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

Approximation of integral operators using product-convolution expansions

Paul Escande · Pierre Weiss

the date of receipt and acceptance should be inserted later

Abstract We consider a class of linear integral operators with impulse responses varying regularly in time or space. These operators appear in a large number of applications ranging from signal/image processing to biology. Evaluating their action on functions is a computationally intensive problem necessary for many practical problems. We analyze a technique called product-convolution expansion: the operator is locally approximated by a convolution, allowing to design fast numerical algorithms based on the fast Fourier transform. We design various types of expansions, provide their explicit rates of approximation and their complexity depending on the time varying impulse response smoothness. This analysis suggests novel wavelet based implementations of the method with numerous assets such as optimal approximation rates, low complexity and storage requirements as well as adaptivity to the kernels regularity. The proposed methods are an alternative to more standard procedures such as panel clustering, cross approximations, wavelet expansions or hierarchical matrices.

Keywords Integral operators, wavelet, spline, structured low rank decomposition, numerical complexity, approximation rate, fast Fourier transform.

AMS Classification 47A58, 41A46, 41A35, 65D07, 65T60, 65T50, 65R32, 94A12

Paul Escande
Département d'Ingénierie des Systèmes Complexes (DISC), Institut Supérieur de l'Aéronautique et de l'Espace (ISAE), Toulouse, France E-mail: paul.escande@gmail.com

Pierre Weiss
Institut des Technologies Avancées en Sciences du Vivant, ITAV-USR3505 and Institut de Mathématiques de Toulouse, IMT-UMR5219, CNRS and Université de Toulouse, Toulouse, France E-mail: pierre.armand.weiss@gmail.com

Acknowledgements We began investigating a much narrower version of the problem in a preliminary version of [20]. An anonymous reviewer however suggested that it would be more interesting to make a general analysis and we therefore discarded the aspects related to product-convolution expansions from [20]. We thank the reviewer for motivating us to initiate this research. We also thank Sandrine Anthoine, Jérémie Bigot, Caroline Chaux, Jérôme Fehrenbach, Hans Feichtinger, Clothilde Mélot, Anh-Tuan Nguyen and Bruno Torrésani for fruitful discussions on related matters. The authors are particularly grateful to Hans Feichtinger for his numerous feedbacks on a preliminary version of this paper and for suggesting the name product-convolution.

1 Introduction

We are interested in the compact representation and fast evaluation of a class of space or time varying linear integral operators with regular variations. Such operators appear in a large number of applications ranging from wireless communications [37,28] to seismic data analysis [23], biology [22] and image processing [38].

In all these applications, a key numerical problem is to efficiently evaluate the action of the operator and its adjoint on given functions. This is necessary - for instance - to design fast inverse problems solvers. The main objective of this paper is to analyze the complexity of a set of approximation techniques coined *product-convolution expansions*.

We are interested in bounded linear integral operators $H : L^2(\Omega) \to L^2(\Omega)$ defined from a kernel K by:

$$Hu(x) = \int_\Omega K(x,y)u(y)\,dy. \tag{1}$$

for all $u \in L^2(\Omega)$, where $\Omega = \mathbb{R} \setminus \mathbb{Z}$ is the one dimensional torus. Extensions to bounded and higher dimensional domains will be mentioned at the end of the pa-

per. Evaluating integrals of type (1) is a major challenge in numerical analysis and many methods have been developed in the literature. Nearly all methods share the same basic principle: decompose the operator kernel as a sum of low rank matrices with a multi-scale structure. This is the case in panel clustering methods [24], hierarchical matrices [6], cross approximations [35] or wavelet expansions [3,13,14]. The method proposed in this paper basically shares the same idea, except that the time varying impulse response T of the operator is decomposed instead of the kernel K. The *time varying impulse response* (TVIR) T of H is defined by:

$$T(x,y) = K(x+y,y), \quad \forall (x,y) \in \Omega \times \Omega. \tag{2}$$

The TVIR representation of H allows formalizing the notion of regularly varying integral operator: the functions $T(x,\cdot)$ should be "smooth" for all $x \in \Omega$. Intuitively, the smoothness assumption means that two neighboring impulse responses should only differ slightly. Under this assumption, it is tempting to approximate H locally by a convolution. Two different approaches have been proposed in the literature to achieve this. The first one is called *convolution-product expansion of order m* and consists of approximating H by an operator H_m of type:

$$H_m u = \sum_{k=1}^{m} w_k \odot (h_k \star u), \tag{3}$$

where h_k and w_k are real-valued functions defined on Ω, $\odot$ denotes the standard multiplication for functions and the Hadamard product for vectors, and $\star$ denotes the convolution operator. The second one, called *product-convolution expansion of order m*, is at the core of this paper and consists of using an expansion of type:

$$H_m u = \sum_{k=1}^{m} h_k \star (w_k \odot u). \tag{4}$$

Function w_k is usually chosen as a windowing function localized in space, while h_k is a kernel describing the operator on the support of w_k. These two types of approximations have been used for a long time in the field of imaging (and to a lesser extent mobile communications and biology) and progressively became more and more refined [40,34,21,22,28,1,27,33,17]. In particular, the recent work [17] provides a nice overview of existing choices for the functions h_k and w_k as well as new ideas leading to significant improvements. Many different names have been used in the literature to describe expansions of type (3) and (4) depending on the communities: sectional methods, overlap-add and overlap-save methods, piecewise convolutions, anisoplanatic convolutions, parallel product-convolution, filter flow, windowed-convolutions... The term product-convolution comes from the field of mathematics [7] [1]. We believe that it precisely describes the set of expansions of type (3) and therefore chose this naming. It was already used in the field of imaging by [1]. Now that product-convolution expansions have been described, natural questions arise:

i) How to choose the functions h_k and w_k?
ii) What is the numerical complexity of evaluating products of type $H_m u$?
iii) What is the resulting approximation error $\|H_m - H\|$, where $\|\cdot\|$ is a norm over the space of operators?
iv) How many operations are needed in order to obtain an approximation H_m such that $\|H_m - H\| \leq \epsilon$?

Elements i) and ii) have been studied thoroughly and improved over the years in the mentioned papers. The main questions addressed herein are points iii) and iv). To the best of our knowledge, they have been ignored until now. They are however necessary in order to evaluate the theoretical performance of different product-convolution expansions and to compare their respective advantages precisely.

The main outcome of this paper is the following: under smoothness assumptions of type $T(x,\cdot) \in H^s(\Omega)$ for all $x \in \Omega$ (the Hilbert space of functions in $L^2(\Omega)$ with s derivatives in $L^2(\Omega)$), most methods proposed in the literature - if implemented correctly - ensure a decay of type $\|H_m - H\|_{HS} = O(m^{-s})$, where $\|\cdot\|_{HS}$ is the Hilbert-Schmidt norm. Moreover, this bound cannot be improved uniformly on the considered smoothness class. By adding a support condition of type $\text{supp}(T(x,\cdot)) \subseteq [-\kappa/2, \kappa/2]$, the bound becomes $\|H_m - H\|_{HS} = O(\sqrt{\kappa} m^{-s})$. More importantly, bounded supports allow reducing the computational burden. After discretization on n time points, we show that the number of operations required to satisfy $\|H_m - H\|_{HS} \leq \epsilon$ vary from $O\left(\kappa^{\frac{1}{2s}} n \log_2(n) \epsilon^{-1/s}\right)$ to $O\left(\kappa^{\frac{2s+1}{2s}} n \log_2(\kappa n) \epsilon^{-1/s}\right)$ depending on the choices of w_k and h_k. We also show that the compressed operator representations of Meyer [32] can be used under additional regularity assumptions.

An important difference of product-convolution expansions compared to most methods in the literature [24,6,35,3,20] is that they are insensitive to the smoothness of $T(\cdot, y)$. The smoothness in the x direction is a useful property to control the discretization error, but not the approximation rate. The proposed methodology might therefore be particularly competitive in applications with irregular impulse responses.

[1] With the terminology of [7], the name product-convolution would have been convolution-product and vice-versa. The name product-convolution seems more appropriate to describe a product followed by a convolution.

The paper is organized as follows. In section 2, we describe the notation and introduce a few standard results of approximation theory. In section 3, we precisely describe the class of operators studied in this paper, show how to discretize them and provide the numerical complexity of evaluating product-convolution expansions of type (4). Sections 4 and 5 contain the full approximation analysis for two different kinds of approaches called linear or adaptive methods. Section 6 contains a summary and a few additional comments.

2 Notation

Let a and b denote functions depending on some parameters. The relationship $a \asymp b$ means that a and b are equivalent, i.e. that there exists $0 < c_1 \leq c_2$ such that $c_1 a \leq b \leq c_2 a$. Constants appearing in inequalities will be denoted by C and may vary at each occurrence. If a dependence on a parameter exists (e.g. ϵ), we will use the notation $C(\epsilon)$.

In most of the paper, we work on the unit circle $\Omega = \mathbb{R}\backslash\mathbb{Z}$ sometimes identified with the interval $\left[-\frac{1}{2}, \frac{1}{2}\right]$. This choice is driven by simplicity of exposition and the results can be extended to bounded domains such as $\Omega = [0,1]^d$ (see section 6.2). Let $L^2(\Omega)$ denote the space of square integrable functions on Ω. The Sobolev space $H^s(\Omega)$ is defined as the set of functions in $L^2(\Omega)$ with weak derivatives up to order s in $L^2(\Omega)$. The k-th weak derivative of $u \in H^s(\Omega)$ is denoted $u^{(s)}$. The norm and semi-norm of $u \in H^s(\Omega)$ are defined by:

$$\|u\|_{H^s(\Omega)} = \sum_{k=0}^{s} \|u^{(k)}\|_{L^2(\Omega)} \quad \text{and} \quad |u|_{H^s(\Omega)} = \|u^{(s)}\|_{L^2(\Omega)}. \tag{5}$$

The sequence of functions $(e_k)_{k\in\mathbb{Z}}$ where $e_k : x \mapsto \exp(-2i\pi kx)$ is an orthonormal basis of $L^2(\Omega)$ (see e.g. [29]).

Definition 1 Let $u \in L^2(\Omega)$ and $e_k : x \mapsto \exp(-2i\pi kx)$ denote the k-th Fourier atom. The Fourier series coefficients $\hat{u}[k]$ of u are defined for all $k \in \mathbb{Z}$ by:

$$\hat{u}[k] = \int_\Omega u(x) e_k(x)\, dx. \tag{6}$$

The space $H^s(\Omega)$ can be characterized through Fourier series.

Lemma 1 (Fourier characterization of Sobolev norms)

$$\|u\|^2_{H^s(\Omega)} \asymp \sum_{k\in\mathbb{Z}} |\hat{u}[k]|^2 (1 + |k|^2)^s. \tag{7}$$

Definition 2 (B-spline of order α) Let $\alpha \in \mathbb{N}$ and $m \geq \alpha + 2$ be two integers. The B-spline of order 0 is defined by

$$B_{0,m} = \mathbb{1}_{[-1/(2m),1/(2m)]}. \tag{8}$$

The B-spline of order $\alpha \in \mathbb{N}^*$ is defined by recurrence by:

$$B_{\alpha,m} = mB_{0,m} \star B_{\alpha-1,m} = m^\alpha \underbrace{B_{0,m} \star \ldots \star B_{0,m}}_{\alpha \text{ times}}. \tag{9}$$

The set of cardinal B-splines of order α is denoted $\mathcal{B}_{\alpha,m}$ and defined by:

$$\mathcal{B}_{\alpha,m} = \left\{ f(\cdot) = \sum_{k=0}^{m-1} c_k B_{\alpha,m}(\cdot - k/m), \quad c_k \in \mathbb{R},\ 0 \leq k \leq m-1 \right\}. \tag{10}$$

In this work, we use the Daubechies wavelet basis for $L^2(\mathbb{R})$ [15]. We let ϕ and ψ denote the scaling and mother wavelets and assume that the mother wavelet ψ has α vanishing moments, i.e.

$$\forall 0 \leq m < \alpha, \quad \int_{[0,1]} t^m \psi(t) dt = 0. \tag{11}$$

Daubechies wavelets satisfy $\text{supp}(\psi) = [-\alpha + 1, \alpha]$, see [31, Theorem 7.9, p. 294]. Translated and dilated versions of the wavelets are defined, for all $j > 0$ by

$$\psi_{j,l}(x) = 2^{j/2} \psi\left(2^j x - l\right). \tag{12}$$

The set of functions $(\psi_{j,l})_{j\in\mathbb{N},l\in Z}$, is an orthonormal basis of $L^2(\mathbb{R})$ with the convention $\psi_{0,l} = \phi(x - l)$. There are different ways to construct a wavelet basis on the interval $[-1/2, 1/2]$ from a wavelet basis on $L^2(\mathbb{R})$. Here, we use boundary wavelets defined in [12]. We refer to [16,31] for more details on the construction of wavelet bases. This yields an orthonormal basis $(\psi_\lambda)_{\lambda\in\Lambda}$ of $L^2(\Omega)$, where

$$\Lambda = \left\{(j,l), j \in \mathbb{N}, 0 \leq l \leq 2^j\right\}. \tag{13}$$

We let $I_\lambda = \text{supp}(\psi_\lambda)$ and for $\lambda \in \Lambda$, we use the notation $|\lambda| = j$.

Let u and v be two functions in $L^2(\Omega)$, the notation $u \otimes v$ will be used both to indicate the function $w \in L^2(\Omega \times \Omega)$ defined by

$$w(x,y) = (u \otimes v)(x,y) = u(x)v(y), \tag{14}$$

or the Hilbert-Schmidt operator $w : L^2(\Omega) \to L^2(\Omega)$ defined for all $f \in L^2(\Omega)$ by:

$$w(f) = (u \otimes v)f = \langle u, f\rangle v. \tag{15}$$

The meaning can be inferred depending on the context. Let $H : L^2(\Omega) \to L^2(\Omega)$ denote a linear integral operators. Its kernel will always be denoted K and its time varying impulse response T. The linear integral operator with kernel T will be denoted J.

The following result is an extension of the singular value decomposition to operators.

Lemma 2 (Schmidt decomposition [36, Theorem 2.2] or [26, Theorem 1 p. 215]) *Let $H : L^2(\Omega) \to L^2(\Omega)$ denote a compact operator. There exists two finite or countable orthonormal systems $\{e_1, \dots\}$, $\{f_1, \dots\}$ of $L^2(\Omega)$ and a finite or infinite sequence $\sigma_1 \geq \sigma_2 \geq \dots$ of positive numbers (tending to zero if it is infinite), such that H can be decomposed as:*

$$H = \sum_{k \geq 1} \sigma_k \cdot e_k \otimes f_k. \tag{16}$$

A function $u \in L^2(\Omega)$ is denoted in regular font whereas its discretized version $\boldsymbol{u} \in \mathbb{R}^n$ is denoted in bold font. The value of function u at $x \in \Omega$ is denoted $u(x)$, while the i-th coefficient of vector $\boldsymbol{u} \in \mathbb{R}^n$ is denoted $\boldsymbol{u}[i]$. Similarly, an operator $H : L^2(\Omega) \to L^2(\Omega)$ is denoted in upper-case regular font whereas its discretized version $\boldsymbol{H} \in \mathbb{R}^{n \times n}$ is denoted in upper-case bold font.

3 Preliminary facts

In this section, we gather a few basic results necessary to derive approximation results.

3.1 Assumptions on the operator and examples

All the results stated in this paper rely on the assumption that the TVIR T of H is a sufficiently simple function. By simple, we mean that i) the functions $T(x, \cdot)$ are smooth for all $x \in \Omega$ and ii) the impulse responses $T(\cdot, y)$ have a bounded support or a fast decay for all $y \in \Omega$.

There are numerous ways to capture the regularity of a function. In this paper, we assume that $T(x, \cdot)$ lives in the Hilbert spaces $H^s(\Omega)$ for all $x \in \Omega$. This hypothesis is deliberately simple to clarify the proofs and the main ideas.

Definition 3 (Class $\mathcal{T}^s$) We let $\mathcal{T}^s$ denote the class of functions $T : \Omega \times \Omega \to \mathbb{R}$ satisfying the smoothness condition: $T(x, \cdot) \in H^s(\Omega)$, $\forall x \in \Omega$ and $\|T(x, \cdot)\|_{H^s(\Omega)}$ is uniformly bounded in x, i.e:

$$\sup_{x \in \Omega} \|T(x, \cdot)\|_{H^s(\Omega)} \leq C < +\infty. \tag{17}$$

Note that if $T \in \mathcal{T}^s$, then H is a Hilbert-Schmidt operator since:

$$\|H\|_{HS}^2 = \int_\Omega \int_\Omega K(x, y)^2 \, dx \, dy \tag{18}$$

$$= \int_\Omega \int_\Omega T(x, y)^2 \, dx \, dy \tag{19}$$

$$= \int_\Omega \|T(x, \cdot)\|_{L^2(\Omega)}^2 \, dx < +\infty. \tag{20}$$

We will often use the following regularity assumption.

Assumption 1 *The TVIR T of H belongs to $\mathcal{T}^s$.*

In many applications, the impulse responses have a bounded support, or at least a fast spatial decay allowing to neglect the tails. This property will be exploited to design faster algorithms. This hypothesis can be expressed by the following assumption.

Assumption 2 $T(x, y) = 0, \forall |x| > \kappa/2$.

3.2 Examples

We provide 3 examples of kernels that may appear in applications. Figure 1 shows each kernel as a 2D image, the associated TVIR and the spectrum of the operator J (the linear integral operator with kernel T) computed with an SVD.

Example 1 A typical kernel that motivates our study is defined by:

$$K(x, y) = \frac{1}{\sqrt{2\pi}\sigma(y)} \exp\left(-\frac{(x-y)^2}{2\sigma^2(y)}\right). \tag{21}$$

The impulse responses $K(\cdot, y)$ are Gaussian for all $y \in \Omega$. Their variance $\sigma(y) > 0$ varies depending on the position y. The TVIR of K is defined by:

$$T(x, y) = \frac{1}{\sqrt{2\pi}\sigma(y)} \exp\left(-\frac{x^2}{2\sigma^2(y)}\right). \tag{22}$$

The impulse responses $T(\cdot, y)$ are not compactly supported, therefore, $\kappa = 1$ in Assumption 2. However, it is possible to truncate them by setting $\kappa = 3 \sup_{y \in \Omega} \sigma(y)$ for instance. This kernel satisfies Assumption 1 only if $\sigma : \Omega \to \mathbb{R}$ is sufficiently smooth. In Figure 1, left column, we set $\sigma(y) = 0.08 + 0.02 \cos(2\pi y)$.

Example 2 The second example is given by:

$$T(x, y) = \frac{2}{\sigma(y)} \max(1 - 2\sigma(y)|x|, 0). \tag{23}$$

The impulse responses $T(\cdot, y)$ are cardinal B-splines of degree 1 and width $\sigma(y) > 0$. They are compactly supported with $\kappa = \sup_{y\in\Omega} \sigma(y)$. This kernel satisfies Assumption 2 only if $\sigma : \Omega \to \mathbb{R}$ is sufficiently smooth. In Figure 1, central column, we set $\sigma(y) = 0.1+0.3(1-|y|)$. This kernel satisfies Assumption 1 with $s = 1$.

Example 3 The last example is a discontinuous TVIR. We set:

$$T(x,y) = g_{\sigma_1}(x)\mathbb{1}_{[-1/4,1/4]}(y) + g_{\sigma_2}(x)(1-\mathbb{1}_{[-1/4,1/4]}(y)), \tag{24}$$

where $g_\sigma(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{\sigma^2}\right)$. This corresponds to the last column in Figure 1, with $\sigma_1 = 0.05$ and $\sigma_2 = 0.1$. For this kernel, both Assumptions 1 and 2 are violated. Notice however that T is the sum of two tensor products and can therefore be represented using only four 1D functions. The spectrum of J should have only 2 non zero elements. This is verified in Figure 1i, where the spectrum is 0 (up to numerical errors of order 10^{-13}), except for the first two elements. .

3.3 Product-convolution expansions as low-rank approximations

Though similar in spirit, convolution-product (3) and product-convolution (4) expansions have a quite different interpretation captured by the following lemma.

Lemma 3 *The TVIR T_m of the convolution-product expansion in* (3) *is given by:*

$$T_m(x,y) = \sum_{k=1}^{m} h_k(x) w_k(x+y). \tag{25}$$

The TVIR T_m of the product-convolution expansion in (4) *is given by:*

$$T_m(x,y) = \sum_{k=1}^{m} h_k(x) w_k(y). \tag{26}$$

Proof We only prove (26) since the proof of (25) relies on the same arguments. By definition:

$$(H_m u)(x) = \left(\sum_{k=1}^{m} h_k \star (w_k \odot u)\right)(x) \tag{27}$$

$$= \int_\Omega \sum_{k=1}^{m} h_k(x-y) w_k(y) u(y)\, dy. \tag{28}$$

By identification, this yields:

$$K_m(x,y) = \sum_{k=1}^{m} h_k(x-y) w_k(y), \tag{29}$$

so that

$$T_m(x,y) = \sum_{k=1}^{m} h_k(x) w_k(y). \tag{30}$$

As can be seen in (26), product-convolution expansions consist of finding low-rank approximations of the TVIR. This interpretation was already proposed in [17] for instance and is the key observation to derive the forthcoming results. The expansion (25) does not share this simple interpretation and we do not investigate it further in this paper.

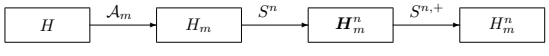
3.4 Discretization

In order to implement a product-convolution expansion of type 4, the problem first needs to be discretized. We address this problem with a Galerkin formalism. Let $(\varphi_1, \ldots, \varphi_n)$ be a basis of a finite dimensional subspace V^n of $L^2(\Omega)$. Given an operator $H : L^2(\Omega) \to L^2(\Omega)$, we can construct a matrix $\boldsymbol{H}^n \in \mathbb{R}^{n\times n}$ defined for all $1 \le i,j \le n$ by $\boldsymbol{H}^n[i,j] = \langle H\varphi_j, \varphi_i\rangle$. Let $S^n : H \mapsto \boldsymbol{H}^n$ denote the discretization operator. From a matrix $\boldsymbol{H}^n$, an operator H^n can be reconstructed using, for instance, the pseudo-inverse $S^{n,+}$ of S^n. We let $H^n = S^{n,+}(\boldsymbol{H}^n)$. For instance, if $(\varphi_1, \ldots, \varphi_n)$ is an *orthonormal* basis of V^n, the operator H^n is given by:

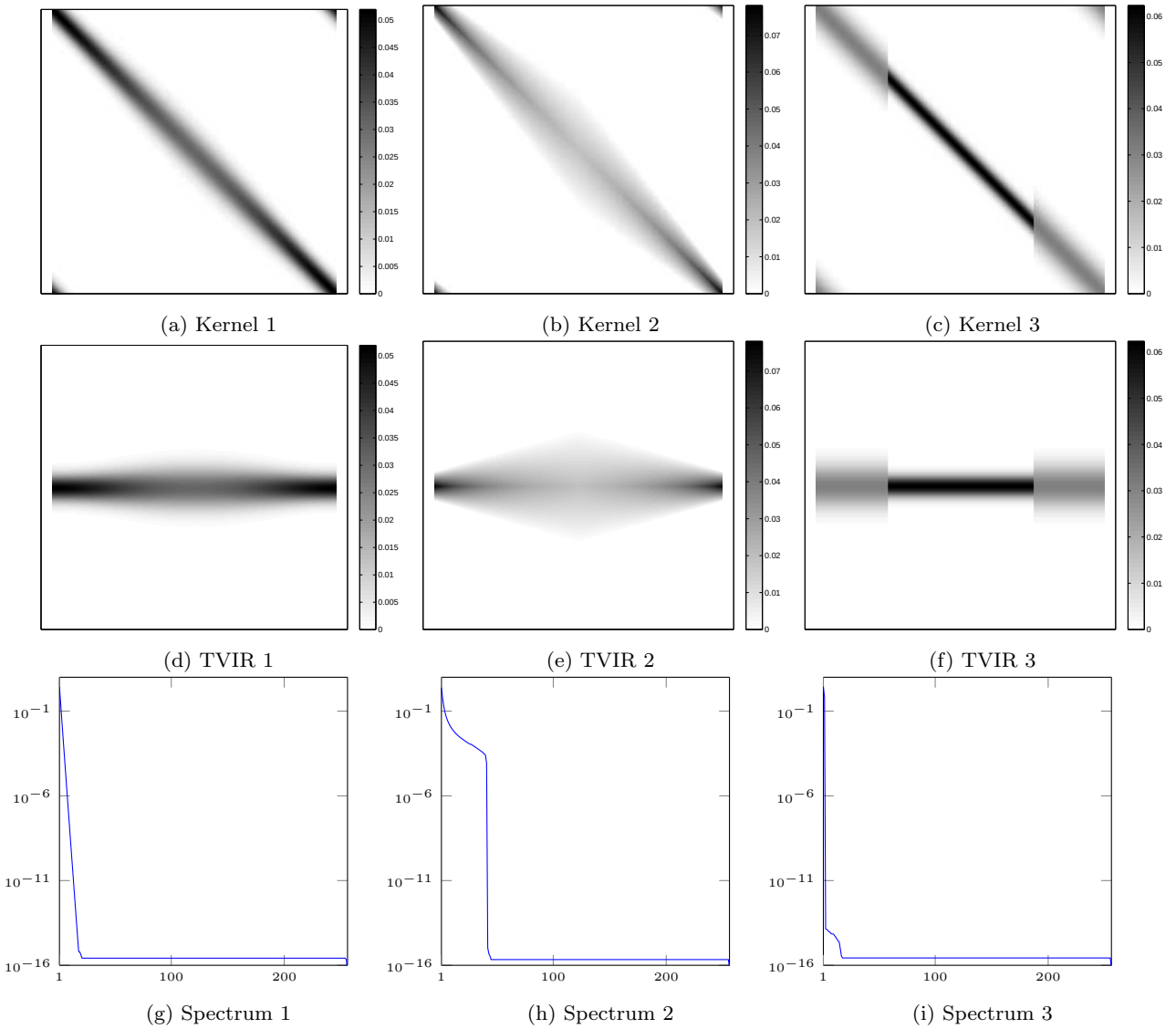
$$H^n = S^{n,+}(\boldsymbol{H}^n) = \sum_{1\le i,j\le n} \boldsymbol{H}^n[i,j] \varphi_i \otimes \varphi_j. \tag{31}$$

This paper is dedicated to analyzing methods denoted $\mathcal{A}_m$ that provide an approximation $H_m = \mathcal{A}_m(H)$ of type (4), given an input operator H. Our analysis provides guarantees on the distance $\|H - H_m\|_{HS}$ depending on m and the regularity properties of the input operator H, for different methods. Depending on the context, two different approaches can be used to implement $\mathcal{A}_m$.

- Compute the matrix $\boldsymbol{H}^n_m = S^n(H_m)$ using numerical integration procedures. Then create an operator $H^n_m = S^{n,+}(\boldsymbol{H}^n_m)$. This approach suffers from two defects. First, it is only possible by assuming that the kernel of H is given analytically. Moreover it might be computationally intractable. It is illustrated below.



- In many applications, the operator H is not given explicitly. Instead, we only have access to its discretization $\boldsymbol{H}^n$. Then it is possible to construct a discrete approximation algorithm $\boldsymbol{\mathcal{A}}_m$ yielding a discrete approximation $\boldsymbol{H}^n_m = \boldsymbol{\mathcal{A}}_m(\boldsymbol{H}^n)$. This matrix



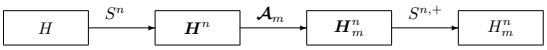
(a) Kernel 1 (b) Kernel 2 (c) Kernel 3

(d) TVIR 1 (e) TVIR 2 (f) TVIR 3

(g) Spectrum 1 (h) Spectrum 2 (i) Spectrum 3

Fig. 1: Different kernels K, the associated TVIR T and the spectrum of the operator J. Left column corresponds to Example 1. Central column corresponds to Example 2. Right column corresponds to Example 3.

can then be mapped back to the continuous world using the pseudo-inverse: $H_m^n = S^{n,+}(\boldsymbol{H}_m^n)$. This is illustrated below. In this paper, we will analyze the construction complexity of $\boldsymbol{H}_m^n$ using this second approach.



Ideally, we would like to provide guarantees on $\|H - H_m^n\|_{HS}$ depending on m and n. In the first approach, this is possible by using the following inequality:

$$\|H - H_m^n\|_{HS} \leq \underbrace{\|H - H_m\|_{HS}}_{\epsilon_a(m)} + \underbrace{\|H_m - H_m^n\|_{HS}}_{\epsilon_d(n)}, \quad (32)$$

where $\epsilon_a(m)$ is the approximation error studied in this paper and $\epsilon_d(n)$ is the discretization error. Under mild regularity assumptions on K, it is possible to obtain results of type $\epsilon_d(n) = O(n^{-\gamma})$, where γ depends on the smoothness of K. For instance, if $K \in H^r(\Omega \times \Omega)$, the error satisfies $\epsilon_d(n) = O(n^{-r/2})$ for many bases including Fourier, wavelets and B-splines [10]. For $K \in BV(\Omega \times \Omega)$, the space of functions with bounded variations, $\epsilon_d(n) = O(n^{-1/4})$, see [31, Theorem 9.3]. As will be seen later, the approximation error $\epsilon_a(m)$ behaves like $O(m^{-s})$. As will be seen later, the proposed approximation technique will be of interest only in the case $m \ll n$, since otherwise, it will require storing too much data. Under this assumption, the discretization error can be considered negligible compared to the approximation error. In all the paper, we assume that $\epsilon_d(n)$ is negligible compared to $\epsilon_a(m)$ without mention.

In the second approach, the error analysis is more complex since there is an additional bias due to the algorithm discretization. This bias is captured by the following inequality:

$$\|H - H_m^n\|_{HS} \leq \underbrace{\|H - H^n\|_{HS}}_{\epsilon_d(n)} + \underbrace{\|H^n - \mathcal{A}_m(H^n)\|_{HS}}_{\epsilon_a(m)} + \underbrace{\|\mathcal{A}_m(H^n) - H_m^n\|_{HS}}_{\epsilon_b(m,n)}. \tag{33}$$

The bias

$$\epsilon_b(m,n) = \|\mathcal{A}_m(S^{n,+}(S^n(H))) - S^{n,+}(\boldsymbol{\mathcal{A}}_m(S^n(H)))\|_{HS} \tag{34}$$

accounts for the difference between using the discrete or continuous approximation algorithm. In this paper, we do not study this bias error and assume that it is negligible compared to the approximation error ϵ_a.

3.5 Implementation and complexity

Let $\boldsymbol{F}_n \in \mathbb{C}^{n\times n}$ denote the discrete inverse Fourier transform and $\boldsymbol{F}_n^*$ denote the discrete Fourier transform. Matrix-vector products $\boldsymbol{F}_n\boldsymbol{u}$ or $\boldsymbol{F}_n^*\boldsymbol{u}$ can be evaluated in $O(n\log_2(n))$ operations using the fast Fourier transform (FFT). The discrete convolution product $\boldsymbol{v} = \boldsymbol{h}\star\boldsymbol{u}$ is defined for all $i \in \mathbb{Z}$ by $\boldsymbol{v}[i] = \sum_{j=1}^n \boldsymbol{u}[i-j]\boldsymbol{h}[j]$, with circular boundary conditions.

Discrete convolution products can be evaluated in $O(n\log_2(n))$ operations by using the following fundamental identity:

$$\boldsymbol{v} = \boldsymbol{F}_n \cdot ((\boldsymbol{F}_n^*\boldsymbol{h}) \odot (\boldsymbol{F}_n^*\boldsymbol{u})). \tag{35}$$

Hence a convolution can be implemented using three FFTs ($O(n\log_2(n))$ operations) and a point-wise multiplication ($O(n)$ operations). This being said, it is straightforward to implement formula (4) with an $O(mn\log_2(n))$ algorithm.

Under the additional assumption that w_k and h_k are supported on bounded intervals, the complexity can be improved. We assume that, after discretization, $\boldsymbol{h}_k$ and $\boldsymbol{w}_k$ are compactly supported, with support length $q_k \leq n$ and $p_k \leq n$ respectively.

Lemma 4 *A matrix-vector product of type* (4) *can be implemented with a complexity that does not exceed*

$$O\left(\sum_{k=1}^m (p_k + q_k)\log_2(\min(p_k, q_k))\right)$$

operations.

Proof A convolution product of type $\boldsymbol{h}_k \star (\boldsymbol{w}_k \odot \boldsymbol{u})$ can be evaluated in $O((p_k+q_k)\log(p_k+q_k))$ operations. Indeed, the support of $\boldsymbol{h}_k \star (\boldsymbol{w}_k \odot \boldsymbol{u})$ has no more than p_k+q_k contiguous non-zeros elements. Using the Stockham sectioning algorithm [39], the complexity can be further decreased to $O((p_k+q_k)\log_2(\min(p_k,q_k)))$ operations. This idea was proposed in [27].

4 Projections on linear subspaces

We now turn to the problem of choosing the functions h_k and w_k in equation (4). The idea studied in this section is to fix a subspace $E_m = \mathrm{span}(e_k, k \in \{1,\dots,m\})$ of $L^2(\Omega)$ and to approximate $T(x,\cdot)$ as:

$$T_m(x,y) = \sum_{k=1}^m c_k(x)e_k(y). \tag{36}$$

For instance, the coefficients c_k can be chosen so that $T_m(x,\cdot)$ is a projection of $T(x,\cdot)$ onto E_m. We propose to analyze three different family of functions e_k: Fourier atoms, wavelets atoms and B-splines. We analyze their complexity and approximation properties as well as their respective advantages.

4.1 Fourier decompositions

It is well known that functions in $H^s(\Omega)$ can be well approximated by linear combination of low-frequency Fourier atoms. This loose statement is captured by the following lemma.

Lemma 5 ([19,18]) *Let $f \in H^s(\Omega)$ and f_m denote its partial Fourier series:*

$$f_m = \sum_{k=-m}^m \hat{f}[k]e_k, \tag{37}$$

where $e_k(y) = \exp(-2i\pi ky)$. Then

$$\|f_m - f\|_{L^2(\Omega)} \leq Cm^{-s}|f|_{H^s(\Omega)}. \tag{38}$$

The so-called Kohn-Nirenberg symbol N of H is defined for all $(x,k) \in \Omega\times\mathbb{Z}$ by

$$N(x,k) = \int_\Omega T(x,y)\exp(-2i\pi ky)\,dy. \tag{39}$$

Illustrations of different Kohn-Nirenberg symbols are provided in Figure 2.

Corollary 1 *Set $e_k(y) = \exp(-2i\pi ky)$ and define T_m by:*

$$T_m(x,y) = \sum_{|k|\leq m} N(x,k)e_k(y). \tag{40}$$

Then, under Assumptions 1 and 2

$$\|H_m - H\|_{HS} \leq C\sqrt{\kappa}m^{-s}. \tag{41}$$

Proof By Lemma 5 and Assumption 1,

$$\|T_m(x,\cdot) - T(x,\cdot)\|_{L^2(\Omega)} \leq Cm^{-s}$$

for some constant C and for all $x \in \Omega$. In addition, by Assumption 2, $\|T_m(x,\cdot) - T(x,\cdot)\|_{L^2(\Omega)} = 0$ for $|x| > \kappa/2$. Therefore:

$$\|H_m - H\|_{HS}^2 = \int_\Omega \int_\Omega (T_m(x,y) - T(x,y))^2 \, dx \, dy \tag{42}$$

$$= \int_\Omega \|T_m(x,\cdot) - T(x,\cdot)\|_{L^2(\Omega)}^2 \, dx \tag{43}$$

$$\leq \kappa C^2 m^{-2s} \, dx \tag{44}$$

As will be seen later, the convergence rate (41) is optimal in the sense that no product-convolution expansion of order m can achieve a better rate under the sole Assumptions 1 and 2.

Corollary 2 *Let $\epsilon > 0$ and set $m = \lceil C\epsilon^{-1/s}\kappa^{1/2s} \rceil$. Under Assumptions 1 and 2, H_m satisfies $\|H - H_m\|_{HS} \leq \epsilon$ and products with H_m and H_m^* can be evaluated with no more than $O(\kappa^{1/2s} n \log n \epsilon^{-1/s})$ operations.*

Proof Since Fourier atoms are not localized in the time domain, the modulation functions $\boldsymbol{w}_k$ are supported on intervals of size $p = n$. The complexity of computing a matrix vector product is therefore $O(mn\log(n))$ operations by Lemma 4.

Finally, let us mention that computing the discrete Kohn-Nirenberg symbol $\boldsymbol{N}$ costs $O(\kappa n^2 \log_2(n))$ operations (κn discrete Fourier transforms of size n). The storage cost of this Fourier representation is $O(m\kappa n)$ since one has to store κn coefficients for each of the m vectors $\boldsymbol{h}_k$.

In the next two sections, we show that replacing Fourier atoms by wavelet atoms or B-splines preserves the optimal rate of convergence in $O(\sqrt{\kappa} m^{-s})$, but has the additional advantage of being localized in space, thereby reducing complexity.

4.2 Spline decompositions

B-Splines form a Riesz basis with dual Riesz basis of form [8]:

$$(\tilde{B}_{\alpha,m}(\cdot - k/m))_{0 \leq k \leq m-1}. \tag{45}$$

The projection f_m of any $f \in L^2(\Omega)$ onto $\mathcal{B}_{\alpha,m}$ can be expressed as:

$$f_m = \operatorname*{arg\,min}_{\tilde{f} \in \mathcal{B}_{\alpha,m}} \|\tilde{f} - f\|_2^2 \tag{46}$$

$$= \sum_{k=0}^{m-1} \langle f, \tilde{B}_{\alpha,m}(\cdot - k/m)\rangle B_{\alpha,m}(\cdot - k/m). \tag{47}$$

Theorem 1 ([4, p. 87] or [19, p. 420]) *Let $f \in H^s(\Omega)$ and $\alpha \geq s$, then*

$$\|f - f_m\|_2 \leq C\sqrt{\kappa} m^{-s} \|f\|_{W^{s,2}}. \tag{48}$$

The following result directly follows.

Corollary 3 *Set $\alpha \geq s$. For each $x \in \Omega$, let $(c_k(x))_{0 \leq k \leq m-1}$ be defined as*

$$c_k(x) = \langle T(x,\cdot), \tilde{B}_{\alpha,m}(\cdot - k/m)\rangle. \tag{49}$$

Define T_m by:

$$T_m(x,y) = \sum_{k=0}^{m-1} c_k(x) B_{\alpha,m}(y - k/m). \tag{50}$$

If $\alpha \geq s$, then, under Assumptions 1 and 2,

$$\|H_m - H\|_{HS} \leq C\sqrt{\kappa} m^{-s}. \tag{51}$$

Proof The proof is similar to that of Corollary (1).

Corollary 4 *Let $\epsilon > 0$ and set $m = \lceil C\epsilon^{-1/s}\kappa^{1/2s} \rceil$. Under Assumptions 1 and 2 H_m satisfies $\|H - H_m\|_{HS} \leq \epsilon$ and products with H_m and H_m^* can be evaluated with no more than*

$$O\left(\left(s + \kappa^{1+1/2s}\epsilon^{-1/s}\right) n \log_2(\kappa n)\right) \tag{52}$$

operations. For small ϵ and large n, the complexity behaves like

$$O\left(\kappa^{1+1/2s} n \log_2(\kappa n) \epsilon^{-1/s}\right). \tag{53}$$

Proof In this approximation, m B-splines are used to cover Ω. B-splines have a compact support of size $(\alpha+1)/m$. This property leads to windowing vector $\boldsymbol{w}_k$ with support of size $p = \lceil (\alpha+1)\frac{n}{m} \rceil$. Furthermore the vectors $(\boldsymbol{h}_k)$ have a support of size $q = \kappa n$. Combining these two results with Lemma 4 and Corollary 3 yields the result for the choice $\alpha = s$.

The complexity of computing the vectors $\boldsymbol{c}_k$ is $O(\kappa n^2 \log(n))$ (κn projections with complexity $n\log(n)$, see e.g. [41]).

As can be seen in Corollary (4), B-splines approximations are preferable over Fourier decompositions whenever the support size κ is small.

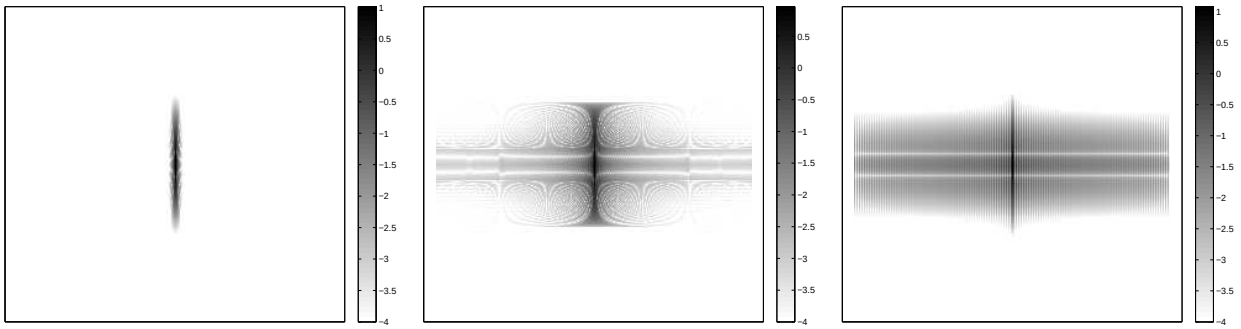
4.3 Wavelet decompositions

Lemma 6 ([31, Theorem 9.5]) *Let $f \in H^s(\Omega)$ and f_m denote its partial wavelet series:*

$$f_m = \sum_{|\mu| \leq \lceil \log_2(m) \rceil} c_\mu \psi_\mu, \tag{54}$$

where ψ is a Daubechies wavelet with $\alpha > s$ vanishing moments and $c_\mu = \langle \psi_\mu, f \rangle$. Then

$$\|f_m - f\|_{L^2(\Omega)} \leq Cm^{-s} |f|_{H^s(\Omega)}. \tag{55}$$



(a) Kernel 1 (b) Kernel 2 (c) Kernel 3

Fig. 2: Kohn-Nirenberg symbols of the kernels given in Examples 1, 2 and 3 in $\log_{10}$ scale. Observe how the decay speed from the center (low frequencies) to the outer parts (high frequencies) changes depending on the TVIR smoothness. Note: the lowest values of the Kohn-Nirenberg symbol have been set to 10^{-4} for visualization purposes.

A direct consequence is the following corollary.

Corollary 5 *Let ψ be a Daubechies wavelet with $\alpha = s+1$ vanishing moments. Define T_m by:*

$$T_m(x,y) = \sum_{|\mu| \leq \lceil \log_2(m) \rceil} c_\mu(x)\psi_\mu(y), \tag{56}$$

where $c_\mu(x) = \langle \psi_\mu, T(x,\cdot)\rangle$. Then, under Assumptions 1 and 2

$$\|H_m - H\|_{HS} \leq C\sqrt{\kappa} m^{-s}. \tag{57}$$

Proof The proof is identical to that of Corollary (1).

Proposition 1 *Let $\epsilon > 0$ and set $m = \lceil C\epsilon^{-1/s}\kappa^{1/2s} \rceil$. Under Assumptions 1 and 2 H_m satisfies $\|H - H_m\|_{HS} \leq \epsilon$ and products with H_m and H_m^* can be evaluated with no more than*

$$O\left(\left(sn\log_2\left(\epsilon^{-1/s}\kappa^{1/2s}\right) + \kappa^{1+1/2s} n \epsilon^{-1/s}\right)\log_2(\kappa n)\right) \tag{58}$$

operations. For small ϵ, the complexity behaves like

$$O\left(\kappa^{1+1/2s} n \log_2(\kappa n)\epsilon^{-1/s}\right). \tag{59}$$

Proof In (56), the windowing vectors $\boldsymbol{w}_k$ are wavelets $\boldsymbol{\psi}_\mu$ of support of size $\min((2s+1)n2^{-|\mu|}, n)$. Therefore each convolution has to be performed on intervals of size $|\boldsymbol{\psi}_\mu| + q + 1$. Since there are 2^j wavelets at scale j, the total number of operations is:

$$\sum_{\mu \,|\, |\mu| < \log_2(m)} (|\boldsymbol{\psi}_\mu| + q + 1)\log_2(\min(|\boldsymbol{\psi}_\mu|, q+1)) \tag{60}$$

$$\leq \sum_{\mu \,|\, |\mu| < \log_2(m)} ((2s+1)n2^{-|\mu|} + \kappa n)\log_2(\kappa n) \tag{61}$$

$$= \sum_{j=0}^{\log_2(m)-1} 2^j \left((2s+1)n2^{-j} + \kappa n\right)\log_2(\kappa n) \tag{62}$$

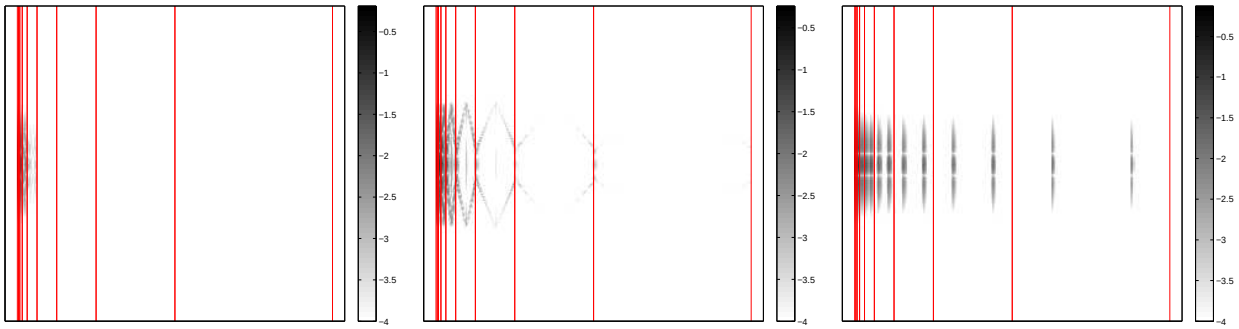
$$= \sum_{j=0}^{\log_2(m)-1} \left((2s+1)n + 2^j\kappa n\right)\log_2(\kappa n) \tag{63}$$

$$\leq ((2s+1)n\log_2(m) + m\kappa n)\log_2(\kappa n) \tag{64}$$

$$= \left((2s+1)n\log_2(\epsilon^{-1/s}\kappa^{1/2s}) + \epsilon^{-1/s}\kappa^{1+1/2s} n\right)\log_2(\kappa n). \tag{65}$$

Computing the vectors $\boldsymbol{c}_\mu$ costs $O(\kappa s n^2)$ operations (κn discrete wavelet transforms of size n). The storage cost of this wavelet representation is $O(m\kappa n)$ since one has to store κn coefficients for each of the m functions $\boldsymbol{h}_k$.

As can be seen from this analysis, wavelet and B-spline approximations roughly have the same complexity over the class $\mathcal{T}^s$. The main advantage of wavelets compared to B-splines with fixed knots is that they are known to characterize much more general function spaces than $H^s(\Omega)$. For instance, if all functions $T(x,\cdot)$ have a single discontinuity at a given $y \in \Omega$, only a few coefficients $c_\mu(x)$ will remain of large amplitude. Wavelets will be able to efficiently encode the discontinuity, while B-splines with fixed knots - which are not localized in nature - will fail to well approximate the TVIR. It is therefore possible to use wavelets in an adaptive way. This effect is visible on Figure 3c: despite



(a) Kernel 1 (b) Kernel 2 (c) Kernel 3

Fig. 3: "Wavelet symbols" of the operators given in Examples 1, 2 and 3 in $\log_{10}$ scale. The red bars indicate separations between scales. Notice that the wavelet coefficients in kernel 1 rapidly decay as scales increase. The decay is slower for kernels 2 and 3 which are less regular. The adaptivity of wavelets can be visualized in Kernel 3: some wavelet coefficients are non zero at large scales, but they are all concentrated around discontinuities. Therefore only a few number of couples (c_μ, ψ_μ) will be necessary to encode the discontinuities. This was not the case with Fourier or B-spline atoms.

discontinuities, only wavelets localized around the discontinuities yield large coefficients. In the next section, we propose two other adaptive methods, in the sense that they are able to automatically adapt to the TVIR regularity.

4.4 Interpolation VS approximation

In all previous results, we constructed the functions w_k and h_k in 4 by projecting $T(x,\cdot)$ onto linear subspaces. This is only possible if the whole TVIR T is available. In very large scale applications, this assumption is unrealistic, since the TVIR contains n^2 coefficients, which cannot even be stored. Instead of assuming a full knowledge of T, some authors (e.g. [34]) assume that the impulse responses $T(\cdot, y)$ are available only at a discrete set of points $y_i = i/m$ for $1 \leq i \leq m$.

In that case, it is possible to *interpolate* the impulse responses instead of approximating them. Given a linear subspace $E_m = \text{span}(e_k, k \in \{1, \dots, m\})$, where the atoms e_k are assumed to be linearly independent, the functions $c_k(x)$ in (36) are chosen by solving the set of linear systems:

$$\sum_{k=1}^{m} c_k(x) e_k(y_i) = T_m(x, y_i) \quad \text{for} \quad 1 \leq i \leq m. \tag{66}$$

In the discrete setting, under Assumption 2, this amounts to solving $\lceil \kappa n \rceil$ linear systems of size $m \times m$. The analysis of such a method requires using very different tools. We refer the interested reader to our recent work [5], where we investigate the rates of convergence with respect to the number of impulse responses, their geometry and the level of noise on the data.

4.5 On Meyer's operator representation

Up to now, we only assumed a regularity of T in the y direction, meaning that the impulse responses vary smoothly in space. In many applications, the impulse responses themselves are smooth. In this section, we show that this additional regularity assumption can be used to further compress the operator. Finding a compact operator representation is a key to treat identification or estimation problems (e.g. blind deblurring in imaging), see e.g. [30].

Since $(\psi_\lambda)_{\lambda \in \Lambda}$ is a Hilbert basis of $L^2(\Omega)$, the set of tensor product functions $(\psi_\lambda \otimes \psi_\mu)_{\lambda \in \Lambda, \mu \in \Lambda}$ is a Hilbert basis of $L^2(\Omega \times \Omega)$. Therefore, any $T \in L^2(\Omega \times \Omega)$ can be expanded as:

$$T(x,y) = \sum_{\lambda \in \Lambda} \sum_{\mu \in \Lambda} c_{\lambda,\mu} \psi_\lambda(x) \psi_\mu(y). \tag{67}$$

The main idea of the construction in this section consists of keeping only the coefficients $c_{\lambda,\mu}$ of large amplitude. A similar idea was proposed in the BCR paper [3][2], except that the kernel K was expanded instead of the TVIR T. Decomposing T was suggested by Beylkin at the end of [2] without a precise analysis.

In this section, we assume that $T \in H^{r,s}(\Omega \times \Omega)$, where

$$H^{r,s}(\Omega \times \Omega) = \{T : \Omega \times \Omega \to \mathbb{R},\ \partial_x^{\alpha_1} \partial_y^{\alpha_2} T \in L^2(\Omega \times \Omega), \\ \forall \alpha_1 \in \{0, \dots, r\}, \forall \alpha_2 \in \{0, \dots, s\}\}. \tag{68}$$

This space arises naturally in applications, where the impulse response regularity r might differ from the reg-

[2] This was also the basic idea in our recent paper [20].

ularity s of their variations. Notice that $H^{2s}(\Omega\times\Omega)\subset H^{s,s}(\Omega\times\Omega)\subset H^s(\Omega)$.

Theorem 2 *Assume that $T\in H^{r,s}(\Omega\times\Omega)$ and satisfies Assumption 2. Assume that ψ has $\max(r,s)+1$ vanishing moments. Let $c_{\lambda,\mu}=\langle T,\psi_\lambda\otimes\psi_\mu\rangle$. Define*

$$H_{m_1,m_2}=\sum_{|\lambda|\leq\log_2(m_1)}\sum_{|\mu|\leq\log_2(m_2)}c_{\lambda,\mu}\psi_\lambda\otimes\psi_\mu. \tag{69}$$

Let $m\in\mathbb{N}$, set $m_1=\lceil m^{s/(r+s)}\rceil$, $m_2=\lceil m^{r/(r+s)}\rceil$ and $H_m=H_{m_1,m_2}$. Then

$$\|H-H_m\|_{HS}\leq C\sqrt{\kappa}m^{-\frac{rs}{r+s}}. \tag{70}$$

Proof First notice that

$$T_{\infty,m_2}=\sum_{|\mu|\leq\lceil\log_2(m_2)\rceil}c_\mu\otimes\psi_\mu, \tag{71}$$

where $c_\mu(x)=\langle T(x,\cdot),\psi_\mu\rangle$. From Corollary 5, we get:

$$\|T_{\infty,m_2}-T\|_{L^2(\Omega\times\Omega)}\leq C\sqrt{\kappa}m_2^{-s}. \tag{72}$$

Now, notice that $c_\mu\in H^r(\Omega)$. Indeed, for all $0\leq k\leq r$, we get:

$$\int_\Omega(\partial_x^k c_\mu(x))^2\,dx \tag{73}$$

$$=\int_\Omega\left(\partial_x^k\int_\Omega T(x,y)\psi_\mu(y)\,dy\right)^2dx \tag{74}$$

$$=\int_\Omega\left(\int_\Omega(\partial_x^k T)(x,y)\psi_\mu(y)\,dy\right)^2dx \tag{75}$$

$$\leq\int_\Omega\|(\partial_x^k T)(x,\cdot)\|^2_{L^2(\Omega)}\|\psi_\mu\|^2_{L^2(\Omega)}\,dx \tag{76}$$

$$=\|(\partial_x^k T)\|_{L^2(\Omega\times\Omega)}<+\infty. \tag{77}$$

Therefore, we can use Lemma 6 again to show:

$$\|T_{\infty,m_2}-T_{m_1,m_2}\|_{L^2(\Omega\times\Omega)}\leq C\sqrt{\kappa}m_1^{-r}. \tag{78}$$

Finally, using the triangle inequality, we get:

$$\|T-T_{m_1,m_2}\|_{HS}\leq C\sqrt{\kappa}(m_1^{-r}+m_2^{-s}). \tag{79}$$

By setting $m_1=m_2^{s/r}$, the two approximation errors in the right-hand side of (79) are balanced. This motivates the choice of m_1 and m_2 indicated in the theorem.

The approximation result in inequality (70) is worse than the previous ones. For instance if $r=s$, then the bound becomes $O(\sqrt{\kappa}m^{-s/2})$ instead of $O(\sqrt{\kappa}m^{-s})$ in all previous theorems. The great advantage of this representation is the operator storage: until now, the whole set of vectors $(\boldsymbol{c}_\mu)$ had to be stored ($O(\kappa nm)$ values), while now, only m coefficients $c_{\lambda,\mu}$ are required. For instance, in the case $r=s$, for an equivalent precision, the storage cost of the new representation is $O(\kappa m^2)$ instead of $O(\kappa nm)$.

In addition, evaluating matrix-vector products can be achieved rapidly by using the following trick:

$$\boldsymbol{H_m u}=\sum_{|\lambda|\leq\log_2(m_1)}\sum_{|\mu|\leq\log_2(m_2)}c_{\lambda,\mu}\boldsymbol{\psi}_\lambda\star(\boldsymbol{\psi}_\mu\odot\boldsymbol{u}) \tag{80}$$

$$=\sum_{|\mu|\leq\log_2(m_2)}\left(\sum_{|\lambda|\leq\log_2(m_1)}c_{\lambda,\mu}\boldsymbol{\psi}_\lambda\right)\star(\boldsymbol{\psi}_\mu\odot\boldsymbol{u}). \tag{81}$$

By letting $\tilde{\boldsymbol{c}}_\mu=\sum_{|\lambda|\leq\log_2(m_1)}c_{\lambda,\mu}\boldsymbol{\psi}_\lambda$, we get

$$\boldsymbol{H}_m\boldsymbol{u}=\sum_{|\mu|\leq\log_2(m_2)}\tilde{\boldsymbol{c}}_\mu\star(\boldsymbol{\psi}_\mu\odot\boldsymbol{u}). \tag{82}$$

which can be can be computed in $O(m_2\kappa n\log_2(\kappa n))$ operations. This remark leads to the following proposition.

Proposition 2 *Assume that $T\in H^{r,s}(\Omega\times\Omega)$ and that it satisfies Assumption 2. Set $m=\left\lceil\left(\frac{\epsilon}{C\sqrt{\kappa}}\right)^{-(r+s)/rs}\right\rceil$. Then the operator H_m defined in Theorem 2 satisfies $\|H-H_m\|_{HS}\leq\epsilon$ and the number of operations necessary to evaluate a product with H_m or H_m^* is bounded above by $O\left(\epsilon^{-1/s}\kappa^{\frac{2s+1}{2s}}n\log_2(n)\right)$.*

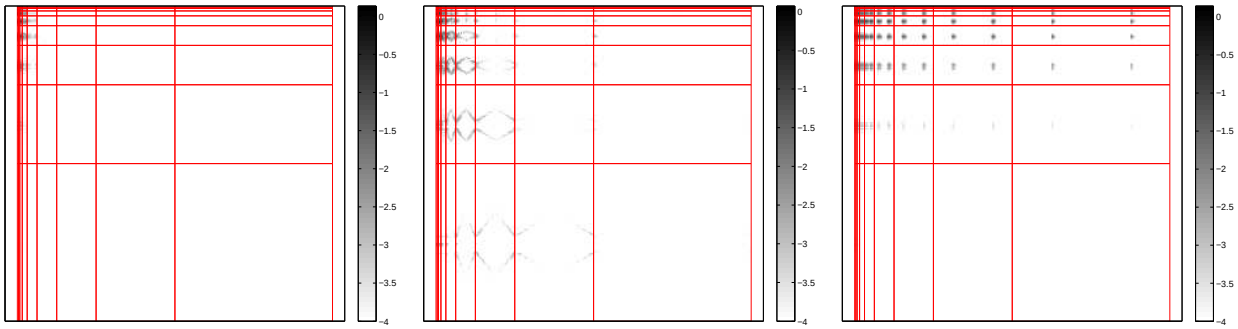
Notice that the complexity of matrix-vector products is unchanged compared to the wavelet or spline approaches with a much better compression ability. However, this method requires a preprocessing to compute $\tilde{\boldsymbol{c}}_\mu$ with complexity $\epsilon^{-1/s}\kappa^{1/2s}n$.

5 Adaptive decompositions

In the last section, all methods shared the same principle: project $T(x,\cdot)$ on a fixed basis for each $x\in\Omega$. Instead of fixing a basis, one can try to find a basis adapted to the operator at hand. This idea was proposed in [21] and [17].

5.1 Singular value decompositions

The authors of [21] proposed to use a singular value decomposition (SVD) of the TVIR in order to construct the functions h_k and w_k. In this section we first detail this idea and then analyze it from an approximation theoretic point of view. Let $J:L^2(\Omega)\to L^2(\Omega)$ denote the linear integral operator with kernel $T\in\mathcal{T}^s$. First notice that J is a Hilbert-Schmidt operator



(a) Kernel 1 (b) Kernel 2 (c) Kernel 3

Fig. 4: Meyer's representations of the operators in Examples 1, 2 and 3 in $\log_{10}$ scale.

since $\|J\|_{HS} = \|H\|_{HS}$. By Lemma 2 and since Hilbert-Schmidt operators are compact, there exists two orthonormal bases (e_k) and (f_k) of $L^2(\Omega)$ such that J can be decomposed as

$$J = \sum_{k\geq 1} \sigma_k \cdot e_k \otimes f_k, \tag{83}$$

leading to

$$T(x,y) = \sum_{k=1}^{+\infty} \sigma_k f_k(x) e_k(y). \tag{84}$$

The following result is a standard.

Theorem 3 *For a given m, a set of functions $(h_k)_{1\leq k\leq m}$ and $(w_k)_{1\leq k\leq m}$ that minimizes $\|H_m - H\|_{HS}$ is given by:*

$$h_k = \sigma_k f_k \quad \text{and} \quad w_k = e_k. \tag{85}$$

Moreover, if $T(x,\cdot)$ satisfies Assumptions 1 and 2, we get:

$$\|H_m - H\|_{HS} = O\left(\sqrt{\kappa} m^{-s}\right). \tag{86}$$

Proof The proof of optimality (86) is standard. Since T_m is the best rank m approximation of T, it is necessarily better than bound (41), yielding (86).

Theorem 4 *For all $\epsilon > 0$ and $m < n$, there exists an operator H with TVIR satisfying 1 and 2 such that:*

$$\|H_m - H\|_{HS} \geq C\sqrt{\kappa} m^{-(s+\epsilon)}. \tag{87}$$

Proof In order to prove, (87), we construct a "worst case" TVIR T. We first begin by constructing a kernel T with $\kappa = 1$ to show a simple pathological TVIR. Define T by:

$$T(x,y) = \sum_{k\in\mathbb{Z}} \sigma_k f_k(x) f_k(y), \tag{88}$$

where $f_k(x) = \exp(2i\pi kx)$ is the k-th Fourier atom, $\sigma_0 = 0$ and $\sigma_k = \sigma_{-k} = \frac{1}{|k|^{s+1/2+\epsilon/2}}$ for $|k| \geq 1$. With this choice,

$$T(x,y) = \sum_{|k|\leq N} 2\sigma_k \cos(2\pi(x+y)) \tag{89}$$

is real for all (x,y). We now prove that $T \in \mathcal{T}^s$. The k-th Fourier coefficient of $T(x,\cdot)$ is given by $\sigma_k f_k(x)$ which is bounded by σ_k for all x. By Lemma 1, $T(x,\cdot)$ therefore belongs to $H^s(\Omega)$ for all $x \in \Omega$. By construction, the spectrum of T is $(|\sigma_k|)_{k\in\mathbb{N}}$, therefore for any rank $2m+1$ approximation of T, we get:

$$\begin{aligned}
\|T - T_{2m+1}\|_{HS}^2 &\geq \sum_{|k|\geq m+1} \frac{1}{|k|^{2s+1+\epsilon}} && (90)\\
&\geq \int_{m+1}^{\infty} \frac{2}{t^{2s+1+\epsilon}}\, dt && (91)\\
&= \frac{1}{2s+\epsilon} \frac{2}{(m+1)^{2s+\epsilon}} && (92)\\
&= O(m^{-2s-\epsilon}), && (93)
\end{aligned}$$

proving the result for $\kappa = 1$. Notice that the kernel K of the operator with TVIR T only depends on x:

$$K(x,y) = \sum_{|k|\leq N} 2\sigma_k \cos(2\pi x). \tag{94}$$

Therefore the worst case TVIR exhibited here is that of a *rank 1* operator H. Obviously, it cannot be well approximated by product-convolution expansions.

Let us now construct a TVIR satisfying Assumption 2. For this, we first construct an orthonormal basis $(\tilde{f}_k)_{k\in\mathbb{Z}}$ of $L^2([-\kappa/2, \kappa/2])$ defined by:

$$\tilde{f}_k(x) = \begin{cases} \frac{1}{\sqrt{\kappa}} f_k\left(\frac{x}{\kappa}\right) & \text{if } |x| \leq \frac{\kappa}{2}, \\ 0 & \text{otherwise.} \end{cases} \tag{95}$$

The worst case operator considered now is defined by:

$$T(x,y) = \sum_{k\in\mathbb{Z}} \tilde{\sigma}_k \tilde{f}_k(x) f_k(y). \tag{96}$$

Its spectrum is $(|\tilde{\sigma}_k|)_{k\in\mathbb{Z}}$, and we get

$$|\langle T(x,\cdot), f_k\rangle| = |\tilde{\sigma}_k \tilde{f}_k(x)| = \frac{1}{\kappa}|\tilde{\sigma}_k|. \tag{97}$$

By Lemma 5, if $\tilde{\sigma}_k = \frac{\kappa}{(1+|k|^2)^s|k|^{1+\epsilon}}$, then $\|T(x,\cdot)\|_{H^s(\Omega)}$ is uniformly bounded by a constant independent of κ. Moreover, by reproducing the reasoning in (90), we get:

$$\|T - T_{2m+1}\|_{HS}^2 = O(\kappa m^{-2s-\epsilon}). \tag{98}$$

Even if the SVD provides an optimal decomposition, there is no guarantee that functions e_k are supported on an interval of small size. As an example, it suffices to consider the "worst case" TVIR given in equation (88). Therefore, vectors $\boldsymbol{w}_k$ are generically supported on intervals of size $p = n$. This yields the following proposition.

Corollary 6 *Let $\epsilon > 0$ and set $m = \lceil C\epsilon^{-1/s}\kappa^{1/2s}\rceil$. Then H_m satisfies $\|H - H_m\|_{HS} \leq \epsilon$ and a product with H_m and H_m^* can be evaluated with no more than $O(\kappa^{1/2s} n \log n \epsilon^{-1/s})$ operations.*

Computing the first m singular vectors in (84) can be achieved in roughly $O(\kappa n^2 \log(m))$ operations thanks to recent advances in randomized algorithms [25]. The storage cost for this approach is $O(mn)$ since the vectors $\boldsymbol{e}_k$ have no reason to be compactly supported.

5.2 The optimization approach in [17]

In [17], the authors propose to construct the windowing functions w_k and the filters h_k using constrained optimization procedures. For a fixed m, they propose solving:

$$\min_{(h_k,w_k)_{1\leq k\leq m}} \left\| T - \sum_{k=1}^{m} h_k \otimes w_k \right\|_{HS}^2 \tag{99}$$

under an additional constraint that $\mathrm{supp}(w_k) \subset \omega_k$ with ω_k chosen so that $\cup_{k=1}^m \omega_k = \Omega$. A decomposition of type 99 is known as structured low rank approximation [9]. This problem is non convex and to the best of our knowledge, there currently exists no algorithm running in a reasonable time to find its global minimizer. It can however be solved approximately using alternating minimization like algorithms.

Depending on the choice of the supports ω_k, different convergence rates can be expected. However, by using the results for B-splines in section 4.2, we obtain the following proposition.

Proposition 3 *Set $\omega_k = [(k-1)/m, k/m + s/m]$ and let $(h_k, w_k)_{1\leq k\leq m}$ denote the global minimizer of (99). Define T_m by $T_m(x,y) = \sum_{k=1}^m h_k(x)w_k(y)$. Then:*

$$\|T - T_m\|_{HS}^2 \leq C\sqrt{\kappa} m^{-s}. \tag{100}$$

Set $m = \lceil \kappa^{1/2s} C\epsilon^{-1/s}\rceil$, then $\|H_m - H\|_{HS} \leq \epsilon$ and the evaluation of a product with H_m or H_m^* is of order

$$O(\kappa^{1+1/2s} n \log(n) \epsilon^{-1/s}). \tag{101}$$

Proof First notice that cardinal B-Splines are also supported on $[(k-1)/m, k/m + s/m]$. Since the method in [17] provides the best choices for (h_k, w_k), the distance $\|H_m - H\|_{HS}$ is necessarily lower than that obtained using B-splines in Corollary 3.

Finally, let us mention that - owing to Corollary 5 - it might be interesting to use the optimization approach (99) with windows of varying sizes.

6 Summary and extensions

6.1 A summary of all results

Table 1 summarizes the results derived so far under Assumptions 1 and 2. In the particular case of Meyer's methods, we assume that $T \in H^{r,s}(\Omega \times \Omega)$ instead of Assumption 1. As can be seen in this table, different methods should be used depending on the application. The best methods are:

- Wavelets: they are adaptive, have a relatively low construction complexity, and matrix-vector products also have the best complexity.
- Meyer: this method has a big advantage in terms of storage. The operator can be represented very compactly with this approach. It has a good potential for problems where the operator should be inferred (e.g. blind deblurring). It however requires stronger regularity assumptions.
- The SVD and the method proposed in [17] both share an optimal adaptivity. The representation however depends on the operator and it is more costly to evaluate it.

6.2 Extensions to higher dimensions

Most of the results provided in this paper are based on standard approximation results in 1D, such as Lemmas 5, 6 and 1. All these lemmas can be extended to higher dimension and we refer the interested reader to [19,31, 18,36] for more details.

Method	*Approximation*	*Product*	*Construction*	*Storage*	*Adaptivity*
Fourier 4.1	$O\left(\kappa^{\frac{1}{2}} m^{-s}\right)$	$O\left(\kappa^{\frac{1}{2s}} n \log(n) \epsilon^{-\frac{1}{s}}\right)$	$O(\kappa n^2 \log(n))$	$O(m\kappa n)$	✗
B-Splines 4.2	$O\left(\kappa^{\frac{1}{2}} m^{-s}\right)$	$O\left(\kappa^{\frac{2s+1}{2s}} n \log(n) \epsilon^{-\frac{1}{s}}\right)$	$O(\kappa n^2 \log(n))$	$O(m\kappa n)$	✗
Wavelets 4.3	$O\left(\kappa^{\frac{1}{2}} m^{-s}\right)$	$O\left(\kappa^{\frac{2s+1}{2s}} n \log(n) \epsilon^{-\frac{1}{s}}\right)$	$O(\kappa s n^2)$	$O(m\kappa n)$	✓
Meyer 4.5	$O\left(\kappa^{\frac{1}{2}} m^{-\frac{rs}{r+s}}\right)$	$O\left(\kappa^{\frac{2s+1}{2s}} n \log(n) \epsilon^{-\frac{1}{s}}\right)$	$O(s n^2)$	$O(m)$	✓
SVD 5.1	$O\left(\kappa^{\frac{1}{2}} m^{-s}\right)$	$O\left(\kappa^{\frac{1}{2s}} n \log(n) \epsilon^{-\frac{1}{s}}\right)$	$O(\kappa n^2 \log(m))$	$O(mn)$	✓
[17] 5.2	$O\left(\kappa^{\frac{1}{2}} m^{-s}\right)$	$O\left(\kappa^{\frac{2s+1}{2s}} n \log(n) \epsilon^{-\frac{1}{s}}\right)$	High (iterative)	$O(m\kappa n)$	✓

Table 1: Summary of the properties of different constructions. Approximation ≡ approximation rates in terms of m. Product ≡ matrix-vector product complexity to get an ϵ approximation. Construction ≡ complexity of the construction of order m representation. Storage ≡ cost of storage of a given representation. Adaptivity ≡ ability to automatically adapt to different input operators.

We now assume that $\Omega = [0,1]^d$ and that the diameter of the impulse responses is bounded by $\kappa \in [0,1]$. Using the mentioned results, it is straightforward to show that the approximation rate of all methods now becomes

$$\|H - H_m\|_{HS} = O(\kappa^{d/2} m^{-s/d}). \tag{102}$$

The space Ω can be discretized on a finite dimensional space of size n^d. Similarly, all complexity results given in Table 1 are still valid by replacing n by n^d, $\epsilon^{-1/s}$ by $\epsilon^{-d/s}$ and κ by κ^d.

6.3 Extensions to least regular spaces

Until now, we assumed that the TVIR T belongs to Hilbert spaces (see e.g. Assumption 1). This assumption was deliberately chosen easy to clarify the presentation. The results can most likely be extended to much more general spaces using nonlinear approximation theory results [18].

For instance, assume that $T \in BV(\Omega \times \Omega)$, the space of functions with bounded variations. Then, it is well known (see e.g. [11]) that T can be expressed compactly on an orthonormal basis of tensor-product wavelets. Therefore, the product-convolution expansion 4 could be used by using the trick proposed in 82.

Similarly, most of the kernels found in partial differential equations (e.g. Calderòn-Zygmund operators) are singular at the origin. Once again, it is well known [32] that wavelets are able to capture the singularities and the proposed methods can most likely be applied to this setting too.

A precise setting useful for applications requires more work and we leave this issue open for future work.

6.4 Controls in other norms

In all the paper we only controlled the Hilbert-Schmidt norm $\|\cdot\|_{HS}$. This choice simplifies the analysis and also allows getting bounds for the spectral norm

$$\|H\|_{2\to 2} = \sup_{\|u\|_{L^2(\Omega)} \leq 1} \|Hu\|_{L^2(\Omega)}, \tag{103}$$

since $\|H\|_{2\to 2} \leq \|H\|_{HS}$. In applications, it often makes sense to consider other operator norms defined by

$$\|H\|_{X\to Y} = \sup_{\|u\|_X \leq 1} \|Hu\|_Y, \tag{104}$$

where $\|\cdot\|_X$ and $\|\cdot\|_Y$ are norms characterizing some function spaces. We showed in [20] that this idea could highly improve practical approximation results.

Unfortunately, it is not clear yet how to extend the proposed results and algorithms to such a setting and we leave this question open for the future. Let us mention that our previous experience shows that this idea can highly change the method's efficiency.

7 Conclusion

In this paper, we analyzed the approximation rates and numerical complexity of product-convolution expansions. This approach was shown to be efficient whenever the time or space varying impulse response of the operator is well approximated by a low rank tensor. We showed that this situation occurs under mild regularity assumptions, making the approach relevant for a large class of applications. We also proposed a few original implementations of this methods based on orthogonal wavelet decompositions and analyzed their respective advantages precisely. Finally, we suggested a few ideas to further improve the practical efficiency of the method.

References

1. Muthuvel Arigovindan, Joshua Shaevitz, John McGowan, John W Sedat, and David A. Agard. A parallel product-convolution approach for representing the depth varying point spread functions in 3d widefield microscopy based on principal component analysis. *Opt. Express*, 18(7):6461–6476, 2010.
2. Gregory Beylkin. On the representation of operators in bases of compactly supported wavelets. *SIAM J. Numer. Anal.*, 29(6):1716–1740, 1992.
3. Gregory Beylkin, Ronald Coifman, and Vladimir Rokhlin. Fast wavelet transforms and numerical algorithms I. *Appl. Comput. Harmon. A.*, 44(2):141–183, 1991.
4. Anatoly Yu Bezhaev and Vladimir Aleksandrovich Vasilenko. *VARIATIONAL THEORY OF SPLINES*. Springer, 2001.
5. Jérémie Bigot, Paul Escande, and Pierre Weiss. Estimation of linear operators from scattered impulse responses. *arXiv preprint arXiv:1610.04056*, 2016.
6. Steffen Börm, Lars Grasedyck, and Wolfgang Hackbusch. Introduction to hierarchical matrices with applications. *Eng. Anal. Bound. Elem.*, 27(5):405–422, 2003.
7. Robert Busby and Harvey Smith. Product-convolution operators and mixed-norm spaces. *T. Am. Math. Soc*, 263(2):309–341, 1981.
8. Ole Christensen. *AN INTRODUCTION TO FRAMES AND RIESZ BASES*. Springer Science & Business Media, 2013.
9. Moody Chu, Robert Funderlic, and Robert Plemmons. Structured low rank approximation. *Linear Algebra Appl.*, 366:157–172, 2003.
10. Philippe G. Ciarlet. *THE FINITE ELEMENT METHOD FOR ELLIPTIC PROBLEMS*, volume 40. SIAM, 2002.
11. Albert Cohen, Wolfgang Dahmen, Ingrid Daubechies, and Ronald DeVore. Harmonic analysis of the space BV. *Rev. Mat. Iberoam.*, 19(1):235–263, 2003.
12. Albert Cohen, Ingrid Daubechies, and Pierre Vial. Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. A.*, 1(1):54–81, 1993.
13. Wolfgang Dahmen, Siegfried Prößdorf, and Reinhold Schneider. Wavelet approximation methods for pseudodifferential equations ii: Matrix compression and fast solution. *Adv. Comput. Math.*, 1(3):259–335, 1993.
14. Wolfgang Dahmen, Siegfried Prössdorf, and Reinhold Schneider. Wavelet approximation methods for pseudodifferential equations: I stability and convergence. *Math. Z.*, 215(1):583–620, 1994.
15. Ingrid Daubechies. Orthonormal bases of compactly supported wavelets.
16. Ingrid Daubechies. *TEN LECTURES ON WAVELETS*, volume 61. SIAM, 1992.
17. Loïc Denis, Eric Thiébaut, Ferréol Soulez, Jean-Marie Becker, and Rahul Mourya. Fast approximations of shift-variant blur. *Int. J. Comput. Vision*, 115(3):253–278, 2015.
18. Ronald A. DeVore. Nonlinear approximation. *Act. Num.*, 7:51–150, 1998.
19. Ronald A. DeVore and George G. Lorentz. *CONSTRUCTIVE APPROXIMATION*, volume 303. Springer Science & Business Media, 1993.
20. Paul Escande and Pierre Weiss. Sparse wavelet representations of spatially varying blurring operators. *SIAM J. Imaging. Sci.*, 8(4):2976–3014, 2015.
21. Ralf C. Flicker and François J. Rigaut. Anisoplanatic deconvolution of adaptive optics images. *J. Opt. Soc. Am. A*, 22(3):504–513, 2005.
22. Erez Gilad and Jost Von Hardenberg. A fast algorithm for convolution integrals with space and time variant kernels. *J. Comput. Phys.*, 216(1):326–336, 2006.
23. L.J. Griffiths, F.R. Smolka, and L.D. Trembly. Adaptive deconvolution: A new technique for processing time-varying seismic data. *Geophysics*, 42(4):742–759, 1977.
24. Wolfgang Hackbusch and Zenon Paul Nowak. On the fast matrix multiplication in the boundary element method by panel clustering. *Numer. Math.*, 54(4):463–491, 1989.
25. Nathan Halko, Per-Gunnar Martinsson, and Joel A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011.
26. Aleksandr Yakovlevich Helemskii. *LECTURES AND EXERCISES ON FUNCTIONAL ANALYSIS*, volume 233. American Mathematical Society Providence, RI, 2006.
27. Michael Hirsch, Suvrit Sra, Bernhard Scholkopf, and Stefan Harmeling. Efficient filter flow for space-variant multiframe blind deconvolution. In *Proc. CVPR IEEE*, 2010.
28. Tomasz Hrycak, Saptarshi Das, Gerald Matz, and Hans G. Feichtinger. Low complexity equalization for doubly selective channels modeled by a basis expansion. *IEEE T. Signal Proces.*, 58(11):5706–5719, 2010.
29. Yitzhak Katznelson. *AN INTRODUCTION TO HARMONIC ANALYSIS*. Cambridge University Press, 2004.
30. Werner Kozek and Götz E. Pfander. Identification of operators with bandlimited symbols. *SIAM J. Math. Anal.*, 37(3):867–888, 2005.
31. Stéphane Mallat. *A WAVELET TOUR OF SIGNAL PROCESSING*. Academic Press, 1999.
32. Yves Meyer. *WAVELETS AND OPERATORS*, volume 1. Cambridge University Press, 1995.
33. David Miraut and Javier Portilla. Efficient shift-variant image restoration using deformable filtering (Part I). *EURASIP J. Adv. Sig. Proc.*, 2012:100, 2012.
34. James G. Nagy and Dianne P. O'Leary. Restoring images degraded by spatially variant blur. *SIAM J. Sci. Comput.*, 19(4):1063–1082, 1998.
35. Ivan Oseledets and Eugene Tyrtyshnikov. TT-cross approximation for multidimensional arrays. *Linear Algebra Appl.*, 432(1):70–88, 2010.
36. Allan Pinkus. *N-WIDTHS IN APPROXIMATION THEORY*, volume 7. Springer Science & Business Media, 2012.
37. June Chul Roh and Bhaskar D. Rao. An efficient feedback method for MIMO systems with slowly time-varying channels. In *IEEE WCNC*, volume 2, pages 760–764. IEEE, 2004.
38. Alexander A. Sawchuk. Space-variant image motion degradation and restoration. *P IEEE*, 60(7):854–861, 1972.
39. Thomas G. Stockham. High-speed convolution and correlation. In *Proceedings of the April 26-28, 1966, Spring joint computer conference*, pages 229–233. ACM, 1966.
40. Joel Trussell and Sergei Fogel. Identification and restoration of spatially variant motion blurs in sequential images. *IEEE T. Image. Process.*, 1(1):123–126, 1992.
41. Michael Unser, Akram Aldroubi, and Murray Eden. B-spline signal processing. Part I - Theory. *IEEE T. Signal Proces.*, 41(2):821–833, 1993.