

Wavelet Compression of Anisotropic Integrodifferential Operators on Sparse Tensor Product Spaces¹

N. Reich

Research Report No. 2008-26

August 2008

Seminar für Angewandte Mathematik

Eidgenössische Technische Hochschule

CH-8092 Zürich

Switzerland

¹Supported in part by scholarships of the Zurich Graduate School in Mathematics (ZGSM) and the Huber Kudlich Foundation at ETH Zurich.

Wavelet Compression of Anisotropic Integrodifferential Operators on Sparse Tensor Product Spaces¹

N. Reich

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

Research Report No. 2008-26

August 2008

Abstract

For a class of anisotropic integrodifferential operators \mathcal{B} arising as semigroup generators of Markov processes, we present a sparse tensor product wavelet compression scheme for the Galerkin finite element discretization of the corresponding integrodifferential equations $\mathcal{B}u = f$ on $[0, 1]^n$ with possibly large n . Under certain conditions on \mathcal{B} , 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. If the conditions on \mathcal{B} are not satisfied, the complexity can be bounded by $\mathcal{O}(h^{-(1+\varepsilon)})$, where $\varepsilon \ll 1$ tends to zero with increasing number of the wavelets' vanishing moments. Here h denotes the width of the corresponding finite element mesh.

The operators under consideration are assumed to be of non-negative (anisotropic) order and admit a non-standard kernel $\kappa(\cdot, \cdot)$ that can be singular on all secondary diagonals.

Practical examples of such operators from Mathematical Finance are given and some numerical results are presented.

Keywords: Wavelet compression, sparse grids, anisotropic integrodifferential operators, norm equivalences

Subject Classification: 47A20, 65F50, 65N12, 65Y20, 68Q25, 45K05, 65N30

¹Supported in part by scholarships of the Zurich Graduate School in Mathematics (ZGSM) and the Huber Kudlich Foundation at ETH Zurich.

1. INTRODUCTION

Sparse tensor product-based wavelet compression for integrodifferential equations (PIDEs) of the form

$$\mathcal{B}u = f \quad \text{on } [0, 1]^n, \quad (1.1)$$

was introduced in [38, 46, 47] for *isotropic* integrodifferential operators \mathcal{B} with distributional Schwartz kernels $\kappa(x, y)$ that are singular only on the diagonal in $[0, 1]^n \times [0, 1]^n$. For such operators essentially dimension-independent and asymptotically optimal complexity results have been shown in [46, 47].

The assumption that \mathcal{B} admits a standard kernel however is not always satisfied in practice. For instance in Mathematical Finance the pricing of contracts on baskets of assets where the underlying is modeled by jump processes leads to equations of the form (1.1), where the integrodifferential operator \mathcal{B} only admits a kernel that can be singular on all secondary diagonals of $[0, 1]^n \times [0, 1]^n$ and all singularities are of different order. In this work we construct a sparse tensor product-based wavelet compression scheme for a wide class of such *anisotropic* operators.

Even though, due to the different singularity structures, the numerical analysis differs significantly (see [47, Section 3] and Section 5 below), from a numerical point of view the challenges arising for the discretization of anisotropic operators are essentially the same as for isotropic ones (cf. [47, Section 1]): Galerkin discretization of integrodifferential equations in general leads to linear systems with densely populated matrices of substantial size. Even on 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 $h \rightarrow 0$. The non-locality of the underlying operator thus implies that the FE stiffness matrix consists of $\mathcal{O}(h^{-2n})$ non-zero entries.

Based on tensor products of univariate wavelet basis functions, in this paper we prove that the complexity of the stiffness matrix can however be reduced to $\mathcal{O}(h^{-(1+\varepsilon)})$ with (for fixed dimension) arbitrarily small $0 < \varepsilon \ll 1$ and, under certain conditions, even to $\mathcal{O}(h^{-1} |\log h|^{2(n-1)})$ without corrupting the convergence of the original FE scheme. Our results are applicable not only to classical pseudodifferential operators but to a wide class of anisotropic integrodifferential operators. To this end, suitable symbol classes are introduced.

The anisotropic discretization technique that we present in this work relies on the following two fundamental approaches (cf. [47]):

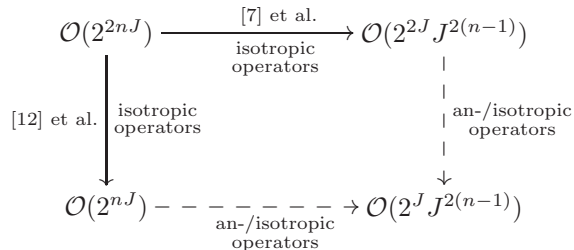
Sparse tensor product spaces as introduced in [7, 26, 27, 43] are used to overcome the ‘‘curse of dimension’’. This approach yields essentially dimension independent $\mathcal{O}(h^{-1} |\log h|^{n-1})$ degrees of freedom as $h \rightarrow 0$ while at the same time (essentially) preserving the approximation 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. [27], these results require greater smoothness of the function to be approximated than the original discretization and this extra regularity increases with the dimension n .

The non-locality of integral operators can be treated by so-called *wavelet compression*. This methodology was introduced by [5] 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. It was shown that wavelet representation yields an almost sparse representation of certain operators. In [14, 15, 57, 60] 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 [50] 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 [12]. Since it somewhat presents a finalization of the isotropic wavelet compression, we refer to [12] 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 [54].

In summary, substituting $h = 2^{-J}$, one finds

- Discretization by *sparse* tensor product spaces yields $\mathcal{O}(2^{2J} J^{2(n-1)})$ non-zero entries in the system matrix.
- Wavelet compression of general *full* tensor product spaces yields $\mathcal{O}(2^{nJ})$ non-zero matrix entries.

The following diagram illustrates the connection between the two approaches and our new results:



We shall use the notion of computational “complexity” exclusively to indicate the number of non-zero entries in a given system matrix. With an efficient implementation and quadrature as in e.g. [29] it can be shown that the overall cost of computing and assembling the system matrix is essentially of the same magnitude as its complexity.

As in [47], 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 [9, 25, 51]. This shows that, in order to solve the corresponding integrodifferential equations one may employ *adaptive* wavelet algorithms as in [8, 9, 24] that converge with the rate of best approximation by an arbitrary linear combination of N wavelets (so-called best N -term approximation, cf. [21]).

The outline of this work is as follows:

In Section 2 the abstract set-up is presented and notation is fixed.

Section 3 provides the main motivation. We briefly introduce finite element asset pricing methods that lead to the abstract anisotropic integrodifferential equations under consideration.

Based on the considerations of Section 3, in Section 4 a new class of anisotropic operator symbols is defined and examples are provided. For the corresponding class of anisotropic operators we then construct the sparse tensor wavelet compression scheme as follows:

In Section 5 fundamental estimates for the entries in the sparse tensor product-based stiffness matrix are derived.

Section 6 provides the consistency requirements that need to be satisfied by the compression scheme in order to preserve the stability and convergence properties of the sparse tensor product setting without compression.

Based on the two previous sections, in Section 7 the actual compression schemes are defined. Consistency with the sparse tensor product setting without compression is proved.

In Section 8 we provide complexity results for the constructed compression schemes. Based on [46], we show that under certain conditions the complexity of the sparse tensor product setting can be reduced to $\mathcal{O}(2^J J^{2(n-1)})$ non-zero matrix entries provided that the number of vanishing moments is sufficiently large. If these conditions are not satisfied, the complexity can be bounded by $\mathcal{O}(2^{(1+\varepsilon)J})$, where $\varepsilon < 1$ tends to zero with increasing number of vanishing moments.

Finally, in Section 9 we briefly provide some numerical experiments that confirm the analytic results.

2. GALERKIN DISCRETIZATION OF MULTIDIMENSIONAL PIDEs

The compression scheme and numerical analysis we present in this work is based on the following generic set-up (cf. [46, 47]): On $[0, 1]^n =: \square$, 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

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

and \mathcal{A} is an integral operator of possibly anisotropic order $\underline{\alpha} \in \mathbb{R}^n$, i.e. $\mathcal{A} : H^{\underline{\alpha}/2}(\square) \rightarrow H^{-\underline{\alpha}/2}(\square)$ continuously. Here, the Sobolev spaces $H^{\underline{\alpha}/2}(\square)$ and its dual $H^{-\underline{\alpha}/2}(\square)$ are defined 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 $\underline{s} \in \mathbb{R}^n$, the space $H^{\underline{s}}(\square)$ is given by

$$H^{\underline{s}}(\square) := \overline{\{\bar{u} \mid u \in C_0^\infty(\square)\}}, \quad (2.4)$$

where the closure is taken with respect to the norm of the anisotropic Sobolev space

$$H^{\underline{s}}(\mathbb{R}^n) := \left\{ f \in \mathcal{S}'(\mathbb{R}^n) : \left\| \sum_{i=1}^n (1 + \xi_i^2)^{s_i/2} \hat{f} \right\|_{L^2(\mathbb{R}^n)} < \infty \right\},$$

where \hat{f} denotes the Fourier transform of $f \in \mathcal{S}'(\mathbb{R}^n)$. We assume that the operator \mathcal{A} admits a kernel representation,

$$\mathcal{A}u(x) = \int_{\square} \kappa(x, y) u(y) dy, \quad (2.5)$$

with a distributional kernel function $\kappa(\cdot, \cdot)$ that is smooth outside the secondary diagonals $S \subset [0, 1]^n \times [0, 1]^n$, i.e.

$$S = \{(x, y) \in [0, 1]^n \times [0, 1]^n : x_i = y_i, \text{ for some } i \in \{1, \dots, n\}\}.$$

The best known example of such integral operators are

Example 2.1. (*Isotropic operators*) Any classical pseudodifferential operator $\mathcal{A} : H^q(\mathbb{R}^n) \rightarrow H^{-q}(\mathbb{R}^n)$ of order $2q \in \mathbb{R}$ with symbol in the Hörmander class $S_{1,0}^{2q}$ in the sense of [34, 55] admits a distributional kernel function $\kappa(\cdot, \cdot)$ as in (2.5).

In this case, by the Schwartz kernel theorem (cf. e.g. [53, Section VI.7]), the function $\kappa(\cdot, \cdot)$ 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

$$\left| \partial_x^{\underline{\sigma}} \partial_y^{\underline{\sigma}'} \kappa(x, y) \right| \leq c_{\underline{\sigma}, \underline{\sigma}'} |x - y|^{-(n+2q+|\underline{\sigma}|+|\underline{\sigma}'|)}, \quad \text{for all } x, y \in [0, 1]^n, \quad (2.6)$$

with some constant $c_{\underline{\sigma}, \underline{\sigma}'}$ independent of $x, y \in [0, 1]^n$. A sparse tensor product-based compression scheme especially for such, so-called *isotropic*, operators has been constructed in [47] and [46, Chapter 2]. The more general compression scheme of this work is, of course, also applicable. Throughout, we refer to estimates of the form (2.6) as Calderón-Zygmund (CZ) estimates.

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 multiindex $\tilde{\underline{\alpha}} \in \mathbb{R}^n$ of the integrodifferential operator $\mathcal{B} = \mathcal{A}_D + \mathcal{A}$ is then given by

$$\tilde{\underline{\alpha}} = \begin{cases} (2, \dots, 2), & \text{if } \mathcal{Q} > 0 \text{ and } \max\{\alpha_1, \dots, \alpha_n\} \leq 2, \\ \underline{\alpha}, & \text{otherwise.} \end{cases} \quad (2.7)$$

For the numerical solution of (2.1), we employ the Galerkin method with respect to a hierarchy of conforming trial spaces $\widehat{V}_J \subset \widehat{V}_{J+1} \subset H^{\tilde{\underline{\alpha}}/2}(\square)$. The variational problem of interest reads: find $u_J \in \widehat{V}_J$ such that,

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

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. [52, Proposition III.2.3].

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

$$\langle \mathcal{B}u, u \rangle \geq \gamma \|u\|_{H^{\tilde{\alpha}/2}(\square)}^2 - C \|u\|_{L^2(\square)}^2, \quad \text{for all } u \in H^{\tilde{\alpha}/2}(\square). \quad (2.9)$$

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

$$|\langle \mathcal{B}u, v \rangle| \leq C' \|u\|_{H^{\tilde{\alpha}/2}(\square)} \|v\|_{H^{\tilde{\alpha}/2}(\square)}. \quad (2.10)$$

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.

To simplify notation, we denote

$$\bar{\alpha} := |\underline{\alpha}|_\infty = \max\{\alpha_1, \dots, \alpha_n\}, \quad \text{for any } \underline{\alpha} \in \mathbb{R}^n.$$

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 [12] based on the construction of [10, 13, 42] and the references therein.

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

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

Using the single-scale bases constructed in [10] based on B-splines adapted to the interval $[0, 1]$ as described in [13], 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 $|\text{supp } \phi_{j,k}| \sim 2^{-j}$.

Associated to these *primal* bases are *dual* bases $\tilde{\Phi}_j = \{\tilde{\phi}_{j,k} : k \in \Delta_j\}$, i.e. there holds $\langle \phi_{j,k}, \tilde{\phi}_{j,k'} \rangle = \delta_{k,k'}$. By \tilde{d} we denote the order of $\tilde{\Phi}_j$ and assume $d \leq \tilde{d}$ for the remainder of this work. In particular, for B-splines of order d and duals of order $\tilde{d} \geq d$ such that $d + \tilde{d}$ is even the bases $\Phi_j, \tilde{\Phi}_j$ as in [13] have approximation orders d and \tilde{d} .

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

$$|\text{supp } \psi_{j,k}| \sim 2^{-j}. \quad (2.12)$$

The dual pair of wavelet bases $\Psi, \tilde{\Psi}$ is defined by $\Psi = \bigcup_{j \geq 0} \Psi_j$, $\tilde{\Psi} = \bigcup_{j \geq 0} \tilde{\Psi}_j$, with $\Psi_0 := \Phi_1$, $\tilde{\Psi}_0 := \tilde{\Phi}_1$. There holds

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

From the biorthogonality of Ψ and $\tilde{\Psi}$ one infers the so-called *cancelation property* of Ψ (see e.g. [6]), i.e.

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

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

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

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

$$\mathcal{V}_{j+1} = \mathcal{W}_{j+1} \oplus \mathcal{V}_j, \quad \text{and } \tilde{\mathcal{V}}_{j+1} = \tilde{\mathcal{W}}_{j+1} \oplus \tilde{\mathcal{V}}_j, \quad \text{for all } j \geq 0, \quad (2.15)$$

and,

$$\mathcal{V}_j = \mathcal{W}_0 \oplus \dots \oplus \mathcal{W}_j, \text{ for all } j \geq 0. \quad (2.16)$$

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. [13, 16], for the one-sided estimates we refer to [60]):

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

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

there holds the norm equivalence,

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

or the one-sided estimate,

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

In case $t = d$ there only holds,

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

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

Example 2.2. The wavelets comprise of piecewise linear continuous functions on $[0, 1]$ vanishing at the endpoints. The mesh for level $j \geq 0$ 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 $s \neq 1, 2$. 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 $s \neq 2k-2, 2k-1, 2k$.

Remark 2.3. Note that there is a strong link between the order of the operator, the approximation order of the multiresolution analysis and the number of vanishing moments of the wavelets which already restricts the possible choice of wavelet bases. In fact, the analysis of the so-called second compression we adapt from [12, 50] refers exclusively to biorthogonal spline wavelets whose singular supports are well defined and not dense in the wavelets' supports. We refer to [28] for more specific illustrations.

2.2. Sparse tensor product spaces

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

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

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,

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

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

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

On $\square = [0, 1]^n$, we define the subspace $V_J \subset H^{\underline{\alpha}/2}(\square)$ as the (full) tensor product of the spaces defined on $[0, 1]$

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

which can be written using (2.16) as

$$\begin{aligned} V_J &= \text{span} \{ \psi_{\mathbf{j},\mathbf{k}} : k_i \in \nabla_{j_i}, 0 \leq j_i \leq J, i = 1, \dots, n \} \\ &= \sum_{j_1, \dots, j_n=0}^J \mathcal{W}_{j_1} \otimes \dots \otimes \mathcal{W}_{j_n}. \end{aligned}$$

We define the regularity $\gamma > |\underline{\alpha}|_\infty/2$ of the trial spaces by

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

It is known that based on the spline wavelets constructed in Example 2.2 the regularity index satisfies $\gamma = d - 1/2$.

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

$$\widehat{V}_J := \text{span} \{ \psi_{\mathbf{j},\mathbf{k}} : k_i \in \nabla_{j_i}, i = 1, \dots, n; 0 \leq |\mathbf{j}|_1 \leq J \} = \sum_{0 \leq |\mathbf{j}|_1 \leq J} \mathcal{W}_{j_1} \otimes \dots \otimes \mathcal{W}_{j_n}. \quad (2.23)$$

One readily infers that $N_J := \dim(V_J) = \mathcal{O}(2^{2nJ})$ 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 $\mathcal{H}^{\underline{s}}([0, 1]^n)$, $\underline{s} \in \mathbb{N}_0^n$, of all measurable functions $u : [0, 1]^n \rightarrow \mathbb{R}$, such that the norm,

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

is finite. That is

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

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

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

there holds the norm equivalence

$$\sum_{(\mathbf{j},\mathbf{k})} 2^{2s_1 j_1 + \dots + 2s_n j_n} |u_{\mathbf{j},\mathbf{k}}|^2 \sim \|u\|_{\mathcal{H}^{\underline{s}}}^2, \quad \text{if } 0 \leq s_i < d - 1/2 \text{ for all } i, \quad (2.25)$$

and the one-sided bounds

$$\sum_{(\mathbf{j},\mathbf{k})} 2^{2s_1 j_1 + \dots + 2s_n j_n} |u_{\mathbf{j},\mathbf{k}}|^2 \lesssim \|u\|_{\mathcal{H}^{\underline{s}}}^2, \quad \text{if } 0 \leq s_i < d \text{ for all } i, \quad (2.26)$$

$$\sum_{\mathbf{k}} 2^{2s_1j_1+\dots+2s_nj_n} |u_{\mathbf{j},\mathbf{k}}|^2 \lesssim \|u\|_{\mathcal{H}^{\underline{s}}}^2, \quad \text{if } s_i = d \text{ for some } i. \quad (2.27)$$

Note that a slightly refined version of the one-sided bounds (2.26), (2.27) can be found in the proof of Theorem [44, Theorem 5.1].

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

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

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

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

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

2.3. Approximation rates for anisotropic operators on sparse tensor product spaces

For the proof of the following convergence and stability results, we refer to [44] and [27,59], respectively. By [59, Propositions 3.1] and [44, Theorem 5.1], the sparse tensor product projection \widehat{P}_J in (2.28) satisfies

Lemma 2.4. *Suppose for each $i = 1, \dots, n$ there holds $0 \leq \alpha_i/2 < \gamma$, with γ given by (2.22), then for $u \in H^{\underline{\alpha}/2}(\square)$ there holds*

(1) *Stability of \widehat{P}_J :*

$$\|\widehat{P}_J u\|_{H^{\underline{\alpha}/2}(\square)} \lesssim \|u\|_{H^{\underline{\alpha}/2}(\square)}. \quad (2.29)$$

(2) *Approximation property of \widehat{P}_J : let $\frac{\alpha_i}{2} \leq t_i \leq d$, $i = 1, \dots, n$, with d given by (2.11). For $u \in \mathcal{H}^{\underline{t}}(\square)$ there holds*

$$\|u - \widehat{P}_J u\|_{H^{\underline{\alpha}/2}(\square)} \lesssim \begin{cases} 2^{\overline{(\frac{\alpha}{2}-t)}J} \|u\|_{\mathcal{H}^{\underline{t}}(\square)} & \text{if } \begin{cases} \overline{\alpha} \neq 0 \text{ or} \\ t_i \neq d \text{ for all } i, \end{cases} \\ 2^{\overline{(\frac{\alpha}{2}-t)}J} J^{\frac{n-1}{2}} \|u\|_{\mathcal{H}^{\underline{t}}(\square)} & \text{otherwise,} \end{cases} \quad (2.30)$$

where we denote $\underline{t} = (t_1, \dots, t_n)$ and $\overline{(\frac{\alpha}{2}-t)} = \max\{\frac{\alpha_1}{2} - t_1, \dots, \frac{\alpha_n}{2} - t_n\}$.

Thus, there holds

Proposition 2.5. *Let $\underline{\alpha}$, as defined in (2.7), denote the order of the integrodifferential operator \mathcal{B} in (2.2). 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*

$$|\langle \mathcal{B}v_J, v_J \rangle| \geq c_1 \|v_J\|_{H^{\underline{\alpha}/2}(\square)}^2 - c_2 \|v_J\|_{L^2(\square)}^2, \quad \text{for all } v_J \in \widehat{V}_J, \quad (2.31)$$

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

$$|\langle \mathcal{B}v_J, w_J \rangle| \leq c_3 \|v_J\|_{H^{\underline{\alpha}/2}(\square)} \|w_J\|_{H^{\underline{\alpha}/2}(\square)}, \quad \text{for all } v_J, w_J \in \widehat{V}_J. \quad (2.32)$$

In particular, the variational problem (2.8) admits a unique solution.

(2) *Let u and u_J denote the solutions of the original equation (2.1) and the variational problem (2.8), respectively. The best convergence of the sparse tensor product Galerkin scheme is determined by*

$$\|u - u_J\|_{H^{\underline{\alpha}/2}(\square)} \lesssim 2^{-(d-|\underline{\alpha}|_{\infty}/2-\nu)J} \|u\|_{\mathcal{H}^{\underline{\alpha}}(\square)}, \quad (2.33)$$

provided $u \in \mathcal{H}^\rho(\square)$. The anisotropic smoothness parameter $\underline{\rho} \in \mathbb{R}_{>0}^n$ is given by

$$\rho_i = d - \left(\frac{|\tilde{\underline{\alpha}}|_\infty}{2} - \frac{\tilde{\alpha}_i}{2} \right), \quad (2.34)$$

for each $i = 1, \dots, n$, and furthermore

$$\nu = \begin{cases} \frac{(n-1)d}{nd-1}, & \text{if } \tilde{\underline{\alpha}} = (0, \dots, 0) \text{ and hence } \underline{\rho} = (d, \dots, d), \\ 0, & \text{otherwise.} \end{cases}$$

Proof. The stability estimates (2.31), (2.32) are obtained exactly as in [47, Proposition 2.1]. The convergence rate (2.33) in the sparse tensor product setting is given by [44, Proposition 5.2]. \square

Having set up the general numerical basis of our approach, in the next section we briefly describe our main applications.

3. MOTIVATION: PRICING OF FINANCIAL DERIVATIVES IN MODELS WITH JUMPS

Even though the results of the present work can be applied to a wide range of non-local operators, our main motivation arises from Mathematical Finance. In this section we illustrate how high-dimensional equations of the form (2.1) naturally occur in this field.

3.1. Pricing equations

Consider arbitrage-free values $u(x, T)$ of contingent claims on baskets of $s \in \mathbb{N}$ assets. The log-returns of the underlying assets are modeled by a Lévy or, more generally, a Feller process X with state space \mathbb{R}^n , $s \leq n$, and $X_0 = x$. For example, the compression techniques constructed in this work can be applied when X is a Lévy copula process (then $n = s \geq 2$, cf. [23]) or the price process of a Barndorff-Nielsen-Shephard (BNS) stochastic volatility model (then $s = 1$, $n = 2$, cf. [3]).

By the fundamental theorem of asset pricing (see e.g. [18]), an arbitrage free price u of an European contingent claim with payoff $g(\cdot)$ is given by the conditional expectation

$$u(x, t) = \mathbb{E}(g(X_t) \mid X_0 = x),$$

under an a-priori chosen risk-neutral martingale measure equivalent to the historical measure (see e.g. [17, 19] for measure selection criteria).

Deterministic methods to compute $u(x, T)$ are based on the solution of the corresponding backward Kolmogorov equation (for the derivation see e.g. [20], [37, Section 7.3] as well as [23, 41, 48]),

$$u_t + \mathcal{B}u = 0, \quad u|_{t=T} = g. \quad (3.1)$$

Here \mathcal{B} denotes the infinitesimal generator of X with domain $\mathcal{D}(\mathcal{B})$. For the Galerkin-based Finite Element implementation, equation (3.1) is converted into variational form as illustrated in Section 2. Formally, the resulting problem reads: find u such that

$$\left\langle \frac{\partial}{\partial t} u, v \right\rangle + \underbrace{\langle \mathcal{B}u, v \rangle}_{\mathcal{E}(u, v)} = 0, \quad \text{for all } v \in \mathcal{D}(\mathcal{B}). \quad (3.2)$$

In the classical setting of Black-Scholes, X is a geometric Brownian Motion and \mathcal{B} is a diffusion operator so that a closed form solution of (3.1) and (3.2) for plain vanilla contracts is possible in certain cases. For more general Lévy or Feller price processes X , \mathcal{B} is in general a pseudodifferential operator with symbol ψ^X , i.e.

$$(\mathcal{B}u)(x) = (\psi^X(x, D)u)(x) = - \int_{\mathbb{R}^n} e^{i\langle \xi, x \rangle} \psi^X(x, \xi) \widehat{u}(\xi) d\xi. \quad (3.3)$$

If X is a pure jump process and hence $\mathcal{B} = \mathcal{A}$ is an integral operator, one obtains the kernel representation (2.5) of $\mathcal{A} = \psi^X(x, D)$ by writing for any $u \in \mathcal{S}(\mathbb{R}^n)$,

$$\mathcal{A}u(x) = - \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} e^{i\langle x-y, \xi \rangle} \psi^X(x, \xi) u(y) dy d\xi.$$

Thus, the kernel $\kappa(\cdot, \cdot)$ in (2.5) can be represented as

$$\kappa(x, y) := \int_{\mathbb{R}^n} e^{i\langle x-y, \xi \rangle} \psi^X(x, \xi) d\xi, \quad (3.4)$$

the inverse Fourier transform (in the sense of oscillatory integrals, see e.g. [35, Equation (18.1.7)]) of ψ^X at $x - y$.

Remark 3.1. If X is a pure jump Lévy process with absolutely continuous Lévy measure then the following relation holds between $\kappa(\cdot, \cdot)$ and the density $k(\cdot)$ of the Lévy measure of X :

$$\kappa(x, y) = \int_{\mathbb{R}^n \setminus \{0\}} \int_{\mathbb{R}^n} \left(e^{i\langle x-y-z, \xi \rangle} - e^{i\langle x-y, \xi \rangle} + \frac{i\langle z, \xi \rangle}{1 + |z|^2} e^{i\langle x-y, \xi \rangle} \right) k(z) d\xi dz, \quad (3.5)$$

in the sense of distributions. By [33, Lemma 2.8], for any $\xi \in \mathbb{R}^n$, $z \in \mathbb{R}^n$ there holds

$$\left| e^{-i\langle z, \xi \rangle} - 1 + \frac{i\langle z, \xi \rangle}{1 + |z|^2} \right| \leq 7 \cdot \frac{|z|^2}{1 + |z|^2} \cdot (1 + |\xi|^2).$$

Hence, the distributional kernel $\kappa(\cdot, \cdot)$ in (3.5) is indeed well defined, since $k(\cdot)$ is a Lévy kernel that satisfies

$$\int_{\mathbb{R}^n} (|z|^2 \wedge 1) k(z) dz < \infty.$$

For an extensive description of Lévy processes we refer to the monographs [4, 49].

For the numerical solution of the variational problem (3.2) we employ variational Galerkin methods developed in [1, 2, 11, 23, 30, 39–41, 46]. We conclude this introductory part by illustrating this approach in case the underlying process X is a Lévy process (for a more general survey we refer to [31, 32]).

3.2. The Finite Element Method for option pricing in multidimensional Lévy models

The considerations of this section are based on [23, 48]. Suppose X is a Lévy process with state space \mathbb{R}^n and characteristic exponent

$$\psi^X(\xi) = -i\langle \gamma, \xi \rangle + \frac{1}{2} \langle \xi, \mathcal{Q}\xi \rangle + \int_{\mathbb{R}^n \setminus \{0\}} \left(1 - e^{i\langle \xi, y \rangle} + \frac{i\langle \xi, y \rangle}{1 + |y|^2} \right) \nu(dy),$$

where $\gamma \in \mathbb{R}^n$ is the drift vector, $\mathcal{Q} \in \mathbb{R}^{n \times n}$ is the covariance matrix and $\nu(dy)$ is the Lévy measure of X .

Assume the risk-neutral dynamics of $s = n > 1$ assets are given by

$$S_t^i = S_0^i e^{rt + X_t^i}, \quad i = 1, \dots, n,$$

under a risk-neutral measure such that $e^{X_t^i}$ is a martingale with respect to the canonical filtration $\mathcal{F}_t^0 := \sigma(X_s, s \leq t)$, $t \geq 0$, of the multivariate process X .

Consider an European option with maturity $T < \infty$ and payoff $g(S)$ which is assumed to be Lipschitz. The value $V(t, S_t)$ of this option is given by

$$V(t, S) = \mathbb{E} \left(e^{-r(T-t)} g(S_T) | S_t = S \right), \quad (3.6)$$

and, sufficient smoothness provided, it can be computed as the solution of a partial integrodifferential equation.

Theorem 3.2. Assume that $V(t, S)$ in (3.6) satisfies

$$V(t, S) \in C^{1,2}((0, T) \times \mathbb{R}_{>0}^n) \cap C^0([0, T] \times \mathbb{R}_{\geq 0}^n).$$

Then $V(t, S)$ is the solution of the following PIDE:

$$\begin{aligned} \frac{\partial V}{\partial t}(t, S) + \frac{1}{2} \sum_{i,j=1}^n S_i S_j \mathcal{Q}_{ij} \frac{\partial^2 V}{\partial S_i \partial S_j}(t, S) + r \sum_{i=1}^n S_i \frac{\partial V}{\partial S_i}(t, S) - rV(t, S) \\ + \int_{\mathbb{R}^n} \left(V(t, S e^z) - V(t, S) - \sum_{i=1}^n S_i (e^{z_i} - 1) \frac{\partial V}{\partial S_i}(t, S) \right) \nu(dz) = 0, \end{aligned} \quad (3.7)$$

in $(0, T) \times \mathbb{R}_{\geq 0}^n$ where $V(t, S e^z) := V(t, S_1 e^{z_1}, \dots, S_n e^{z_n})$, and the terminal condition is given by

$$V(T, S) = g(S) \quad \forall S \in \mathbb{R}_{\geq 0}^n. \quad (3.8)$$

Proof. [48, Theorem 4.2]. □

If the marginal Lévy measures ν_i , $i = 1, \dots, n$, of ν are absolutely continuous and admit densities $\nu_i(dz) = k_i(z) dz$ with constants $G_i > 0$, $M_i > 0$, $i = 1, \dots, n$, such that

$$k_i(z) \lesssim \begin{cases} e^{G_i z}, & \text{for all } z < -1, \\ e^{-M_i z}, & \text{for all } z > 1, \end{cases} \quad (3.9)$$

the PIDE (3.7) can be transformed into a simpler form.

Corollary 3.3. Suppose the marginal Lévy measures ν_i , $i = 1, \dots, n$, satisfy (3.9) with $M_i > 1$, $G_i > 0$, $i = 1, \dots, n$. Furthermore, let

$$u(\tau, x) = e^{r\tau} V\left(T - \tau, e^{x_1 + (\gamma_1 - r)\tau}, \dots, e^{x_n + (\gamma_n - r)\tau}\right),$$

where

$$\gamma_i = \frac{\mathcal{Q}_{ii}}{2} + \int_{\mathbb{R}} (e^{z_i} - 1 - z_i) \nu_i(dz_i).$$

Then, u satisfies the PIDE

$$\frac{\partial u}{\partial \tau} + \mathcal{A}_D[u] + \mathcal{A}[u] = 0, \quad (3.10)$$

in $(0, T) \times \mathbb{R}^n$ with initial condition $u(0, x) := u_0$. The differential operator is defined for $\varphi \in C_0^2(\mathbb{R}^n)$ by

$$\mathcal{A}_D[\varphi] = -\frac{1}{2} \sum_{i,j=1}^n \mathcal{Q}_{ij} \frac{\partial^2 \varphi}{\partial x_i \partial x_j}, \quad (3.11)$$

and the integrodifferential operator by

$$\mathcal{A}[\varphi] = - \int_{\mathbb{R}^n} \left(\varphi(x + z) - \varphi(x) - \sum_{i=1}^n z_i \frac{\partial \varphi}{\partial x_i}(x) \right) \nu(dz). \quad (3.12)$$

The initial condition is given by

$$u_0 = g(e^x) := g(e^{x_1}, \dots, e^{x_n}).$$

Proof. [48, Corollary 4.3] □

For any $u, v \in C_0^\infty(\mathbb{R}^n)$ we associate with the diffusion part \mathcal{A}_D the bilinear form

$$\mathcal{E}_D(u, v) = \frac{1}{2} \sum_{i,j=1}^n \mathcal{Q}_{ij} \int_{\mathbb{R}^n} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j} dx.$$

To the jump part \mathcal{A} we associate the so-called canonical bilinear jump form

$$\mathcal{E}_J(u, v) = - \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \left(u(x+z) - u(x) - \sum_{i=1}^n z_i \frac{\partial u}{\partial x_i}(x) \right) v(x) dx \nu(dz),$$

and set

$$\mathcal{E}(u, v) = \mathcal{E}_D(u, v) + \mathcal{E}_J(u, v). \quad (3.13)$$

Herewith, we can now formulate the formal parabolic problem (3.2) rigorously:

$$\begin{aligned} &\text{Find } u \in L^2((0, T); \mathcal{D}(\mathcal{E})) \cap H^1((0, T); \mathcal{D}(\mathcal{E})^*) \text{ such that} \\ &\left\langle \frac{\partial u}{\partial \tau}, v \right\rangle_{\mathcal{D}(\mathcal{E})^*, \mathcal{D}(\mathcal{E})} + \mathcal{E}(u, v) = 0, \quad \tau \in (0, T), \quad \forall v \in \mathcal{D}(\mathcal{E}), \\ &u(0) = u_0. \end{aligned} \quad (3.14)$$

Here $\mathcal{D}(\mathcal{E})$ denotes the domain of the bilinear form $\mathcal{E}(\cdot, \cdot)$ of X . The well-posedness of (3.14) has been analyzed in [23, 48] where it also has been shown that $\mathcal{D}(\mathcal{E})$ can explicitly be characterized in terms of anisotropic Sobolev spaces.

For implementation, the variational problem (3.14) needs to be localized to a bounded domain, discretized in space, and a time stepping scheme has to be applied. More precisely, these three steps are accomplished as follows:

1. Localization. For the localization we find that in Finance truncation of the original x -domain \mathbb{R}^n to $\Omega_R := [-R, R]^n$, $R > 0$, corresponds to approximating the solution u of (3.7) by the price u_R of a barrier option on Ω_R . In log-price u_R is given by

$$u_R(t, x) = \mathbb{E} \left(g(e^{X_T}) 1_{\{T < \tau_{\Omega_R, t}\}} \mid X_t = x \right),$$

where $\tau_{\Omega_R, t} = \inf\{s \geq t \mid X_s \notin \Omega_R\}$ denotes the first exit time of X_t from Ω_R after time t . In case of semiheavy tails (3.9), the solution of the localized problem u_R converges pointwise exponentially to the solution u of the original problem.

Theorem 3.4. *Suppose the payoff function $g : \mathbb{R}^n \rightarrow \mathbb{R}$ satisfies*

$$g(S) \lesssim \sum_{i=1}^n S_i + 1, \quad \forall S \in \mathbb{R}_{\geq 0}^n,$$

and the marginal measures ν_i satisfy (3.9) with $M_i > 1$, $G_i > 1$, $i = 1, \dots, n$. Then, there exist constants $\alpha, \beta > 0$ such that

$$|u(t, x) - u_R(t, x)| \lesssim e^{-\alpha R + \beta \|x\|_\infty}.$$

Proof. [48, Theorem 4.15]. □

For any function u with support in Ω_R we denote by \bar{u} its extension by zero to the whole of \mathbb{R}^n and define

$$\mathcal{E}_R(u, v) = \mathcal{E}(\bar{u}, \bar{v}),$$

with

$$\mathcal{D}(\mathcal{E}_R) = \overline{\{\bar{u} \mid u \in C_0^\infty(\Omega_R)\}},$$

where the closure is taken with respect to the natural norm $\|\cdot\|_{\mathcal{E}}$ of $\mathcal{D}(\mathcal{E})$, i.e. $\|u\|_{\mathcal{E}}^2 = \mathcal{E}(u, u) + \|u\|_{L^2}^2$. Thus, we can restate the variational form (3.14) on the bounded domain and the existence and uniqueness results for (3.14) remain valid.

$$\text{Find } u_R \in L^2((0, T); \mathcal{D}(\mathcal{E}_R)) \cap H^1((0, T); \mathcal{D}(\mathcal{E}_R)^*) \text{ such that}$$

$$\begin{aligned} \left\langle \frac{\partial u_R}{\partial \tau}, v \right\rangle + \mathcal{E}_R(u_R, v) &= 0, \quad \forall \tau \in (0, T), \forall v \in \mathcal{D}(\mathcal{E}_R), \\ u_R(0) &= u_0|_{\Omega_R}. \end{aligned} \quad (3.15)$$

2. Space discretization. For the discretization of $\mathcal{D}(\mathcal{E}_R)$ we choose the sparse tensor product spaces \widehat{V}_J , $J > 0$, defined in Section 2.2 – assuming that $\widehat{V}_J \subset \mathcal{D}(\mathcal{E}_R)$, which for instance is shown in [23] for Lévy copula processes. The corresponding semidiscrete problem reads:

$$\begin{aligned} &\text{First choose an approximation } u_{0,J} \in \widehat{V}_J \text{ for the initial data } u_0|_{\Omega_R}. \\ &\text{Then find } u_J \in H^1((0, T); \widehat{V}_J) \text{ such that} \\ &\left\langle \frac{\partial u_J}{\partial \tau}, v_J \right\rangle + \mathcal{E}_R(u_J, v_J) = 0, \quad \forall \tau \in (0, T), \forall v_J \in \widehat{V}_J, \\ &u_J(0) = u_{0,J}. \end{aligned} \quad (3.16)$$

The approximation of the initial data could be chosen as $u_{0,J} = \widehat{P}_J(u_0|_{\Omega_R})$ or as an interpolant of $u_0|_{\Omega_R}$. The semidiscrete problem (3.16) is an initial value problem for $\widehat{N}_J = \dim \widehat{V}_J$ ordinary differential equations

$$\mathbb{K}_J \frac{\partial}{\partial \tau} \underline{u} + \mathbb{A}_J \underline{u} = 0, \quad \underline{u}(0) = \underline{u}_0, \quad (3.17)$$

where $\underline{u}(t)$ denotes the coefficient vector of $u_J(t)$ with respect to the wavelet basis of \widehat{V}_J . Likewise \underline{u}_0 denotes the coefficient vector of $u_{0,J}$, and $\mathbb{K}_J, \mathbb{A}_J$ denote the mass- and stiffness matrix, respectively, with respect to the basis of \widehat{V}_J .

3. Time discretization using the θ -scheme. Let $0 \leq \theta \leq 1$. For $T < \infty$ and $I \in \mathbb{N}$, define the time step

$$k = \frac{T}{I},$$

and $t^i = ik$, $i = 0, \dots, I$. The fully discrete θ -scheme reads:

$$\begin{aligned} &\text{First find } u_J^0 \in \widehat{V}_J \text{ satisfying } u_J^0 = u_{0,J}. \\ &\text{Then for } i = 0, 1, \dots, I-1, \text{ find } u_J^{i+1} \in \widehat{V}_J \text{ such that} \\ &\left\langle \frac{u_J^{i+1} - u_J^i}{k}, v_J \right\rangle + \mathcal{E}_R(u_J^{i+\theta}, v_J) = 0, \quad \forall v_J \in \widehat{V}_J. \end{aligned} \quad (3.18)$$

Here $u_J^{i+\theta} := \theta u_J^{i+1} + (1-\theta)u_J^i$. In matrix form, the fully discrete problem (3.18) reads

$$(k^{-1}\mathbb{K}_J + \theta\mathbb{A}_J)\underline{u}^{i+1} = k^{-1}\mathbb{K}_J\underline{u}^i - (1-\theta)\mathbb{A}_J\underline{u}^i, \quad i = 0, 1, \dots, I-1. \quad (3.19)$$

Standard analysis of the θ -scheme (3.18) assumes that the bilinear form $\mathcal{E}_R(\cdot, \cdot)$ can be evaluated exactly, i.e. that the corresponding stiffness matrix \mathbb{A}_J is available. In practice this is unrealistic, since most often one only obtains approximations of \mathbb{A}_J . Since in this work we are interested in wavelet compression of \mathbb{A}_J resulting in a compressed matrix \mathbb{A}_J^{compr} , we conclude this section by illustrating how the impact of the resulting consistency error can be analyzed. To this end, we follow [58]: With the compressed matrix \mathbb{A}_J^{compr} we associate the perturbed bilinear form $\widetilde{\mathcal{E}}_R$. From (3.18) we herewith obtain the perturbed θ -scheme

$$\widetilde{u}_J^0 = u_{0,J}, \quad (3.20)$$

$$\left\langle \frac{\widetilde{u}_J^{i+1} - \widetilde{u}_J^i}{k}, v_J \right\rangle + \widetilde{\mathcal{E}}_R(\widetilde{u}_J^{i+\theta}, v_J) = 0, \quad (3.21)$$

for $i = 0, 1, \dots, I-1$ and every $v_J \in \widehat{V}_J$, where $\widetilde{u}_J^{i+\theta} := \theta \widetilde{u}_J^{i+1} + (1-\theta)\widetilde{u}_J^i$. In matrix form, (3.21) reads

$$(k^{-1}\mathbb{K}_J + \theta\mathbb{A}_J^{compr})\widetilde{\underline{u}}^{i+1} = k^{-1}\mathbb{K}_J\widetilde{\underline{u}}^i - (1-\theta)\mathbb{A}_J^{compr}\widetilde{\underline{u}}^i,$$

for $i = 0, 1, \dots, I-1$. In order to define *consistency conditions* for the perturbed θ -scheme, as in (2.7) we denote by $\underline{\alpha} \in (0, 2]^n$ the order of the integrodifferential operator $\mathcal{B} = \mathcal{A}_D + \mathcal{A}$ corresponding to $\mathcal{E}(\cdot, \cdot)$, i.e.

$$\tilde{\alpha} = \begin{cases} \underline{\alpha}, & \text{if } X \text{ is a pure jump process, i.e. } \mathcal{A}_D \equiv 0, \\ (2, \dots, 2), & \text{otherwise,} \end{cases}$$

where $\underline{\alpha} \in (0, 2)^n$ denotes the order of \mathcal{A} . Then the perturbed form $\tilde{\mathcal{E}}_R$ needs to satisfy the following consistency conditions, cf. [58, Section 3.3]:

- (1) There is $\delta < 1$ independent of $J \geq 0$ such that

$$\left| \mathcal{E}_R(u_J, v_J) - \tilde{\mathcal{E}}_R(u_J, v_J) \right| \leq \delta \|u_J\|_{H^{\tilde{\alpha}/2}} \|v_J\|_{H^{\tilde{\alpha}/2}}, \quad \forall u_J, v_J \in \hat{V}_J. \quad (3.22)$$

- (2) There exists a constant $C > 0$ independent of $J \geq 0$ such that

$$\left| \mathcal{E}_R(\hat{P}_J u, v_J) - \tilde{\mathcal{E}}_R(\hat{P}_J u, v_J) \right| \leq C 2^{J(|\underline{\alpha}|_\infty/2-d)} J^\nu \|u\|_{\mathcal{H}^{(d, \dots, d)}} \|v_J\|_{H^{\tilde{\alpha}/2}}, \quad (3.23)$$

for all $u \in \mathcal{H}^{(d, \dots, d)}(\Omega_R)$, $v_J \in \hat{V}_J$ with some $\nu \geq 0$. Here, as in Section 2, d denotes the approximation order (2.11) of \hat{V}_J .

If $\tilde{\mathcal{E}}_R(\cdot, \cdot)$ satisfies (3.22), the stability of the perturbed θ -scheme (3.20), (3.21) can be obtained from [58, Proposition 4.3]. Note that, by [58, Remark 4.4] for $\theta < 1/2$ there exists a positive constant C_* independent of the FE meshwidth $h = 2^{-J}$ and θ such that the time-step restriction

$$k \leq C_* \frac{h^{|\underline{\alpha}|_\infty}}{1-2\theta}, \quad (3.24)$$

is sufficient for stability. For $\theta \geq 1/2$ the scheme is unconditionally stable.

Assuming (3.22) and (3.23), the convergence of the perturbed θ -scheme is determined by [58, Theorem 5.4]. Since we give an explicit version of this result for anisotropic operators in Section 6.2, Theorem 6.6, we do not repeat the statement here.

Finally, note that in order to obtain the semidiscrete problem (3.16), (3.17), we need to solve a variational problem of the form (2.8) with $\mathcal{B} = \mathcal{A}_D + \mathcal{A}$ with non-local operator \mathcal{A} as abstractly described in Section 2.

Based on the semidiscrete formulations (3.16), (3.17), in the following sections we will only consider elliptic integrodifferential equations in space with the understanding that the developed methods can also be applied in the context of parabolic problems such as (3.14).

4. ANISOTROPIC OPERATORS AND SYMBOLS

We return to the generic set-up of Section 2. As before, we are interested in the efficient discretization of the integrodifferential equation (2.1) of the form

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

with an integrodifferential operator

$$\mathcal{B} = \mathcal{A}_D + \mathcal{A}, \quad (4.1)$$

given by (2.2). The sparse tensor product stiffness matrix of the differential operator \mathcal{A}_D in wavelet coordinates is of essentially optimal complexity $\mathcal{O}(2^J J^{(n-1)})$, since \mathcal{A}_D is local. Hence, the goal of this section is to reduce the complexity of the stiffness matrix of the integral operator \mathcal{A} . More precisely, we provide a generic compression scheme for continuous operators

$$\mathcal{A} : H^{\underline{\alpha} + \underline{s}}(\mathbb{R}^n) \rightarrow H^{\underline{s}}(\mathbb{R}^n), \quad \underline{s} \in \mathbb{R}^n,$$

for any multiindex $\underline{\alpha} \in \mathbb{R}_{\geq 0}^n$. To specify such operators we introduce a suitable class of anisotropic symbols.

Even though the compression scheme constructed in this section is also applicable to isotropic operators of Hörmander-type as in Example 2.1, it is not an extension of the compression scheme defined [47] which was constructed especially for such operators. Since the anisotropic operators that we consider here are allowed a more complex singularity structure than isotropic operators, the matrix entry estimates that we derive in Section 5 below differ significantly from those of [47]. For isotropic operators the specialized compression scheme of [47] is more efficient, i.e. optimal complexity can be proved under weaker assumptions on the number of vanishing moments (see also [46, Chapter 2]). For anisotropic operators however only the compression scheme that we define in the following is applicable.

Recall that for any symbol $p : \mathbb{R}^n \rightarrow \mathbb{R}$, the corresponding operator $p(D)$ is defined by

$$p(D)u(x) = - \int_{\mathbb{R}^n} e^{i\langle x, \xi \rangle} p(\xi) \widehat{u}(\xi) d\xi, \quad u \in \mathcal{S}(\mathbb{R}^n). \quad (4.2)$$

Furthermore, denote the axes in \mathbb{R}^n by $\Lambda := \{x \in \mathbb{R}^n : x_i = 0 \text{ for some } i \in \{1, \dots, n\}\}$. Herewith we can define a suitable class of anisotropic symbols and corresponding operators.

Definition 4.1. A function $p : \mathbb{R}^n \rightarrow \mathbb{R}$ is called a symbol in class $\Gamma^\alpha(\mathbb{R}^n)$, $\alpha \in \mathbb{R}^n$, if $p \in C^\infty(\mathbb{R}^n \setminus \Lambda) \cap C(\mathbb{R}^n)$ such that for any $\underline{\tau} \in \mathbb{N}_0^n$ there holds

$$\left| \partial_{\xi}^{\underline{\tau}} p(\xi) \right| \lesssim \prod_{i \in \mathcal{I}_{\underline{\tau}}} |\xi_i|^{\alpha_i - \tau_i} \cdot \prod_{k \notin \mathcal{I}_{\underline{\tau}}} (1 + |\xi_k|^2)^{\frac{\alpha_k}{2}}, \quad \text{for all } \xi \in \mathbb{R}^n, \quad (4.3)$$

where we set $\mathcal{I}_{\underline{\tau}} := \{i : \tau_i > 0\}$. The multiindex α is called the (anisotropic) order of the symbol p and the operator $\mathcal{A} = p(D)$.

Note that similar classes have already been presented in the context of Lévy processes in [44, 45]. Some possible realizations of operators \mathcal{A} with symbols $p \in \Gamma^\alpha(\mathbb{R}^n)$ are:

Example 4.2. If for any $\underline{\tau} \in \mathbb{N}_0^n$ the function $p \in C^\infty(\mathbb{R}^n \setminus \Lambda) \cap C(\mathbb{R}^n)$ satisfies

$$\left| \partial_{\xi}^{\underline{\tau}} p(\xi) \right| \lesssim \prod_{i=1}^n (1 + |\xi_i|^2)^{\frac{\alpha_i - \tau_i}{2}}, \quad \text{for all } \xi \in \mathbb{R}^n,$$

then $p \in \Gamma^\alpha(\mathbb{R}^n)$ and $\mathcal{A} = p(D)$ is admissible in this setting.

Example 4.3. By [44, Theorem 4.7], the infinitesimal generator \mathcal{A} of a pure jump Lévy copula process X with tempered stable margins admits a symbol $\psi^X \in \Gamma^\alpha(\mathbb{R}^n)$ and is hence admissible. In this case α_i , $i = 1, \dots, n$, are determined by the jump intensities of the marginal Lévy measures of X .

Example 4.4. Consider any isotropic symbol $p \in C^\infty(\mathbb{R}^n)$ of Hörmander-type with non-negative order, i.e. there exists some $\alpha \in \mathbb{R}_{\geq 0}$ such that for all $\underline{\tau} \in \mathbb{N}_0^n$ there holds

$$\left| \partial_{\xi}^{\underline{\tau}} p(\xi) \right| \lesssim (1 + |\xi|^2)^{\frac{\alpha - |\underline{\tau}|}{2}}, \quad \text{for all } \xi \in \mathbb{R}^n. \quad (4.4)$$

Then $p \in \Gamma^\alpha(\mathbb{R}^n)$ with $\alpha_1 = \dots = \alpha_n = \alpha$. To see this, one may use that for $\underline{\tau} \in \mathbb{N}_0^n$ there holds

$$\prod_{i=1}^n (1 + |\xi_i|^2)^{\frac{\tau_i}{2}} \leq \prod_{i=1}^n \left(1 + \sum_{j=1}^n |\xi_j|^2 \right)^{\frac{\tau_i}{2}} = (1 + |\xi|^2)^{\frac{|\underline{\tau}|}{2}},$$

and thus

$$(1 + |\xi|^2)^{-\frac{|\underline{\tau}|}{2}} \leq \prod_{i=1}^n (1 + |\xi_i|^2)^{-\frac{\tau_i}{2}}. \quad (4.5)$$

Furthermore,

$$(1 + |\xi|^2)^{\frac{\alpha}{2}} \leq \left(\prod_{i=1}^n (1 + |\xi_i|^2) \right)^{\frac{\alpha}{2}} = \prod_{i=1}^n (1 + |\xi_i|^2)^{\frac{\alpha}{2}}, \quad (4.6)$$

since $\alpha \geq 0$. Clearly, (4.5) and (4.6) imply that (4.3) holds for any symbol $p \in C^\infty(\mathbb{R}^n)$ that satisfies (4.4). Note that this statement does not remain true if $\alpha < 0$ in (4.4).

5. FUNDAMENTAL ANISOTROPIC MATRIX ESTIMATES

Throughout this section, we consider an arbitrary but fixed pair of two n -dimensional tensor product wavelets $\psi_{\mathbf{j}, \mathbf{k}} = \psi_{j_1, k_1} \otimes \dots \otimes \psi_{j_n, k_n}$, $\psi_{\mathbf{j}', \mathbf{k}'} = \psi_{j'_1, k'_1} \otimes \dots \otimes \psi_{j'_n, k'_n}$. For any coordinate direction $s = 1, \dots, n$, we denote

$$\delta_{x_s} := \text{dist}(\text{supp}\{\psi_{j_s, k_s}\}, \text{supp}\{\psi_{j'_s, k'_s}\}),$$

and

$$\sigma_{x_s} := \begin{cases} \text{dist}(\text{singsupp}\{\psi_{j_s, k_s}\}, \text{supp}\{\psi_{j'_s, k'_s}\}), & \text{if } j_s \leq j'_s, \\ \text{dist}(\text{supp}\{\psi_{j_s, k_s}\}, \text{singsupp}\{\psi_{j'_s, k'_s}\}), & \text{if } j'_s \leq j_s. \end{cases}$$

In the following section we collect a number of auxiliary estimates which might be interesting in their own right, but more importantly enable us to prove Theorem 5.5 in Section 5.2.

5.1. Auxiliary results

Throughout this section, fix a coordinate direction $i \in \{1, \dots, n\}$. Given $\xi \in \mathbb{R}^n$, we write

$$\tilde{\xi}(i) = (\xi_1, \dots, \xi_{i-1}, \xi_{i+1}, \dots, \xi_n).$$

For fixed $\tilde{\xi}(i)$, we define the partial symbol $p_i^{\tilde{\xi}(i)} : \mathbb{R} \rightarrow \mathbb{R}$ by

$$p_i^{\tilde{\xi}(i)}(\cdot) := p(\xi_1, \dots, \xi_{i-1}, \cdot, \xi_{i+1}, \dots, \xi_n). \quad (5.1)$$

The corresponding operator is given by

$$p_i^{\tilde{\xi}(i)}(D)u(x) := p(\xi_1, \dots, \xi_{i-1}, D, \xi_{i+1}, \dots, \xi_n)u(x) = \int_{\mathbb{R}} e^{ix\xi_i} p(\xi) \hat{u}(\xi_i) d\xi_i. \quad (5.2)$$

To prove fundamental matrix entry estimates in Section 5.2 below, we need to collect some auxiliary results on symbols $p : \mathbb{R}^n \rightarrow \mathbb{R}$ that satisfy for fixed $\tilde{\xi}(i)$,

$$|p(\xi_1, \dots, \xi_n)| \lesssim c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(i)) \cdot (1 + |\xi_i|^2)^{\frac{m}{2}}, \quad \text{for all } \xi_i \in \mathbb{R}, \quad (5.3)$$

with suitable $m \in \mathbb{R}$ and where $c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(i))$ denotes some given function that may depend on \mathbf{j} , \mathbf{j}' and $\tilde{\xi}(i)$ but is independent of $\xi_i \in \mathbb{R}$.

Remark 5.1. The $(\mathbf{j}, \mathbf{j}')$ -dependence of $c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(i))$ implies that the symbol $p : \mathbb{R}^n \rightarrow \mathbb{R}$ itself may depend on the level indices \mathbf{j} , \mathbf{j}' of the wavelets that were fixed in the beginning. Inequality (5.3) implies that this dependence can be characterized by the given function $c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(i))$. For example, in the proof of Theorem 5.5 below, we will encounter $c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(i))$ of the form

$$c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(i)) = \prod_{\substack{1 \leq t \leq n \\ t \neq i}} |\xi_t|^{\alpha_t - \tau_t} \cdot 2^{-\alpha_i \min\{j_i, j'_i\}},$$

which is independent of $\xi_i \in \mathbb{R}$.

Lemma 5.2. *If $p : \mathbb{R}^n \rightarrow \mathbb{R}$ satisfies (5.3) then for $u \in H^{s+m}(\mathbb{R})$ there holds*

$$\|p_i^{\tilde{\xi}(i)}(D)u\|_{H^s(\mathbb{R})} \lesssim c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(i)) \cdot \|u\|_{H^{s+m}(\mathbb{R})}. \quad (5.4)$$

Proof. At first assume $s = 0$ and $m = 0$ in (5.3). Then, $p_i^{\tilde{\xi}(i)}(\cdot) \in L^\infty(\mathbb{R})$. Using Plancherel's theorem, one thus obtains for $u \in L^2(\mathbb{R})$,

$$\|p_i^{\tilde{\xi}(i)}(D)u\|_{L^2(\mathbb{R})} = \|(p_i^{\tilde{\xi}(i)} \cdot \hat{u})^\vee\|_{L^2(\mathbb{R})} = \|(p_i^{\tilde{\xi}(i)} \cdot \hat{u})\|_{L^2(\mathbb{R})} \leq \|p_i^{\tilde{\xi}(i)}\|_{L^\infty(\mathbb{R})} \cdot \|\hat{u}\|_{L^2(\mathbb{R})}.$$

By (5.3) there holds $\|p_i^{\tilde{\xi}(i)}\|_{L^\infty(\mathbb{R})} \lesssim c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(i))$ and thus we have

$$\|p_i^{\tilde{\xi}(i)}(D)u\|_{L^2(\mathbb{R})} \lesssim c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(i)) \cdot \|u\|_{L^2(\mathbb{R})}. \quad (5.5)$$

Now we lift this result to arbitrary $m, s \in \mathbb{R}$ using the classical lifting procedure, cf. e.g. [55, Section II.6] or [36, Theorem 2.5.4]. Let $s = 0$ and $m \in \mathbb{R}$ arbitrary and write

$$p_i^{\tilde{\xi}(i)}(D)u(x) = \int_{\mathbb{R}} e^{ix\xi_i} p(\xi) (1 + |\xi_i|^2)^{-\frac{m}{2}} (1 + |\xi_i|^2)^{\frac{m}{2}} \hat{u}(\xi_i) d\xi_i.$$

With this representation and the above arguments one obtains

$$\|p_i^{\tilde{\xi}(i)}(D)u\|_{L^2(\mathbb{R})} \leq \|p_i^{\tilde{\xi}(i)} \cdot (1 + |\cdot|^2)^{-\frac{m}{2}}\|_{L^\infty(\mathbb{R})} \cdot \|(1 + |D|^2)^{\frac{m}{2}} u\|_{L^2(\mathbb{R})} \lesssim c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(i)) \cdot \|u\|_{H^m(\mathbb{R})}.$$

For arbitrary $s \in \mathbb{R}$, note that by e.g. [36, Corollary 2.4.19] the symbol of $(1 + |D|^2)^{\frac{s}{2}} \circ p_i^{\tilde{\xi}(i)}(D)$ satisfies (5.3) with $s + m$ instead of m . Thus,

$$\|p_i^{\tilde{\xi}(i)}(D)u\|_{H^s(\mathbb{R})} = \|(1 + |D|^2)^{\frac{s}{2}} \circ p_i^{\tilde{\xi}(i)}(D)u\|_{L^2(\mathbb{R})} \lesssim c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(i)) \cdot \|u\|_{H^{s+m}(\mathbb{R})}.$$

□

Lemma 5.3. *Suppose, in addition to (5.3), the symbol $p : \mathbb{R}^n \rightarrow \mathbb{R}$ satisfies for any $\sigma \in \mathbb{N}_0$,*

$$|\partial_{\xi_i}^\sigma p(\xi_1, \dots, \xi_n)| \lesssim c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(i)) \cdot |\xi_i|^{m-\sigma}, \quad \text{for } \xi_i \in \mathbb{R} \setminus \{0\}, \quad (5.6)$$

with the same function $c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(i))$ as in (5.3). Then the associated distributional kernel

$$\kappa_i^{\tilde{\xi}(i)}(x_i, x'_i) = \int e^{i(x_i - x'_i)\xi_i} p_i^{\tilde{\xi}(i)}(\xi_i) d\xi_i,$$

satisfies for any $\tau, \tau' \in \mathbb{N}_0$,

$$\left| \partial_{x_i}^\tau \partial_{x'_i}^{\tau'} \kappa_i^{\tilde{\xi}(i)}(x_i, x'_i) \right| \lesssim c_{\tau, \tau'} \cdot c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(i)) \cdot |x_i - x'_i|^{-(1+m+\tau+\tau')}, \quad (5.7)$$

where $c_{\tau, \tau'} > 0$ is some constant depending on τ, τ' but not on \mathbf{j}, \mathbf{j}' or $\tilde{\xi}(i)$.

Proof. The result follows directly from the proof of the classical Calderón-Zygmund estimates, cf. [50, Lemma 3.0.2] or [46, Lemma 3.4.1], since the constant factor $c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(i))$ can simply be taken out of the integral defining $\kappa_i^{\tilde{\xi}(i)}(\cdot, \cdot)$. □

5.2. Matrix entry estimates

In this section we prove estimates for the matrix entries $[\mathbb{A}_J]_{(\mathbf{j}, \mathbf{k})(\mathbf{j}', \mathbf{k}')} = \langle \mathcal{A}\psi_{\mathbf{j}, \mathbf{k}}, \psi_{\mathbf{j}', \mathbf{k}'} \rangle$. From now on suppose $\mathcal{A} = p(D)$ denotes an operator with symbol $p \in \Gamma^\alpha(\mathbb{R}^n)$. To simplify notation, for $s = 1, \dots, n$, and $m \in \mathbb{R}$ we introduce

$$\omega(s, m) := \begin{cases} 2^{-\frac{1}{2}(j_s + j'_s)} 2^{-\tilde{d}(j_s + j'_s)} \delta_{x_s}^{-(1+m+2\tilde{d})}, & \text{if } \delta_{x_s} > 2^{-\min\{j_s, j'_s\}}, \\ 2^{m - \min\{j_s, j'_s\}}, & \text{if } \delta_{x_s} \leq 2^{-\min\{j_s, j'_s\}}, \end{cases} \quad (5.8)$$

and

$$\tilde{\omega}(s, m) := \begin{cases} 2^{-\frac{1}{2}(j_s + j'_s)} 2^{-\tilde{d}(j_s + j'_s)} \delta_{x_s}^{-(1+m+2\tilde{d})}, & \text{if } \delta_{x_s} > 2^{-\min\{j_s, j'_s\}}, \\ 2^{m - \min\{j_s, j'_s\}}, & \text{if } \sigma_{x_s} \leq 2^{-\max\{j_s, j'_s\}}, \\ 2^{\frac{\min\{j_s, j'_s\} - \max\{j_s, j'_s\}}{2}} 2^{-\tilde{d} \max\{j_s, j'_s\}} \sigma_{x_s}^{-(m+\tilde{d})}, & \text{otherwise.} \end{cases} \quad (5.9)$$

Note that $\tilde{\omega}(s, m)$ is indeed well-defined since $\delta_{x_s} \leq \sigma_{x_s}$. Clearly there holds $\tilde{\omega}(s, m) \leq \omega(s, m)$ for all $s = 1, \dots, n$. The definition of $\tilde{\omega}(s, m)$ is motivated by the following profound result.

Proposition 5.4. *Let $m \in \mathbb{R}$ and suppose for some $i \in \{1, \dots, n\}$ there holds $\psi_{j_i, k_i}, \psi_{j'_i, k'_i} \in H^m(\mathbb{R})$. If a symbol $p : \mathbb{R}^n \rightarrow \mathbb{R}$ satisfies (5.3) and (5.6) then the one-dimensional operator*

$$p_i^{\tilde{\xi}^{(i)}}(D) : H^m(\mathbb{R}) \rightarrow L^2(\mathbb{R}),$$

defined in (5.2) satisfies

$$\left| \left\langle p_i^{\tilde{\xi}^{(i)}}(D) \psi_{j_i, k_i}, \psi_{j'_i, k'_i} \right\rangle \right|, \left| \left\langle p_i^{\tilde{\xi}^{(i)}}(D) \psi_{j'_i, k'_i}, \psi_{j_i, k_i} \right\rangle \right| \lesssim c(\mathbf{j}, \mathbf{j}', \tilde{\xi}^{(i)}) \cdot \tilde{\omega}(i, m).$$

Proof. Case 1. If $\delta_{x_i} > 2^{-\min\{j_i, j'_i\}}$ the result follows from the classical compression estimate [12, Theorem 6.1] using the Calderón-Zygmund estimate (5.7) of $\kappa_i^{\tilde{\xi}^{(i)}}(\cdot, \cdot)$, i.e. Lemma 5.3.

Case 2. If $\delta_{x_i} \leq 2^{-\min\{j_i, j'_i\}}$ and $\sigma_{x_i} > 2^{-\max\{j_i, j'_i\}}$ the continuity result, Lemma 5.2, implies that [12, Theorem 6.3] can be applied. This yields the required estimate.

Case 3. If $\sigma_{x_i} \leq 2^{-\max\{j_i, j'_i\}}$ we assume without loss of generality $j_i \leq j'_i$. With Schwarz' inequality and Lemma 5.2 one obtains

$$\begin{aligned} \left| \left\langle p_i^{\tilde{\xi}^{(i)}}(D) \psi_{j_i, k_i}, \psi_{j'_i, k'_i} \right\rangle \right| &\leq \|p_i^{\tilde{\xi}^{(i)}}(D) \psi_{j_i, k_i}\|_{L^2(\mathbb{R})} \cdot \|\psi_{j'_i, k'_i}\|_{L^2(\mathbb{R})} \\ &\lesssim c(\mathbf{j}, \mathbf{j}', \tilde{\xi}^{(i)}) \cdot \|\psi_{j_i, k_i}\|_{H^m(\mathbb{R})} \lesssim c(\mathbf{j}, \mathbf{j}', \tilde{\xi}^{(i)}) \cdot 2^{j_i m}, \end{aligned}$$

where the last inequality follows from the wavelet norm equivalence (2.17). Since the same arguments apply to the adjoint integral operator $p_i^{\tilde{\xi}^{(i)}}(D)^*$, the result follows. \square

From (2.22), recall the notion of the wavelets' regularity parameter

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

Using Proposition 5.4, one obtains the main matrix entry estimate.

Theorem 5.5. *Suppose $\bar{\alpha} \leq \gamma$ and let $\mathcal{A} = p(D)$ be an operator with symbol $p \in \Gamma^\alpha(\mathbb{R}^n)$. There holds*

$$\left| \langle \mathcal{A} \psi_{\mathbf{j}, \mathbf{k}}, \psi_{\mathbf{j}', \mathbf{k}'} \rangle \right|, \left| \langle \mathcal{A} \psi_{\mathbf{j}', \mathbf{k}'}, \psi_{\mathbf{j}, \mathbf{k}} \rangle \right| \lesssim \prod_{i=1}^n \tilde{\omega}(i, \alpha_i). \quad (5.10)$$

Proof. The proof is based on an iterative argument reducing the dimensionality of the symbol $p : \mathbb{R}^n \rightarrow \mathbb{R}$. For this, we introduce some notation: to describe that p depends on the variables ξ_1, \dots, ξ_n we denote $p_{1..n} := p$. Correspondingly, for any coordinate direction i the partial symbol (5.1) is denoted by $p_{1..n, i}^{\tilde{\xi}^{(i)}} : \mathbb{R} \rightarrow \mathbb{R}$. By Fubini's theorem there holds

$$\begin{aligned} \langle \mathcal{A} \psi_{\mathbf{j}, \mathbf{k}}, \psi_{\mathbf{j}', \mathbf{k}'} \rangle &= \int_{\mathbb{R}^n} p_{1..n}(\xi_1, \dots, \xi_n) \prod_{s=1}^n \overline{\widehat{\psi_{j_s, k_s}(\xi_s)}} \widehat{\psi_{j'_s, k'_s}(\xi_s)} d\xi_1 \dots d\xi_n \\ &= \int_{\mathbb{R}^{n-1}} \left\langle p_{1..n, 1}^{\tilde{\xi}^{(1)}}(D) \psi_{j_1, k_1}, \psi_{j'_1, k'_1} \right\rangle \cdot \prod_{s=2}^n \overline{\widehat{\psi_{j_s, k_s}(\xi_s)}} \widehat{\psi_{j'_s, k'_s}(\xi_s)} d\xi_2 \dots d\xi_n, \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ denotes the $L^2(\mathbb{R})$ scalar product and $\tilde{\xi}^{(1)} = (\xi_2, \dots, \xi_n) \in \mathbb{R}^{n-1}$. Defining

$$p_{2..n}(\xi_2, \dots, \xi_n) := \left\langle p_{1..n, 1}^{\tilde{\xi}^{(1)}}(D) \psi_{j_1, k_1}, \psi_{j'_1, k'_1} \right\rangle, \quad (5.11)$$

one obtains

$$\langle \mathcal{A} \psi_{\mathbf{j}, \mathbf{k}}, \psi_{\mathbf{j}', \mathbf{k}'} \rangle = \int_{\mathbb{R}^{n-1}} p_{2..n}(\xi_2, \dots, \xi_n) \prod_{s=2}^n \overline{\widehat{\psi_{j_s, k_s}(\xi_s)}} \widehat{\psi_{j'_s, k'_s}(\xi_s)} d\xi_2 \dots d\xi_n,$$

and we have reduced the dimension of the integral by one. Furthermore, for any $\tau_s \in \mathbb{N}_0$, $s = 2, \dots, n$, consider the symbol $\partial_{\xi_2}^{\tau_2} \dots \partial_{\xi_n}^{\tau_n} p_{1..n} = \partial_{\xi_2}^{\tau_2} \dots \partial_{\xi_n}^{\tau_n} p$. Using (4.3) with $\underline{\tau} = (0, \tau_2, \dots, \tau_n)$ one obtains

$$\left| \left[\partial_{\xi_2}^{\tau_2} \dots \partial_{\xi_n}^{\tau_n} p_{1..n} \right]_1^{\tilde{\xi}(1)}(\xi_1, \dots, \xi_n) \right| \lesssim c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(1)) \cdot (1 + |\xi_1|^2)^{\frac{\alpha_1}{2}}, \quad (5.12)$$

where

$$c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(1)) = c_1(\mathbf{j}, \mathbf{j}', \tilde{\xi}(1)) := \prod_{\substack{2 \leq t \leq n \\ t \in \underline{\mathcal{I}}_{\underline{\tau}}}} |\xi_t|^{\alpha_t - \tau_t} \cdot \prod_{\substack{2 \leq k \leq n \\ k \notin \underline{\mathcal{I}}_{\underline{\tau}}}} (1 + |\xi_k|^2)^{\frac{\alpha_k}{2}}. \quad (5.13)$$

Hence, instead of the function p in (5.3), the symbol $\partial_{\xi_2}^{\tau_2} \dots \partial_{\xi_n}^{\tau_n} p_{1..n}$ satisfies (5.3) with $i = 1$, $m = \alpha_1$ and $c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(1))$ as in (5.13). Analogously, using (4.3) with $\underline{\tau} = (\sigma, \tau_2, \dots, \tau_n)$ one infers that for any $\sigma \in \mathbb{N}_0$ there holds

$$\left| \partial_{\xi_1}^{\sigma} \left[\partial_{\xi_2}^{\tau_2} \dots \partial_{\xi_n}^{\tau_n} p_{1..n} \right]_1^{\tilde{\xi}(1)}(\xi_1, \dots, \xi_n) \right| \lesssim c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(1)) \cdot |\xi_1|^{\alpha_1 - \sigma}, \quad (5.14)$$

again with $c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(1))$ as in (5.13). Thus, for any choice of $\tau_s \in \mathbb{N}_0$, $s = 2, \dots, n$, the symbol $\partial_{\xi_2}^{\tau_2} \dots \partial_{\xi_n}^{\tau_n} p_{1..n}$ satisfies (5.3) and (5.6) for $i = 1$ with $m = \alpha_1$ and $c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(1))$ as in (5.13). Hence, with the representation (5.11) of $p_{2..n} : \mathbb{R}^{n-1} \rightarrow \mathbb{R}$, Proposition 5.4 implies

$$\begin{aligned} \left| \partial_{\xi_2}^{\tau_2} \dots \partial_{\xi_n}^{\tau_n} p_{2..n}(\xi_2, \dots, \xi_n) \right| &= \left| \left\langle \partial_{\xi_2}^{\tau_2} \dots \partial_{\xi_n}^{\tau_n} p_{1..n,1}^{\tilde{\xi}(1)}(D) \psi_{j_1, k_1}, \psi_{j'_1, k'_1} \right\rangle \right| \\ &\lesssim \prod_{\substack{2 \leq t \leq n \\ t \in \underline{\mathcal{I}}_{\underline{\tau}}}} |\xi_t|^{\alpha_t - \tau_t} \cdot \prod_{\substack{2 \leq k \leq n \\ k \notin \underline{\mathcal{I}}_{\underline{\tau}}}} (1 + |\xi_k|^2)^{\frac{\alpha_k}{2}} \cdot \tilde{\omega}(1, \alpha_1). \end{aligned}$$

As in (5.12) and (5.14), one herewith obtains that for all $\tau_t \in \mathbb{N}_0$, $t = 3, \dots, n$, the symbol $\partial_{\xi_3}^{\tau_3} \dots \partial_{\xi_n}^{\tau_n} p_{2..n} : \mathbb{R}^{n-1} \rightarrow \mathbb{R}$ satisfies (5.3) and (5.6) in its first coordinate direction $i = 1$ with $m = \alpha_2$ and

$$c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(1)) = c_2(\mathbf{j}, \mathbf{j}', \tilde{\xi}(1)) := \prod_{\substack{3 \leq t \leq n \\ t \in \underline{\mathcal{I}}_{\underline{\tau}}}} |\xi_t|^{\alpha_t - \tau_t} \cdot \prod_{\substack{3 \leq k \leq n \\ k \notin \underline{\mathcal{I}}_{\underline{\tau}}}} (1 + |\xi_k|^2)^{\frac{\alpha_k}{2}} \cdot \tilde{\omega}(1, \alpha_1),$$

where now $\tilde{\xi}(1) = (\xi_3, \dots, \xi_n) \in \mathbb{R}^{n-2}$, since $p_{2..n} : \mathbb{R}^{n-1} \rightarrow \mathbb{R}$. Proposition 5.4 again implies

$$\begin{aligned} \left| \partial_{\xi_3}^{\tau_3} \dots \partial_{\xi_n}^{\tau_n} p_{3..n}(\xi_2, \dots, \xi_n) \right| &:= \left| \left\langle \partial_{\xi_3}^{\tau_3} \dots \partial_{\xi_n}^{\tau_n} p_{2..n,1}^{\tilde{\xi}(1)}(D) \psi_{j_2, k_2}, \psi_{j'_2, k'_2} \right\rangle \right| \\ &\lesssim \prod_{\substack{3 \leq t \leq n \\ t \in \underline{\mathcal{I}}_{\underline{\tau}}}} |\xi_t|^{\alpha_t - \tau_t} \cdot \prod_{\substack{3 \leq k \leq n \\ k \notin \underline{\mathcal{I}}_{\underline{\tau}}}} (1 + |\xi_k|^2)^{\frac{\alpha_k}{2}} \cdot \tilde{\omega}(2, \alpha_2) \cdot \tilde{\omega}(1, \alpha_1). \end{aligned}$$

Consequently, $\partial_{\xi_4}^{\tau_4} \dots \partial_{\xi_n}^{\tau_n} p_{3..n}(\xi_2, \dots, \xi_n)$ is again a symbol satisfying (5.3) and (5.6) with $i = 1$, $m = \alpha_3$ and

$$c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(1)) = c_3(\mathbf{j}, \mathbf{j}', \tilde{\xi}(1)) := \prod_{\substack{4 \leq t \leq n \\ t \in \underline{\mathcal{I}}_{\underline{\tau}}}} |\xi_t|^{\alpha_t - \tau_t} \cdot \prod_{\substack{4 \leq k \leq n \\ k \notin \underline{\mathcal{I}}_{\underline{\tau}}}} (1 + |\xi_k|^2)^{\frac{\alpha_k}{2}} \cdot \tilde{\omega}(2, \alpha_2) \cdot \tilde{\omega}(1, \alpha_1),$$

where now $\tilde{\xi}(1) = (\xi_4, \dots, \xi_n)$. Iterating this procedure yields

$$\begin{aligned} |\langle \mathcal{A} \psi_{\mathbf{j}, \mathbf{k}}, \psi_{\mathbf{j}', \mathbf{k}'} \rangle| &= \left| \int_{\mathbb{R}^{n-1}} p_{2..n}(\xi_2, \dots, \xi_n) \prod_{s=2}^n \overline{\psi_{j_s, k_s}(\xi_s)} \widehat{\psi_{j'_s, k'_s}(\xi_s)} d\xi_2 \dots d\xi_n \right| \\ &= \left| \int_{\mathbb{R}^{n-2}} \left\langle p_{2..n,1}^{\tilde{\xi}(1)}(D) \psi_{j_2, k_2}, \psi_{j'_2, k'_2} \right\rangle \prod_{s=3}^n \overline{\psi_{j_s, k_s}(\xi_s)} \widehat{\psi_{j'_s, k'_s}(\xi_s)} d\xi_3 \dots d\xi_n \right| \end{aligned}$$

$$\begin{aligned}
&= \left| \int_{\mathbb{R}^{n-2}} p_{3..n}(\xi_3, \dots, \xi_n) \prod_{s=3}^n \overline{\widehat{\psi_{j_s, k_s}(\xi_s)} \widehat{\psi_{j'_s, k'_s}(\xi_s)}} d\xi_3 \dots d\xi_n \right| \\
&= \dots \\
&= \left| \int_{\mathbb{R}} \left\langle p_{n-1..n,1}^{\tilde{\xi}(1)}(D) \psi_{j_{n-1}, k_{n-1}}, \psi_{j'_{n-1}, k'_{n-1}} \right\rangle \overline{\widehat{\psi_{j_n, k_n}(\xi_n)} \widehat{\psi_{j'_n, k'_n}(\xi_n)}} d\xi_n \right| \\
&= \left| \int_{\mathbb{R}} p_{n..n}(\xi_n) \overline{\widehat{\psi_{j_n, k_n}(\xi_n)} \widehat{\psi_{j'_n, k'_n}(\xi_n)}} d\xi_n \right|,
\end{aligned}$$

with $p_{n..n} : \mathbb{R} \rightarrow \mathbb{R}$ satisfying (5.3) and (5.6) in its first coordinate direction $i = 1$ with $m = \alpha_n$ and

$$c(\mathbf{j}, \mathbf{j}', \tilde{\xi}(1)) = c_n(\mathbf{j}, \mathbf{j}', \tilde{\xi}(1)) := \tilde{\omega}(n-1, \alpha_{n-1}) \cdot \dots \cdot \tilde{\omega}(1, \alpha_1),$$

where now $\tilde{\xi}(1) = \emptyset$ is the empty vector, since $p_{n..n}$ is univariate. Finally, since

$$\int_{\mathbb{R}} p_{n..n}(\xi_n) \overline{\widehat{\psi_{j_n, k_n}(\xi_n)} \widehat{\psi_{j'_n, k'_n}(\xi_n)}} d\xi_n = \left\langle p_{n..n,1}^{\tilde{\xi}(1)}(D) \psi_{j_n, k_n}, \psi_{j'_n, k'_n} \right\rangle,$$

Proposition 5.4 implies

$$|\langle \mathcal{A} \psi_{\mathbf{j}, \mathbf{k}}, \psi_{\mathbf{j}', \mathbf{k}'} \rangle| \lesssim \tilde{\omega}(n, \alpha_n) \cdot \dots \cdot \tilde{\omega}(1, \alpha_1).$$

□

6. ANISOTROPIC CONSISTENCY FRAMEWORK

In this section we define the consistency requirements for the compression of anisotropic operators $\mathcal{A} = p(D)$, $p \in \Gamma^\alpha(\mathbb{R})$. We extend the results of [47, Section 4] that have been obtained for the compression of isotropic operators.

6.1. Fundamental consistency estimates

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

$$X_{\theta, \underline{\alpha}/2, \underline{\rho}} := (H^{\underline{\alpha}/2}(\square), \mathcal{H}^{\underline{\rho}}(\square))_{\theta, 2}, \quad 0 \leq \theta \leq 1, \quad (6.1)$$

using the K -method of interpolation (cf. e.g. [56, Section 1.3]). Here $\underline{\rho}$ denotes the smoothness multiindex defined by (2.34) depending on $\underline{\alpha}$. For the wavelet norm estimated for $H^{\underline{\alpha}/2}$ and $\mathcal{H}^{\underline{\rho}}$ one infers the following estimate for the norm of $X_{\theta, \underline{\alpha}/2, \underline{\rho}}$:

Proposition 6.1. *Let $0 \leq \theta \leq 1$ and $\mathbf{j} \in \mathbb{N}_0^n$ be a fixed level index. For any $u \in X_{\theta, \underline{\alpha}/2, \underline{\rho}}$ 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*

$$\sum_{\mathbf{k} \in \nabla_{\mathbf{j}}} 2^{(1-\theta)|\underline{\alpha} \cdot \mathbf{j}|_\infty} 2^{2\theta \sum_{i=1}^n \rho_i j_i} |u_{\mathbf{j}, \mathbf{k}}|^2 \lesssim \|u\|_{X_{\theta, \underline{\alpha}/2, \underline{\rho}}}^2. \quad (6.2)$$

Proof. Recall that the norm equivalences and estimates (2.17)–(2.19) of the univariate wavelets imply that

$$\sum_{\substack{\mathbf{j}' \in \mathbb{N}_0^n \\ \mathbf{k}' \in \nabla_{\mathbf{j}'}}} (2^{\alpha_1 j_1} + \dots + 2^{\alpha_n j_n}) |v_{\mathbf{j}', \mathbf{k}'}|^2 \sim \|v\|_{H^{\underline{\alpha}/2}(\square)}^2, \quad (6.3)$$

$$\sum_{\mathbf{k} \in \nabla_{\mathbf{j}}} 2^{2 \sum_{i=1}^n \rho_i j_i} |v_{\mathbf{j}, \mathbf{k}}|^2 \lesssim \|v\|_{\mathcal{H}^{\underline{\rho}}(\square)}^2, \quad (6.4)$$

for any sufficiently smooth function v with wavelet coefficients $v_{\mathbf{j},\mathbf{k}}$ (see e.g. [46, Theorem 3.2.2]).

Herewith we may now proceed as in the proof of [47, Proposition 4.1]: by [56, Theorem 1.3.3 (c)], there exists a positive constant $c > 0$ only depending on θ such that for all $t \in \mathbb{R}_{>0}$ there holds

$$t^{-\theta} K(t, u) \leq c \|u\|_{X_{\theta, \underline{\alpha}/2, \underline{\rho}}}, \quad (6.5)$$

where $K(t, u) = \inf_{g \in \mathcal{H}^L} \{ \|u - g\|_{H^{\underline{\alpha}/2}(\square)} + t \|g\|_{\mathcal{H}^L} \}$ denotes the K -functional. Furthermore, using the wavelet norm estimates (6.3)–(6.4), one obtains that

$$T(\mathbf{j}, t, u) := \inf_{\substack{g \in \mathcal{H}^{(d,d)} \\ g = \sum_{\mathbf{k} \in \nabla_{\mathbf{j}}} g_{\mathbf{j},\mathbf{k}} \psi_{\mathbf{j},\mathbf{k}}}} \left\{ \left(\sum_{\mathbf{k} \in \nabla_{\mathbf{j}}} \left(1 + \sum_{i=1}^n 2^{\alpha_i j_i} \right) |u_{\mathbf{j},\mathbf{k}} - g_{\mathbf{j},\mathbf{k}}|^2 \right)^{\frac{1}{2}} + t \left(\sum_{\mathbf{k} \in \nabla_{\mathbf{j}}} 2^{2 \sum_{i=1}^n \rho_i j_i} |g_{\mathbf{j},\mathbf{k}}|^2 \right)^{\frac{1}{2}} \right\},$$

satisfies $T(\mathbf{j}, t, u) \lesssim K(t, u)$ uniformly for all $t \in \mathbb{R}_{>0}$ and all $u \in X_{\theta, \underline{\alpha}/2, \underline{\rho}}$. Thus, by (6.5), it suffices to show that there exists some $t \in \mathbb{R}_{>0}$ such that

$$\sum_{\mathbf{k} \in \nabla_{\mathbf{j}}} 2^{(1-\theta)|\underline{\alpha} \cdot \mathbf{j}|_{\infty}} 2^{2\theta \sum_{i=1}^n \rho_i j_i} |u_{\mathbf{j},\mathbf{k}}|^2 \lesssim t^{-2\theta} T(\mathbf{j}, t, u)^2, \quad (6.6)$$

Choosing

$$t = \frac{\left(1 + \sum_{i=1}^n 2^{\alpha_i j_i} \right)^{\frac{1}{2}}}{2^{\sum_{i=1}^n \rho_i j_i}},$$

the validity of (6.6) can be verified easily, exactly as in the proof of [47, Proposition 4.1]. \square

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 (2.7), the order of \mathcal{B} is given by

$$\tilde{\underline{\alpha}} = \begin{cases} (2, \dots, 2), & \text{if } \mathcal{Q} > 0 \text{ and } \bar{\alpha} \leq 2, \\ \underline{\alpha}, & \text{otherwise,} \end{cases} \quad (6.7)$$

where $\mathcal{Q} \in \mathbb{R}^{n \times n}$ denotes the coefficient matrix of \mathcal{A}_D .

For any $\tilde{\underline{\alpha}} \in \mathbb{R}_{\geq 0}^n$, Proposition 2.5 implies that the convergence rate of the sparse tensor product approximation without compression is determined entirely by the maximum $|\tilde{\underline{\alpha}}|_{\infty} = \max\{\tilde{\alpha}_1, \dots, \tilde{\alpha}_n\}$. Thus, in order to be consistent with the sparse tensor product Galerkin discretization, the compressed scheme has to satisfy

Requirement 6.1. *The operator $\mathcal{A}_J^{\text{compr}}$ corresponding to a compressed matrix $\mathbb{A}_J^{\text{compr}}$ must fulfill*

$$\left| \left\langle (\mathcal{A} - \mathcal{A}_J^{\text{compr}}) \hat{P}_J u, \hat{P}_J v \right\rangle \right| \lesssim \varepsilon \cdot 2^{\frac{1}{n} J \sum_{i=1}^n [\tilde{\alpha}_i - (1-\theta_1) \frac{\tilde{\alpha}_i}{2} - \theta_1 \rho_i - (1-\theta_2) \frac{\tilde{\alpha}_i}{2} - \theta_2 \rho_i]} \|u\|_{X_{\theta_1, \tilde{\underline{\alpha}}/2, \underline{\rho}}} \|v\|_{X_{\theta_2, \tilde{\underline{\alpha}}/2, \underline{\rho}}}, \quad (6.8)$$

for any $0 \leq \theta_1, \theta_2 \leq 1$ and some suitable constant $\varepsilon > 0$ uniformly with respect to $J \geq 0$. Here $\underline{\rho} \in \mathbb{R}_{\geq 0}^n$ is as defined in (2.34) depending on $\tilde{\underline{\alpha}}$ via

$$\rho_i = d - \frac{|\tilde{\underline{\alpha}}|_{\infty} - \tilde{\alpha}_i}{2}, \quad i = 1, \dots, n.$$

Remark 6.2. In the isotropic setting $\alpha_1 = \dots = \alpha_n$, Requirement 6.1 coincides with [47, Requirement 4.1], because all the summands on the right hand side of (6.8) are equal.

To see that Requirement 6.1 ensures the stability and convergence results of the sparse tensor product scheme without compression (see e.g. [23]), one may proceed as follows: setting $\theta_1 = \theta_2 = 0$ in (6.8) implies

$$\left| \left\langle (\mathcal{A} - \mathcal{A}_J^{\text{compr}}) \hat{P}_J u, \hat{P}_J v \right\rangle \right| \lesssim \varepsilon \cdot \|u\|_{H^{\tilde{\underline{\alpha}}/2}} \|v\|_{H^{\tilde{\underline{\alpha}}/2}}, \quad \text{for all } u, v \in H^{\tilde{\underline{\alpha}}/2}. \quad (6.9)$$

By definition (2.34) of $\underline{\rho}$, setting $\theta_1 = 1$, $\theta_2 = 0$ implies

$$\left| \left\langle (\mathcal{A} - \mathcal{A}_J^{\text{compr}}) \widehat{P}_J u, \widehat{P}_J v \right\rangle \right| \lesssim \varepsilon \cdot 2^{-J(d-\frac{\alpha}{2})} \|u\|_{\mathcal{H}^\rho} \|v\|_{H^{\underline{\alpha}/2}}, \quad (6.10)$$

for all $u \in \mathcal{H}^\rho$, $v \in H^{\underline{\alpha}/2}$. Herewith, one obtains

Theorem 6.3. *Suppose the solution u of (2.1) satisfies $u \in \mathcal{H}^\rho$ and $\mathcal{A}_J^{\text{compr}}$ satisfies Requirement 6.1. 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*

$$\langle \mathcal{B}_J^{\text{compr}} u_J, u_J \rangle \geq c'_1 \|u_J\|_{H^{\underline{\alpha}/2}}^2 - c_2 \|u_J\|_{L^2}^2, \quad \text{for all } u_J \in \widehat{V}_J, \quad (6.11)$$

and there exists some $c'_3 > 0$ such that for all $J \geq J_0$,

$$|\langle \mathcal{B}_J^{\text{compr}} u_J, v_J \rangle| \leq c'_3 \|u_J\|_{H^{\underline{\alpha}/2}} \|v_J\|_{H^{\underline{\alpha}/2}}, \quad \text{for all } u_J, v_J \in \widehat{V}_J. \quad (6.12)$$

Furthermore, the convergence rate (2.33) of the Galerkin scheme without compression is preserved (cf. Proposition 2.5).

Proof. Note that since $\mathcal{B} = \mathcal{A}_D + \mathcal{A}$ there holds $\mathcal{B} - \mathcal{B}_J^{\text{compr}} = \mathcal{A} - \mathcal{A}_J^{\text{compr}}$ for all $J > 0$. Thus, inequality (6.11) may be verified by inserting (6.9) into (2.31). This yields,

$$\langle \mathcal{B}_J^{\text{compr}} u_J, u_J \rangle \geq c_1 \|u_J\|_{H^{\underline{\alpha}/2}}^2 - c_2 \|u_J\|_{L^2}^2 - 2\varepsilon \|u_J\|_{H^{\underline{\alpha}/2}}^2 = (c_1 - 2\varepsilon) \|u_J\|_{H^{\underline{\alpha}/2}}^2 - c_2 \|u_J\|_{L^2}^2,$$

and $c'_1 := c_1 - 2\varepsilon > 0$ for sufficiently small $\varepsilon > 0$ from Requirement 6.1. The constants c_1, c_2 are obtained from (2.31). For the continuity inequality (6.12) one obtains from (2.32) and Requirement 6.1 with $\theta_1 = \theta_2 = 0$,

$$|\langle \mathcal{B}_J^{\text{compr}} u_J, v_J \rangle| \leq |\langle \mathcal{B} u_J, v_J \rangle| + |\langle (\mathcal{B}_J^{\text{compr}} - \mathcal{B}) u_J, v_J \rangle| \leq c_3 \|u_J\|_{H^{\underline{\alpha}/2}} \|v_J\|_{H^{\underline{\alpha}/2}} + \varepsilon \|u_J\|_{H^{\underline{\alpha}/2}} \|v_J\|_{H^{\underline{\alpha}/2}},$$

with c_3 from (2.32). Setting $c'_3 = c_3 + \varepsilon$ one obtains (6.12).

Finally, noting that $\underline{\alpha}/2 < \gamma$ (with γ as in (2.22)), the convergence result follows from (6.10) in conjunction with Strang's first lemma (see e.g. [22, Lemma 2.27]). \square

The following theorem provides (lower) bounds for the cut-off parameters and hence will enable us to define the compression scheme.

Theorem 6.4. *The compressed matrix $\mathbb{A}_J^{\text{compr}}$ fulfills Requirement 6.1 if its block matrices $\mathbb{A}_{\mathbf{j}, \mathbf{j}'}$ satisfy,*

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

with

$$\sigma_{\mathbf{j}, \mathbf{j}'} := \begin{cases} \sum_{i=1}^n \left[\frac{2}{n} J \left(\rho_i - \frac{\tilde{\alpha}_i}{2} \right) - \rho_i (j_i + j'_i) \right], & \text{if } \begin{cases} L(\mathbf{j}', \underline{\rho}) \geq R(\mathbf{j}', \tilde{\underline{\alpha}}), \\ L(\mathbf{j}, \underline{\rho}) \geq R(\mathbf{j}, \tilde{\underline{\alpha}}), \end{cases} \\ -\frac{1}{2} (|\tilde{\underline{\alpha}} \cdot \mathbf{j}'|_\infty + |\tilde{\underline{\alpha}} \cdot \mathbf{j}|_\infty), & \text{if } \begin{cases} L(\mathbf{j}', \underline{\rho}) < R(\mathbf{j}', \tilde{\underline{\alpha}}), \\ L(\mathbf{j}, \underline{\rho}) < R(\mathbf{j}, \tilde{\underline{\alpha}}), \end{cases} \\ \sum_{i=1}^n \left[\frac{1}{n} J \left(\rho_i - \frac{\tilde{\alpha}_i}{2} \right) - \rho_i j_i \right] - \frac{1}{2} |\tilde{\underline{\alpha}} \cdot \mathbf{j}'|_\infty, & \text{if } \begin{cases} L(\mathbf{j}', \underline{\rho}) < R(\mathbf{j}', \tilde{\underline{\alpha}}), \\ L(\mathbf{j}, \underline{\rho}) \geq R(\mathbf{j}, \tilde{\underline{\alpha}}), \end{cases} \\ \sum_{i=1}^n \left[\frac{1}{n} J \left(\rho_i - \frac{\tilde{\alpha}_i}{2} \right) - \rho_i j'_i \right] - \frac{1}{2} |\tilde{\underline{\alpha}} \cdot \mathbf{j}|_\infty, & \text{if } \begin{cases} L(\mathbf{j}', \underline{\rho}) \geq R(\mathbf{j}', \tilde{\underline{\alpha}}), \\ L(\mathbf{j}, \underline{\rho}) < R(\mathbf{j}, \tilde{\underline{\alpha}}), \end{cases} \end{cases} \quad (6.13)$$

where

$$L(\mathbf{j}, \underline{\rho}) := \sum_{i=1}^n \rho_i \left(\frac{1}{n} J - j_i \right),$$

$$R(\mathbf{j}, \underline{\alpha}) := \frac{1}{n} \sum_{i=1}^n J \frac{\tilde{\alpha}_i}{2} - \frac{1}{2} |\underline{\alpha} \cdot \mathbf{j}|_\infty.$$

Proof. The proof follows the lines of [47, Theorem 4.2]. Let $0 \leq \theta_1, \theta_2 \leq 1$ be as in Requirement 6.1. Using Proposition 6.1 one obtains,

$$\begin{aligned} \left| \left\langle (\mathcal{A} - \mathcal{A}_J^{compr}) \widehat{P}_J u, \widehat{P}_J v \right\rangle \right| &\lesssim \sum_{\substack{0 \leq |\mathbf{j}|_1 \leq J \\ 0 \leq |\mathbf{j}'|_1 \leq J}} \left| \sum_{\substack{\mathbf{k} \in \nabla_{\mathbf{j}} \\ \mathbf{k}' \in \nabla_{\mathbf{j}'}}} u_{\mathbf{j}, \mathbf{k}} v_{\mathbf{j}', \mathbf{k}'} \left\langle (\mathcal{A} - \mathcal{A}_J^{compr}) \psi_{\mathbf{j}, \mathbf{k}}, \psi_{\mathbf{j}', \mathbf{k}'} \right\rangle \right| \\ &\lesssim \sum_{\substack{0 \leq |\mathbf{j}|_1 \leq J \\ 0 \leq |\mathbf{j}'|_1 \leq J}} \left(\left\| [u_{\mathbf{j}, \mathbf{k}}]_{\mathbf{k} \in \nabla_{\mathbf{j}}} \right\|_2 \left\| [v_{\mathbf{j}', \mathbf{k}'}]_{\mathbf{k}' \in \nabla_{\mathbf{j}'}} \right\|_2 \left\| \mathbb{A}_{\mathbf{j}, \mathbf{j}'} - \mathbb{A}_{\mathbf{j}, \mathbf{j}'}^{compr} \right\|_2 \right) \\ &\lesssim \|u\|_{X_{\theta_1, \underline{\alpha}/2, \underline{\rho}}} \|v\|_{X_{\theta_2, \underline{\alpha}/2, \underline{\rho}}} \cdot \sum_{\substack{0 \leq j_1 + j_2 \leq J \\ 0 \leq j'_1 + j'_2 \leq J}} \left\| 2^{-\frac{1}{2}(1-\theta_1)|\underline{\alpha} \cdot \mathbf{j}|_\infty} 2^{-\theta_1 \sum_{i=1}^n \rho_i j_i} \right. \\ &\quad \left. \times 2^{-\frac{1}{2}(1-\theta_2)|\underline{\alpha} \cdot \mathbf{j}'|_\infty} 2^{-\theta_2 \sum_{i=1}^n \rho_i j'_i} \left(\mathbb{A}_{\mathbf{j}, \mathbf{j}'} - \mathbb{A}_{\mathbf{j}, \mathbf{j}'}^{compr} \right) \right\|_2. \end{aligned}$$

Thus, Requirement 6.1 is satisfied if

$$\left\| 2^{-\frac{1}{2}(1-\theta_1)|\underline{\alpha} \cdot \mathbf{j}|_\infty - \theta_1 \sum_{i=1}^n \rho_i j_i - \frac{1}{2}(1-\theta_2)|\underline{\alpha} \cdot \mathbf{j}'|_\infty - \theta_2 \sum_{i=1}^n \rho_i j'_i} \left(\mathbb{A}_{\mathbf{j}, \mathbf{j}'} - \mathbb{A}_{\mathbf{j}, \mathbf{j}'}^{compr} \right) \right\|_2 \leq \varepsilon.$$

One therefore needs to impose

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

with some σ such that for all $0 \leq \theta_1, \theta_2 \leq 1$ there holds,

$$\begin{aligned} \sigma &\geq -\frac{1}{n} J \sum_{i=1}^n \left[\alpha_i - (1-\theta_1) \frac{\alpha_i}{2} - \theta_1 \rho_i - (1-\theta_2) \frac{\alpha_i}{2} - \theta_2 \rho_i \right] \\ &\quad - \frac{1}{2} (1-\theta_1) |\underline{\alpha} \cdot \mathbf{j}|_\infty - \theta_1 \sum_{i=1}^n \rho_i j_i - \frac{1}{2} (1-\theta_2) |\underline{\alpha} \cdot \mathbf{j}'|_\infty - \theta_2 \sum_{i=1}^n \rho_i j'_i. \end{aligned} \tag{6.14}$$

Differentiation of the right hand side of (6.14) with respect to θ_1 and θ_2 , resp., shows its monotonicity with respect to these parameters. Herewith one obtains that for any $\mathbf{j}, \mathbf{j}' \in \mathbb{N}_0^n$ the parameter $\sigma_{\mathbf{j}, \mathbf{j}'}$ defined by (6.13) satisfies (6.14). \square

As already indicated in [47], in case one is only interested in sparse tensor product based wavelet compression for the fast evaluation of integral expressions as in e.g. [38], then Requirement 6.1 can be relaxed to

Requirement 6.2. *The operator \mathcal{A}_J^{compr} corresponding to a compressed matrix \mathbb{A}_J^{compr} has to satisfy*

$$\left| \left\langle (\mathcal{A} - \mathcal{A}_J^{compr}) \widehat{P}_J u, \widehat{P}_J v \right\rangle \right| \lesssim \varepsilon 2^{J\bar{\alpha}} 2^{-J(\overline{t+t'})} \|u\|_{\mathcal{H}^{\underline{t}}} \|v\|_{\mathcal{H}^{\underline{t}'}}, \tag{6.15}$$

for any multiindices $\underline{t}, \underline{t}' \in [\frac{\bar{\alpha}}{2}, d]^n$ and some suitable constant $\varepsilon > 0$ uniformly with respect to J . Here $\overline{t+t'} = \max\{t_i + t'_i : i = 1, \dots, n\}$.

In this case, Theorem 6.4 can be simplified to

Corollary 6.5. *The compressed matrix \mathbb{A}_J^{compr} fulfills Requirement 6.2 if its 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

$$\sigma'_{j,j'} := 2J(d' - \frac{\bar{\alpha}}{2}) - d'(|j|_1 + |j'|_1), \quad (6.16)$$

for given $d \leq d' < \tilde{d} + \bar{\alpha}$.

6.2. Consistency of the perturbed θ -scheme

As outlined in Section 3.2, in Financial Mathematics one often encounters parabolic equations of the form (3.14) arising from a given Lévy process X . Using the sparse tensor product spaces \widehat{V}_J to discretize in space and the θ -scheme described in Section 3.2 to discretize in time, one obtains a fully discrete problem of the form (3.18). Applying a compression scheme to the stiffness matrix \mathbb{A}_J of the infinitesimal generator \mathcal{A} of X then yields a perturbed θ -scheme as described in (3.20)–(3.21). From (6.9) and (6.10) one directly infers the validity of (3.22) and (3.23).

Thus, suppose the compression scheme satisfies Requirement 6.1. Then [58, Theorem 5.4] yields the following convergence result of the perturbed θ -scheme (3.20)–(3.21) for anisotropic operators. For sake of brevity, we employ the notation of Section 3.2:

Theorem 6.6. *Suppose the operator \mathcal{A}_J^{compr} corresponding to the compressed matrix \mathbb{A}_J^{compr} satisfies Requirement 6.1. For $\theta \in [0, \frac{1}{2})$ assume that the time step k satisfies (3.24). Assume further that the approximation $u_{0,J} \in \widehat{V}_J$ of the initial data u_0 is quasi-optimal in $L^2(\square)$.*

Then the following error estimates hold for the perturbed θ -scheme (3.20)–(3.21) with $\theta \in [0, 1]$:

$$\begin{aligned} \|u^I - \tilde{u}_J^I\|_{L^2}^2 + k \sum_{i=0}^{I-1} \|u^{i+\theta} - \tilde{u}_J^{i+\theta}\|_{H^{\alpha/2}}^2 &\lesssim 2^{-2J(d-\frac{\bar{\alpha}}{2})} \max_{0 \leq t \leq T} \|u(t)\|_{\mathcal{H}^{\alpha}}^2 \\ &+ \begin{cases} k^2 \int_0^T \|\ddot{u}(s)\|_*^2 ds, & \text{for all } \theta \in [0, 1] \\ k^4 \int_0^T \|\ddot{u}(s)\|_*^2 ds, & \text{for } \theta = \frac{1}{2} \end{cases}, \\ &+ 2^{-2J(d-\frac{\bar{\alpha}}{2})} \int_0^T \|\dot{u}(s)\|_{\mathcal{H}^{\alpha}}^2 ds, \end{aligned}$$

where $\|u\|_* := \sup_{v_J \in \widehat{V}_J} \frac{\langle u, v_J \rangle}{\|v_J\|_{H^{\alpha/2}}}$.

7. ANISOTROPIC COMPRESSION SCHEME

Based on the estimate (5.10) of Theorem 5.5, in this section we define two compression schemes and show that the resulting compressed matrices \mathbb{A}_J^{compr} satisfy Requirement 6.1 and Requirement 6.2, respectively. The schemes are split into two parts based on the distinction of *first* and *second* compression as defined in [12, 50]:

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.

Due to Theorem 5.5, from now on we suppose $\bar{\alpha} \leq \gamma$.

To simplify notation, for any fixed pair of tensor product wavelets $\psi_{\mathbf{j}, \mathbf{k}} = \psi_{j_1, k_1} \otimes \dots \otimes \psi_{j_n, k_n}$, $\psi_{\mathbf{j}', \mathbf{k}'} = \psi_{j'_1, k'_1} \otimes \dots \otimes \psi_{j'_n, k'_n}$ and each $i = 1, \dots, n$, we denote

$$m_i := \max\{j_i, j'_i\} - \min\{j_i, j'_i\} = j_i + j'_i - 2 \min\{j_i, j'_i\} \geq 0,$$

and

$$s_i := \sum_{s \neq i} \alpha_s \min\{j_s, j'_s\}. \quad (7.1)$$

Furthermore, denote by $\tilde{\sigma}_{\mathbf{j}, \mathbf{j}'}$ 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 (6.13) and (6.16).

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

$$C_{\mathbf{j}, \mathbf{j}'}^{i, \mathcal{I}} := c_i \max \left\{ 2^{-\min\{j_i, j'_i\}}, 2^{\frac{\tilde{\sigma}_{\mathbf{j}, \mathbf{j}'} - \tilde{d}(j_i + j'_i) + s_i + \frac{1}{2} \sum_{t \in \mathcal{I}} m_t - \tilde{d} \sum_{l \notin \mathcal{I} \cup \{i\}} m_l}{2\tilde{d} + \alpha_i}} \right\},$$

with $c_i > 0$. In addition, 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

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

Herewith, for $\bar{\alpha}/2 \leq d \leq \tilde{d}$, the first compression scheme is defined by

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

Analogously, for each $i = 1, \dots, n$ and any index set $\mathcal{I} \subset \{1, \dots, n\}$, the cut-off parameters of the second compression are defined by

$$E_{\mathbf{j}, \mathbf{j}'}^{i, \mathcal{I}} := e_i \max \left\{ 2^{-\max\{j_i, j'_i\}}, 2^{\frac{\tilde{\sigma}_{\mathbf{j}, \mathbf{j}'} - \tilde{d} \max\{j_i, j'_i\} + s_i + \frac{1}{2} \sum_{t \in \mathcal{I} \setminus \{i\}} m_t - \tilde{d} \sum_{l \notin \mathcal{I}} m_l}{\tilde{d} + \alpha_i}} \right\},$$

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

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

Finally, the fully compressed matrix \mathbb{A}_J^{compr} is defined by

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

There holds

Theorem 7.1. *If $\tilde{\sigma}_{\mathbf{j}, \mathbf{j}'} = \sigma_{\mathbf{j}, \mathbf{j}'}$ as defined in (6.13) then the compressed matrix \mathbb{A}_J^{compr} satisfies Requirement 6.1 and is thus consistent with the sparse tensor product discretization of Section 2. Furthermore, in (6.8) there holds*

$$\varepsilon = \max_{i=1, \dots, n} \left\{ c_i^{-(2\tilde{d} + \alpha_i)}, e_i^{-(\tilde{d} + \alpha_i)} \right\}.$$

If $\tilde{\sigma}_{\mathbf{j}, \mathbf{j}'} = \sigma'_{\mathbf{j}, \mathbf{j}'}$ as defined in (6.16) then the compressed matrix \mathbb{A}_J^{compr} satisfies Requirement 6.2.

Proof. For sake of brevity, we only prove the result in case $\tilde{\sigma}_{\mathbf{j}, \mathbf{j}'} = \sigma_{\mathbf{j}, \mathbf{j}'}$. For $\tilde{\sigma}_{\mathbf{j}, \mathbf{j}'} = \sigma'_{\mathbf{j}, \mathbf{j}'}$ the result follows analogously by replacing Theorem 6.4 with Corollary 6.5 in the analysis below.

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

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

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

satisfies the requirements of Theorem 6.4. To simplify notation, we introduce the index set

$$\mathcal{D}_{\mathbf{j}, \mathbf{j}'} := \{ \mathbf{k} \in \nabla_{\mathbf{j}} : \delta_{x_i} > C_{\mathbf{j}, \mathbf{j}'}^{i, \mathcal{I}}, \delta_{x_l} > 2^{-\min(j_l, j'_l)} \forall l \notin \mathcal{I} \}.$$

By Theorem 5.5, each matrix entry can be estimated by $\prod_{i=1}^n \tilde{\omega}(i, \alpha_i)$ with $\tilde{\omega}(i, \alpha_i)$ defined by (5.9) corresponding to the wavelets of the matrix entry. Since $\tilde{\omega}(i, \alpha_i) \leq \omega(i, \alpha_i)$, $i = 1, \dots, n$, with $\omega(i, \alpha_i)$ defined in (5.8), one may thus estimate the column sums of each $\mathbb{T}_{\mathbf{j}, \mathbf{j}'}$ by

$$\begin{aligned} \sum_{\mathbf{k} \in \nabla_{\mathbf{j}}} |t_{(\mathbf{j}, \mathbf{k})(\mathbf{j}', \mathbf{k}')}^x| &\lesssim \sum_{\mathbf{k} \in \mathcal{D}_{\mathbf{j}, \mathbf{j}'}} 2^{-\frac{1}{2}(j_i + j'_i)} 2^{-\tilde{d}(j_i + j'_i)} \delta_{x_i}^{-(1 + \alpha_i + 2\tilde{d})} \prod_{s \in \mathcal{I}} 2^{\alpha_s j'_s} \prod_{l \notin \mathcal{I}} 2^{-(\frac{1}{2} + \tilde{d})(j_l + j'_l)} \delta_{x_l}^{-(1 + \alpha_l + 2\tilde{d})} \\ &\lesssim 2^{-\frac{1}{2}(j_i + j'_i)} 2^{-\tilde{d}(j_i + j'_i)} \int_{|x_i| > C_{\mathbf{j}, \mathbf{j}'}^{i, \mathcal{I}}} |x_i|^{-(1 + \alpha_i + 2\tilde{d})} dx_i \\ &\quad \times \prod_{l \notin \mathcal{I}} \left[2^{-(\frac{1}{2} + \tilde{d})(j_l + j'_l)} \int_{|x_l| > 2^{-j'_l}} \delta_{x_l}^{-(1 + \alpha_l + 2\tilde{d})} dx_l \right] \cdot \sum_{\substack{k_s: s \in \mathcal{I}, \\ \mathbf{k} \in \mathcal{D}_{\mathbf{j}, \mathbf{j}'}}} \left(\prod_{s \in \mathcal{I}} 2^{\alpha_s j'_s} \right), \end{aligned}$$

where we have used the fact that the sum over the indices k_i and k_l , $l \notin \mathcal{I}$, is taken only over those matrix entries that satisfy $\delta_{x_l} > \max\{2^{-j_l}, 2^{-j'_l}\}$ and can therefore be estimated by the product of the integrals. Since for each $s \in \mathcal{I}$ there are $\mathcal{O}(2^{j_s - j'_s})$ non-zero column entries, one finally obtains

$$\begin{aligned} \sum_{\mathbf{k} \in \nabla_{\mathbf{j}}} |t_{(\mathbf{j}, \mathbf{k})(\mathbf{j}', \mathbf{k}')}^x| &\lesssim 2^{\frac{1}{2}(|\mathbf{j}|_1 - |\mathbf{j}'|_1)} 2^{-\tilde{d}(j_i + j'_i)} (C_{\mathbf{j}, \mathbf{j}'}^{i, \mathcal{I}})^{-(\alpha_i + 2\tilde{d})} \prod_{l \notin \mathcal{I}} 2^{-\tilde{d}(j_l + j'_l)} 2^{(\alpha_l + 2\tilde{d})j'_l} \prod_{s \in \mathcal{I}} 2^{\alpha_s j'_s} \\ &\lesssim c_i^{-(\alpha_i + 2\tilde{d})} 2^{\frac{1}{2}(|\mathbf{j}|_1 - |\mathbf{j}'|_1)} 2^{-\sigma_{\mathbf{j}, \mathbf{j}'}}. \end{aligned}$$

In the same way one obtains for the row sums

$$\sum_{\mathbf{k}' \in \nabla_{\mathbf{j}'}} |t_{(\mathbf{j}, \mathbf{k})(\mathbf{j}', \mathbf{k}')}^x| \lesssim c_i^{-(\alpha_i + 2\tilde{d})} 2^{\frac{1}{2}(|\mathbf{j}'|_1 - |\mathbf{j}|_1)} 2^{-\sigma_{\mathbf{j}, \mathbf{j}'}}.$$

Hence, 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 that $\mathbb{T}^{i, \mathcal{I}}$ satisfies the requirements of Theorem 6.4 with $\varepsilon = c_i^{-(\alpha_i + 2\tilde{d})}$. The consistency of \mathbb{A}_f^{cpr-1} follows.

To analyze \mathbb{A}_f^{cpr-2} , for any $i \in \{1, \dots, n\}$ and $\mathcal{I} \subset \{1, \dots, n\}$, define the perturbation matrix $\mathbb{S}^{i, \mathcal{I}}$ by its blocks

$$\left[\mathbb{S}_{\mathbf{j}, \mathbf{j}'}^{i, \mathcal{I}} \right]_{(\mathbf{k}, \mathbf{k}')} := \begin{cases} [\mathbb{A}_J]_{(\mathbf{j}, \mathbf{k})(\mathbf{j}', \mathbf{k}')} , & \text{if } \begin{cases} \sigma_{x_i} > E_{\mathbf{j}, \mathbf{j}'}^{i, \mathcal{I}}, \\ \delta_{x_i} \leq 2^{-\min\{j_i, j'_i\}}, \\ \delta_{x_k} > 2^{-\min(j_k, j'_k)} \forall k \notin \mathcal{I}, \end{cases} \\ 0, & \text{otherwise.} \end{cases}$$

By Theorem 5.5, one finds for the entries of $\mathbb{S}^{i, \mathcal{I}}$:

$$|s_{(\mathbf{j}, \mathbf{k})(\mathbf{j}', \mathbf{k}')}^x| \lesssim 2^{\frac{1}{2}(j'_i - j_i)} 2^{\tilde{d}j_i} \sigma_x^{-(\alpha_i + \tilde{d})} \prod_{\substack{s \in \mathcal{I}, \\ s \neq i}} 2^{\alpha_s j'_s} \prod_{l \notin \mathcal{I}} 2^{-(\frac{1}{2} + \tilde{d})(j_l + j'_l)} \delta_{x_l}^{-(1 + \alpha_l + 2\tilde{d})}$$

$$\begin{aligned}
&\lesssim 2^{\frac{1}{2}(j'_i - j_i)} 2^{\tilde{d}j_i} (E_{\mathbf{j}, \mathbf{j}'}^{i, \mathcal{I}})^{-(\alpha_i + \tilde{d})} \prod_{\substack{s \in \mathcal{I}, \\ s \neq i}} 2^{\alpha_s j'_s} \prod_{l \notin \mathcal{I}} 2^{-(\frac{1}{2} + \tilde{d})(j_l + j'_l)} \delta_{x_l}^{-(1 + \alpha_l + 2\tilde{d})} \\
&\lesssim e_i^{-(\alpha_i + \tilde{d})} 2^{\frac{1}{2}(|\mathbf{j}|_1 - |\mathbf{j}'|_1)} 2^{-\sigma_{\mathbf{j}, \mathbf{j}'}} \prod_{l \notin \mathcal{I}} \delta_{x_l}^{-(1 + \alpha_l + 2\tilde{d})} \cdot 2^{-(\alpha_l + 2\tilde{d})j'_l} 2^{-j_l}.
\end{aligned}$$

Thus, using the same arguments as above, the weighted column sums of $\mathbb{S}^{i, \mathcal{I}}$ may be estimated by

$$\begin{aligned}
\sum_{\mathbf{k}' \in \nabla_{\mathbf{j}'}} 2^{\frac{1}{2}(|\mathbf{j}|_1 - |\mathbf{j}'|_1)} |s_{(\mathbf{j}, \mathbf{k})(\mathbf{j}', \mathbf{k}')}^x| &\lesssim e_i^{-(\alpha_i + \tilde{d})} 2^{-\sigma_{\mathbf{j}, \mathbf{j}'}} \prod_{l \notin \mathcal{I}} 2^{-(\alpha_l + 2\tilde{d})j'_l} \int_{|x_l| > 2^{-j'_l}} |x_l|^{-(1 + \alpha_l + 2\tilde{d})} dx_l \\
&\lesssim e_i^{-(\alpha_i + \tilde{d})} 2^{-\sigma_{\mathbf{j}, \mathbf{j}'}} ,
\end{aligned}$$

and analogously

$$\sum_{\mathbf{k} \in \nabla_{\mathbf{j}}} 2^{\frac{1}{2}(|\mathbf{j}'|_1 - |\mathbf{j}|_1)} |s_{(\mathbf{j}, \mathbf{k})(\mathbf{j}', \mathbf{k}')}^x| \lesssim e_i^{-(\alpha_i + \tilde{d})} 2^{-\sigma_{\mathbf{j}, \mathbf{j}'}} .$$

Thus, Schur's lemma yields that $\mathbb{S}^{i, \mathcal{I}}$ fulfills the requirements of Theorem 6.4 for any $i \in \{1, \dots, n\}$ and $\mathcal{I} \subset \{1, \dots, n\}$ with $\varepsilon = e_i^{-(\alpha_i + \tilde{d})}$. As above, this ensures the consistency of \mathbb{A}_J^{cpr-2} . \square

8. COMPLEXITY ESTIMATES

In this section, we turn to the complexity analysis of the compression scheme of Section 7. We split the analysis subject to the choice of $\tilde{\sigma}_{\mathbf{j}, \mathbf{j}'}$.

8.1. Complexity of the compression scheme with $\tilde{\sigma}_{\mathbf{j}, \mathbf{j}'} = \sigma'_{\mathbf{j}, \mathbf{j}'}$

In short, for arbitrary $n \geq 2$, $\underline{\alpha} \in \mathbb{R}_{\geq 0}^n$ and $d \in \mathbb{N}$, the main results of this section reads:

There exists $\tilde{d} \in \mathbb{N}$ such that for any operator \mathcal{A} of order $\underline{\alpha}$ the compression scheme of Section 7 with $\tilde{\sigma}_{\mathbf{j}, \mathbf{j}'} = \sigma'_{\mathbf{j}, \mathbf{j}'}$ as defined in (6.16) yields essentially optimal complexity $\mathcal{O}(2^J J^{2(n-1)})$.

More precisely, we prove

Theorem 8.1. *Let $\underline{\alpha} \in \mathbb{R}_{\geq 0}^n$ and $0 < \bar{\alpha} \leq 2d \leq 2d'$. Suppose further*

$$2nd' < \tilde{d} + (n+1)\bar{\alpha}. \quad (8.1)$$

Then for any operator $\mathcal{A} = p(D)$ with $p \in \Gamma^{\underline{\alpha}}(\mathbb{R}^n)$ the number of non-zero entries in the matrix \mathbb{A}_J^{cpr} defined by the compression scheme in Section 7 with $\tilde{\sigma}_{\mathbf{j}, \mathbf{j}'} = \sigma'_{\mathbf{j}, \mathbf{j}'}$ is $\mathcal{O}(2^J J^{2(n-1)})$.

Proof. Without loss of generality we assume the worst but admissible case $\alpha_1 = \dots = \alpha_n = \bar{\alpha} > 0$. Fix any $\mathcal{I} \subset \{1, \dots, n\}$. Since \mathcal{I} is arbitrary, it suffices to show that there are $\mathcal{O}(2^J J^{2(n-1)})$ entries $\langle \mathcal{A}\psi_{\mathbf{j}, \mathbf{k}}, \psi_{\mathbf{j}', \mathbf{k}'} \rangle$ of \mathbb{A}_J^{cpr} with

$$\delta_{x_i} \leq 2^{-\min\{j_i, j'_i\}}, \quad \text{for all } i \in \mathcal{I}, \quad \text{and} \quad \delta_{x_i} \geq 2^{-\min\{j_i, j'_i\}}, \quad \text{for all } i \notin \mathcal{I}. \quad (8.2)$$

Based on the compression scheme of Section 7, in each matrix block $\mathbb{A}_{\mathbf{j}, \mathbf{j}'}$ of \mathbb{A}_J^{cpr} we divide the coordinate directions into four groups. Let

$$D_1 := \left\{ r \in \mathcal{I} : 2^{-\min\{j_r, j'_r\}} \leq E_{\mathbf{j}, \mathbf{j}'}^{r, \mathcal{I}} \right\}, \quad (8.3)$$

$$D_2 := \left\{ s \in \mathcal{I} : E_{\mathbf{j}, \mathbf{j}'}^{s, \mathcal{I}} \leq 2^{-\min\{j_s, j'_s\}} \text{ and } 2^{-\max\{j_s, j'_s\}} = E_{\mathbf{j}, \mathbf{j}'}^{s, \mathcal{I}} \right\}, \quad (8.4)$$

$$D_3 := \left\{ t \in \mathcal{I} : E_{\mathbf{j}, \mathbf{j}'}^{t, \mathcal{I}} \leq 2^{-\min\{j_t, j'_t\}} \text{ and } 2^{-\max\{j_t, j'_t\}} < E_{\mathbf{j}, \mathbf{j}'}^{t, \mathcal{I}} \right\}, \quad (8.5)$$

$$D_4 := \left\{ i \in \{1, \dots, n\} \setminus \mathcal{I} \right\}. \quad (8.6)$$

Obviously $D_1 \cup D_2 \cup D_3 \cup D_4 = \{1, \dots, n\}$. By definition of the compression scheme, the number $\#\mathbb{A}_{\mathbf{j}, \mathbf{j}'}$ of non-zero entries in each matrix block $\mathbb{A}_{\mathbf{j}, \mathbf{j}'}$ can be bounded by

$$\begin{aligned} \#\mathbb{A}_{\mathbf{j}, \mathbf{j}'} &= \mathcal{O} \left(\prod_{r \in D_1} 2^{j_r + j'_r} 2^{-\min\{j_r, j'_r\}} \cdot \prod_{s \in D_2} 2^{j_s + j'_s} E_{\mathbf{j}, \mathbf{j}'}^{s, \mathcal{I}} \cdot \prod_{t \in D_3} 2^{j_t + j'_t} E_{\mathbf{j}, \mathbf{j}'}^{t, \mathcal{I}} \cdot \prod_{i \in D_4} 2^{j_i + j'_i} C_{\mathbf{j}, \mathbf{j}'}^{i, \mathcal{I}} \right) \\ &= \mathcal{O} \left(\prod_{r \in D_1} 2^{\max\{j_r, j'_r\}} \cdot \prod_{s \in D_2} 2^{\min\{j_s, j'_s\}} \right. \\ &\quad \times \prod_{t \in D_3} 2^{j_t + j'_t} 2^{\frac{2J(d' - \frac{\bar{\alpha}}{2}) - d'(|\mathbf{j}|_1 + |\mathbf{j}'|_1) - \bar{d} \max\{j_t, j'_t\} + s_t + \frac{1}{2} \sum_{l \in \mathcal{I} \setminus \{t\}} m_l - \bar{d} \sum_{l \notin \mathcal{I}} m_l}{d + \bar{\alpha}}} \\ &\quad \left. \times \prod_{i \in D_4} 2^{j_i + j'_i} 2^{\frac{2J(d' - \frac{\bar{\alpha}}{2}) - d'(|\mathbf{j}|_1 + |\mathbf{j}'|_1) - \bar{d}(j_i + j'_i) + s_i + \frac{1}{2} \sum_{l \in \mathcal{I}} m_l - \bar{d} \sum_{l \notin \mathcal{I} \cup \{i\}} m_l}{2d + \bar{\alpha}}} \right), \end{aligned} \quad (8.7)$$

with $s_l, m_l, l = 1, \dots, n$, as in Section 7. To simplify this notation, from now on we assume without loss of generality that $j'_l \leq j_l$ for all $l = 1, \dots, n$. The result for all other index combinations follows analogously.

Regrouping the single factors in (8.7) corresponding to their level index yields

$$\#\mathbb{A}_{\mathbf{j}, \mathbf{j}'} = \mathcal{O} \left(C_0 \cdot \prod_{r \in D_1} R_r \cdot \prod_{s \in D_2} S_s \cdot \prod_{t \in D_3} T_t \cdot \prod_{i \in D_4} I_i \right), \quad (8.8)$$

where, with $N := \frac{\#D_3}{d + \bar{\alpha}} + \frac{\#D_4}{2d + \bar{\alpha}}$, we have set

$$\begin{aligned} C_0 &= 2^{N(2J(d' - \frac{\bar{\alpha}}{2}))}, \\ R_m &= 2^{j_m} 2^{-N(d'(j_m + j'_m) - \bar{\alpha}j'_m - \frac{1}{2}(j_m - j'_m))}, \\ S_m &= 2^{j'_m} 2^{-N(d'(j_m + j'_m) - \bar{\alpha}j'_m - \frac{1}{2}(j_m - j'_m))}, \\ T_m &= 2^{j_m + j'_m} 2^{-\frac{\bar{d}}{d + \bar{\alpha}} j_m} 2^{-N(d'(j_m + j'_m) - \bar{\alpha}j'_m - \frac{1}{2}(j_m - j'_m))} 2^{-\frac{1}{d + \bar{\alpha}}(\bar{\alpha}j'_m + \frac{1}{2}(j_m - j'_m))}, \\ I_m &= 2^{j_m + j'_m} 2^{-\frac{\bar{d}}{2d + \bar{\alpha}}(j_m + j'_m)} 2^{-N(d'(j_m + j'_m) - \bar{\alpha}j'_m + \bar{d}(j_m - j'_m))} \cdot 2^{-\frac{1}{2d + \bar{\alpha}}(\bar{\alpha}j'_m - \bar{d}(j_m - j'_m))}, \end{aligned} \quad (8.9)$$

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

By definition of D_1, D_3 , there holds $R_r \leq T_r$ for all $r \in D_1$. In order to keep notation feasible (see [46, Remark 4.6.2] for a more detailed approach) we estimate

$$\#\mathbb{A}_{\mathbf{j}, \mathbf{j}'} = \mathcal{O} \left(C_0 \cdot \prod_{s \in D_2} S_s \cdot \prod_{t \in D_1 \cup D_3} T_t \cdot \prod_{i \in D_4} I_i \right). \quad (8.10)$$

Denote $\tilde{j}_m := (j_m + j'_m)/2$, $m = 1, \dots, n$. By (8.1) there holds

$$(1 - \bar{\alpha})N \leq (1 - \bar{\alpha}) \frac{n}{d + \bar{\alpha}} \leq 1,$$

which implies $N/2 \leq 1 - N(1/2 - \bar{\alpha})$. For each $s \in D_2$ one thus obtains

$$\begin{aligned} S_s &= 2^{(1 - N(\frac{1}{2} - \bar{\alpha}))j'_s} 2^{\frac{1}{2}Nj_s} 2^{-Nd'(j_2 + j'_2)} \\ &= 2^{(1 - N(\frac{1}{2} - \bar{\alpha}))j'_s} 2^{\frac{1}{2}Nj_s} 2^{-2Nd'\tilde{j}_s} \end{aligned}$$

$$\begin{aligned} &\leq 2^{\tilde{j}_s} 2^{-N(2d' - \bar{\alpha})\tilde{j}_s} \\ &=: \tilde{S}_s. \end{aligned}$$

Analogously, one infers from (8.1) that $n \leq \tilde{d} + (n-1)\bar{\alpha} + 1$ which implies for any $t \in D_1 \cup D_3$,

$$\begin{aligned} T_t &= 2^{\left(1 - \frac{\tilde{d}}{d+\bar{\alpha}} + \frac{1}{2}N - \frac{1}{2d+2\bar{\alpha}}\right)j_t} 2^{\left(1+N(\bar{\alpha}-\frac{1}{2}) - \frac{\bar{\alpha}-1/2}{d+\bar{\alpha}}\right)j'_t} 2^{-Nd'(j_t+j'_t)} \\ &\leq 2^{2\tilde{j}_t} 2^{-\frac{\tilde{d}}{d+\bar{\alpha}}\tilde{j}_t} 2^{N\bar{\alpha}\tilde{j}_t} 2^{-\frac{\bar{\alpha}}{d+\bar{\alpha}}\tilde{j}_t} 2^{-N2d'\tilde{j}_t} \\ &=: \tilde{T}_t. \end{aligned}$$

In the same way, one immediately obtains for $i \in D_4$ that $I_i \leq 2^{\tilde{j}_i} 2^{-N(2d' - \bar{\alpha})\tilde{j}_i} =: \tilde{I}_i$. But, there holds $\tilde{S}_m = \tilde{T}_m = \tilde{I}_m$ for $m = 1, \dots, n$ and from (8.10) one infers

$$\begin{aligned} \#\mathbb{A}_{\mathbf{j}, \mathbf{j}'} &= \mathcal{O}\left(C_0 \cdot \prod_{s \in D_2} \tilde{S}_s \cdot \prod_{t \in D_1 \cup D_3} \tilde{T}_t \cdot \prod_{i \in D_4} \tilde{I}_i\right) \\ &= \mathcal{O}\left(2^{N(2J(d' - \frac{\bar{\alpha}}{2}))} \cdot \prod_{m=1}^n 2^{\tilde{j}_m} 2^{-N(2d' - \bar{\alpha})\tilde{j}_m}\right) = \mathcal{O}\left(2^{N(2d' - \bar{\alpha})(J - |\tilde{\mathbf{j}}|_1)} 2^{|\tilde{\mathbf{j}}|_1}\right), \end{aligned}$$

where $|\tilde{\mathbf{j}}|_1 = \tilde{j}_1 + \dots + \tilde{j}_n$. Using $N \leq \frac{n}{d+\bar{\alpha}}$ one finally obtains

$$\#\mathbb{A}_{\mathbf{j}, \mathbf{j}'} = \mathcal{O}\left(2^J \cdot 2^{\left(\frac{n(2d' - \bar{\alpha})}{d+\bar{\alpha}} - 1\right)(J - |\tilde{\mathbf{j}}|_1)}\right). \quad (8.11)$$

Since (8.1) implies $\frac{n(2d' - \bar{\alpha})}{d+\bar{\alpha}} < 1$, summing over all matrix blocks yields that there are $\mathcal{O}(2^J J^{2(n-1)})$ entries satisfying (8.2). \square

Remark 8.2. Clearly, for any d' and $\bar{\alpha}$, requirement (8.1) is satisfied for some sufficiently large \tilde{d} . Since in (8.10) we have introduced some sub-optimality, this parameter restriction is sufficient but not necessary, see [46, Remark 4.6.2].

8.2. Complexity of the compression scheme with $\tilde{\sigma}_{\mathbf{j}, \mathbf{j}'} = \sigma_{\mathbf{j}, \mathbf{j}'}$

Based on the methodology for a detailed complexity analysis of the previous section, we now turn to the complexity estimates for the compression scheme of Section 7 with $\tilde{\sigma}_{\mathbf{j}, \mathbf{j}'} = \sigma_{\mathbf{j}, \mathbf{j}'}$, where $\sigma_{\mathbf{j}, \mathbf{j}'}$ is given by (6.13). In short, with arbitrary $n \geq 2$, $\tilde{\underline{\alpha}} \geq \underline{\alpha} \in \mathbb{R}_{\geq 0}^n$ and $d \in \mathbb{N}$, one finds:

For any $\varepsilon > 0$ there exists $\tilde{d} \in \mathbb{N}$ such that for any operator \mathcal{A} of order $\underline{\alpha}$ the compression scheme of Section 7 with $\tilde{\sigma}_{\mathbf{j}, \mathbf{j}'} = \sigma_{\mathbf{j}, \mathbf{j}'}$ as defined in (6.16) yields complexity $\mathcal{O}(2^{(1+\varepsilon)J})$.

Furthermore, the compression scheme with $\tilde{\sigma}_{\mathbf{j}, \mathbf{j}'} = \sigma_{\mathbf{j}, \mathbf{j}'}$ yields optimal complexity $\mathcal{O}(2^J J^{2(n-1)})$ if the integrodifferential operator $\mathcal{B} = \mathcal{A}_D + \mathcal{A}$ admits a non-vanishing differential part \mathcal{A}_D and the order $\underline{\alpha}$ of the integral part \mathcal{A} is sufficiently small, see Theorem 8.3.

At first, we extend the rather conservative estimates and methodology of Section 8.1 to the case $\tilde{\sigma}_{\mathbf{j}, \mathbf{j}'} = \sigma_{\mathbf{j}, \mathbf{j}'}$. For sake of simplicity, for now assume that the differential operator \mathcal{A}_D in (4.1) does not vanish or, more precisely, that the order $\underline{\alpha}$ of the integral operator \mathcal{A} is less than the order $\tilde{\underline{\alpha}}$ of the integrodifferential operator $\mathcal{B} = \mathcal{A}_D + \mathcal{A}$, i.e.

$$\underline{\alpha} < \tilde{\underline{\alpha}} = (2, \dots, 2) \in \mathbb{R}^n, \quad (8.12)$$

as defined in (6.7). The case $\underline{\alpha} = \tilde{\underline{\alpha}}$ is discussed afterwards.

Theorem 8.3. *Suppose that (8.12) holds. Let $\mathbb{A}_J^{\text{compr}}$ be the matrix defined by the compression scheme of Section 7 with $\tilde{\sigma}_{\mathbf{j}, \mathbf{j}'} = \sigma_{\mathbf{j}, \mathbf{j}'}$ as defined in (6.13).*

If (8.1) holds, the number of non-zero matrix entries in each block $\mathbb{A}_{\mathbf{j}, \mathbf{j}'}$ of $\mathbb{A}_J^{\text{compr}}$ can be bounded by

$$\#\mathbb{A}_{\mathbf{j}, \mathbf{j}'} \lesssim 2^J 2^{\frac{2-n\bar{\alpha}}{2d+\bar{\alpha}}J}. \quad (8.13)$$

In particular, the complexity of \mathbb{A}_J^{compr} is $\mathcal{O}(2^J J^{2(n-1)})$ if $\bar{\alpha} \leq 2/n$.

Remark 8.4. The complexity bound (8.13) is only valid under (8.1) - which implies that the number of vanishing moments \tilde{d} has to be increased with the dimension. In particular, the rather conservative estimate (8.13) does not imply the ‘‘curse of dimension’’ even though the dimension n occurs in the exponent on the right hand side. With sufficiently large \tilde{d} the right hand side of (8.13) can be chosen arbitrarily close to the optimality estimate $\#\mathbb{A}_{\mathbf{j},\mathbf{j}'} \lesssim 2^J$ regardless of the choice of $\underline{\alpha}$, cf. [46, Section 4.6.2].

Proof. By Theorem 8.1, it suffices to prove the result only for those matrix blocks that satisfy $\sigma_{\mathbf{j},\mathbf{j}'} \neq \sigma'_{\mathbf{j},\mathbf{j}'}$. Hence, from the definition (6.13) of $\sigma_{\mathbf{j},\mathbf{j}'}$ one obtains that there are essentially two cases: either

$$\sigma_{\mathbf{j},\mathbf{j}'} = -\frac{1}{2} (|\tilde{\underline{\alpha}} \cdot \mathbf{j}'|_\infty + |\tilde{\underline{\alpha}} \cdot \mathbf{j}|_\infty), \quad (8.14)$$

or

$$\sigma_{\mathbf{j},\mathbf{j}'} = \sum_{i=1}^n \left[\frac{1}{n} J \left(\rho_i - \frac{\tilde{\alpha}_i}{2} \right) - \rho_i j'_i \right] - \frac{1}{2} |\tilde{\underline{\alpha}} \cdot \mathbf{j}|_\infty \quad (8.15)$$

The remaining possible case $\sigma_{\mathbf{j},\mathbf{j}'} = \sum_{i=1}^n \left[\frac{1}{n} J (\rho_i - \tilde{\alpha}_i/2) - \rho_i j_i \right] - \frac{1}{2} |\tilde{\underline{\alpha}} \cdot \mathbf{j}'|_\infty$ in (6.13) follows analogously to the case (8.15).

At first, suppose (8.14) holds. By (8.12), this simplifies to $\sigma_{\mathbf{j},\mathbf{j}'} = -(|\mathbf{j}|_\infty + |\mathbf{j}'|_\infty)$. As above, assume without loss of generality that $j'_i \leq j_i$, $i = 1, \dots, n$. The proof for all other index constellations follows analogously. Let D_1, \dots, D_4 be as in (8.3)–(8.6). Using exactly the same arguments as in the proof of Theorem 8.1, analogously to (8.8) one obtains that the number of non-zero entries in each matrix block $\mathbb{A}_{\mathbf{j},\mathbf{j}'}$ satisfies

$$\#\mathbb{A}_{\mathbf{j},\mathbf{j}'} = \mathcal{O} \left(C_0 \cdot \prod_{r \in D_1} R_r \cdot \prod_{s \in D_2} S_s \cdot \prod_{t \in D_3} T_t \cdot \prod_{i \in D_4} I_i \right), \quad (8.16)$$

where, with $N = \frac{\#D_3}{d+\bar{\alpha}} + \frac{\#D_4}{2d+\bar{\alpha}}$, according to the definition of the cut-off parameters $E_{\mathbf{j},\mathbf{j}'}^{i,\mathcal{I}}, C_{\mathbf{j},\mathbf{j}'}^{i,\mathcal{I}}$ we now have

$$\begin{aligned} C_0 &= 2^{-N(|\mathbf{j}|_\infty + |\mathbf{j}'|_\infty)}, \\ R_m &= 2^{j_m} 2^{N(\bar{\alpha}j'_m + \frac{1}{2}(j_m - j'_m))}, \\ S_m &= 2^{j'_m} 2^{N(\bar{\alpha}j'_m + \frac{1}{2}(j_m - j'_m))}, \\ T_m &= 2^{j_m + j'_m} 2^{-\frac{\tilde{d}}{d+\bar{\alpha}}j_m} 2^{N(\bar{\alpha}j'_m + \frac{1}{2}(j_m - j'_m))} 2^{-\frac{1}{d+\bar{\alpha}}(\bar{\alpha}j'_m + \frac{1}{2}(j_m - j'_m))}, \\ I_m &= 2^{j_m + j'_m} 2^{-\frac{\tilde{d}}{2d+\bar{\alpha}}(j_m + j'_m)} 2^{N(\bar{\alpha}j'_m - \tilde{d}(j_m - j'_m))} 2^{-\frac{1}{2d+\bar{\alpha}}(\bar{\alpha}j'_m - \tilde{d}(j_m - j'_m))}, \end{aligned}$$

for $m = 1, \dots, n$. By definition of D_1, D_3 , there holds $R_r \leq T_r$ for all $r \in D_1$. Thus, one may estimate

$$\#\mathbb{A}_{\mathbf{j},\mathbf{j}'} = \mathcal{O} \left(C_0 \cdot \prod_{s \in D_2} S_s \cdot \prod_{t \in D_1 \cup D_3} T_t \cdot \prod_{i \in D_4} I_i \right). \quad (8.17)$$

Denoting

$$\tilde{j}_m := \frac{j_m + j'_m}{2}, \quad m = 1, \dots, n. \quad (8.18)$$

for each $s \in D_2$ one obtains

$$S_s = 2^{(1-N(\frac{1}{2}-\bar{\alpha}))j'_s} 2^{\frac{1}{2}Nj_s} \leq 2^{\tilde{j}_s} 2^{N\bar{\alpha}\tilde{j}_s} =: \tilde{S}_s.$$

Analogously one finds $T_t \leq \tilde{S}_t$ for each $t \in D_1 \cup D_3$ and $I_i \leq \tilde{S}_i$ for each $i \in D_4$. Note that equality holds in these estimates if and only if $j_m = j'_m$, i.e. the ‘‘worst case’’ is obtained when $j_m = j'_m$ for all $m = 1, \dots, n$. From (8.17), one obtains

$$\#\mathbb{A}_{\mathbf{j},\mathbf{j}'} \lesssim C_0 \cdot \prod_{s \in D_2} \tilde{S}_s \cdot \prod_{t \in D_1 \cup D_3} \tilde{S}_t \cdot \prod_{i \in D_4} \tilde{S}_i \lesssim 2^{-N(|\mathbf{j}|_\infty + |\mathbf{j}'|_\infty)} \cdot \prod_{m=1}^n 2^{\tilde{j}_m} 2^{N\bar{\alpha}\tilde{j}_m}. \quad (8.19)$$

The right hand side of (8.19) reaches its maximum when $|\tilde{j}|_1 = J$. By (8.18), this implies $|\mathbf{j}|_\infty = |\mathbf{j}'|_\infty = J/n$ and one obtains

$$\#\mathbb{A}_{\mathbf{j},\mathbf{j}'} \lesssim 2^J 2^{N(\bar{\alpha} - \frac{2}{n})J}. \quad (8.20)$$

Recall that this maximal complexity can only be obtained when $|\tilde{j}|_1 = J$ and correspondingly $j_m = j'_m = J/n$, $m = 1, \dots, n$. In this case $D_3 = \emptyset$ and hence

$$N = \frac{\#D_4}{2\tilde{d} + \bar{\alpha}} \leq \frac{n}{2\tilde{d} + \bar{\alpha}}.$$

Thus, in case (8.14) holds, one finds the desired estimate

$$\#\mathbb{A}_{\mathbf{j},\mathbf{j}'} \lesssim 2^J 2^{\frac{n\bar{\alpha}-2}{2\tilde{d}+\bar{\alpha}}J}.$$

Finally, suppose (8.15) holds. By (8.12), this simplifies to $\sigma_{\mathbf{j},\mathbf{j}'} = J(d-1) - d|\mathbf{j}'|_1 - |\mathbf{j}|_\infty$. Clearly, $\sigma_{\mathbf{j},\mathbf{j}'}$ becomes maximal, when $j'_i \leq j_i$, $i = 1, \dots, n$. With this choice of \mathbf{j} , \mathbf{j}' and $\sigma_{\mathbf{j},\mathbf{j}'}$ one may proceed as above to re-derive (8.17), where now

$$\begin{aligned} C_0 &= 2^{N(J(d-1) - |\mathbf{j}|_\infty)}, \\ S_m &= 2^{j'_m} 2^{N(\bar{\alpha}j'_m + \frac{1}{2}(j_m - j'_m) - dj'_m)}, \\ T_m &= 2^{j_m + j'_m} 2^{-\frac{\tilde{d}}{d+\bar{\alpha}}j_m} 2^{N(\bar{\alpha}j'_m + \frac{1}{2}(j_m - j'_m) - dj'_m)} 2^{-\frac{1}{d+\bar{\alpha}}(\bar{\alpha}j'_m + \frac{1}{2}(j_m - j'_m))}, \\ I_m &= 2^{j_m + j'_m} 2^{-\frac{\tilde{d}}{2\tilde{d}+\bar{\alpha}}(j_m + j'_m)} 2^{N(\bar{\alpha}j'_m - \tilde{d}(j_m - j'_m) - dj'_m)} 2^{-\frac{1}{2\tilde{d}+\bar{\alpha}}(\bar{\alpha}j'_m - \tilde{d}(j_m - j'_m))}, \end{aligned}$$

for $m = 1, \dots, n$. Now with $\tilde{S}_m := 2^{\tilde{j}_m} 2^{-N(d-\bar{\alpha})\tilde{j}_m}$ and \tilde{j}_m as in (8.18), $m = 1, \dots, n$, one finds $S_s \leq \tilde{S}_s$ for each $s \in D_2$, $T_t \leq \tilde{S}_t$ for each $t \in D_1 \cup D_3$ and $I_i \leq \tilde{S}_i$ for each $i \in D_4$. Thus, there holds

$$\#\mathbb{A}_{\mathbf{j},\mathbf{j}'} \lesssim 2^{N(J(d-1) - |\mathbf{j}|_\infty)} \cdot \prod_{m=1}^n 2^{\tilde{j}_m} 2^{-N(d-\bar{\alpha})\tilde{j}_m} \sim 2^{N(d-1)J} 2^{(1-N(d-\bar{\alpha}))|\tilde{j}|_1} 2^{-N|\mathbf{j}|_\infty}. \quad (8.21)$$

Since by definition $N \leq \frac{n}{d+\bar{\alpha}}$, condition (8.1) implies $N/n \leq 1 - N(d - \bar{\alpha})$, and hence (8.21) reaches its maximum when $|\tilde{j}|_1 = J$ and hence $|\mathbf{j}|_\infty = J/n$. This yields

$$\#\mathbb{A}_{\mathbf{j},\mathbf{j}'} \lesssim 2^J 2^{N(\bar{\alpha} - 1 - \frac{1}{n})J}, \quad (8.22)$$

and the result follows, since the right hand side of (8.22) can be bounded by the right hand side of (8.20). \square

Now, consider the case $\underline{\alpha} = \tilde{\underline{\alpha}}$, i.e. $\mathcal{A}_D = 0$ and hence $\mathcal{B} = \mathcal{A}$ in (4.1) has no differential component. We assume without loss of generality that in each direction we have the same (strongest) marginal singularity, i.e.

$$\alpha_1 = \dots = \alpha_n = \bar{\alpha}. \quad (8.23)$$

The structure of the cut-off parameters defined in Section 7 implies that assumption (8.23) provides a “worst” but admissible case. Using the arguments of Theorem 8.3 one infers that the matrix block with the greatest asymptotic complexity in \mathbb{A}_J^{compr} is given when $\mathbf{j} = \mathbf{j}' = \underline{J}/n$, i.e.

$$j_1 = \dots = j_n = j'_1 = \dots = j'_n = \frac{1}{n}J. \quad (8.24)$$

For this index combination one obtains

$$L(\mathbf{j}, d) = d(J - |\mathbf{j}|_1) = 0 < (1 - \frac{1}{n})\frac{\bar{\alpha}}{2}J = \frac{\bar{\alpha}}{2}J - \frac{\bar{\alpha}}{2}|\mathbf{j}|_\infty = R(\mathbf{j}, \tilde{\underline{\alpha}}),$$

and analogously $L(\mathbf{j}', d) < R(\mathbf{j}', \tilde{\alpha})$, with $R(\cdot, \cdot)$ and $L(\cdot, \cdot)$ as in Theorem 6.4. Thus, by (6.13) there holds

$$\sigma_{\underline{J}/n, \underline{J}/n} = -\frac{1}{n}\tilde{\alpha}J.$$

Herewith, one obtains that for any coordinate direction $i = 1, \dots, n$, there holds

$$2^{-\min\{j_i, j'_i\}} = 2^{-\frac{1}{n}J} \leq E_{\mathbf{j}, \mathbf{j}'}^{i, \mathcal{I}}$$

and hence, using (8.3)–(8.6), one finds $D_2 = D_3 = \emptyset$ and $D_1 = \mathcal{I}$, $D_4 = \{1, \dots, n\} \setminus D_1$. Thus, in this one matrix block, the number $\#\mathbb{A}_{\underline{J}/n, \underline{J}/n}$ of all non-zero entries that satisfy $\delta_{x_i} < 2^{-\min\{j_i, j'_i\}}$ for all $i \in \mathcal{I}$, is given by

$$\#\mathbb{A}_{\underline{J}/n, \underline{J}/n} = \mathcal{O}\left(\prod_{r \in D_1} 2^{\max\{j_r, j'_r\}} \cdot \prod_{i \in D_4} 2^{j_1 + j'_1} C_{\mathbf{j}, \mathbf{j}'}^{i, \mathcal{I}}\right) = \mathcal{O}\left(\prod_{r \in D_1} 2^{\frac{1}{n}J} \cdot \prod_{i \in D_4} \left[2^{\frac{2}{n}J} 2^{-\frac{\tilde{\alpha} + 2\tilde{d}}{n(2d + \tilde{\alpha})}J} 2^{\frac{s_i}{2d + \tilde{\alpha}}}\right]\right),$$

with $s_i = \frac{n-1}{n}\tilde{\alpha}J$ as defined in (7.1). Clearly (8.25) reaches its maximum when $D_4 = \{1, \dots, n\}$, i.e. $\mathcal{I} = \emptyset$ and hence one considers all matrix entries that satisfy $\delta_{x_i} > 2^{-\min\{j_i, j'_i\}}$ in all coordinate directions $i = 1, \dots, n$. The number of non-zero entries of such wavelets in $\mathbb{A}_{\underline{J}/n, \underline{J}/n}$ hence satisfies

$$\#\mathbb{A}_{\underline{J}/n, \underline{J}/n} \lesssim \prod_{i=1}^n \left[2^{\frac{2}{n}J} 2^{-\frac{\tilde{\alpha} + 2\tilde{d}}{n(2d + \tilde{\alpha})}J} 2^{\frac{s_i}{2d + \tilde{\alpha}}}\right] \sim 2^J 2^{\frac{1}{2d + \tilde{\alpha}} \sum_{i=1}^n s_i} \sim 2^J 2^{\frac{(n-1)\tilde{\alpha}}{2d + \tilde{\alpha}}J}.$$

Thus, due to the existence of the correction terms s_i in the cut-off parameters in Section 7 the anisotropic compression scheme does in general not yield optimal complexity when $\mathcal{B} = \mathcal{A}$ is an integral operator. Nonetheless, by choosing a wavelet basis with a sufficiently large number of vanishing moments the scheme yields complexity $\mathcal{O}(2^{(1+\varepsilon)J})$ for any given $\varepsilon > 0$.

9. NUMERICAL RESULTS

In this final section, we provide some basic numerical illustrations regarding the accuracy of the presented matrix entry estimates as well as the complexity of the compressed sparse tensor product stiffness matrix. For further illustrations e.g. on the impact of the first and the second compression techniques we refer to [46, Chapter 6].

In short, the numerical results show that the compression scheme of Section 7 provides a very accurate prediction of the structure of the stiffness matrix of anisotropic operators. Hence, one may conclude that the consistency analysis and the matrix entry estimates of the previous sections are accurate. The numerical results also confirm that the complexity of the compressed sparse tensor product stiffness matrix is (asymptotically) of the magnitude that was proved in Section 8.

The numerical results have been obtained using tensor products of the piecewise linear biorthogonal spline wavelets $\psi_{j,k}$ constructed in Example 2.2. Recall that with this choice of basis one has $d = \tilde{d} = 2$ and $|\text{supp } \psi_{j,k}| = 2 \cdot 2^{-j}$ for all $\psi_{j,k} \in \Psi$.

Remark 9.1. Note that the absolute complexity values presented in the following are only valid for the wavelets of Example 2.2. While the asymptotic behavior of the matrix complexity is independent of the particular choice of basis, the complexity constants naturally depend on this choice. For example, reducing the size of the wavelets' supports may decrease this constant significantly. Similar observations have already been made in the context of isotropic wavelet compression, cf. e.g. [28].

9.1. Accuracy

To analyze the accuracy of our compression predictions and the fundamental estimates given in Section 5, we consider the following model problem:

Find the numerical solution of the integrodifferential equation

$$\mathcal{A}u = 0, \quad \text{on } [0, 1]^2,$$

where \mathcal{A} denotes the infinitesimal generator of a bivariate Lévy copula process with tempered stable margins and Clayton-type Lévy copula F as defined in [48]. The parameters for the marginal Lévy measures are chosen as follows: $c_1 = c_2 = 1$, $\beta_1 = \beta_2 = 8$ and $\alpha_1 = \alpha_2 = 1$. The parameters of the Lévy copula of Clayton-type are $\theta = 10$, $\eta = 0.5$.

The structure of the stiffness matrix of \mathcal{A} in the full tensor product space V_J of (2.21) is illustrated in Figure 1, right. Naturally, this matrix is never computed in practice.

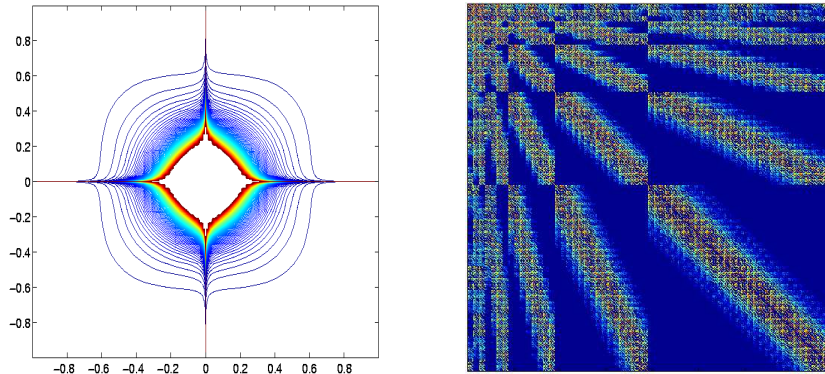


FIGURE 1. Left: Contour plot of the density of the infinitesimal generator \mathcal{A} . Right: Stiffness matrix of the Galerkin discretization of \mathcal{A} in the full tensor product space V_J as in (2.21) with wavelet basis. Level $J = 5$, 4096^2 entries, color coded according to $\log_{10} |\langle \mathcal{A}\psi_{j,\mathbf{k}}, \psi_{j',\mathbf{k}'} \rangle|$.

On level $J = 5$, the structure of stiffness matrix of \mathcal{A} in the sparse tensor product space \widehat{V}_J of (2.23) (without any compression) is shown in Figure 2. On the right hand side the prediction of the compression scheme of Section 7 with $\tilde{\sigma}_{j,j'} = \sigma_{j,j'}$ is given. In practice, only the black entries on the right hand side of Figure 2 need to be computed.

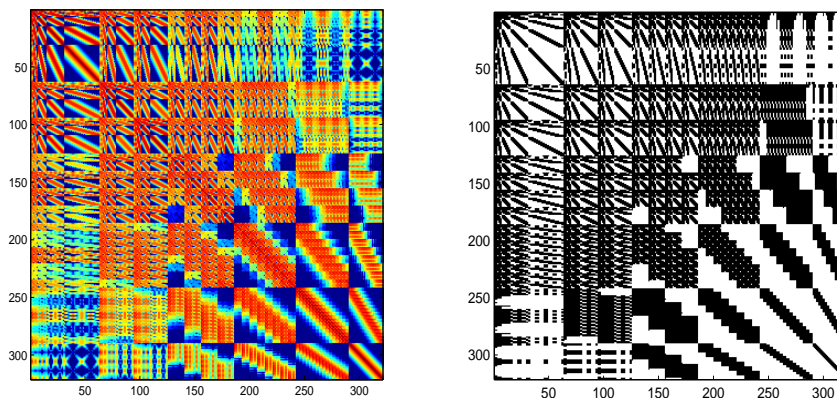


FIGURE 2. Accurate sparsity pattern prediction by the compression scheme. Left: Actual stiffness matrix of \mathcal{A} in the sparse tensor product space \widehat{V}_J of (2.23), 320^2 non-zero entries, color coded according to $\log_{10} |\langle \mathcal{A}\psi_{j,\mathbf{k}}, \psi_{j',\mathbf{k}'} \rangle|$. Right: A-priori sparsity pattern prediction by the compression scheme of Section 7.

The full and sparse tensor product stiffness matrices without compression in Figures 1 & 2 were taken from [61]. We also refer to this source for further numerical results.

9.2. Complexity

In Figure 3, the number of non-zero matrix entries of the non-compressed and the compressed sparse tensor product matrix are compared. Hereto, we set $n = 2$ and choose all constants $c_i, e_i, i = 1, 2$, in the compression scheme of Section 7 to be $c_i = e_i = 1, i = 1, 2$.

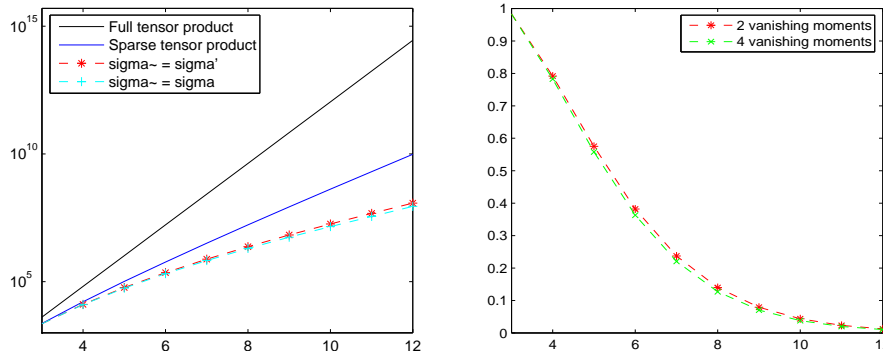


FIGURE 3. Left: Number of non-zero matrix entries of the different stiffness matrices of \mathcal{A} per level $J = 3, \dots, 12$. Right: Percentage of the remaining complexity of the sparse tensor product matrix after compression on levels $J = 3, \dots, 12$.

On the left hand side of Figure 3, the absolute number of non-zero entries is plotted per level $J = 3, \dots, 12$. Here the red line represents the matrix complexity based on the wavelets of Example 2.2 after compression by the scheme of Section 7 with $\tilde{\sigma}_{j,j'} = \sigma'_{j,j'}$. The cyan line represents the scheme with $\tilde{\sigma}_{j,j'} = \sigma_{j,j'}$, where we have assumed (8.12). On the right hand side, the percentage of matrix entries of the sparse tensor product matrix that remain in the compressed matrix (with $\tilde{\sigma}_{j,j'} = \sigma'_{j,j'}$) is given.

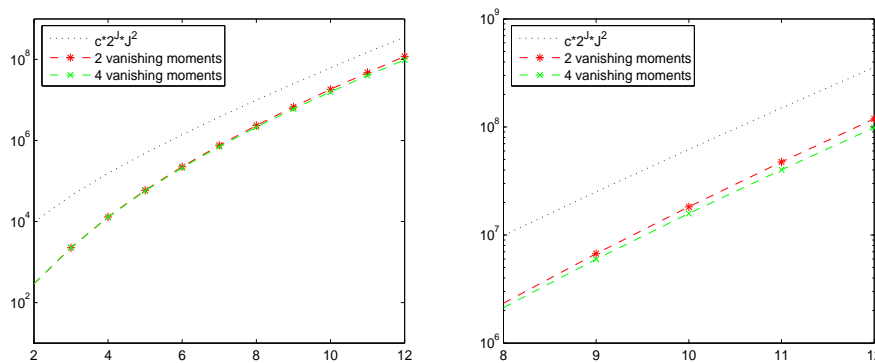


FIGURE 4. Comparison of the number of non-zero entries in the compressed sparse tensor product stiffness matrix and the predicted rate $\mathcal{O}(2^J J^2)$ of Theorem 8.1. Left: On levels $J = 2, \dots, 12$. Right: On levels $J = 8, \dots, 12$.

In Figure 4, the complexity growth of the compressed sparse tensor product matrix is compared with the results predicted by Theorem 8.1. The increasing sparsity of the compressed matrices is illustrated in Figure 5. Here we have discretized an operator $\mathcal{A} = p(D)$ with $p \in \Gamma^\alpha(\mathbb{R}^2)$, $\underline{\alpha} = (0.5, 1.5)$, using the wavelets of Example 2.2.

REFERENCES

- [1] Y. Achdou and O. Pironneau. A numerical procedure for calibration of volatility with american options. *Appl. Math. Finance*, 12(3):201–241, 2005.

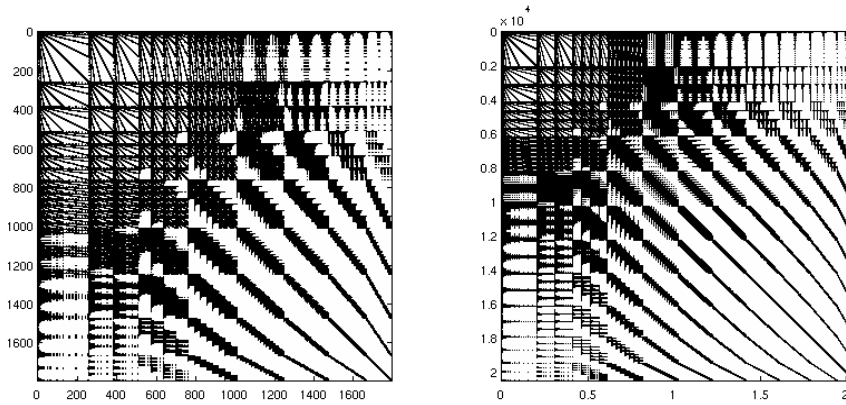


FIGURE 5. Sparsity pattern of the compressed matrices. Left: Level $J = 7$. Right: Level $J = 10$.

- [2] Y. Achdou and N. Tchou. Variational analysis for the Black and Scholes equation with stochastic volatility. *M2AN Math. Model. Numer. Anal.*, 36(3):373–395, 2002.
- [3] O.E. Barndorff-Nielsen and N. Shephard. Non-gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society*, 63:167–241, 2001.
- [4] J. Bertoin. Lévy processes. Cambridge University Press, Cambridge, 1996.
- [5] G. Beylkin, R. Coifman, and V. Rokhlin. The fast wavelet transform and numerical algorithms. *Comm. Pure and Appl. Math.*, 44:141–183, 1991.
- [6] 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.
- [7] 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.
- [8] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods for elliptic operator equations: convergence rates. *Math. Comp.*, 70:27–75 (electronic), 2001.
- [9] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods. II. Beyond the elliptic case. *Found. Comput. Math.*, 2:203–245, 2002.
- [10] A. Cohen, I. Daubechies, and J.-C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 45(5):485–560, 1992.
- [11] R. Cont and E. Voltchkova. A finite difference scheme for option pricing in jump diffusion and exponential Lévy models. *SIAM J. Numer. Anal.*, 43(4):1596–1626, 2005.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] W. Dahmen and R. Schneider. Wavelets with complementary boundary conditions—function spaces on the cube. *Results Math.*, 34(3-4):255–293, 1998.
- [17] F. Delbaen, P. Grandits, T. Rheinländer, D. Samperi, M. Schweizer, and C. Stricker. Exponential hedging and entropic penalties. *Mathematical Finance*, 12:99–123, 2002.
- [18] F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Math. Ann.*, 300:463–520, 1994.
- [19] F. Delbaen and W. Schachermayer. The variance-optimal martingale measure for continuous processes. *Bernoulli*, 2:81–105, 1996.
- [20] M. Demuth and J. van Casteren. *Stochastic Spectral Theory for Selfadjoint Feller Operators*. Birkhäuser Verlag, Basel, 2000.
- [21] R. DeVore. Nonlinear approximation. In *Acta numerica, 1998*, volume 7 of *Acta Numer.*, pages 51–150. Cambridge Univ. Press, Cambridge, 1998.
- [22] A. Ern and J.-L. Guermond. *Theory and practice of Finite Elements*. Springer Verlag, New York, 2004.
- [23] 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.
- [24] T. Gantumur, H. Harbrecht, and R. Stevenson. An optimal adaptive wavelet method without coarsening of the iterands. *Math. Comp.*, 76:615–629 (electronic), 2007.

- [25] T. Gantumur and R. Stevenson. Computation of differential operators in wavelet coordinates. *Math. Comp.*, 75:697–709 (electronic), 2006.
- [26] 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.
- [27] M. Griebel, P. Oswald, and T. Schiekofer. Sparse grids for boundary integral equations. *Numer. Math.*, 83(2):279–312, 1999.
- [28] H. Harbrecht and R. Schneider. Biorthogonal wavelet bases for the boundary element method. *Math. Nachr.*, 269/270:167–188, 2004.
- [29] H. Harbrecht and R. Schneider. Wavelet Galerkin schemes for boundary integral equations—implementation and quadrature. *SIAM J. Sci. Comput.*, 27:1347–1370 (electronic), 2006.
- [30] N. Hilber, A.-M. Matache, and C. Schwab. Sparse wavelet methods for option pricing under stochastic volatility. *Journal of Computational Finance*, 8:1–42, 2005.
- [31] N. Hilber, N. Reich, C. Schwab, and C. Winter. Numerical methods for Lévy processes. *Finance and Stochastics*. Special Issue on Computational Methods in Finance. To appear.
- [32] N. Hilber, N. Reich, and C. Winter. Wavelet methods. In R. Cont, editor, *Encyclopedia of Quantitative Finance*. John Wiley & Sons Ltd., Chichester. To appear.
- [33] W. Hoh. *Pseudo Differential Operators generating Markov Processes*. Habilitationsschrift, University of Bielefeld, 1998.
- [34] L. Hörmander. *Linear partial differential operators*, volume 116 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1963.
- [35] L. Hörmander. *The analysis of linear partial differential operators. III*, volume 274 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1985. Pseudodifferential operators.
- [36] N. Jacob. *Pseudo Differential Operators and Markov Processes, Vol. 2: Generators and their potential theory*. Imperial College Press, London, 2002.
- [37] N. Jacob. *Pseudo Differential Operators and Markov Processes, Vol. 3: Markov processes and applications*. Imperial College Press, London, 2005.
- [38] S. Knapek and F. Koster. Integral operators on sparse grids. *SIAM J. Numer. Anal.*, 39(5):1794–1809 (electronic), 2001/02.
- [39] F. Liu, N. Reich, and A. Zhou. *Two-scale Finite Element Discretizations for Infinitesimal Generators of Jump Processes in Finance*. 2008. Submitted. Research report 2008-23 Seminar for Applied Mathematics, ETH Zürich.
- [40] A.-M. Matache, P.A. Nitsche, and C. Schwab. Wavelet Galerkin pricing of American contracts on Lévy driven assets. *Quantitative Finance*, 5:403–424, 2005.
- [41] A.-M. Matache, T. von Petersdorff, and C. Schwab. Fast deterministic pricing of options on Lévy driven assets. *Mathematical Modelling and Numerical Analysis*, 38:37–71, 2004.
- [42] H. Nguyen and R. Stevenson. Finite Elements on manifolds. *IMA J. Numer. Math.*, 23:149–173, 2003.
- [43] P. Oswald. On N -term approximation by Haar functions in H^s -norms. In S.M. Nikol'skij, B.S. Kashin, and A.D. Izaak, editors, *Metric Function Theory and Related Topics in Analysis*, pages 137–163. AFC, Moscow, 1999.
- [44] N. Reich. *Anisotropic Operator Symbols Arising From Multivariate Jump Processes*. Submitted. Research report 2008-25, Seminar for Applied Mathematics, ETH Zürich, 2008.
- [45] 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.
- [46] 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>.
- [47] N. Reich. *Wavelet Compression of Integral Operators on Sparse Tensor Spaces – Construction, Consistency and Asymptotically Optimal Complexity*. Submitted. Research report 2008-24, Seminar for Applied Mathematics, ETH Zürich, 2008.
- [48] N. Reich, C. Schwab, and C. Winter. *On Kolmogorov Equations for Anisotropic Multivariate Lévy Processes*. submitted. Research report 2008-3, Seminar for Applied Mathematics, ETH Zürich, 2008.
- [49] K.-I. Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge, 1999.
- [50] R. Schneider. *Multiskalen- und Wavelet-Matrixkompression: Analysisbasierte Methoden zur Lösung großer vollbesetzter Gleichungssysteme*. B.G. Teubner, Stuttgart, 1998.
- [51] C. Schwab and R. Stevenson. Adaptive wavelet algorithms for elliptic PDE's on product domains. *Math. Comp.*, 77(261):71–92 (electronic), 2008.
- [52] R.E. Showalter. *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*. American Mathematical Society, Rhode Island, 1997.
- [53] E.M. Stein. *Harmonic Analysis*. Princeton University Press, Princeton, 1993.
- [54] R. Stevenson. On the compressibility of operators in wavelet coordinates. *SIAM J. Math. Anal.*, 35(5):1110–1132 (electronic), 2004.
- [55] M.E. Taylor. *Pseudodifferential operators*. Princeton University Press, Princeton, 1981.
- [56] H. Triebel. *Interpolation theory, function spaces, differential operators*. Johann Ambrosius Barth Verlag, Heidelberg, second edition, 1995.
- [57] 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.
- [58] T. von Petersdorff and C. Schwab. Wavelet discretizations of parabolic integrodifferential equations. *SIAM J. Numer. Anal.*, 41(1):159–180 (electronic), 2003.

- [59] T. von Petersdorff and C. Schwab. Numerical solution of parabolic equations in high dimensions. *M2AN Math. Model. Numer. Anal.*, 38(1):93–127, 2004.
- [60] T. von Petersdorff, C. Schwab, and R. Schneider. Multiwavelets for second-kind integral equations. *SIAM J. Numer. Anal.*, 34(6):2212–2227, 1997.
- [61] C. Winter. *Wavelet Galerkin schemes for option pricing in multidimensional Lévy models* (tentative title). PhD Thesis, ETH Zürich, in preparation.