, L.: Existence and computation of low Kronecker-rank approximations for large linear systems of tensor product structure. Computing **72**(3–4), 247–265 (2004)
19. Greengard, L., Rokhlin, V.: A fast algorithm for particle simulation. J. Comput. Phys. **73**, 325–348 (1987)
20. Griebel, M., Knapek, S.: Optimized tensor-product approximation spaces. Constr. Approx. **16**(4), 525–540 (2000)
21. Griebel, M., Oswald, P.: Tensor product type subspace splittings and multilevel iterative methods for anisotropic problems. Adv. Comput. Math. **4**(1–2), 171–206 (1995)
22. Hackbusch, W.: Iterative Solution of Large Sparse Systems of Equations. Applied Mathematical Sciences, vol. 95. Springer, New York (1994)
23. Hoang, V.H., Schwab, Ch.: High-dimensional finite elements for elliptic problems with multiple scales. SIAM J. Multiscale Model. Simul. **3**(1), 168–194 (2005)
24. Metselaar, A.: Handling wavelet expansions in numerical methods. PhD thesis, University of Twente (2002)
25. Nitsche, P.-A.: Sparse approximation of singularity functions. Constr. Approx. **21**(1), 63–81 (2005)
26. Nitsche, P.-A.: Best N-term approximation spaces for tensor product wavelet bases. Constr. Approx. **24**(1), 49–70 (2006)
27. Schwab, Ch., Stevenson, R.P.: Adaptive wavelet algorithms for elliptic PDEs on product domains. Math. Comput. **77**, 71–92 (2008)
28. Todor, R.A., Schwab, Ch.: Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients. IMA J. Numer. Anal. **27**(2), 232–261 (2007)
29. Yserentant, H.: Sparse grid spaces for the numerical solution of the electronic Schrödinger equation. Numer. Math. **101**(2), 381–389 (2005)
30. Zenger, Ch.: Sparse grids. In: Parallel Algorithms for Partial Differential Equations, Kiel, 1990. Notes Numer. Fluid Mech., vol. 31, pp. 241–251. Vieweg, Braunschweig (1991)