Numerical Computation of Spatially Varying Blur Operators A Review of Existing Approaches with a New One

Paul Escande[a], Pierre Weiss[b]

[a] *Département Mathématiques, Informatique, Automatique (DMIA)*
Institut Supérieur de l'Aéronautique et de l'Espace (ISAE)
Toulouse, France
[b] *Institut des Technologies Avancées en Sciences du Vivant*
ITAV-USR3505
Toulouse, France

Abstract

Restoring images degraded by spatially varying blur is a problem encountered in many disciplines such as astrophysics, computer vision or biomedical imaging. One of the main challenges to perform this task is to design efficient numerical algorithms to approximate integral operators.

We review the main approaches developped so far and detail their pros and cons. We then analyze the numerical complexity of the mainstream approach based on piecewise convolutions. We show that this method provides an ϵ-approximation of the matrix-vector product in $\mathcal{O}\left(N\log(N)\epsilon^{-1}\right)$ operations where N is the number of pixels. Moreover, we show that this bound cannot be improved even if further assumptions on the kernel regularity are made.

We then introduce a new method based on a sparse approximation of the blurring operator in the wavelet domain. This method requires $\mathcal{O}\left(N\log(N)\epsilon^{-1/M}\right)$ operations to provide ϵ-approximations, where $M \geq 1$ is a scalar describing the regularity of the blur kernel. We then propose variants to further improve the method by exploiting the fact that both images and operators are sparse in the same wavelet basis.

We finish by numerical experiments to illustrate the practical efficiency of the proposed algorithms.

Keywords:
Image deblurring, spatially varying blur, integral operator approximation, wavelet compression, piecewise convolution

Email addresses: paul.escande@gmail.com (Paul Escande), pierre.armand.weiss@gmail.com (Pierre Weiss)

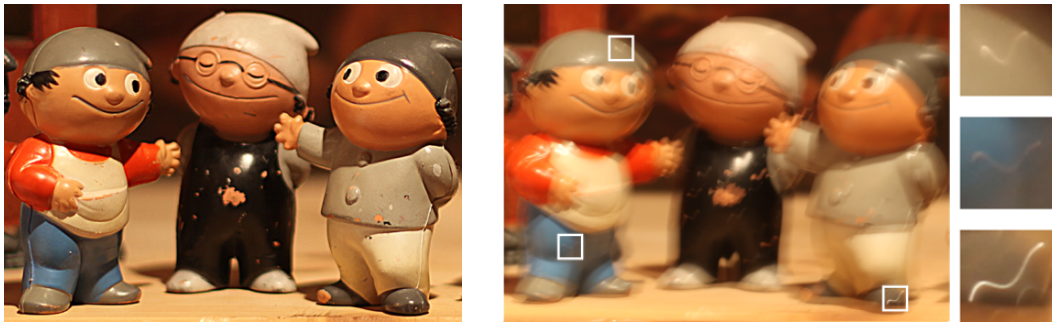
Contents

1. Introduction

The problem of image restoration in the presence of spatially varying blur appears in many domains. Examples of applications in computer vision, biomedical imaging and astronomy are shown in Figures 1, 2 and 3 respectively. This problem has a long history that probably started at the beginning of the seventies (see e.g. [1]). However, despite being present in most imaging systems, it seems to have drawn little attention in the literature compared to the space invariant models. There are probably two important facts explaining this phenomenon: i) the difficulty to estimate the Point Spread Function (PSF) at every point of the image domain and ii) the difficulty to efficiently compute spatially varying blurs. In this paper, we focus on problem ii) even though this work might open new perspectives to solve problem i).

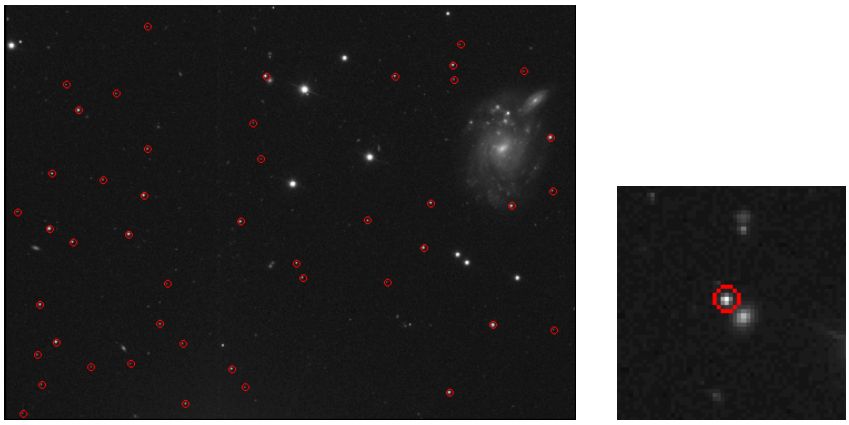
Spatially invariant blurs can be evaluated efficiently since convolution operators are diagonal in the Fourier domain. This allows to blur images using the fast Fourier transform (FFT), leading to algorithms with an $\mathcal{O}(N \log N)$ complexity, where N denotes the number of pixels of the image. To the best of our knowledge, no such tool exists in the case of spatially varying blurs. The main result of this paper is to show that under mild technical assumptions, spatially varying blur operators can be approximated efficiently by sparse matrices in the wavelet domain. We show that approximations of the blurred images can be obtained with a precision ϵ in no more than $\mathcal{O}\left(N \log\left(N\right) \epsilon^{-1/M}\right)$ operations, where $M \geq 1$ is an integer that describes the smoothness of the blur kernel. This result is a consequence of the seminal works [2, 3]. Surprisingly, these have received little attention in the imaging community despite the success of wavelets in image compression.

The outline of this paper is as follows. We introduce the notation use throughout the paper in Section 2. We propose an original mathematical description of blurring operators appearing in image processing in Section 3. We review the main existing computational strategies and analyse their pros and cons in Section 4. By far, in the literature, the most commonly approach used to spatially blur images consists in using piecewise convolutions. [4, 5, 6]. Since this technique is so widespread, we propose an original analysis of its theoretical efficiency in Section 5. We introduce the proposed method and analyze its theoretical efficiency Section 6. We then propose various algorithms to design good sparsity patterns in Section 7. Finally, we perform numerical tests to analyze the proposed method and compare it to the standard methods based on piecewise convolutions in Section 8.



(a) Sharp image (b) Blurred image and the associated PSF

Figure 1: An example in computer vision. Image degraded by spatially varying blur due to a camera shake. Images are from [7] and used here by courtesy of Michael Hirsch.



(a) (b)

Figure 2: An example in astronomy. Image degraded by spatially varying blur due to atmoshpere turbulence `http://www.sdss.org/`.

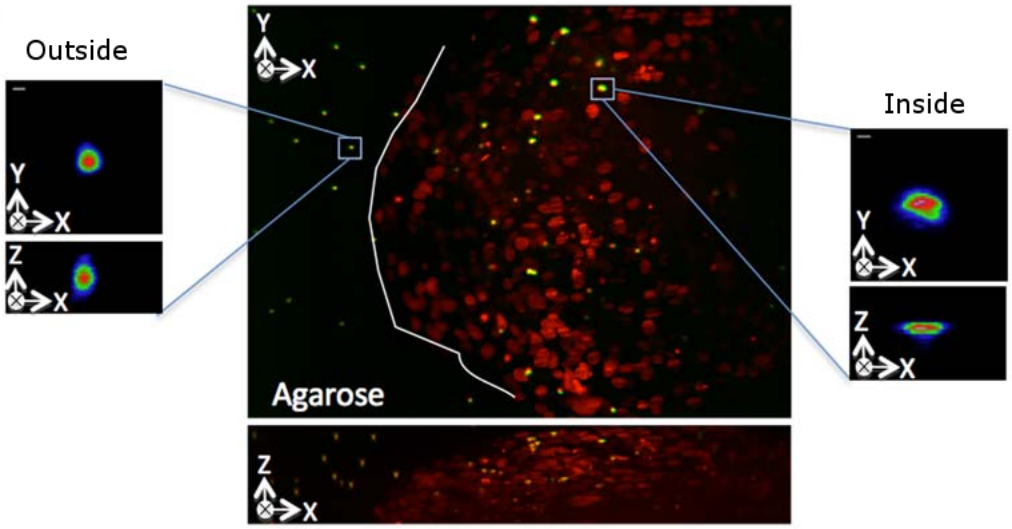


Figure 3: An example in biology. Image of a multicellular tumor spheroid imaged in 3D using Selective Plane Illumination Microscope (SPIM). Fluorescence beads (in green) are inserted in the tumor model and allow the observation of the PSF at different locations. Nuclei are stained in red. On the left-hand-side the 3D PSF outside of the sample is observed. On the right-hand-side the 3D PSF inside the sample is observed. This image is from [8] and used here by courtesy of Corinne Lorenzo et al.

2. Notation

In this paper, we consider d dimensional images defined on a domain $\Omega = [0,1]^d$. The space $\mathbb{L}^2(\Omega)$ will denote the space of squared integrable functions defined on Ω. The set of compactly supported functions in $\mathcal{C}^\infty(\Omega)$ is denoted $\mathcal{D}(\Omega)$. Its dual space (the set of distributions) is denoted $\mathcal{D}'(\Omega)$.

The Sobolev spaces $W^{M,p}$ are defined as the set of functions $f \in \mathbb{L}^p$ with partial derivatives up to order M in $\mathbb{L}^p$ where $p \in [1, +\infty]$ and $M \in \mathbb{N}$. These spaces, equipped with the following norm are Banach spaces

$$\|f\|_{W^{M,p}} = \|f\|_{\mathbb{L}^p} + |f|_{W^{M,p}}, \quad \text{where,} \quad |f|_{W^{M,p}} = \sum_{|\alpha|=M} \|\partial^\alpha f\|_{\mathbb{L}^p}. \tag{1}$$

Let X and Y denote two metric spaces endowed with their respective norms $\|\cdot\|_X$ and $\|\cdot\|_Y$. In all the paper $H : X \to Y$ will denote a linear operator and H^* its adjoint operator. The subordinate operator norm is defined by $\|H\|_{X\to Y} = \sup_{x\in X, \|x\|_X=1} \|Hx\|_Y$. The notation $\|H\|_{p\to q}$ corresponds to the case where X and Y are endowed with the standard $\mathbb{L}^p$ and $\mathbb{L}^q$ norms. In all the paper, operators acting in a continuous domain are written in plain text format H. Finite dimensional matrices are written in bold fonts $\mathbf{H}$. Approximation operators will be denoted $\widetilde{H}$ in the continuous domain or $\widetilde{\mathbf{H}}$ in the discrete domain.

In this paper we consider a wavelet basis of $\mathbb{L}^2([0,1])$. We let ϕ and ψ denote the scaling and mother wavelets. We assume that the mother-wavelet ψ has M vanishing moments, i.e.

$$\text{for all } 0 \leq m < M, \quad \int_{\mathbb{R}} t^m \psi(t) dt = 0.$$

Furthermore, it will be assumed that the wavelets ψ and all its derivatives up to the M-th order belong to $\mathbb{L}^\infty(\Omega) \cap \mathbb{L}^2(\Omega)$

We define translated and dilated versions of the wavelets as follows

$$\begin{aligned} \phi_{j,m} &= 2^{j/2} \phi\left(2^j \cdot - m\right), \\ \psi_{j,m} &= 2^{j/2} \psi\left(2^j \cdot - m\right). \end{aligned} \tag{2}$$

We assume that every function f of $\mathbb{L}^2(\Omega)$ can be written as

$$u = \sum_{m\in\mathcal{T}_{l_0}} \langle u, \phi_{l_0,m}\rangle \, \phi_{l_0,m} + \sum_{j\geq l_0} \sum_{m\in\mathcal{T}_j} \langle u, \psi_{j,m}\rangle \, \psi_{j,m}$$

where $\mathcal{T}_j = \{0 \leq m < 2^j\}$. This is a slight abuse since wavelets defined in (2) do not define a Hilbert basis of $\mathbb{L}^2([0,1])$. There are various ways to define wavelet

bases on the interval [9] and wavelets having a support intersecting the boundary should be given a different definition. We stick to these definitions to keep the proofs arguments as simple as can be.

We let $\Psi^* : \mathbb{L}^2(\Omega) \to l^2(\mathbb{Z})$ denote the wavelet decomposition operator and $\Psi : l^2(\mathbb{Z}) \to \mathbb{L}^2(\Omega)$ its associated reconstruction operator. In what follows, we will always set $l_0 = 0$. We refer to [10, 11, 9] for more details on the construction of wavelet bases.

The discrete wavelet transform is denoted $\mathbf{\Psi} : \mathbb{R}^N \to \mathbb{R}^N$. The analyses we propose can be extended to $[0,1]^d$ using tensor-products wavelet bases for instance. We provide proofs in the one-dimensional setting for the ease of exposition, but they can be extended to the d-dimensional setting using almost the same arguments.

3. Blurring operators and their mathematical properties

3.1. A mathematical description of blurring operators

In this paper, we consider d-dimensional real-valued images defined on a domain $\Omega = [0,1]^d$, where d denotes the space dimension. We consider a blurring operator $H : \mathbb{L}^2(\Omega) \to \mathbb{L}^2(\Omega)$ defined for any $u \in \mathbb{L}^2(\Omega)$ by the following integral operator:

$$\forall x \in \Omega, \quad Hu(x) = \int_{y \in \Omega} K(x,y)u(y)dy. \tag{3}$$

The function $K : \Omega \times \Omega \to \mathbb{R}$ is a kernel that defines the Point Spread Function (PSF) $K(\cdot, y)$ at each location $y \in \Omega$. The image Hu is the blurred version of u. By the Schwartz kernel theorem, a linear operator of kind (3) can basically represent any linear operator if K is a generalized function. We thus need to determine properties of K specific to blurring operators that will allow to design efficient numerical algorithms to approximate the integral (3). In most practical applications, the kernel satisfies the following properties:

Spatial decay. The PSFs usually have a bounded support (e.g. motion blurs, convolution with the ccd sensors support) or at least a fast spatial decay (Airy pattern, Gaussian blurs,...). This property can be modelled as

$$|K(x,y)| \leq \frac{C}{\|x-y\|_2^\alpha} \tag{4}$$

whenever $\|x-y\|_2$ is sufficiently large and where $\alpha > 0$ describes the decay speed and C is a positive constant. For instance, the 2D Airy disk describing the PSF due to diffraction of light in a circular aperture satisfies (4) with $\alpha = 4$ [12].

Boundedness of the operator. Throughout the paper, we will assume that the operator $H : \mathbb{L}^2(\Omega) \to \mathbb{L}^2(\Omega)$ is bounded. A sufficient condition for this condition to hold is that $K(x, y)$ be bounded at every point $(x, y) \in \Omega \times \Omega$. In fact we will see that $K(x, y)$ can be unbounded on the diagonal $x = y$. From a physical point of view, this assumption only means that the acquisition system keeps the energy finite.

PSF smoothness. This assumption means that the function $x \mapsto K(x, y)$ is smooth for all $y \in \Omega \setminus \{x\}$. This hypothesis is usually satisfied since the PSF is the result of a convolution with the acquisition device impulse response. In most imaging applications, this impulse response is smooth (e.g. Airy disk) and thus leads to a regular PSF. From a mathematical point of view, we may thus assume that $K(\cdot, y)$ is $\mathcal{C}^m(\Omega)$ for a certain $m \geq 1$ and for all $y \in \Omega \setminus \{x\}$. Since the PSF decays spatially, we may also assume that:

$$|\partial_x^m K(x, y)| \leq \frac{C}{\|x - y\|_2^\beta} \tag{5}$$

for a certain $\beta > 0$ and all $y \neq x$.

PSFs variations are smooth The function $y \mapsto K(x, y)$ is smooth for all $x \in \Omega$. More precisely, we assume that $K(x, \cdot)$ is $\mathcal{C}^m(\Omega)$ for a certain $m \geq 1$, for all $x \in \Omega$ and that

$$\left|\partial_y^m K(x, y)\right| \leq \frac{C}{\|x - y\|_2^\gamma} \tag{6}$$

for a certain $\gamma > 0$. From a practical point of view, it means that the PSF does not vary abruptly on the domain. This hypothesis is not true in all applications. For instance, when objects move in front of a still background, the PSF can only be considered as piecewise regular. The PSFs in astronomy under very turbulent atmosphere may also vary abruptly. This assumption simplifies the analysis of numerical procedures to approximate H. Moreover, it seems reasonable in many settings. For instance, in fluorescence microscopy, the PSF width (or Strehl ratio) mostly depends on the optical thickness, i.e. the quantity of matter laser light has to go through, and this quantity is intrinsically continuous. Even in cases where the PSFs variations are not smooth (e.g. spatially varying motion blur), the discontinuities locations are usually known only approximately and it seems important to smooth the transitions in order to avoid reconstruction artifacts.

Remark 1. A standard assumption in image processing is that the constant functions are preserved by the operator H. This hypothesis ensures that brightness is

preserved on the image domain. In this paper we do not make this assumption and thus encompass image formation models comprising blur and attenuation. Handling attenuation is an important problem in domains such as fluroescence microscopy.

Remark 2. The above properties are important to derive mathematical theories, but only represent an approximation of real systems. The methods proposed in this paper may be applied even if the above properties are not satisfied and are likely to perform well. Let us stress out that in many practical devices even linearity of the acquisition device should be call into question.

3.2. Link with Calderón-Zygmund operators

The properties we just described allow to classify blurring operators as Calderón-Zygmund operators. The latter have the following definition:

Definition 1 (Calderón-Zygmund operators). A continuous linear operator $H : \mathcal{D}(\mathbb{R}^d) \to \mathcal{D}'(\mathbb{R}^d)$ corresponds to a singular operator of regularity $M \geq 1$, if its distribution-kernel, restricted to $\mathbb{R}^d \times \mathbb{R}^d \setminus \{x = y\}$ is a function $K(x, y)$ satisfying the following properties,

$$|K(x,y)| \leq C_0 \frac{1}{|x-y|^d},$$
$$|\partial_x^m K(x,y)| + \left|\partial_y^m K(x,y)\right| \leq C_1 \frac{1}{|x-y|^{d+m}}, \quad \text{for all } m \leq M.$$

For a test function $u \in \mathcal{D}(\mathbb{R}^d)$, $Hu(x)$ is defined, for all x outside the support of u, by

$$Hu(x) = \int_{\mathbb{R}^d} K(x,y)u(y)dy.$$

The operator H is called a Calderón-Zygmund operator if it can be extended to a continuous linear operator from $\mathbb{L}^2(\mathbb{R}^d) \to \mathbb{L}^2(\mathbb{R}^d)$.

Calderón-Zygmund operators initially appeared to generalize the Riesz transform which plays an important role in harmonic analysis. They also model many physical phenomena in acoustics or electromagnetism. Calderón-Zygmund kernels can be singular on the diagonal of the domain (see e.g. Biot-Savart law). These operators have been studied extensively. For example, David and Journé [13] studied sufficient conditions ensuring the boundedness of the operator from $\mathbb{L}^2(\mathbb{R}^d) \to \mathbb{L}^2(\mathbb{R}^d)$. For the sake of efficient operator computation, important results have been derived in different papers [3, 2]. The authors investigated the operator compression in wavelet bases. We will provide a numerical study of these ideas in Section 6.

Remark 3. Definition 1 might look awkward to non mathematicians. It can be highly simplified when K is sufficiently regular. Definition 2 below is probably enough for most practical problems encountered in imaging.

Definition 2 (Calderón-Zygmund operators - for beginners). An integral operator $H : \mathbb{L}^2(\Omega) \to \mathbb{L}^2(\Omega)$ with a kernel $K \in \mathcal{C}^M(\Omega \times \Omega)$ is a Calderón-Zygmund operator of regularity $M \geq 1$ if the following properties are satisfied

$$|K(x,y)| \leq C_0 \frac{1}{|x-y|^d},$$
$$|\partial_x^m K(x,y)| + \left|\partial_y^m K(x,y)\right| \leq C_1 \frac{1}{|x-y|^{d+m}}, \quad \text{for all } m \leq M.$$

4. A brief review of existing approximation methods

Various approaches have been proposed in the literature to approximate numerically the integral operator (3). In this section, we review some of these methods. We briefly discuss their pros and cons. In section 5, we will focus and analyze the most commonly used approach based on piecewise convolutions. In many situations, $K(x,\cdot)$ has bounded support and this property can be exploited to accelerate computations. In the following we define,

$$\kappa = \sup_{x \in \Omega} \operatorname{diam}\left(K(x,\cdot)\right), \tag{7}$$

where $\operatorname{diam}\left(K(x,\cdot)\right) = \sup_{x_1,x_2 \in \operatorname{supp} K(x,\cdot)} \|x_1 - x_2\|_\infty$.

4.1. Direct discretization

The most straightforward approach to approximate the product Hu numerically consists of discretizing (3) using the rectangle rule. Let $\mathbf{\Omega} = \{k/N\}_{k \in \{0,\dots,N\}}^d$ denote a Euclidean discretization of Ω and $\mathbf{u} \in \mathbb{R}^{N^d}$ denote a discretization of u on $\mathbf{\Omega}$. The product Hu can be approximated by $\mathbf{Hu}$ where $\mathbf{H}$ is an $N^d \times N^d$ matrix defined by

$$\mathbf{H} = \frac{1}{N^d} \left(K(x,y)\right)_{x \in \mathbf{\Omega}, y \in \mathbf{\Omega}}. \tag{8}$$

The rationale behind this discretization is that if u and H are sufficiently smooth

$$\forall x \in \mathbf{\Omega}, \quad Hu(x) \simeq \frac{1}{N^d} \sum_{y \in \mathbf{\Omega}} K(x,y) u(y). \tag{9}$$

This approach is simple to implement but a matrix vector product costs $\mathcal{O}(N^{2d})$ arithmetic operations. By taking into account the boundedness of the kernel support, one can easily reduce the complexity to $\mathcal{O}(\kappa^d N^{2d})$ where $\kappa \in]0,1]$ is defined in (7). Indeed, for each of the N^d pixels, the method performs $(\kappa N)^d$ operations (additions and mulitplications).

This method has the advantages of being straightforward to understand and implement. It is also easily parallelizable. It is thus a suitable method when the PSF width remains small over the image domain. However, it becomes unusable whenever the image size and the PSF sizes become large, i.e. in most practical settings.

Note that other quadrature formulas can be used and improve the approximation quality.

4.2. Composition of a diffeomorphism and a convolution

One of the first alternative method proposed to reduce the computational complexity, is based on first applying a diffeomorphism to the image domain [1, 14, 15, 16] followed by a convolution using FFTs and an inverse diffeomorphism. The diffeomorphism is chosen in order to transform the spatially varying blur into an invariant one. This approach suffers from two important drawbacks:

- first it was shown that not all spatially varying kernel can be approximated by this approach [15],
- second, this method requires good interpolation methods and the use of Euclidean grids with small grid size in order to correctly estimate integrals.

4.3. Approximation by separable kernels

This approach was described in [17, 18]. The main idea is to approximate the kernel K of H by a separable kernel $\widetilde{K}$ that reads:

$$\widetilde{K}(x,y) = \prod_{k=1}^{d} \widetilde{K}_k(x_k, y_k)$$

where each $\widetilde{K}_k : [0,1] \times [0,1] \to \mathbb{R}$ operates in only one direction. With this assumption, the approximation operator $\widetilde{H}$ can be decomposed as the product of d one-dimensional operators $\widetilde{H}^{(k)}$

$$\widetilde{H} = \widetilde{H}^{(1)} \circ \ldots \circ \widetilde{H}^{(d)}. \tag{10}$$

with

$$\forall x \in \Omega, \quad \widetilde{H}^{(k)}u(x) = \int_{y_k \in [0,1]} \widetilde{K}_k(x_k, y_k) u\left((x_1, \ldots, y_k, \ldots, x_d)\right) dy_k.$$

In the discrete setting, the computational complexity of a product $\widetilde{\mathbf{H}}\mathbf{u}$ is $\mathcal{O}(d\kappa N^{d+1})$ operations where κ denotes the PSF extent in pixels. The complexity of a product is thus reduced by a factor $\kappa^{d-1}N^{d-1}$ compared to the standard discretization described in paragraph 4.1.

The separability assumption (10) implies that $y \mapsto K(\cdot, y)$ is a separable function, meaning that the PSFs are separable. Moreover, it implies that $x \mapsto K(x, \cdot)$ is a separable function, meaning that the PSFs variations are also separable. Unfortunately, most physically realistic PSFs are not separable (see e.g. Figure 4). Furthermore, the separability of the PSFs variations is an extremely restrictive assumption. There are however a few cases where this approximation might be sound. For instance, in 3D fluorescence microscopy, it is common to approximate the PSFs by anisotropic Gaussians [19], and to assume that the Gaussian variances only vary along one direction (e.g. the propagation of light direction) [20, 21, 22].

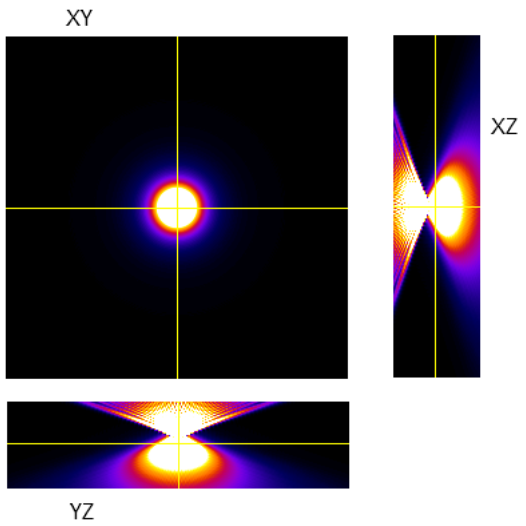


Figure 4: An orthogonal view of a Variable Refractive Index Gibson and Lanni PSF obtained with the PSF Generator [23]

4.4. Diagonal approximations of blurring operator in wavelet or wavelet packet bases

Some works [24, 25] proposed to approximate blurring operators $\mathbf{H}$ using operators diagonal in wavelet or wavelet packet bases. This idea consists in defining

an approximation $\widetilde{\mathbf{H}}$ of kind $\widetilde{\mathbf{H}} = \mathbf{\Psi}\mathbf{\Sigma}\mathbf{\Psi}^*$, where $\mathbf{\Psi}^*$ and $\mathbf{\Psi}$ are wavelet (packet) transforms and $\mathbf{\Sigma}$ is a diagonal matrix. The wavelet transforms might be redundant in order to ensure translation invariance. This diagonal approximation mimics the fact that shift-invariant operators are diagonal in the Fourier domain. These approaches lead to fast $\mathcal{O}(N^d)$ algorithms to compute matrix vector products. In [25], the authors proposed to deblur images using diagonal approximations of the blurring operators in redundant wavelet packet bases. This approximation was shown to be fast and efficient in deblurring images when the exact operator was scarcely known or in high noise levels. It is however too coarse for applications with low noise levels.

4.5. Piecewise convolutions

Probably the most commonly used approach is the piecewise convolution approximation of the kernel [4, 5, 26, 6]. Due to its wide use, we propose a detailed analysis of the method in the next section.

5. An analysis of piecewise convolutions algorithms

The main idea of this approach is to decompose the image domain into subregions and perform a convolution on each subregion. The results are then gathered together to obtain the blurred image. In its simplest form, this approach consists in partitioning the domain Ω in squares of equal sizes. More advanced strategies consist in decomposing the domain with overlapping subregions. The blurred image can then be obtained by using windowing functions that interpolate the kernel between subregions (see e.g. [4]).

In this section, we analyze this approach from a theoretical point of view: we derive upper and lower complexity bounds on the approximation error with respect to the subregions sizes.

5.1. Description of the method

From now on, we work in 1D, i.e. we assume that $d = 1$. We consider the simple case where the region Ω is partitioned into non-overlapping subregions. We decompose the segment $\Omega = [0, 1]$ in m subregions of length $1/m$. We denote the subregions $\omega_k = \left[\frac{k-1}{m}, \frac{k}{m}\right]$ for $k \in \{1, \dots, m\}$ (see Figure 5). The operator H is approximated by $\widetilde{H}_m$ defined by

$$\widetilde{H}_m = \sum_{k=1}^{m} \mathbb{1}_{\omega_k} \widetilde{H}^{(k)} \tag{11}$$

where $\widetilde{H}^{(k)}$ is a convolution operator representing the spatially invariant blur in the k-th subregion and

$$\mathbb{1}_{\omega_k}(x) = \begin{cases} 1 & \text{if } x \in \omega_k \\ 0 & \text{otherwise.} \end{cases}$$

The convolution kernel associated to $\widetilde{H}^{(k)}$ is set to $\widetilde{h}_k(y) = K(c_k, c_k - y)$ where c_k denotes the center of the set ω_k. This choice ensures that $\widetilde{H}_m u(c_k) = Hu(c_k)$.

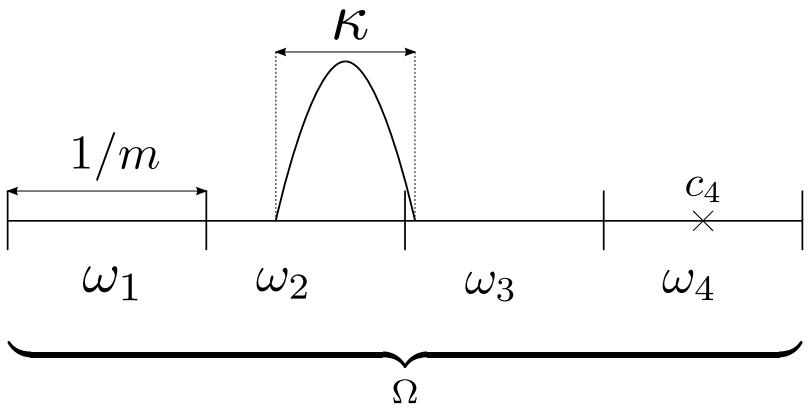


Figure 5: Illustration of the notations. The domain Ω is decomposed in $m = 4$ subregions of length $1/m$. The PSF support has a maximal width κ. The center of each subregion ω_k is denoted c_k.

The structure of the approximation operator allows to use Fast Fourier Transforms on each subregion and leads to fast matrix-vector products.

5.2. Theoretical analysis

We first study this approach in the continuous setting. We derive upper and lower bounds on the approximation error of the method defined as $\left\|H - \widetilde{H}_m\right\|_{p\to q}$. We examine the cases $p = q = 2$ and $p = 1$, $q = +\infty$. We assume that the kernel K is L-Lipschitz, i.e,

$$\forall (x_1, y_1), (x_2, y_2) \in \Omega, \quad |K(x_1, y_1) - K(x_2, y_2)| \leq L \left\|(x_1, y_1) - (x_2, y_2)\right\|_2 . \tag{12}$$

Using this assumption, we can derive an upper-bound controlling the approximation error.

Lemma 1 (Upper-Bounds)**.** *For all L-Lipschitz kernels K, the approximation errors are bounded above by,*

$$\left\|H - \widetilde{H}_m\right\|_{1\to\infty} \leq \frac{L}{2m}$$

and

$$\left\|H - \widetilde{H}_m\right\|_{2\to 2} \leq \frac{L}{2m}. \tag{13}$$

Proof. Proof in Appendix A. □

Remark 4. The same result holds if instead of the global Lipschitz assumption (12), we assume the less stringent

$$\forall (x,y) \in \Omega \times \Omega \text{ and for all } h \quad |K(x+h, y+h) - K(x,y)| \leq Lh.$$

This assumption basically indicates that the kernel K can be considered as a convolution locally since convolution kernels satisfy, $K(x+h, y+h) = K(x,y)$ for all x, y, h.

Lemma 2 indicates that the bounds in Lemma 1 are tight for all $\mathcal{C}^1$ kernels that are not space invariant.

Lemma 2 (Lower-Bounds)**.** *Let K be a $\mathcal{C}^1(\Omega \times \Omega)$ kernel of a space varying integral operator. Then there exists constants $c_1 > 0$ and $c_2 > 0$ such that,*

$$\left\|H - \widetilde{H}_m\right\|_{1\to\infty} \geq \frac{c_1}{m}$$

and

$$\left\|H - \widetilde{H}_m\right\|_{2\to 2} \geq \frac{c_2}{m}. \tag{14}$$

Proof. Proof in Appendix A. □

5.3. Numerical complexity

In this paragraph, we assume that the image consists of $N \in \mathbb{N}$ pixels. The PSF width in the discrete setting is thus approximately equal to κN. We let $\widetilde{\mathbf{H}}_m$ denote the discrete piecewise convolution approximation of $\mathbf{H}$. It is defined similarly to (11).

The complexity results of the piecewise convolution approach in the discrete setting is summarized in the following theorem:

Theorem 3. *Let K denote a Lipschitz kernel that is not a convolution. Let $\mathbf{H}$ denote the discretized operator defined in equation* (8)*. Let $\widetilde{\mathbf{H}}_m$ denote the discrete piecewise convolution approximation of $\mathbf{H}$. The following results hold:*

i) A product $\widetilde{\mathbf{H}}_m\mathbf{u}$ *implemented with FFTs with* $\mathbf{u} \in \mathbb{R}^N$ *is performed in a number of operations proportional to*

$$(N + \kappa N m) \log\left(\frac{N}{m} + \kappa N\right). \tag{15}$$

ii) For sufficiently large m *and* N *with* $m < N$*, there exists constants* $0 \leq c1 \leq c2$ *such that*

$$\left\|\mathbf{H} - \widetilde{\mathbf{H}}_m\right\|_{2\to 2} \leq \frac{c_2}{m} \tag{16}$$

and

$$\left\|\mathbf{H} - \widetilde{\mathbf{H}}_m\right\|_{2\to 2} \geq \frac{c_1}{m}. \tag{17}$$

iii) For sufficiently large $N \in \mathbb{N}$ *and sufficiently small* $\epsilon > 0$*, the number of operations required to obtain* $\|\mathbf{H} - \widetilde{\mathbf{H}}_m\|_{2\to 2} \leq \epsilon$ *is proportional to*

$$\frac{L\kappa N \log(\kappa N)}{\epsilon}. \tag{18}$$

Proof. Proof in Appendix A. □

One cannot hope to improve this complexity result due to Lemma 2. In particular, the method efficiency is insensitive to higher degrees of regularity of the kernel. It is possible to show that the method complexity in dimension d scales as $\mathcal{O}(\frac{Ld(\kappa N)^d \log(\kappa N)}{\epsilon})$ using similar arguments to those employed in the appendices.

Remark 5. Many authors proposed to decompose the domain into overlapping subregions and use linear interpolations of PSF on the overlapping domains. We do not analyze this technique in this paper. It is not clear yet whether it would improve the complexity result (18). Our numerical experiments do not show a significant improvement between piecewise convolutions with or without overlap.

Remark 6. The complexity analysis proposed above does not reflect an important property of fast discrete Fourier transforms: their efficiency depend greatly on the image size and is usually lower for sizes that have a prime factorization comprising only small primes (e.g. less than 7). This is illustrated in Figure 6. The time needed to compute an FFT on an image of size 499×499 is about 5 times higher than the time needed to compute the FFT on an image of size 512×512. This phenomenon explains the fact that the practical complexity of piecewise convolution algorithms may increase in a chaotic manner with respect to m.

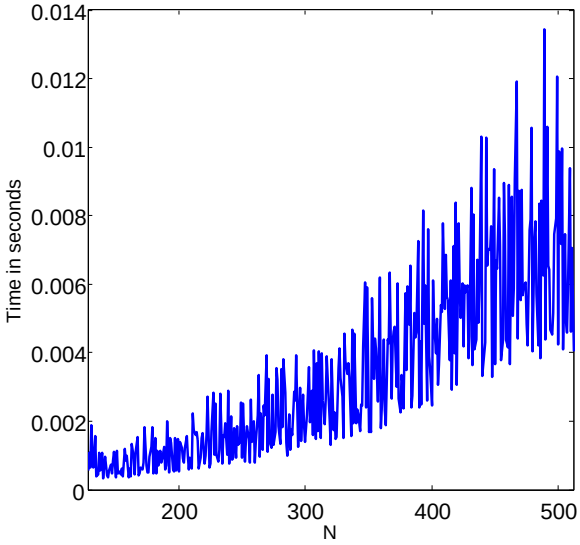


Figure 6: In this experiment, we evaluate the computing time of 2D FFTs applied to images of size $N \times N$. We use the FFTW in Matlab and do not include the time needed to compute the plans.

6. Wavelet representation of the blurring operator

In this section, we show that blurring operators can be well approximated by sparse representations in the wavelet domain. Since H is a linear operator in a Hilbert space, it can be written as $H = \Psi\Theta\Psi^*$, where $\Theta : l^2(\mathbb{Z}) \to l^2(\mathbb{Z})$ is the (infinite dimensional) matrix representation of the blur operator in the wavelet domain. Matrix Θ is characterized by the coefficients:

$$\theta_{j,m,k,n} = \langle H\psi_{j,m}, \psi_{k,n} \rangle , \qquad \forall j, m, k, n \in \mathbb{Z}. \tag{19}$$

In their seminal papers [3, 27, 2], Y. Meyer, R. Coifman, G. Beylkin and V. Rokhlin prove that the coefficients of Θ decrease fastly away from its diagonal. A natural approach to obtain fast numerical algorithms to approximate H thus consists in thresholding Θ to obtain a sparse matrix S.

This section is organized as follows: first we discuss the interest of approximating H in a wavelet basis rather than using the standard discretization described in paragraph 4.1. Second, we provide various theoretical results concerning the number of coefficients necessary to obtain an ϵ-approximation of H in different metrics.

6.1. Discretization of the operator by projection

The proposed method relies on a discretization of H different from that of paragraph 4.1. The main idea is to use a projection on a finite dimensional linear subspace $V_N = \mathrm{Span}(\psi_1, \ldots, \psi_N)$ of $\mathbb{L}^2(\Omega)$ where $(\psi_1, \psi_2, \ldots)$ is an orthonormal basis of $\mathbb{L}^2(\Omega)$. We define a projected operator H_N by $H_N u = P_{V_N} H P_{V_N} u$. We can associate an $N \times N$ matrix $\mathbf{\Theta}$ to this operator defined by $\mathbf{\Theta} = (\langle H\psi_i, \psi_j \rangle)_{1 \leq i,j \leq N}$.

It is very common in image processing to assume that natural images belong to functional spaces containing functions with some degree of regularity. For instance, images are well represented by bounded variation functions [28] (even though this assumption is only partially true [29]) or by functions belonging to fractional Sobolev spaces [30]. This hypothesis can be expressed by assuming that

$$\|u - P_{V_N} u\|_2 = \mathcal{O}(N^{-\alpha}) \tag{20}$$

for a certain $\alpha > 0$. For instance, in 1D, if $(\psi_1, \psi_2, \ldots)$ is a wavelet or a Fourier basis and $u \in H^1(\Omega)$ then $\alpha = 2$. For $u \in BV(\Omega)$ (the space of bounded variation functions), $\alpha = 1$ in 1D and $\alpha = 1/2$ in 2D [10, 31].

Moreover, if we assume that H is a regularizing operator, meaning that $\|Hu - P_{V_N} Hu\|_2 = \mathcal{O}(N^{-\beta})$ with $\beta \geq \alpha$ for all u satisfying (20), then we have:

$$\begin{aligned}
&\|Hu - H_N u\|_2 \\
&= \|Hu - P_{V_N} H(u + P_{V_N} u - u)\|_2 \\
&\leq \|Hu - P_{V_N} Hu\|_2 + \|P_{V_N} u - u\|_2 \\
&= \mathcal{O}(N^{-\alpha}).
\end{aligned}$$

This simple analysis shows that under mild assumptions, the Galerkin approximation of the operator converges and that the convergence rate can be controlled. The situation is not as easy for standard discretization using finite elements for instance (see e.g. [32, 33] where a value $\alpha = 1/6$ is obtained in 2D for BV functions, while the simple analysis above leads to $\alpha = 1/2$).

6.2. Discretization by projection on a wavelet basis

In this section, we set $l_0 = 0$ and thus assume that every $u \in \mathbb{L}^2([0,1])$ can be decomposed as

$$u = \langle u, \phi_0 \rangle \, \phi_0 + \sum_{j=0}^{+\infty} \sum_{m=0}^{2^j - 1} \langle u, \psi_{j,m} \rangle \, \psi_{j,m}.$$

In order to get a representation of the operator in a finite dimensional setting, we truncate the wavelet representation at scale J. This way, we obtain an operator

$\mathbf{H}$ acting on a space of dimension N, where $N = 1 + \sum_{j=0}^{J} 2^j$ denotes the numbers of wavelets kept to represent images. It can be written in the following convenient form:

$$\mathbf{H} = \mathbf{\Psi}\mathbf{\Theta}\mathbf{\Psi}^*. \tag{21}$$

In this equation, $\mathbf{\Psi} : \mathbb{R}^N \to \mathbb{R}^N$ is defined by

$$\mathbf{\Psi x} = x_0 \boldsymbol{\phi}_0 + \sum_{j=0}^{J} \sum_{m=0}^{2^j-1} x_{j,m} \boldsymbol{\psi}_{j,m},$$

$\mathbf{\Theta}$ is an $N \times N$ matrix which corresponds to a truncated version (also called finite section) of the matrix Θ defined in (19) and $\boldsymbol{\phi}_0$ and $\boldsymbol{\psi}_{j,m}$ are discrete wavelets.

6.3. Theoretical guarantees with sparse approximations

Sparse approximations of integral operators have been studied theoretically in [2, 3]. To the best of our knowledge, this approach was never applied to the approximation of spatially varying blur operators. This is surprizing since image processing is perhaps the domain where wavelets were the most successful. One exception is [24], where the authors show that wavelets can be useful to approximate foveation operators. This study was however limited to diagonal approximations which are too coarse to properly deblur images [25].

Let us recall a typical result that motivates the proposed approach. We stick to the one-dimensional case for the ease of exposition.

Lemma 4 (Decay of $\theta_{j,m,k,n}$ – [2, 3])**.** *Assume that H belongs to the Calderón-Zygmund class (see Definition 1). Assume that K belongs to $W^{M,\infty}(\Omega \times \Omega)$. Assume that the mother wavelet is compactly supported with M vanishing moments.*

Then, the coefficients of Θ satisfy the following inequality:

$$|\theta_{j,m,k,n}| \leq \gamma_{j,m,k,n} \tag{22}$$

where

$$\gamma_{j,m,k,n} = C_M 2^{-\left(M+\frac{1}{2}\right)|j-k|} \left(\frac{2^{-k} + 2^{-j}}{2^{-k} + 2^{-j} + |2^{-j}m - 2^{-k}n|} \right)^{M+1}. \tag{23}$$

and C_M is a constant that depends on the kernel regularity M, but not on j, k, m, n.

Proof. See Appendix C. □

Remark 7. The hypothesis $K \in W^{M,\infty}(\Omega \times \Omega)$ is by no means necessary. We made this assumption in order to make the proof arguments as simple as possible. In particular, the operator can have a singularity on its diagonal and have a much weaker Hölder regularity [27]. We refer to the pionneering book [27] and subsequent works [34, 35] for more refinements.

Lemma (22) basically indicates that the coefficients of $\mathbf{\Theta}$ decrease away from its diagonal. A more precise analysis (see Appendix D) allows to show the following proposition.

Proposition 1. *Matrix $\mathbf{\Theta}$ contains no more than $\mathcal{O}\left(N\eta^{-\frac{1}{M+1}}\right)$ coefficients above η.*

Said differently, $\mathbf{\Theta}$ can be approximated by a matrix $\mathbf{\Theta}_\eta$ satisfying $\|\mathbf{\Theta} - \mathbf{\Theta}_\eta\|_{1\to\infty} \leq \eta$ using only $\mathcal{O}\left(N\eta^{-\frac{1}{M+1}}\right)$ coefficients. This is a first result showing that blur operators can be highly compressed in the wavelet domain and that wavelets are capable of capturing the kernel regularity automatically.

The following theorem provides complexity results regarding the spectral norm

Theorem 5. *Let $\mathbf{\Theta}_\eta$ be the matrix obtained by zeroing all coefficients in $\mathbf{\Theta}$ such that*

$$\left(\frac{2^{-j}+2^{-k}}{2^{-j}+2^{-k}+|2^{-j}m-2^{-k}n|}\right)^{M+1} \leq \eta.$$

Let $\widetilde{\mathbf{H}}_\eta = \mathbf{\Psi}\mathbf{\Theta}_\eta\mathbf{\Psi}^*$ denote the resulting operator. The following results hold:

i) The number of non zero coefficients in $\mathbf{\Theta}_\eta$ is bounded above by

$$C'_M N \log_2(N)\, \eta^{-\frac{1}{M+1}} \tag{24}$$

where $C'_M > 0$ depends only on M.

ii) The approximation $\widetilde{\mathbf{H}}_\eta$ satisfies $\left\|\mathbf{H} - \widetilde{\mathbf{H}}_\eta\right\|_{2\to 2} \lesssim \eta^{\frac{M}{M+1}}$.

iii) The number of coefficients needed to satisfy $\left\|\mathbf{H} - \widetilde{\mathbf{H}}_\eta\right\|_{2\to 2} \leq \epsilon$ is therefore bounded above by

$$C''_M N \log_2(N)\, \epsilon^{-\frac{1}{M}} \tag{25}$$

where $C''_M > 0$ depends only on M.

Proof. See Appendix E. □

Remark 8. We presented a simple thresholding strategy that provides a good idea of the approximation abilities of the method. However, it is possible to use different thresholding strategies to discard the $\log_2(N)$ term. The matrix is then compressed to $\mathcal{O}(N\epsilon^{-\frac{1}{M}})$ coefficients while still satisfying $\left\|\mathbf{H} - \widetilde{\mathbf{H}}_\eta\right\|_{2\to 2} \lesssim \epsilon$. See for instance [35] for more advanced techniques. We do not present those schemes as controlling the spectral norm might not be really relevant in image processing. Let us mention that similar results can be obtained in dimension d using proofs very similar to those of the appendix. The method complexity in dimension d scales as $\mathcal{O}(N^d\epsilon^{-\frac{1}{M}})$.

Remark 9. There are a few differences making the wavelet approach more attractive than piecewise convolutions from a theoretical point of view:

- A discretization in the wavelet domain provides better theoretical guarantees than the standard quadrature rules (see section 6.1).
- A comparison between the upper-bound (25) and the bound (18) is instructive. In the piecewise convolution approach, there is no hope to obtain a better approximation rate than $\mathcal{O}\left(\frac{1}{m}\right)$ (see Proposition 14). For the wavelet approach, the situation is different: the method is capable of handling *automatically* the degree of smoothness of the integral kernel K since there is a dependency in $\epsilon^{-\frac{1}{M}}$ where M is the smoothness level of the integral operator.
- We will see in the next section that the method is quite versatile since different sparsity patterns can be chosen depending on the knowledge of the blur kernel and on the regularity of the signals that are to be processed.
- The method can also handle more general singular operators as was shown in the seminal papers [3, 27, 2].

7. How to define sparsity patterns?

A key step to control the approximation quality is the selection of the coefficients in the matrix Θ that should be kept. For instance, a simple thresholding of $\mathbf{\Theta}$ leads to sub-optimal and somewhat disappointing results. In this section we propose algorithms to select the most relevant coefficients for images belonging to functional spaces such as that of bounded variation functions. We study the case where $\mathbf{\Theta}$ is known completely and the case where only an upper-bound such as (22) is available.

7.1. Problem formalization

Let $\mathbf{H}$ be the $N \times N$ matrix defined in equation (21). We wish to approximate $\mathbf{H}$ by a matrix $\widetilde{\mathbf{H}}_K$ of kind $\mathbf{\Psi}\mathbf{S}_K\mathbf{\Psi}^*$ where $\mathbf{S}_K$ is a matrix with at most K non zero coefficients. Let $\mathbb{S}_K$ denote the space of $N \times N$ matrices with at most K non zeros coefficients. The problem we address in this paragraph reads

$$\begin{aligned}&\min_{\mathbf{S}_K \in \mathbb{S}_K} \|\mathbf{H} - \widetilde{\mathbf{H}}_K\|_{X \to 2} \\ &= \min_{\mathbf{S}_K \in \mathbb{S}_K} \max_{\|\mathbf{u}\|_X \leq 1} \|\mathbf{H}\mathbf{u} - \mathbf{\Psi}\mathbf{S}_K\mathbf{\Psi}^*\mathbf{u}\|_2.\end{aligned}$$

The solution of this problem provides the best K-sparse matrix $\mathbf{S}_K$, in the sense that no other choice provides a better SNR uniformly on the unit ball $\{\mathbf{u} \in \mathbb{R}^N, \|\mathbf{u}\|_X \leq 1\}$.

The norm $\|\cdot\|_X$ should be chosen depending on the type of images that have to be blurred. For instance, it is well-known that natural images are highly compressible in the wavelet domain. Therefore, a natural choice could be to set $\|\mathbf{u}\|_X = \|\mathbf{\Psi}^*\mathbf{u}\|_1$. This choice will ensure a good reconstruction of images that have a wavelet decomposition with a low ℓ^1-norm.

Another very common assumption in image processing is that images have a bounded total variation. Functions in $BV(\Omega)$ can be characterized by their wavelet coefficients [31]. For instance, if $u \in BV([0,1])$, then

$$|\langle u, \phi_0 \rangle| + \sum_{j=0}^{+\infty} \sum_{m=0}^{2^j - 1} 2^j \, |\langle u, \psi_{j,m} \rangle| < +\infty \tag{26}$$

for all wavelet bases. This result motivated us to consider norms defined by

$$\|\mathbf{u}\|_X = \|\mathbf{\Sigma}\mathbf{\Psi}^*\mathbf{u}\|_1$$

where $\mathbf{\Sigma} = diag(\sigma_1, \ldots, \sigma_N)$ is a diagonal matrix. Depending on the regularity level of the images considered, different diagonal coefficients can be used. For instance, for BV images in 1D, one could set $\sigma_i = 2^{j(i)}$ where $j(i)$ is the scale of the i-th wavelet, owing to (26).

We can now take advantage of the fact that images and operators are sparse in the same wavelet basis. Let $\mathbf{z} = \mathbf{\Psi}^*\mathbf{u}$ and $\mathbf{\Delta} = \mathbf{\Theta} - \mathbf{S}_K$. Since we consider orthogonal wavelet transforms:

$$\begin{aligned}\left\|\mathbf{H} - \widetilde{\mathbf{H}}_K\right\|_{X \to 2} &= \max_{\|\mathbf{u}\|_X \leq 1} \|\mathbf{\Psi}(\mathbf{\Theta} - \mathbf{S}_K)\mathbf{\Psi}^*\mathbf{u}\|_2 \\ &= \max_{\|\mathbf{\Sigma}\mathbf{z}\|_1 \leq 1} \|(\mathbf{\Theta} - \mathbf{S}_K)\mathbf{z}\|_2\end{aligned}$$

$$= \max_{\|\mathbf{z}\|_1 \leq 1} \left\| \mathbf{\Delta} \mathbf{\Sigma}^{-1} \mathbf{z} \right\|_2 .$$

By letting $\mathbf{\Delta}^{(i)}$ denote the i-th column of $\mathbf{\Delta}$, we finally get the following simple expression for the operator norm:

$$\left\| \mathbf{H} - \widetilde{\mathbf{H}} \right\|_{X \to 2} = \max_{1 \leq i \leq N} \frac{1}{\sigma_i} \left\| \mathbf{\Delta}^{(i)} \right\|_2 . \tag{27}$$

Our goal is thus to find the solution of:

$$\min_{\mathbf{S}_K \in \mathbb{S}_K} \max_{1 \leq i \leq N} \frac{1}{\sigma_i} \left\| \mathbf{\Delta}^{(i)} \right\|_2 . \tag{28}$$

7.2. An algorithm when $\mathbf{\Theta}$ is known

Finding the minimizer of problem (28) can be achieved using a simple greedy algorithm: the matrix $\mathbf{S}_{k+1}$ is obtained by adding the largest coefficient of the column Δ_i with largest Euclidean norm to $\mathbf{S}_k$. This procedure can be implemented efficiently by using quick sort algorithms. The complete procedure is described in Algorithm 1. The overall complexity of this algorithm is $\mathcal{O}(N^2 \log(N))$. The most computationally intensive step is the sorting procedure in the initialisation. The loop on k can be accelerated by first sorting the set $(\gamma_j)_{1 \leq j \leq N}$, but the algorithm's complexity remains essentially unchanged.

7.3. An algorithm when $\mathbf{\Theta}$ is unknown

In the previous paragraph, we assumed that the full matrix $\mathbf{\Theta}$ was known. There are at least two reasons that make this assumption irrelevant. First, computing $\mathbf{\Theta}$ is very computationally intensive and it is not even possible to store this matrix in RAM for medium sized images (e.g. 512×512). Second, in blind deblurring problems, the operator $\mathbf{H}$ needs to be inferred from the data and adding priors on the sparsity pattern of $\mathbf{S}_K$ might be an efficient choice to improve the problem identifiability.

When $\mathbf{\Theta}$ is unknown, we may take advantage of equation (22) to define sparsity patterns. A naive approach would consist in applying Algorithm (1) directly on the upper enveloppe $\mathbf{\Gamma} = (\gamma_{j,m,k,n})_{j,m,k,n}$. However, this matrix cannot be stored and this approach is applicable only for small images. In order to reduce the computational burden, one may take advantage of the special structure of the upper-bound: equation (22) indicates that the coefficients $\theta_{j,m,k,n}$ can be discarded for sufficiently large $|j-k|$ and $\left|2^{-j}m - 2^{-k}n\right|$. The quantity $|2^{-j}m - 2^{-k}n| \in [0,1]$ measures the shift between wavelets at different scales. Equation (22) thus means that for a given wavelet $\psi_{j,m}$, only its spatial neighbours in neighbouring scales have significant correlation coefficients $\langle H\psi_{j,m}, \psi_{k,n} \rangle$. We may thus construct sparsity patterns using the notion of multiscale neighbourhoods defined below.

Algorithm 1: An algorithm to find the minimizer of (28).

Input:
$\mathbf{\Theta}$: $N \times N$ matrix;
$\mathbf{\Sigma}$: Diagonal matrix;
K: the number of elements in the thresholded matrix;
Output:
$\mathbf{S}_K$: Matrix minimizing (28)
Initialization:
Set $\mathbf{S}_K = 0 \in \mathbb{R}^{N \times N}$;
Sort the coefficients of each column $\mathbf{\Theta}^{(j)}$ of $\mathbf{\Theta}$ in decreasing order;
The sorted columns $\mathbf{\Theta}^{(j),S}$ and index set I_j satisfy $\mathbf{\Theta}^{(j),S}(i) = \mathbf{\Theta}^{(j)}(I_j(i))$;
Compute the norms $\gamma_j = \frac{\|\mathbf{\Theta}^{(j)}\|_2^2}{\sigma_j^2}$;
Define $\mathbf{O} = (1, \ldots, 1) \in \mathbb{R}^N$;
$\mathbf{O}(j)$ is the index of the largest coefficient in $\Theta^{(j),S}$ not yet added to $\mathbf{S}_K$;
begin
  for $k = 1$ **to** K **do**
    Find $l = \arg\max_{j=1\ldots N} \gamma_j$;
    Update $\gamma_l = \gamma_l - \left(\frac{\Theta^{(l),S}(\mathbf{O}(l))}{\sigma_l} \right)^2$;
    Set $\mathbf{O}(l) = \mathbf{O}(l) + 1$;
    Set $\mathbf{S}_K(I_l(\mathbf{O}(l)), l) = \mathbf{\Theta}(I_l(\mathbf{O}(l)), l)$;
  end
end

Definition 3 (Multiscale shift). The multiscale shift $s \in \mathbb{Z}$ between two wavelets $\psi_{j,m}$ and $\psi_{k,m}$ is defined by

$$s = \left\lfloor \frac{n}{2^{\max(k-j,0)}} \right\rfloor - \left\lfloor \frac{m}{2^{\max(j-k,0)}} \right\rfloor. \tag{29}$$

The divisions by $2^{\max(k-j,0)}$ and $2^{\max(j-k,0)}$ allow to rescale the shifts at the coarsest level.

Definition 4 (Multiscale neighborhood). Let $\boldsymbol{\mathcal{N}}$ denote the set of all neighborhood relationships, i.e. the set of all possible couples of type (scale, (scale,shift)). A multiscale neigborhood $\mathcal{N}$ is an element of the powerset $\mathcal{P}(\boldsymbol{\mathcal{N}})$. Given a multiscale neigborhood $\mathcal{N}$, two wavelets $\psi_{j,m}$ and $\psi_{k,n}$ will be said to be $\mathcal{N}$-neighbors if $(j,(k,s)) \in \mathcal{N}$ where s is defined in equation (29).

This definition is illustrated in Figure 7.

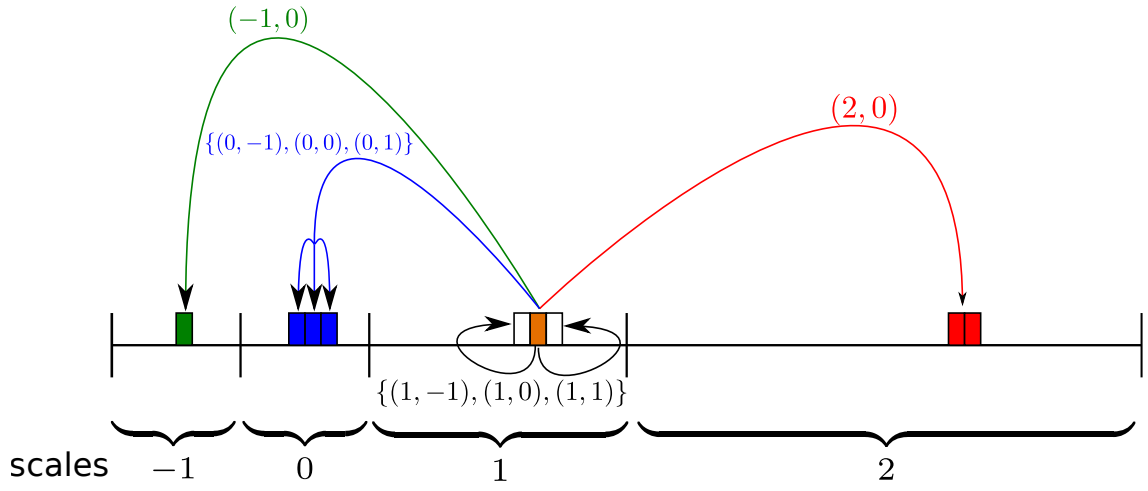


Figure 7: Illustration of a multiscale neighborhood on a 1D signal. In this example, the neighborhood at scale 1 is $\mathcal{N}(1) = \{(-1,0),(0,-1),(0,0),(0,1),(1,-1),(1,0),(1,1),(2,0)\}$. Notice that the two red wavelets at scale 2 are neighbors of the orange wavelet at scale 1 and that this relationship is described through only one shift.

The problem of finding a sparsity pattern is now reduced to finding a good multiscale neighborhood. In what follows, we let $\boldsymbol{\mathcal{N}}(j)$ denote the set of all possible

neighborhood relationships at scale j. Let $\mathcal{N} \in \mathcal{P}(\boldsymbol{\mathcal{N}})$ denote a multiscale neighborhood. We define the matrix $\mathbf{S}_{\mathcal{N}}$ as follows:

$$\mathbf{S}_{\mathcal{N}}(j,m,k,n) = \begin{cases} \theta_{j,m,k,n} & \text{if } \psi_{j,m} \text{ is an } \mathcal{N}\text{-neighbor of } \psi_{k,n} \\ 0 & \text{otherwise.} \end{cases}$$

Equation (22) indicates that

$$|\theta_{j,m,k,n}| \leq u(j,k,s)$$

with

$$u(j,k,s) = C_M 2^{-\left(M+\frac{1}{2}\right)|j-k|} \left(\frac{2^{-k}+2^{-j}}{2^{-k}+2^{-j}+2^{-\min(j,k)}\,|s|} \right)^{M+1}.$$

and where s is defined in (29). Let $\mathbf{U}$ be the matrix defined by $\mathbf{U}(j,m,k,n) = u(j,k,s)$. Finding a good sparsity pattern can now be achieved by solving the following problem:

$$\min_{\mathcal{N}\in\mathcal{P}(\boldsymbol{\mathcal{N}}),|\mathcal{N}|=K} \max_{1\leq i\leq N} \frac{1}{\sigma_i} \left\| (\mathbf{U}-\mathbf{S}_{\mathcal{N}})^{(i)} \right\|_2 \tag{30}$$

where $(\mathbf{U}-\mathbf{S}_{\mathcal{N}})^{(i)}$ denotes the i-th column of $(\mathbf{U}-\mathbf{S}_{\mathcal{N}})$.

In what follows, we assume that σ_i only depends on the scale $j(i)$ of the i-th wavelet. Similarly to the previous section, finding the optimal sparsity pattern can be performed using a greedy algorithm. A multiscale neighborhood is constructed by iteratively adding the couple (scale, (scale,shift)) that minimizes a residual. This technique is described in Algorithm 2.

Note that the norms γ_k only depend on the scale $j(k)$, so that the initialisation step only requires $\mathcal{O}(N\log_2(N))$ operations. Similarly to Algorithm 1, this algorithm can be accelerated by first sorting the elements of $u(j,k,s)$ in decreasing order. The overall complexity for this algorithm is $\mathcal{O}(N\log(N)^2)$ operations.

7.4. The algorithm in higher dimension

The algorithm described in the previous section can also be applied in 2D with minor modifications. For each node of the wavelet decomposition quadtree, a neighbourhood is defined. It describes which correlation coefficients shall be preserved to generate the sparsity pattern. This principle is illustrated on Figure 8a. In this example, we consider a wavelet transform of depth 2. A neighbourhood $\mathcal{N}(i)$ is associated to each sub-band. For instance $\mathcal{N}(1)$ can be represented as follows:

$$\mathcal{N}(1) = \left(\begin{array}{cccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ l & l & l & l & l & h & v & d \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \end{array}\right) \begin{array}{l} \text{scale} \\ \text{orientation} \\ \text{vertical translation} \\ \text{horizontal translation} \end{array}$$

Algorithm 2: An algorithm to find the minimizer of (30).

Input:
u: Upper-bound ;
$\mathbf{\Sigma}$: Diagonal matrix;
K: the number of elements of the neighborhood;
Output:
$\mathcal{N}$: multiscale neighborhood minimizing (30)
Initialization:
Set $\mathcal{N} = \emptyset$;
Compute the norms $\gamma_k = \frac{\|\mathbf{U}^{(k)}\|_2^2}{\sigma_k^2}$;
begin
  for $k = 1$ **to** K **do**
    Find $j^* = \underset{j=1...N}{\arg\max}\, \gamma_j$;
    (The column with largest norm)
    Find $(k^*, s^*) = \underset{(k,s)\in\boldsymbol{\mathcal{N}}(j^*)}{\arg\max}\ u^2(j^*, k, s)2^{\max(j^*-k,0)}$;
    (The best scale and shift for this column is (k^*, s^*))
    (The number of elements in the neighborhood relationship $(j^*, (k, s))$ is $2^{\max(j^*-k,0)}$)
    Update $\mathcal{N} = \mathcal{N} \cup \{(j^*, (k^*, s^*))\}$;
    Set $\gamma_k = \gamma_k - u^2(j^*, k^*, s^*) \cdot 2^{\max(j^*-k,0)}$
  end
end

where l stands for the low frequency wavelet, h, v, d for the horizontal, vertical and diagonal orientations respectively. Figure 8b illustrates the neighbourhood $\mathcal{N}(1)$ for a given wavelet at the center of the image.

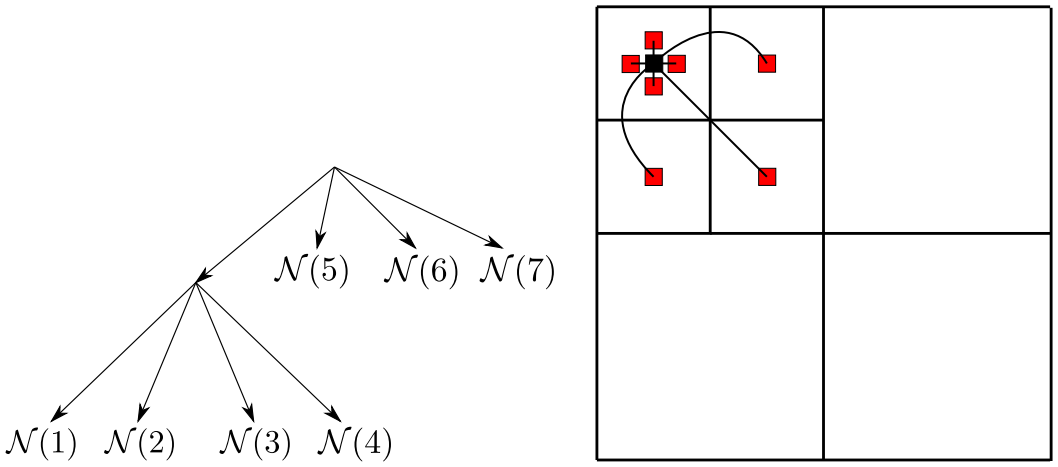


Figure 8: Illustration of the neighbourhood $\mathcal{N}(1)$ in the case of a wavelet transform of depth 2. (8a) The quadtree corresponding to a wavelet transform of depth 2. (8b) An illustration of the neighbourhood $\mathcal{N}(1)$ in the wavelet coefficients representation. In black, the given wavelet at the center of the image, and in red all the preserved neighbouring wavelets.

The same algorithm as that described in the previous section can be used if an upper-bound u of kind $|\theta_{j_1,o_1,m_1,j_2,o_2,m_2}| \leq u(j_1, o_1, j_2, o_2, s)$ provided. In this equation o_1 and o_2 are parameters describing the wavelet orientation and s is the 2-dimensional shift between wavelets at different scale. Using arguments similar to those of Appendix C, it is possible to obtain the following upper-bound:

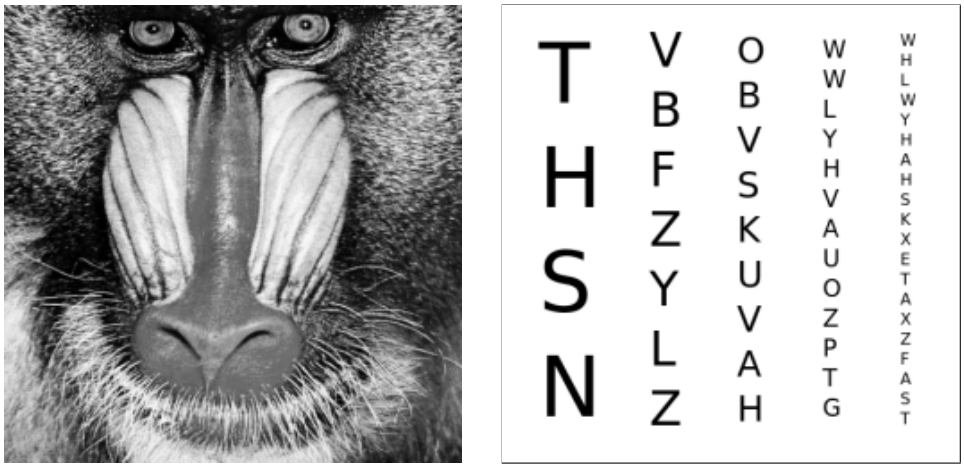
$$|\theta_{j_1,o_1,m_1,j_2,o_2,m_2}| \leq C_M 2^{-\left(M+\frac{d}{2}\right)|j_1-j_2|} \left(\frac{2^{-j_1} + 2^{-j_2}}{2^{-j_1} + 2^{-j_2} + |2^{-j_1} m_1 - 2^{-j_2} m_2|} \right)^{M+d}. \quad (31)$$

8. Numerical experiments

In this section we perform various numerical experiments in order to evaluate the practical efficiency of wavelet based methods. We also perform comparisons between the piecewise convolution approach and the wavelet based approach. We first evaluate the method's efficiency on the direct problem:

- we analyze $\left\|\mathbf{H} - \widetilde{\mathbf{H}}\right\|$ for various operator norms.
- we compare $\mathbf{Hu}$ and $\widetilde{\mathbf{H}}\mathbf{u}$ for real images $\mathbf{u}$.

We then study the performance of the different methods for deblurring problems. Two different blur kernels and two images will be considered, see Figures 9 and 10. The images are rescaled in $[0, 1]$ to ease the visualization of residuals. Due to memory limitations, we only consider images of size $N = 256 \times 256$. Note that a full matrix of size N^2 stored in double precision weighs around 32 gigabytes.



(a) Mandrill (b) Letters

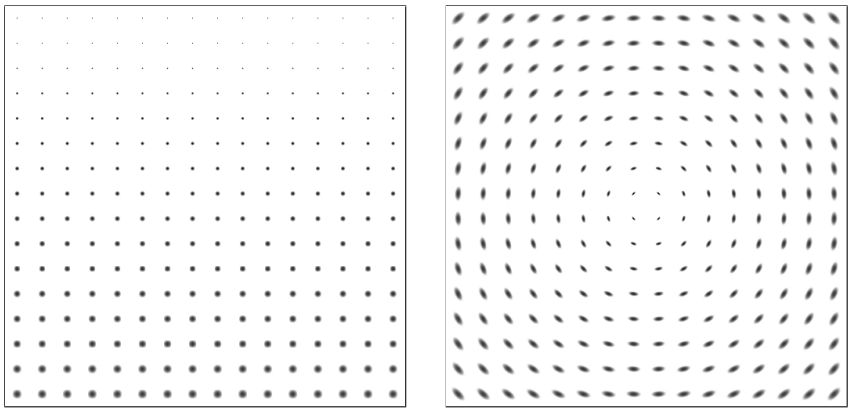
Figure 9: The two images of size 256×256 used in these numerical experiments

8.1. Computation of the full $\mathbf{\Theta}$ matrix

Before applying our approximation methods, matrix $\mathbf{\Theta}$ needs to be computed explicitly. The coefficients $\langle H\psi_{j,m}, \psi_{k,n}\rangle$ are approximated by their discrete counterparts. If $\boldsymbol{\psi}_{j,m}$ and $\boldsymbol{\psi}_{k,n}$ denote discrete wavelets, we simply compute the wavelet transform of $\mathbf{H}\psi_{j,m}$ and store it into the (j, m)-th column of $\mathbf{\Theta}$. This computation scheme is summarized in Algorithm 3. This algorithm corresponds to the use of rectangle methods to evaluate the dot-products:

$$\int_\Omega \int_\Omega K(x, y)\psi_{j,m}(y)\psi_{k,n}(x)dydx \simeq \frac{1}{N^{2d}} \sum_{x\in X} \sum_{y\in X} K(x, y)\psi_{j,m}(y)\psi_{k,n}(x). \tag{32}$$

In all our numerical experiments, we used Daubechies wavelets with 10 vanishing moments decomposed at the fourth level. We made several tests not reported here



(a) (b)

Figure 10: The PSFs associated to the blur kernels. The PSFs in Figure (10a) are Gaussians with equal variances increasing in the vertical direction. The PSFs in Figure (10b) are anisotropic Gaussians with covariance matrices that depend on the polar coordinates.

to find the best combination. It appears that for the considered operators, using as many vanishing moments as possible was preferable. Using more than 10 vanishing moments however led to insignificant performance increase while making the numerical complexity higher.

8.2. Evaluation of $\left\|\mathbf{H}-\widetilde{\mathbf{H}}\right\|_{2\to 2}$ *with a simple thresholding strategy*

In this first numerical experiment, we evaluate $\left\|\mathbf{H}-\widetilde{\mathbf{H}}\right\|_{2\to 2}$ where $\widetilde{\mathbf{H}}$ is obtained by piecewise convolutions or sparse approximations in the wavelet domain.

The sparse approximation of the operator is constructed by thresholding the matrix $\mathbf{\Theta}$ in order to keep the K largest coefficients. The values $K = 2^l \times N^2$ with $l \in \{0 \dots 2\log_2 N\}$ have been chosen. This way K is a multiple of the number of pixels in the image. The piecewise convolution approximation is constructed by partitioning the image into $2^l \times 2^l$ sub-images where $l \in \{0 \dots \log_2 N\}$. We also studied the case where sub-images overlap and linearly interpolated the blur between sub-images as proposed in [4, 6]. The overlap has been fixed to 50% of the sub-images sizes.

For each sub-image size, and each overlap, the norm $\left\|\mathbf{H}-\widetilde{\mathbf{H}}\right\|_{2\to 2}$ is approximated using a power method [36]. We stop the iterative process when the difference between the eigenvalues of two successive iterations is smaller than $10^{-8}\|\mathbf{H}\|_{2\to 2}$. The number of operations associated to each type of approximation is computed

Algorithm 3: An algorithm to compute $\boldsymbol{\Theta}$

Output:
$\boldsymbol{\Theta}$: the full matrix of $\mathbf{H}$
begin
  forall the (j,m) **do**
    Compute the wavelet $\boldsymbol{\psi}_{j,m}$ using an inverse wavelet transform
    Compute the blurred wavelet $\mathbf{H}\boldsymbol{\psi}_{j,m}$
    Compute $\left(\left\langle \mathbf{H}\boldsymbol{\psi}_{j,m}, \boldsymbol{\psi}_{k,n}\right\rangle\right)_{k,n}$ using one forward wavelet transform
    Set $\left(\left\langle \mathbf{H}\boldsymbol{\psi}_{j,m}, \boldsymbol{\psi}_{k,n}\right\rangle\right)_{k,n}$ in the (j,m)-th column of $\boldsymbol{\Theta}$.
  end
end

using theoretical complexities. For sparse matrix-vector product the number of operations is proportional to the number of non-zero coefficients in the matrix. For piecewise convolutions, the number of operations is proportional to the number of windows $(2^l \times 2^l)$ multiplied by the cost of a discrete convolution over a window $\left(\frac{N}{2^l} + N\kappa\right)^2 \log_2\left(\frac{N}{2^l} + N\kappa\right)$. [1]

Figure 11 shows the results of this experiment. The wavelet based method seems to perform much better than piecewise convolutions. The total number of operations is reduced by a factor roughly equal to 100 for a precision of 0.1.

8.3. Numerical comparison of different sparsity patterns.

In this numerical experiment, we obtain a K-sparse matrix $\boldsymbol{\Theta}_K$ using either a simple thresholding strategy or Algorithm 1. We evaluate the error $\left\|\mathbf{H} - \widetilde{\mathbf{H}}\right\|_{X\to 2}$ defined in (27) for both methods. In this experiment, we set $\sigma_i = 2^{j(i)}$. It is readily seen from Figure 12 that Algorithm provides a much better error decay for both operators.

8.4. Quality of matrix vector products for real images

In this experiment, we compare $\widetilde{\mathbf{H}}\mathbf{u}$ to $\mathbf{H}\mathbf{u}$, where $\mathbf{u}$ is the image in Figure 9b and where $\widetilde{\mathbf{H}}$ is obtained either by piecewise convolutions or by sparse wavelet

[1] We also carried out numerical computations of the operator norm error when images are supposed to be in Sobolev spaces $H^s(\Omega)$ for different s. However, since the results do not differ too much, we do not present them.

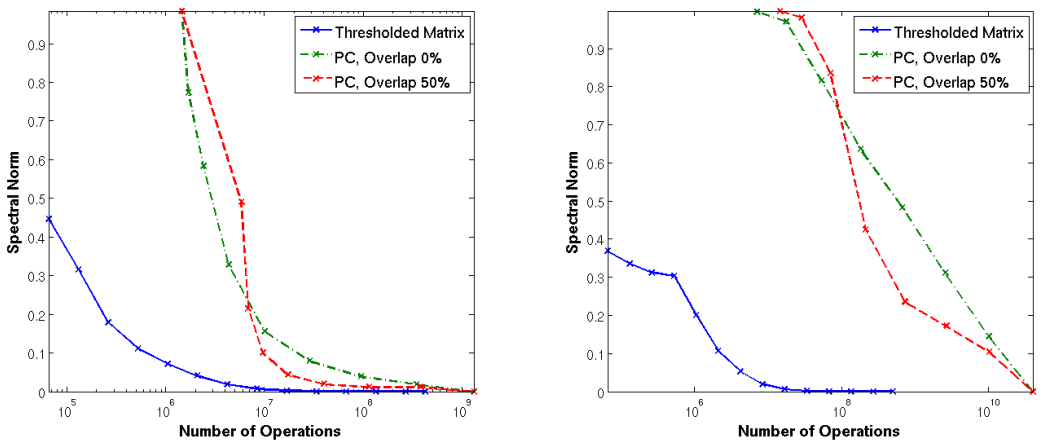


Figure 11: The operator norms $\left\|\mathbf{H}-\widetilde{\mathbf{H}}\right\|_{2\to 2}$ are displayed for the three proposed kernels. (From left to right, kernels corresponding to Figures 10a and 10b) and with respect to the number of operations needed to compute $\widetilde{\mathbf{H}}\mathbf{u}$. Plots are diplayed in a semilogx scale.

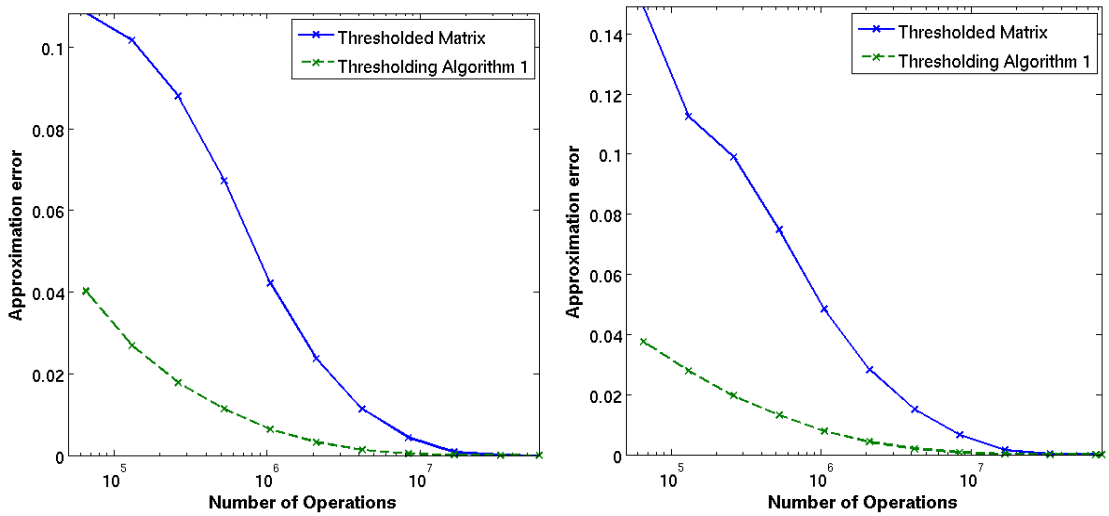


Figure 12: The operator norms $\left\|\mathbf{H}-\widetilde{\mathbf{H}}\right\|_{X\to 2}$ are displayed for kernels Figure 10a (left) and Figure 10b (right); and with respect to the number of operations needed to compute $\widetilde{\mathbf{H}}u$. Plots are diplayed in a semilogx scale. Daubechies wavelets with 10 vanishing moments have been used.

approximations. We plot the pSNR between the exact blurred image $\mathbf{Hu}$ and the blurred image using the approximated operator $\widetilde{\mathbf{H}}\mathbf{u}$. The two methods are tested with the following choices:

- We only test the piecewise-convolution with overlap since it produces better pSNR.
- Different sparsity patterns are tested. The first one is obtained by thresholding of $\mathbf{\Theta}$. The second one is obtained using Algorithm 1. We test three different diagonal matrices $\mathbf{\Sigma} = diag(\sigma_1, \ldots, \sigma_{N^2})$:
  1. $\mathbf{\Sigma}$ contains only ones on its diagonal: $\sigma_i = 1, \ \forall i$.
  2. σ is piecewise constant with a value increasing with respect to the scale of the wavelet coefficients: $\sigma_i = 2^{j(i)}$.
  3. $\sigma_i = 2^{j(i)}$, and $\sigma_i = \frac{1}{2}$ for indices i corresponding to the low frequency sub-band.

  The third one is obtained using Algorithm 2. The algorithm finds multi-scale neighbourhoods until $K = l \times N^2$ coefficients populate the matrix, with $l \in \{1, \ldots, 200\}$.

  We also performed experiments with $\sigma_i = 4^{j(i)}$, which corresponds to the assumption that images belong to $BV(\Omega)$. We do not report the results since they were not good.

The results of this experiment are displayed in Figure 13 for the two kernels from Figures 10b and 10a. Approximations built with Algorithm 1 and with $\sigma_i = 2^{j(i)}$ increase the pSNR of the blurred image $\widetilde{\mathbf{H}}\mathbf{u}$ by almost 10dBs compared to the naive thresholding approaches. This experiment highlights the relevance of Algorithm 1, used with the second and third $\mathbf{\Sigma}$ matrices. It allows to construct accurate sparse approximations of operators when applied to images in $BV(\Omega)$.

The piecewise convolution approach performs slightly better than wavelet based methods for the simple kernel 10a on a small intervall. It corresponds to the partitioning of Ω in 16×16 and 32×32 sub-windows (i.e. sub-windows of size 16×16 or 8×8 pixels). However, for more complex kernels, wavelet methods perform better.

This experiment also show the qualities and limits of the "blind" Algorithm 2. In this algorithm, the structure of the approximating matrix in deduced from the upper-bound (31). Matrices constructed using Algorithm 2 perform similarly to Algorithm 1 (that has a full knowledge of $\mathbf{\Theta}$) up to approximately $K = 30 \times N^2$ coefficients. Above this number, the approximation quality increases very slowly. This is probably due to the fact that the upper-bound (31) is too rough: the operator

might be much sparser than what is predicted by the theory. We will see that in deblurring applications, an approximation made of $K = 30 \times N^2$ non zero coefficients is more than enough.

Figure 17 shows the sparsity patterns of matrices obtained with Algorithms 1 and 2 for $K = 30N^2$ and $K = 128N^2$ coefficients. It is readily seen that the sparsity patterns look very similar and tend to confirm the soundness of Algorithm 2.

Finally, for kernel Figure 10b, we show blurred images $\widetilde{\mathbf{H}}\mathbf{u}$ in Figures 14 and 15 for the different sparsity patterns. Figure 14 shows the blurred images $\widetilde{\mathbf{H}}\mathbf{u}$ obtained with Algorithm 1 and with the three different $\mathbf{\Sigma}$ matrices. Figure 15 displays the blurred images $\widetilde{\mathbf{H}}\mathbf{u}$ obtained with the simple thresholding scheme and Algorithm 2. Finally, Figure 16 provides a comparison of the piecewise convolution approach and the wavelet based approach in terms of approximation quality and computing times. The following conclusions can be drawn from this experiment:

- The residual artefacts appearing in the piecewise convolution and wavelet based approach are different. They are localized at the interfaces between sub-images for the piecewise convolution approach while they span the whole image domain for the wavelet based approach. It is likely that using translation and/or rotation invariant wavelet would improve substancially the reconstruction.

- The approximation using the third $\mathbf{\Sigma}$ matrix produces the best results and should be preferred over more simple approaches.

- The sparsity pattern obtained using Algorithm 2 suffers from more artifacts than the other approaches. The quality is however acceptable from a visual point of view.

- In our implementation, the piecewise convolution approach (implemented in C) is outperformed by the wavelet based method (implemented in Matlab with C-mex files). For instance, for a precision of 45dBs, the wavelet based approach is about 10 times faster. Note that no method is multi-threaded, so that the comparison seems fair.

- The computing time of 1.21 seconds for the piecewise convolution approach with a 2×2 partition might look awkward since the computing times are significantly lower for finer partitions. This can be explained by Remark 6. In our numerical experiments we implemented two versions of the piecewise convolution approach: one based on the FFTW `http://www.fftw.org/` and the other based on Kiss FFT `http://sourceforge.net/projects/kissfft/`. The latter revealed to provide lower computing times (probably to the fact

that no plan is pre-computed) and we thus reported computing times using this method.

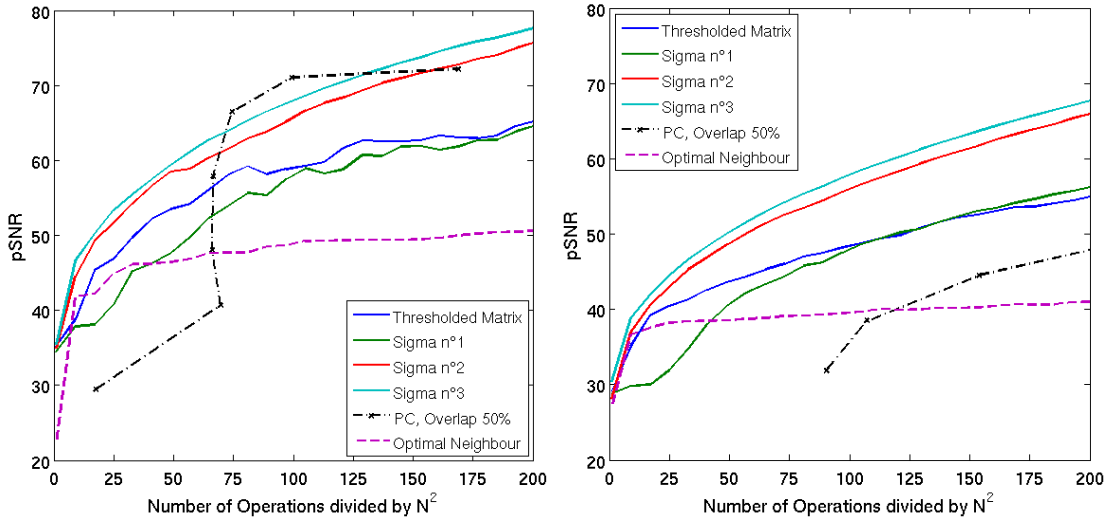


Figure 13: pSNR of the blurred image using the approximated operators $\widetilde{\mathbf{H}}\mathbf{u}$ with respect to the blurred image using the exact operator $\mathbf{H}\mathbf{u}$. The results have been obtained using the letters image Figure 9b.

8.5. Deblurring problem

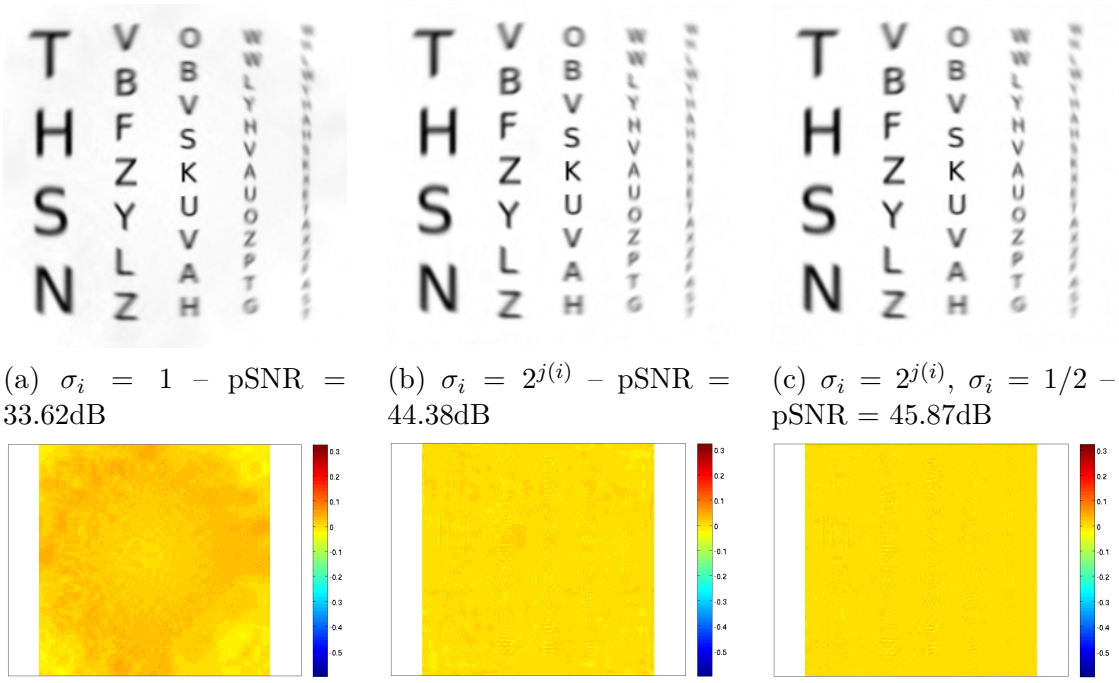
In this experiment we compare the methods efficiency in deblurring problems. We assume the following classical image degradation model

$$\mathbf{v} = \mathbf{H}\mathbf{u} + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}\left(0, \sigma^2 \mathrm{Id}\right), \tag{33}$$

where $\mathbf{v}$ is the degraded image observed, $\mathbf{u}$ is the image to restore, $\mathbf{H}$ in the blurring operator and σ^2 is the noise variance. A standard TV-L2 optimization problem is solved to restore the image $\mathbf{u}$:

$$\text{Find } \mathbf{u}^* \in \underset{\mathbf{u}\in\mathbb{R}^{Nd}, \left\|\widetilde{\mathbf{H}}\mathbf{u}-\mathbf{v}\right\|_2^2 \leq \alpha}{\arg\min} TV(\mathbf{u}), \tag{34}$$

where $\widetilde{\mathbf{H}}$ is an approximating operator and TV is the isotropic total variation of $\mathbf{u}$. The optimization problem is solved using the primal-dual algorithm proposed in [37]. We do not detail the resolution method since it is now well documented in the literature.

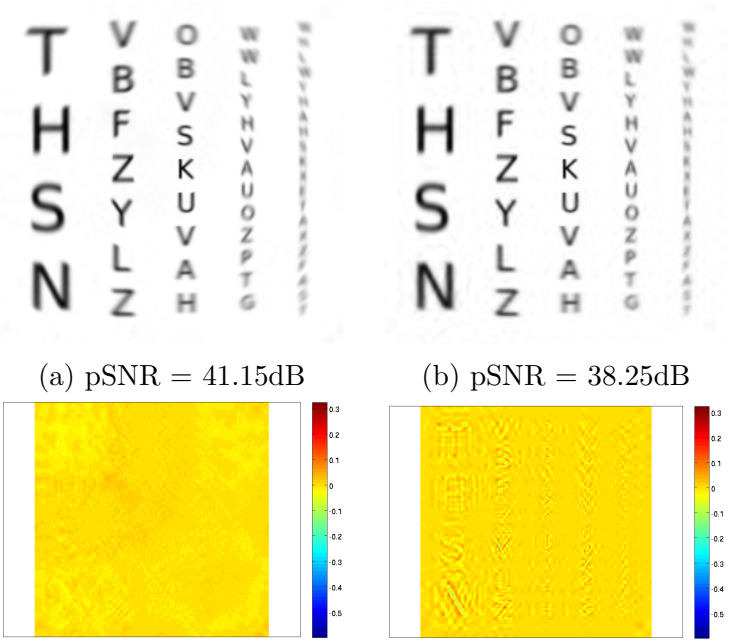


(a) $\sigma_i = 1$ – pSNR = 33.62dB

(b) $\sigma_i = 2^{j(i)}$ – pSNR = 44.38dB

(c) $\sigma_i = 2^{j(i)}$, $\sigma_i = 1/2$ – pSNR = 45.87dB

Figure 14: Blurred images using matrices formed with Algorithm 1 and for the kernel Figure 10b. $K = 30N^2$ coefficients are kept in the matrices. Algorithm 1 has been applied with the three different $\mathbf{\Sigma}$ matrices. The blurred images $\widetilde{\mathbf{H}}\mathbf{u}$ are shown on top. The differences $\mathbf{H}\mathbf{u} - \widetilde{\mathbf{H}}\mathbf{u}$ are also displayed on the bottom. They all have the same color range.



(a) pSNR = 41.15dB (b) pSNR = 38.25dB

Figure 15: Blurred images using the wavelet based method with two different sparsity patterns and the kernel in Figure 10b. $K = 30N^2$ coefficients are kept in the matrices. Matrices are obtained with the simple threshodling (left) and Algorithm 2 (right). The differences $\mathbf{Hu} - \widetilde{\mathbf{H}}\mathbf{u}$ are displayed within the same color range.

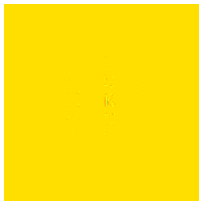
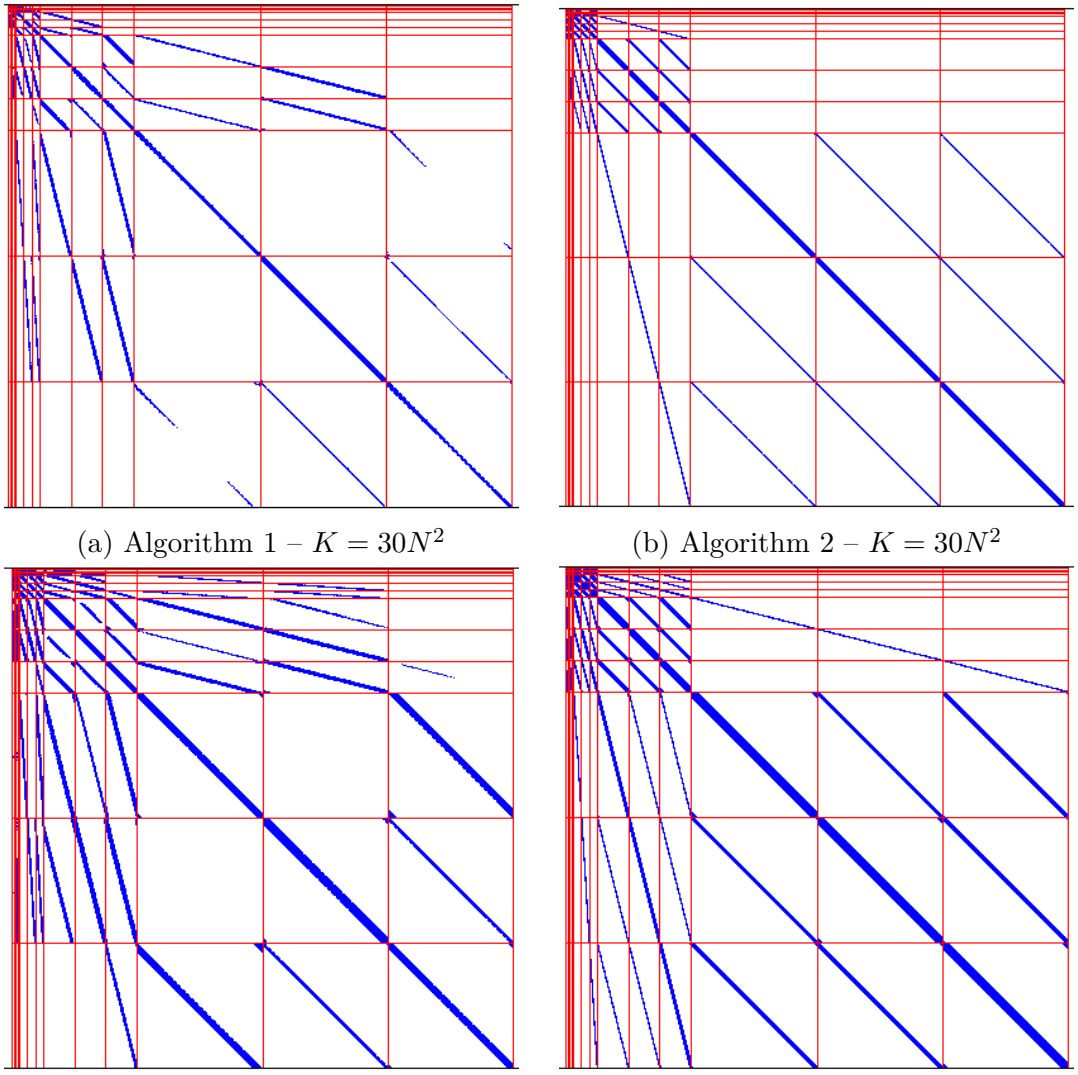
	Piece. Conv.	Difference	Algorithm 1	Difference	$l =$
2×2	31.90 dB		36.66 dB		5
1.21 sec					0.039s
4×4	38.49 dB		45.87 dB		30
0.17 sec					0.040s
8×8	44.51 dB		50.26 dB		50
0.36 sec					0.048s
16×16	53.75 dB		57.79 dB		100
0.39 sec					0.058s

Figure 16: Blurred images and the differences $\mathbf{Hu} - \widetilde{\mathbf{Hu}}$ for the kernel Figure 10b. Results on the left are obtained using piecewise convolution approximations with 2×2, 4×4, 8×8 and 16×16 partitionings all with 50% overlap. Results on the right are obtained using Algorithm 1 with the third $\mathbf{\Sigma}$ matrix keeping $K = lN^2$ coefficients. The pSNR and the time needed for the computation for the matrix-vector product are shown. The differences are displayed within the same color range as Figures 14 and 15



(a) Algorithm 1 – $K = 30N^2$

(b) Algorithm 2 – $K = 30N^2$

(c) Algorithm 1 – $K = 128N^2$

(d) Algorithm 2 – $K = 128N^2$

Figure 17: The structure of the wavelet matrices of $\mathbf{\Theta}_K$ are displayed for Algorithms 1 and 2 and for $K = 30N^2$ and $K = 128N^2$ coefficients. Algorithm 1 has been applied using the third $\mathbf{\Sigma}$ matrix.

An important remark is that the interest of the total variation term is not only used to regularize the ill-posed inverse problem, but also to handle the errors in the operator approximation. In practice we found that setting $\alpha = (1+\epsilon)\sigma^2 N$ where $\epsilon > 0$ is a small parameter provides good experimental results.

In Figures 18 to 21, we present some deblurring results with or without noise for image Figure 9a with kernel Figure 10a and for image Figure 9b with kernel 10b. Figure 18 shows that without noise and the simple kernel 10a, 4×4 piecewise convolutions perform better than wavelet approaches with $30N^2$. Piecewise convolution achieve better pSNR but are 4 times slower. For equivalent computation times, wavelet approaches should be constructed with $100N^2$ coefficients.

In Figures 20 and 21 we can notice that wavelet methods perform better than piecewise convolution. For this experiment they are roughly 10 times faster. Also notice that Algorithm 1 is strongly preferable to a simple thresholding since it reduces deblurring artefacts.

With noise, all methods perfom nearly the same as the exact operator. It suggests that it is not necessary to construct accurate approximations of the operators in practical problems. This observation is also supported by the experiment in Figure 22. In this experiment, we plot the pSNR of the deblurred image in presence of noise with respect to the number of elements in $\mathbf{\Theta}_K$. Interestingly, a matrix containing only $20N^2$ coefficients leads to deblurred images close to the results obtained with the exact operator. In this experiment, a total of $K = 5N^2$ coefficients in $\mathbf{\Theta}_K$ is enough to retrieve satisfactory results. This is a very encouraging result for blind deblurring problems.

9. Conclusion

9.1. Brief summary

In this paper, we analyzed standard approaches to efficiently approximate spatially varying blur. We showed that the standard piecewise convolution approach has an $\mathcal{O}(N \log(N) \epsilon^{-1})$ complexity where ϵ denotes the desired accuracy in l^2. Wavelet based methods have a greater adaptivity to the smoothness of the operator and exhibit an $\mathcal{O}(N \log(N) \epsilon^{-1/M})$ complexity, where M denotes the kernel regularity. This method is versatile since it is possible to adapt it to the kind of images that have to be treated. We showed that much better performance can be obtained by leveraging the fact that natural signals exhibit some structure in the wavelet domain. These theoretical results were confirmed by practical experiments on real images. Even though our conclusions are still preliminary since we tested only small 256×256 images, the wavelet based methods seem to significantly outperform the piecewise

(a) Degraded image

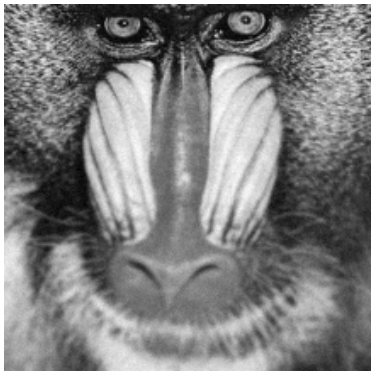
(b) Exact operator – 35.38dB

(c) Simple thresholding – 31.54dB

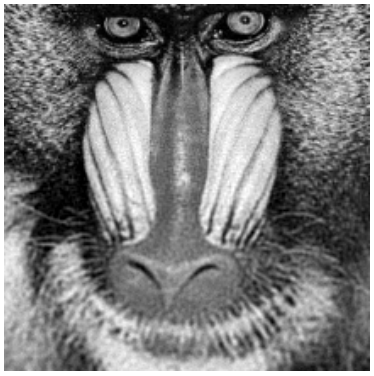
(d) Algorithm 1 – 31.03dB

(e) Piecewise convolutions – 33.81dB

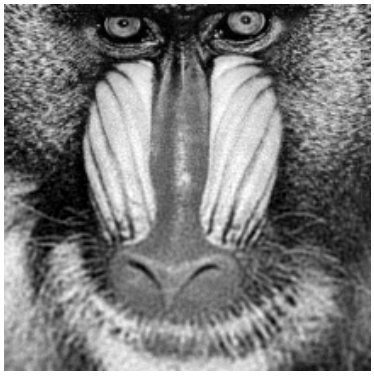
Figure 18: Deblurring results for kernel Figure 10a and without noise. Top-left: degraded image. Top-right: deblurred using the exact operator. Middle-left: deblurred by the wavelet based method and a simple thresholding. Middle-right: deblurred by the wavelet based method and Algorithm 2 with the third $\mathbf{\Sigma}$ matrix. Bottom: deblurred using a 4×4 piecewise convolution algorithm with 50% overlap. For wavelet methods $K = 30N^2$ coefficients are kept in matrices. pSNR are displayed for each restoration.



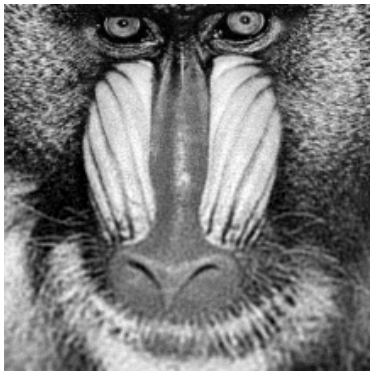
(a) Degraded image



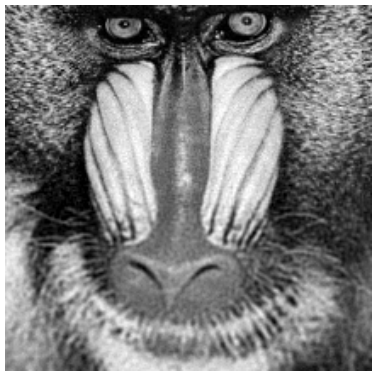
(b) Exact operator – 25.85dB



(c) Simple thresholding – 25.83dB



(d) Algorithm 1 – 25.83dB



(e) Piecewise convolutions – 25.81dB

Figure 19: Deblurring results for kernel Figure 10a and with $\sigma = 0.02$. Top-left: degraded image. Top-right: deblurred using the exact operator. Middle-left: deblurred by the wavelet based method and a simple thresholding. Middle-right: deblurred by the wavelet based method and Algorithm 2 with the third $\mathbf{\Sigma}$ matrix. Bottom: deblurred using a 4×4 piecewise convolution algorithm with 50% overlap. For wavelet methods $K = 30N^2$ coefficients are kept in matrices. pSNR are displayed for each restoration.

(a) Degraded image

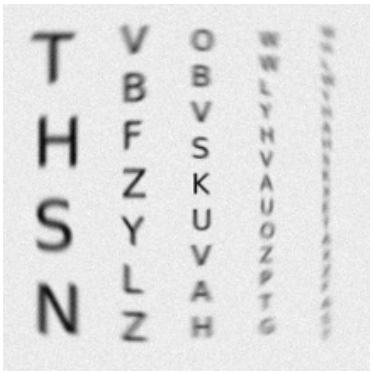
(b) Exact operator – 34.53dB

(c) Simple thresholding – 30.96dB

(d) Algorithm 1 – 30.56dB

(e) Piecewise convolutions – 28.37dB

Figure 20: Deblurring results for kernel Figure 10b and without noise. Top-left: degraded image. Top-right: deblurred using the exact operator. Middle-left: deblurred by the wavelet based method and a simple thresholding. Middle-right: deblurred by the wavelet based method and Algorithm 2 with the third $\mathbf{\Sigma}$ matrix. Bottom: deblurred using a 4×4 piecewise convolution algorithm with 50% overlap. For wavelet methods $K = 30N^2$ coefficients are kept in matrices. pSNR are displayed for each restoration.



(a) Degraded image

(b) Exact operator – 28.97dB

(c) Simple thresholding – 27.89dB

(d) Algorithm 1 – 28.02dB

(e) Piecewise convolutions – 27.12dB

Figure 21: Deblurring results for kernel Figure 10b and with $\sigma = 0.02$ noise. Top-left: degraded image. Top-right: deblurred using the exact operator. Middle-left: deblurred by the wavelet based method and a simple thresholding. Middle-right: deblurred by the wavelet based method and Algorithm 2 with the third $\mathbf{\Sigma}$ matrix. Bottom: deblurred using a 4×4 piecewise convolution algorithm with 50% overlap. For wavelet methods $K = 30N^2$ coefficients are kept in matrices. pSNR are displayed for each restoration.

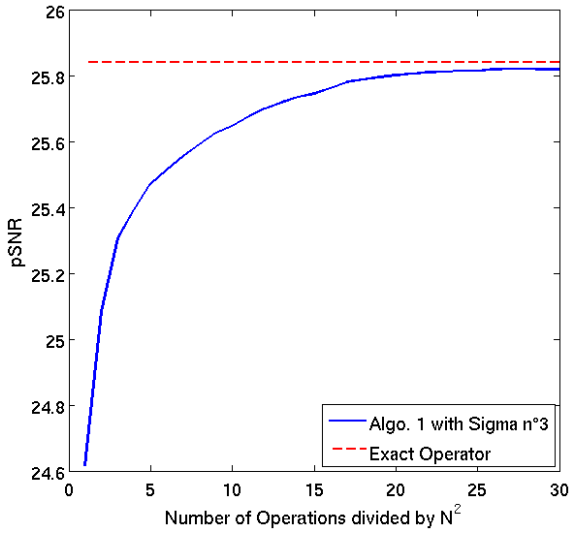


Figure 22: pSNR of the deblurred image with respect to the number of coefficients in the matrix divided by N^2 for the image Figure 9a and the kernel Figure 10a. The matrix is constructed using Algorithm 1 with the third $\mathbf{\Sigma}$ matrix with $K = lN^2$ coefficients for l from 1 to 30. Deblurred imaged using these matrices are compared with the one obtained with the exact operator.

convolutions approaches with or without overlap. Moreover, they seem to provide satisfactory deblurring results on practical problems with a complexity no greater than $5N^2$ operations, where N^2 denotes the pixels number.

9.2. Outlook

We provided a simple complexity analysis based solely on the *global* regularity of the kernel function. It is well known that wavelets are able to adapt locally to the structures of images or operators [38]. The method should thus provide an efficient tool for piecewise regular blurs appearing in computer vision for instance. It could be interesting to evaluate precisely the complexity of wavelet based approximations for piecewise regular blurs.

An important problem of the wavelet based approach is the need to project the operator on a wavelet basis. In this paper we performed this operation using the computationally intensive Algorithm 3. It could be interesting to derive fast projection methods for certain blur families.

Another exciting research perspective is the problem of blind deconvolution. Expressing the unknown operator as a sparse matrix in the wavelet domain is a good way to improve the problem identifiability. This is however far from being sufficient since the blind deconvolution problem has far more unknowns (a full operator and an image) than data (a single image). Further assumptions should thus be made on the wavelet coefficients regularity, and we plan to study this problem in a forthcoming work.

Finally let us mention that we observed some artifacts when using the wavelet based methods with high sparsity levels. This is probably due to their non translation and rotation invariance. It could be interesting to study sparse approximations in redundant wavelet bases or curvelet bases. It was shown for instance in [39] that curvelets are near optimal to represent Fourier integral operators.

Acknowledgements

The authors wish to thank François Malgouyres and Mathieu Bouyrie for stimulating discussions in the early preparation of this work. They wish to thank Jérémie Bigot and Guillermo Cabrera for providing the SDSS database address as an illustration of a deblurring problem with a spatially variant PSF. They wish to thank Jérôme Fehrenbach and the IP3D team of ITAV for their comments and support. They wish to thank Michael Hirsch for providing the pictures in Figure 1. This work was supported by ANR SPH-IM-3D (ANR-12-BSV5-0008). Paul Escande is pursuing a PhD degree supported by the MODIM project funded by the PRES of Toulouse University and Midi-Pyrénées region.

Bibliography

[1] A. Sawchuk, Space-variant image motion degradation and restoration, Proceedings of the IEEE 60 (1972) 854–861.

[2] G. Beylkin, R. Coifman, V. Rokhlin, Fast wavelet transform and numerical algorithm, Commun. Pure and Applied Math. 44 (1991) 141–183.

[3] Y. Meyer, Wavelets and operators, Vol. 2, Cambridge Univ Press, 1992.

[4] J. G. Nagy, D. P. O'leary, Fast iterative image restoration with a spatially varying psf, in: Optical Science, Engineering and Instrumentation'97, International Society for Optics and Photonics, 1997, pp. 388–399.

[5] J. Nagy, D. O'Leary, Restoring images degraded by spatially variant blur, SIAM Journal on Scientific Computing 19 (1998) 1063.

[6] M. Hirsch, S. Sra, B. Scholkopf, S. Harmeling, Efficient filter flow for space-variant multiframe blind deconvolution, in: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE, 2010, pp. 607–614.

[7] M. Hirsch, C. J. Schuler, S. Harmeling, B. Scholkopf, Fast removal of non-uniform camera shake, in: Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE, 2011, pp. 463–470.

[8] R. Jorand, G. Le Corre, J. Andilla, A. Maandhui, C. Frongia, V. Lobjois, B. Ducommun, C. Lorenzo, Deep and clear optical imaging of thick inhomogeneous samples, PLoS ONE 7.

[9] A. Cohen, I. Daubechies, P. Vial, Wavelets on the interval and fast wavelet transforms, Applied and Computational Harmonic Analysis 1 (1) (1993) 54–81.

[10] S. Mallat, A Wavelet Tour of Signal Processing - The Sparse Way, Third Edition, 2008.

[11] I. Daubechies, Ten Lectures on Wavelets, SIAM, 1992.

[12] M. Born, E. Wolf, Principles of optics: electromagnetic theory of propagation, interference and diffraction of light, CUP Archive, 1999.

[13] G. David, J.-L. Journé, A boundedness criterion for generalized calderón-zygmund operators, The Annals of Mathematics 120 (1984) 371–397.

[14] A. A. Sawchuk, Space-variant image restoration by coordinate transformations, Journal of the Optical Society of America 64 (1974) 138–144.

[15] S. R. McNown, B. R. Hunt, Approximate shift-invariance by warping shift-variant systems, in: SPIE's 1994 International Symposium on Optics, Imaging, and Instrumentation, International Society for Optics and Photonics, 1994, pp. 156–167.

[16] A. Tabernero, J. Portilla, R. Navarro, Duality of log-polar image representations in the space and spatial-frequency domains, Signal Processing, IEEE Transactions on 47 (1999) 2469–2479.

[17] E. Angel, A. K. Jain, Restoration of images degraded by spatially varying pointspread functions by a conjugate gradient method, Applied Optics 17 (1978) 2186–2190.

[18] J. Kamm, J. G. Nagy, Kronecker product and svd approximations for separable spatially variant blurs, in: SPIE's International Symposium on Optical Science, Engineering, and Instrumentation, International Society for Optics and Photonics, 1998, pp. 358–369.

[19] B. Zhang, J. Zerubia, J. Olivo-Marin, Gaussian approximations of fluorescence microscope point-spread function models, Applied Optics 46 (2007) 1819–1829.

[20] C. Preza, J.-A. Conchello, Depth-variant maximum-likelihood restoration for three-dimensional fluorescence microscopy, JOSA A 21 (2004) 1593–1601.

[21] E. Maalouf, B. Colicchio, A. Dieterlen, Fluorescence microscopy three-dimensional depth variant point spread function interpolation using zernike moments, JOSA A 28 (2011) 1864–1870.

[22] S. Ben Hadj, L. Blanc-Féraud, Restoration mehod for spatially variant blurred images, Research report, INRIA (June 2011).

[23] H. Kirshner, D. Sage, M. Unser, 3D PSF models for fluorescence microscopy in ImageJ, in: Proceedings of the Twelfth International Conference on Methods and Applications of Fluorescence Spectroscopy, Imaging and Probes (MAF'11), 2011, p. 154.

[24] E.-C. Chang, S. Mallat, C. Yap, Wavelet foveation, Applied and Computational Harmonic Analysis 9 (1999) 312–335.

[25] P. Escande, P. Weiss, F. Malgouyres, Spatially varying blur recovery. diagonal approximations in the wavelet domain, Proceedings of ICPRAM.

[26] P. Hansen, J. Nagy, D. O'leary, Deblurring images: matrices, spectra, and filtering, Siam, 2006.

[27] R. Coifman, Y. Meyer, Wavelets, Calderón-Zygmund and multilinear operators, Cambridge Studies in Advanced Math 48.

[28] L. I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D: Nonlinear Phenomena 60 (1992) 259–268.

[29] Y. Gousseau, J.-M. Morel, Are natural images of bounded variation?, SIAM Journal on Mathematical Analysis 33 (2001) 634–648.

[30] D. Mumford, B. Gidas, Stochastic models for generic images, Quarterly of applied mathematics 59 (2001) 85–112.

[31] P. Petrushev, A. Cohen, H. Xu, R. A. DeVore, Nonlinear approximation and the space bv (r 2), American Journal of Mathematics 121 (1999) 587–628.

[32] J. Wang, B. J. Lucier, Error bounds for finite-difference methods for rudin-osher-fatemi image smoothing, SIAM Journal on Numerical Analysis 49 (2011) 845–868.

[33] S. Bartels, Total variation minimization with finite elements: Convergence and iterative solution, SIAM Journal on Numerical Analysis 50 (2012) 1162–1180.

[34] A. Cohen, Numerical analysis of wavelet methods, Vol. 32, Elsevier, 2003.

[35] W. Dahmen, S. Prössdorf, R. Schneider, Wavelet approximation methods for pseudodifferential equations II: matrix compression and fast solution, Advances in Computational Mathematics 1 (1993) 259–335.

[36] G. H. Golub, C. F. Van Loan, Matrix computations, Vol. 3, JHU Press, 2012.

[37] A. Chambolle, T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, Journal of Mathematical Imaging and Vision 40 (2011) 120–145.

[38] A. Cohen, W. Dahmen, R. A. DeVore, Adaptive wavelet methods ii-beyond the elliptic case., Foundations of Computational Mathematics 2 (2002) 203–245.

[39] E. Candes, L. Demanet, Curvelets and fourier integral operators, Comptes Rendus Mathematique 336 (2003) 395–398.

[40] J. Deny, J.-L. Lions, Les espaces du type de beppo levi, in: Annales de l'institut Fourier, Vol. 5, Institut Fourier, 1954, pp. 305–370.

Appendix A. Proof of Lemma 1 and 2

Before proving lemmas 1 and 2, we need additional results.

Lemma 6. *The integral operator $\widetilde{H}_m$ can be written as follows*

$$\widetilde{H}_m u = \int_\Omega \widetilde{K}(x,y)u(y)dy$$

with

$$\widetilde{K}(x,y) = K\left(c(x), c(x) - x + y\right)$$

and where $c(x)$ denotes the center of the subregion containing x.

$$c(x) = \frac{\lfloor mx \rfloor}{m} + \frac{1}{2m}.$$

Proof. By construction, $\widetilde{K}$ should lead to piecewise convolutions. Therefore, we can associate a convolution kernel $\tilde{h}_k$ for each ω_k defined for all $x \in \omega_k$ by $\widetilde{K}(x,y) = \tilde{h}_k(x-y)$. Assuming that $\tilde{h}$ corresponds to the kernel K at the center of ω_k we get in each ω_k,

$$\widetilde{K}(c(x),\cdot) = \quad K(c(x),\cdot) \quad \Leftrightarrow \quad K(c(x),\cdot) = \quad \tilde{h}_k(c(x)-\cdot).$$

Hence, $\tilde{h}_k(y) = K(c(x), c(x)-y)$. □

Lemma 7 (Convolution)**.** *An integral operator H with a $\mathcal{C}^1(\Omega,\Omega)$ kernel K is a convolution iff*

$$\forall (x,y) \in \Omega \times \Omega, \quad \left\langle \nabla K(x,y), \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\rangle = 0.$$

Proof. First suppose that H is a convolution. It means that $K(x,y) = k(x-y)$. A simple differentiation leads to

$$\partial_x K(x,y) + \partial_y K(x,y) = k'(x-y) - k'(x-y) = 0.$$

Second, suppose that

$$\forall (x,y) \in \Omega \times \Omega, \quad \left\langle \nabla K(x,y), \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\rangle = 0.$$

It means that $K(x+h, y+h) = K(x,y)$ for all x, y, h. Therefore, choosing $h = -y$ leads to

$$K(x,y) = K(x-y,0) = k(x-y), \quad \forall x, y.$$

□

Proof of Lemma 1. First we study $\left|K(x,y) - \widetilde{K}(x,y)\right|$. Since K is L-Lipschitz (see (12)) we get

$$\left|K(x,y) - \widetilde{K}(x,y)\right| = \left|K(x,y) - K(c(x), c(x) - x + y)\right| \leq \sqrt{2}L \left|x - c(x)\right|.$$

$$\begin{aligned}
\left\|H - \widetilde{H}_m\right\|_{1\to\infty} &= \sup_{x\in\Omega}\sup_{y\in\Omega}\left|K(x,y) - \widetilde{K}(x,y)\right| \\
&\leq \sup_{x\in\Omega}\sqrt{2}L\left|x - c(x)\right| \\
&\leq \frac{L}{\sqrt{2}m}
\end{aligned}$$

To study the spectral norm, we consider a function u with unit Euclidean norm.

$$\begin{aligned}
\left\|Hu - \widetilde{H}_m u\right\|_2^2 &= \int_{x\in\Omega}\left(\int_{y\in\Omega}\left(K(x,y) - \widetilde{K}(x,y)\right)u(y)dy\right)^2 dx \\
&\leq \int_{x\in\Omega}\left(\int_{y\in\Omega}\left(K(x,y) - \widetilde{K}(x,y)\right)^2 dy\right)\left(\int_{y\in\Omega}u^2(y)dy\right)dx \\
&\leq \int_{x\in\Omega}\int_{y\in\Omega}\left(K(x,y) - \widetilde{K}(x,y)\right)^2 dydx \\
&\leq \int_{x\in\Omega}\int_{y\in\Omega}2L^2\left|x - c(x)\right|^2 dydx \\
&= 2L^2\int_{x\in\Omega}\left|x - c(x)\right|^2 dx, \quad \text{since} \quad |\Omega| = 1 \\
&= 2L^2\sum_{k=1}^{m}\int_{\frac{k-1}{m}}^{\frac{k}{m}}\left|x - c(x)\right|^2 dx \\
&\leq 2L^2\sum_{k=1}^{m}\frac{1}{4m^2}\frac{1}{m} \\
&\leq \frac{L^2}{2m^2}.
\end{aligned}$$

□

Proof of Lemma 2. If H is a $\mathcal{C}^1$ kernel that is not a convolution then, Lemma 7 ensures that there exists $(x_0, y_0) \in \Omega\times\Omega$ such that $\left\langle \nabla K(x_0,y_0), \begin{pmatrix}1\\1\end{pmatrix}\right\rangle \neq 0$. Since K is $\mathcal{C}^1$, there exists $h > 0$ and $\epsilon > 0$ such that $\forall (x,y) \in I\times J := [x_0, x_0+h]\times[y_0, y_0+h]$

either $\left\langle \nabla K(x,y), \begin{pmatrix}1\\1\end{pmatrix} \right\rangle > \epsilon$ or $\left\langle \nabla K(x,y), \begin{pmatrix}1\\1\end{pmatrix} \right\rangle < -\epsilon$. We assume without loss of generality that the first inequality holds. We define

$$I_m = \bigcup_{k=1}^{m} \left[\frac{k}{m} - \frac{1}{4m}, \frac{k}{m} \right] \cap I.$$

Assuming that $\lfloor mh \rfloor \geq 2$, ensures that I is partitioned in at least two subregions and that $\frac{2h}{11} \leq |I_m| \leq \frac{h}{4}$. In order to prove (14), we exhibit a pathological vector u with unit Euclidean norm:

$$u = \frac{\mathbb{1}_J}{\sqrt{h}}.$$

We first bound $\left\| H - \widetilde{H}_m \right\|_2$ from below as follows:

$$\begin{aligned}
\left\| H - \widetilde{H}_m \right\|_2^2 &\geq \left\| Hu - \widetilde{H}_m u \right\|_2^2 \\
&\geq \int_{x \in I_m} \left(Hu - \widetilde{H}_m u \right)^2 (x) dx \\
&= \int_{x \in I_m} \left(\int_{y \in \Omega} K(x,y) - K(c(x), c(x) - x + y) u(y) \right)^2 dx \\
&= \int_{x \in I_m} \frac{1}{h} \left(\int_{y \in J} K(x,y) - K(c(x), c(x) - x + y) \right)^2 dx
\end{aligned}$$

Since

$$K(x_2, y_2) = K(x_1, y_1) + \int_{t=0}^{1} \left\langle \nabla K(x_1 + t(x_2 - x_1), y_1 + t(y_2 - y_1)), \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \end{pmatrix} \right\rangle dt$$

we get:

$$\begin{aligned}
&K(x,y) - K(c(x), c(x) - x + y) \\
&= \int_{t=0}^{1} \left\langle \nabla K \left(c(x) + t(x - c(x)), c(x) - x + y + t(x - c(x)) \right), \begin{pmatrix} x - c(x) \\ x - c(x) \end{pmatrix} \right\rangle dt \\
&= (x - c(x)) \int_{t=0}^{1} \left\langle \nabla K \left(a + th, b + th \right), \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\rangle dt
\end{aligned}$$

with $h = x - c(x)$, $a = c(x)$, $b = c(x) - x + y$. Since $\left\langle \nabla K\left(a + th, b + th\right), \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\rangle > \epsilon$, we have:

$$\left(\int_{y \in J} K(x,y) - K(c(x), c(x) - x + y) \right)^2 \geq \left(x - c(x)\right)^2 \epsilon^2 h^2$$

On I_m, $(x - c(x))^2 \geq \frac{1}{16m^2}$. Thus

$$\left(\int_{y \in J} K(x,y) - K(c(x), c(x) - x + y) \right)^2 \geq \frac{1}{16} \frac{\epsilon^2 h^2}{m^2}.$$

A lower bound of the error can be derived as follows:

$$\begin{aligned} \left\| Hu - \widetilde{H}_m u \right\|_2^2 &\geq \frac{1}{16} \frac{\epsilon^2 h}{m^2} \left| I_m \right| \\ &\geq \frac{2}{11} \frac{1}{16} \frac{\epsilon^2 h^2}{m^2} \end{aligned}$$

Similarly, choosing $v = \frac{1}{h} \mathbb{1}_J$, we get

$$\begin{aligned} \left\| Hv - \widetilde{H}_m v \right\|_\infty &= \sup_{x \in I_m} \left| Hv - \widetilde{H}_m v \right| \\ &= \sup_{x \in I_m} \frac{1}{h} \left| \int_{y \in J} K(x,y) - \widetilde{K}(x,y) dy \right| \\ &\geq \frac{1}{4m}. \end{aligned}$$

□

Appendix B. Proof of Theorem 3

Proof of i). The discrete convolutions have to be performed on extended regions of size $N/m + \kappa N$ in order to correctly handle the boundary conditions. Each convolution is thus performed on a sub-region with $N/m + \kappa N$ pixels. This can be performed in $\mathcal{O}\left((N/m + \kappa N) \log(N/m + \kappa N)\right)$ arithmetic operations. Since there are m subwindows, we get the announced result. □

Proof of ii). The proof of this fact is similar to the one in the continuous setting. The only additional requirement is that N and m are sufficiently large to partition I and J in at least two subregions. □

Proof of iii). This is a simple consequence of complexity bound (15) and the bounds estimates provided in equations (16) and (17). In order to satisfy $\left\|\mathbf{H} - \widetilde{\mathbf{H}}_m\right\|_{2\to 2} \leq \epsilon$, it is necessary and sufficient to set $\frac{L}{2m} \leq \epsilon$. The smallest integer ensuring this condition is $m = \lceil \frac{L}{2\epsilon} \rceil$. Therefore, the numerical complexity of the method becomes

$$N\left(1 + \frac{\kappa L}{\epsilon}\right) \log\left(\frac{N\epsilon}{L} + \kappa N\right).$$

The asymptotics of this complexity for small ϵ is

$$\frac{L\kappa N \log(\kappa N)}{\epsilon}.$$

□

Appendix C. Proof of Lemma 4

We let Π_M denote the set of polynomials of degree less or equal to M.

Lemma 8 below is a common result in numerical analysis [40] (see also Theorem 3.2.1 in [34]). It ensures that the approximation error of a function by a polynomial of degree M is bounded by the Sobolev semi-norm $W^{M,p}$.

Lemma 8 (Polynomial approximation)**.** *For $1 \leq p \leq +\infty$, $M \in \mathbb{N}^*$ and $\Omega \subset \mathbb{R}^d$ a bounded domain, the following bound holds*

$$\inf_{g\in\Pi_M} \|f - g\|_{\mathbb{L}^p(\Omega)} \leq C\,|f|_{W^{M+1,p}(\Omega)}\,, \tag{C.1}$$

where C is a constant that depends on d, M, p and Ω only.

Moreover, if $I_h \subset \Omega \subset \mathbb{R}^d$ is a cube of sidelength h, the following estimate holds

$$\inf_{g\in\Pi_M} \|f - g\|_{\mathbb{L}^p(I_h)} \leq Ch^{M+1}\,|f|_{W^{M+1,p}(I_h)}\,, \tag{C.2}$$

where C is a constant only depending on d, M, p and Ω.

Let $I_{j,m}$ denote the support of the wavelet $\psi_{j,m}$ and $\lambda_{j,m}$ denote its center. From the wavelets definition, we get

$$I_{j,m} = [2^{-j}(m-1), 2^{-j}(m+1)]$$

therefore $|I_{j,m}| = 2\cdot 2^{-j}$ and $\lambda_{j,m} = 2^{-j}m$. We will now prove Lemma 4 in the simple case where $K \in W^{M,\infty}$ for the ease of exposition. This assumption ensures that K is non singular on its diagonal.

Proof of Lemma 4. Since the mapping $(x,y) \mapsto K(x,y)\psi_{j,m}(y)\psi_{k,n}(x)$ is bounded, it is also absolutely integrable on compact domains. Therefore $\langle H\psi_{j,m}, \psi_{k,n}\rangle$ is well-defined for all (j,m,k,n). Moreover Fubini's theorem can be applied and we get

$$\begin{aligned}\langle H\psi_{j,m}, \psi_{k,n}\rangle &= \int_{I_{k,n}} \int_{I_{j,m}} K(x,y)\psi_{j,m}(y)\psi_{k,n}(x)dydx \\ &= \int_{I_{j,m}} \int_{I_{k,n}} K(x,y)\psi_{j,m}(y)\psi_{k,n}(x)dxdy.\end{aligned}$$

To prove the result, we distinguish the cases $j \leq k$ and $j > k$. In this proof, we focus on the case $j \leq k$. The other one can be obtained by symmetry, using the facts that $\langle H\psi_{j,m}, \psi_{k,n}\rangle = \langle \psi_{j,m}, H^*\psi_{k,n}\rangle$ and that H and H^* are both Calderòn-Zygmund operators in the same class.

To exploit the regularity of K and ψ, note that for all $g \in \Pi_{M-1}$, $\int_{I_{k,n}} g(x)\psi_{k,n}(x)dx = 0$ since ψ has M vanishing moments. Therefore,

$$\langle H\psi_{j,m}, \psi_{k,n}\rangle = \int_{I_{j,m}} \inf_{g\in\Pi_{M-1}} \int_{I_{k,n}} \left(K(x,y) - g(x)\right) \psi_{j,m}(y)\psi_{k,n}(x)dxdy,$$

and

$$\begin{aligned}|\langle H\psi_{j,m}, \psi_{k,n}\rangle| &\leq \int_{I_{j,m}} \inf_{g\in\Pi_{M-1}} \int_{I_{k,n}} |K(x,y) - g(x)|\, |\psi_{j,m}(y)|\, |\psi_{k,n}(x)|\, dxdy \\ &\leq \int_{I_{j,m}} \inf_{g\in\Pi_{M-1}} \|K(\cdot,y) - g\|_{\mathbb{L}^\infty(I_{k,n})}\, \|\psi_{k,n}\|_{\mathbb{L}^1(I_{k,n})}\, |\psi_{j,m}(y)|\, dy.\end{aligned}$$

By Lemma 8, $\inf_{g\in\Pi_{M-1}} \|K(\cdot,y) - g\|_{\mathbb{L}^\infty(I_{k,n})} \lesssim 2^{-kM} |K(\cdot,y)|_{W^{M,\infty}(I_{k,n})}$ since $I_{k,n}$ is a cube of sidelength $2\cdot 2^{-k}$. We thus obtain

$$\begin{aligned}|\langle H\psi_{j,m}, \psi_{k,n}\rangle| &\lesssim 2^{-kM} \|\psi_{k,n}\|_{\mathbb{L}^1(I_{k,n})}\, \|\psi_{j,m}\|_{\mathbb{L}^1(I_{j,m})} \operatorname*{ess\,sup}_{y\in I_{j,m}} |K(\cdot,y)|_{W^{M,\infty}(I_{k,n})} \\ &\lesssim 2^{-kM} 2^{-\frac{dj}{2}} 2^{-\frac{dk}{2}} \operatorname*{ess\,sup}_{y\in I_{j,m}} |K(\cdot,y)|_{W^{M,\infty}(I_{k,n})} \qquad \text{(C.3)}\end{aligned}$$

since $\|\psi_{j,m}\|_{\mathbb{L}^1} = 2^{-\frac{dj}{2}} \|\psi\|_{\mathbb{L}^1}$. To conclude the proof, we distinguish the cases where $\psi_{j,m}$ and $\psi_{k,n}$ have supports distant from more than $2\cdot 2^{-j}$ or not.

Case 1: $|\lambda_{j,m} - \lambda_{k,n}| \geq 4\cdot 2^{-j}$. This is a sufficient condition for the supports to be disjoints. More precisely, $\forall x \in I_{k,n}$ and $\forall y \in I_{j,m}$, we get $|x-y| \geq 2\cdot 2^{-j}$. Indeed,

$|x-\lambda_{k,n}| \leq 2^{-k} \leq 2^{-j} \leq \frac{1}{4}|\lambda_{j,m}-\lambda_{k,n}|$ and $|y-\lambda_{j,m}| \leq 2^{-j} \leq \frac{1}{4}|\lambda_{j,m}-\lambda_{k,n}|$. Hence, by triangle inequality,

$$\begin{aligned}|\lambda_{j,m}-\lambda_{k,n}| &= |(\lambda_{j,m}-y)+(x-\lambda_{k,n})+(y-x)| \\ &\leq |\lambda_{j,m}-y|+|\lambda_{k,n}-x|+|x-y| \\ &\leq \frac{1}{2}|\lambda_{j,m}-\lambda_{k,n}|+|x-y|,\end{aligned}$$

so that $|x-y| \geq \frac{1}{2}|\lambda_{j,m}-\lambda_{k,n}| \geq 2\cdot 2^{-j}$. We can end the proof as follows

$$\begin{aligned}\operatorname*{ess\,sup}_{y\in I_{j,m}} |K(\cdot,y)|_{W^{M,\infty}(I_{k,n})} &\lesssim \operatorname*{ess\,sup}_{|x-y|\geq\frac{1}{2}|\lambda_{j,m}-\lambda_{k,n}|} \frac{1}{|x-y|^{M+d}} \text{[check def]} \\ &\leq \left(\frac{2}{|\lambda_{j,m}-\lambda_{k,n}|}\right)^{M+d}.\end{aligned}$$

This inequality is tighter than the one given in equation (22). Indeed:

$$\frac{2}{|\lambda_{j,m}-\lambda_{k,n}|} \leq \frac{2\cdot(1+2^{j-k})}{|\lambda_{j,m}-\lambda_{k,n}|} \leq \frac{4\cdot(1+2^{j-k})}{2^{-j}+2^{-k}+|\lambda_{j,m}-\lambda_{k,n}|}.$$

Combining this upper-bound with (C.3), we obtain

$$|\langle H\psi_{j,m},\psi_{k,n}\rangle| \lesssim 2^{(j-k)(M+d/2)}\left(\frac{2^{-j}+2^{-k}}{2^{-j}+2^{-k}+|\lambda_{j,m}-\lambda_{k,n}|}\right)^{M+d}.$$

Case 2: $|\lambda_{j,m}-\lambda_{k,n}| \leq 4\cdot 2^{-j}$. In that case, the support may intersect and we get

$$\frac{2^{-j}+2^{-k}}{2^{-j}+2^{-k}+|\lambda_{j,m}-\lambda_{k,n}|} \geq \frac{2^{-j}}{2\cdot 2^{-j}+4\cdot 2^{-j}} \geq \frac{1}{6}.$$

Therefore it suffices to prove that $|\langle H\psi_{j,m},\psi_{k,n}\rangle| \lesssim 2^{(j-k)(\frac{d}{2}+M)}$. Since $K \in W^{M,\infty}$,

$$\operatorname*{ess\,sup}_{y\in I_{j,m}} |K(\cdot,y)|_{W^{M,\infty}(I_{k,n})} < +\infty$$

and we get

$$|\langle H\psi_{j,m},\psi_{k,n}\rangle| \lesssim 2^{-\frac{dj}{2}}2^{-\frac{dk}{2}}2^{-kM} \leq 2^{-\frac{dj}{2}}2^{-\frac{dk}{2}}2^{-kM}2^{jM}, \quad \text{since } j\geq 0.$$

□

Appendix D. Proof of Proposition 1

Proof of Proposition 1. Let $\mathcal{S} = \{0 \dots, \log_2(N)\}$ denote the set of admissible scales and $\mathcal{T}_j = \{0, \dots, 2^j - 1\}$ denote the set of admissible shifts at scale j. Let N_η denote the total number of coefficients in $\mathbf{\Theta}$ above a threshold $\eta \geq 0$. This number is given by

$$N_\eta = \sum_{i=0}^{\log_2(N)} \sum_{j=0}^{\log_2(N)} \# \left\{ m \in \mathcal{T}_j, n \in \mathcal{T}_k, |\theta_{j,m,k,n}| \leq \eta \right\}.$$

From Lemma 4, we get that $|\theta_{j,m,k,n}| \leq C_M 2^{-\left(M+\frac{1}{2}\right)|j-k|} \left(\frac{\omega_{j,k}}{\omega_{j,k} + |2^{-j}m - 2^{-k}n|} \right)^{M+1}$ with $\omega_{j,k} = 2^{-j} + 2^{-k}$. From this upper-bound, we infer that

$$\# \left\{ m \in \mathcal{T}_j, n \in \mathcal{T}_k, |\theta_{j,m,k,n}| \leq \eta \right\} \leq \# \mathcal{G}_{j,k}$$

where

$$\mathcal{G}_{j,k} = \{ m \in \mathcal{T}_j, n \in \mathcal{T}_k, \ \left| 2^{-j}m - 2^{-k}n \right| \geq B_{j,k}^\eta \}$$

and

$$B_{j,k}^\eta = \omega_{j,k} \left[\left(\frac{\eta}{C_M} \right)^{-\frac{1}{M+1}} 2^{-\left(\frac{M+1/2}{M+1}\right)|j-k|} - 1 \right].$$

In the following we will denote $a_M = \left(\frac{M+1/2}{M+1}\right)$. The cardinal of $\mathcal{G}_{j,k}$ can be evaluated by noting that

$$(m,n) \in \mathcal{G}_{j,k} \ \Leftrightarrow \ -B_{j,k}^\eta \leq 2^{-k}n \leq B_{j,k}^\eta \quad \text{and} \quad n \in \mathcal{T}_k$$

Thus for fixed m, j and k, there are at most $2^{k+1}B_{j,k}^\eta$ indices belonging to $\mathcal{G}_{j,k}$ and $\#\mathcal{G}_{j,k} \leq 2^j 2^{k+1} B_{j,k}^\eta$. We can finish the proof as follows

$$\begin{aligned}
N_\eta &\lesssim \sum_{j=0}^{\log_2(N)} \sum_{k=0}^{\log_2(N)} 2^j 2^k B_{j,k}^\eta \\
&\lesssim \left(\frac{\eta}{C_M} \right)^{-\frac{1}{M+1}} \sum_{j=0}^{\log_2(N)} \sum_{k=j}^{\log_2(N)} 2^j 2^k (2^{-j} + 2^{-k}) 2^{a_M(j-k)} \\
&= \left(\frac{\eta}{C_M} \right)^{-\frac{1}{M+1}} \left(\sum_{j=0}^{\log_2(N)} \sum_{k=j}^{\log_2(N)} 2^k 2^{a_M(j-k)} + \sum_{j=0}^{\log_2(N)} \sum_{k=j}^{\log_2(N)} 2^j 2^{a_M(j-k)} \right) \\
&= \left(\frac{\eta}{C_M} \right)^{-\frac{1}{M+1}} \left(\sum_{j=0}^{\log_2(N)} 2^{j a_M} \sum_{k=j}^{\log_2(N)} 2^{k(1-a_M)} + \sum_{j=0}^{\log_2(N)} 2^{j(1+a_M)} \sum_{k=j}^{\log_2(N)} 2^{-k a_M} \right)
\end{aligned}$$

$$
\begin{aligned}
&\lesssim \left(\frac{\eta}{C_M}\right)^{-\frac{1}{M+1}} \left(\sum_{j=0}^{\log_2(N)} 2^{j a_M} N^{1-a_M} + \sum_{j=0}^{\log_2(N)} 2^{j(1+a_M)} 2^{-j a_M} \right) \\
&= \left(\frac{\eta}{C_M}\right)^{-\frac{1}{M+1}} \left(\sum_{j=0}^{\log_2(N)} 2^{j a_M} \sum_{j=0}^{\log_2(N)} 2^{j} N^{1-a_M} \right) \\
&\lesssim \left(\frac{\eta}{C_M}\right)^{-\frac{1}{M+1}} \left(N + N^{1-a_M} N^{a_M} \right) \\
&\lesssim \left(\frac{\eta}{C_M}\right)^{-\frac{1}{M+1}} N.
\end{aligned}
$$

We used the fact that $a_M = \left(\frac{M+1/2}{M+1}\right) < 1$. □

Appendix E. Proof of Theorem 5

Proof of i). The notations in this proof are the same as in Appendix D. First note that the thresholding rule in Theorem 5 is equivalent to zeroing all coefficients such that $\left|2^{-j}m - 2^{-k}n\right| \geq C^{\eta}_{j,k}$ with $C^{\eta}_{j,k} := \omega_{j,k}\left(\eta^{-\frac{1}{M+1}} - 1\right)$.

Let us define

$$
\mathcal{G} = \left\{ (j,m,k,n) \in \mathcal{S} \times \mathcal{T}_j \times \mathcal{S} \times \mathcal{T}_k,\ \left|2^{-j}m - 2^{-k}n\right| \leq C^{\eta}_{j,k} \right\}.
$$

and

$$
\mathcal{G}_{j,k} := \left\{ m \in \mathcal{T}_j, n \in \mathcal{T}_k,\ \left|2^{-j}m - 2^{-k}n\right| \leq C^{\eta}_{j,k} \right\}.
$$

Using an elementary argument described in Appendix D, we get that the number of coefficients in $\mathcal{G}_{j,k}$ is proportional to $2^j 2^k B^{\eta}_{j,k}$. Hence the overall number of non zero coefficients $\#\mathcal{G}$ in $\mathbf{\Theta}_{\eta}$ satisfies

$$
\begin{aligned}
\#\mathcal{G} &\lesssim \sum_{j=0}^{\log_2(N)} \sum_{k=j}^{\log_2(N)} 2^j 2^k B^{\eta}_{j,k} \\
&\lesssim \eta^{-\frac{1}{M+1}} \sum_{j=0}^{\log_2(N)} \sum_{k=j}^{\log_2(N)} 2^j 2^k (2^{-j} + 2^{-k}) \\
&= \eta^{-\frac{1}{M+1}} \left(\sum_{j=0}^{\log_2(N)} \sum_{k=j}^{\log_2(N)} 2^k + \sum_{j=0}^{\log_2(N)} \sum_{k=j}^{\log_2(N)} 2^j \right) \\
&\lesssim \eta^{-\frac{1}{M+1}} \left(\sum_{j=0}^{\log_2(N)} N + \sum_{j=0}^{\log_2(N)} 2^j \log_2(N) \right)
\end{aligned}
$$

$$\lesssim \eta^{-\frac{1}{M+1}} N \log_2(N).$$

□

Proof of ii). Since $\mathbf{\Psi}$ is an orthogonal wavelet transform

$$\left\|\mathbf{H} - \widetilde{\mathbf{H}}_\eta\right\|_{2\to 2} = \|\mathbf{\Theta} - \mathbf{\Theta}_\eta\|_{2\to 2}\,.$$

Let $\mathbf{\Delta}_\eta = \mathbf{\Theta} - \mathbf{\Theta}_\eta$. We will make use of the following inequality

$$\|\mathbf{\Delta}_\eta\|_{2\to 2}^2 \leq \|\mathbf{\Delta}_\eta\|_{1\to 1}\|\mathbf{\Delta}_\eta\|_{\infty\to\infty} \tag{E.1}$$

Since the upper-bound (23) is symmetric and $\|\mathbf{\Delta}_\eta\|_{1\to 1} = \max_{j,m}\sum_{k,n}|\Delta_{j,m,k,n}|$ and $\|\mathbf{\Delta}_\eta\|_{\infty\to\infty} = \max_{k,n}\sum_{j,m}|\Delta_{j,m,k,n}|$, it suffices to find an upper-bound on $\|\mathbf{\Delta}_\eta\|_{1\to 1}$.

By definition of $\mathbf{\Theta}_\eta$ we get that $\sum_{k,n}|\Delta_{j,m,k,n}| = \sum_{k\in\mathcal{S}}\sum_{\substack{n\\ s.t.(m,n)\in\mathcal{G}_{j,k}}}|\theta_{j,m,k,n}|$. We split the cases $k \geq j$ and $k < j$. We begin with the case $k \geq j$ and we note $\delta = k - j$. We have

$$\begin{aligned}|\theta_{j,m,k,n}| &\leq C_M 2^{-\left(M+\frac{1}{2}\right)\delta}\left(\frac{\omega_{j,k}}{\omega_{j,k} + |2^{-j}m - 2^{-k}n|}\right)^{M+1}\\ &= C_M 2^{-\left(M+\frac{1}{2}\right)\delta}\left(\frac{1+2^{-\delta}}{1+2^{-\delta}+|m-2^{-\delta}n|}\right)^{M+1}.\end{aligned}$$

The first sum on n is then bounded above by

$$\begin{aligned}\sum_{\substack{n\\ s.t.(m,n)\in\mathcal{G}_{j,k}}}|\theta_{j,m,k,n}| &\leq C_M 2^{-\left(M+\frac{1}{2}\right)\delta}\sum_{\substack{n\\ s.t.(m,n)\in\mathcal{G}_{j,k}}}\left(\frac{1+2^{-\delta}}{1+2^{-\delta}+|m-2^{-\delta}n|}\right)^{M+1}\\ &\leq 2C_M 2^{-\left(M+\frac{1}{2}\right)\delta}\sum_{n>2^\delta(m+2^j C_{j,k}^\eta)}\left(\frac{1+2^{-\delta}}{1+2^{-\delta}-m+2^{-\delta}n}\right)^{M+1}.\end{aligned}$$

We can end up as follows

$$\sum_{\substack{n\\ s.t.(m,n)\in\mathcal{G}_{j,k}}}|\theta_{j,m,k,n}|$$

$$
\begin{aligned}
&\lesssim 2^{-\left(M+\frac{1}{2}\right)\delta}(1+2^{-\delta})^{M+1}\int_{x\geq 2^{\delta}(m+2^{j}C_{j,k}^{\eta})}\left(\frac{1}{1+2^{-\delta}-m+2^{-\delta}x}\right)^{M+1}dx\\
&\leq 2^{-\left(M-\frac{1}{2}\right)\delta}(1+2^{-\delta})^{M+1}\frac{1}{M}\left(\frac{1}{1+2^{-\delta}+2^{j}C_{j,k}^{\eta}}\right)^{M}\\
&= 2^{-\left(M-\frac{1}{2}\right)\delta}(1+2^{-\delta})^{M+1}\frac{1}{M}\left(\frac{1}{1+2^{-\delta}+(1+2^{-\delta})\left(\eta^{-\frac{1}{M+1}}-1\right)}\right)^{M}\\
&\leq 2^{-\left(M-\frac{1}{2}\right)\delta}(1+2^{-\delta})\eta^{\frac{M}{M+1}} \lesssim 2^{-\left(M-\frac{1}{2}\right)\delta}\eta^{\frac{M}{M+1}}.
\end{aligned}
$$

By summing on $k \geq j$ we get

$$
\sum_{k=j}^{\log_2 N}\sum_{\substack{n\\ s.t.(m,n)\in\mathcal{G}_{j,k}}}|\theta_{j,m,k,n}| \leq \sum_{k=j}^{\log_2(N)} C_M 2^{-\left(M-\frac{1}{2}\right)\delta}\eta^{\frac{M}{M+1}} \lesssim \eta^{\frac{M}{M+1}}
$$

Finally, we can see that there exists a constant C_M depending only on M such that

$$
\|\mathbf{\Delta}_\eta\|_{1\to 1} \leq C_M\eta^{\frac{M}{M+1}} \quad \text{and} \quad \|\mathbf{\Delta}_\eta\|_{\infty\to\infty} \leq C_M\eta^{\frac{M}{M+1}}.
$$

It suffices to use inequality (E.1) to conclude. □

Proof of iii). This is a direct consequence of point i) and ii). □