IMA Journal of Numerical Analysis (2011) Page 1 of 28
doi:10.1093/imanum/drn000

The adaptive tensor product wavelet scheme: Sparse matrices and the application to singularly perturbed problems

NABI CHEGINI† AND ROB STEVENSON‡
Korteweg-de Vries Institute for Mathematics, University of Amsterdam, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands.

[Received on 4 January 2011]

Locally supported biorthogonal wavelets are constructed on the unit interval with respect to which second order constant coefficient differential operators are sparse. As a result, the representation of second order differential operators on the hypercube with respect to the resulting tensor product wavelet coordinates is again sparse. The advantage of tensor product approximation is that it yields (nearly) dimension independent rates. An adaptive tensor product wavelet method is applied to solve various singularly perturbed boundary value problems. The numerical results indicate robustness with respect to the singular perturbations. For a two-dimensional model problem, this will be supported by theoretical results.

Keywords: Adaptive method, tensor product wavelets, optimal computational complexity, sparse matrices, singular perturbations.

1. Introduction

1.1 *Boundary value problem and its representation in tensor product wavelet coordinates*

With $\mathrm{I} := (0,1)$ and $\square := \mathrm{I}^n$ for some $n \in \mathbb{N}$, given $f \in H^{-1}(\square)$ we consider the problem of finding $u \in H^1_0(\square)$ such that

$$a(u,v) := \int_{\square} \sum_{i,j=1}^{n} a_{ij}\partial_i u \partial_j v + \sum_{i=1}^{n} b_i \partial_i uv + c_0 uv = f(v) \qquad (v \in H^1_0(\square)), \tag{1.1}$$

where the a_{ij}, b_i and $c_0 \geqslant 0$ are *constants*.

Here, in order that we can equip the corresponding function spaces with tensor product wavelet bases, we consider as domain the hypercube, being a prototype of a product domain. The application of piecewise multi-level tensor product bases on general domains via a domain decomposition approach is currently under investigation.

We consider convection-diffusion-reaction equations that may have an anisotropic diffusion tensor directed essentially along the coordinate axes, and a non-dominant convection part. The following result gives conditions under which such equations are well-posed.

LEMMA 1.1 For some constants $\delta, \Delta > 0$, and $K \geqslant 0$, and $c_1, \ldots, c_n > 0$, let

$$\delta \sum_{m=1}^{n} c_m \xi_m^2 \leqslant \sum_{i,j=1}^{n} a_{ij}\xi_i\xi_j \leqslant \Delta \sum_{m=1}^{n} c_m \xi_m^2 \quad (\xi \in \mathbb{R}^n) \tag{1.2}$$

†`Email`: N.GodarzvandChegini@uva.nl. This author was supported by the Netherlands Organisation for Scientific Research (NWO) under contract No. 613.000.902.
‡Corresponding author. `Email`: R.P.Stevenson@uva.nl

$$\sum_{i=1}^{n} \frac{b_i^2}{c_i} \leqslant K \max(c_0, \sum_{i=1}^{n} c_i). \tag{1.3}$$

Then A, defined by $(Au)(v) = a(u,v)$, defines a boundedly invertible operator between $H_0^1(\square)$, equipped with "energy-norm",

$$|||v||| := \sqrt{\int_\square c_0 v^2 + \sum_{m=1}^{n} c_m (\partial_m v)^2},$$

and its dual $H^{-1}(\square)$, equipped with the corresponding dual norm. The operator norms $|||A|||$, $|||A^{-1}|||$ can be bounded in terms of δ, Δ, K (and the Poincaré constant) only, thus *independently* of the a_{ij}, b_i, and c_i that satisfy the inequalities (1.2) and (1.3).

Proof. For $u \in H_0^1(\square)$, $a(u,u) \geqslant \min(1,\delta)|||u|||^2$. Since the spectral norm of $\left(\frac{a_{ij}}{\sqrt{c_i}\sqrt{c_j}}\right)_{ij}$ is $\leqslant \Delta$, we have $|\int_\square \sum_{i,j} a_{ij}\partial_i u \partial_j v| \leqslant \Delta \int_\square \sqrt{\sum_i c_i(\partial_i u)^2}\sqrt{\sum_j c_j(\partial_j v)^2} \leqslant \Delta \sqrt{\int_\square \sum_i c_i(\partial_i u)^2}\sqrt{\int_\square \sum_j c_j(\partial_j v)^2}$. Obviously $|\int_\square c_0 uv| \leqslant \sqrt{c_0}\|u\|_{L_2(\square)}\sqrt{c_0}\|v\|_{L_2(\square)}$. By Poincaré's inequality, for some constant $C > 0$, $\|\cdot\|^2_{L_2(\square)} \leqslant C\|\partial_i \cdot\|^2_{L_2(\square)}$ on $H_0^1(\square)$, or $\|\cdot\|^2_{L_2(\square)} \leqslant \frac{C}{\sum_{m=1}^n c_m}\sum_{i=1}^n c_i\|\partial_i \cdot\|^2_{L_2(\square)}$, and so $\|\cdot\|^2_{L_2(\square)} \leqslant \frac{\max(1,C)}{\max(c_0,\sum_{m=1}^n c_m)}|||\cdot|||^2$. With that, $\|\sum_i b_i \partial_i uv\|_{L_2(\square)} \leqslant \|v\|_{L_2(\square)}\sqrt{\sum_i \frac{b_i^2}{c_i}}\sqrt{\sum_i c_i\|\partial_i u\|^2_{L_2(\square)}} \leqslant \max(1,C)^{\frac{1}{2}}\sqrt{K}|||v|||||u|||$. Now the proof is completed by an application of the Lax-Milgram lemma. □

Using that

$$H_0^1(\square) = H_0^1(\mathrm{I}) \otimes L_2(\mathrm{I}) \otimes \cdots \otimes L_2(\mathrm{I}) \cap \quad \cdots \quad \cap L_2(\mathrm{I}) \otimes \cdots \otimes L_2(\mathrm{I}) \otimes H_0^1(\mathrm{I}),$$

a *Riesz basis* for $H_0^1(\square)$ can be constructed by tensorizing univariate Riesz bases of wavelet type. Indeed, let $\Psi = \{\psi_\lambda : \lambda \in \nabla\}$ be a normalized Riesz basis for $L_2(\mathrm{I})$ that, when re-normalized in $H^1(\mathrm{I})$, is a Riesz basis for $H_0^1(\mathrm{I})$. Then, when normalized in $|||\cdot|||$, $\Psi \otimes \cdots \otimes \Psi$ is a Riesz basis for $H_0^1(\square)$ (Griebel & Oswald (1995)). In particular, with the *Riesz constants*

$$\lambda_{L_2(\mathrm{I})} := \inf_{0\neq \mathbf{d}\in \ell_2(\nabla)} \frac{\|\sum_{\lambda\in\nabla} d_\lambda \psi_\lambda\|^2_{L_2(\mathrm{I})}}{\|\mathbf{d}\|^2_{\ell_2(\nabla)}}, \Lambda_{L_2(\mathrm{I})} := \sup_{0\neq \mathbf{d}\in \ell_2(\nabla)} \frac{\|\sum_{\lambda\in\nabla} d_\lambda \psi_\lambda\|^2_{L_2(\mathrm{I})}}{\|\mathbf{d}\|^2_{\ell_2(\nabla)}},$$

$$\lambda_{H_0^1(\mathrm{I})} := \inf_{0\neq \mathbf{d}\in \ell_2(\nabla)} \frac{|\sum_{\lambda\in\nabla} d_\lambda \psi_\lambda|^2_{H^1(\mathrm{I})}}{\sum_{\lambda\in\nabla}|d_\lambda|^2|\psi_\lambda|^2_{H^1(\mathrm{I})}}, \Lambda_{H_0^1(\mathrm{I})} := \sup_{0\neq \mathbf{d}\in \ell_2(\nabla)} \frac{|\sum_{\lambda\in\nabla} d_\lambda \psi_\lambda|^2_{H^1(\mathrm{I})}}{\sum_{\lambda\in\nabla}|d_\lambda|^2|\psi_\lambda|^2_{H^1(\mathrm{I})}},$$

we have

$$\min(\lambda_{L_2(\mathrm{I})}, \lambda_{H_0^1(\mathrm{I})})\lambda^{n-1}_{L_2(\mathrm{I})} \leqslant \frac{|||\sum_{\boldsymbol{\lambda}\in\nabla^n} d_{\boldsymbol{\lambda}} \psi_{\lambda_1}\otimes\cdots\otimes\psi_{\lambda_n}|||^2}{\sum_{\boldsymbol{\lambda}\in\nabla^n}|d_{\boldsymbol{\lambda}}|^2|||\psi_{\lambda_1}\otimes\cdots\otimes\psi_{\lambda_n}|||^2} \leqslant \max(\Lambda_{L_2(\mathrm{I})}, \Lambda_{H_0^1(\mathrm{I})})\Lambda^{n-1}_{L_2(\mathrm{I})} \tag{1.4}$$

($\mathbf{d} \in \ell_2(\nabla^n)$) (Dijkema *et al.* (2009)).

REMARK 1.1 This result shows that the Riesz constants of $\Psi\otimes\cdots\otimes\Psi$, normalized in $|||\cdot|||$, with respect to $H_0^1(\square)$ equipped with $|||\cdot|||$ are even independent of the space dimension n if (and actually only if) Ψ is an *orthonormal* basis for $L_2(\mathrm{I})$. In the current paper we will not consider such univariate wavelets.

REMARK 1.2 Actually, here and in the following, we can allow one or more c_i to be zero (when the corresponding b_i are zero as well) as long as $c_0 + \sum_{m=1}^n c_m > 0$. In that case $H_0^1(\square)$ should be read as the completion of $C_0^\infty(\square)$ with respect to the energy-norm (cf. Hochmuth (2001)).

Let us denote the normalized tensor product basis for $H_0^1(\square)$ as

$$\Psi := \{\boldsymbol{\psi}_{\boldsymbol{\lambda}} := \otimes_{m=1}^n \psi_{\lambda_m} / |||\otimes_{m=1}^n \psi_{\lambda_m}||| : \boldsymbol{\lambda} \in \boldsymbol{\nabla} := \nabla^n\}.$$

Note that $|||\otimes_{m=1}^n \psi_{\lambda_m}||| = \sqrt{c_0 + \sum_{m=1}^n c_m |\psi_{\lambda_m}|^2_{H^1(\mathrm{I})}}$. By writing

$$u = \mathbf{u}^\top \Psi := \sum_{\boldsymbol{\lambda}\in\boldsymbol{\nabla}} \mathbf{u}_{\boldsymbol{\lambda}} \boldsymbol{\psi}_{\boldsymbol{\lambda}},$$

and with

$$\mathbf{f} := [f(\boldsymbol{\psi}_{\boldsymbol{\lambda}})]_{\boldsymbol{\lambda}\in\boldsymbol{\nabla}},$$

an equivalent formulation of (1.1) is

$$\mathbf{A}\mathbf{u} = \mathbf{f}. \tag{1.5}$$

where $\mathbf{A} = [a(\boldsymbol{\psi}_{\boldsymbol{\mu}}, \boldsymbol{\psi}_{\boldsymbol{\lambda}})]_{\boldsymbol{\lambda},\boldsymbol{\mu}\in\boldsymbol{\nabla}}$ is the bi-infinite stiffness matrix of $a(\cdot,\cdot)$ with respect to Ψ.

Since $A : H_0^1(\square) \to H^{-1}(\square)$ is boundedly invertible, and Ψ is a Riesz basis for $H_0^1(\square)$, we have that $\mathbf{A} : \ell_2(\boldsymbol{\nabla}) \to \ell_2(\boldsymbol{\nabla})$ is boundedly invertible. In particular, from (1.4) we infer that

$$\|\mathbf{A}\| \leqslant |||A||| \max(\Lambda_{L_2(\mathrm{I})}, \Lambda_{H_0^1(\mathrm{I})}) \Lambda_{L_2(\mathrm{I})}^{n-1}, \ \|\mathbf{A}^{-1}\| \leqslant |||A^{-1}||| (\min(\lambda_{L_2(\mathrm{I})}, \lambda_{H_0^1(\mathrm{I})}) \lambda_{L_2(\mathrm{I})}^{n-1})^{-1}. \tag{1.6}$$

1.2 *Linear and nonlinear multi-level tensor product approximation*

The advantage of the representation of u in tensor product wavelet coordinates is that highly efficient approximations exist. Let $d \geqslant 2$ denote the order of the univariate wavelets. It is well known that if u has mixed derivatives of a sufficiently high order in $L_2(\square)$, then it can be approximated from the span of N tensor product wavelets with indices in the so-called optimized *sparse-grid* index set, such that the error in $H_0^1(\square)$ (with standard norm) is $\mathscr{O}(N^{-(d-1)})$ (Griebel & Knapek (2000); Bungartz & Griebel (2004)). So different than with non-tensor product approximation of order d, where one can expect at best a rate $(d-1)/n$, the obtained rate $d-1$ does *not deteriorate* with increasing n, i.e., the so-called *curse of dimensionality* is avoided.

For u being the solution of an elliptic PDE, generally the required boundedness of its mixed derivatives *cannot* be expected. For the Poisson problem, it can be enforced by requiring that the right-hand side f vanishes to a sufficiently high order at the non-smooth parts of $\partial\square$. For generic smooth f, however, the aforementioned rate $d-1$ reduces to $\frac{1}{2} + \frac{1}{n}$ (Dauge & Stevenson, 2010, Sect. 6). Yet, as is shown in Dauge & Stevenson (2010), the best possible rate $d-1$ can be retrieved by a proper modification of the optimized sparse grid index sets involving *local refinements* towards the boundary. This result was shown under the assumption that the corresponding *dual wavelets* are *uniformly local*.

The approximation results mentioned so far concern *linear approximation*. If u can be approximated from the spans of a priori selected sets of wavelets with rate s – that, as we have seen, can be as large as $(d-1)$ – then, obviously, it can be approximated within a tolerance $\mathscr{O}(N^{-s})$ by the best approximation from the span of the best N wavelets, dependent on u. Due to the Riesz basis property, the latter is equivalent, *only dependent on* $\kappa(\mathbf{A})$, to the property that the representation $\mathbf{u}$ of u with respect to Ψ can be approximated in $\ell_2(\boldsymbol{\nabla})$ within tolerance $\mathscr{O}(N^{-s})$ by the best possible vector of length N, known as a *best N-term approximation* for $\mathbf{u}$, and denoted as $\mathbf{u}_N$.

For $s < d$, the class of functions that can be approximated within a tolerance $\mathscr{O}(N^{-s})$ by a linear combination of best N tensor product wavelets has been shown to be equal to the intersection of certain tensor product of Besov spaces. We refer to Nitsche (2006); Sickel & Ullrich (2009).

In this paper, we study families of *singularly perturbed* boundary value problems. In particular, we consider the families where $a(u,v)$ is $\int_\square \eta^2 \nabla u \cdot \nabla v + uv$, $\int_\square \eta^2 \partial_1 u \partial_1 v + \sum_{i=2}^n \partial_i u \partial_i v$, or $\int_\square \eta^2 \nabla u \cdot \nabla v - \eta(\partial_1 u)v + uv$. The solutions of these problems have η-dependent boundary and corners layers that we expect can be efficiently approximated by tensor product approximation.

For the first family in two space dimensions, a splitting of the solution into a smooth part and edge and corners layers will be used to derive linear approximation rates (and with that upper bounds for the nonlinear rates) in $H_0^1(\square)$ equipped with the (η-dependent) energy norm $|||\cdot|||$. The rates will be robust with respect to η. Numerical results will indicate that the same holds true for the other two families.

1.3 *The adaptive solution of the boundary value problem*

So far we discussed (best) approximation of u in the energy-norm $|||\cdot|||$ from $\operatorname{span}\Psi$ or, in view of (1.4), equivalently, that of $\mathbf{u}$ in $\ell_2(\boldsymbol{\nabla})$ by finitely supported vectors. Both u and $\mathbf{u}$, however, are only implicitly given as solutions of (1.1) and (1.5), respectively. In Cohen *et al.* (2001, 2002), adaptive wavelet schemes were developed for solving (1.5) that, under some conditions, are *quasi-optimal*, meaning that they solve (1.5) with an $\ell_2(\boldsymbol{\nabla})$-rate equal to that of best N-term approximations, in linear complexity. We consider a modified version proposed in Gantumur *et al.* (2007) of the Adaptive Wavelet-Galerkin Method from Cohen *et al.* (2001), which avoids the so-called *coarsening* step.

For ease of presentation, let us assume that $\mathbf{A} = \mathbf{A}^T > 0$, i.e., that the b_i in (1.1) are zero. Otherwise, the scheme can be applied to the normal equations $\mathbf{A}^\top \mathbf{A}\mathbf{u} = \mathbf{A}^\top \mathbf{f}$, in which case in the following discussion, $\mathbf{A}$ and $\mathbf{f}$ should be read as $\mathbf{A}^\top \mathbf{A}$ and $\mathbf{A}^\top \mathbf{f}$ on all places.

Given a finite set $\Lambda \subset \boldsymbol{\nabla}$, let $\mathbf{I}_\Lambda : \ell_2(\Lambda) \to \ell_2(\boldsymbol{\nabla})$ denote the trivial embedding, so that its adjoint $\mathbf{P}_\Lambda : \ell_2(\boldsymbol{\nabla}) \to \ell_2(\Lambda)$ is the restriction of a vector to its indices in Λ. With $\mathbf{A}_\Lambda := \mathbf{P}_\Lambda \mathbf{A} \mathbf{I}_\Lambda$ and $\mathbf{f}_\Lambda := \mathbf{P}_\Lambda \mathbf{f}$, the solution of $\mathbf{A}_\Lambda \mathbf{u}_\Lambda = \mathbf{f}_\Lambda$ is known as the Galerkin approximation to $\mathbf{u}$ from $\ell_2(\Lambda)$. For some sufficiently small parameter $\mu \in (0,1)$, the idealized Adaptive Wavelet-Galerkin Method reads as follows:

$\Lambda_0 := \emptyset$, $\mathbf{u}_{\Lambda_0} := 0$,
`for` $i = 1,2,\ldots$ `do`
  Find a $\Lambda_{i+1} \supset \Lambda_i$ *such that* $\|\mathbf{P}_{\Lambda_{i+1}}(\mathbf{f} - \mathbf{A}\mathbf{u}_{\Lambda_i})\| \geqslant \mu \|\mathbf{f} - \mathbf{A}\mathbf{u}_{\Lambda_i}\|$, *and such that, up to some absolute multiple,* $\#(\Lambda_{i+1} \setminus \Lambda_i)$ *is minimal among all such* Λ_{i+1}.
  Solve $\mathbf{A}_{\Lambda_{i+1}} \mathbf{u}_{\Lambda_{i+1}} = \mathbf{f}_{\Lambda_{i+1}}$.
`enddo`

Note that in this scheme the residual $\mathbf{f} - \mathbf{A}\mathbf{u}_{\Lambda_i}$, being zero on Λ_i, plays the role of an a posteriori error estimator to guide a proper expansion of the set Λ_i.

Generally, the above scheme cannot be performed exactly. First of all, generally $\mathbf{f}$ will be infinitely supported and thus has to be approximated. Secondly, generally any column of $\mathbf{A}$ has infinitely many non-zeros. Thanks to the properties of wavelets, however, as being smooth and having vanishing moments, the sizes of the entries of $\mathbf{A}$ decay rapidly away from the diagonal. This property has been used to design an adaptive approximate matrix-vector multiplication routine **APPLY** in which, dependent to some prescribed overall tolerance, the accuracy with which any column is approximated increases with the modulus of the corresponding entry in the vector. This **APPLY** routine is used both for the approximate computation of the residual $\mathbf{f} - \mathbf{A}\mathbf{u}_{\Lambda_i}$, and for the repeated approximate multiplications with $\mathbf{A}_{\Lambda_{i+1}}$ for the iterative solving of the Galerkin problem $\mathbf{A}_{\Lambda_{i+1}} \mathbf{u}_{\Lambda_{i+1}} = \mathbf{f}_{\Lambda_{i+1}}$. Concerning the latter, note that generally the number of non-zero entries in $\mathbf{A}_{\Lambda_{i+1}}$ is not of the order of $\#\Lambda_{i+1}$.

Near-sparsity of $\mathbf{A}$ can be quantified by the concept of s^*-computability. It was shown that if the value of s^* is larger than the rate s of best N-term approximations for $\mathbf{u}$, and (quasi-) best N-term

approximations for $\mathbf{f}$ can be computed in linear complexity, then the approximations produced by the adaptive scheme converge with the best possible rate s in linear complexity (cf. Stevenson (2009)).

In the specific setting of tensor product wavelets, for uniformly local, sufficiently smooth univariate wavelets that have sufficiently many vanishing moments, in Schwab & Stevenson (2008) it was demonstrated that s^* exceeds the best possible rate s that can be expected, so that the practical adaptive wavelet scheme is quasi-optimal. This holds also true for differential operators with non-constant, but sufficiently smooth coefficients.

Practical results with the Adaptive (Tensor Product) Wavelet-Galerkin Method were reported on in Dijkema *et al.* (2009). There L_2-orthonormal univariate wavelets were used to arrive at a method that is quasi-optimal, uniformly in the space dimension n (cf. Remark 1.1).

1.4 *Custom designed wavelets that lead to sparse stiffness matrices*

Although the adaptive wavelet scheme described above is quasi-optimal, our practical experience with implementing the method learned us that quantitatively the application of the **APPLY** routine is very demanding, where this routine is also not easy to implement. In the specific setting of tensor product wavelets, the bi-infinite stiffness matrix corresponding to (1.1) reads as

$$\mathbf{A} = \mathbf{D}^{-1}\Big[\sum_{i,j=1}^{n} a_{ij}\bigotimes_{m=1}^{n}\vec{Z}(\delta_{im},\delta_{jm}) + \sum_{i=1}^{n} b_i\bigotimes_{m=1}^{n}\vec{Z}(\delta_{im},0) + c_0\bigotimes_{m=1}^{n}\vec{Z}(0,0)\Big]\mathbf{D}^{-1},$$

where $\mathbf{D} := \operatorname{diag}\big[\|\|\otimes_{m=1}^{n}\psi_{\lambda_m}\|\| : \boldsymbol{\lambda}\in\boldsymbol{\nabla}\big]$,

$$\vec{Z}(s,t) := \begin{cases} \vec{M} & \text{when } (s,t)=(0,0)\\ \vec{A} & \text{when } (s,t)=(1,1)\\ \vec{G} & \text{when } (s,t)=(1,0)\\ -\vec{G}^{\top} & \text{when } (s,t)=(0,1)\end{cases}$$

and

$$\vec{A} := \left[\int_{\mathrm{I}}\dot{\psi}_\mu\dot{\psi}_\lambda\right]_{\lambda,\mu\in\nabla},\quad \vec{M} := \left[\int_{\mathrm{I}}\psi_\mu\psi_\lambda\right]_{\lambda,\mu\in\nabla}, \text{and}\quad \vec{G} := \left[\int_{\mathrm{I}}\dot{\psi}_\mu\psi_\lambda\right]_{\lambda,\mu\in\nabla}.$$

Here, and on other places, a (double) "dot" on top of a univariate function denotes its (second) derivative. With a (double) "dot" on top of a linear space of univariate functions we will denote the linear space of (second) derivatives of these functions.

The difficulties with the **APPLY** routine led us in Dijkema & Stevenson (2010) to construct a univariate wavelet basis that generates Riesz bases for a range of Sobolev spaces, including $L_2(\mathrm{I})$ and $H^1_0(\mathrm{I})$, and which has the special property to give rise to matrices $\vec{M}$, $\vec{A}$, and $\vec{G}$ that all are truly *sparse*. As a result, the representation $\mathbf{A}$ of (1.1) with respect to the n-fold tensor product wavelet basis is sparse. So in this case, $\mathbf{A}$ can be applied *exactly* to any (finitely supported) vector at a cost that is linear in its support length, and there is no need for a complicated, and quantitatively demanding approximate adaptive matrix-vector multiplication routine **APPLY**.

The wavelets constructed in Dijkema & Stevenson (2010) have globally supported duals. A number of theoretical results require uniformly local dual wavelets, although usually, with some more efforts, they can be extended to duals that decay sufficiently fast. We mention the characterization of approximation classes as (intersections of tensor products of) Besov spaces, and the proof from Dauge & Stevenson (2010) that these approximation classes contain (intersections of tensor products of) certain

weighted Sobolev spaces. Furthermore, for a number of applications, uniformly local dual wavelets are either necessary or in any case convenient. We think of the generalization of the adaptive wavelet scheme to nonlinear equations, and the application of wavelets in data analysis and image compression.

In view of these observations, in this paper we construct wavelets that share the properties with those from Dijkema & Stevenson (2010), i.e., they will yield matrices $\vec{M}$, $\vec{A}$, and $\vec{G}$ that all are truly *sparse*, and that have the additional property that their duals are uniformly local. Moreover, although the order of the wavelets here will be 5 with 5 vanishing moments, vs. 4 with 4 vanishing moments in Dijkema & Stevenson (2010), their condition numbers with respect to L_2 and H^1 norms will even be slightly better.

REMARK 1.3 True sparsity of $\mathbf{A}$ is only valid for differential operators with constant coefficients. This was the only reason to consider constant coefficients from the start in (1.1). With the exception of sparsity of $\mathbf{A}$, however, all other results in this paper directly generalize to variable, possibly sufficiently smooth coefficients. Concerning sparsity, for smooth, variable coefficients, the additional non-zeros outside the sparsity pattern of a constant coefficient operator will be orders of magnitude smaller than those inside this pattern, depending on the levels of the wavelets involved. For the residual computation inside the adaptive wavelet scheme, which is the quantitatively most demanding part, it can be envisaged that they can be ignored, possibly apart from those corresponding to some coarsest levels. A precise analysis of this phenomenon is outside the scope of this work.

The remainder of this paper is organized as follows: In Sect. 2 it is shown that our desire to have sparse stiffness matrices can essentially only be fulfilled by continuously differentiable wavelets. In Sect. 3, the general framework is recalled how to construct biorthogonal, uniformly local wavelets that generate Riesz bases for a scale of Sobolev spaces. In Sect. 4, this framework is used to construct a wavelet basis that has the additional property to give rise to sparse stiffness matrices. The quantitative stability of the basis is investigated in Sect. 5. In Sect. 6, the Adaptive Wavelet-Galerkin Method is described. In Sect. 7, approximation rates are derived for a model reaction-diffusion equation in two space dimensions. Finally, in Sect. 8 numerical results are presented of the Adaptive Wavelet-Galerkin method applied to various singularly perturbed problems.

For completeness, here and in the remainder of this work, with $C \lesssim D$ we will mean that C can be bounded by a multiple of D, independently of parameters on which C and D may depend, with the exception of δ, Δ and K from Lemma 1.1, and the space dimension n (cf. remark 1.1). Obviously, $C \gtrsim D$ is defined as $D \lesssim C$, and $C \eqsim D$ as $C \lesssim D$ and $C \gtrsim D$.

2. A minimal smoothness requirement

We would like to find a collection of univariate wavelets $\Psi = \{\psi_\lambda : \lambda \in \nabla\}$ such that, with $|\lambda| \in \mathbb{N}_0$ denoting the level of ψ_λ or that of λ, in any case the following conditions are satisfied:

(1) $\operatorname{diam}\operatorname{supp}\psi_\lambda \lesssim 2^{-|\lambda|}$,

(2) $\sup_{j,k\in\mathbb{N}_0} \#\{|\lambda| = j : [k2^{-j},(k+1)2^{-j}] \cap \operatorname{supp}\psi_\lambda \neq \emptyset\} < \infty$,

(3) Ψ is Riesz basis for $L_2(\mathrm{I})$,

(4) $\{\psi_\lambda / \|\dot{\psi}_\lambda\|_{L_2(\mathrm{I})} : \lambda \in \nabla\}$ is a Riesz basis for $H^1_0(\mathrm{I})$,

(5) $\int_{\mathrm{I}} \dot{\psi}_\lambda \dot{\psi}_\mu = 0$ when $||\lambda| - |\mu|| > M$,

where $M \in \mathbb{N}_0$ is some constant, that later will be chosen to be 1. As a consequence, with respect to a level-wise partition of the wavelets, $\vec{A}$ will be block tridiagonal with, because of (1) and (2), sparse non-zero blocks. Note that under the assumptions (1) and (2), $\vec{A}$ is sparse if and only if (5) is valid.

The additional requirements that $\vec{M}$ and $\vec{G}$ be sparse play no role in this section.

We will refer to the properties (1) and (2) by saying that the wavelets are *(uniformly) local*. We will use this definition of locality also for other collections of functions, in particular for scaling functions.

The following proposition shows that (1)-(5) can only be satisfied by piecewise smooth wavelets when they are *globally continuously differentiable*.

PROPOSITION 2.1 (Dijkema & Stevenson, 2010, Prop.1) If, in addition to (1) – (4), each wavelet is piecewise smooth with bounded piecewise first and second derivatives, then (5) requires that they are globally C^1.

3. Biorthogonal multi-resolution analyses and wavelets

In order to construct wavelets that, properly scaled, generate Riesz bases for a range of Sobolev spaces, in particular for $L_2(\mathrm{I})$ and $H^1_0(\mathrm{I})$ (cf. (3) and (4)), we will use the following well-known theorem (cf. Dahmen (1996); Dahmen & Stevenson (1999); Cohen (2003)).

THEOREM 3.1 (Biorthogonal space decompositions) Let

$$V_0 \subset V_1 \subset \cdots \subset L_2(\mathrm{I}), \qquad \tilde{V}_0 \subset \tilde{V}_1 \subset \cdots \subset L_2(\mathrm{I})$$

be sequences of primal and dual spaces such that

$$\dim V_j = \dim \tilde{V}_j < \infty \quad \text{and} \quad \alpha_j := \inf_{0\neq \tilde{v}_j \in \tilde{V}_j} \sup_{0\neq v_j \in V_j} \frac{|\langle \tilde{v}_j, v_j\rangle_{L_2(\mathrm{I})}|}{\|\tilde{v}_j\|_{L_2(\mathrm{I})}\|v_j\|_{L_2(\mathrm{I})}} \gtrsim 1. \tag{3.1}$$

In addition, for some $0 < \gamma < d$, let

$$\inf_{v_j \in V_j} \|v - v_j\|_{L_2(\mathrm{I})} \lesssim 2^{-jd}\|v\|_{\mathscr{H}^d(\mathrm{I})} \quad (v \in \mathscr{H}^d(\mathrm{I}))$$

(Jackson estimate), and

$$\|v_j\|_{\mathscr{H}^s(\mathrm{I})} \lesssim 2^{js}\|v_j\|_{L_2(\mathrm{I})} \quad (v_j \in V_j,\, s \in [0,\gamma))$$

(Bernstein estimate), where $\mathscr{H}^s(\mathrm{I})$ $(s \in [0,d])$ are Hilbert spaces such that $\mathscr{H}^d(\mathrm{I}) \hookrightarrow L_2(\mathrm{I})$ dense, and, for $s \in (0,d)$, $\mathscr{H}^s(\mathrm{I}) := [L_2(\mathrm{I}), \mathscr{H}^d(\mathrm{I})]_{s/d}$, and let similar estimates be valid at the dual side with $((V_j)_j, d, \gamma, \mathscr{H}^s(\mathrm{I}))$ reading as $((\tilde{V}_j)_j, \tilde{d}, \tilde{\gamma}, \tilde{\mathscr{H}}^s(\mathrm{I}))$.

Then, with $\Phi_0 = \{\phi_{0,k} \colon k \in I_0\}$ being a basis for V_0 (scaling functions) and $\Psi_j = \{\psi_{j,k} \colon k \in J_j\}$ $(j \in \mathbb{N})$ being uniform $L_2(\mathrm{I})$-Riesz bases for $V_j \cap \tilde{V}_{j-1}^{\perp_{L_2(\mathrm{I})}}$ (wavelets), for $s \in (-\tilde{\gamma}, \gamma)$ the collection

$$\Phi_0 \cup \cup_{j\in\mathbb{N}} 2^{-sj}\Psi_j$$

is a Riesz basis for $\mathscr{H}^s(\mathrm{I})$, where, for $s < 0$, $\mathscr{H}^s(\mathrm{I}) := (\tilde{\mathscr{H}}^{-s}(\mathrm{I}))'$.

In view of the notations introduced earlier, we denote (j,k) also as λ, where $|\lambda| = j$, $\phi_{0,k}$ as $\psi_{0,k}$ and $I_0 \cup \cup_{j\in\mathbb{N}} J_j$ as ∇.

The existence of the Riesz basis $\Psi := \Phi_0 \cup \cup_{j\in\mathbb{N}} \Psi_j$ for $L_2(\mathrm{I})$ as in Theorem 3.1 is equivalent to the existence of another, dual Riesz basis $\tilde{\Psi}$ for $L_2(\mathrm{I})$. Writing this basis correspondingly as $\tilde{\Phi}_0 \cup \cup_{j\in\mathbb{N}} \tilde{\Psi}_j$, $\tilde{\Phi}_0$ is a basis for $\tilde{V}_0$, and $\tilde{\Psi}_j$ for $\tilde{V}_j \cap V_{j-1}^{\perp_{L_2(\mathrm{I})}}$. For $s \in (-\gamma, \tilde{\gamma})$, the collection

$$\tilde{\Phi}_0 \cup \cup_{j\in\mathbb{N}} 2^{-sj} \tilde{\Psi}_j$$

is a Riesz basis for $\tilde{\mathscr{H}}^s(\mathrm{I})$.

The property (3.1) can be verified by an application of the following proposition (cf. e.g. Stevenson (2003)).

PROPOSITION 3.2 Let $(V_j)_j, (\tilde{V}_j)_j \subset L_2(\mathrm{I})$ be nested sequences of finite dimensional spaces, then (3.1) is equivalent to

- the existence of uniformly bounded (biorthogonal) projectors $Q_j : L_2(\mathrm{I}) \to V_j$ with $\mathrm{Im} Q_j = V_j$, $\mathrm{Im}(I - Q_j) = \tilde{V}_j^{\perp_{L_2(\mathrm{I})}}$, where $\|Q_j\|_{L_2(\mathrm{I})\to L_2(\mathrm{I})} = \alpha_j^{-1}$,
- the existence of biorthogonal, uniform $L_2(\mathrm{I})$-Riesz bases Φ_j and $\tilde{\Phi}_j$ for V_j and $\tilde{V}_j$, respectively.

We will verify (3.1) by constructing biorthogonal scaling functions as meant in the second item of Proposition 3.2. In particular, we will construct V_j and $\tilde{V}_j$ such that they can be equipped with uniformly *local* biorthogonal bases. With such bases at hand, under a mild additional condition, uniformly local primal *and* dual wavelets become available. This is shown in the following proposition. For a proof we refer to Carnicer *et al.* (1996); Stevenson (2003).

PROPOSITION 3.3 (a). Let the union of Φ_j and Ξ_{j+1} ("initial stable completion") be a uniform $L_2(\mathrm{I})$-Riesz basis for V_{j+1}, then

$$\Psi_{j+1} := \Xi_{j+1} - \langle \Xi_{j+1}, \tilde{\Phi}_j \rangle_{L_2(\mathrm{I})} \Phi_j$$

is a uniform $L_2(\mathrm{I})$-Riesz basis for $V_{j+1} \cap \tilde{V}_j^{\perp_{L_2(\mathrm{I})}}$.

(b). Writing $\begin{bmatrix} \Phi_j^\top & \Xi_{j+1}^\top \end{bmatrix} = \Phi_{j+1}^\top \check{\mathbf{M}}_j$ and $\check{\mathbf{M}}_j^{-1} = \begin{bmatrix} \mathbf{G}_{j,0} \\ \mathbf{G}_{j,1} \end{bmatrix}$, i.e., $\Phi_{j+1}^\top = \Phi_j^\top \mathbf{G}_{j,0} + \Xi_{j+1}^\top \mathbf{G}_{j,1}$, then

$$\tilde{\Psi}_{j+1}^\top = \tilde{\Phi}_{j+1}^\top \mathbf{G}_{j,1}^\top.$$

(c). If Φ_j, $\tilde{\Phi}_j$, and Ξ_{j+1} are uniformly local, then Ψ_{j+1} is uniformly local.

If $\tilde{\Phi}_{j+1}$ is uniformly local, and $\check{\mathbf{M}}_j^{-1}$ and thus $\mathbf{G}_{j,1}$ are uniformly local (i.e., entries corresponding to basis functions are zero whenever their distance is larger than some absolute multiple of 2^{-j}), then $\tilde{\Psi}_{j+1}$ is uniformly local.

4. Biorthogonal multi-resolution analyses that lead to sparse stiffness matrices

We shall select $(V_j)_j$, $(\tilde{V}_j)_j$ that satisfy the conditions of Theorem 3.1 for some $\gamma > 1$, where $\mathscr{H}^d(\mathrm{I}) = H_0^1(\mathrm{I}) \cap H^d(\mathrm{I})$. The pairs $(V_j, \tilde{V}_j)$ will be equipped with biorthogonal, uniformly local, uniform $L_2(\mathrm{I})$-Riesz bases Φ_j, $\tilde{\Phi}_j$. In addition, and that will turn out to be the main point, $(V_j)_j$, $(\tilde{V}_j)_j$ will be selected such that $V_j \subset C^1(\mathrm{I}) \cap H_0^1(\mathrm{I})$ and

$$V_j + \dot{V}_j + \ddot{V}_j \subset \tilde{V}_{j+1}. \tag{4.1}$$

Here, with $\dot{V}_j$ ($\ddot{V}_j$) we mean the linear space of (second) derivatives of functions in V_j.

Furthermore, a uniformly local "initial stable completion" Ξ_{j+1} will be constructed such that $\Phi_j \cup \Xi_{j+1}$ is a uniform $L_2(\mathrm{I})$-Riesz basis for V_{j+1}, and such that the basis transformation $\check{\mathbf{M}}_j^{-1}$ is uniformly local. As a result, we obtain a wavelet collection $\Psi = \Phi_0 \cup \cup_{j\in\mathbb{N}}\Psi_j$ that, properly scaled, generates Riesz bases for a range of Sobolev spaces, including $L_2(\mathrm{I})$ and $H_0^1(\mathrm{I})$. Furthermore, both Ψ_j and the corresponding duals $\tilde{\Psi}_j$ will be uniformly local.

As a consequence of (4.1), we have that for $|\mu| > |\lambda| + 1$,

$$\begin{aligned}
\langle \psi_\lambda, \psi_\mu \rangle_{L_2(\mathrm{I})} &= 0, \\
\langle \dot{\psi}_\lambda, \dot{\psi}_\mu \rangle_{L_2(\mathrm{I})} &= -\langle \ddot{\psi}_\lambda, \psi_\mu \rangle_{L_2(\mathrm{I})} = 0, \\
\langle \psi_\lambda, \dot{\psi}_\mu \rangle_{L_2(\mathrm{I})} &= -\langle \dot{\psi}_\lambda, \psi_\mu \rangle_{L_2(\mathrm{I})} = 0,
\end{aligned}$$

showing that the univariate system matrices $\vec{M}$, $\vec{A}$, and $\vec{G}$ are *sparse* as desired.

4.1 *A realization*

We do not have a general procedure to construct biorthogonal multi-resolution analyses that satisfy all aforementioned conditions, in particular (4.1), but we construct one particular instance. We take

$$V_j = \prod_{k=0}^{2^{j+1}-1} P_4(k2^{-(j+1)}, (k+1)2^{-(j+1)}) \cap C^1(\mathrm{I}) \cap H_0^1(\mathrm{I})$$

of dimension $5 \cdot 2^{j+1} - 2(2^{j+1}-1) - 2 = 3 \cdot 2^{j+1}$. Then

$$V_j + \dot{V}_j + \ddot{V}_j \subset Z_j := \prod_{k=0}^{2^{j+1}-1} P_4(k2^{-(j+1)}, (k+1)2^{-(j+1)})$$

of dimension $5 \cdot 2^{j+1}$. Following the idea of a so-called *intertwining multiresolution analysis* (Donovan *et al.* (1996)), we will select $\tilde{V}_{j+1}$ as the direct sum of Z_j and a subspace of Z_{j+1} of dimension 2^{j+1} so that $\dim \tilde{V}_{j+1} = \dim V_{j+1}$. Since $(Z_j)_j$ is nested, so is $(\tilde{V}_j)_j$, and Bernstein and Jackson estimates are satisfied at primal and dual side with parameters $d = \tilde{d} = 5$, $\gamma = \frac{5}{2}$, $\tilde{\gamma} = \frac{1}{2}$, where $\mathscr{H}^5(\mathrm{I}) = H_0^1(\mathrm{I}) \cap H^5(\mathrm{I})$ and $\tilde{\mathscr{H}}^5(\mathrm{I}) = H^5(\mathrm{I})$.

We start our construction of $(\tilde{V}_j)_j$ and that of biorthogonal scaling functions Φ_j and $\tilde{\Phi}_j$ on a *reference macro-element* $\hat{\mathrm{I}} = (-1,1)$. Since at each level, the functions at the primal side are piecewise polynomials with respect to a partition that is twice as fine as the corresponding partition at the dual side, this reference macro-element comprises two primal elements, and one dual element.

Let $H = \{-1,0,1\}$ and $L = \{-1,-\frac{1}{2},0,\frac{1}{2},1\}$. For $i \in H$ and $j \in L$, let the interpolatory basis functions $h_i, \ell_j \in P_4(-1,0) \times P_4(0,1) \cap C^1(\hat{\mathrm{I}})$ and $\tilde{\ell}_j \in P_4(\hat{\mathrm{I}})$ be defined by

$$\begin{aligned}
\dot{h}_i(\hat{\imath}) &= \delta_{i\hat{\imath}}, \qquad & \dot{\ell}_j(\hat{\imath}) &= 0, \\
h_i(\hat{\jmath}) &= 0, \qquad & \ell_j(\hat{\jmath}) &= \delta_{j\hat{\jmath}},
\end{aligned} \qquad \tilde{\ell}_j(\hat{\jmath}) = \delta_{j\hat{\jmath}}, \qquad (\hat{\imath} \in H, \hat{\jmath} \in L),$$

cf. Figure 1.

With the aim to construct biorthogonal collections, we will apply a number of transformations to these bases, and enrich the collection at the dual side with an additional function.

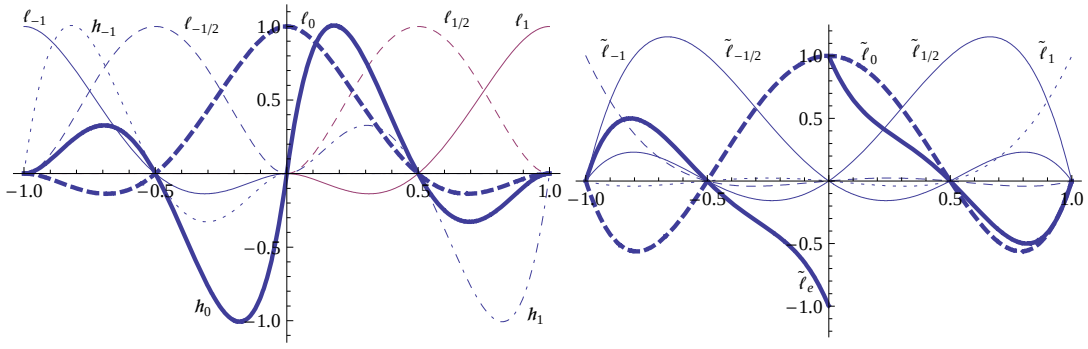

FIG. 1. Interpolatory basis of $P_4(-1,0)\times P_4(0,1)\cap C^1(\hat{\mathrm{I}})$ (left picture) and that of $P_4(\hat{\mathrm{I}})$ (right picture), the latter together with the additional, discontinuous function $\tilde{\ell}_e \in P_4(-1,0)\times P_4(0,1)$ defined in step 2. In this figure, the functions h_{-1}, h_0, and h_1 were multiplied by a factor 13, and the function $\tilde{\ell}_e$ with a factor $\frac{14}{9}$.

Step 1. We define $\hat{h}_{-1}$ and $\hat{\ell}_{-1}$ as linear combinations of h_{-1}, $\ell_{-\frac{1}{2}}$, h_0, ℓ_0, $\ell_{\frac{1}{2}}$ and ℓ_{-1}, $\ell_{-\frac{1}{2}}$, h_0, ℓ_0, $\ell_{\frac{1}{2}}$, respectively, such that

$$\langle \hat{h}_{-1}, \tilde{\ell}_{\hat{\jmath}} \rangle_{L_2(\hat{\mathrm{I}})} = \langle \hat{\ell}_{-1}, \tilde{\ell}_{\hat{\jmath}} \rangle_{L_2(\hat{\mathrm{I}})} = \tfrac{1}{2}\delta_{-1,\hat{\jmath}} \quad (\hat{\jmath} \in L).$$

We set $\hat{h}_1(x) = -\hat{h}_{-1}(-x)$ and $\hat{\ell}_1(x) = \hat{\ell}_{-1}(-x)$.

Step 2. We select $\tilde{\ell}_e$ from the 5 dimensional space $\{f \in P_4(-1,0)\times P_4(0,1) : f(-1) = f(-\frac{1}{2}) = f(0^-) + f(0^+) = f(\frac{1}{2}) = f(1) = 0\}$ such that

$$\langle \begin{bmatrix} \hat{h}_{-1} & \hat{\ell}_{-1} & \hat{h}_1 & \hat{\ell}_1 \end{bmatrix}^\top, \tilde{\ell}_e \rangle_{L_2(\hat{\mathrm{I}})} = 0.$$

Fixing some scaling of $\tilde{\ell}_e$, it has values $\frac{531}{1792}$, $-\frac{63}{256}$, $\frac{63}{256}$, $-\frac{531}{1792}$ in $-\frac{3}{4}$, $-\frac{1}{4}$, $\frac{1}{4}$, $\frac{3}{4}$, respectively, and $-\tilde{\ell}_e(0^-) = \tilde{\ell}_e(0^+) = \frac{9}{14}$, cf. Figure 1.

Step 3. We set

$$\begin{bmatrix} \tilde{\hat{\ell}}_{-\frac{1}{2}} \\ \tilde{\hat{\ell}}_e \\ \tilde{\hat{\ell}}_0 \\ \tilde{\hat{\ell}}_{\frac{1}{2}} \end{bmatrix} = \left\langle \begin{bmatrix} \tilde{\ell}_{-\frac{1}{2}} \\ \tilde{\ell}_e \\ \tilde{\ell}_0 \\ \tilde{\ell}_{\frac{1}{2}} \end{bmatrix}, \begin{bmatrix} \ell_{-\frac{1}{2}} \\ h_0 \\ \ell_0 \\ \ell_{\frac{1}{2}} \end{bmatrix} \right\rangle_{L_2(\hat{\mathrm{I}})}^{-1} \begin{bmatrix} \tilde{\ell}_{-\frac{1}{2}} \\ \tilde{\ell}_e \\ \tilde{\ell}_0 \\ \tilde{\ell}_{\frac{1}{2}} \end{bmatrix}.$$

Step 4. For $j \in \{-1,1\}$, we define

$$\tilde{\hat{\ell}}_j = \tilde{\ell}_j - \langle \tilde{\ell}_j, \ell_{-\frac{1}{2}} \rangle_{L_2(\hat{\mathrm{I}})} \tilde{\hat{\ell}}_{-\frac{1}{2}} - \langle \tilde{\ell}_j, h_0 \rangle_{L_2(\hat{\mathrm{I}})} \tilde{\hat{\ell}}_e - \langle \tilde{\ell}_j, \ell_0 \rangle_{L_2(\hat{\mathrm{I}})} \tilde{\hat{\ell}}_0 - \langle \tilde{\ell}_j, \ell_{\frac{1}{2}} \rangle_{L_2(\hat{\mathrm{I}})} \tilde{\hat{\ell}}_{\frac{1}{2}}$$

Now setting

$$\hat{\Phi} := \begin{bmatrix} \hat{h}_{-1} & \hat{\ell}_{-1} & \ell_{-\frac{1}{2}} & h_0 & \ell_0 & \ell_{\frac{1}{2}} & \hat{h}_1 & \hat{\ell}_1 \end{bmatrix}^\top,$$
$$\tilde{\hat{\Phi}} := \begin{bmatrix} \tilde{\hat{\ell}}_{-1} & \tilde{\hat{\ell}}_{-\frac{1}{2}} & \tilde{\hat{\ell}}_e & \tilde{\hat{\ell}}_0 & \tilde{\hat{\ell}}_{\frac{1}{2}} & \tilde{\hat{\ell}}_1 \end{bmatrix}^\top,$$

from the definitions in Steps 1-4, we have

$$\langle \hat{\boldsymbol{\Phi}}, \tilde{\hat{\boldsymbol{\Phi}}} \rangle_{L_2(\hat{\mathrm{I}})} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix} \tag{4.2}$$

The collections $\hat{\boldsymbol{\Phi}}$ and $\tilde{\hat{\boldsymbol{\Phi}}}$ are illustrated in Figure 2, and their coefficients in terms of the interpolatory basis or in terms of $\{\tilde{\ell}_{-1}, \tilde{\ell}_{-\frac{1}{2}}, \tilde{\ell}_e, \tilde{\ell}_0, \tilde{\ell}_{\frac{1}{2}}, \tilde{\ell}_1\}$ can be found in Tables 1 and 2, respectively.

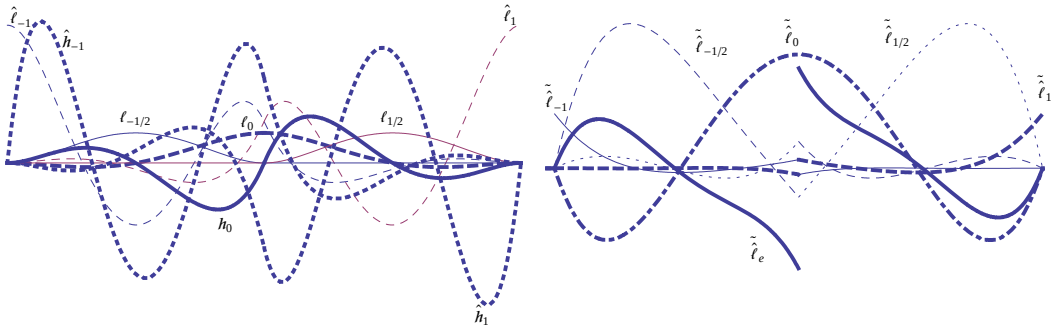

FIG. 2. Biorthogonal collections $\hat{\boldsymbol{\Phi}}$ and $\tilde{\hat{\boldsymbol{\Phi}}}$ on the reference element, with h_0 and $\tilde{\ell}_e$ being multiplied by a factor 20 or $\frac{1}{20}$, respectively.

	h_{-1}	ℓ_{-1}	$\ell_{-\frac{1}{2}}$	h_0	ℓ_0	$\ell_{\frac{1}{2}}$	h_1	ℓ_1
$\hat{h}_{-1}$	$\frac{165}{2}$	0	$-\frac{945}{256}$	$-\frac{285}{8}$	$\frac{45}{16}$	$-\frac{105}{256}$	0	0
$\hat{\ell}_{-1}$	0	$\frac{55}{12}$	$-\frac{6335}{3072}$	$-\frac{295}{16}$	$\frac{265}{192}$	$-\frac{655}{3072}$	0	0
$\ell_{-\frac{1}{2}}$	0	0	1	0	0	0	0	0
h_0	0	0	0	1	0	0	0	0
ℓ_0	0	0	0	0	1	0	0	0
$\ell_{\frac{1}{2}}$	0	0	0	0	0	1	0	0
$\hat{h}_1$	0	0	$\frac{105}{256}$	$-\frac{285}{8}$	$-\frac{45}{16}$	$\frac{945}{256}$	$\frac{165}{2}$	0
$\hat{\ell}_1$	0	0	$-\frac{655}{3072}$	$\frac{295}{16}$	$\frac{265}{192}$	$-\frac{6335}{3072}$	0	$\frac{55}{12}$

Table 1. Coefficients of $\hat{\boldsymbol{\Phi}}$.

Having defined primal and dual collections $\hat{\boldsymbol{\Phi}}$ and $\tilde{\hat{\boldsymbol{\Phi}}}$ on the reference macro-element, the collections $\boldsymbol{\Phi}_j$ and $\tilde{\boldsymbol{\Phi}}_j$ are assembled in the way known from finite elements. First, the collections $\hat{\boldsymbol{\Phi}}$ and $\tilde{\hat{\boldsymbol{\Phi}}}$ are

	$\tilde{\ell}_{-1}$	$\tilde{\ell}_{-\frac{1}{2}}$	$\tilde{\ell}_e$	$\tilde{\ell}_0$	$\tilde{\ell}_{\frac{1}{2}}$	$\tilde{\ell}_1$
$\hat{\tilde{\ell}}_{-1}$	1	$-\frac{27}{352}$	$\frac{7}{33}$	$\frac{1}{44}$	$\frac{1}{352}$	0
$\hat{\tilde{\ell}}_{-\frac{1}{2}}$	0	$\frac{8775}{4096}$	$\frac{35}{44}$	$-\frac{105}{5632}$	$-\frac{1395}{45056}$	0
$\hat{\tilde{\ell}}_e$	0	$-\frac{45}{88}$	$\frac{630}{11}$	0	$\frac{45}{88}$	0
$\hat{\tilde{\ell}}_0$	0	$-\frac{315}{2816}$	$\frac{1715}{528}$	$\frac{735}{352}$	$-\frac{315}{2816}$	0
$\hat{\tilde{\ell}}_{\frac{1}{2}}$	0	$-\frac{1395}{45056}$	$\frac{35}{44}$	$-\frac{105}{5632}$	$\frac{8775}{4096}$	0
$\hat{\tilde{\ell}}_1$	0	$\frac{1}{352}$	$\frac{7}{33}$	$\frac{1}{44}$	$-\frac{27}{352}$	1

Table 2. Coefficients of $\tilde{\hat{\Phi}}$.

lifted to any macro-element $[k2^{-(j+1)},(k+2)2^{-(j+1)}]$, and multiplied by $2^{(j+1)/2}$ to compensate for the change in length.

At the primal side, functions at an interface between two macro-elements that correspond to the same degree of freedom are connected, i.e., $\hat{h}_1$ from the left is connected to $\hat{h}_{-1}$ from the right, and $\hat{\ell}_1$ from the left is connected to $\hat{\ell}_{-1}$ from the right. In view of the boundary conditions, at the left boundary the function $\hat{\ell}_{-1}$ is dropped and at the right boundary so is the function $\hat{\ell}_1$.

In view of the discontinuity between elements, at the dual side no degrees of freedom are identified. Instead a simple basis transformation is applied. At each internal interface, the pair of functions consisting of $\hat{\ell}_1$ from the left and $\hat{\ell}_{-1}$ from the right are replaced by the symmetric and anti-symmetric functions $\hat{\ell}_1+\hat{\ell}_{-1}$ and $-\hat{\ell}_1+\hat{\ell}_{-1}$, with double support lengths. Finally, at the left boundary $\hat{\ell}_{-1}$ is multiplied by 2, and at the right boundary $\hat{\ell}_1$ is multiplied by -2. In view of (4.2), one verifies that the resulting primal and dual collections, denoted as $\boldsymbol{\Phi}_j$ and $\tilde{\boldsymbol{\Phi}}_j$, are *biorthogonal*.

By construction, $\operatorname{span}\boldsymbol{\Phi}_j = V_j$ and $Z_j \subset \tilde{V}_j := \operatorname{span}\tilde{\boldsymbol{\Phi}}_j \subset Z_{j+1}$. The collections $\boldsymbol{\Phi}_j$ and $\tilde{\boldsymbol{\Phi}}_j$ are uniformly local and uniformly $L_2(\mathrm{I})$-bounded. As a consequence of these properties, it holds that $\|\mathbf{c}_j^\top\boldsymbol{\Phi}_j\|_{L_2(\mathrm{I})} \lesssim \|\mathbf{c}_j\|_{\ell_2}$ and similar at the dual side. Together with the biorthogonality, this shows that $\boldsymbol{\Phi}_j$ and $\tilde{\boldsymbol{\Phi}}_j$ are uniform $L_2(\mathrm{I})$-Riesz bases. E.g., at the dual side, we have $\|\tilde{\mathbf{c}}_j\|_{\ell_2} = \sup_{\mathbf{c}_j\neq 0}\frac{|\langle \mathbf{c}_j^\top\boldsymbol{\Phi}_j,\tilde{\mathbf{c}}_j^\top\tilde{\boldsymbol{\Phi}}_j\rangle_{L_2(\mathrm{I})}|}{\|\mathbf{c}_j\|_{\ell_2}} \lesssim \|\tilde{\mathbf{c}}_j^\top\tilde{\boldsymbol{\Phi}}_j\|_{L_2(\mathrm{I})}\sup_{\mathbf{c}_j\neq 0}\frac{\|\mathbf{c}_j^\top\boldsymbol{\Phi}_j\|_{L_2(\mathrm{I})}}{\|\mathbf{c}_j\|_{\ell_2}} \lesssim \|\tilde{\mathbf{c}}_j^\top\tilde{\boldsymbol{\Phi}}_j\|_{L_2(\mathrm{I})}$.

Finally, we note that the basis transformations between the interpolatory basis for V_j and the basis $\boldsymbol{\Phi}_j$ are uniformly local.

4.2 *Wavelets*

In view of Proposition 3.3, in order to define the wavelets, it is sufficient to construct $\boldsymbol{\Xi}_{j+1}$ such that $\boldsymbol{\Phi}_j\cup\boldsymbol{\Xi}_{j+1}$ is a uniform $L_2(\mathrm{I})$-Riesz basis for V_{j+1}. Moreover, in order to obtain uniformly local wavelets with uniformly local duals, we would like to select $\boldsymbol{\Xi}_{j+1}$ such that the basis transformations between $\boldsymbol{\Phi}_{j+1}$ and $\boldsymbol{\Phi}_j\cup\boldsymbol{\Xi}_{j+1}$ are uniformly local. Since the basis transformations between the interpolatory basis of V_j and the basis $\boldsymbol{\Phi}_j$ are uniformly local, the latter condition is equivalent to the locality of the basis transformations between the interpolatory basis for V_{j+1} and the union of the interpolatory basis for V_j and $\boldsymbol{\Xi}_{j+1}$.

A natural choice for $\boldsymbol{\Xi}_{j+1}$ is the subset of interpolatory basis functions for V_{j+1} that correspond to the new degrees of freedom. With this choice, the last mentioned basis transformations are uniformly local.

Indeed, with I_ℓ being the canonical interpolation operator onto V_ℓ, the argument is that for $u_{j+1} \in V_{j+1}$ the computation of the splitting $u_{j+1} = I_j u_{j+1} + I_{j+1}(u_{j+1} - I_j u_{j+1})$ requires local quantities only.

Yet, in order to reduce the support size of most of the resulting wavelets, we do not simply take Ξ_{j+1} to be the above collection, but we construct it from that collection by applying a uniformly local transformation with uniformly local inverse. Our aim is to ensure that most functions in Ξ_{j+1} are orthogonal to those duals scaling functions that correspond to primal scaling functions that have supports that extend to more than one macro-element. The resulting wavelets will then have no components in the directions of those scaling functions.

Similar to the previous subsection, it is sufficient to specify $\hat{\Xi}$ such that $\hat{\Phi} \cup \hat{\Xi}$ is a basis for $\sum_{p=-2}^{1} P_4(p/2,(p+1)/2) \cap C(\hat{\mathrm{I}})$. Then by lifting $\hat{\Xi}$ to any macro-element $[k2^{-(j+1)},(k+2)2^{-(j+1)}]$ multiplying it by $2^{(j+1)/2}$, and taking the union over all macro-elements, the collection Ξ_{j+1} is obtained. Since the function values and first order derivatives of any function in $\hat{\Xi}$ will vanish at $\partial\hat{\mathrm{I}}$, no degrees of freedom will have to be identified over the interfaces.

Let $\bar{H} = \{-1,-\frac{1}{2},0,\frac{1}{2},1\}$ and $\bar{L} = \{-1,-\frac{3}{4},-\frac{1}{2},-\frac{1}{4},0,\frac{1}{4},\frac{1}{2},\frac{3}{4},1\}$. We define the collection $\hat{\Sigma} := \{\bar{h}_i : \bar{H} \setminus H\} \cup \{\bar{\ell}_j : j \in \bar{L} \setminus L\} \subset \sum_{p=-2}^{1} P_4(p/2,(p+1)/2) \cap C(\hat{\mathrm{I}})$ as follows

$$\begin{aligned} \overleftarrow{h}_i(\hat{\imath}) &= \delta_{\hat{\imath}\hat{\imath}}, & \overleftarrow{\ell}_j(\hat{\imath}) &= 0, \\ \overline{h}_i(\hat{\jmath}) &= 0, & \overline{\ell}_j(\hat{\jmath}) &= \delta_{\hat{\jmath}\hat{\jmath}}, \end{aligned} \qquad (\hat{\imath} \in \bar{H}, \hat{\jmath} \in \bar{L}),$$

cf. Figure 3. So $\hat{\Sigma}$ consists of the interpolatory basis functions corresponding to the new degrees of freedom.

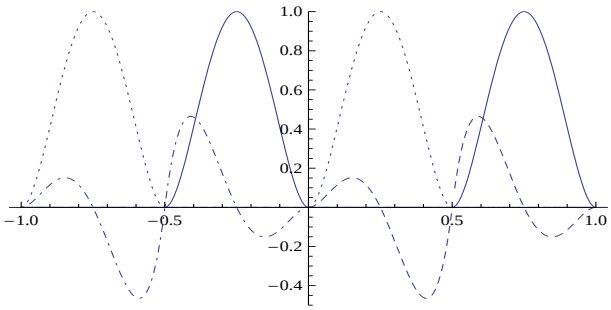

FIG. 3. The "preliminary" initial stable completion $\hat{\Sigma}$ on the reference element.

We will define $\hat{\Xi}$ by applying some basis transformations to $\hat{\Sigma}$. In view of our aforementioned aim, we look for $\hat{\Xi}$ such that most of its elements are orthogonal to $\tilde{\hat{\ell}}_{\pm 1}$.

Step 1. We determine $\hat{\xi}_{\pm\frac{1}{2}}$ as the best approximation to $\tilde{\hat{\ell}}_{\pm 1}$ from span $\hat{\Sigma}$, i.e.,

$$\hat{\xi}_{\pm\frac{1}{2}} = \langle \hat{\Sigma}, \tilde{\hat{\ell}}_{\pm 1} \rangle_{L_2(\hat{\mathrm{I}})} \langle \hat{\Sigma}, \hat{\Sigma} \rangle^{-1}_{L_2(\hat{\mathrm{I}})} \hat{\Sigma},$$

and then redefine the obtained $\{\hat{\xi}_{\pm\frac{1}{2}}\}$ by biorthogonalizing it with $\{\tilde{\hat{\ell}}_{\pm 1}\}$, i.e.,

$$\begin{bmatrix}\hat{\xi}_{-\frac{1}{2}}\\ \hat{\xi}_{\frac{1}{2}}\end{bmatrix} \leftarrow \left\langle \begin{bmatrix}\hat{\xi}_{-\frac{1}{2}}\\ \hat{\xi}_{\frac{1}{2}}\end{bmatrix}, \begin{bmatrix}\tilde{\hat{\ell}}_{-1}\\ \tilde{\hat{\ell}}_{1}\end{bmatrix} \right\rangle^{-1}_{L_2(\hat{\mathrm{I}})} \begin{bmatrix}\hat{\xi}_{-\frac{1}{2}}\\ \hat{\xi}_{\frac{1}{2}}\end{bmatrix}.$$

Step 2. We select $\{\hat{\xi}_{\pm\frac{3}{4}}, \hat{\xi}_{\pm\frac{1}{4}}\}$ from $\operatorname{span}\hat{\Sigma} \cap \operatorname{span}\{\tilde{\hat{\ell}}_{\pm 1}\}^{\perp_{L_2(\hat{\mathrm{I}})}}$ by means of

$$\hat{\xi}_p := \bar{\ell}_p - \langle \bar{\ell}_p, \tilde{\hat{\ell}}_{-1}\rangle_{L_2(\bar{\mathrm{I}})}\hat{\xi}_{-\frac{1}{2}} - \langle \bar{\ell}_p, \tilde{\hat{\ell}}_{1}\rangle_{L_2(\bar{\mathrm{I}})}\hat{\xi}_{\frac{1}{2}} \quad (p \in \{\pm\tfrac{3}{4}, \pm\tfrac{1}{4}\}).$$

Now it holds that $\operatorname{span}\hat{\Xi} = \operatorname{span}\hat{\Sigma}$, and the coefficients of $\hat{\Xi}$ in terms of $\hat{\Sigma}$ are given in Table 3.

	$\bar{\ell}_{-\frac{3}{4}}$	$\bar{h}_{-\frac{1}{2}}$	$\bar{\ell}_{-\frac{1}{4}}$	$\bar{\ell}_{\frac{1}{4}}$	$\bar{h}_{\frac{1}{2}}$	$\bar{\ell}_{\frac{3}{4}}$
$\hat{\xi}_{-\frac{3}{4}}$	316773	−27794688	236519	138180	2619904	103140
$\hat{\xi}_{-\frac{1}{2}}$	67599950	638640640	−5457650	−3219590	−65697280	−1779910
$\hat{\xi}_{-\frac{1}{4}}$	22235	366336	3253977	−11556	−1234432	127356
$\hat{\xi}_{\frac{1}{4}}$	127356	1234432	−11556	3253977	−366336	22235
$\hat{\xi}_{\frac{1}{2}}$	−1779910	65697280	−3219590	−5457650	−638640640	67599950
$\hat{\xi}_{\frac{3}{4}}$	103140	−2619904	138180	236519	27794688	316773

Table 3. The coefficients of the initial stable completion $\hat{\Xi}$ on the reference element, all multiplied by 3262142, in terms of $\hat{\Sigma}$.

At this point, we have specified Φ_j, $\tilde{\Phi}_j$ and Ξ_{j+1}. Using these ingredients, the collections of wavelets Ψ_{j+1} are determined by Proposition 3.3. Next, we redefine this Ψ_{j+1} by applying some basis transformations to this collection in order to improve its conditioning.

By construction, we have $\hat{\xi}_{-\frac{1}{2}} \perp_{L_2(\hat{\mathrm{I}})} \tilde{\hat{\ell}}_1$, $\hat{\xi}_{\frac{1}{2}} \perp_{L_2(\hat{\mathrm{I}})} \tilde{\hat{\ell}}_{-1}$ and $\hat{\xi}_{\pm\frac{3}{4}}, \hat{\xi}_{\pm\frac{1}{4}} \perp_{L_2(\hat{\mathrm{I}})} \tilde{\hat{\ell}}_{\pm 1}$. This means that the support of a wavelet resulting from $\hat{\xi}_{-\frac{1}{2}}$, $\hat{\xi}_{\frac{1}{2}}$ or $\hat{\xi}_{\pm\frac{3}{4}}, \hat{\xi}_{\pm\frac{1}{4}}$ is contained in the union of a macro-element and its left neighbor (if any), the union of a macro-element and its right neighbor (if any), or in the macro-element itself. In other words, each interface between macro-elements is contained in the interior of the supports of two wavelets that have their supports in the union of the macro-elements at both sides of the interface (the "interface wavelets"), each macro-element contains the supports of four wavelets (the "internal wavelets"), and both macro-elements at left and right boundary contain the support of an additional wavelet (the "boundary wavelets").

In order to improve the conditioning of the resulting wavelets, first we orthogonalize each group of four internal wavelets. Then we make the interface and boundary wavelets orthogonal to the internal wavelets by subtracting from each interface or boundary wavelet $\psi_{j,k}$ all (non-zero) terms $\frac{\langle \psi_{j,k}, \psi_{j,\hat{k}}\rangle_{L_2(\mathrm{I})}}{\langle \psi_{j,\hat{k}}, \psi_{j,\hat{k}}\rangle_{L_2(\mathrm{I})}}\psi_{j,\hat{k}}$ where $\hat{k}$ runs over the (neighboring) internal wavelets. Finally, we orthogonalize each resulting group of two interface wavelets. All these transformations are local and have local inverses, meaning that the resulting dual wavelets are still uniformly local.

We end up with 8 types of wavelets $\psi_{(i)}$ ($1 \leqslant i \leqslant 8$): Four internal ones ($1 \leqslant i \leqslant 4$), two interface wavelets ($5 \leqslant i \leqslant 6$), and a left ($i = 7$) and right boundary wavelet ($i = 8$). These wavelets, scaled

arbitrarily, are illustrated in Figure 4. To efficiently specify these wavelets in terms of the interpolatory basis of $\sum_{p\in\mathbb{Z}} P_4(p/2,(p+1)/2)\cap C^1(\mathbb{R})$, we introduce $\check{\psi}_{(6)}(x) := \psi_{(5)}(2-x)$. The functions $\psi_{(5)}$, $\check{\psi}_{(6)}$ were the interface wavelets before $\check{\psi}_{(6)}$ was modified by making it orthogonal to ψ_5, i.e.,

$$\psi_{(6)} = \check{\psi}_{(6)} - \frac{\langle\check{\psi}_{(6)},\psi_{(5)}\rangle_{L_2(\mathbb{R})}}{\langle\psi_{(5)},\psi_{(5)}\rangle_{L_2(\mathbb{R})}}\psi_{(5)}, \text{ where } \frac{\langle\check{\psi}_{(6)},\psi_{(5)}\rangle_{L_2(\mathbb{R})}}{\langle\psi_{(5)},\psi_{(5)}\rangle_{L_2(\mathbb{R})}} = \frac{22164193}{46639227}.$$

Coefficients of $\{\psi_{(i)} : 1 \leqslant i \leqslant 8, i \neq 6\} \cup \{\check{\psi}_{(6)}\}$ in terms of the interpolatory basis, i.e., their function values at $\frac{1}{4}\mathbb{Z}$ and derivatives at $\frac{1}{2}\mathbb{Z}$ are given in Tables 4 and 5, and Table 6, respectively. All missing function values or derivatives can be found from the following relations: $\psi_{(1)}(x) = \psi_{(1)}(-x)$, $\psi_{(2)}(x) = \psi_{(2)}(-x)$, $\psi_{(3)}(x) = -\psi_{(3)}(-x)$, $\psi_{(4)}(x) = -\psi_{(4)}(-x)$, $\check{\psi}_{(6)}(x) = \psi_{(5)}(2-x)$, and $\psi_{(7)}(x) = \psi_{(8)}(-x)$.

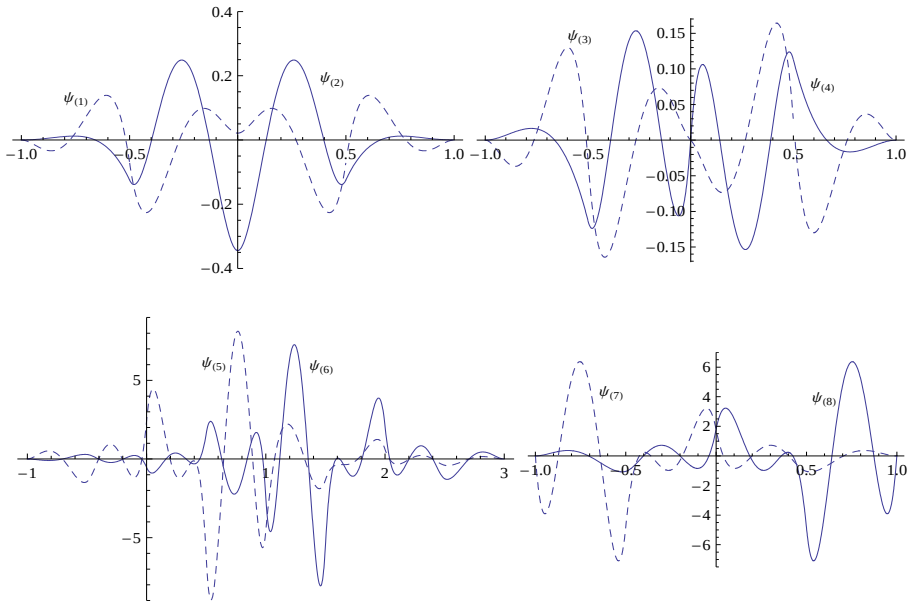

FIG. 4. The 4 internal wavelets (top row), the 2 interface wavelets (bottom left), and the left- and right boundary wavelets (bottom right).

Finally, in order to improve the conditioning of the basis in norms other than the $L_2(\mathrm{I})$-norm, in our definition of the wavelet collection Ψ, we replace the single-scale basis Φ_0 for V_0 by a biorthogonal three-scale basis. We set $V_{-1} := P_4(\mathrm{I})\cap H_0^1(\mathrm{I})$, $V_{-2} := \mathrm{span}\{x(1-x)\}$, $\tilde{V}_{-1} := P_2(\mathrm{I})$, and $\tilde{V}_{-2} := P_0(\mathrm{I})$, and construct a basis for V_0 as the union of the $L_2(\mathrm{I})$-orthogonal bases $\{\phi_1 := x \mapsto x(1-x)\}$ for V_{-2}, $\{\phi_2,\phi_3\}$ for $V_{-1}\cap\tilde{V}_{-2}^{\perp_{L_2(\mathrm{I})}}$ and $\{\phi_4,\phi_5,\phi_6\}$ for $V_0\cap\tilde{V}_{-1}^{\perp_{L_2(\mathrm{I})}}$. The resulting basis $\{\phi_1,\ldots,\phi_6\}$, scaled arbitrarily, is illustrated in Figure 5, and the coefficients of $\{\phi_2,\ldots,\phi_6\}$ in terms of the interpolatory basis for V_0 are given in Table 7.

At this point, we have fully specified the wavelet basis $\Psi = \{\phi_1,\ldots,\phi_6\}\cup\cup_{j\in\mathbb{N}}\Psi_j$.

	$-\frac{3}{4}$	$-\frac{2}{4}$	$-\frac{1}{4}$	0
$\psi_{(1)}$	$\frac{308451}{13889536}$	$-\frac{60949}{868096}$	$\frac{415475}{13889536}$	$\frac{1155}{54256}$
$\psi_{(2)}$	$\frac{2499}{246272}$	$-\frac{109}{3848}$	$\frac{4711}{246272}$	0
$\psi_{(3)}$	$\frac{443061983}{36627756800}$	$-\frac{285143417}{2289234800}$	$\frac{362748295}{1465110272}$	$-\frac{9856717}{28615435}$
$\psi_{(4)}$	$\frac{148837}{9425168}$	$-\frac{66512}{589073}$	$\frac{1415825}{9425168}$	0
$\psi_{(5)}$	$\frac{53708555285}{166405095424}$	$-\frac{14980898075}{10400318464}$	$\frac{327725985455}{499215286272}$	$\frac{3690665}{1534272}$
$\breve{\psi}_{(6)}$	$\frac{47396075155}{499215286272}$	$-\frac{12393524605}{31200955392}$	$\frac{72032214595}{499215286272}$	$\frac{315095}{511424}$
$\psi_{(7)}$	$\frac{794099332235}{124803821568}$	$-\frac{37693124645}{7800238848}$	$-\frac{118474648325}{124803821568}$	$\frac{686345}{383568}$
$\psi_{(8)}$	$\frac{28432397675}{124803821568}$	$-\frac{8137292405}{7800238848}$	$\frac{63923442715}{124803821568}$	$\frac{686345}{383568}$

Table 4. Values of $\{\psi_{(i)}, 1 \leqslant i \leqslant 8, i \neq 6\} \cup \{\breve{\psi}_{(6)}\}$ at $-\frac{3}{4}, -\frac{1}{2}, -\frac{1}{4}, 0$.

	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
$\psi_{(5)}$	$-\frac{544659795185}{499215286272}$	$-\frac{192581213825}{31200955392}$	$\frac{1324982041845}{166405095424}$	$-\frac{55}{12}$
$\breve{\psi}_{(6)}$	$-\frac{23587067295}{166405095424}$	$-\frac{13936238415}{10400318464}$	$\frac{798548796595}{499215286272}$	$-\frac{55}{12}$

Table 5. Values of $\{\psi_{(5)}, \breve{\psi}_{(6)}\}$ at $\frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1$.

5. Riesz constants

As shown by Theorem 3.1, for $s \in (-\frac{1}{2}, \frac{5}{2})$, our collection Ψ, normalized in $\mathscr{H}^s(\mathrm{I})$, is a Riesz basis for that space, and so in particular for $L_2(\mathrm{I})$ and $H_0^1(\mathrm{I})$. In various estimates, the values of (the quotients of) the corresponding Riesz constants $\lambda_{L_2(\mathrm{I})}$, $\Lambda_{L_2(\mathrm{I})}$, $\lambda_{H_0^1(\mathrm{I})}$, and $\Lambda_{H_0^1(\mathrm{I})}$ play an important role. In Figure 6, we present numerically computed values of $\lambda_{L_2(\mathrm{I}),J}$, $\Lambda_{L_2(\mathrm{I}),J}$, $\kappa_{L_2(\mathrm{I}),J} := \frac{\Lambda_{L_2(\mathrm{I}),J}}{\lambda_{L_2(\mathrm{I}),J}}$, $\lambda_{H_0^1(\mathrm{I}),J}$, $\Lambda_{H_0^1(\mathrm{I}),J}$, and $\kappa_{H_0^1(\mathrm{I}),J} := \frac{\Lambda_{H_0^1(\mathrm{I}),J}}{\lambda_{H_0^1(\mathrm{I}),J}}$, where $\lambda_{L_2(\mathrm{I}),J} := \inf_{0 \neq \mathbf{d} \in \ell_2(\{\lambda \in \nabla : |\lambda| \leqslant J\})} \frac{\| \sum_{\lambda \in \nabla} d_\lambda \psi_\lambda \|^2_{L_2(\mathrm{I})}}{\sum_{\lambda \in \nabla} |d_\lambda|^2 \| \psi_\lambda \|^2_{L_2(\mathrm{I})}}$ with analogous definitions of the other quantities.

6. Adaptive Wavelet-Galerkin Method

The Adaptive Wavelet-Galerkin Method (**AWGM**), already mentioned in Sect. 1, is described here in somewhat more detail. For ease of exposition, we shall assume that the right-hand side $\mathbf{f} = [f(\boldsymbol{\psi}_{\boldsymbol{\lambda}})]_{\boldsymbol{\lambda} \in \boldsymbol{\nabla}}$ is finite. We shall comment on the general case in Remark 6.1. More details about the Adaptive Wavelet-Galerkin Method, we refer to Cohen *et al.* (2001); Stevenson (2009).

AWGM$[\varepsilon] \to \mathbf{w}_\varepsilon$:
% Input: $\varepsilon > 0$.
% Parameters: $\mu \in (0, \kappa(\mathbf{A})^{-\frac{1}{2}})$ *and* $\gamma \in (0, \mu\,\kappa(\mathbf{A})^{-1})$.

$i := 0$, $\Lambda_i := \emptyset$, $\mathbf{w}^{(i)} := 0$, $\mathbf{r}^{(i)} := \mathbf{f}$
`while` $\|\mathbf{r}^{(i)}\| > \varepsilon$ `do`
  $\Lambda_{i+1} := \mathbf{EXPAND}[\Lambda_i, \mathbf{r}^{(i)}, \mu \|\mathbf{r}^{(i)}\|]$
  $\mathbf{w}^{(i+1)} := \mathbf{GALERKIN}[\Lambda_{i+1}, \mathbf{w}^{(i)}, \|\mathbf{r}^{(i)}\|, \gamma \|\mathbf{r}^{(i)}\|]$
  $\mathbf{r}^{(i+1)} := \mathbf{f} - \mathbf{A}\mathbf{w}^{(i+1)}$
  $i := i + 1$
`enddo`
$\mathbf{w}_\varepsilon := \mathbf{w}^{(i)}$

	-1	$-\frac{1}{2}$	0	$\frac{1}{2}$	1
$\dot{\psi}_{(1)}$	0	$-\frac{502391}{108512}$	0	$\frac{502391}{108512}$	0
$\dot{\psi}_{(2)}$	0	$-\frac{29149}{7696}$	$-\frac{547}{1924}$	$-\frac{29149}{7696}$	0
$\dot{\psi}_{(3)}$	0	$-\frac{412896003}{286154350}$	0	$\frac{412896003}{286154350}$	0
$\dot{\psi}_{(4)}$	0	$-\frac{667982}{589073}$	$\frac{2402104}{589073}$	$-\frac{667982}{589073}$	0
$\dot{\psi}_{(5)}$	0	$\frac{6471062785}{1300039808}$	$\frac{1697905}{20336}$	$-\frac{190667608285}{1300039808}$	$\frac{165}{2}$
$\dot{\tilde{\psi}}_{(6)}$	0	$\frac{2249357165}{1300039808}$	$\frac{447205}{20336}$	$-\frac{30059475545}{1300039808}$	$-\frac{165}{2}$
$\dot{\psi}_{(7)}$	-165	$\frac{40152033185}{325009952}$	$-\frac{312675}{5084}$	$-\frac{1055426405}{325009952}$	0
$\dot{\psi}_{(8)}$	0	$\frac{1055426405}{325009952}$	$\frac{312675}{5084}$	$-\frac{40152033185}{325009952}$	165

Table 6. Values of $\{\dot{\psi}_{(i)}, 1 \leqslant i \leqslant 8, i \neq 6\} \cup \{\dot{\tilde{\psi}}_{(6)}\}$ at $-1, -\frac{1}{2}, 0, \frac{1}{2}, 1$.

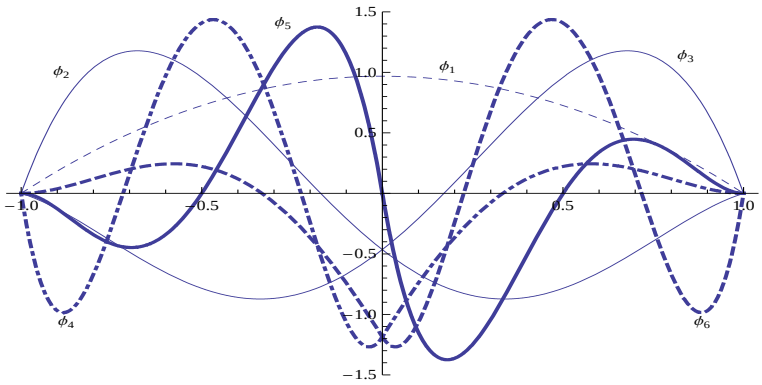

FIG. 5. The three-scale basis $\{\phi_i : 1 \leqslant i \leqslant 6\}$ for V_0.

By a call of **GALERKIN**, the Galerkin system corresponding to the current active wavelet index set is solved within some prescribed tolerance. The previous approximate Galerkin solution is used as a starting value of an iterative solver.

GALERKIN$[\Lambda, \mathbf{w}, \delta, \varepsilon] \to \bar{\mathbf{w}}$:
% Input: $\varepsilon, \delta > 0$, $\Lambda \subset \nabla$, *a* $\mathbf{w} \in \ell_2(\Lambda)$ *with* $\|\mathbf{f}|_\Lambda - \mathbf{A}|_{\Lambda\times\Lambda}\mathbf{w}\| \leqslant \delta$.
% Output: $\bar{\mathbf{w}} \in \ell_2(\Lambda)$ *with* $\|\mathbf{f}|_\Lambda - \mathbf{A}|_{\Lambda\times\Lambda}\bar{\mathbf{w}}\| \leqslant \varepsilon$ *in* $\mathcal{O}(\log(\delta/\varepsilon)\#\Lambda)$ *operations.*

Using that $\mathbf{A}_\Lambda$ is well-conditioned, uniformly in Λ, we have implemented this routine as a Conjugate Residual iteration with starting vector $\mathbf{w}$.

By a call of **EXPAND**, the current active wavelet index set is expanded with those indices where the current approximate residual has large values. A formal description of this procedure is given below.

EXPAND$[\Lambda, \mathbf{g}, \sigma] \to \bar{\Lambda}$:
% Input: $\Lambda \subset \nabla$, *a finitely supported* $\mathbf{g} \in \ell_2(\nabla)$, *and a scalar* $\sigma \in [0, \|\mathbf{g}\|_{\ell_2(\nabla)}]$.
% Output: $\Lambda \subset \bar{\Lambda} \subset \nabla$ *with* $\|\mathbf{P}_{\bar{\Lambda}}\mathbf{g}\| \geqslant \sigma$ *and such that, up to some absolute multiple,*
% $\#(\bar{\Lambda} \setminus \Lambda)$ *is minimal over all such* $\bar{\Lambda}$, *and the cost of the call is* $\mathcal{O}(\#\Lambda \cup \operatorname{supp}\mathbf{g})$.

Noting that $\|\mathbf{P}_{\bar{\Lambda}}\mathbf{g}\| \geqslant \sigma$ is equivalent to $\|\mathbf{P}_{\bar{\Lambda}\setminus\Lambda}\mathbf{g}\|^2 \geqslant \sigma^2 - \|\mathbf{P}_\Lambda\mathbf{g}\|^2$, a set $\bar{\Lambda}$ with minimal $\#(\bar{\Lambda} \setminus \Lambda)$

	h_{-1}	ℓ_{-1}	$\ell_{-\frac{1}{2}}$	h_0	ℓ_0	$\ell_{\frac{1}{2}}$	h_1	ℓ_1
ϕ_2	$\frac{18}{5}$	0	$\frac{33}{80}$	-1	$-\frac{1}{5}$	$-\frac{27}{80}$	$\frac{2}{5}$	0
ϕ_3	$-\frac{2}{5}$	0	$-\frac{27}{80}$	1	$-\frac{1}{5}$	$\frac{33}{80}$	$-\frac{18}{5}$	0
ϕ_4	0	0	0	-1	0	0	0	0
ϕ_5	-4	0	$\frac{19}{64}$	1	$-\frac{1}{4}$	$\frac{3}{64}$	0	0
ϕ_6	0	0	$\frac{3}{64}$	-1	$-\frac{1}{4}$	$\frac{19}{64}$	4	0

Table 7. Coefficients of $\{\phi_i : 2 \leqslant i \leqslant 6\}$.

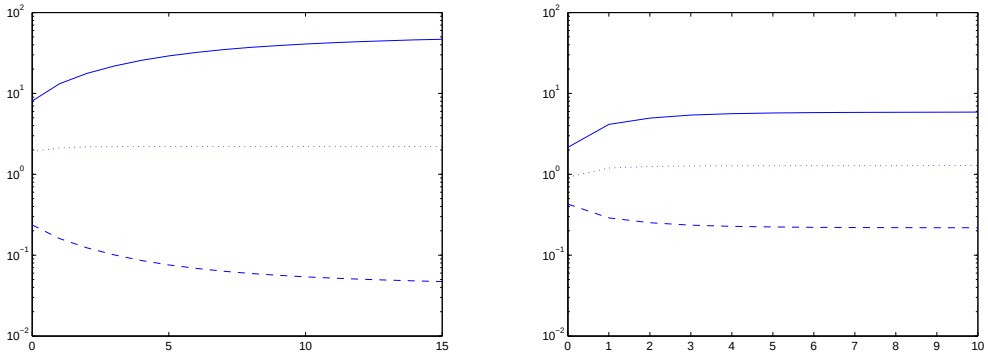

FIG. 6. Smallest (dashed) and largest (dotted) eigenvalues and condition number (solid) of the univariate (scaled) "mass" $\left[\int_{\mathrm{I}} \frac{\psi_\mu \psi_\lambda}{\|\psi_\mu\|_{L_2(\mathrm{I})}\|\psi_\lambda\|_{L_2(\mathrm{I})}}\right]_{\lambda,\mu\in\nabla,|\lambda|,|\mu|\leqslant J}$ (left) and "stiffness" $\left[\int_{\mathrm{I}} \frac{\dot\psi_\mu \dot\psi_\lambda}{\|\dot\psi_\mu\|_{L_2(\mathrm{I})}\|\dot\psi_\lambda\|_{L_2(\mathrm{I})}}\right]_{\lambda,\mu\in\nabla|\lambda|,|\mu|\leqslant J}$ (right) matrices on levels $J = 0,1,\ldots$.

can be found by ordering $\{\mathbf{g}_{\boldsymbol{\lambda}} : \boldsymbol{\lambda} \in \bar\Lambda \setminus \Lambda\}$ by non-increasing modulus, and then by selecting elements from the head to the tail of this ordered sequence until the criterion is met. The loglinear complexity of this implementation can be reduced to a linear complexity by performing an approximate sorting, at the expense of obtaining a $\bar\Lambda$ for which $\bar\Lambda \setminus \Lambda$ is only minimal up to some absolute factor.

The following result shows that the **AWGM** is quasi-optimal.

THEOREM 6.1 For $\mathbf{w}_\varepsilon := \mathbf{AWGM}[\varepsilon]$, we have that $\|\mathbf{f} - \mathbf{A}\mathbf{w}_\varepsilon\| \leqslant \varepsilon$. If for some $s > 0$,

$$\mathbf{u} \in \mathscr{A}^s := \{\mathbf{v} \in \ell_2(\boldsymbol{\nabla}) : |\mathbf{v}|_{\mathscr{A}^s} := \sup_{N\in\mathbb{N}_0} N^s \|\mathbf{v} - \mathbf{v}_N\| < \infty\},$$

i.e., $\mathbf{u}$ is approximated by its best N-term approximations at rate s, then both $\#\operatorname{supp}\mathbf{w}_\varepsilon$ and, assuming $\varepsilon \lesssim \|\mathbf{f}\|$, the number of operations used by the call are of order $\varepsilon^{-1/s}|\mathbf{u}|_{\mathscr{A}^s}^{1/s}$, only dependent on s when it tends to 0 or ∞, and on upper bounds for $\|\mathbf{A}\|$ and $\|\mathbf{A}^{-1}\|$, cf. (1.6) and Lemma 1.1.

REMARK 6.1 For general functions f, the vector $\mathbf{f}$ has infinite support. Inside the Adaptive Wavelet-Galerkin Method, this $\mathbf{f}$ has to be approximated with increasing accuracy, see Stevenson (2009) for details. Assuming $\mathbf{u} \in \mathscr{A}^s$, then if these approximations of $\mathbf{f}$ converge at rate s, then the overall rate of the adaptive method is still s. Using that $\mathbf{A}$ is sparse, it can be shown that if $\mathbf{u} \in \mathscr{A}^s$, then also $\mathbf{f} \in \mathscr{A}^s$, meaning that best N-term approximations of $\mathbf{f}$ converge at rate s.

7. Rates for the reaction-diffusion problem

In this section, we consider the problem of finding $u_\eta \in H^1_0(\square)$ such that

$$a_\eta(u_\eta, v) := \int_\square \eta^2 \nabla u_\eta \cdot \nabla v + u_\eta v = f(v) \quad (v \in H^1_0(\square)). \tag{7.1}$$

The bilinear form a_η defines a boundedly invertible operator, uniformly in $\eta > 0$, between $H^1_0(\square)$ equipped with the energy norm

$$|||\cdot||| = |||\cdot|||_\eta = \sqrt{\eta^2 |\cdot|^2_{H^1(\square)} + \|\cdot\|^2_{L_2(\square)}}$$

and its dual equipped with the corresponding dual norm. The case $\eta = 0$ is included when $H^1_0(\square)$ is read as the closure of $C^\infty_0(\square)$ with respect to the energy norm.

We consider a univariate wavelet collection $\Psi = \{\psi_\lambda : \lambda \in \nabla\}$ that, properly scaled, is a Riesz basis for $L_2(\mathrm{I})$ and $H^1_0(\mathrm{I})$ (and $1 = \|\psi_\lambda\|_{L_2(\mathrm{I})} \eqsim 2^{-|\lambda|}\|\psi_\lambda\|_{H^1(\mathrm{I})}$), has *order* $d > 1$, and has a corresponding collection of dual wavelets $\tilde\Psi = \{\tilde\psi_\lambda : \lambda \in \nabla\}$ that is *uniformly local* (i.e., the dual wavelets satisfy (1) and (2)). An example of such a collection Ψ is given by the one constructed in Section 4, in which case $d = 5$.

We will make the following assumption on the solution of (7.1):

Assumption 7.1 The solution u_η of (7.1) allows for a splitting into

$$u_\eta = U_\eta + \sum_{k=1}^{n} \sum_{p=1}^{\binom{n}{k}2^k} V_{\eta,k,p}$$

where for $\|\alpha\|_\infty \leqslant d$, $\|\partial^\alpha U_\eta\|_{L_\infty(\square)} \lesssim 1$, and $V_{\eta,k,p}$ is a layer function associated to the pth $(n-k)$-face of $\square$. Note that a 0-face is a vertex and an $(n-1)$-face is a side. For some constants $\gamma, \tau > 0$ the layer functions $V_{\eta,k} = V_{\eta,k,p}$ associated to the $(n-k)$-face $x_1 = \cdots = x_k = 0$ satisfy for $\|\alpha\|_\infty \leqslant d$ and $\eta > 0$,

$$|\partial^\alpha V_{\eta,1}(\mathbf{x})| \lesssim \eta^{-\alpha_1} e^{-\gamma x_1/\eta}, \tag{7.2}$$

and for $k \in \{2, \ldots, n\}$,

$$|\partial^\alpha V_{\eta,k}(\mathbf{x})| \lesssim \eta^{-(\alpha_1+\cdots+\alpha_k)} \min\left(1, \frac{\sqrt{x_1^2+\cdots+x_k^2}}{\eta}\right)^{1-\frac{k}{2}+\tau-(\alpha_1+\cdots+\alpha_k)} e^{-\gamma\frac{\sqrt{x_1^2+\cdots+x_k^2}}{\eta}}, \tag{7.3}$$

where the other layer functions satisfy analogous decay estimates as function of the distance to their faces. For $\eta = 0$, the layer functions are zero.

For the two-dimensional case $n = 2$ and sufficiently smooth right-hand side f, Assumption 7.1 has been verified in Han & Kellogg (1990). The final corollary in Han & Kellogg (1990) suggests that α_1 at the right-hand side of (7.2) should read as $|\alpha|$. Inspection of the proof (in particular (2.6c)) reveals that (7.2) is correct (cf. also (Apel & Lube, 1998, Lemma 2)), which will be essential. The upper bounds from Han & Kellogg (1990) for the derivatives of the corner layer functions are actually orders of magnitude smaller than those allowed by (7.3).

REMARK 7.1 A generalization of the results from Han & Kellogg (1990) to general polygons and with a variable zeroth order part of the differential operator is given is given in Kellogg (1995) (cf. (Apel, 1999, §5.3.1) for a discussion of (7.2) with α_1 instead of $|\alpha|$ in these cases). The bounds for the corner singularities in Kellogg (1995) are less favorable than in Han & Kellogg (1990), but still orders of magnitude smaller than those that allowed by (7.3).

To the best of our knowledge, complete proofs of Assumption 7.1 in more than two dimensions are not available in the literature.

For $\beta \in [0,1)$, and $\ell \in \mathbb{N}_0$, following Nitsche (2005) now let

$$\nabla_\ell^\beta := \{\lambda \in \nabla : |\lambda| \leqslant \tfrac{\ell}{1-\beta}, \operatorname{dist}(\operatorname{supp} \tilde{\psi}_\lambda, \partial \mathrm{I})^\beta \leqslant 2^{\ell-|\lambda|}\}. \tag{7.4}$$

This set ∇_ℓ^β contains all λ's with $|\lambda| \leqslant \ell$ as well as, for $\beta > 0$, λ's on higher levels, up to $\frac{\ell}{1-\beta}$, when the supports of the corresponding dual wavelets are sufficiently close to 0 or 1. Nevertheless, for any fixed $\beta \in [0,1)$, $\#\nabla_\ell^\beta \eqsim 2^\ell$.

Considering the corresponding collection of tensor product wavelets $\Psi = \{\boldsymbol{\psi}_{\boldsymbol{\lambda}} : \boldsymbol{\lambda} \in \boldsymbol{\nabla} = \nabla^n\}$, let, for $\nu \geqslant 1$,

$$\boldsymbol{\nabla}_{\ell,\nu}^\beta := \{\textstyle\prod_{i=1}^n \nabla_{m_i}^\beta : \mathbf{m} \in \mathbb{N}_0^n, \nu \sum_{i=1}^n m_i + (1-\nu)\max_i m_i \leqslant \ell\}.$$

For $\nu = 1$, $\beta = 0$, this set $\boldsymbol{\nabla}_{\ell,\nu}^\beta$ is the familiar *sparse grid* index set. The set shrinks with increasing ν and grows with increasing β. For any fixed $\beta \in [0,1)$,

$$\#\boldsymbol{\nabla}_{\ell,\nu}^\beta \eqsim \begin{cases} \ell^{n-1}2^\ell & \text{if } \nu = 1, \\ 2^\ell & \text{if } \nu > 1. \end{cases}$$

Let us denote with $\mathbf{P}_{\ell,\nu}^\beta$ the biorthogonal projector $u = \sum_{\boldsymbol{\lambda}\in\boldsymbol{\nabla}} u_{\boldsymbol{\lambda}} \boldsymbol{\psi}_{\boldsymbol{\lambda}} \to \sum_{\boldsymbol{\lambda}\in\boldsymbol{\nabla}_{\ell,\nu}^\beta} u_{\boldsymbol{\lambda}} \boldsymbol{\psi}_{\boldsymbol{\lambda}}$ defined on $L_2(\square)$.

For $k \in \mathbb{N}_0$, $\theta \geqslant 0$, the *weighted Sobolev space* $H_\theta^k(\mathrm{I})$ is defined as the space of all measurable functions u for which, with weight $w_\theta(x) := (x(1-x))^\theta$, the norm

$$\|u\|_{H_\theta^k(\mathrm{I})} = \|u\|_{H_{w_\theta}^k(\mathrm{I})} := \left[\sum_{j=0}^k \int_{\mathrm{I}} |w_\theta(x) u^{(j)}(x)|^2 dx\right]^{\frac{1}{2}}$$

is finite. We have the following error estimates:

THEOREM 7.2 ((Dauge & Stevenson, 2010, Thm. 4.3)) (a). For $\theta \in [0,d)$ and $\beta \in (\frac{\theta}{d}, 1)$,

$$\|u - \mathbf{P}_{\ell,1}^\beta u\|_{L_2(\square)} \lesssim \ell^{\frac{n-1}{2}} 2^{-d\ell} \|u\|_{\otimes_{i=1}^n H_\theta^d(\mathrm{I})} \quad (u \in \otimes_{i=1}^n H_\theta^d(\mathrm{I})),$$

if $\beta \geqslant 1 - \frac{1}{2d}$ or $u \in H_0^1(\square)$.

(b). For $\theta \in [0,d)$, $\beta \in (\frac{\theta}{d}, 1)$, and $\nu \in [1, \frac{d}{d-1})$,

$$\|u - \mathbf{P}_{\ell,\nu}^\beta u\|_{H^1(\square)} \lesssim 2^{-(d-1)\ell} \sqrt{\sum_{m=1}^n \|u\|^2_{\otimes_{i=1}^n H^d_{\theta-\delta_{mi}\min(1,\theta)}(\mathrm{I})}}$$

for all $u \in \cap_{m=1}^n \otimes_{i=1}^n H^d_{\theta-\delta_{mi}\min(1,\theta)}(\mathrm{I}) \cap H_0^1(\square)$.

The bounds from this theorem involve weighted Sobolev norms of u with weights that, for $\theta > 1$, vanish at the whole of $\partial\square$. Whereas for non-singularly perturbed elliptic problems with smooth coefficients and a smooth right-hand side – which have solutions that are smooth up to the non-singular parts of the boundary–, there is no advantage of weights that vanish at those parts of the boundary, to arrive at favorable results for singularly perturbed elliptic problems, it is essential that these weights vanish at the whole of the boundary because of the developments of boundary layers. By combining the decay estimates (7.2), (7.3) with Theorem 7.2, we arrive at the following result:

THEOREM 7.3 Let Assumption 7.1 be valid. Then for any $\varsigma > 0$, there exists a sequence $(\boldsymbol{\nabla}_N)_{N\in\mathbb{N}} \subset \boldsymbol{\nabla}$ with $\#\boldsymbol{\nabla}_N \lesssim N$ such that for the solution of (7.1) it holds that

$$\inf_{w_\eta \in \operatorname{span}\{\boldsymbol{\psi}_{\boldsymbol{\lambda}} : \boldsymbol{\lambda}\in\boldsymbol{\nabla}_N\}} |||u_\eta - w_\eta|||_\eta \lesssim (\log N)^{(n-1)(\frac{1}{2}+d)} N^{-d} + \eta^{\frac{1}{2}-\varsigma} N^{-(d-1)}.$$

Proof. Theorem 7.2 implies that for any $\theta \in [1,d)$, for $i \in \{1,2\}$ there exists a sequence $(\boldsymbol{\nabla}_N^{(i)})_{N\in\mathbb{N}} \subset \boldsymbol{\nabla}$ with $\#\boldsymbol{\nabla}_N^{(i)} \lesssim N$ such that for the corresponding biorthogonal projector, let us call it $\mathbf{P}_N^{(i)}$, it holds that

$$\|u - \mathbf{P}_N^{(1)} u\|_{L_2(\square)} \lesssim (\log N)^{(n-1)(\frac{1}{2}+d)} N^{-d} \|u\|_{\otimes_{i=1}^n H^d_\theta(\mathrm{I})} \quad (u \in \otimes_{i=1}^n H^d_\theta(\mathrm{I})),$$

$$\|u - \mathbf{P}_N^{(2)} u\|_{H^1(\square)} \lesssim N^{-(d-1)} \sqrt{\sum_{m=1}^n \|u\|^2_{\otimes_{i=1}^n H^d_{\theta-\delta_{mi}}(\mathrm{I})}} \quad (u \in \cap_{m=1}^n \otimes_{i=1}^n H^d_{\theta-\delta_{mi}}(\mathrm{I}) \cap H^1_0(\square)).$$

To see how the first estimate follows from Thm. 7.2(a), note that with $N := 2\#\boldsymbol{\nabla}^\beta_{\ell,1} \eqsim \ell^{n-1}2^\ell$, we have $N^{\frac{1}{2}} \lesssim 2^\ell \lesssim N$, and so $\ell^{\frac{n-1}{2}} 2^{-d\ell} \eqsim \ell^{(n-1)(\frac{1}{2}+d)} N^{-d} \eqsim (\log N)^{(n-1)(\frac{1}{2}+d)} N^{-d}$.

The first estimate gives the proof of the theorem for $\eta = 0$ (which is not fully trivial since all wavelets vanish at the boundary), and in the following we consider $\eta > 0$.

Since Ψ is, properly scaled, a Riesz basis for $L_2(\square)$ and $H^1_0(\square)$, we infer that for the biorthogonal projector $\mathbf{P}_N$ onto the span of the wavelets $\boldsymbol{\psi}_{\boldsymbol{\lambda}}$ with $\boldsymbol{\lambda} \in \boldsymbol{\nabla}_N^{(1)} \cup \boldsymbol{\nabla}_N^{(2)}$, it holds that

$$|||u - \mathbf{P}_N u|||_\eta \lesssim (\log N)^{(n-1)(\frac{1}{2}+d)} N^{-d} \|u\|_{\otimes_{i=1}^n H^d_\theta(\mathrm{I})} + \eta N^{-(d-1)} \sqrt{\sum_{m=1}^n \|u\|^2_{\otimes_{i=1}^n H^d_{\theta-\delta_{mi}}(\mathrm{I})}} \tag{7.5}$$

$(u \in \cap_{m=1}^n \otimes_{i=1}^n H^d_{\theta-\delta_{mi}}(\mathrm{I}) \cap H^1_0(\square))$.

We apply (7.5) to the solution u_η of (7.1). In view of Assumption 7.1, it suffices to show that for suitable $\theta \in [1,d)$, and all layer functions $V_{\eta,k}$,

$$\|V_{\eta,k}\|_{\otimes_{i=1}^n H^d_\theta(\mathrm{I})} \lesssim 1, \tag{7.6}$$

$$\sqrt{\sum_{p=1}^n \|V_{\eta,k}\|^2_{\otimes_{i=1}^n H^d_{\theta-\delta_{pi}}(\mathrm{I})}} \lesssim \eta^{-\frac{1}{2}-\varsigma}. \tag{7.7}$$

Setting

$$c_d = \frac{2d-1}{2\gamma},$$

we have $\int_{c_d\eta|\log\eta|}^\infty |\eta^{-d} e^{-\gamma y/\eta}|^2 dy = \frac{1}{2\gamma}$. Taking $\theta > d - \frac{1}{2}$, also $\int_0^{c_d\eta|\log\eta|} |y^\theta \eta^{-d}|^2 dy \lesssim 1$, and so in view of (7.2), we conclude that $V_{\eta,1}$ satisfies (7.6).

To show that for $k \geqslant 2$, $V_{k,\eta}$ satisfies (7.6), in view of (7.3) it suffices to show that

$$\int_{(0,1)^k} |y_1^\theta \cdots y_k^\theta \eta^{-(1-\frac{k}{2}+\tau)} \|y\|_2^{1-\frac{k}{2}+\tau-kd} e^{-\gamma\|y\|_2/\eta}|^2 dy$$
$$\lesssim \int_0^{\sqrt{n}} |r^{k\theta+1-\frac{k}{2}+\tau-kd} \eta^{-(1-\frac{k}{2}+\tau)} e^{-\gamma r/\eta}|^2 r^{k-1} dr \lesssim 1,$$

or that, for some suitable $c > 0$ independent of η,

$$\int_0^{c\eta|\log\eta|} |r^{k\theta+1-\frac{k}{2}+\tau-kd} \eta^{-(1-\frac{k}{2}+\tau)}|^2 r^{k-1} dr \lesssim 1,$$

which holds true when $\theta > \max(d - \frac{1+\tau}{k}, d - \frac{1}{2})$.

Similarly, since for $\theta > d - \varsigma$, $\int_0^{c_d\eta|\log\eta|} |y^{\theta-1}\eta^{-d}|^2 dy \lesssim \eta^{-1-2\varsigma}$, in view of (7.2) $V_{\eta,1}$ satisfies (7.7).

Since for $\theta > \max(d-\tau, d-\frac{1}{2}+\frac{1}{2k})$, $\int_0^1 |r^{k\theta-\frac{k}{2}+\tau-kd} \eta^{-(1-\frac{k}{2}+\tau)} e^{-\gamma r/\eta}|^2 r^{k-1} dr \lesssim \eta^{-1}$, in view of (7.3) $V_{\eta,k}$ for $k \in \{2, \cdots, n\}$ satisfies (7.7), even for $\varsigma = 1$. □

REMARK 7.2 Theorem 7.3 can be slightly improved to the statement that

$$\inf_{w_\eta \in \operatorname{span}\{\boldsymbol{\psi}_{\boldsymbol{\lambda}} : \boldsymbol{\lambda} \in \boldsymbol{\nabla}_N\}} |||u_\eta - w_\eta|||_\eta \lesssim (\log N)^{(n-1)(\frac{1}{2}+d)} N^{-d} + \eta^{\frac{1}{2}} |\log\eta|^{d-\frac{1}{2}} N^{-(d-1)}. \tag{7.8}$$

To see this, let ϕ be a smooth function on $[0,\infty)$, with $\phi \equiv 0$ on $[0,1]$ and $\phi \equiv 1$ on $[2,\infty)$. With $r_p = r_p(\mathbf{x})$ being the distance of $\mathbf{x}$ to the pth $(n-2)$ face of $\square$, we split

$$u_\eta = u_\eta^R + u_\eta^S, \quad \text{where } u_\eta^R(\mathbf{x}) := u_\eta(\mathbf{x}) \prod_{p=1}^{2\binom{n}{2}} \phi(r_p/\eta). \tag{7.9}$$

From (7.2) and (7.3), we infer that u_η^R, u_η^S allow for a splitting into

$$u_\eta^R = U_\eta^R + \sum_{k=1}^n \sum_{p=1}^{\binom{n}{k}2^k} V_{\eta,k,p}^R, \quad u_\eta^S = \sum_{k=2}^n \sum_{p=1}^{\binom{n}{k}2^k} V_{\eta,k,p}^S,$$

where for $\|\alpha\|_\infty \leqslant d$, $\|\partial^\alpha U_\eta^R\|_{L_\infty(\square)} \lesssim 1$, and $V_{\eta,k,p}^R, V_{\eta,k,p}^S$ are layer functions associated to the pth $(n-k)$ face of $\square$. The layer functions $V_{\eta,k}^R = V_{\eta,k,p}^R$, $V_{\eta,k}^S = V_{\eta,k,p}^S$ associated to the $(n-k)$-face $x_1 = \cdots = x_k = 0$ satisfy for $\|\alpha\|_\infty \leqslant d$ and $\eta > 0$,

$$|\partial^\alpha V_{\eta,1}^R(\mathbf{x})| \lesssim \eta^{-\alpha_1} e^{-\gamma x_1/\eta}, \tag{7.10}$$

and for $k \in \{2, \ldots, n\}$,

$$|\partial^\alpha V_{\eta,k}^R(\mathbf{x})| \lesssim \eta^{-(\alpha_1+\cdots+\alpha_k)} \tag{7.11}$$

where $V_{\eta,k}^R$ vanishes outside an $\mathscr{O}(\eta)$ neighborhood of $x_1 = \cdots = x_k = 0$, and

$$|\partial^\alpha V_{\eta,k}^S(\mathbf{x})| \lesssim \eta^{-(\alpha_1+\cdots+\alpha_k)} \min\Big(1, \frac{\sqrt{x_1^2+\cdots+x_k^2}}{\eta}\Big)^{1-\frac{k}{2}+\tau-(\alpha_1+\cdots+\alpha_k)} e^{-\gamma\frac{\sqrt{x_1^2+\cdots+x_k^2}}{\eta}}, \tag{7.12}$$

where the other layer functions satisfy analogous decay estimates as function of the distance to their faces.

(Note that other than with say the splitting of u_η into $U_\eta + \sum_{p=1}^{2n} V_{\eta,1,p}$ and $\sum_{k=2}^{n}\sum_{p=1}^{\binom{n}{k}2^k} V_{\eta,k,p}$, the splitting of u_η into u_η^R and u_η^S preserves the zero boundary conditions.)

In addition to $(\boldsymbol{\nabla}^{(1)})_N$ and $(\boldsymbol{\nabla}^{(2)})_N$, we construct a sequence $(\boldsymbol{\nabla}^{(3)})_N \subset \boldsymbol{\nabla}$ with $\#\boldsymbol{\nabla}_N^{(3)} \lesssim N$, and consider the biorthogonal projector $\mathbf{P}_N$ onto the span of the wavelets $\boldsymbol{\psi}_{\boldsymbol{\lambda}}$ with $\boldsymbol{\lambda} \in \boldsymbol{\nabla}_N^{(1)} \cup \boldsymbol{\nabla}_N^{(2)} \cup \boldsymbol{\nabla}_N^{(3)}$. Since Ψ is a Riesz basis for $L_2(\square)$, in view of the results from the proof of Theorem 7.3, in order to establish the improved bound (7.8) it suffices to show that

$$\|(I-\mathbf{P}_N)u_\eta\|_{H^1(\square)} \lesssim \eta^{-\frac{1}{2}}|\log\eta|^{d-\frac{1}{2}}N^{-(d-1)}.$$

Similarly, since Ψ is, properly scaled, a Riesz basis for $H_0^1(\square)$, for the latter it suffices to show that

$$\begin{aligned}\|(I-\mathbf{P}_N^{(2)})u_\eta^S\|_{H^1(\square)} &\lesssim \eta^{-\frac{1}{2}}N^{-(d-1)},\\ \|(I-\mathbf{P}_N^{(3)})u_\eta^R\|_{H^1(\square)} &\lesssim \eta^{-\frac{1}{2}}|\log\eta|^{d-\frac{1}{2}}N^{-(d-1)},\end{aligned} \tag{7.13}$$

where $\mathbf{P}_N^{(3)}$ is the biorthogonal projector onto the span of the wavelets $\boldsymbol{\psi}_{\boldsymbol{\lambda}}$ with $\boldsymbol{\lambda} \in \boldsymbol{\nabla}_N^{(3)}$. Because the expansion of u_η^S does not contain layer functions associated to $(n-1)$-faces, the first result follows from the proof of Thm. 7.3 (cf. the last line of its proof).

It is sufficient to show (7.13) for $N \eqsim 2^\ell$ and $\ell \in \mathbb{N}_0$. Following the idea of a so-called *Shishkin mesh* (Shishkin (1987)), with $L \in \mathbb{N}_0$ such that

$$2^{L-\ell}c_d\eta|\log\eta| \eqsim 1,$$

we set

$$\check{\nabla}_\ell^\eta := \{\lambda \in \nabla : |\lambda| \leqslant \ell \text{ or } |\lambda| \leqslant L \text{ and } \operatorname{dist}(\operatorname{supp}\tilde{\psi}_\lambda, \partial\mathrm{I}) \leqslant c_d\eta|\log\eta|\}.$$

Then $\#\check{\nabla}_\ell^\eta \eqsim 2^\ell$ (uniformly in $\eta > 0$). Note that the difference between the "regular" level ℓ and the highest level L of the the indices in $\check{\nabla}_\ell^\eta$ is independent of ℓ (but dependent on η), whereas in the definition of ∇_ℓ^β in (7.4) this difference was a multiple of ℓ (independent of η).

With $\check{P}_\ell^\eta$ denoting the biorthogonal projector onto $\operatorname{span}\{\psi_\lambda : \lambda \in \check{\nabla}_\ell^\eta\}$, for $u \in H^d(\mathrm{I}) \cap H_0^1(\mathrm{I})$ we have

$$\begin{aligned}\|u-\check{P}_\ell^\eta u\|_{L_2(\mathrm{I})}^2 &\eqsim \sum_{|\lambda|>L} |\langle\tilde{\psi}_\lambda, u\rangle_{L_2(\mathrm{I})}|^2 + \sum_{\{\ell<|\lambda|\leqslant L:\operatorname{supp}\tilde{\psi}_\lambda \in (c_d\eta|\log\eta|,1-c_d\eta|\log\eta|)\}} |\langle\tilde{\psi}_\lambda, u\rangle_{L_2(\mathrm{I})}|^2\\ &\lesssim 4^{-Ld}\|u\|_{H^d(\mathrm{I})}^2 + 4^{-\ell d}\|u\|_{H^d(c_d\eta|\log\eta|,1-c_d\eta|\log\eta|)}^2\\ &\eqsim 4^{-\ell d}\|u\|_{H^d_{w_\eta^d}(\mathrm{I})}^2\end{aligned} \tag{7.14}$$

with weight w_η^d being the dth power of $w_\eta := \begin{cases} 1 & \text{on } (c_d\eta|\log\eta|, 1-c_d\eta|\log\eta|),\\ \eta|\log\eta| & \text{elsewhere on I.}\end{cases}$ Similarly, for $u \in H^d(\mathrm{I}) \cap H_0^1(\mathrm{I})$,

$$\begin{aligned}\|u-\check{P}_\ell^\eta u\|_{H^1(\mathrm{I})}^2 &\eqsim \sum_{|\lambda|>L} 4^{|\lambda|}|\langle\tilde{\psi}_\lambda, u\rangle_{L_2(\mathrm{I})}|^2 + \sum_{\{\ell<|\lambda|\leqslant L:\operatorname{supp}\psi_\lambda \in (c_d\eta|\log\eta|,1-c_d\eta|\log\eta|)\}} 4^{|\lambda|}|\langle\tilde{\psi}_\lambda, u\rangle_{L_2(\mathrm{I})}|^2\\ &\lesssim 4^{-\ell(d-1)}\|u\|_{H^d_{w_\eta^{d-1}}(\mathrm{I})}^2.\end{aligned} \tag{7.15}$$

For $\nu \geqslant 1$, we set

$$\check{\mathbf{\nabla}}^{\eta}_{\ell,\nu} := \{\prod_{i=1}^{n} \check{\nabla}^{\eta}_{m_i} : \mathbf{m} \in \mathbb{N}_0^n,\, \nu\sum_{i=1}^{n} m_i + (1-\nu)\max_i m_i \leqslant \ell\},$$

and let $\check{\mathbf{P}}^{\eta}_{\ell,\nu}$ denote the biorthogonal projector onto $\mathrm{span}\{\boldsymbol{\psi}_{\boldsymbol{\lambda}} : \boldsymbol{\lambda} \in \check{\mathbf{\nabla}}^{\eta}_{\ell,\nu}\}$. Using (7.14) and (7.15), the arguments that led to Theorem 7.2 show that for $u \in \otimes_{i=1}^{n} H^d(\mathrm{I}) \cap H_0^1(\square)$,

$$\|u - \check{\mathbf{P}}^{\eta}_{\ell,1} u\|_{L_2(\square)} \lesssim \ell^{\frac{n-1}{2}} 2^{-d\ell} \|u\|_{\otimes_{i=1}^{n} H^d_{w^d_\eta}(\mathrm{I})}$$

and, in the current setting most importantly,

$$\|u - \check{\mathbf{P}}^{\eta}_{\ell,\nu} u\|_{H^1(\square)} \lesssim 2^{-(d-1)\ell} \sqrt{\sum_{m=1}^{n} \|u\|^2_{\otimes_{i=1}^{n} H^d_{w_\eta^{d-\delta_{mi}}}(\mathrm{I})}}$$

when $\nu \in [1, \frac{d}{d-1})$.

Since for any $\nu > 1$, $\#\check{\mathbf{\nabla}}^{\eta}_{\ell,\nu} \eqsim 2^{\ell}$ uniformly in $\eta > 0$, for proving(7.13) it remains to show that for all layer functions $V^R_{\eta,k}$

$$\sum_{m=1}^{n} \|V^R_{\eta,k}\|^2_{\otimes_{i=1}^{n} H^d_{w_\eta^{d-\delta_{mi}}}(\mathrm{I})} \lesssim \eta^{-1} |\log \eta|^{2d-1}. \tag{7.16}$$

For $k \geqslant 2$, (7.11) shows that the left-hand side is $\mathcal{O}(\eta^{k-2}|\log\eta|^{2(kd-1)})$. For $k = 1$, recalling that $\int_{c_d \eta|\log\eta|}^{\infty} |\eta^{-d} e^{-\gamma y/\eta}|^2 dy = \frac{1}{2\gamma}$, from (7.10) one infers the bound (7.16).

Let us briefly compare (7.8) with error bounds from the literature for the same problem. In Apel & Lube (1998), a Galerkin finite element method of order d is considered with respect to an anisotropic grid of Shishkin type. The upper bound for the error in $|||\cdot|||$ that is derived reads as a multiple of

$$N^{-\frac{d}{n}} + \eta^{\frac{1}{2}} |\log\eta|^{d-\frac{1}{2}} N^{-\frac{d-1}{n}}. \tag{7.17}$$

The fact that our error bound (7.8) is obtained by essentially replacing N by N^n is due to the use of tensor product approximation.

The curse of dimensionality visible in (7.17) is partly circumvented in Liu *et al.* (2009) where a *two*-scale sparse grid method is considered for problem (7.1). The upper bound for the error from that paper derived for $d = 2$ reads as a multiple of

$$N^{-\frac{2d}{n+1}} + \eta^{\frac{1}{2}} |\log\eta|^{d-\frac{1}{2}} N^{-\frac{2(d-1)}{n+1}}. \tag{7.18}$$

The results from both Apel & Lube (1998) and Liu *et al.* (2009) are derived under Assumption 7.1 with the *additional condition* that the right-hand side f of (7.1) satisfies certain compatibility conditions of sufficiently high order – for $n = 2$ the first being that the right-hand side f vanishes at the corners of $\square$ –, which ensure that $u \in C^{k,\alpha}(\overline{\square})$ for sufficiently large k. In effect, it means that in that case it can be assumed that $u^S_\eta = 0$ in (7.9) so that (7.12) does not have to be taken into account. We emphasize that our results are valid *without compatibility conditions* on the data.

Finally, we note that the rates demonstrated given by (7.8) are realized automatically by the *Adaptive* Wavelet Galerkin Method, which converges with the best possible rate in linear complexity.

8. Numerical Results

We consider the families of singularly perturbed boundary value problems

$$\begin{cases} (-\eta^2\Delta + I)u_\eta = f & \text{on } \square = (0,1)^2, \\ u_\eta = 0 & \text{on } \partial\square, \end{cases} \tag{8.1}$$

$$\begin{cases} -(\eta^2\partial_1^2 + \partial_2^2)u_\eta = f & \text{on } \square, \\ u_\eta = 0 & \text{on } \partial\square, \end{cases} \tag{8.2}$$

and

$$\begin{cases} (-\eta^2\Delta - \eta\partial_1 + I)u_\eta = f & \text{on } \square, \\ u_\eta = 0 & \text{on } \partial\square, \end{cases} \tag{8.3}$$

where $f(x,y) = (1-x)(1-y)$.

Solutions of these problems for $\eta = 10^{-2}$ are illustrated in Figure 7. Note that f does not vanish

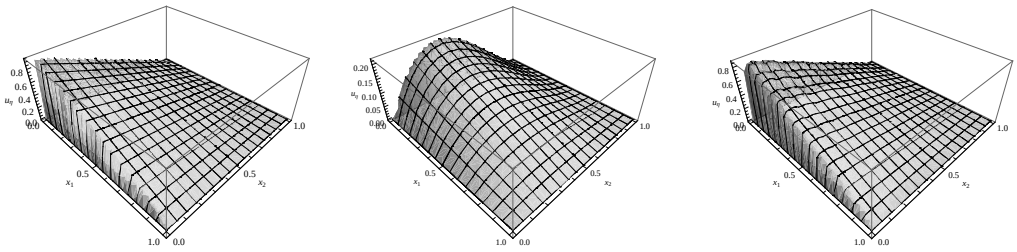

FIG. 7. (Approximate) solutions of the problems (8.1), (8.2), and (8.3) for $\eta = 10^{-2}$.

at all corners, and therefore does not satisfy compatibility conditions. So the solution u_η exhibits all possible kinds of edge and corner singularities.

In Figures 8, 9, and 10, the results are presented of the Adaptive Tensor Product Wavelet Method applied to the problems (8.1), (8.2), and (8.3). In all cases, the support length of the computed approximation $\mathbf{w}$ vs. the $\ell_2(\nabla)$-norm of its residual $\|\mathbf{f} - \mathbf{A}\mathbf{w}\|$ is given, where $\mathbf{f}$ is the representation of the right-hand side and $\mathbf{A} = \mathbf{A}_\eta$ is the bi-infinite stiffness matrix with respect to the tensor product wavelets. Since we apply the univariate wavelets from Sect. 4, the matrix $\mathbf{A}$ is *sparse*. Note that $\|\mathbf{f} - \mathbf{A}\mathbf{w}\|$ is equivalent to the error in $\mathbf{w}^T\Psi$ in the energy-norm $|||\cdot|||$ corresponding to the problem, uniformly in $\eta \geqslant 0$. For the third, non-symmetric problem, the Adaptive Tensor Product Wavelet Method is applied to the normal equations $\mathbf{A}^\top\mathbf{A}\mathbf{u} = \mathbf{A}^\top\mathbf{f}$.

The numerical results illustrate the robustness of the adaptive method with respect to the singular perturbation proven for problem (8.1), and indicate this robustness for problems (8.2) and (8.3). Since for problem (8.1), only starting from approximately 10^5 unknowns the error is decreasing with decreasing η, the results give no indication about the sharpness of the bound (7.8) as function of *N and* η.

For our convenience, we have chosen f to be a low order polynomial so that the resulting vector $\mathbf{f}$ is finitely supported. A a consequence, for $\eta = 0$, the representation of the solution $u = f$ for (8.1)

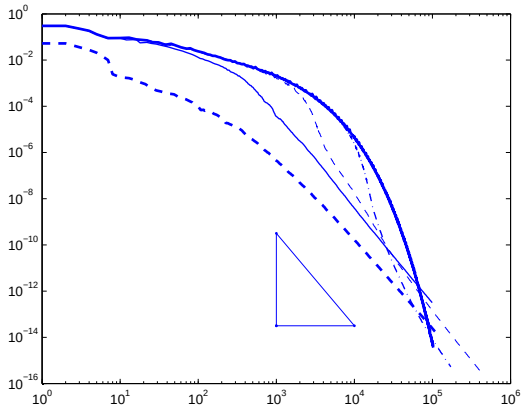

FIG. 8. #supp $\mathbf{w}$ vs. $\|\mathbf{f}-\mathbf{A}\mathbf{w}\|$ for problem (8.1) for, at height 10^{-6} from left-to-right, $\eta = 1, 10^{-2}, 10^{-4}, 10^{-8}, 0$. The slope of the triangle is the, for $\eta \neq 0$, best possible asymptotic rate -4.

and (8.3), or $u(x,y) = \frac{1}{6}(1-x)y(1-y)(y-2)$ for (8.2), is somewhat exceptionally close to a sparse one since the integral of u vanishes against all dual wavelets above the lowest level that have their supports inside the open box $(0,1)^2$. Therefore, our results for $\eta = 0$, and so for small η and N not very large, are somewhat better than generally can be expected. To get an impression of the error with a generic smooth f, the harmless approximation error for a smooth function that vanishes at the boundary should be added.

Figure 11 illustrates which wavelets were selected by the Adaptive Tensor Product Wavelet Method. To give an impression of the strength of the local refinement; for problem (8.2), $\eta = 10^{-4}$ and #supp $\mathbf{w} =$ 65635, the maximum level of a univariate wavelet present in the expansion is 29. The cardinality of the smallest "full-grid" approximation that contains this wavelet, i.e., an approximation from span$\{\boldsymbol{\psi}_{\boldsymbol{\lambda}} : |\lambda_1|, |\lambda_2| \leqslant 29\}$, is $\approx 10^{19}$.

REFERENCES

APEL, T. (1999) *Anisotropic finite elements: local estimates and applications*. Advances in Numerical Mathematics. Stuttgart: B. G. Teubner, p. 261.

APEL, T. & LUBE, G. (1998) Anisotropic mesh refinement for a singularly perturbed reaction diffusion model problem. *Appl. Numer. Math.*, **26**, 415–433.

BUNGARTZ, H. & GRIEBEL, M. (2004) Sparse grids. *Acta Numer.*, **13**, 147–269.

CARNICER, J., DAHMEN, W. & PEÑA, J. (1996) Local decomposition of refinable spaces and wavelets. *Appl. and Comp. Harm. Anal.*, **3**, 127–153.

COHEN, A., DAHMEN, W. & DEVORE, R. (2001) Adaptive wavelet methods for elliptic operator equations – Convergence rates. *Math. Comp*, **70**, 27–75.

COHEN, A., DAHMEN, W. & DEVORE, R. (2002) Adaptive wavelet methods II - Beyond the elliptic case. *Found. Comput. Math.*, **2**, 203–245.

COHEN, A. (2003) *Numerical Analysis of Wavelet Methods*. Amsterdam: Elsevier.

DAHMEN, W. (1996) Stability of multiscale transformations. *J. Fourier Anal. Appl.*, **2**, 341–362.

DAHMEN, W. & STEVENSON, R. (1999) Element-by-element construction of wavelets satisfying stability and

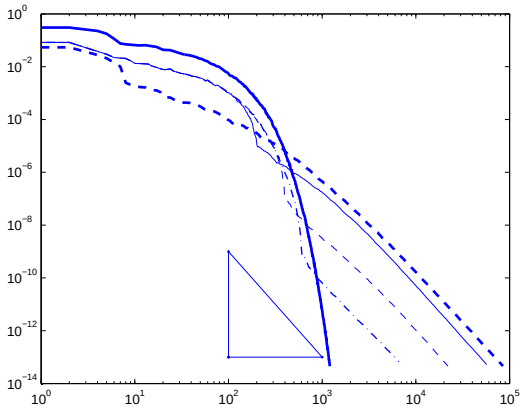

FIG. 9. #supp $\mathbf{w}$ vs. $\|\mathbf{f}-\mathbf{A}\mathbf{w}\|$ for problem (8.2) for, at height 10^{-12} from *right-to-left*, $\eta = 1, 10^{-4}, 10^{-8}, 10^{-12}, 0$. The slope of the triangle is the, for $\eta \neq 0$, best possible asymptotic rate -4.

moment conditions. *SIAM J. Numer. Anal.*, **37**, 319–352.

DAUGE, M. & STEVENSON, R. (2010) Sparse tensor product wavelet approximation of singular functions. *SIAM J. Math. Anal.*, **42**, 2203–2228.

DIJKEMA, T., SCHWAB, C. & STEVENSON, R. (2009) An adaptive wavelet method for solving high-dimensional elliptic PDEs. *Constr. Approx.*, **30**, 423–455.

DIJKEMA, T. & STEVENSON, R. (2010) A sparse Laplacian in tensor product wavelet coordinates. *Numer. Math.*, **115**, 433–449.

DONOVAN, G., GERONIMO, J. & HARDIN, D. (1996) Intertwining multiresolution analyses and the construction of piecewise-polynomial wavelets. *SIAM J. Math. Anal.*, **27**, 1791–1815.

GANTUMUR, T., HARBRECHT, H. & STEVENSON, R. (2007) An optimal adaptive wavelet method without coarsening of the iterands. *Math. Comp.*, **76**, 615–629.

GRIEBEL, M. & KNAPEK, S. (2000) Optimized tensor-product approximation spaces. *Constr. Approx.*, **16**, 525–540.

GRIEBEL, M. & OSWALD, P. (1995) Tensor product type subspace splittings and multilevel iterative methods for anisotropic problems. *Adv. Comput. Math.*, **4**, 171–206.

HAN, H. & KELLOGG, R. B. (1990) Differentiability properties of solutions of the equation $-\varepsilon^2\Delta u + ru = f(x,y)$ in a square. *SIAM J. Math. Anal.*, **21**, 394–408.

HOCHMUTH, R. (2001) Nonlinear anisotropic boundary value problems—regularity results and multiscale discretizations. *Nonlinear Anal.*, **46**, 1–18.

KELLOGG, B. (1995) Boundary layers and corner singularities for a self-adjoint problem. *Boundary value problems and integral equations in nonsmooth domains (Luminy, 1993)*. Lecture Notes in Pure and Appl. Math., vol. 167. New York: Dekker, pp. 121–149.

LIU, F., MADDEN, N., STYNES, M. & ZHOU, A. (2009) A two-scale sparse grid method for a singularly perturbed reaction-diffusion problem in two dimensions. *IMA J. Numer. Anal.*, **29**, 986–1007.

NITSCHE, P.-A. (2005) Sparse approximation of singularity functions. *Constr. Approx.*, **21**, 63–81.

NITSCHE, P.-A. (2006) Best N-term approximation spaces for tensor product wavelet bases. *Constr. Approx.*, **24**, 49–70.

SCHWAB, C. & STEVENSON, R. (2008) Adaptive wavelet algorithms for elliptic PDEs on product domains. *Math.*

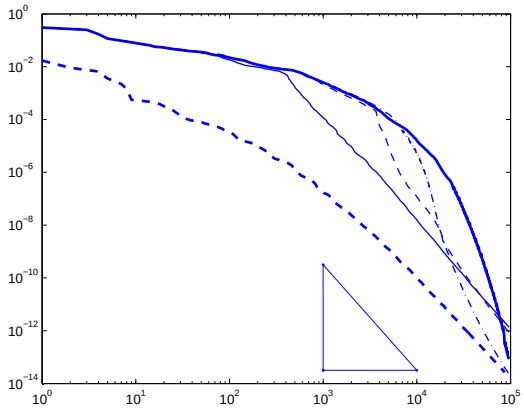

FIG. 10. #supp $\mathbf{w}$ vs. $\|\mathbf{f}-\mathbf{A}\mathbf{w}\|$ for problem (8.3) for, at height 10^{-6} from left-to-right, $\eta = 1, 10^{-2}, 10^{-4}, 10^{-8}, 0$. The slope of the triangle is the, for $\eta \neq 0$, best possible asymptotic rate -4.

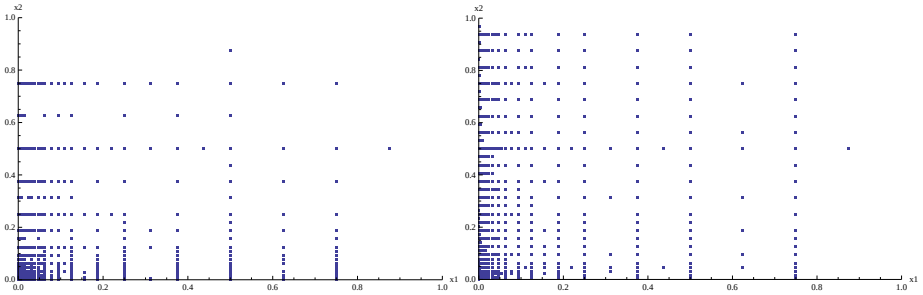

FIG. 11. Centers of the supports of the wavelets selected by the Adaptive Tensor Product Wavelet Method. Left problem (8.1), $\eta = 10^{-4}$ and #supp $\mathbf{w} = 10291$. Right problem (8.2), $\eta = 10^{-4}$ and #supp $\mathbf{w} = 10190$.

Comp., **77**, 71–92.

SHISHKIN, G. I. (1987) Approximation of solutions of singularly perturbed boundary value problems with a corner boundary layer. *Zh. Vychisl. Mat. i Mat. Fiz.*, **27**, 1360–1374, 1438.

SICKEL, W. & ULLRICH, T. (2009) Tensor products of Sobolev-Besov spaces and applications to approximation from the hyperbolic cross. *J. Approx. Theory*, **161**, 748–786.

STEVENSON, R. (2003) Locally supported, piecewise polynomial biorthogonal wavelets on non-uniform meshes. *Constr. Approx.*, **19**, 477–508.

STEVENSON, R. (2009) Adaptive wavelet methods for solving operator equations: An overview. *Multiscale, Nonlinear and Adaptive Approximation: Dedicated to Wolfgang Dahmen on the Occasion of his 60th Birthday* (R. DeVore & A. Kunoth eds). Berlin: Springer, pp. 543–598.