Eidgenössische Technische Hochschule Zürich
Ecole polytechnique fédérale de Zurich
Politecnico federale di Zurigo
Swiss Federal Institute of Technology Zurich

Wavelet Compression of Integral Operators on Sparse Tensor Spaces

Construction, Consistency and Asymptotically Optimal Complexity

Nils Reich

Research Report No. 2008-24
June 2008

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Wavelet Compression of Integral Operators on Sparse Tensor Spaces

Construction, Consistency and Asymptotically Optimal Complexity

Nils Reich

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Research Report No. 2008-24 June 2008

Abstract

For the Galerkin finite element discretization of integrodifferential equations $\mathcal{B}u = f$ on $[0,1]^n$, we present a sparse tensor product wavelet compression scheme. The scheme is of essentially optimal and dimension independent complexity $\mathcal{O}(h^{-1}|\log h|^{2(n-1)})$ without corrupting the convergence or smoothness requirements of the original sparse tensor finite element scheme. The operators under consideration are assumed to be of non-negative order and admit a standard kernel $\kappa(\cdot,\cdot)$ (singular only on the diagonal).

1 Introduction

Due to the possible non-locality of integrodifferential operators, Galerkin discretization of integrodifferential equations in general leads to linear systems with densely populated matrices of substantial size. For instance in Mathematical Finance the pricing of contracts on baskets of assets where the underlying is modeled by jump processes results in high dimensional partial integrodifferential equations (PIDEs) on $[0,1]^n$. Even on such tensor product domains, the straightforward application of standard numerical schemes fails due to the "curse of dimension": the number of degrees of freedom on a tensor product Finite Element (FE) mesh of width h in dimension n grows like $\mathcal{O}(h^{-n})$ as $h \to 0$. The non-locality of the underlying operator thus implies that the stiffness matrix of the Galerkin discretization based on tensor product FE spaces consists of $\mathcal{O}(h^{-2n})$ non-zero entries.

In this work we show that, using tensor products of univariate wavelet basis functions, the complexity of the stiffness matrix can be reduced to $\mathcal{O}(h^{-1}|\log h|^{2(n-1)})$ without corrupting the rate of convergence and smoothness requirements of the original FE scheme.

Up to now, several approaches to resolve the curse of dimension and the non-locality have been introduced. Two of them are of special importance for our analysis:

Firstly, to overcome the exponential growth of complexity, *sparse tensor product* spaces were introduced at the end of the 1990s, see e.g. [3, 16, 17, 25] and the references therein. This methodology yields $\mathcal{O}(h^{-1}|\log h|^{n-1})$ degrees of freedom as $h \to 0$ while at the same time (essentially) preserving the approximation rate. Discretizing integrodifferential equations on a sparse tensor product space thus yields matrices containing $\mathcal{O}(h^{-2}|\log h|^{2(n-1)})$ entries. As shown in e.g. [17], these results require greater smoothness of the function to be approximated than the original discretization and this extra regularity increases with the dimension n.

Secondly, to cope with the non-locality of integral operators, in the very different setting of *isotropic* (or *standard*) wavelet representation, i.e. the FE basis functions consist of tensor products of scaling functions and wavelets only on the *same* level, so-called wavelet compression has been introduced in [1]. There it was shown that wavelet representation yields an almost sparse representation of certain operators. In [9, 10, 39, 37] this approach was advanced further (on not necessarily tensor product domains) and given a rigorous mathematical foundation based on the requisite that the compressed system has to preserve the stability and convergence properties of the unperturbed discretization. In [30] it was shown that wavelet compression techniques may yield asymptotically optimal complexity (on not necessarily tensor product domains) in the sense that the number of non-zero entries in the resulting matrices grows linearly with the number of degrees of freedom. In contrast to sparse tensor product approximation, this methodology does not require additional smoothness of the approximated function. But, since the number of non-zero matrix entries grows linearly with the degrees of freedom, there still is exponential growth of the number of non-trivial matrix entries as the dimension n tends to infinity. The results on isotropic wavelet compression have been unified in a sophisticated way in [7]. Since it somewhat presents a finalization of the isotropic wavelet compression, we refer to [7] for a more detailed description of the development in this field. Note that, with a slightly different approach but based on analogous principles similar complexity results for the isotropic setting have been presented in [34]. In summary, substituting $h = 2^{-J}$, there holds:

- Using *sparse* tensor product spaces one obtains $\mathcal{O}(2^{2J}J^{2(n-1)})$ non-zero entries in the system matrix.

- Wavelet compression of general *full* tensor product spaces yields $\mathcal{O}(2^{nJ})$ non-zero matrix entries.

In this work the notion of computational "complexity" is used exclusively to indicate the number of non-zero entries in a given system matrix. With an efficient implementation and quadrature as in [19] it can be shown that the overall cost of computing and assembling the system matrix is essentially of the same magnitude as its complexity.

Note that the complexity results of this work also imply that, under certain conditions, the stiffness matrices of the anisotropic non-local operators under consideration are s^*-compressible in the sense of [5, 15, 31] with essentially *dimension independent* s^*. This shows that, in order to solve the corresponding integrodifferential equations one may employ *adaptive* wavelet algorithms as in [4, 5, 14] that converge with the rate of best approximation by an arbitrary linear combination of N wavelets (so-called best N-term approximation).

The outline of this work is as follows:

In Section 3 we derive fundamental estimates of the entries in the sparse tensor product-based matrix. These estimates form the basis for the consistency analysis and hence enable us to define cut-off parameters. The estimates rely on techniques developed for isotropic wavelets by [7, 9, 30, 37]. We intensively exploit the tensor product structure of our wavelets to reduce our considerations to their one-dimensional counterparts.

Section 4 defines the framework on which the compression is based. In this section we provide consistency requirements that have to be satisfied by the compressed matrix in order to preserve stability and convergence properties of the sparse tensor product setting without compression.

Combining the results from the two previous sections, in Section 5 the actual cut-off parameters are defined. The scheme also exploits the tensor product structure in the sense that all dropping criteria of matrix entries are given in terms of the wavelets defined in one particular coordinate direction.

In Section 6 we provide complexity results for the constructed compression schemes. Based on [28], we show that the complexity of the sparse tensor product setting can be reduced to $\mathcal{O}\left(2^J J^{2(n-1)}\right)$ non-zero matrix entries provided that the number of vanishing moments or the order of the underlying operator is sufficiently large.

Finally, in Section 7 we briefly illustrate how the results of the previous sections imply s^*-compressibility.

2 Galerkin discretization of multidimensional PIDEs

On the n-dimensional unit cube $\square := [0,1]^n$, we consider an integrodifferential equation

$$\mathcal{B}u = f, \tag{2.1}$$

with an integrodifferential operator

$$\mathcal{B} = \mathcal{A}_D + \mathcal{A}, \tag{2.2}$$

where $\mathcal{A}_D$ denotes a (possibly vanishing) differential operator

$$\mathcal{A}_D u = -\frac{1}{2}\sum_{i,j=1}^{n} \mathcal{Q}_{ij}\frac{\partial^2 u}{\partial x_i \partial x_j}, \qquad \mathcal{Q}_{ij} \in \mathbb{R},\ i,j = 1,\dots,n, \tag{2.3}$$

and $\mathcal{A}$ is an integral operator operator of order $2q \in \mathbb{R}$ with symbol in the Hörmander class $S^{2q}_{1,0}$ in the sense of [20, 35]. It is well known that $\mathcal{A}$ acts continuously on Sobolev spaces and admits a kernel representation

$$\mathcal{A}u(x) = \int_{\square} \kappa(x,y)u(y)dy, \tag{2.4}$$

with a distributional kernel function $\kappa(\cdot,\cdot)$. By the Schwartz kernel theorem (cf. e.g. [33, Section VI.7]), the kernel $\kappa(\cdot,\cdot)$ is singular only on the diagonal in $[0,1]^n \times [0,1]^n$ and for any $\underline{\sigma}$, $\underline{\sigma}' \in \mathbb{N}_0^n$ there holds

$$\left|\partial_x^{\underline{\sigma}}\partial_y^{\underline{\sigma}'}\kappa(x,y)\right| \leq c_{\underline{\sigma},\underline{\sigma}'}|x-y|^{-(n+2q+|\underline{\sigma}|+|\underline{\sigma}'|)}, \qquad \text{for all } x,y \in [0,1]^n, \tag{2.5}$$

with some constant $c_{\underline{\sigma},\underline{\sigma}'}$ independent of x, $y \in [0,1]^n$.

Denoting by $\mathcal{Q} = (\mathcal{Q}_{ij})_{1\leq i,j\leq n}$ the coefficient matrix of the differential operator $\mathcal{A}_D$ in (2.3) we shall assume that either $\mathcal{Q} = 0$ or $\mathcal{Q} > 0$. The order $2\widetilde{q} \in \mathbb{R}$ of the integrodifferential operator $\mathcal{B} = \mathcal{A}_D + \mathcal{A}$ is then given by

$$2\widetilde{q} = \begin{cases} 2, & \text{if } \mathcal{Q} > 0 \text{ and } 2q \leq 2, \\ 2q, & \text{otherwise.} \end{cases} \tag{2.6}$$

Corresponding to the operator order $2\widetilde{q} \in \mathbb{R}$ we define the Sobolev spaces $H^{\widetilde{q}}(\square)$ as follows: for $u \in C_0^\infty(\square)$, define $\bar{u}$ to be the zero extension of u to all of $\mathbb{R}^n$. Then, for $s \in \mathbb{R}$, the space $H^s(\square)$ is given by

$$H^s(\square) := \overline{\{\bar{u} \,|\, u \in C_0^\infty(\square)\}}, \tag{2.7}$$

where the closure is taken with respect to the norm of $H^s(\mathbb{R}^n)$, the classical Sobolev space on $\mathbb{R}^n$.

For the numerical solution of (2.1), we employ the Galerkin method with respect to a hierarchy of conforming trial spaces $\widehat{V}_J \subset \widehat{V}_{J+1} \subset \ldots \subset H^{\widetilde{q}}(\square)$. The variational problem of interest reads: find $u_J \in \widehat{V}_J$ such that,

$$\langle \mathcal{B}u_J,\, v_J\rangle = \langle f,\, v_J\rangle \ \text{ for all } \ v_J \in \widehat{V}_J. \tag{2.8}$$

The index J represents the meshwidth of order 2^{-J}. We shall make the following assumptions on the operator $\mathcal{B}$ to ensure that the variational problem (2.8) is well posed - for details we refer to e.g. [32, Proposition III.2.3].

1. $\mathcal{B}$ satisfies a Gårding inequality, i.e. there exist constants $\gamma > 0$, $C \geq 0$ such that

$$\langle \mathcal{B}u, u\rangle \geq \gamma\|u\|^2_{H^{\widetilde{q}}(\square)} - C\|u\|^2_{L^2(\square)}, \qquad \text{for all } u \in H^{\widetilde{q}}(\square). \tag{2.9}$$

2. $\mathcal{B} : H^{\widetilde{q}}(\square) \to H^{-\widetilde{q}}(\square)$ is continuous, i.e. there exists a constant $C' > 0$ such that for all $u, v \in H^{\widetilde{q}}(\square)$ there holds

$$|\langle \mathcal{B}u, v\rangle| \leq C'\|u\|_{H^{\widetilde{q}}(\square)}\|v\|_{H^{\widetilde{q}}(\square)}. \tag{2.10}$$

The nested trial spaces $\widehat{V}_J \subset \widehat{V}_{J+1}$ we employ in (2.8) shall be sparse tensor product spaces based on a wavelet multiresolution analysis described in the next sections.

We shall frequently write $a \lesssim b$ to express that a is bounded by a constant multiple of b, uniformly with respect to all parameters on which a and b may depend. Then $a \sim b$ means $a \lesssim b$ and $b \lesssim a$.

2.1 Wavelets on the interval

On $[0,1]$ we shall use the same scaling functions and wavelets as described in [7] based on the construction of [6, 8, 23] and the references therein.

Denoting by Δ_j some suitable index sets, the trial spaces $\mathcal{V}_j$ are spanned by single-scale bases $\Phi_j = \{\phi_{j,k} : k \in \Delta_j\}$. The approximation order of the trial spaces we denote by d, i.e.

$$d = \sup\left\{ s \in \mathbb{R} : \sup_{j\geq 0} \left\{ \frac{\inf_{v_j\in\mathcal{V}_j} \|v - v_j\|_0}{2^{-js}\|v\|_s} \right\} < \infty\,,\ \forall\, v \in H^s([0,1]) \right\}. \tag{2.11}$$

Using the single-scale bases constructed in [6] based on B-splines adapted to the interval $[0,1]$ as described in [8], we assume that for each $j \geq 0$, the basis functions $\phi_{j,k} \in \Phi_j$ have compact supports and admit two important properties: $\|\phi_{j,k}\|_{L^2([0,1])} = 1$ and $|\operatorname{supp}\, \phi_{j,k}| \sim 2^{-j}$.

Associated to these *primal* bases are *dual* bases $\widetilde{\Phi}_j = \{\widetilde{\phi}_{j,k} : k \in \Delta_j\}$, i.e. there holds $\langle \phi_{j,k}, \widetilde{\phi}_{j,k'} \rangle = \delta_{k,k'}$. By $\widetilde{d}$ we denote the order of $\widetilde{\Phi}_j$ and assume $d \leq \widetilde{d}$. In particular, for B-splines of order d and duals of order $\widetilde{d} \geq d$ such that $d + \widetilde{d}$ is even the bases Φ_j, $\widetilde{\Phi}_j$ as in [8] have approximation orders d and $\widetilde{d}$.

To these single-scale bases there exist biorthogonal complement or wavelet bases $\Psi_j = \{\psi_{j,k} : k \in \nabla_j\}$, $\widetilde{\Psi}_j = \{\widetilde{\psi}_{j,k} : k \in \nabla_j\}$, where $\nabla_j := \Delta_{j+1} \backslash \Delta_j$. Inherited from $\phi_{j,k}$, the wavelets $\psi_{j,k}$ have compact supports and there holds

$$|\operatorname{supp}\, \psi_{j,k}| \sim 2^{-j}. \tag{2.12}$$

The dual pair of wavelet bases Ψ, $\widetilde{\Psi}$ is defined by $\Psi = \bigcup_{j\geq 0} \Psi_j$, $\widetilde{\Psi} = \bigcup_{j\geq 0} \widetilde{\Psi}_j$, with $\Psi_0 := \Phi_0$, $\widetilde{\Psi}_0 := \widetilde{\Phi}_0$. There holds

$$\|\psi_{j,k}\|_{L^2([0,1])} \sim 1, \qquad \text{for all } \psi_{j,k} \in \Psi.$$

From the biorthogonality of Ψ and $\widetilde{\Psi}$ one infers the so-called *cancelation property* of Ψ (see e.g. [2]), i.e.

$$|\langle \psi_{j,k},\, f\rangle| \lesssim 2^{-j(\widetilde{d}+1/2)} |f|_{W^{\widetilde{d},\infty}(\operatorname{supp}\psi_{j,k})}, \qquad \text{for each } \psi_{j,k} \in \Psi. \tag{2.13}$$

Here $|f|_{W^{\widetilde{d},\infty}(\Omega)} := \sup_{x\in\Omega} |\partial^{\widetilde{d}} f(x)|$. The *mother wavelet* of Ψ we denote by ψ, i.e. for any j and $k \in \nabla_j$,

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), \qquad x \in [0,1]. \tag{2.14}$$

Denoting by $\mathcal{W}_j$, $\widetilde{\mathcal{W}}_j$ the span of Ψ_j, $\widetilde{\Psi}_j$, there holds

$$\mathcal{V}_{j+1} = \mathcal{W}_{j+1} \oplus \mathcal{V}_j, \ \text{ and } \widetilde{\mathcal{V}}_{j+1} = \widetilde{\mathcal{W}}_{j+1} \oplus \widetilde{\mathcal{V}}_j, \ \text{ for all } j \geq 0, \tag{2.15}$$

and,

$$\mathcal{V}_j = \mathcal{W}_0 \oplus \ldots \oplus \mathcal{W}_j, \ \text{ for all } j \geq 0. \tag{2.16}$$

Crucial for the consistency of our compression scheme is the fact that the wavelets on $[0,1]$ satisfy the following norm estimates (cf. e.g. [8, 11], for the one-sided estimates we refer to [39]):

For an arbitrary $u \in H^t([0,1])$, $0 \leq t \leq d$, with wavelet decomposition

$$u = \sum_{j=0}^{\infty} \sum_{k\in\nabla_j} u_{j,k}\psi_{j,k} = \sum_{j=0}^{\infty} \sum_{k\in\nabla_j} \langle u,\, \widetilde{\psi}_{j,k}\rangle \psi_{j,k},$$

there holds the norm equivalence,

$$\sum_{(j,k)} 2^{2tj}|u_{j,k}|^2 \sim \|u\|^2_{H^t([0,1])}\,, \quad \text{if } 0 \le t < d - 1/2, \tag{2.17}$$

or the one-sided estimate,

$$\sum_{(j,k)} 2^{2tj}|u_{j,k}|^2 \lesssim \|u\|^2_{H^t([0,1])}\,, \quad \text{if } d - 1/2 \le t < d. \tag{2.18}$$

In case $t = d$ there only holds,

$$\sum_{\substack{(j,k)\\ j\le J}} 2^{2tj}|u_{j,k}|^2 \lesssim J\,\|u\|^2_{H^t([0,1])}\,, \quad \text{if } t = d. \tag{2.19}$$

We conclude this section by an explicit example of wavelets on $[0,1]$ with approximation order $d = 2$:

Example 2.1. *The wavelets comprise of piecewise linear continuous functions on $[0,1]$ vanishing at the endpoints. The mesh for level $j \ge 0$ is defined by the nodes $x_{j,k} := k2^{-(j+1)}$ with $k \in \nabla_j := \{0,\dots,2^{j+1}\}$. There holds $N_j := \dim \mathcal{V}_j = 2^{j+1} - 1$ and therefore $M_j := \dim \mathcal{V}_j - \dim \mathcal{V}_{j-1} = 2^j$.*

On level $j = 0$ we have $N_0 = M_0 = 1$ and $\psi_{0,1}$ is defined as the piecewise linear function with value $c_0 > 0$ at $x_{0,1} = \frac{1}{2}$ and 0 at the endpoints 0, 1.

For $j > 0$ we firstly define $c_j := 2^{j/2}$. Then the wavelet $\psi_{j,1}$ is defined as the piecewise linear function such that $\psi_{j,1}(x_{j,1}) = 2c_j$, $\psi_{j,1}(x_{j,2}) = -c_j$ and $\psi_{j,1}(x_{j,s}) = 0$ for all other $s \ne 1,\, 2$. Similarly, the wavelet ψ_{j,M_j} takes the values $\psi_{j,M_j}(x_{j,N_j}) = 2c_j$, $\psi_{j,M_j}(x_{j,N_j-1}) = -c_j$ and zero at all other nodes. For $1 < k < M_j$ the wavelet $\psi_{j,k}$ is defined by $\psi_{j,k}(x_{j,2k-2}) = -c_j$, $\psi_{j,k}(x_{j,2k-1}) = 2c_j$, $\psi_{j,k}(x_{j,2k}) = -c_j$ and $\psi_{j,k}(x_{j,s}) = 0$ for all other $s \ne 2k-2,\, 2k-1,\, 2k$.

Remark 2.2. *Note that the analysis of the so-called second compression that we adapt from [7, 30] refers exclusively to biorthogonal spline wavelets whose singular supports are well defined and not dense in the wavelets' supports. We refer to [18] for more specific illustrations.*

2.2 Sparse tensor product spaces

For $x = (x_1,\dots,x_n) \in [0,1]^n$, we denote,

$$\psi_{\mathbf{j},\mathbf{k}}(x) := \psi_{j_1,k_1} \otimes \ldots \otimes \psi_{j_n,k_n}(x_1,\dots,x_n) = \psi_{j_1,k_1}(x_1)\ldots\psi_{j_n,k_n}(x_n).$$

Using Fubini's theorem one infers that the scaling and cancelation properties (2.12), (2.13) of the univariate wavelets carry forward to their tensor products. In particular,

$$\big|\mathrm{supp}\ \psi_{\mathbf{j},\mathbf{k}}\big| = \prod_{i=1}^{n} |\mathrm{supp}\ \psi_{j_i,k_i}| \sim 2^{-(j_1+\ldots+j_n)},$$

and each $\psi_{\mathbf{j},\mathbf{k}}$ has $\widetilde{d}$ vanishing moments which implies the cancelation property

$$|\langle v, \psi_{\mathbf{j},\mathbf{k}}\rangle| \lesssim 2^{-\frac{1}{2}|\mathbf{j}|_1} 2^{-\widetilde{d}\max\{j_1,\dots,j_n\}} |v|_{W^{\widetilde{d},\infty}(\mathrm{supp}\,\psi_{\mathbf{j},\mathbf{k}})}. \tag{2.20}$$

On $[0,1]^n$, we define the subspace $V_J \subset H^q(\square)$ as the (full) tensor product of the spaces defined on $[0,1]$

$$V_J := \bigotimes_{i=1}^{n} \mathcal{V}_J, \tag{2.21}$$

which can be written using (2.16) as

$$V_J = \operatorname{span}\left\{\psi_{\mathbf{j},\mathbf{k}} \,:\, k_i \in \nabla_{j_i},\, 0 \le j_i \le J,\, i = 1,\dots,n\right\} = \sum_{j_1,\dots,j_n=0}^{J} \mathcal{W}_{j_1} \otimes \ldots \otimes \mathcal{W}_{j_n}.$$

We define the regularity $\gamma > \widetilde{q}$ of the trial spaces by

$$\gamma = \sup\left\{s \in \mathbb{R} \,:\, V_J \subset H^s(\square)\right\}. \tag{2.22}$$

It is known that based on the spline wavelets constructed in Example 2.1 the regularity index satisfies $\gamma = d - 1/2$.

The sparse tensor product spaces $\widehat{V}_J$ are now defined by

$$\widehat{V}_J := \operatorname{span}\left\{\psi_{\mathbf{j},\mathbf{k}} \,:\, k_i \in \nabla_{j_i},\, i = 1,\dots,n;\, 0 \le |\mathbf{j}|_1 \le J\right\} = \sum_{0 \le |\mathbf{j}|_1 \le J} \mathcal{W}_{j_1} \otimes \ldots \otimes \mathcal{W}_{j_n}. \tag{2.23}$$

One readily infers that $N_J := \dim(V_J) = \mathcal{O}(2^{nJ})$ whereas $\widehat{N}_J := \dim(\widehat{V}_J) = \mathcal{O}(2^J J^{n-1})$ as J tends to infinity. However, both spaces have similar approximation properties in terms of the Finite Element meshwidth $h = 2^{-J}$, provided the function to be approximated is sufficiently smooth. To characterize the necessary extra smoothness we introduce the spaces $\mathcal{H}^s([0,1]^n)$, $s \in \mathbb{N}_0$, of all measurable functions $u : [0,1]^n \to \mathbb{R}$, such that the norm,

$$\|u\|_{\mathcal{H}^s(\square)} := \Bigg(\sum_{\substack{0 \le \alpha_i \le s,\\ i=1,\dots,n}} \|\partial_1^{\alpha_1} \ldots \partial_n^{\alpha_n} u\|^2_{L^2(\square)}\Bigg)^{1/2},$$

is finite. That is

$$\mathcal{H}^s([0,1]^n) = \bigotimes_{i=1}^{n} H^s([0,1]). \tag{2.24}$$

For arbitrary $s \in \mathbb{R}_{\ge 0}$, we define $\mathcal{H}^s$ by interpolation. Because of the underlying tensor product structure (2.24), one infers from (2.17)–(2.19) that for

$$u = \sum_{(\mathbf{j},\mathbf{k})} u_{\mathbf{j},\mathbf{k}} \psi_{\mathbf{j},\mathbf{k}} = \sum_{(\mathbf{j},\mathbf{k})} u_{\mathbf{j},\mathbf{k}} \psi_{j_1 k_1} \otimes \ldots \otimes \psi_{j_n k_n},$$

there holds the norm equivalence

$$\sum_{(\mathbf{j},\mathbf{k})} 2^{2s(j_1+\ldots+j_n)} |u_{\mathbf{j},\mathbf{k}}|^2 \sim \|u\|^2_{\mathcal{H}^s}, \quad \text{if } 0 \le s < d - 1/2, \tag{2.25}$$

and the one-sided bounds

$$\sum_{(\mathbf{j},\mathbf{k})} 2^{2s(j_1+\ldots+j_n)} |u_{\mathbf{j},\mathbf{k}}|^2 \lesssim \|u\|^2_{\mathcal{H}^s}, \quad \text{if } 0 \le s < d, \tag{2.26}$$

$$\sum_{\mathbf{k}} 2^{2s(j_1+\ldots+j_n)}|u_{\mathbf{j},\mathbf{k}}|^2 \quad \lesssim \quad \|u\|^2_{\mathcal{H}^s}\,, \quad \text{if } s=d. \tag{2.27}$$

By (2.21), one may decompose any $u \in L^2(\square)$ into

$$u(x) \quad = \quad \sum_{\substack{j_i\geq 0\\ i=1,\ldots,n}} \sum_{k_i\in\nabla_{j_i}} u_{\mathbf{j},\mathbf{k}}\psi_{\mathbf{j},\mathbf{k}}(x) = \sum_{\substack{j_i\geq 0\\ i=1,\ldots,n}} \sum_{k_i\in\nabla_{j_i}} u_{\mathbf{j},\mathbf{k}}\psi_{j_1,k_1}(x_1)\ldots\psi_{j_n,k_n}(x_n).$$

In this style, the sparse grid projection $\widehat{P}_J : L^2(\square) \to \widehat{V}_J$ is defined by truncation of the wavelet expansion:

$$(\widehat{P}_J u)(x) := \sum_{0\leq|\mathbf{j}|_1\leq J} \sum_{\mathbf{k}\in\nabla_{\mathbf{j}}} u_{\mathbf{j},\mathbf{k}}\psi_{\mathbf{j},\mathbf{k}}(x), \tag{2.28}$$

where $\nabla_{\mathbf{j}} = \nabla_{(j_1,\ldots,j_n)} := \nabla_{j_1}\times\ldots\times\nabla_{j_n}$.

2.3 Approximation rates for sparse tensor product spaces

For the proof of the following convergence and stability results, we refer to [38] and [17], respectively. By [38, Propositions 3.1 & 3.2], the sparse tensor product projection $\widehat{P}_J$ in (2.28) satisfies

Lemma 2.3. *Suppose $0 \leq t < \gamma$, then for $u \in H^t$, there holds*

1. *Stability of $\widehat{P}_J$:*
$$\|\widehat{P}_J u\|_{H^t(\square)} \lesssim \|u\|_{H^t(\square)}. \tag{2.29}$$

2. *Approximation property of $\widehat{P}_J$: for $0 \leq t < \gamma$ and $t < t' \leq d$,*
$$\|u - \widehat{P}_J u\|_{H^t(\square)} \lesssim \begin{cases} 2^{-Jd}J^{\frac{1}{2}}\|u\|_{\mathcal{H}^d(\square)} & \text{if } t=0 \text{ and } t'=d,\\ 2^{J(t-t')}\|u\|_{\mathcal{H}^{t'}(\square)} & \text{otherwise.}\end{cases} \tag{2.30}$$

Herewith, we can summarize:

Proposition 2.4. *Let $\mathcal{A}$ in* (2.2) *be an operator of order $2q \in \mathbb{R}$ as in* (2.4)*. Denote the order of $\mathcal{B}$ in* (2.2) *by $2\widetilde{q}$ as defined in* (2.6)*. Then the sparse tensor product spaces $\widehat{V}_J$ in* (2.23) *based on the wavelets introduced in Section 2.1 satisfy:*

1. *The Galerkin discretization of (2.1) based on sparse tensor product spaces $\widehat{V}_J$ as defined in (2.23) is stable, i.e. there exist $J_0 > 0$ and $c_1 > 0$, $c_2 \geq 0$ such that for any $J \geq J_0$ there holds*
$$|\langle \mathcal{B}v_J,\, v_J\rangle| \geq c_1\|v_J\|^2_{H^{\widetilde{q}}(\square)} - c_2\|v_J\|^2_{L^2(\square)}, \qquad \text{for all } v_J \in \widehat{V}_J, \tag{2.31}$$
and there exists some $c_3 > 0$ such that for all $J \geq J_0$,
$$|\langle \mathcal{B}v_J,\, w_J\rangle| \leq c_3\|v_J\|_{H^{\widetilde{q}}(\square)}\|w_J\|_{H^{\widetilde{q}}(\square)}, \qquad \text{for all } v_J, w_J \in \widehat{V}_J. \tag{2.32}$$
In particular, the variational problem (2.8) admits a unique solution.

2. *For $\widetilde{q} < \gamma$ and $\widetilde{q} < t' \leq d$, the convergence of the sparse tensor product Galerkin scheme is determined by*
$$\|u - u_J\|_{H^{\widetilde{q}}(\square)} \lesssim 2^{J(\widetilde{q}-t'+\nu)} \|u\|_{\mathcal{H}^{t'}(\square)}, \tag{2.33}$$
where u, u_J denote the solutions of the original equation $\mathcal{B}u = f$ and the variational problem (2.8), respectively. Here
$$\nu = \begin{cases} \dfrac{(n-1)d}{nd-1}, & \text{if } \widetilde{q} = 0 \text{ and } t' = d, \\ 0, & \text{otherwise}. \end{cases}$$

Proof. The stability estimates (2.31), (2.32) are satisfied on the full tensor product space V_J. This carries forward to the sparse tensor product space $\widehat{V}_J$ with sufficiently large $J > 0$, because $\widehat{V}_J \subset V_J$ and $\bigcup_{J=0}^{\infty} \widehat{V}_J$ is dense in $H^{\widetilde{q}}(\square)$. To see the latter, note that $V_{J/2} \subset \widehat{V}_J$ and $\bigcup_{J=0}^{\infty} V_{J/2} = \bigcup_{J=0}^{\infty} V_J$ is dense in $H^{\widetilde{q}}(\square)$.

The convergence rate (2.33) in the sparse tensor product setting can be obtained directly from [38, Section 3.5]. □

3 Fundamental estimates

In this section we derive fundamental estimates for the entries of the stiffness matrix $\mathbb{A}_J$ of $\mathcal{A}$, i.e.
$$[\mathbb{A}_J]_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')} = \langle \mathcal{A}\psi_{\mathbf{j},\mathbf{k}},\, \psi_{\mathbf{j}',\mathbf{k}'} \rangle,$$
for $0 \leq |\mathbf{j}|_1, |\mathbf{j}'|_1 \leq J$, $\mathbf{k} \in \nabla_{\mathbf{j}}$, $\mathbf{k}' \in \nabla_{\mathbf{j}'}$. Throughout this section, we consider an arbitrary but fixed pair of n-variate tensor product wavelets $\psi_{\mathbf{j},\mathbf{k}} = \psi_{j_1,k_1} \otimes \ldots \otimes \psi_{j_n,k_n}$, $\psi_{\mathbf{j}',\mathbf{k}'} = \psi_{j_1',k_1'} \otimes \ldots \otimes \psi_{j_n',k_n'}$. For any coordinate direction $s = 1, \ldots, n$, we denote
$$\delta_{x_s} := \operatorname{dist}\left(\operatorname{supp}\{\psi_{j_s,k_s}\},\, \operatorname{supp}\{\psi_{j_s',k_s'}\}\right),$$
and
$$\sigma_{x_s} := \begin{cases} \operatorname{dist}\left(\operatorname{singsupp}\{\psi_{j_s,k_s}\},\, \operatorname{supp}\{\psi_{j_s',k_s'}\}\right), \text{ if } j_s \leq j_s', \\ \operatorname{dist}\left(\operatorname{supp}\{\psi_{j_s,k_s}\},\, \operatorname{singsupp}\{\psi_{j_s',k_s'}\}\right), \text{ if } j_s' \leq j_s. \end{cases}$$
In addition, we set
$$\delta_{xy} := \operatorname{dist}\left(\operatorname{supp}\{\psi_{\mathbf{j},\mathbf{k}}\},\, \operatorname{supp}\{\psi_{\mathbf{j}',\mathbf{k}'}\}\right).$$
Usually, we simply write δ_{x_s}, σ_{x_s} without indicating the dependence of these terms on $\mathbf{j}$, $\mathbf{j}'$, $\mathbf{k}$ and $\mathbf{k}'$, because this is clear from the context.

3.1 Auxiliary wavelet-based estimates

Before we give the actual estimates for the matrix entries we need to collect the following lemmas.

Lemma 3.1. *Let $i \in \{1, \ldots, n\}$ denote a coordinate direction. Consider a linear continuous operator*
$$\mathcal{A}_i : H^{m/2}([0,1]) \to H^{-m/2}([0,1]),$$

of order $m \in \mathbb{R}$. Suppose there exists some given function $c_i(\mathbf{j},\mathbf{j}')$ that may depend on all level indices $\mathbf{j},\mathbf{j}'$ except $j_i,\ j_i'$ and some universal constant $c>0$ independent of the underlying wavelets such that

$$|\mathcal{A}_i\psi_{j_i',k_i'}(x)| \le c\cdot c_i(\mathbf{j},\mathbf{j}')\cdot 2^{-j_i'(\widetilde{d}+1/2)}\,\mathrm{dist}(\mathrm{supp}(\psi_{j_i',k_i'}),\ x)^{-(1+m+\widetilde{d})}, \tag{3.1}$$

for all $x\in[0,1]$. Then for any $g\in L^\infty([0,1])$ with $\mathrm{supp}(g)\cap\mathrm{supp}(\psi_{j_i',k_i'})=\emptyset$ there holds

$$\begin{aligned} &|\langle \mathcal{A}_i\psi_{j_i',k_i'},\ g\rangle| \\ &\lesssim c_i(\mathbf{j},\mathbf{j}')\cdot\|g\|_{L^\infty([0,1])}2^{-j_i'(\widetilde{d}+1/2)}\,\mathrm{dist}(\mathrm{supp}(\psi_{j_i',k_i'}),\mathrm{supp}(g))^{-(m+\widetilde{d})}. \end{aligned} \tag{3.2}$$

Proof. Denote $S_g:=\mathrm{supp}(g)$. Applying (3.1) one obtains

$$\begin{aligned} |\langle \mathcal{A}_i\psi_{j_i',k_i'},\ g\rangle| &\lesssim \|g\|_{L^\infty([0,1])}\int_{S_g}|\mathcal{A}_i\psi_{j_i',k_i'}(x)|dx \\ &\lesssim c_i(\mathbf{j},\mathbf{j}')\cdot\|g\|_{L^\infty([0,1])}2^{-j_i'(\widetilde{d}+1/2)}\int_{S_g}\mathrm{dist}(\mathrm{supp}(\psi_{j_i',k_i'}),\ x)^{-(1+m+\widetilde{d})}dx. \end{aligned}$$

Since S_g and $\mathrm{supp}(\psi_{j_i',k_i'})$ are disjoint,

$$\int_{S_g}\mathrm{dist}(\mathrm{supp}(\psi_{j_i',k_i'}),\ x)^{-(1+m+\widetilde{d})}dx \lesssim \mathrm{dist}(\mathrm{supp}(\psi_{j_i',k_i'}),S_g)^{-(m+\widetilde{d})},$$

and hence,

$$|\langle \mathcal{A}_i\psi_{j_i',k_i'},\ g\rangle| \lesssim\ c_i(\mathbf{j},\mathbf{j}')\cdot\|g\|_{L^\infty([0,1])}2^{-j_i'(\widetilde{d}+1/2)}\,\mathrm{dist}(\mathrm{supp}(\psi_{j_i',k_i'}),S_g)^{-(m+\widetilde{d})}.$$

Note that for n-variate isotropic wavelets a similar estimate has already been shown in [7, Lemma 6.4]. ☐

Remark 3.2. *To obtain the desired matrix entry estimates, in the following section we set $c_i(\mathbf{j},\mathbf{j}') = 2^{-\frac{1}{2}(|\mathbf{j}|_1+|\mathbf{j}'|_1-j_i-j_i')}$ which is independent of j_i, j_i'.*

Lemma 3.3. *For a smooth piece f (not necessarily defined on the whole $[0,1]$) of the wavelet ψ_{j_i,k_i} denote by $\overline{f}\in C_0^\infty(\mathbb{R})$ a smooth extension of f to the whole of $\mathbb{R}$. There holds*

$$\|\overline{f}\|_{W^{\infty,\widetilde{d}+m}(\mathrm{supp}\,\psi_{j_i',k_i'})} \lesssim 2^{j_i(\frac{1}{2}+\widetilde{d}+m)}. \tag{3.3}$$

Proof. Using integration by parts, one obtains

$$\begin{aligned} \|\overline{f}\|_{W^{\infty,\widetilde{d}+m}(\mathrm{supp}\,\psi_{j_i',k_i'})} &\lesssim \|2^{j_i/2}\psi(2^{j_i}\cdot -k_i)\|_{W^{\infty,\widetilde{d}+m}([0,1])} \\ &\lesssim 2^{j_i(\frac{1}{2}+\widetilde{d}+m-1)}\|\psi^{(-(\widetilde{d}+m))}(2^{j_i}\cdot -k_i)\|_{L^\infty([0,1])} \\ &\lesssim 2^{j_i(\frac{1}{2}+\widetilde{d}+m)}. \end{aligned}$$

Here ψ denotes the mother wavelet as in (2.14) and $\psi^{(-(\widetilde{d}+m))}$ is its $(\widetilde{d}+m)$-th antiderivative. ☐

Lemma 3.4. *Let $\mathcal{A}_i^\sharp : H^s(\mathbb{R}) \to H^{s-m}(\mathbb{R})$, $s \in \mathbb{R}$, be compactly supported and continuous of order $m \in \mathbb{R}$. Let $\overline{f}$ be as in Lemma 3.3 such that*

$$\left\| \mathcal{A}_i^\sharp \overline{f} \right\|_{H^{s-m}(\mathbb{R})} \leq c \cdot c_i(\mathbf{j}, \mathbf{j}') \cdot \left\| \overline{f} \right\|_{H^s(\mathbb{R})}, \qquad s \in \mathbb{R}, \tag{3.4}$$

with some suitable universal constant $c > 0$ and $c_i(\mathbf{j}, \mathbf{j}')$ as in Lemma 3.1. There holds

$$|\langle \mathcal{A}_i^\sharp \overline{f}, \, \psi_{j_i', k_i'} \rangle| \lesssim c_i(\mathbf{j}, \mathbf{j}') \cdot 2^{-j_i'(\widetilde{d}+1/2)} 2^{j_i(\frac{1}{2}+\widetilde{d}+m)}. \tag{3.5}$$

Proof. Denote $\Omega_{j_i', k_i'} := \operatorname{supp} \psi_{j_i', k_i'}$. Employing the cancelation property (2.13) one obtains

$$\begin{aligned} |\langle \mathcal{A}_i^\sharp \overline{f}, \, \psi_{j_i', k_i'} \rangle| &= \left| \int_{\mathbb{R}} \mathcal{A}_i^\sharp \overline{f}(x) \psi_{j_i', k_i'}(x) dx \right| \\ &\lesssim 2^{-j_i'(\widetilde{d}+1/2)} \cdot \sup_{x \in \Omega_{j_i', k_i'}} \left| \partial^{\widetilde{d}} \mathcal{A}_i^\sharp \overline{f}(x) \right|. \end{aligned}$$

Since $\overline{f} \in C_0^\infty(\mathbb{R})$, the fact that (3.4) holds for each $s \in \mathbb{R}$ implies

$$\begin{aligned} |\langle \mathcal{A}_i^\sharp \overline{f}, \, \psi_{j_i', k_i'} \rangle| &\lesssim c_i(\mathbf{j}, \mathbf{j}') \cdot 2^{-j_i'(\widetilde{d}+1/2)} \left| \overline{f} \right|_{W^{\widetilde{d}+m, \infty}(\Omega_{j_i', k_i'})} \\ &\lesssim c_i(\mathbf{j}, \mathbf{j}') \cdot 2^{-j_i'(\widetilde{d}+1/2)} 2^{j_i(\frac{1}{2}+\widetilde{d}+m)}, \end{aligned}$$

where the last line follows from Lemma 3.3. □

3.2 Matrix entry estimates

In order to exploit the tensor product structure of our discretization the following very simple lemma is crucial.

Lemma 3.5. *Let $i \in \{1, \ldots, n\}$ denote any coordinate direction. For a standard kernel $\kappa(\cdot, \cdot)$ of order $2q$ as in (2.5) and any two wavelets $\psi_{\mathbf{j}, \boldsymbol{k}} = \psi_{j_1, k_1} \otimes \ldots \otimes \psi_{j_n, k_n}$, $\psi_{\mathbf{j}', \boldsymbol{k}'} = \psi_{j_1', k_1'} \otimes \ldots \otimes \psi_{j_n', k_n'}$, the "dimensionally reduced" kernels,*

$$\kappa_i(x_i, x_i') := \int_{[0,1]^{n-1}} \int_{[0,1]^{n-1}} \kappa\left(x, x'\right) \prod_{\substack{s=1 \\ s \neq i}}^{n} \psi_{j_s, k_s}(x_s) \psi_{j_s', k_s'}(x_s') dx dx',$$

inherit a Calderón-Zygmund-type estimate, i.e. for $\alpha, \beta \in \mathbb{N}_0$ there holds

$$\left| \partial_{x_i}^\alpha \partial_{x_i'}^\beta \kappa_i(x_i, x_i') \right| \lesssim 2^{-\frac{1}{2}(|\mathbf{j}|_1 + |\mathbf{j}'|_1 - j_i - j_i')} \left| x_i - x_i' \right|^{-(n+2q+\alpha+\beta)}. \tag{3.6}$$

Proof. Without loss of generality, we assume $i = 1$. The proof is a straightforward application of the standard estimate (2.5) for $\kappa(\cdot, \cdot)$ and Fubini's theorem:

$$|\partial_{x_1}^\alpha \partial_{x_1'}^\beta \kappa_1(x_1, x_1')| \;=\; \left| \int_{[0,1]^{n-1}} \int_{[0,1]^{n-1}} \partial_{x_1}^\alpha \partial_{x_1'}^\beta \kappa\left(x, x'\right) \right.$$

$$
\begin{aligned}
&\qquad\qquad \cdot \prod_{s=2}^{n} \psi_{j_s,k_s}(x_s)\psi_{j_s',k_s'}(x_s')\,dx_2dx_2'\dots dx_ndx_n' \Big| \\
\lesssim\ & \Big| \int_{[0,1]^{n-1}} \int_{[0,1]^{n-1}} \Big(\sum_{m=1}^{n} (x_m - x_m')^2 \Big)^{-\frac{n+2q+\alpha+\beta}{2}} \\
&\qquad\qquad \cdot \prod_{s=2}^{n} \psi_{j_s,k_s}(x_s)\psi_{j_s',k_s'}(x_s')\,dx_2dx_2'\dots dx_ndx_n' \Big| \\
\lesssim\ & |x_1 - x_1'|^{-(n+2q+\alpha+\beta)} \prod_{s=2}^{n} \Big| \int_0^1 \psi_{j_s,k_s}(x_s)dx_s \cdot \int_0^1 \psi_{j_s',k_s'}(x_s')dx_s' \Big| \\
\lesssim\ & |x_1 - x_1'|^{-(n+2q+\alpha+\beta)} \prod_{s=2}^{n} 2^{-\frac{1}{2}j_s} 2^{-\frac{1}{2}j_s'}.
\end{aligned}
$$

□

Remark 3.6. *Obviously, κ_i depends on the particular choice of ψ_{j_s,k_s}, $\psi_{j_s',k_s'}$, $s \neq i$. To keep notation as simple as possible, we will usually conceal this dependence and simply write κ_i. The corresponding wavelets are always clear from the context.*

Now we turn to the actual fundamental estimates of the matrix entries. Let $\mathcal{A}_i : H^{q+n-1}([0,1]) \to H^{-q}([0,1])$ denote the reduction of $\mathcal{A}$ to the operator that canonically corresponds to the dimensionally reduced kernel κ_i defined above. Using the cancelation property of the one-dimensional wavelets $\psi_{j_i',k_i'}$, for $x \in [0,1]$ one infers from Lemma 3.5,

$$
\begin{aligned}
|\mathcal{A}_i \psi_{j_i',k_i'}(x)| &= |\langle \kappa_i(x,\cdot), \psi_{j_i',k_i'} \rangle| \\
&\lesssim 2^{-j_i'(\widetilde{d}+1/2)} |\kappa_i(x,\cdot)|_{W^{\infty,\widetilde{d}}(\mathrm{supp}(\psi_{j_i',k_i'}))} \qquad (3.7)\\
&\lesssim 2^{-j_i'(\widetilde{d}+1/2)} 2^{-\frac{1}{2}(|\mathbf{j}|_1 + |\mathbf{j}'|_1 - j_i - j_i')} \operatorname{dist}(\mathrm{supp}(\psi_{j_i',k_i'}),\, x)^{-(n+2q+\widetilde{d})}.
\end{aligned}
$$

In particular, $\mathcal{A}_i$ satisfies (3.1) with $c_i(\mathbf{j},\mathbf{j}') = 2^{-\frac{1}{2}(|\mathbf{j}|_1 + |\mathbf{j}'|_1 - j_i - j_i')}$ and $m = 2q + n - 1$.

Theorem 3.7. *Assume $j_i \leq j_i'$. If $0 < \sigma_{x_i} \lesssim 2^{-j_i}$, there holds*

$$
\left.\begin{array}{l} |\langle \mathcal{A}_i \psi_{j_1,k_1},\, \psi_{j_1',k_1'} \rangle| \\ |\langle \mathcal{A}_i \psi_{j_1',k_1'},\, \psi_{j_1,k_1} \rangle| \end{array}\right\} \lesssim 2^{\frac{j_i - j_i'}{2}} 2^{-j_i'\widetilde{d}} 2^{-\frac{1}{2}(|\mathbf{j}|_1 + |\mathbf{j}'|_1 - j_i - j_i')} \sigma_{x_i}^{-(2q+\widetilde{d}+n-1)},
$$

uniformly with respect to J.

Proof. Since $0 < \sigma_{x_i}$ we may denote by f (not defined on the whole $[0,1]$) the smooth piece of ψ_{j_i,k_i} whose support contains $\mathrm{supp}(\psi_{j_i',k_i'})$. Clearly, f may be extended to some $\overline{f} \in C_0^\infty(\mathbb{R})$ and one may decompose,

$$
\psi_{j_i,k_i} = \overline{f} + \overline{f}^C,
$$

with some suitable $\overline{f}^C$. We split

$$
|\langle \mathcal{A}_i \psi_{j_i',k_i'},\, \psi_{j_i,k_i} \rangle| \leq |\langle \mathcal{A}_i \psi_{j_i',k_i'},\, \overline{f} \rangle| + |\langle \mathcal{A}_i \psi_{j_i',k_i'},\, \overline{f}^C \rangle|, \qquad (3.8)
$$

and estimate both terms separately. On the one hand,

$$\mathrm{supp}(\overline{f}^C) \cap \mathrm{supp}(\psi_{j_1',k_1'}) = \emptyset,$$

and thus (3.7) implies that Lemma 3.1 with $g = \overline{f}^C$ yields the required estimate for the second term in (3.8). On the other hand, consider the extension $\mathcal{A}_i^\sharp : H^{q+n-1}(\mathbb{R}) \to H^{-q}(\mathbb{R})$ of $\mathcal{A}_i$ defined by

$$\mathcal{A}_i^\sharp f(x) = \int_{\mathbb{R}} \chi(x)\chi(x')\kappa_i(x,x')f(x')dx',$$

with some suitable C^∞-cut-off function χ that is 1 on $[0,1]$ and 0 outside $[-1,2]$. By construction, $\mathcal{A}_i^\sharp$ is compactly supported (in the sense of [33, 35]) and, as shown in e.g. [35, Section II.6], the operator $\mathcal{A}_i^\sharp$ acts continuously on the whole scale of Sobolev spaces. In fact, since $\mathcal{A}_i^\sharp$ is of order $2q+n-1$, for each $s>0$ there holds

$$\begin{aligned} \left\|\mathcal{A}_i^\sharp \overline{f}\right\|_{H^{s-2q}(\mathbb{R})} &\lesssim 2^{-\frac{1}{2}(|\mathbf{j}|_1+|\mathbf{j}'|_1-j_i-j_i')}\|\overline{f}\|_{H^{s+n-1}(\mathbb{R})} \\ &\lesssim 2^{-\frac{1}{2}(|\mathbf{j}|_1+|\mathbf{j}'|_1-j_i-j_i')}2^{j_i(s+n-1)}, \end{aligned}$$

where the factor $2^{-\frac{1}{2}(|\mathbf{j}|_1+|\mathbf{j}'|_1-j_i-j_i')}$ results from Lemma 3.5, because it is a scalar. Herewith, Lemma 3.4 with $c_i(\mathbf{j},\mathbf{j}') = 2^{-\frac{1}{2}(|\mathbf{j}|_1+|\mathbf{j}'|_1-j_i-j_i')}$ and $m = 2q+n-1$ yields

$$|\langle \mathcal{A}_i^\sharp \overline{f},\, \psi_{j_i',k_i'}\rangle_{L^2(\mathbb{R})}| \lesssim 2^{-\frac{1}{2}(|\mathbf{j}|_1+|\mathbf{j}'|_1-j_i-j_i')}2^{-j_i'(\widetilde{d}+1/2)}2^{\frac{1}{2}j_i}2^{j_i(\widetilde{d}+2q+n-1)}.$$

Since $\sigma_{x_i} \lesssim 2^{-j_i}$, this implies the required estimate for the first term in (3.8). □

The following theorem corresponds to Theorem 6.1 in [7]. But here, we exploit the tensor product structure and anisotropic nature of our wavelets.

Theorem 3.8. *There holds*

$$\left.\begin{array}{l} |\langle \mathcal{A}\psi_{\mathbf{j},\boldsymbol{k}},\, \psi_{\mathbf{j}',\boldsymbol{k}'}\rangle| \\ |\langle \mathcal{A}\psi_{\mathbf{j}',\boldsymbol{k}'},\, \psi_{\mathbf{j},\boldsymbol{k}}\rangle| \end{array}\right\} \lesssim 2^{-\frac{1}{2}(|\mathbf{j}|_1+|\mathbf{j}'|_1)}2^{-\widetilde{d}(j^{(1)}+j^{(2)})}\delta_{xy}^{-(n+2q+2\widetilde{d})}, \tag{3.9}$$

where $j^{(1)}$, $j^{(2)}$ may be any distinct two out of the four indices.

Proof. One may assume without loss of generality that $j^{(1)} = j_1$ and $j^{(2)} = j_2'$, because in the analysis below all the wavelets are interchangeable. In fact, one only requires that all one-dimensional wavelets admit the same number of vanishing moments.

Consider arbitrary but fixed points $x_s \in [0,1]$, $x_l' \in [0,1]$ in all coordinate directions $s \neq 1$, $l \neq 2$. Using integration by parts and the standard kernel estimate (2.5) for $\kappa(\cdot,\cdot)$ one finds

$$\begin{aligned} &\left|\int_0^1\int_0^1 \psi_{j_1,k_1}(x_1)\psi_{j_2',k_2'}(x_2')\kappa((x_1,\ldots,x_n),\,(x_1',\ldots,x_n'))dx_1dx_2'\right| \\ \lesssim\;& 2^{-\widetilde{d}(j_1+j_2')}\int_0^1\int_0^1 |\psi_{j_1,k_1}^{(-\widetilde{d})}(x_1)\psi_{j_2',k_2'}^{(-\widetilde{d})}(x_2')|\left(\sum_{s=1}^n (x_s-x_s')^2\right)^{-\frac{n+2q+2\widetilde{d}}{2}}dx_1dx_2'. \end{aligned}$$

Using Fubini's Theorem, one herewith obtains

$$
\begin{aligned}
|\langle \mathcal{A}\psi_{\mathbf{j},\mathbf{k}},\, \psi_{\mathbf{j}',\mathbf{k}'}\rangle| &\lesssim \int_{[0,1]^{n-1}}\int_{[0,1]^{n-1}} \left|\prod_{s=2}^{n}\psi_{j_s,k_s}(x_s)\prod_{\substack{l=1\\l\neq 2}}^{n}\psi_{j_l',k_l'}(x_l')\right| \\
&\qquad \times \int_0^1\int_0^1 \left|\psi_{j_1,k_1}(x_1)\psi_{j_2',k_2'}(x_2')\right|\left|\kappa(x,\,x')\right| dx dx' \\
&\lesssim 2^{-\widetilde{d}(j_1+j_2')}\int_{[0,1]^n}\int_{[0,1]^n}\left|\psi_{j_1,k_1}^{(-\widetilde{d})}(x_1)\psi_{j_2',k_2'}^{(-\widetilde{d})}(x_2')\right| \\
&\qquad \times \left|\prod_{s=2}^{n}\psi_{j_s,k_s}(x_s)\prod_{\substack{l=1\\l\neq 2}}^{n}\psi_{j_l',k_l'}(x_l')\right| \cdot |x-x'|^{-(n+2q+2\widetilde{d})} dx dx' \\
&\lesssim 2^{-(\widetilde{d}+\frac{1}{2})(j_1+j_2')}2^{-\frac{1}{2}(|\mathbf{j}|_1+|\mathbf{j}'|_1-j_1-j_2')}\delta_{xy}^{-(n+2q+2\widetilde{d})},
\end{aligned}
$$

where the last line follows from the standard bounds on the volume of the supports of the wavelets and their suprema as described in Section 2.1. □

Theorem 3.9. *Assume* $j_i \leq j_i'$. *If* $0 < \sigma_{x_i} \lesssim 2^{-j_i}$ *then there holds*

$$
\left.\begin{array}{l}
|\langle \mathcal{A}\psi_{\mathbf{j},\boldsymbol{k}},\, \psi_{\mathbf{j}',\boldsymbol{k}'}\rangle| \\
|\langle \mathcal{A}\psi_{\mathbf{j}',\boldsymbol{k}'},\, \psi_{\mathbf{j},\boldsymbol{k}}\rangle|
\end{array}\right\} \lesssim 2^{-\widetilde{d}j_i'}2^{j_i}2^{-\frac{1}{2}(|\mathbf{j}|_1+|\mathbf{j}'|_1)}\sigma_{x_i}^{-(2q+\widetilde{d}+n-1)}. \tag{3.10}
$$

Proof. Using Fubini's Theorem one obtains

$$
\begin{aligned}
\left|\langle \mathcal{A}\psi_{\mathbf{j},\mathbf{k}},\, \psi_{\mathbf{j}',\mathbf{k}'}\rangle\right| &= \left|\int_{[0,1]^n}\int_{[0,1]^n}\psi_{\mathbf{j},\mathbf{k}}(x)\psi_{\mathbf{j}',\mathbf{k}'}(x')\kappa(x,\,x')dx dx'\right| \\
&= \left|\int_0^1\int_0^1 \psi_{j_i,k_i}(x_i)\psi_{j_i',k_i'}(x_i')\right. \\
&\qquad \left. \times \int_{[0,1]^{n-1}}\int_{[0,1]^{n-1}}\kappa\left(x,x'\right)\prod_{\substack{s=1\\s\neq i}}^{n}\psi_{j_s,k_s}(x_s)\psi_{j_s',k_s'}(x_s')dx dx'\right| \\
&= \left|\int_0^1\int_0^1 \psi_{j_i,k_i}(x_i)\psi_{j_i',k_i'}(x_i')\kappa_i(x_i,x_i')dx_i dx_i'\right| \\
&= \left|\langle \mathcal{A}_i\psi_{j_i,k_i},\, \psi_{j_i',k_i'}\rangle\right|,
\end{aligned}
$$

with κ_i and $\mathcal{A}_i$ defined as above. Thus, the result follows directly from Theorem 3.7. □

4 Consistency framework

In order to analyze the impact of a compression scheme for $\mathcal{A}$ on the discretization of the original problem

$$
\mathcal{B}u = \mathcal{A}_D u + \mathcal{A}u = f,
$$

recall that by (2.6), the order of $\mathcal{B}$ is given by

$$2\widetilde{q} = \begin{cases} 2, & \text{if } \mathcal{Q} > 0 \text{ and } 2q \leq 2, \\ 2q, & \text{otherwise,} \end{cases} \tag{4.1}$$

where $\mathcal{Q} \in \mathbb{R}^{n \times n}$ denotes the coefficient matrix of $\mathcal{A}_D$. The stability and convergence results of the sparse tensor product FE discretization without compression are given by Proposition 2.4.

To characterize the consistency requirements that need to be satisfied by a compression scheme for $\mathcal{A}$, we define the following scale of interpolation spaces

$$X_{\theta,\widetilde{q},d} := (H^{\widetilde{q}}(\square),\, \mathcal{H}^d(\square))_{\theta,2}, \quad 0 \leq \theta \leq 1, \tag{4.2}$$

using the K-method of interpolation (cf. e.g. [36, Section 1.3]). From the wavelet norm equivalences and estimates of the spaces $H^{\widetilde{q}}(\square)$ and $\mathcal{H}^d(\square)$ one obtains the following estimate for the norm of $X_{\theta,\widetilde{q},d}$.

Proposition 4.1. *Let* $0 \leq \theta \leq 1$ *and* $\mathbf{j} \in \mathbb{N}_0^n$ *be a fixed level index. For any* $u \in X_{\theta,\widetilde{q},d}$ *with wavelet representation*

$$u = \sum_{\mathbf{j}'} \sum_{\mathbf{k}' \in \nabla_{\mathbf{j}'}} u_{\mathbf{j}',\mathbf{k}'} \psi_{\mathbf{j}',\mathbf{k}'},$$

there holds

$$\sum_{\mathbf{k} \in \nabla_{\mathbf{j}}} 2^{2(1-\theta)\widetilde{q}|\mathbf{j}|_\infty} 2^{2\theta d |\mathbf{j}|_1} |u_{\mathbf{j},\mathbf{k}}|^2 \lesssim \|u\|^2_{X_{\theta,\widetilde{q},d}}. \tag{4.3}$$

Proof. By e.g. [36, Theorem 1.3.3 (c)], there exists a positive number $c > 0$ only depending on θ such that for all $t \in \mathbb{R}_{>0}$ there holds

$$t^{-\theta} K(t,u) \leq c\|u\|_{X_{\theta,\widetilde{q},d}}, \qquad \text{for all } u \in X_{\theta,\widetilde{q},d},$$

where

$$K(t,u) = \inf_{g \in \mathcal{H}^d} \left\{ \|u-g\|_{H^{\widetilde{q}}} + t\|g\|_{\mathcal{H}^d} \right\},$$

denotes the K-functional. Thus, it suffices to show that there exists some constant $c > 0$ such that for any fixed $\mathbf{j} \in \mathbb{N}_0^n$ there exists $0 < t < \infty$ such that

$$\sum_{\mathbf{k} \in \nabla_{\mathbf{j}}} 2^{2(1-\theta)q|\mathbf{j}|_\infty} 2^{2\theta d |\mathbf{j}|_1} |u_{\mathbf{j},\mathbf{k}}|^2 \leq c \cdot t^{-2\theta} K(t,u)^2, \qquad \text{for all } u \in X_{\theta,\widetilde{q},d}. \tag{4.4}$$

Define

$$T(\mathbf{j},t,u) := \inf_{\substack{g \in \mathcal{H}^d \\ g = \sum_{\mathbf{k} \in \nabla_{\mathbf{j}}} g_{\mathbf{j},\mathbf{k}} \psi_{\mathbf{j},\mathbf{k}}}} \left\{ \left(\sum_{\mathbf{k} \in \nabla_{\mathbf{j}}} (2^{2\widetilde{q} j_1} + \ldots + 2^{2\widetilde{q} j_n}) \times |u_{\mathbf{j},\mathbf{k}} - g_{\mathbf{j},\mathbf{k}}|^2 \right)^{\frac{1}{2}} + t \left(\sum_{\mathbf{k} \in \nabla_{\mathbf{j}}} 2^{2d|\mathbf{j}|_1} |g_{\mathbf{j},\mathbf{k}}|^2 \right)^{\frac{1}{2}} \right\}.$$

Using the wavelet norm estimate (2.27) for $\|\cdot\|_{\mathcal{H}^d}$ and the well-known norm equivalence for $\|\cdot\|_{H^{\widetilde{q}}}$ one obtains

$$\begin{aligned} T(\mathbf{j},t,u) &\lesssim \inf_{g\in\mathcal{H}^d}\left\{\Big(\sum_{\substack{\mathbf{j}'\in\mathbb{N}_0^n\\ \mathbf{k}'\in\nabla_{\mathbf{j}'}}}(2^{2\widetilde{q}j_1'}+\ldots+2^{2\widetilde{q}j_n'})|u_{\mathbf{j}',\mathbf{k}'}-g_{\mathbf{j}',\mathbf{k}'}|^2\Big)^{\frac{1}{2}}+t\|g\|_{\mathcal{H}^d}\right\}\\ &\sim \inf_{g\in\mathcal{H}^d}\left\{\|u-g\|_{H^{\widetilde{q}}}+t\|g\|_{\mathcal{H}^d}\right\}\\ &\sim K(t,u), \end{aligned}$$

for any $0<t<\infty$. Thus, to prove (4.3) it suffices to show that there exists some $0<t<\infty$ such that

$$\sum_{\mathbf{k}\in\nabla_{\mathbf{j}}}2^{2(1-\theta)\widetilde{q}|\mathbf{j}|_\infty}2^{2\theta d|\mathbf{j}|_1}|u_{\mathbf{j},\mathbf{k}}|^2\lesssim t^{-2\theta}T(\mathbf{j},t,u)^2. \tag{4.5}$$

With $t=\frac{(2^{2\widetilde{q}j_1}+\ldots+2^{2\widetilde{q}j_n})^{\frac{1}{2}}}{2^{d(j_1+\cdots+j_n)}}$, inequality (4.5) reads

$$\begin{aligned} \sum_{\mathbf{k}\in\nabla_{\mathbf{j}}}2^{2(1-\theta)\widetilde{q}|\mathbf{j}|_\infty}2^{2\theta d|\mathbf{j}|_1}|u_{\mathbf{j},\mathbf{k}}|^2 &\lesssim \frac{2^{2\theta d(j_1+\ldots+j_n)}}{(2^{2\widetilde{q}j_1}+\ldots+2^{2\widetilde{q}j_n})^{\theta-1}}\\ &\qquad\times\inf_{g\in\mathcal{H}^d}\left\{\Big(\sum_{\mathbf{k}\in\nabla_{\mathbf{j}}}|u_{\mathbf{j},\mathbf{k}}-g_{\mathbf{j},\mathbf{k}}|^2\Big)^{\frac{1}{2}}+\Big(\sum_{\mathbf{k}\in\nabla_{\mathbf{j}}}|g_{\mathbf{j},\mathbf{k}}|^2\Big)^{\frac{1}{2}}\right\}^2\\ &\sim \frac{2^{2\theta d(j_1+\ldots+j_n)}}{(2^{2\widetilde{q}j_1}+\ldots+2^{2\widetilde{q}j_n})^{\theta-1}}\sum_{\mathbf{k}\in\nabla_{\mathbf{j}}}|u_{\mathbf{j},\mathbf{k}}|^2. \end{aligned}$$

Since the validity of this inequality follows immediately from the trivial estimate $2^{2(1-\theta)\widetilde{q}|\mathbf{j}|_\infty}\leq(2^{2\widetilde{q}j_1}+\ldots+2^{2\widetilde{q}j_n})^{1-\theta}$, this implies the validity of (4.5) with this particular choice of $0<t<\infty$. Hence the result follows. □

For a given compression scheme, denote by $\mathbb{A}_J^{compr}$ the compressed matrix and let the corresponding operator be given by

$$\mathcal{A}_J^{compr}u:=\sum_{\mathbf{j},\mathbf{j}'}\sum_{\mathbf{k},\mathbf{k}'}\left[\mathbb{A}_J^{compr}\right]_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')}\langle\widetilde{\psi}_{\mathbf{j},\mathbf{k}},u\rangle\,\widetilde{\psi}_{\mathbf{j}',\mathbf{k}'},$$

where the first sum is taken over all level indices such that $0\leq|\mathbf{j}|_1,|\mathbf{j}'|_1\leq J$. There holds $\langle\mathcal{A}_J^{compr}\psi_{\mathbf{j},\mathbf{k}},\psi_{\mathbf{j}',\mathbf{k}'}\rangle=\left[\mathbb{A}_J^{compr}\right]_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')}$, for all $\psi_{\mathbf{j},\mathbf{k}},\psi_{\mathbf{j}',\mathbf{k}'}\in\Psi$.

As shown below, in order to preserve stability and convergence results from the unperturbed (but sparse tensor product) Galerkin discretization the compression scheme has to satisfy the following consistency requirement.

Requirement 1. *For any $0\leq\theta_1,\theta_2\leq 1$ and all $u\in X_{\theta_1,\widetilde{q},d}$, $v\in X_{\theta_2,\widetilde{q},d}$, the operator $\mathcal{A}_J^{compr}$ corresponding to a compressed matrix $\mathbb{A}_J^{compr}$ has to satisfy*

$$\begin{aligned} &\left|\left\langle\left(\mathcal{A}-\mathcal{A}_J^{compr}\right)\widehat{P}_Ju,\,\widehat{P}_Jv\right\rangle\right|\\ &\lesssim\varepsilon\cdot 2^{2J\widetilde{q}}2^{-J((1-\theta_1)\widetilde{q}+\theta_1 d)}2^{-J((1-\theta_2)\widetilde{q}+\theta_2 d)}\|u\|_{X_{\theta_1,\widetilde{q},d}}\|v\|_{X_{\theta_2,\widetilde{q},d}}, \end{aligned} \tag{4.6}$$

with some suitable constant $\varepsilon > 0$ uniformly with respect to $J \geq 0$.

Setting $\theta_1 = \theta_2 = 0$ one directly obtains that if the compressed matrix satisfies Requirement 1 there holds

$$\left|\left\langle \left(\mathcal{A} - \mathcal{A}_J^{compr}\right) u_J,\, u_J \right\rangle\right| \lesssim \varepsilon \|u_J\|_{H^{\widetilde{q}}}^2, \qquad u_J \in \widehat{V}_J. \tag{4.7}$$

Note that, since $\mathcal{B} = \mathcal{A}_D + \mathcal{A}$ and $\mathcal{A}_D$ is local, there holds

$$\mathcal{B} - \mathcal{B}_J^{compr} = \mathcal{A} - \mathcal{A}_J^{compr} \quad \text{for all } J > 0. \tag{4.8}$$

Thus, one obtains

Theorem 4.2. *Suppose the solution u of* (2.1) *satisfies $u \in \mathcal{H}^d$ and $\mathcal{A}_J^{compr}$ satisfies Requirement 1. Then for sufficiently small $\varepsilon > 0$ the compressed Galerkin scheme is stable, i.e. there exist $J_0 > 0$ and $c_1' > 0$, $c_2 \geq 0$ such that for any $J \geq J_0$ there holds*

$$\langle \mathcal{B}_J^{compr} u_J,\, u_J \rangle \geq c_1' \|u_J\|_{H^{\widetilde{q}}}^2 - c_2 \|u_J\|_{L^2}^2, \quad \textit{for all } u_J \in \widehat{V}_J, \tag{4.9}$$

and there exists some $c_3' > 0$ such that for all $J \geq J_0$,

$$|\langle \mathcal{B}_J^{compr} u_J,\, v_J \rangle| \leq c_3' \|u_J\|_{H^{\widetilde{q}}} \|v_J\|_{H^{\widetilde{q}}}, \quad \textit{for all } u_J, v_J \in \widehat{V}_J. \tag{4.10}$$

Furthermore, the convergence rate (2.33) *of the Galerkin scheme without compression is preserved (cf. Proposition 2.4).*

Proof. Inequality (4.9) may be verified by inserting (4.7) into (2.31) and using (4.8). This yields,

$$\begin{aligned} \langle \mathcal{B}_J^{compr} u_J,\, u_J \rangle &\geq c_1 \|u_J\|_{H^{\widetilde{q}}}^2 - c_2 \|u_J\|_{L^2}^2 - 2\varepsilon \|u_J\|_{H^{\widetilde{q}}}^2 \\ &= (c_1 - 2\varepsilon) \|u_J\|_{H^{\widetilde{q}}}^2 - c_2 \|u_J\|_{L^2}^2, \end{aligned}$$

and $c_1' := c_1 - 2\varepsilon > 0$ for sufficiently small $\varepsilon > 0$ from Requirement 1. The constants c_1, c_2 are obtained from (2.31). For the continuity inequality (4.10) one obtains from (2.32) and Requirement 1 with $\theta_1 = \theta_2 = 0$,

$$\begin{aligned} |\langle \mathcal{B}_J^{compr} u_J,\, v_J \rangle| &\leq |\langle \mathcal{B} u_J,\, v_J \rangle| + \left|\left\langle \left(\mathcal{B}_J^{compr} - \mathcal{B}\right) u_J,\, v_J \right\rangle\right| \\ &\leq c_3 \|u_J\|_{H^{\widetilde{q}}} \|v_J\|_{H^{\widetilde{q}}} + \varepsilon \|u_J\|_{H^{\widetilde{q}}} \|v_J\|_{H^{\widetilde{q}}}, \end{aligned}$$

with c_3 from (2.32). Setting $c_3' = c_3 + \varepsilon$ one obtains (4.10).

Finally, noting that $\widetilde{q} < \gamma$ (with γ as in (2.22)), the convergence result follows by setting $\theta_1 = 1$, $\theta_2 = 0$ in (4.6) from Strang's first lemma (see e.g. [12, Lemma 2.27]). ☐

By the above results, Requirement 1 indeed provides the correct framework for the anisotropic compression scheme. Incorporating this idea, the following theorem provides (lower) bounds for the cut-off parameters and hence will enable us to define the compression scheme (cf. Section 5). To simplify notation, for any two level indices $\mathbf{j}, \mathbf{j}' \in \mathbb{N}_0^n$ we introduce

$$\sigma_{\mathbf{j},\mathbf{j}'} := \begin{cases} 2J(d'-\widetilde{q}) - d'(|\mathbf{j}|_1 + |\mathbf{j}'|_1), \text{ if } \begin{cases} (J-|\mathbf{j}'|_1)d' \geq (J-|\mathbf{j}'|_\infty)\widetilde{q}, \\ (J-|\mathbf{j}|_1)d' \geq (J-|\mathbf{j}|_\infty)\widetilde{q}, \end{cases} \\ -\widetilde{q}(|\mathbf{j}|_\infty + |\mathbf{j}'|_\infty), \qquad \text{if } \begin{cases} (J-|\mathbf{j}'|_1)d' < (J-|\mathbf{j}'|_\infty)\widetilde{q}, \\ (J-|\mathbf{j}|_1)d' < (J-|\mathbf{j}|_\infty)\widetilde{q}, \end{cases} \\ J(d'-\widetilde{q}) - \widetilde{q}|\mathbf{j}'|_\infty - d'|\mathbf{j}|_1, \text{ if } \begin{cases} (J-|\mathbf{j}'|_1)d' < (J-|\mathbf{j}'|_\infty)\widetilde{q}, \\ (J-|\mathbf{j}|_1)d' \geq (J-|\mathbf{j}|_\infty)\widetilde{q}, \end{cases} \\ J(d'-\widetilde{q}) - d'|\mathbf{j}'|_1 - \widetilde{q}|\mathbf{j}|_\infty, \text{ if } \begin{cases} (J-|\mathbf{j}'|_1)d' \geq (J-|\mathbf{j}'|_\infty)\widetilde{q}, \\ (J-|\mathbf{j}|_1)d' < (J-|\mathbf{j}|_\infty)\widetilde{q}, \end{cases} \end{cases} \tag{4.11}$$

with some given control parameter $d \leq d' < \widetilde{d} + 2\widetilde{q}$.

Theorem 4.3. *Let $\mathbb{A}_J^{compr}$ denote the matrix compressed with respect to some cut-off parameter $E_{\mathbf{j},\mathbf{j}'}$, i.e.*

$$\left[\mathbb{A}_J^{compr}\right]_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')} := \begin{cases} 0, & \text{if } \delta_{\mathbf{j},\mathbf{j}'} > E_{\mathbf{j},\mathbf{j}'}, \\ \left[\mathbb{A}_J\right]_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')}, & \text{otherwise,} \end{cases}$$

with some suitable $\delta_{\mathbf{j},\mathbf{j}'} \in \{\delta_{x_i}, \sigma_{x_i} : i = 1, \ldots, n\}$. Then $\mathbb{A}_J^{compr}$ satisfies Requirement 1 if the block matrices $\mathbb{A}_{\mathbf{j},\mathbf{j}'}^{compr}$ satisfy

$$\left\|\mathbb{A}_{\mathbf{j},\mathbf{j}'} - \mathbb{A}_{\mathbf{j},\mathbf{j}'}^{compr}\right\|_2 \lesssim \varepsilon\, 2^{-\sigma_{\mathbf{j},\mathbf{j}'}},$$

with $\sigma_{\mathbf{j},\mathbf{j}'}$ as in (4.11).

Proof. Let $0 \leq \theta_1, \theta_2 \leq 1$ be as in Requirement 1. There holds

$$\begin{aligned} \left|\left\langle \left(\mathcal{A} - \mathcal{A}_J^{compr}\right) \widehat{P}_J u,\, \widehat{P}_J v\right\rangle\right| &= \left| \sum_{\substack{0\leq|\mathbf{j}|_1\leq J \\ \mathbf{k}\in\nabla_\mathbf{j}}} \sum_{\substack{0\leq|\mathbf{j}'|_1\leq J \\ \mathbf{k}'\in\nabla_{\mathbf{j}'}}} \left\langle \left(\mathcal{A} - \mathcal{A}_J^{compr}\right) u_\mathbf{k}^\mathbf{j} \psi_{\mathbf{j},\mathbf{k}},\, v_{\mathbf{k}'}^{\mathbf{j}'} \psi_{\mathbf{j}',\mathbf{k}'} \right\rangle \right| \\ &\lesssim \sum_{\substack{0\leq|\mathbf{j}|_1\leq J \\ 0\leq|\mathbf{j}'|_1\leq J}} \left| \sum_{\substack{\mathbf{k}\in\nabla_\mathbf{j} \\ \mathbf{k}'\in\nabla_{\mathbf{j}'}}} u_\mathbf{k}^\mathbf{j} v_{\mathbf{k}'}^{\mathbf{j}'} \left\langle \left(\mathcal{A} - \mathcal{A}_J^{compr}\right) \psi_{\mathbf{j},\mathbf{k}},\, \psi_{\mathbf{j}',\mathbf{k}'} \right\rangle \right| \\ &\lesssim \sum_{\substack{0\leq|\mathbf{j}|_1\leq J \\ 0\leq|\mathbf{j}'|_1\leq J}} \left(\left\| \left[u_\mathbf{k}^\mathbf{j}\right]_{\mathbf{k}\in\nabla_\mathbf{j}} \right\|_2 \left\| \left[v_{\mathbf{k}'}^{\mathbf{j}'}\right]_{\mathbf{k}'\in\nabla_{\mathbf{j}'}} \right\|_2 \left\| \mathbb{A}_{\mathbf{j},\mathbf{j}'} - \mathbb{A}_{\mathbf{j},\mathbf{j}'}^{compr} \right\|_2 \right) \\ &\lesssim \|u\|_{X_{\theta_1,\widetilde{q},d}} \|v\|_{X_{\theta_2,\widetilde{q},d}} \\ &\qquad \times \sum_{\substack{0\leq|\mathbf{j}|_1\leq J \\ 0\leq|\mathbf{j}'|_1\leq J}} \Big\| 2^{-(1-\theta_1)\widetilde{q}|\mathbf{j}'|_\infty} 2^{-\theta_1 d|\mathbf{j}'|_1} \\ &\qquad \times 2^{-(1-\theta_2)\widetilde{q}|\mathbf{j}|_\infty} 2^{-\theta_2 d|\mathbf{j}|_1} \left(\mathbb{A}_{\mathbf{j},\mathbf{j}'} - \mathbb{A}_{\mathbf{j},\mathbf{j}'}^{compr} \right) \Big\|_2, \end{aligned}$$

where the last inequality follows from Proposition 4.1.

Thus, Requirement 1 is satisfied if for all $0 \le \theta_1, \theta_2 \le 1$ there holds

$$\left\| \overline{\mathbb{A}_{\mathbf{j},\,\mathbf{j}'}} \right\|_2 \lesssim \varepsilon,$$

where

$$\begin{aligned} \overline{\mathbb{A}_{\mathbf{j},\,\mathbf{j}'}} \;&:=\; 2^{-2J\widetilde{q}} \cdot 2^{J((1-\theta_1)\widetilde{q}+\theta_1 d)} 2^{J((1-\theta_2)\widetilde{q}+\theta_2 d)} \\ &\quad\times 2^{-(1-\theta_1)\widetilde{q}|\mathbf{j}'|_\infty} 2^{-\theta_1 d|\mathbf{j}'|_1} 2^{-(1-\theta_2)\widetilde{q}|\mathbf{j}|_\infty} 2^{-\theta_2 d|\mathbf{j}|_1} \left(\mathbb{A}_{\mathbf{j},\,\mathbf{j}'} - \mathbb{A}^{compr}_{\mathbf{j},\,\mathbf{j}'}\right). \end{aligned}$$

With this definition, the row sums of $\overline{\mathbb{A}} := \left(\overline{\mathbb{A}_{\mathbf{j},\,\mathbf{j}'}}\right)_{(\mathbf{j},\mathbf{j}')}$ can trivially be estimated by

$$\begin{aligned} \sum_{0 \le |\mathbf{j}'|_1 \le J} \left|\overline{\mathbb{A}_{\mathbf{j},\,\mathbf{j}'}}\right| \;&\le\; \sum_{0 \le |\mathbf{j}'|_1 \le J} 2^{(1-\theta_1)\widetilde{q}(J-|\mathbf{j}'|_\infty)} 2^{\theta_1 d(J-|\mathbf{j}'|_1)} \\ &\qquad\times 2^{(1-\theta_2)\widetilde{q}(J-|\mathbf{j}|_\infty)} 2^{\theta_2 d(J-|\mathbf{j}|_1)} 2^{-2J\widetilde{q}} \left\| \mathbb{A}_{\mathbf{j},\,\mathbf{j}'} - \mathbb{A}^{compr}_{\mathbf{j},\,\mathbf{j}'} \right\|_2 . \end{aligned}$$

Since the same estimate is also valid for the column sums, by the Cauchy-Schwarz inequality (or the Schur Lemma with unit weight, cf. [22, Section VIII.4]) one obtains that Requirement 1 is satisfied if

$$\left\| \mathbb{A}_{\mathbf{j},\,\mathbf{j}'} - \mathbb{A}^{compr}_{\mathbf{j},\,\mathbf{j}'} \right\|_2 \lesssim \varepsilon\, 2^{-\sigma},$$

with σ such that

$$(1-\theta_1)\widetilde{q}(J-|\mathbf{j}'|_\infty) + \theta_1 d(J-|\mathbf{j}'|_1) + (1-\theta_2)\widetilde{q}(J-|\mathbf{j}|_\infty) + \theta_2 d(J-|\mathbf{j}|_1) - 2J\widetilde{q} \le \sigma, \tag{4.12}$$

for all $0 \le \theta_1, \theta_2 \le 1$. Differentiation of the left hand side of (4.12) with respect to θ_1 and θ_2, resp., shows its monotonicity with respect to these parameters. Thus using a monotonicity argument one obtains that (4.12) is satisfied by $\sigma_{\mathbf{j},\mathbf{j}'}$ defined in (4.11). □

The definition of the cut-off parameters will directly rely on the following modification of Theorem 4.3.

Corollary 4.4. *If there exists some parameter $g = g(\mathbf{j},\,\mathbf{j}')$ and a constant c, such that,*

$$2^{-\sigma_{\mathbf{j},\mathbf{j}'}} = 2^{-g} E^{-c}_{\mathbf{j},\,\mathbf{j}'},$$

then Theorem 4.3 boils down to requiring that

$$E_{\mathbf{j},\,\mathbf{j}'} \ge \varepsilon^{-c}\, 2^{\frac{\tau}{c}}, \text{ with } \tau \ge \sigma_{\mathbf{j},\mathbf{j}'} - g.$$

Remark 4.5. *Note that in case one is interested in using sparse tensor product based wavelet compression for the fast evaluation of integral expressions as in e.g. [21], then (because the stability estimates* (4.9)–(4.10) *are not necessary) Requirement 1 can be relaxed to*

Requirement 2. *For $d < \widetilde{d} + 2q$, the operator $\mathcal{A}^{compr}_J$ corresponding to a compressed matrix $\mathbb{A}^{compr}_J$ has to satisfy*

$$\left| \left\langle \left(\mathcal{A} - \mathcal{A}^{compr}_J\right) \widehat{P}_J u,\, \widehat{P}_J v \right\rangle \right| \lesssim 2^{2Jq} 2^{-J(t+t')} \|u\|_{\mathcal{H}^{t'}} \|v\|_{\mathcal{H}^t}, \tag{4.13}$$

for any $q \le t,\, t' \le d$ uniformly with respect to $J \ge 0$.

Analogously to Theorem 4.3 one then obtains

Corollary 4.6. *Let $\mathbb{A}_J^{compr}$ denote the matrix compressed with respect to some cut-off parameter $E_{\mathbf{j},\mathbf{j}'}$, i.e.*

$$\left[\mathbb{A}_J^{compr}\right]_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')} := \begin{cases} 0, & \text{if } \delta_{\mathbf{j},\mathbf{j}'} > E_{\mathbf{j},\mathbf{j}'}, \\ \left[\mathbb{A}_J\right]_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')}, & \text{otherwise,} \end{cases}$$

with some suitable $\delta_{\mathbf{j},\mathbf{j}'} \in \{\delta_{x_i}, \sigma_{x_i} : i = 1, \dots, n\}$. Then $\mathbb{A}_J^{compr}$ satisfies Requirement 2 if the block matrices $\mathbb{A}_{\mathbf{j},\mathbf{j}'}^{compr}$ satisfy,

$$\left\|\mathbb{A}_{\mathbf{j},\mathbf{j}'} - \mathbb{A}_{\mathbf{j},\mathbf{j}'}^{compr}\right\|_2 \lesssim 2^{-\sigma'_{\mathbf{j},\mathbf{j}'}},$$

with

$$\sigma'_{\mathbf{j},\mathbf{j}'} := 2J(d' - q) - d'(|\mathbf{j}|_1 + |\mathbf{j}'|_1), \tag{4.14}$$

for given $d \le d' < \widetilde{d} + 2q$.

Remark 4.7. *By definition, $\sigma'_{\mathbf{j},\mathbf{j}'} \le \sigma_{\mathbf{j},\mathbf{j}'}$ for all $\mathbf{j}, \mathbf{j}' \in \mathbb{N}_0^n$.*

5 Compression scheme

In this section, we define two compression schemes and show that the resulting compressed matrices $\mathbb{A}_J^{compr}$ satisfy Requirement 1 and Requirement 2, respectively. The schemes are split into two parts based on the distinction of *first* and *second* compression as defined in [7, 30]:

In the first compression the cut-off criteria are based on the distance of the wavelets' supports. The second compression employs cut-off criteria based on the distance of the support of smaller wavelets to the singular support of larger ones, i.e. it is based on σ_{x_s} defined above. Note that here matrix entries can be dropped even if the supports of their wavelets intersect. The consistency of the first compression relies on Theorem 3.8 whereas the second compression scheme results from Theorem 3.9.

Denote by $\widetilde{\sigma}_{\mathbf{j},\mathbf{j}'}$ some parameter depending on $\mathbf{j}, \mathbf{j}'$ which can be chosen to be either $\sigma_{\mathbf{j},\mathbf{j}'}$ or $\sigma'_{\mathbf{j},\mathbf{j}'}$ as defined in (4.11) and (4.14).

For each coordinate direction $i = 1, \dots, n$ and any index set $\mathcal{I} \subset \{1, \dots, n\}$, the corresponding cut-off parameter of the *first compression* is defined by

$$C_{\mathbf{j},\mathbf{j}'}^{i,\mathcal{I}} := c_i \max\left\{2^{-\min\{j_i, j'_i\}}, 2^{\frac{\widetilde{\sigma}_{\mathbf{j},\mathbf{j}'} - \widetilde{d}(j^{(1)} + j^{(2)}) - \sum_{k \in \mathcal{I}} \min\{j_k, j'_k\}}{2\widetilde{d} + 2q + |\mathcal{I}|}}\right\},$$

with $c_i > 0$. Here $|\mathcal{I}|$ denotes the cardinality of the index set $\mathcal{I} \subset \{1, \dots, n\}$. Clearly, one has to pick

$$j^{(1)} = \max\{j_1, \dots, j_n\}, \quad j^{(2)} = \max(\{j_1, \dots, j_n\} \backslash \{j^{(1)}\}) \tag{5.1}$$

for the best compression results. We therefore fix $j^{(1)}$ and $j^{(2)}$ in this way for the remainder of this work.

Furthermore, to each pair of wavelets $\psi_{\mathbf{j},\mathbf{k}}, \psi_{\mathbf{j}',\mathbf{k}'}$ corresponding to one matrix entry, we associate the index subset

$$\mathcal{I}(\mathbf{j}, \mathbf{k}, \mathbf{j}', \mathbf{k}') := \left\{s \in \{1, \dots, n\} : \delta_{x_s} \le 2^{-\min\{j_s, j'_s\}}\right\}.$$

Herewith, for $q \le d \le \widetilde{d}$, the first compression scheme is defined by

$$\left[\mathbb{A}_J^{cpr-1}\right]_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')} := \begin{cases} 0, \text{ if } \begin{cases} \exists\, i \in \{1,\ldots,n\}, \text{ s.t.} \\ \delta_{x_i} > C_{\mathbf{j},\mathbf{j}'}^{i,\mathcal{I}(\mathbf{j},\mathbf{k},\mathbf{j}',\mathbf{k}')}, \end{cases} \\ \left[\mathbb{A}_J\right]_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')}, \text{ otherwise.} \end{cases}$$

The cut-off parameters of the *second compression* are defined by

$$E_{\mathbf{j},\mathbf{j}'}^{i,\mathcal{I}} := e_i 2^{\frac{\widetilde{\sigma}_{\mathbf{j},\mathbf{j}'} - \widetilde{d}\max\{j_i,j_i'\} - \sum_{k\in\mathcal{I}\setminus\{i\}} \min\{j_k,j_k'\}}{\widetilde{d}+2q+|\mathcal{I}|-1}},$$

with $e_i > 0$. The second compression scheme is thus defined by

$$\left[\mathbb{A}_J^{cpr-2}\right]_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')} := \begin{cases} 0, \text{ if } \begin{cases} \exists\, i \in \{1,\ldots,n\}, \text{ s.t.} \\ i \in \mathcal{I}(\mathbf{j},\mathbf{k},\mathbf{j}',\mathbf{k}'), \\ \sigma_{x_i} > E_{\mathbf{j},\mathbf{j}'}^{i,\mathcal{I}(\mathbf{j},\mathbf{k},\mathbf{j}',\mathbf{k}')}, \end{cases} \\ \left[\mathbb{A}_J\right]_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')}, \text{ otherwise.} \end{cases}$$

The fully compressed matrix $\mathbb{A}_J^{compr}$ is defined by

$$\left[\mathbb{A}_J^{compr}\right]_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')} := \begin{cases} 0, \text{ if } \begin{cases} \left[\mathbb{A}_J^{cpr-m}\right]_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')} = 0, \\ \text{for some } m \in \{1,2\}, \end{cases} \\ \left[\mathbb{A}_J\right]_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')}, \text{ otherwise.} \end{cases}$$

The consistency of this scheme is ensured by

Theorem 5.1. *If* $\widetilde{\sigma}_{\mathbf{j},\mathbf{j}'} = \sigma_{\mathbf{j},\mathbf{j}'}$ *as defined in* (4.11) *then the compressed matrix* $\mathbb{A}_J^{compr}$ *satisfies Requirement 1 with*

$$\varepsilon = \max\left\{ c_i^{-(2\widetilde{d}+2q+n-1)},\ e_i^{-(\widetilde{d}+2q+n-1)} \ :\ i = 1,\ldots,n \right\}. \tag{5.2}$$

If $\widetilde{\sigma}_{\mathbf{j},\mathbf{j}'} = \sigma'_{\mathbf{j},\mathbf{j}'}$ *as defined in* (4.14) *then the compressed matrix* $\mathbb{A}_J^{compr}$ *satisfies Requirement 2.*

Proof. For sake of brevity, we only prove the result in case $\widetilde{\sigma}_{\mathbf{j},\mathbf{j}'} = \sigma_{\mathbf{j},\mathbf{j}'}$. For $\widetilde{\sigma}_{\mathbf{j},\mathbf{j}'} = \sigma'_{\mathbf{j},\mathbf{j}'}$ the result follows analogously by replacing Theorem 4.3 by Corollary 4.6 in the analysis below.

Throughout this proof, we further assume without loss of generality that $j_s' \le j_s$, $s = 1,\ldots,n$. For all other index combinations, the result follows in the same fashion.

To analyze $\mathbb{A}_J^{cpr-1}$ it is sufficient to show that, for arbitrary but fixed $i \in \{1,\ldots,n\}$ and $\mathcal{I} \subset \{1,\ldots,n\}$, the perturbation matrix $\mathbb{T}^{i,\mathcal{I}}$ with blocks $\mathbb{T}_{\mathbf{j},\mathbf{j}'}^{i,\mathcal{I}}$ defined by

$$\left[\mathbb{T}_{\mathbf{j},\mathbf{j}'}^{i,\mathcal{I}}\right]_{(\mathbf{k},\mathbf{k}')} := \begin{cases} \left[\mathbb{A}_J\right]_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')}, & \text{if } \begin{cases} \delta_{x_i} > C_{\mathbf{j},\mathbf{j}'}^{i,\mathcal{I}}, \\ \delta_{x_s} > 2^{-\min\{j_s,j_s'\}}\ \forall\, s \notin \mathcal{I}, \\ \delta_{x_s} \le 2^{-\min\{j_s,j_s'\}}\ \forall\, s \in \mathcal{I}, \end{cases} \\ 0, & \text{otherwise.} \end{cases}$$

satisfies the requirements of Theorem 4.3. To this end, we shall apply Schur's Lemma (cf. [22, Section VIII.4]) to estimate $\big\|\mathbb{T}^{i,\mathcal{I}}_{\mathbf{j},\mathbf{j}'}\big\|_2$.

Denote $\mathbb{T}^{i,\mathcal{I}}_{\mathbf{j},\mathbf{j}'} = \big(t^{i,\mathcal{I}}_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')}\big)_{(\mathbf{k},\mathbf{k}')}$ and, to simplify notation, introduce the index set

$$\mathcal{D}^1_{\mathbf{j},\mathbf{j}'} := \{\mathbf{k} \in \nabla_{\mathbf{j}} : \delta_{x_i} > C^{i,\mathcal{I}}_{\mathbf{j},\mathbf{j}'},\, \delta_{x_s} > 2^{-j_s}\,\forall\, s \notin \mathcal{I},\, \delta_{x_s} \le 2^{-j_s}\,\forall\, s \in \mathcal{I}\}.$$

Recall that by δ_{xy} we denote the distance of the supports of two wavelets corresponding to a matrix entry. By Theorem 3.8, the column sums of $\mathbb{T}^{i,\mathcal{I}}_{\mathbf{j},\mathbf{j}'}$ may be estimated by

$$\begin{aligned}\sum_{\mathbf{k}\in\nabla_{\mathbf{j}}} \big|t^{i,\mathcal{I}}_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')}\big| &\lesssim \sum_{\mathbf{k}\in\mathcal{D}^1_{\mathbf{j},\mathbf{j}'}} 2^{-\frac{1}{2}(|\mathbf{j}|_1+|\mathbf{j}'|_1)}2^{-\widetilde{d}(j^{(1)}+j^{(2)})}\delta_{xy}^{-(n+2q+2\widetilde{d})}\\ &\lesssim 2^{-\frac{1}{2}(|\mathbf{j}|_1+|\mathbf{j}'|_1)}2^{-\widetilde{d}(j^{(1)}+j^{(2)})}\cdot\prod_{l\in\mathcal{I}}2^{j_l-j'_l}\cdot\sum_{\substack{\mathbf{k}\in\mathcal{D}^1_{\mathbf{j},\mathbf{j}'}\\ k_s:s\notin\mathcal{I}}}\delta_{xy}^{-(n+2q+2\widetilde{d})},\end{aligned}$$

where we have used that for each $l \in \mathcal{I}$ there holds $\delta_{x_l} \le 2^{-j_l}$ and hence there are only $\mathcal{O}(2^{j_l-j'_l})$ non-zero entries per column.

Furthermore, for each $s \in \mathcal{I}^c := \{1,\dots,n\}\backslash\mathcal{I}$ there holds $\delta_{x_s} > 2^{-j_s}$, which implies that the distance of the wavelets' supports is larger than the longest edge length in coordinate direction s. Thus, denoting

$$\mathcal{X}_1 := \{x \in \mathbb{R}^{n-|\mathcal{I}|} : |x| > C^{i,\mathcal{I}}_{\mathbf{j},\mathbf{j}'}\},$$

the remaining sum may be estimated by an integral (cf. [7]) to obtain

$$\sum_{\mathbf{k}\in\nabla_{\mathbf{j}}} \big|t^{i,\mathcal{I}}_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')}\big| \lesssim 2^{-\frac{1}{2}(|\mathbf{j}|_1+|\mathbf{j}'|_1)}2^{-\widetilde{d}(j^{(1)}+j^{(2)})}\cdot\prod_{l\in\mathcal{I}}2^{j_l-j'_l}\prod_{s\in\mathcal{I}^c}2^{j_s}\cdot\int_{\mathcal{X}_1}|x|^{-(n+2q+2\widetilde{d})}dx,$$

since the sum is only taken over those matrix entries that satisfy $\delta_{xy} > C^{i,\mathcal{I}}_{\mathbf{j},\mathbf{j}'}$. Hence,

$$\begin{aligned}\sum_{\mathbf{k}\in\nabla_{\mathbf{j}}} \big|t^{i,\mathcal{I}}_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')}\big| &\lesssim 2^{-\frac{1}{2}(|\mathbf{j}|_1+|\mathbf{j}'|_1)}2^{-\widetilde{d}(j^{(1)}+j^{(2)})}\cdot\prod_{l\in\mathcal{I}}2^{j_l-j'_l}\cdot\prod_{s\in\mathcal{I}^c}2^{j_s}\cdot(C^{i,\mathcal{I}}_{\mathbf{j},\mathbf{j}'})^{-(|\mathcal{I}|+2q+2\widetilde{d})}\\ &\lesssim 2^{\frac{1}{2}(|\mathbf{j}|_1-|\mathbf{j}'|_1)}2^{-\widetilde{d}(j^{(1)}+j^{(2)})}2^{-\sum_{l\in\mathcal{I}}j'_l}(C^{i,\mathcal{I}}_{\mathbf{j},\mathbf{j}'})^{-(|\mathcal{I}|+2q+2\widetilde{d})}.\end{aligned}$$

Using that for each $l \in \mathcal{I}$ there are only $\mathcal{O}(1)$ non-zero row-entries, one analogously obtains that the row sums can be estimated by

$$\sum_{\mathbf{k}'\in\nabla_{\mathbf{j}'}} \big|t^{i,\mathcal{I}}_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')}\big| \lesssim 2^{\frac{1}{2}(|\mathbf{j}'|_1-|\mathbf{j}|_1)}2^{-\widetilde{d}(j^{(1)}+j^{(2)})}2^{-\sum_{l\in\mathcal{I}}j'_l}(C^{i,\mathcal{I}}_{\mathbf{j},\mathbf{j}'})^{-(|\mathcal{I}|+2q+2\widetilde{d})}.$$

Thus, by Schur's lemma with weights $2^{\frac{1}{2}(|\mathbf{j}'|_1-|\mathbf{j}|_1)}$, $2^{\frac{1}{2}(|\mathbf{j}|_1-|\mathbf{j}'|_1)}$ one obtains

$$\Big\|\mathbb{T}^{i,\mathcal{I}}_{\mathbf{j},\mathbf{j}'}\Big\|_2 \lesssim 2^{-(j^{(1)}+j^{(2)})\widetilde{d}}2^{-\sum_{l\in\mathcal{I}}j'_l}\cdot(C^{x}_{\mathbf{j},\mathbf{j}'})^{-(2q+2\widetilde{d})} \lesssim c_i^{-(2q+2\widetilde{d}+|\mathcal{I}|)}2^{-\sigma_{\mathbf{j},\mathbf{j}'}}. \tag{5.3}$$

Hence, by Theorem 4.3, the matrix $\mathbb{A}^{cpr-1}_J$ satisfies Requirement 1.

To analyze $\mathbb{A}_J^{cpr-2}$, consider the corresponding perturbation matrix $\mathbb{S}^{i,\mathcal{I}}$ given by

$$\left[\mathbb{S}^{i,\mathcal{I}}_{\mathbf{j},\mathbf{j}'}\right]_{(\mathbf{k},\mathbf{k}')} := \begin{cases} [\mathbb{A}_J]_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')}, & \text{if } \begin{cases} \sigma_{x_i} > E^{i,\mathcal{I}}_{\mathbf{j},\mathbf{j}'},\ i \in \mathcal{I}, \\ \delta_{x_s} > 2^{-\min\{j_s,j'_s\}}\ \forall\, s \notin \mathcal{I}, \\ \delta_{x_s} \le 2^{-\min\{j_s,j'_s\}}\ \forall\, s \in \mathcal{I}, \end{cases} \\ 0, & \text{otherwise.} \end{cases}$$

By Theorem 3.9, the entries of $\mathbb{S}^{i,\mathcal{I}}$ can be estimated by

$$\left|s^{i,\mathcal{I}}_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')}\right| \lesssim 2^{-\widetilde{d}j_i} 2^{j'_i} 2^{-\frac{1}{2}(|\mathbf{j}|_1+|\mathbf{j}'|_1)} \sigma_{x_i}^{-(2q+\widetilde{d}+n-1)}.$$

Introduce the index set

$$\mathcal{D}^2_{\mathbf{j},\mathbf{j}'} := \{\mathbf{k} \in \nabla_{\mathbf{j}} : \sigma_{x_i} > E^{i,\mathcal{I}}_{\mathbf{j},\mathbf{j}'},\ \delta_{x_s} > 2^{-j_s}\ \forall\, s \notin \mathcal{I},\ \delta_{x_s} \le 2^{-j_s}\ \forall\, s \in \mathcal{I}\}.$$

Denoting $\mathcal{X}_2 := \{x \in \mathbb{R}^{n-|\mathcal{I}|} : |x| > E^{i,\mathcal{I}}_{\mathbf{j},\mathbf{j}'}\}$, as for the first compression one herewith obtains

$$\begin{aligned} &\sum_{\mathbf{k}\in\nabla_{\mathbf{j}}} 2^{\frac{1}{2}(|\mathbf{j}'|_1-|\mathbf{j}|_1)} \left|s^{i,\mathcal{I}}_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')}\right| \\ \lesssim\ & 2^{|\mathbf{j}'|_1-|\mathbf{j}|_1} 2^{-(|\mathbf{j}'|_1-j'_i)} 2^{-\widetilde{d}j_i} \sum_{\mathbf{k}\in\mathcal{D}^2_{\mathbf{j},\mathbf{j}'}} \sigma_{x_i}^{-(2q+\widetilde{d}+n-1)} \\ \lesssim\ & 2^{|\mathbf{j}'|_1-|\mathbf{j}|_1} 2^{-(|\mathbf{j}'|_1-j'_i)} 2^{-\widetilde{d}j_i} \cdot \prod_{l\in\mathcal{I}} 2^{j_l-j'_l} \cdot \prod_{s\in\mathcal{I}^c} 2^{j_s} \int_{\mathcal{X}_2} |x|^{-(2q+\widetilde{d}+n-1)} dx, \end{aligned}$$

since for all $s \in \mathcal{I}^c$ there holds $\delta_{x_s} = \sigma_{x_s}$. Thus, for the weighted column sums of $\mathbb{S}^{i,\mathcal{I}}_{\mathbf{j},\mathbf{j}'}$ one finds

$$\begin{aligned} \sum_{\mathbf{k}\in\nabla_{\mathbf{j}}} 2^{\frac{1}{2}(|\mathbf{j}'|_1-|\mathbf{j}|_1)} \left|s^{i,\mathcal{I}}_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')}\right| &\lesssim 2^{-\widetilde{d}j_i} 2^{-\sum_{l\in\mathcal{I}\setminus\{i\}} j'_l} (E^{i,\mathcal{I}}_{\mathbf{j},\mathbf{j}'})^{-(2q+\widetilde{d}+|\mathcal{I}|-1)} \\ &\lesssim e_i^{-(2q+\widetilde{d}+|\mathcal{I}|-1)} 2^{-\sigma_{\mathbf{j},\mathbf{j}'}}. \end{aligned}$$

Analogously, as above one may estimate the row sums by

$$\sum_{\mathbf{k}'\in\nabla_{\mathbf{j}'}} 2^{\frac{1}{2}(|\mathbf{j}|_1-|\mathbf{j}'|_1)} \left|s^{i,\mathcal{I}}_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')}\right| \lesssim e_i^{-(2q+\widetilde{d}+|\mathcal{I}|-1)} 2^{-\sigma_{\mathbf{j},\mathbf{j}'}}.$$

Hence, by Theorem 4.3, the matrix $\mathbb{A}_J^{cpr-2}$ satisfies Requirement 1 and the overall consistency of $\mathbb{A}_J^{compr}$ follows. □

6 Complexity estimates

In this section we analyze the complexity of the compression scheme assuming that we have a large number of vanishing moments at hand. For more sophisticated methods and detailed results (which are less restrictive on the vanishing moments) we refer to [28, Sections 2.4, 4.6].

Theorem 6.1. *For arbitrary dimension $n \geq 2$, any operator of order $2\widetilde{q} \geq 0$ and approximation order $d \geq 2$ there exists a number $\widetilde{d}_0 \geq d$ such that the number of non-zero entries in the matrix $\mathbb{A}_J^{compr}$ defined by the compression scheme in Section 5 with $\widetilde{\sigma}_{\mathbf{j},\mathbf{j}'} = \sigma_{\mathbf{j},\mathbf{j}'}$ is $\mathcal{O}\left(2^J J^{2(n-1)}\right)$, provided that the underlying wavelets admit $\widetilde{d} \geq \widetilde{d}_0$ vanishing moments.*

Proof. Fix any $\mathcal{I} \subset \{1, \ldots, n\}$. Since $\mathcal{I}$ is arbitrary, it suffices to show that there are $\mathcal{O}\left(2^J J^{2(n-1)}\right)$ entries $\langle \mathcal{A}\psi_{\mathbf{j},\mathbf{k}},\, \psi_{\mathbf{j}',\mathbf{k}'}\rangle$ of $\mathbb{A}_J^{compr}$ that satisfy

$$\begin{aligned} \delta_{x_i} &\leq 2^{-\min\{j_i, j_i'\}}, \qquad &\text{for all } i \in \mathcal{I},\\ \delta_{x_i} &\geq 2^{-\min\{j_i, j_i'\}}, \qquad &\text{for all } i \notin \mathcal{I}. \end{aligned} \tag{6.1}$$

Based on the compression scheme of Section 5, in each matrix block $\mathbb{A}_{\mathbf{j},\mathbf{j}'}$ of $\mathbb{A}_J^{compr}$ we divide the coordinate directions into three groups. Let

$$D_1 \; := \; \left\{ s \in \mathcal{I} \,:\, 2^{-\min\{j_s, j_s'\}} \leq E_{\mathbf{j},\mathbf{j}'}^{s,\mathcal{I}} \right\}, \tag{6.2}$$

$$D_2 \; := \; \left\{ t \in \mathcal{I} \,:\, E_{\mathbf{j},\mathbf{j}'}^{t,\mathcal{I}} \leq 2^{-\min\{j_t, j_t'\}} \right\}, \tag{6.3}$$

$$D_3 \; := \; \left\{ i \in \{1, \ldots, n\} \backslash \mathcal{I} \right\}. \tag{6.4}$$

Obviously $D_1 \cup D_2 \cup D_3 = \{1, \ldots, n\}$. By definition of the compression scheme, the number $\#\mathbb{A}_{\mathbf{j},\mathbf{j}'}$ of non-zero entries that satisfy (6.1) in each matrix block $\mathbb{A}_{\mathbf{j},\mathbf{j}'}$ can be bounded by

$$\begin{aligned} \#\mathbb{A}_{\mathbf{j},\mathbf{j}'} &= \mathcal{O}\left(\prod_{s\in D_1} 2^{j_s + j_s'} 2^{-\min\{j_s, j_s'\}} \cdot \prod_{t \in D_2} 2^{j_t + j_t'} E_{\mathbf{j},\mathbf{j}'}^{t,\mathcal{I}} \cdot \prod_{i \in D_3} 2^{j_i + j_i'} C_{\mathbf{j},\mathbf{j}'}^{i,\mathcal{I}} \right)\\ &= \mathcal{O}\left(\prod_{s\in D_1} 2^{\max\{j_s, j_s'\}} \cdot \prod_{t \in D_2} 2^{j_t + j_t'} 2^{\frac{\widetilde{\sigma}_{\mathbf{j},\mathbf{j}'} - \widetilde{d}\max\{j_t, j_t'\} - \sum_{k\in\mathcal{I}\backslash\{t\}} \min\{j_k, j_k'\}}{\widetilde{d} + 2q + |\mathcal{I}| - 1}} \right.\\ &\qquad \left. \times \prod_{i \in D_3} 2^{j_i + j_i'} 2^{\frac{\widetilde{\sigma}_{\mathbf{j},\mathbf{j}'} - \widetilde{d}(j^{(1)} + j^{(2)}) - \sum_{k \in \mathcal{I}} \min\{j_k, j_k'\}}{2\widetilde{d} + 2q + |\mathcal{I}|}} \right). \end{aligned} \tag{6.5}$$

To simplify this notation, from now on we assume without loss of generality that $j_l' \leq j_l$ for all $l = 1, \ldots, n$. The result for all other index combinations follows analogously.

Regrouping the single factors in (6.5) corresponding to their level index yields

$$\#\mathbb{A}_{\mathbf{j},\mathbf{j}'} = \mathcal{O}\Big(C_0 \cdot \prod_{s \in D_1} S_s \cdot \prod_{t \in D_2} T_t \cdot \prod_{i \in D_3} I_i \Big), \tag{6.6}$$

where now the particular form of C_0, S_s, T_t, I_i depends on the particular form of $\widetilde{\sigma}_{\mathbf{j},\mathbf{j}'} = \sigma_{\mathbf{j},\mathbf{j}'}$ as defined in (4.11). We analyze the different possible cases separately: denote

$$N := \frac{\#D_2}{\widetilde{d} + 2q + |\mathcal{I}| - 1} + \frac{\#D_3}{2\widetilde{d} + 2q + |\mathcal{I}|}. \tag{6.7}$$

In case $\widetilde{\sigma}_{\mathbf{j},\mathbf{j}'} = 2J(d' - \widetilde{q}) - d'(|\mathbf{j}|_1 + |\mathbf{j}'|_1)$ one obtains

$$\begin{aligned} C_0 &= 2^{N(2J(d'-\widetilde{q}))}, \\ S_m &= 2^{j_m} 2^{-N(d'(j_m+j'_m)+j'_m)}, \\ T_m &= 2^{j_m+j'_m} 2^{-\frac{\widetilde{d}}{\widetilde{d}+2q+|\mathcal{I}|-1} j_m} 2^{-N(d'(j_m+j'_m)+j'_m)} 2^{\frac{1}{\widetilde{d}+2q+|\mathcal{I}|-1} j'_m}, \\ I_m &= 2^{j_m+j'_m} 2^{-\frac{\widetilde{d}}{2\widetilde{d}+2q+|\mathcal{I}|}(j^{(1)}+j^{(2)})} 2^{-Nd'(j_m+j'_m)}, \end{aligned} \tag{6.8}$$

for $m = 1, \ldots, n$. Here, each factor (except for C_0) depends on exactly one coordinate direction.

By definition of D_1, D_2, there holds $S_s \leq T_s$ for all $s \in D_1$. In order to keep notation feasible we can thus assume $\#D_2 = |\mathcal{I}|$, $D_1 = \emptyset$ and estimate

$$\#\mathbb{A}_{\mathbf{j},\mathbf{j}'} = \mathcal{O}\Bigg(C_0 \cdot \prod_{t \in D_2} T_t \cdot \prod_{i \in D_3} I_i \Bigg). \tag{6.9}$$

Note that the estimate (6.9) introduces a certain suboptimality in the assumptions on the number of vanishing moments that we shall need below. For slightly improved estimates we refer to [28].

Denote

$$\widetilde{j}_m := \frac{j_m + j'_m}{2}, \qquad m = 1, \ldots, n. \tag{6.10}$$

For sufficiently large $\widetilde{d} \geq N - 1/(\widetilde{d} + 2q + \mathcal{I} - 1)$ one obtains for any $t \in D_2$,

$$\begin{aligned} T_t &= 2^{\left(1-\frac{\widetilde{d}}{\widetilde{d}+2q+|\mathcal{I}|-1}\right) j_t} 2^{\left(1-N+\frac{1}{\widetilde{d}+2q+|\mathcal{I}|-1}\right) j'_t} 2^{-Nd'(j_t+j'_t)} \\ &\leq 2^{2\widetilde{j}_t} 2^{-\frac{\widetilde{d}-1}{\widetilde{d}+2q+|\mathcal{I}|-1}\widetilde{j}_t} 2^{-N\widetilde{j}_t} 2^{-N2d'\widetilde{j}_t} =: \widetilde{T}_t. \end{aligned}$$

In addition, one immediately obtains for $i \in D_3$,

$$I_i \leq 2^{2\widetilde{j}_i} 2^{-\frac{2\widetilde{d}}{2\widetilde{d}+2q+|\mathcal{I}|}\widetilde{j}_i} 2^{-N2d'\widetilde{j}_i} =: \widetilde{I}_i.$$

Thus,

$$\#\mathbb{A}_{\mathbf{j},\mathbf{j}'} = \mathcal{O}\Bigg(C_0 \cdot 2^{2|\widetilde{\mathbf{j}}|_1} 2^{-N2d'|\widetilde{\mathbf{j}}|_1} \cdot \prod_{t \in D_2} 2^{-\frac{\widetilde{d}-1}{\widetilde{d}+2q+|\mathcal{I}|-1}\widetilde{j}_t} \cdot \prod_{i \in D_3} 2^{-\frac{2\widetilde{d}}{2\widetilde{d}+2q+|\mathcal{I}|}\widetilde{j}_i} \Bigg). \tag{6.11}$$

Since $D_2 = \mathcal{I}$ and $D_3 = \{1, \ldots, n\} \backslash \mathcal{I}$, there holds

$$N = \frac{|\mathcal{I}|}{\widetilde{d} + 2q + |\mathcal{I}| - 1} + \frac{n - |\mathcal{I}|}{2\widetilde{d} + 2q + |\mathcal{I}|}.$$

Herewith, a basic calculation shows that for a sufficiently large number of vanishing moments $\widetilde{d}$, the maximum of (6.11) is obtained when $\#D_2 = n$ and hence $N = n/(\widetilde{d} + 2q + n - 1)$. Then (6.11) reads

$$\#\mathbb{A}_{\mathbf{j},\mathbf{j}'} = \mathcal{O}\Bigg(2^{N(2d'-2\widetilde{q})J} 2^{2|\widetilde{\mathbf{j}}|_1} 2^{-N2d'|\widetilde{\mathbf{j}}|_1} 2^{-\frac{\widetilde{d}-1}{\widetilde{d}+2q+n-1}|\widetilde{\mathbf{j}}|_1} \Bigg), \tag{6.12}$$

and (for sufficiently large $\widetilde{d}$) the right hand side of (6.12) is monotonically increasing in $|\widetilde{\mathbf{j}}|_1 \leq J$. Hence,

$$\begin{aligned} \#\mathbb{A}_{\mathbf{j},\mathbf{j}'} &= \mathcal{O}\left(2^{N(2d'-2\widetilde{q})J}2^{2J}2^{-N2d'J}2^{-\frac{\widetilde{d}-1}{\widetilde{d}+2q+n-1}J}\right), \\ &= \mathcal{O}\left(2^{2J}2^{-\frac{\widetilde{d}+2\widetilde{q}+n-1}{\widetilde{d}+2q+n-1}J}\right) = \mathcal{O}\big(2^J\big), \end{aligned} \tag{6.13}$$

since $q \leq \widetilde{q}$. In case $\widetilde{\sigma}_{\mathbf{j},\mathbf{j}'} = 2J(d'-\widetilde{q}) - d'(|\mathbf{j}|_1 + |\mathbf{j}'|_1)$, summing over all matrix blocks yields that there are indeed $\mathcal{O}\left(2^J J^{2(n-1)}\right)$ entries satisfying (6.1), because the maximum $\mathcal{O}(2^J)$ in (6.13) is only obtained when $|\widetilde{\mathbf{j}}|_1 = J$ (i.e. $j_m = j'_m = J/n$ for all $m = 1,\ldots,n$) and the number of non-zero entries per block is monotonically decreasing as $|\widetilde{\mathbf{j}}|_1$ decreases.

To prove the claimed result in case $\widetilde{\sigma}_{\mathbf{j},\mathbf{j}'} = -\widetilde{q}(|\mathbf{j}|_\infty + |\mathbf{j}'|_\infty)$ we proceed similarly: note that in this case (6.6) holds with

$$\begin{aligned} C_0 &= 2^{-N\widetilde{q}(|\mathbf{j}|_\infty + |\mathbf{j}'|_\infty)}, \\ S_s &= 2^{j_s}2^{-Nj'_s}, \\ T_t &= 2^{j_t+j'_t}2^{-\frac{\widetilde{d}}{\widetilde{d}+2q+|\mathcal{I}|-1}j_t}2^{-Nj'_t}2^{\frac{1}{\widetilde{d}+2q+|\mathcal{I}|-1}j'_t}, \\ I_i &= 2^{j_i+j'_i}2^{-\frac{\widetilde{d}}{2\widetilde{d}+2q+|\mathcal{I}|}(j^{(1)}+j^{(2)})}. \end{aligned} \tag{6.14}$$

As above, we assume without loss of generality that $D_1 = \emptyset$, because this provides a worst but admissible case. Herewith

$$\begin{aligned} \#\mathbb{A}_{\mathbf{j},\mathbf{j}'} &= \mathcal{O}\left(C_0 \cdot 2^{|\mathbf{j}|_1+|\mathbf{j}'|_1} \cdot \prod_{t\in D_2} 2^{-\frac{\widetilde{d}}{\widetilde{d}+2q+|\mathcal{I}|-1}j_t}2^{-Nj'_t}2^{\frac{1}{\widetilde{d}+2q+|\mathcal{I}|-1}j'_t} \cdot \prod_{i\in D_3} 2^{-\frac{2\widetilde{d}}{2\widetilde{d}+2q+|\mathcal{I}|}j_i}\right) \\ &= \mathcal{O}\left(C_0 \cdot 2^{2|\widetilde{\mathbf{j}}|_1} \cdot \prod_{t\in D_2} 2^{-\frac{\widetilde{d}-1}{\widetilde{d}+2q+|\mathcal{I}|-1}\widetilde{j}_t}2^{-N\widetilde{j}'_t} \prod_{i\in D_3} \cdot 2^{-\frac{2\widetilde{d}}{2\widetilde{d}+2q+|\mathcal{I}|}\widetilde{j}_i}\right), \end{aligned} \tag{6.15}$$

with $\widetilde{j}_m$, $m = 1,\ldots,n$, as in (6.10). Since $N \to 0$ as $\widetilde{d} \to \infty$, provided a sufficiently large number of vanishing moments the right hand side of (6.15) reaches its maximum when $\#D_2 = |\mathcal{I}| = n$. Thus,

$$\#\mathbb{A}_{\mathbf{j},\mathbf{j}'} = \mathcal{O}\left(2^{-N\widetilde{q}(|\mathbf{j}|_\infty+|\mathbf{j}'|_\infty)} \cdot 2^{2|\widetilde{\mathbf{j}}|_1} \cdot 2^{-\frac{\widetilde{d}-1}{\widetilde{d}+2q+n-1}|\widetilde{\mathbf{j}}|_1}2^{-N|\widetilde{\mathbf{j}}|_1}\right). \tag{6.16}$$

Clearly (6.16) becomes maximal only when $|\mathbf{j}|_\infty + |\mathbf{j}'|_\infty$ is minimal, i.e. $\widetilde{j}_1 = \ldots = \widetilde{j}_n$. Thus,

$$\begin{aligned} \#\mathbb{A}_{\mathbf{j},\mathbf{j}'} &= \mathcal{O}\left(2^{2n\widetilde{j}_1} \cdot 2^{-N2\widetilde{q}\widetilde{j}_1} \cdot 2^{-\frac{n(\widetilde{d}-1)}{\widetilde{d}+2q+n-1}\widetilde{j}_1}2^{-Nn\widetilde{j}_1}\right) \\ &= \mathcal{O}\left(2^{2n\widetilde{j}_1} \cdot 2^{-\frac{n2\widetilde{q}}{\widetilde{d}+2q+n-1}\widetilde{j}_1} \cdot 2^{-\frac{n(\widetilde{d}-1)}{\widetilde{d}+2q+n-1}\widetilde{j}_1}2^{-\frac{n^2}{\widetilde{d}+2q+n-1}\widetilde{j}_1}\right) \\ &= \mathcal{O}\big(2^{n\widetilde{j}_1}\big) = \mathcal{O}\big(2^J\big), \end{aligned} \tag{6.17}$$

where the maximum is only obtained when $\widetilde{j}_1$ is maximal, i.e. $\widetilde{j}_1 = \ldots = \widetilde{j}_n = J/n$.

Hence, as above, in case $\widetilde{\sigma}_{\mathbf{j},\mathbf{j}'} = -\widetilde{q}(|\mathbf{j}|_\infty + |\mathbf{j}'|_\infty)$, summing over all matrix blocks yields that there are indeed $\mathcal{O}\left(2^J J^{2(n-1)}\right)$ entries satisfying (6.1).

Finally, the cases $\widetilde{\sigma}_{\mathbf{j},\mathbf{j}'} = J(d'-\widetilde{q}) - \widetilde{q}|\mathbf{j}'|_\infty - d'|\mathbf{j}|_1$ and $\widetilde{\sigma}_{\mathbf{j},\mathbf{j}'} = J(d'-\widetilde{q}) - d'|\mathbf{j}'|_1 - \widetilde{q}|\mathbf{j}|_\infty$ follow in the same fashion as the above, since these choices of $\widetilde{\sigma}_{\mathbf{j},\mathbf{j}'}$ are simply a combination of the two previous ones. For sake of brevity we omit the details here. □

7 s^*-compressibility

In this final section, we briefly illustrate that the above complexity results also imply that the stiffness matrices of the non-local operators under consideration are s^*-compressible in the sense of [5, 15, 31] – with essentially dimension independent s^*. This shows that, in order to solve the corresponding integrodifferential equations one may employ *adaptive* wavelet algorithms as in [4, 5, 14] that converge with the rate of best approximation by an arbitrary linear combination of N wavelets (so-called best N-term approximation).

More precisely, to introduce the required notation we convert the original operator equation (2.1) into an infinite matrix-vector system

$$\mathbb{B}\mathbf{u} = \mathbf{f}, \tag{7.1}$$

where $\mathbb{B}$ denotes the infinite stiffness matrix of $\mathcal{B}$. We shall use $\|\cdot\|$ to denote $\|\cdot\|_{\ell^2\to\ell^2}$ and define

Definition 7.1. *For $s^* > 0$, an infinite matrix $\mathbb{B}$ is called s^*-compressible if for each $J \in \mathbb{N}_0$ an infinite matrix $\mathbb{B}_J$ can be constructed that has in each row and column $\mathcal{O}(2^J)$ non-zero entries and satisfies*

$$\|\mathbb{B} - \mathbb{B}_J\| \lesssim 2^{-Js}, \quad \text{for any constant } s < s^*. \tag{7.2}$$

From the consistency Theorem 5.1 and the complexity Theorem 6.1 one obtains that $\mathbb{B}$ in (7.1) is indeed s^*-compressible:

Corollary 7.2. *Suppose the solution u of* (2.1) *satisfies $u \in \mathcal{H}^d$. Then the infinite matrix $\mathbb{B}$ of the non-local operator $\mathcal{B}$ is s^*-compressible. If in addition the number $\widetilde{d}$ of the wavelets' vanishing moments is sufficiently large, there holds*

$$d - \widetilde{q} \leq s^*, \tag{7.3}$$

where, as above, d denotes the approximation order of the wavelet basis Ψ and $2\widetilde{q}$ is the order of $\mathcal{B}$.

Proof. *(Sketch)* For $J \in \mathbb{N}_0$, let $\mathbb{B}_J$ be the infinite matrix given by

$$[\mathbb{B}_J]_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')} := \begin{cases} 0, & \text{if } |\mathbf{j}|_1 > J \text{ or } |\mathbf{j}'|_1 > J, \\ 0, & \text{if } \begin{cases} \psi_{\mathbf{j},\mathbf{k}}, \psi_{\mathbf{j}',\mathbf{k}'} \in \widehat{V}_J, \\ \text{but } \mathbb{B}_J^{compr} = 0, \end{cases} \\ [\mathbb{B}]_{(\mathbf{j},\mathbf{k})(\mathbf{j}',\mathbf{k}')}, & \text{otherwise,} \end{cases}$$

for any $\mathbf{j}, \mathbf{j}' \in \mathbb{N}_0^n$ and $\mathbf{k} \in \nabla_{\mathbf{j}}$, $\mathbf{k}' \in \nabla_{\mathbf{j}'}$. Here $\mathbb{B}_J^{compr}$ denotes the (finite) compressed stiffness matrix of $\mathcal{B}$ at level J as given by the compression scheme of Section 5 with $\widetilde{\sigma}_{\mathbf{j},\mathbf{j}'} = \sigma_{\mathbf{j},\mathbf{j}'}$ as defined in (4.11).

By Theorem 6.1, in each row and column of $\mathbb{B}_J$ there are $\mathcal{O}(2^J)$ non-zero entries (provided $\widetilde{d}$ is chosen sufficiently large, depending on the dimension n). Furthermore, by Theorem 5.1, the finite matrix $\mathbb{B}_J^{compr}$ satisfies Requirement 1. Setting $\theta_1 = 1$, $\theta_2 = 0$ in (4.6) one may thus proceed as in the proofs of [34, Theorems 2.3, 3.3] to obtain

$$\|\mathbb{B} - \mathbb{B}_J\| \lesssim 2^{-Js}, \qquad \text{for all } s < d - \widetilde{q}.$$

For sake of brevity, we omit the details here. □

Remark 7.3. *One central advantage of best N-term approximations in the sense of [4, 5, 14] is that it can be based on* Besov *regularity which in general is a much milder condition than assuming the corresponding Sobolev regularity. Since we are working on sparse tensor product spaces based on mixed* Sobolev *smoothness, future work is required to exploit this advantage here. For best N-term approximations on sparse tensor products based on Besov regularity we refer to [24].*

Remark 7.4. *In this work we have provided a compression scheme exclusively for operators of Hörmander type. There are however applications (e.g. Mathematical Finance) where* anisotropic *integrodifferential operators occur that are not covered by the described schemes. In [28, 29] a wavelet compression scheme is constructed especially for such anisotropic operators (cf. [13, 26, 27]).*

References

[1] G. Beylkin, R. Coifman, and V. Rokhlin. The fast wavelet transform and numerical algorithms. *Comm. Pure and Appl. Math.*, 44:141–183, 1991.

[2] J. H. Bramble, A. Cohen, and W. Dahmen. *Multiscale problems and methods in numerical simulations*, volume 1825 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2003.

[3] H.-J. Bungartz and M. Griebel. A note on the complexity of solving Poisson's equation for spaces of bounded mixed derivatives. *J. Complexity*, 15(2):167–199, 1999.

[4] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods for elliptic operator equations: convergence rates. *Math. Comp.*, 70:27–75 (electronic), 2001.

[5] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods. II. Beyond the elliptic case. *Found. Comput. Math.*, 2:203–245, 2002.

[6] A. Cohen, I. Daubechies, and J.-C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 45(5):485–560, 1992.

[7] W. Dahmen, H. Harbrecht, and R. Schneider. Compression techniques for boundary integral equations—asymptotically optimal complexity estimates. *SIAM J. Numer. Anal.*, 43(6):2251–2271 (electronic), 2006.

[8] W. Dahmen, A. Kunoth, and K. Urban. Biorthogonal spline wavelets on the interval—stability and moment conditions. *Appl. Comput. Harmon. Anal.*, 6(2):132–196, 1999.

[9] W. Dahmen, S. Prössdorf, and R. Schneider. Wavelet approximation methods for pseudodifferential equations. II. Matrix compression and fast solution. *Adv. Comput. Math.*, 1(3-4):259–335, 1993.

[10] W. Dahmen, S. Prössdorf, and R. Schneider. Multiscale methods for pseudo-differential equations on smooth closed manifolds. In *Wavelets: theory, algorithms, and applications (Taormina, 1993)*, volume 5 of *Wavelet Anal. Appl.*, pages 385–424. Academic Press, San Diego, CA, 1994.

[11] W. Dahmen and R. Schneider. Wavelets with complementary boundary conditions—function spaces on the cube. *Results Math.*, 34(3-4):255–293, 1998.

[12] A. Ern and J.-L. Guermond. *Theory and practice of Finite Elements*. Springer Verlag, New York, 2004.

[13] W. Farkas, N. Reich, and C. Schwab. Anisotropic stable Lévy copula processes—analytical and numerical aspects. *Math. Models Methods Appl. Sci.*, 17(9):1405–1443, 2007.

[14] T. Gantumur, H. Harbrecht, and R. Stevenson. An optimal adaptive wavelet method without coarsening of the iterands. *Math. Comp.*, 76:615–629 (electronic), 2007.

[15] T. Gantumur and R. Stevenson. Computation of differential operators in wavelet coordinates. *Math. Comp.*, 75:697–709 (electronic), 2006.

[16] M. Griebel and S. Knapek. *Optimized General Sparse Grid Approximation Spaces for Operator Equations*. Submitted. Available as SFB611 preprint No 402, University of Bonn, 2008.

[17] M. Griebel, P. Oswald, and T. Schiekofer. Sparse grids for boundary integral equations. *Numer. Math.*, 83(2):279–312, 1999.

[18] H. Harbrecht and R. Schneider. Biorthogonal wavelet bases for the boundary element method. *Math. Nachr.*, 269/270:167–188, 2004.

[19] H. Harbrecht and R. Schneider. Wavelet Galerkin schemes for boundary integral equations—implementation and quadrature. *SIAM J. Sci. Comput.*, 27:1347–1370 (electronic), 2006.

[20] L. Hörmander. *Linear partial differential operators*. Grundlehren der mathematischen Wissenschaften, Vol. 116, Springer Verlag, Berlin, 1963.

[21] S. Knapek and F. Koster. Integral operators on sparse grids. *SIAM J. Numer. Anal.*, 39(5):1794–1809 (electronic), 2001/02.

[22] Y. Meyer. *Opérateurs de Calderón-Zygmund*. Hermann, Paris, 1990.

[23] H. Nguyen and R. Stevenson. Finite Elements on manifolds. *IMA J. Numer. Math.*, 23:149–173, 2003.

[24] P.-A. Nitsche. Best N term approximation spaces for tensor product wavelet bases. *Constr. Approx.*, 24(1):49–70, 2006.

[25] P. Oswald. On N-term approximation by Haar functions in H^s-norms. In S.M. Nikolskij, B.S. Kashin, and A.D. Izaak, editors, *Metric Function Theory and Related Topics in Analysis*, pages 137–163. AFC, Moscow, 1999.

[26] N. Reich. *Anisotropic Operator Symbols Arising From Multivariate Jump Processes*. Submitted. Research report 2008-25, Seminar for Applied Mathematics, ETH Zürich, 2008.

[27] N. Reich. Multiscale analysis for jump processes in finance. In K. Kunisch, G. Of, and O. Steinbach, editors, *Proceedings of ENUMATH 2007*. Springer Verlag, 2008. To appear.

[28] N. Reich. *Wavelet Compression of Anisotropic Integrodifferential Operators on Sparse Tensor Product Spaces*. PhD Thesis 17661, ETH Zürich, 2008. `http://e-collection.ethbib.ethz.ch/view/eth:30174`.

[29] N. Reich. *Wavelet Compression of Anisotropic Integrodifferential Operators on Sparse Tensor Product Spaces*. In preparation. Research report, Seminar for Applied Mathematics, ETH Zürich, 2008.

[30] R. Schneider. *Multiskalen- und Wavelet-Matrixkompression: Analysisbasierte Methoden zur Lösung großer vollbesetzter Gleichungssysteme*. B.G. Teubner, Stuttgart, 1998.

[31] C. Schwab and R. Stevenson. Adaptive wavelet algorithms for elliptic PDE's on product domains. *Math. Comp.*, 77(261):71–92 (electronic), 2008.

[32] R.E. Showalter. *Monotone Operators in Banach Space and Nonliner Partial Differential Equations*. American Mathematical Society, Rhode Island, 1997.

[33] E.M. Stein. *Harmonic Analysis*. Princeton University Press, Princeton, 1993.

[34] R. Stevenson. On the compressibility operators in wavelet coordinates. *SIAM J. Math. Anal.*, 35(5):1110–1132 (electronic), 2004.

[35] M.E. Taylor. *Pseudodifferential operators*. Princeton University Press, Princeton, 1981.

[36] H. Triebel. *Interpolation theory, function spaces, differential operators*. Johann Ambrosius Barth Verlag, Heidelberg, second edition, 1995.

[37] T. von Petersdorff and C. Schwab. Fully discrete multiscale Galerkin BEM. In W. Dahmen, A. Kurdila, and P. Oswald, editors, *Multiscale wavelet methods for PDEs*, pages 287–346. Academic Press, San Diego, 1997.

[38] T. von Petersdorff and C. Schwab. Numerical solution of parabolic equations in high dimensions. *M2AN Math. Model. Numer. Anal.*, 38(1):93–127, 2004.

[39] T. von Petersdorff, C. Schwab, and R. Schneider. Multiwavelets for second-kind integral equations. *SIAM J. Numer. Anal.*, 34(6):2212–2227, 1997.