Demystifying the asymptotic behavior of global denoising

Antoine Houdard, Andrès Almansa, Julie Delon

Demystifying the asymptotic behavior of global denoising

Antoine Houdard · Andrés Almansa · Julie Delon

Abstract In this work, we revisit the global denoising framework recently introduced by Talebi & Milanfar. We analyze the asymptotic behavior of its mean-squared error restoration performance in the oracle case when the image size tends to infinity. We introduce precise conditions both on the image and the global filter to ensure and quantify this convergence. We also make a clear distinction between two different levels of oracle that are used in that framework. By reformulating global denoising with the classical formalism of diagonal estimation, we conclude that the second-level oracle can be avoided by using Donoho and Johnstone's theorem, whereas the first-level oracle is mostly required in the sequel. We also discuss open issues concerning the most challenging aspect, namely the extension of these results to the case where neither oracle is required.

Keywords Diagonal Estimation, · Global Denoising, · Wiener Filtering, · Asymptotic study

A. Houdard
LTCI, Télécom ParisTech, France
CNRS MAP5, Université Paris Descartes, France
E-mail: antoine.houdard@telecom-paristech.fr

A. Almansa
CNRS MAP5, Université Paris Descartes, France
E-mail: andres.almansa@parisdescartes.fr

J. Delon
CNRS MAP5, Université Paris Descartes, France
E-mail: delon@parisdescartes.fr

1 Introduction

Image denoising usually amounts to estimate an image $\mathbf{u}$ from the observation

$$\widetilde{\mathbf{u}} = \mathbf{u} + \epsilon, \tag{1}$$

where ϵ is an additive noise. Classically, ϵ is assumed to be the realization of a random vector following a Gaussian distribution $\mathcal{N}(0, \sigma^2 I_N)$. This noise model is widely used since the Anscombe transform permits to transform the more realistic Poisson noise in a nearly Gaussian noise with fixed variance.

The story of image denoising, and more generally of image restoration, is probably as old as the story of image processing. If classical image denoising methods have always gathered various mathematical tools, such as neighborhood filters [27], variational models [22], non linear Partial Differential Equations [12] or transform domain estimation [16], the last true revolution came with the introduction of patch-based methods in 2004 [2,5]. At the time, trendy restoration approaches (total variation or wavelet thresholding) were not able to provide a satisfying trade off between texture preservation, flat areas restoration, detail reconstruction, and apparition of oscillating artifacts. By exploiting the self-similarity in images, patch-based approaches permitted to significantly improve the performance of image denoising algorithms in this regard.

Leading methods nowadays are all patch-based [7, 14,28,21]. These extremely popular approaches have been adopted in a huge range of applications. Their underlying assumption being that similar patches can be seen as independent realizations of the same distribution, the performance of a denoising algorithm should increase when the number of realization increases. Theoretically, this should lead to a form of asymptotic opti-

mality when the image size tends toward infinity. Consistency results, under stationnarity hypotheses, have been shown for instance for the DUDE algorithm [19, 26] and for the Non Local Means [5]. Now, despite their non-local nature, most of these algorithms limit the search area for similar patches to a medium-sized neighborhood around each pixel. Doing otherwise would confront them to a dilemma [10]. A larger search size means potentially more similar patches, reducing the variance of the denoising estimator. However increasing the search area in natural images also tends to increase the risk to consider dissimilar patches as similar, thus increasing the bias of the denoising estimator. Most authors found the best compromise in relatively small search areas. As a consequence, increasing the image size does not necessarily improve denoising performance. This observation was supported by extensive experimentation in [15], who showed that even if an infinite database of natural image examples was available, non-local denoising performance would attain an asymptotic performance that does not tend to infinite signal to noise ratio. Non-local methods seemed to be doomed to fundamental limits that could not be overcome.

In 2012, Talebi and Milanfar [24, 23] proposed a truly global denoising approach where each pixel is used to denoise every other pixel. They claimed in a subsequent paper [25] that this approach is asymptotically optimal, in the sense that *"the mean-squared error monotonically decays with increasing image size"*, regardless of image content, at least in an oracle scenario. In this context, this paper raises again the question: can denoising methods be fixed in such a way that they attain infinite PSNR when given an infinite number of examples (or an image of infinite size) ? They opened the debate by showing that given an oracle, such an asymptotic performance seems to be possible. However two questions are still left open:

1. What conditions has to satisfy an infinite image for the asymptotic result to hold?
2. Do these conclusions extend to the non-oracle case?

This article tries to give a precise answer to the first question, and some elements of response to the second one. To do so, we revisit the theory of diagonal estimation (refered to as Wiener filtering in Talebi's paper) that was first developed for wavelet bases. Considering images as vectors of $\mathbf{R}^N$, a diagonal estimator $\widehat{\mathbf{u}} = W\widetilde{\mathbf{u}}$ is a non linear estimator of $\mathbf{u}$ that is diagonal in a given orthonormal basis $V = \{V_i\}_{i=1,\ldots,N}$, which means that it can be written

$$\widehat{\mathbf{u}} = W\widetilde{\mathbf{u}} = V\Lambda V^T\widetilde{\mathbf{u}} = \sum_{k=1}^{N} \lambda_k(\widetilde{\mathbf{u}}) \cdot \langle \widetilde{\mathbf{u}}, V_k \rangle \cdot V_k, \qquad (2)$$

where Λ is a diagonal matrix whose k^{th} coefficient $\lambda_k(\widetilde{\mathbf{u}})$ depends on the observation $\widetilde{\mathbf{u}}$ (otherwise, the resulting estimator would be linear). The diagonal estimation framework is widely used in image processing: the basis V is often chosen as a Fourier or a wavelet basis [9, 8,16], or can for instance be built up as an orthonormal dictionnary from the image itself [20]. The success of diagonal estimation stems partly from the fact that if the image $\mathbf{u}$ is sparse in the orthonormal basis V, these "diagonal estimators are nearly optimal among all non linear estimators", as stated in [16]. The *global denoising* formalism introduced by Talebi and Milanfar [24, 23] can be reinterpreted in this context. Indeed, the idea of global denoising boils down to build V as an orthonormal basis that diagonalizes a given denoising filter (such as NLmeans [5]) computed on $\widetilde{\mathbf{u}}$. In the context of diagonal estimation, we will derive a novel asymptotic study of global denoising. Basically, we introduce precise conditions both on the image and the global filter to ensure that the mean-squared error

$$\text{MSE}(\widehat{\mathbf{u}}|\mathbf{u}) \stackrel{def}{=} \frac{1}{N} \mathbb{E}(\|\widehat{\mathbf{u}} - \mathbf{u}\|^2), \qquad (3)$$

for global image denoising decays toward zero for increasing image size. We will see that classical results of the diagonal estimation theory also permit to envision possible answers to the question of the extension of global denoising to the non oracle case.

The paper is organized as follows. In Section 2 we provide a short reminder on the theory of diagonal estimation in an orthonormal basis. The first contribution of this paper is to revisit this framework to present the global denoising formalism and to put it into perspective relatively to classical diagonal estimation results. The second and main contribution is the novel asymptotic study of global denoising presented in Section 3. Finally, in Section 4, we also discuss and show experiments on several open issues, including the extension of these results to the non-oracle case.

2 Global filtering revisited

2.1 Diagonal estimation : a short reminder

We recall here the basic properties of a diagonal estimator in terms of quadratic risk minimization, before revisiting the theory of global denoising in this context.

2.1.1 Quadratic risk

Assume that W is deterministic, *i.e.* that the coefficients λ_k are independent of the random noise ϵ and only rely on the unknown image $\mathbf{u}$. In this case, the

mean quadratic risk or mean squared error (MSE) of the diagonal estimator given by Equation (2) can be easily derived. Let us denote by $\mathbf{b}$ the projection of the unknown image $\mathbf{u}$ in the orthogonal basis V, that is $\mathbf{b} = V^T\mathbf{u}$. The MSE between $\mathbf{u}$ and $\widehat{\mathbf{u}}$ can be written as a function of the eigenvalues (λ_k) and the projection $\mathbf{b}$, as a sum of a variance and bias terms.

Proposition 1 *Let* $\widehat{\mathbf{u}} = V\Lambda V^T\widetilde{\mathbf{u}}$, *with* $V\Lambda V^T$ *a deterministic filter. Then,*

$$\text{MSE}(\widehat{\mathbf{u}}|\mathbf{u}) = \frac{1}{N}\sum_{j=1}^{N}\left((1-\lambda_j)^2 b_j^2 + \sigma^2\lambda_j^2\right). \quad (4)$$

Proof (Proof of Proposition 1)

$$\begin{aligned} N \cdot \text{MSE}(\widehat{\mathbf{u}}|\mathbf{u}) &\overset{def}{=} \mathbb{E}(\|\widehat{\mathbf{u}} - \mathbf{u}\|^2) \\ &= \underbrace{\mathbb{E}(\|V\Lambda^T\widetilde{\mathbf{u}} - V\Lambda V^T\mathbf{u}\|^2)}_{\text{variance term}} \\ &+ \underbrace{\mathbb{E}(\|V\Lambda V^T\mathbf{u} - \mathbf{u}\|^2)}_{\text{bias term}} \\ &= \mathbb{E}(\|V\Lambda V^T\epsilon\|^2) + \mathbb{E}(\|(\Lambda - I_N)V^T\mathbf{u}\|^2) \\ &= (\sum_{i=1}^{N}\lambda_i^2)\sigma^2 + \sum_{i=1}^{N}(\lambda_i - 1)^2\,\mathbb{E}[(V^T\mathbf{u})_i^2]. \end{aligned}$$

$\square$

Observe that the last equality holds only because the filter $W = V\Lambda V^T$ does not depend on the noise ϵ.

2.1.2 Oracle quadratic risk minimization

Minimization of the MSE w.r.t. *the* $\{\lambda_i\}$*'s* For a fixed orthonormal basis V, the previous MSE is a convex function of the eigenvalues λ_i, and reaches its global minimum for

$$\lambda_i^\star = \frac{b_i^2}{\sigma^2 + b_i^2}. \quad (5)$$

The corresponding minimal value of the MSE is

$$\text{MSE}^\star := \text{MSE}(\lambda^\star) = \frac{\sigma^2}{N}\sum_{j=1}^{N}\lambda_j^\star = \frac{\sigma^2}{N}\sum_{j=1}^{N}\frac{b_i^2}{\sigma^2 + b_i^2}. \quad (6)$$

This formula shares similarities with Wiener filters, with the difference that the coordinates $\{b_i\}$ are not expected values but actually depends on the oracle image $\mathbf{u}$, which is assumed to be deterministic. This oracle MSE cannot be attained in practice but only represents a lower bound for the quadratic risk of diagonal estimators in the basis V. However, it can be shown that some well chosen thresholding estimators have a risk which is not too far from the oracle one [16].

Minimization w.r.t. *the* $\{b_i\}$*'s* The previous oracle diagonal estimation is done in a given basis V, which could for instance be chosen as a Discrete Cosine Transform basis or a Wavelet basis. Obviously, the final estimation strongly depends on this choice, and one might wonder in practice how to optimize the selection of the basis V for a given image $\mathbf{u}$. The quantity $\text{MSE}^\star$ from equation (6) depends only on $\mathbf{b} = V^T\mathbf{u}$, the projection of the oracle image $\mathbf{u}$ on the basis V. The following Proposition describes the form of the $\mathbf{b}$ minimizing (6). The matrix V^T being orthonormal, the minimization is constrained by $\|\mathbf{b}\|_2 = \|\mathbf{u}\|_2$.

Proposition 2 *Minimizing* $\mathbf{b} \mapsto \text{MSE}^\star(\mathbf{b})$ *under the constraint* $\|\mathbf{b}\|_2 = \|\mathbf{u}\|_2$ *provides the following* $2N$ *global minimums*

$$\mathbf{b}^\star = \pm\|\mathbf{u}\|_2\mathbf{e}_i$$

where e_i *is the* i-*th vector of* $\mathbf{R}^N$ *basis.*

Proof (Proof of Proposition 2) Let define the function Ψ by

$$\Psi(\mathbf{b},\mu) = \frac{\sigma^2}{N}\sum_{j=1}^{N}\frac{b_i^2}{\sigma^2 + b_i^2} - \mu\left(\sum_{i=j}^{N}b_j^2 - \|\mathbf{u}\|_2^2\right),$$

where μ is a Lagrange multiplier.

The derivation with respect to the b_i yields

$$\partial_{b_i}\psi(\mathbf{b},\mu) = \frac{2\sigma^4}{N}\frac{b_i}{(\sigma^2 + b_i^2)^2} - 2\mu b_i, \quad (7)$$

and the derivation w.r.t. μ

$$\partial_\mu\psi(\mathbf{b},\mu) = \|\mathbf{u}\|_2^2 - \sum_{j=1}^{N}b_j^2. \quad (8)$$

Setting (8) to zero implies the existence of i_0 such that $b_{i_0} \neq 0$ (otherwise $\|\mathbf{u}\|_2^2$ would be zero). Then, setting (7) to zero yields for i_0

$$\mu = \frac{\sigma^4}{N}\frac{1}{(\sigma^2 + b_{i_0}^2)^2},$$

and for each $i \neq i_0$

$$b_i\frac{\sigma^4}{N}\left[\frac{(\sigma^2 + b_{i_0}^2)^2 - (\sigma^2 + b_i^2)^2}{(\sigma^2 + b_{i_0}^2)^2(\sigma^2 + b_i^2)^2}\right] = 0, \quad (9)$$

which implies $b_i = 0$ or $b_i^2 = b_{i_0}^2$. Using (8) again gives the following generic form for the critical points of Ψ

$$b_i = \begin{cases} \pm\sqrt{\dfrac{\|\mathbf{u}\|_2^2}{\#I}} & \text{if } i \in I \\ 0 & \text{otherwise,} \end{cases}$$

where $I \in \mathcal{P}\left(\{1, \ldots, N\}\right) \setminus \emptyset$ is the support of $\mathbf{b}$.

Let $\mathbf{b}$ be a critical point and I its support, we have

$$\mathrm{MSE}^{\star}(\mathbf{b}) = \frac{\sigma^2}{N} \frac{\|\mathbf{u}\|_2^2}{\sigma^2 + \frac{\|\mathbf{u}\|_2^2}{\#I}} \geqslant \frac{\sigma^2}{N} \frac{\|\mathbf{u}\|_2^2}{\sigma^2 + \|\mathbf{u}\|_2^2}$$

where the equality occurs if and only if $\#I = 1$. Thus, among all critical points, the minimal ones are the $\mathbf{b}^{\star} = \pm\|\mathbf{u}\|_2 \mathbf{e}_i$.

Finally, they are also global minima for Ψ as we have for all $\mathbf{b}$

$$\mathrm{MSE}^{\star}(\mathbf{b}) \geqslant \frac{\sigma^2}{N} \frac{\sum b_j^2}{\sigma^2 + \|\mathbf{u}\|_2^2} = \frac{\sigma^2}{N} \frac{\|\mathbf{u}\|_2^2}{\sigma^2 + \|\mathbf{u}\|_2^2} = \mathrm{MSE}^{\star}(\mathbf{b}^{\star}).$$

$\square$

The previous Proposition means that an optimal basis $V_{\star}$ should be such that $V_{\star}^T \mathbf{u} = \|\mathbf{u}\|_2 \mathbf{e}_i$, for a given i in $\{1, \ldots, N\}$. It follows that $V_{\star}$ should be composed of the vector $\frac{\mathbf{u}}{\|\mathbf{u}\|_2}$ and simply completed in an orthonormal basis. The resulting oracle filter would be

$$(W_{\star})_{ij} = u_i \frac{u_j}{\sigma^2 + \|\mathbf{u}\|_2^2}. \tag{10}$$

Again, even if this filter is not reachable since it depends on the unknown oracle $\mathbf{u}$, this results strongly support the intuitive idea that ideal bases should provide a sparse representation of $\mathbf{u}$. In practice, diagonal estimation should be applied in a well-adapted basis for each image, typically a basis V that provides a fast decrease of the $\{b_j\}$. The principle of global filtering [24], described in Section 2, is to rely on classical non linear filters from the denoising literature to choose V.

2.1.3 Non-oracle case

The oracle $\mathbf{u}$ and its projection $\mathbf{b} = V^T \mathbf{u}$ being unavailable, we need a way to approximate the previous estimation from the knowledge of $\widetilde{\mathbf{b}} = V^T \widetilde{\mathbf{u}}$. A classical solution is to consider (hard or soft) thresholding estimators in a given orthonormal basis V, in order to discard irrelevant components. As illustrated by Figure 1, the result of hard thresholding is far from being as satisfying as the oracle estimation (5), at least on a Discrete Cosine Basis. However, for a specific value of the threshold T, the mean squared error obtained with a hard or soft thresholding can still be controlled by the one of the oracle attenuation (5).

Theorem 1 (Donoho-Johnstone [16,9])
Let $T = \sigma\sqrt{2\log N}$. The MSE provided by the thresholded eigenvalues λ^{th} (with hard or soft thresholding) satisfies for $N \geq 4$

$$\mathrm{MSE}(\lambda^{th}) \leqslant (2\ln N + 1)\left(\frac{\sigma^2}{N} + 2\,\mathrm{MSE}(\lambda^{\star})\right).$$
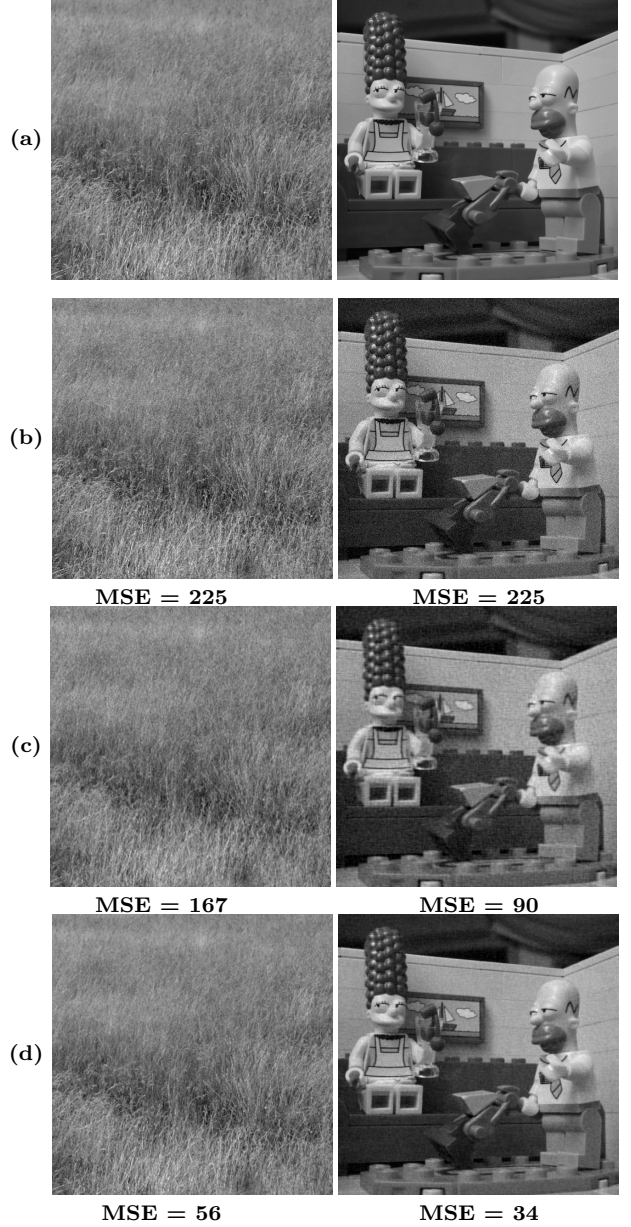

Fig. 1 (a) Original images $\mathbf{u}$, (b) Noisy images $\widetilde{\mathbf{u}}$ with $\sigma = 15$, (c) Images $\widetilde{\mathbf{u}}$ denoised by hard-thresholding in a DCT basis with a threshold $T = \sigma\sqrt{2\ln N}$, (d) Images $\widetilde{\mathbf{u}}$ denoised by diagonal estimation with the oracle λ_i^* in a DCT basis.

Proofs of this theorem can be found in [9] or [16]. It helps to predict what kind of images can be well denoised by hard thresholding in a given basis. For a DCT basis for instance, we can expect a lower oracle $\mathrm{MSE}(\lambda^{\star})$ for smoother images, and the same property should hold for thresholding. This is illustrated by Figure 1, which shows two noisy images and their respective denoised versions by oracle attenuation and hard thresholding. At the same noise level, the second image has a better oracle result (d) than the first one, and this is also true for the hard thresholding result (c). We

can also conclude that a basis V nearly optimal for the oracle should also be a good choice for the thresholding estimation. Observe that the threshold $T = \sigma\sqrt{2\ln N}$ is not really optimal in practice. A good way to fix T for soft thresholding is to resort to the SURE estimator of the MSE [16].

2.2 Global filtering in this context

Global denoising [24,23] draws on the concept of diagonal estimation in order to improve current denoising filters. As described in the previous section, a well chosen basis should provide a sparse representation of $\mathbf{u}$ and a general basis obviously cannot fit well for all natural images. Global denoising builds V as an orthonormal basis that diagonalizes a classical non linear denoising filter (such as NLmeans [5]), computed on $\widetilde{\mathbf{u}}$. The underlying assumption is that if the chosen denoising filter is well adapted to the image, the coefficients b_j will decrease relatively quickly and the diagonal estimate will be all the more efficient.

2.2.1 Principle of global denoising

Assuming the same image formation model (1), numerous classic denoising filters, such as Gaussian or bilateral filters as well as NL-means [5] type filters, can be written under the form

$$\widehat{\mathbf{u}} = W\widetilde{\mathbf{u}}, \tag{11}$$

where $W = D^{-1}K$, with K the positive definite kernel from the filter and D a diagonal matrix with entries $D_{ii} = \sum_j K_{ij}$, $i \in \{1, \dots, N\}$ [1]. Starting from a given denoising filter W, the idea of global denoising, made popular by Milanfar in [18], [17], is to modify this filter W, in order to decrease the mean square error between $\widehat{\mathbf{u}}$ and $\mathbf{u}$. For instance, if we assume that W can be diagonalized in an orthonormal basis V (this can be ensured by symmetrizing it, as described in the next section), the oracle attenuation (5) of the eigenvalues can be applied to improve the filter.

2.2.2 Symmetrizing the filter W

To ensure the fact that W can be diagonalized in an orthonormal basis, the authors of [24] propose to replace

---

[1] For instance, for NL-means we would have $K_{i,j} = e^{-\frac{\|P_i - P_j\|^2}{2h^2}}$, with P_i and P_j the patches centered at i and j and h a parameter

---

this filter by a symmetric doubly stochastic version W^s of W which minimizes the cross-entropy

$$\sum_{i,j} W_{ij}^s \log \frac{W_{ij}^s}{W_{ij}}. \tag{12}$$

In practice, this minimization problem can be solved numerically with the Sinkhorn algorithm, which consists in iteratively normalizing the rows and the columns of W until convergence. Starting from a positive definite kernel K, it can be shown that the resulting filter W^s is positive definite, symmetric and doubly stochastic, and that its eigenvalues are very close to those of W [18]. In practice, the denoising results obtained with this symmetrized filter appear to be equivalent or slightly better than the ones obtained with W [18,6]. In the following, we always consider the filter W in its symmetric and doubly stochastic version.

2.2.3 Deterministic filter

The mean-squared error formulation (4) is valid only if the filter $W = V\Lambda V^T$ is deterministic. This assumption is sensible when V is fixed, as a DCT or wavelet basis for instance. However when the filter W comes from the noisy image $\widetilde{\mathbf{u}}$, this hypothesis does not hold. To illustrate this fact, we compare, for different choices of the filter W, the theoretical $\mathrm{MSE}_{\mathrm{theo}}$ computed by formula (4) with the experimental mean-squared error

$$\mathrm{MSE}_{\mathrm{eval}} = \frac{1}{N}\sum_{j=1}^{N}|\mathbf{u}_j - W\widetilde{\mathbf{u}}_j|^2. $$
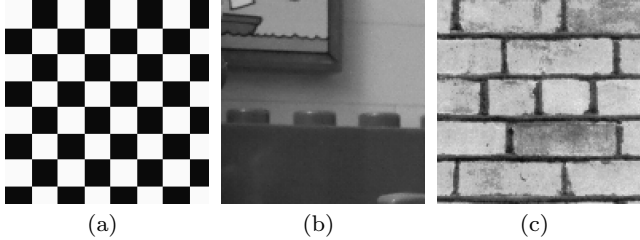
Figure 2 shows the relative error

$$\frac{|\mathrm{MSE}_{\mathrm{eval}} - \mathrm{MSE}_{\mathrm{theo}}|}{\mathrm{MSE}_{\mathrm{theo}}} \tag{13}$$

for the following filters W, computed on three different images:

1. a Non-Local Means filter [5] computed on the original image $\mathbf{u}$ (called Oracle-NLM or O-NLM);
2. a Non-Local Means filter computed on the noisy image $\widetilde{\mathbf{u}}$ (called NLM);
3. a Non-Local Means filter computed on a version of $\widetilde{\mathbf{u}}$ already denoised by NL means (called pre-filtered NLM or P-NLM);

The first filter is independent from the noise present in $\widetilde{\mathbf{u}}$, so the relative error is very small. On the contrary, the NL-means filter computed directly on the noisy image strongly depends on the noise realization, and the relative error between the theoretical and experimental MSE remains above 10% for all three images. Finally, observe that if the NLM is computed on a version of

$\widetilde{\mathbf{u}}$ that has already been denoised in a first step (by a NLM kernel or another denoising procedure), the resulting W seems to be partly decoupled from the noise, at least enough for the theoretical $\mathrm{MSE}_{\mathrm{theo}}$ to be a good predictor of $\mathrm{MSE}_{\mathrm{eval}}$.

	(a)	(b)	(c)
images	**(a)**	**(b)**	**(c)**
O-NLM	6.6 % (± 3.0)	0.3 % (± 1.9)	0.4 % (± 0.9)
NLM	34.2 % (± 3.0)	24.6 % (± 1.9)	11.2 % (± 0.9)
P-NLM	6.8 % (± 1.7)	2.7 % (± 0.3)	1.0 % (± 0.3)

Fig. 2 Relative error (13) between the theoretical MSE provided by formula (4) and the experimental MSE, for three images and three different filters W (NL-Means computed on the $\mathbf{u}$, NL-Means computed on the $\widetilde{\mathbf{u}}$, NL-Means computed on a prefiltered version of $\widetilde{\mathbf{u}}$. The mean and standard deviation have been computed on 5 different realisations of noise with $\sigma = 15$ for each image.

2.2.4 Two oracle levels

In the previous section we used the noiseless image $\mathbf{u}$ as an oracle to compute the weights of the O-NLM filter W. Note that this use of the oracle is different from the one introduced in equation (5) (section 2.1.2) to compute the optimal eigenvalues $\lambda_j^\star$. In general it will be clear from the context which level of oracle we refer to. In ambiguous cases we shall refer to the first one as W-oracle and to the second one as λ-oracle.

2.2.5 Discussion

By producing a basis V that is well-adapted to the image we want to denoise, global image denoising usually produces better results than a diagonal estimation on a DCT or wavelet basis. However, global denoising still suffers from two major issues:

- first, the λ-oracle $\mathbf{u}$ is needed in order to optimize the eigenvalues;
- second, memory cost and computation time are untractable because of the eigendecomposition of the filter W of size $N \times N$.

In order to bypass the first issue, we saw in section 2.1.3 that hard or soft thresholding could provide MSE re-

sults controlled by the optimal MSE*. Another possibility would be to try multiple sets of eigenvalues and keep the ones minimizing a SURE estimator of the MSE. This is the solution proposed by the GLIDE algorithm [23]. The second issue can be solved by computing only a small percentage of eigenvectors. In GLIDE, Talebi and Milanfar make use of the Nyström extension in order to approximate the filter W and its first eigenvalues.

3 Asymptotic study

In this part, we study the asymptotic behavior of the MSE given by formula (4) when the image size increases. In [25], the authors claim that global denoising is asymptotically optimal, in the sense that the MSE in (3) tends to zero. Before going further, let us mention that this decay of the global MSE may occur while some local areas of the image remain poorly denoised even when the image size tends to infinity, as it is shown on Figure 3. In order to explore the precise conditions of this convergence, we define in Section 3.2 a reasonable model for an image whose size grows to infinity. We also assume a parametric model for the decay of the coefficients b_j, and we derive in Section 3.3 different conditions of convergence for the MSE and its corresponding decay rate. Finally, in Section 4.2, we discuss and illustrate these different results and the realism of these models for different choices of images and filters W.

In the following, we always consider that the filter W is independent from the noise ϵ.

3.1 Upper bound on the optimal MSE

We have seen in Section 2.1 that the oracle risk for diagonal estimation was given by

$$\mathrm{MSE}^\star = \frac{\sigma^2}{N} \sum_{j=1}^{N} \frac{b_j^2}{\sigma^2 + b_j^2},$$

with $\mathbf{b} = V^T \mathbf{u}$ the projection of the oracle image $\mathbf{u}$ in the eigenbasis V. Now, this MSE can be upper bounded by the l^1-norm of $\mathbf{b}$ divided by N:

$$
\begin{aligned}
\mathrm{MSE}^\star &= \frac{\sigma^2}{N} \sum_{j=1}^{N} \frac{b_j^2}{\sigma^2 + b_j^2} \\
&\leqslant \frac{\sigma^2}{N} \sum_{j=1}^{N} \frac{b_j^2}{2\sigma |b_j|} \\
&= \frac{\sigma}{2N} \|\mathbf{b}\|_1.
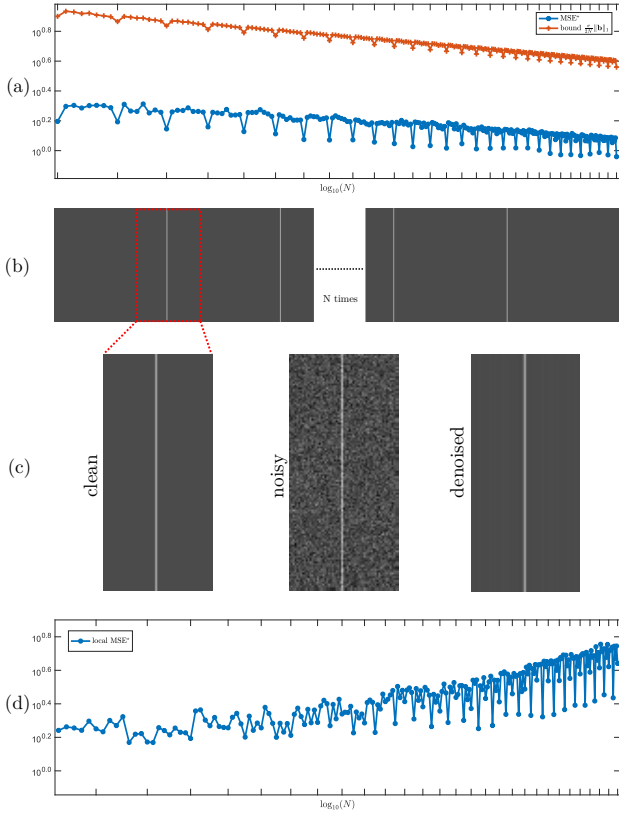\end{aligned}
\tag{14}
$$

Fig. 3 Example of denoising with a global MSE decaying to zero when image size grows to infinity, while a local MSE increases. **(b)** is an image constructed by repeating N times a pattern of a constant image with a vertical line. **(a)** presents the behavior of the MSE between the denoised image obtained by optimal diagonal estimation in a DCT basis and the clean image. The MSE is shown in $\log_{10}$ scale, when the pattern is repeated N times with N increasing. **(c)** presents a zoom on the structured part of the image and the result of the denoising for $N = N_{\max}$. The denoised image presents important ringing artifacts. Finally, **(d)** shows the behavior of the local MSE on the part presented in **(c)** when N grows. This shows that even with a global MSE converging to zero, the restoration can remain locally bad and even get worse when the image size increases.

The authors of [25] suggest that this upper bound might converge towards 0 when N grows to infinity. In order to prove this convergence, they assume that the sorted coefficients $|b_j|$ drop off at a given rate $\alpha > 0$

$$|b_j| \leqslant \frac{C}{j^\alpha}.$$

We shall see below and in section 3.2 that this models requires C to depend on N to make sense. Nevertheless for a fixed image size, this hypothesis seems quite reasonable for different existing filters, as illustrated by Figure 4. When working with Fourier or space-frequency decompositions, the value of α was shown to be related to the regularity of the image [16], and values of α be-
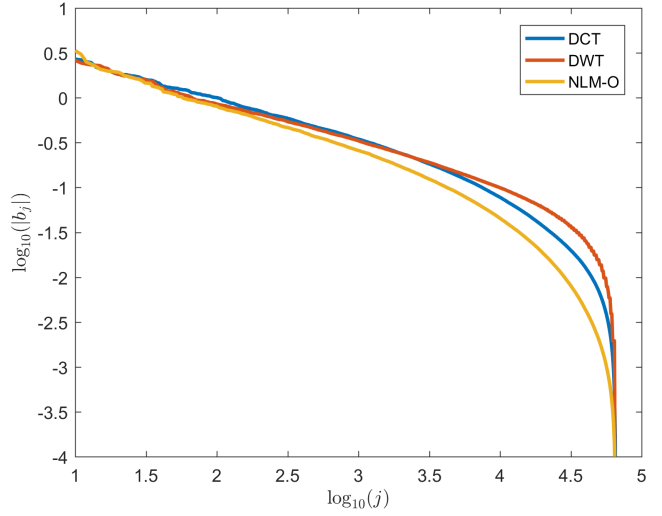

Fig. 4 Decay rate of the coefficients b_j for the image **synthetic** from Figure 7, in a loglog scale graph. In blue: coefficients in a DCT basis. In red: coefficients in a wavelet basis. In yellow: coefficients in the eigenbasis of the oracle NLM filter.

tween 0.5 and 1 were shown to be in agreement to actual image data [11]. Such models have also been used in asymptotic studies where the image *resolution* tends to infinity, but here we are interested in the asymptotic behavior when image *size* grows to infinity at constant resolution. In this particular kind of asymptotic study, we cannot expect the rate α and the constant C to remain constant when the image size grows towards infinity. Put another way, there is no reason that we can bound the b_j coefficients independently of the image size N. To demonstrate this claim, we propose a model for an image whose size grows to infinity. Under this model we will show that the coefficients b_j^N actually depend on N. Then we propose a more complete parametric model for the coefficients decay[2]

3.2 Proposed models

Infinite image model Consider an image of infinite size

$$\mathbf{U} : \mathbb{Z}^2 \to \{l, \dots, L\},$$

taking values in a discrete set of gray levels $\{l, \dots, L\} \subseteq \mathbb{N}$. For typical 8 bit images $l = 0$ and $L = 255$.

From this image we construct an infinite sequence of images of growing size N

$$\mathbf{u}^N \overset{def}{=} \left(U_{\varphi(1)}, \dots, U_{\varphi(N)} \right),$$

---

[2] From now on, we will write b_j^N instead of b_j to remember that the behavior of these coefficients strongly depend on the image size N.

by truncating the infinite image to size N, for all $N \in \mathbb{N}$. The function $\varphi : \mathbb{N}^+ \to \mathbb{Z}^2$ sweeps the plane in spiral starting from the origin.

Since the image gray level values are bounded, the L^2-norm of $\mathbf{u}^N$ satisfies the following inequality

$$l\sqrt{N} \leqslant \|\mathbf{u}^N\|_2 \leqslant L\sqrt{N}, \qquad (15)$$

which means that the energy of the growing image increases at most like $\mathcal{O}(\sqrt{N})$. This information on the L^2-norm of $\mathbf{u}^N$ is important because it constrains the behavior of $\mathbf{b}^N$ as $\|\mathbf{b}^N\|_2 = \|\mathbf{u}^N\|_2$.

Because generally the lowest value l is zero, the energy of the image may not grow as fast as $\sqrt{N}$. However we show that if $\|\mathbf{u}^N\|_2 = o(\sqrt{N})$, the image is becoming sparse with increasing size: because $\mathbf{U}$ is taking values in a discrete finite set, by setting $c = \min\{U_i \neq 0, i \in \mathbb{N}\}$ we have

$$c^2 \frac{\#\{u_i^N \neq 0\}}{N} \leqslant \frac{\|u^N\|_2^2}{N} \xrightarrow[N \to \infty]{} 0. \qquad (16)$$

This shows that the ratio of non-zero pixels collapses when the image size goes to infinity. This case will not be considered in the following. Indeed if we consider an infinite image $\mathbf{U}$ such that $\|\mathbf{u}^N\|_2 = o(\sqrt{N})$ then, the upper-bound (14) tends to zero when N goes to infinity:

$$\|\mathbf{b}^N\|_1 \leqslant \sqrt{N}\|\mathbf{b}^N\|_2 = \sqrt{N}\|\mathbf{u}^N\|_2 = o(N),$$

which implies the convergence. We provide in Section 4 an experiment with an image padded with zeros illustrating this case.

This leads us to define the widespread infinite image model as follows.

Hypothesis 1 (Widespread infinite image model) *Let* $\mathbf{U}$ *be an infinite image and denote* $\mathbf{u}^N$ *its truncation of size* N. *Then* $\mathbf{U}$ *is said to be* non sparse *if there exists* $m > 0$ *and* $M > 0$ *such that*

$$m\sqrt{N} \leqslant \|\mathbf{u}^N\|_2 \leqslant M\sqrt{N}. \qquad (17)$$

Domination decay model Now consider a sequence of orthogonal bases V^N (the eigenbases of symmetric filtering operators W^N). Recall that we denote by $\mathbf{b}^N = V^N \mathbf{u}^N$ the projection of the image of size N on the corresponding eigenbasis. We need a realistic model on the asymptotic behaviour of b_j^N when N, j go to infinity. In this part we design an upper bound for $|b_j^N|$ which is both

- simple and easy to manipulate to prove convergence results;
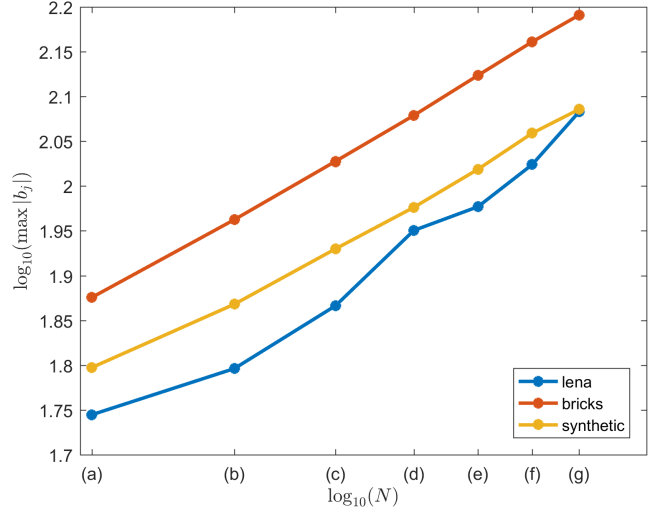- adapted to the data, in the sense it has the same shape as the $|b_j^N|$.

Fig. 5 Behaviour of $\max_j(|b_j^N|)$ with increasing size N for three different images from Figure 7 in loglog scale. Here $\mathbf{b}^N$ is the image $\mathbf{u}$ projected in the DCT basis.

In order to design it we start with the model from [23] namely

$$|b_j| \leqslant \frac{C}{j^\alpha},$$

with $\alpha > 0$. If we consider such a model for all N with an image verifying Hypothesis 1 then we have

$$m^2 N \leqslant \|\mathbf{u}\|_2^2 = \|\mathbf{b}\|_2^2 \leqslant C^2 \sum_{j=1}^N \frac{1}{j^{2\alpha}}.$$

This implies the divergence of the sum in the right term, which is thus equivalent to $N^{1-2\alpha}$ when N goes to infinity. This yields $m^2 = \mathcal{O}(N^{-2\alpha})$ and so $\alpha \leqslant 0$ which is a contradiction. As a consequence, the constant C should depend on N. In the following, we consider the model

$$|b_j^N| \leqslant \frac{C_N}{j^\alpha},$$

and we discuss how to simplify it based on information given by numerical experiments. We need to define how the "constant" C_N grows with the image size N. Figure 5 shows that $\max_j(|b_j^N|)$ is increasing linearly with N in loglog scale. That leads us to consider N^γ as a model for C_N. Finally, we consider the following decay model for $\mathbf{b}^N$:

Hypothesis 2 (Domination decay model) *Let* $\mathbf{U}$ *be an infinite image and denote by* $\mathbf{u}^N$ *its truncation of size* N. *Let* $\mathbf{V} = (V^N)_N$ *be a family of orthogonal bases of increasing size. Then the pair* $(\mathbf{U}, \mathbf{V})$ *is said to fit*

the domination decay model with parameters C, α and $\gamma > 0$ *if for all* $N, j \in \mathbb{N}$

$$|b_j^N| \leqslant C \frac{N^\gamma}{j^\alpha}. \tag{18}$$

where $\mathbf{b}^N = V^N \mathbf{u}^N$ *is the projection of* $\mathbf{u}^N$ *on the basis* V^N.

A case where hypotheses 1 and 2 are trivially satisfied is the case of constant images with a DCT filter. Indeed, the corresponding **b** is the optimal one from Proposition 2. In the next section, we study the convergence of the upper-bound of the MSE under these hypotheses.

3.3 Conditions of convergence

In Section 2.1 we showed that the optimal diagonal estimator on a given basis V^N could be bounded in terms of the ℓ^1-norm of the coefficients $\mathbf{b}^N$ in that basis. In the following, we show that under Hypotheses 1 and 2, this ℓ^1 norm can in turn be upper-bounded by a decreasing function of N,

$$\mathrm{MSE}(\lambda^\star) \leqslant \frac{\sigma}{2N} \|\mathbf{b}\|_1 \leqslant C' \frac{1}{N^r}, \tag{19}$$

thus ensuring convergence of the optimal MSE at a rate r that depends on the parameters α and γ of the decay model. When this rate is positive then we can use this second upper bound to prove the asymptotic optimality of diagonal estimation on that basis.

Theorem 2 (Asymptotic optimality) *Consider*

- *an infinite image* **U** *that satisfies Hypothesis 1 (*i.e. *non-sparsity) and*
- *a sequence of orthogonal bases* **V** *such that the pair* **(U, V)** *satisfies Hypothesis 2 (*i.e. (C, α, γ) *decay rate of the image projection on that basis).*

  If the decay rate is fast enough, i.e. *if*

$$\frac{1}{2} \leqslant \gamma < 1 \quad and \quad \alpha > \gamma,$$

then the denoising provided by oracle optimization of diagonal estimation on that basis is asymptotically optimal *meaning that the* MSE *tends to 0 when the image size* N *goes to infinity.*

The proof of this result is the combination of the two following Lemmas. The first one shows that the hypothesis on image energy (17) constrains the parameters α and γ of the domination criterion and the second one further restricts the values of these parameters to ensure convergence of the upper-bound of the MSE.
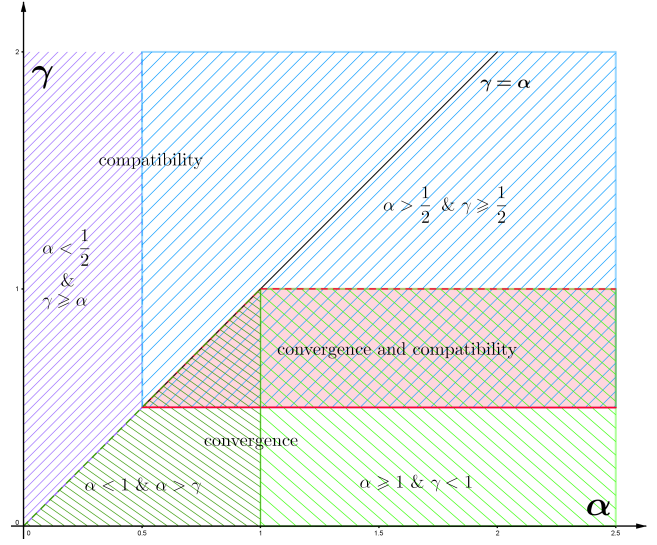

Fig. 6 Illustration of the domain of compatibility and the domain of convergence provided by the two lemmas 1 and 2. The intersection in red represent the set of parameters that provides the result in Theorem 2.

Figure 6 illustrates the results provided by Lemma 1 and lemma 2 on the parameters α and γ. The resulting parameters for Theorem 2 are given by the intersection of the two domains.

Lemma 1 (Compatibility with image model) *Assume that* **U** *satifies Hypothesis 1. If the projection of* **U** *on* **V** *satisfies the decay model of Hypothesis 2 with parameters* (C, α, γ), *then either*

$$\gamma \geqslant \frac{1}{2} \ and \ \alpha \geq \frac{1}{2}$$
$$or$$
$$\gamma \geqslant \alpha \ and \ \alpha < \frac{1}{2}.$$

This lemma emphasizes the fact that we actually cannot bound the $|b_j^N|$ independently of N as long as we have images that are not loosing energy with increasing size. The only way to obtain $\gamma = 0$ (a bound independent of N) is to impose $\alpha = 0$ which leads to the pathological case $\mathbf{b} \propto (1, \dots, 1)$.

Proof (Proof of Lemma 1) Because $\|\mathbf{b}^N\|_2 = \|\mathbf{u}^N\|_2$, the model (17) on the image energy gives

$$m^2 N \leqslant \|\mathbf{b}^N\|_2^2 \leqslant M^2 N.$$

Applying the decay criterion $|b_j^N| \leqslant C \frac{N^\gamma}{j^\alpha}$ in the previous equation yields

$$m^2 N \leqslant C^2 N^{2\gamma} \sum_{j=1}^N \frac{1}{j^{2\alpha}}.$$

The behavior when N goes to infinity of the sum in the right term differs depending on α:

– if $\alpha < \frac{1}{2}$ the sum diverges and there exists a constant C' such that

$$\sum_{j=1}^{N} \frac{1}{j^{2\alpha}} \underset{N \to \infty}{\sim} C' N^{1-2\alpha}$$

– if $\alpha = \frac{1}{2}$ the sum diverges and there exists a constant C' such that

$$\sum_{j=1}^{N} \frac{1}{j^{2\alpha}} \underset{N \to \infty}{\sim} C' \ln N$$

– if $\alpha > \frac{1}{2}$ the sum converges to a constant $C' \in \mathbb{R}$

The first case leads to $m^2 = \mathcal{O}\left(N^{2\gamma-2\alpha}\right)$ and so $\alpha \leqslant \gamma$. The second and the third cases lead to $m^2 = \mathcal{O}\left(N^{2\gamma-1} \ln N\right)$ and $m^2 = \mathcal{O}\left(N^{2\gamma-1}\right)$ respectively and so $\gamma \geqslant \frac{1}{2}$.   □

Lemma 2 (condition of convergence) *Considering the model (18) we have convergence to zero of the bound (14) only if*

$$\alpha \geqslant 1 \ and \ \gamma < 1$$
$$or$$
$$\alpha < 1 \ and \ \alpha > \gamma.$$

This lemma shows that the convergence can actually occur with all $\alpha > 0$ as long as γ is not too large. We also notice that the model proposed in [23] in $\frac{C}{j^\alpha}$ satisfies the convergence hypothesis. However, we saw with that this model is not compatible with Hypothesis 1.

Proof (Proof of Lemma 2) We have $|b_j^N| \leqslant C \dfrac{N^\gamma}{j^\alpha}$ so

$$\frac{\sigma}{2N} \|\mathbf{b}\|_1 \leqslant \frac{C\sigma}{2N} \sum_{j=1}^{N} \frac{N^\gamma}{j^\alpha} = \frac{C\sigma}{2} N^{\gamma-1} \sum_{j=1}^{N} \frac{1}{j^\alpha}.$$

The behavior when N goes to infinity of the sum in the right term differs depending on α:

– if $\alpha < 1$ the sum diverge and there exists a constant C' such that

$$\sum_{j=1}^{N} \frac{1}{j^\alpha} \underset{N \to \infty}{\sim} C' N^{1-\alpha}.$$

– if $\alpha = 1$ the sum diverges and there exists a constant C' such that

$$\sum_{j=1}^{N} \frac{1}{j^\alpha} \underset{N \to \infty}{\sim} C' \ln N.$$

– if $\alpha > 1$ the sum converges to a constant $C' \in \mathbb{R}$.

The first case leads to

$$\frac{\sigma^2}{N} \|\mathbf{b}\|_1 = \mathcal{O}\left(N^{\gamma-\alpha}\right),$$

and convergence occurs only if $\alpha > \gamma$. The second and the third cases lead respectively to

$$\frac{\sigma^2}{N} \|\mathbf{b}\|_1 = \mathcal{O}\left(N^{\gamma-1} \ln N\right),$$

and

$$\frac{\sigma^2}{N} \|\mathbf{b}\|_1 = \mathcal{O}\left(N^{\gamma-1}\right),$$

and convergence occurs only if $\gamma < 1$.   □

The proof of Lemma 2 also provides a decay rate that we summarize in the following corollary.

Corollary 1 (Decay rate) *Under conditions of convergence in Theorem 2, that is $\frac{1}{2} \leqslant \gamma < 1$ and $\alpha > \gamma$ the MSE of optimal diagonal oracle estimation satisfies*

$$\mathrm{MSE}(\lambda^\star) \underset{N \to \infty}{=} \mathcal{O}\left(\frac{1}{N^r}\right)$$

with $r \in]0, \frac{1}{2}]$ defined by

– *$r = \alpha - \gamma$ when $\gamma < \alpha < 1$*
– *$r = 1 - \gamma$ when $\alpha > 1$*

The particular case $\alpha = 1$ yields convergence in $\mathcal{O}\left(\frac{\log N}{N^{1-\gamma}}\right)$.

This result shows that the decay is always slower than $\frac{1}{\sqrt{N}}$ and it can be really slow when r is close to zero. Thus, even though we can have an asymptotic optimal filtering, the decay rate can be so small that we cannot actually see it even if we work with huge images. Moreover, this asymptotic study is performed on the oracle diagonal filter. This result is by itself essentially theoretical. However, in combination with Donoho-Johnstone Theorem 1, we might further use this result to prove, under specific conditions on the infinite image, the asymptotic optimality of non-oracle filtering.

Corollary 2 (Decay rate of thresholding) *Assume that the convergence conditions of Theorem 2 are satisfied. From the Donoho-Jonhstone Theorem 1, the MSE obtained by thresholding the coefficients b_j^N satisfies*

$$\mathrm{MSE}(\lambda^{th}) \underset{N \to \infty}{=} \mathcal{O}\left(\frac{\log N}{N^r}\right)$$

with $r \in]0, \frac{1}{2}]$ defined as in Corollary 1. The particular case $\alpha = 1$ yields convergence in $\mathcal{O}\left(\frac{(\log N)^2}{N^{1-\gamma}}\right)$.

Proof (Proof of Corollary 2) Let consider the case $\alpha \neq 1$. By Donoho-Jonhstone Theorem we have

$$\mathrm{MSE}(\lambda^{th}) \leqslant (2\log N + 1)\left(\frac{\sigma^2}{N} + 2\,\mathrm{MSE}(\lambda^\star)\right).$$

Then by Corollary 1 there exists a constant C such that

$$\mathrm{MSE}(\lambda^\star) \leqslant \frac{C}{N^r},$$

with $r \in]0, \frac{1}{2}]$. It follows that

$$\mathrm{MSE}(\lambda^{th}) \leqslant 2C\frac{\log N}{N^r} + (2\log N + 1)\frac{\sigma^2}{N} + \frac{2C}{N^r}.$$

The two last terms in the previous inequality are $o\left(\frac{\log N}{N}\right)$ when N goes to infinity that yield the announced result. A similar proof can be done for the case $\alpha = 1$. $\square$

3.4 Special cases

In the following two paragraphs we discuss some simple particular cases in which the asymptotic behaviour of global denoising can be directly deduced.

In more realistic cases we need to experimentally fit our image model to natural images for different bases in order to predict what would happen when the image size tends to infinity. This experimental study is deferred to section 4.

3.4.1 Optimal basis with optimal eigenvalues

When an oracle is used both to choose the optimal basis V and the optimal eigenvalues λ of the filter $W = V\Lambda V^T$, we showed in Proposition 2 that the optimal MSE decays like σ^2/N, so we have $r = 1$, a much faster convergence than in the more realistic cases based on an image model. In this case only b_1 is non zero for all values of N, so computing α and γ does not make any sense.

3.4.2 Gaussian textures on DCT basis

Another case of interest is the case when the image $\mathbf{u}$ is a Gaussian texture, meaning that

$$\mathbf{u} = \mathbf{h} * \mathbf{m}$$

is generated by convolving a known kernel $\mathbf{h}$ with a white noise image $\mathbf{m}$ where $m_i \sim \mathcal{N}(0, \tau^2)$ iid.

In this case when choosing V as a Fourier or DCT basis, a straightforward calculation (see appendix A)

shows that the MSE upper bound in equation (14) for global filtering with optimal λ_j in that basis becomes

$$\begin{aligned}\mathrm{MSE}^*_{\mathrm{bound}} &= \frac{\sigma^2}{2N}\|\mathbf{b}\|_1 \\ &= \frac{1}{N^2}\frac{\sigma}{2}\sum_k |\widehat{\mathbf{h}_N}(k)||\widehat{\mathrm{m}_N}(k)| =: A_N.\end{aligned}$$

Thus asymptotically we have a strictly positive MSE bound

$$\mathrm{MSE}^*_{\mathrm{bound}} = A_N \xrightarrow[N\to\infty]{} A_\infty = \frac{\sigma\tau\|\hat{\mathbf{h}}\|_1}{\sqrt{2\pi}} > 0$$

for Gaussian textures when using the Fourier or DCT basis. Our experiments (see Section 4.3) confirm this finding. Indeed, when choosing a Fourier or DCT basis $V = F$, then MSE^*_N remains constant when $N \to \infty$, in this case $r \approx 0$. Nevertheless, when choosing an adaptive basis V from the diagonalization of the non-local means filtering operator, then MSE^*_N does experimentally tend to 0 for Gaussian textures. This shows that the NLM basis may better exploit the self-similarity in Gaussian textures.

3.4.3 Oracle vs. non oracle filters

Hypothesis 2 on the domination decay model assumes that the family of orthogonal bases $\mathbf{V}$ is well adapted to the infinite image $\mathbf{U}$. This happens in particular when the chosen filters are oracle filters, which means that they are computed in the image itself. Consider for instance the case of a very simple oracle filter which consists in denoising $\tilde{\mathbf{u}}$ by averaging at pixel i all values $\tilde{\mathbf{u}}_j$ such that $|u_i - u_j| \leq \varepsilon$ for a given threshold $\varepsilon > 0$. If the infinite image $\mathbf{U}$ is bounded, for instance with values in $[0, 1[$, we can show that the value $\mathrm{MSE}(\widehat{\mathbf{u}^N}|\mathbf{u}^N)$ converges to a limit smaller than ε^2 when the image size increases. This result being satisfied for every $\varepsilon > 0$, the MSE of this oracle filter is naturally asymptotically optimal. A proof of this example is given in appendix B.

The case of non oracle filters is of course far more ambiguous. Consider for instance the case of a dead leaves infinite image model studied in [15]. The previous argument shows that a well chosen oracle filter would denoise this image perfectly. However, because of the independence between the leaves, it is clearly not possible to achieve a null asymptotic MSE for a non oracle filter, since for a given leaf, the values observed outside of the leaf are useless to denoise the pixels inside the leaf.
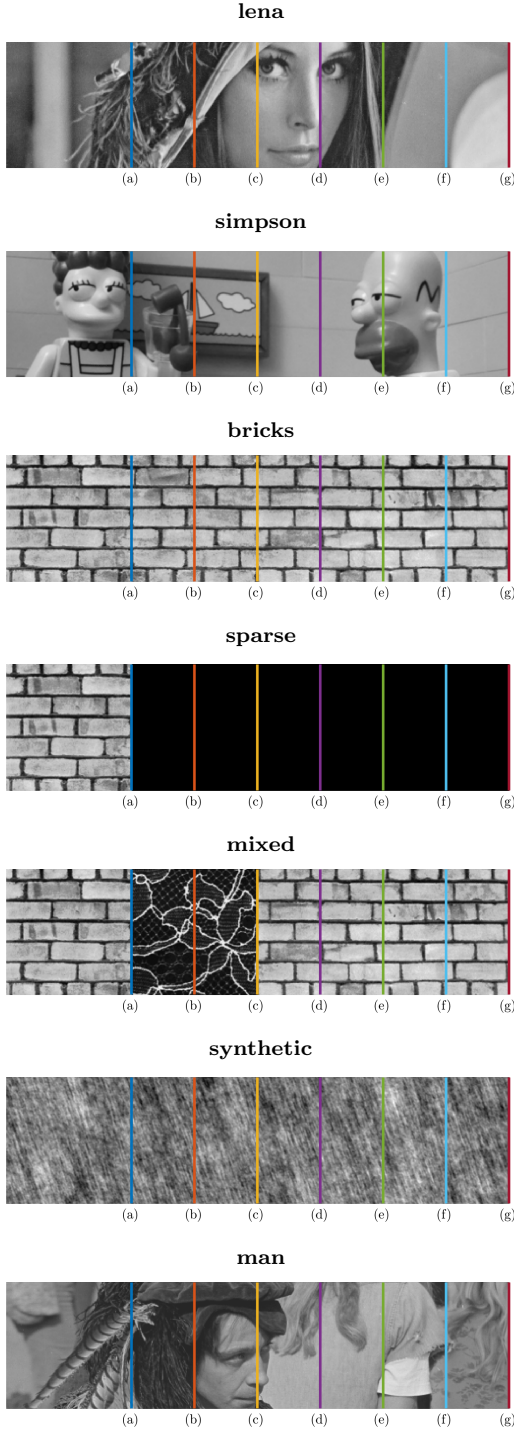
lena

(a) (b) (c) (d) (e) (f) (g)

simpson

(a) (b) (c) (d) (e) (f) (g)

bricks

(a) (b) (c) (d) (e) (f) (g)

sparse

(a) (b) (c) (d) (e) (f) (g)

mixed

(a) (b) (c) (d) (e) (f) (g)

synthetic

(a) (b) (c) (d) (e) (f) (g)

man

(a) (b) (c) (d) (e) (f) (g)

Fig. 7 The images used for the experiments. The sub-images sizes are (a) 128×128, (b) 128×192, (c) 128×256, (d) 128×320, (e) 128×384, (f) 128×448, (g) 128×512. Images credits: **lena** and **man** are standard images used in image processing, **simpson** is from Julie Delon, **brick**, **sparse** and **mixed** are Brodatz textures [4] and **synthetic** is a random generated gaussian texture.

4 Experiments

In the previous section we introduced a decay model for the b_j^N coefficients of natural image sequences decomposed on the orthonormal basis given by a symmetric filtering algorithm.

We gave precise conditions on the (γ, α) parameters of this decay model. These conditions may be used to determine whether optimal diagonal estimation on this basis can yield asymptotically optimal denoising performance when applied to a certain family of image sequences.

In practice, answering this question requires to estimate these coefficients from a particular filter/basis based on a truncated image sequence. The next Section 4.1 explains how these model parameters are estimated from real data. Then in Section 4.2 we analyze the asymptotic performance of several denoising algorithms based on the estimated parameters. Finally, in Section 4.3 we provide a discussion about some specific cases that are of particular interest to illustrate our analysis.

4.1 Estimating model parameters (C, γ, α)

Theorem 2 gave us a sufficient condition for asymptotic optimality of a filter on an image sequence. This condition is based on the assumption that the $|b_j^N|$ coefficients follow a particular model, namely:

$$|b_j^N| \approx C \frac{N^\gamma}{j^\alpha}. \tag{20}$$

Observing different curves $j \mapsto |b_j^N|$ for various images, sizes N and orthonormal bases in loglog scale (see the first column of Figure 11 for an example), we notice that the model (20) holds except for the first few largest coefficients and for a significant proportion of the smallest coefficients. This behaviour can be expected, since we sorted the coefficients. It appears even when the $|b_j^N|$ coefficients are only white noise (as illustrated in Figure 8). Thus we exclude the values of $j < d = 5$ and $j > N^p$ (for $p = 0.6$) from the bilinear regression that allows to fit the values of C, α and γ to the $|b_j^N|$ coefficients.

Put another way we find α, γ and C that minimize

$$\| \log(|b_j^N|) - (\gamma \log(N) - \alpha \log(j) + \log(C)) \|_2,$$

with N from $N_{\min}$ to $N_{\max}$ and j from d to $\lfloor N^p \rfloor$.
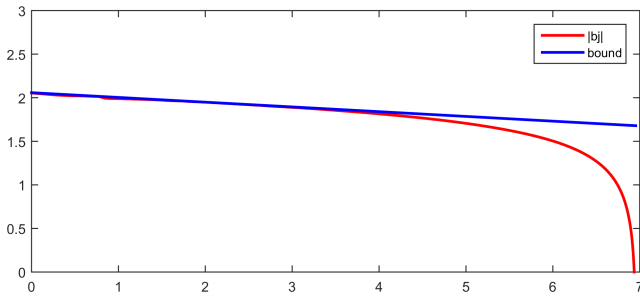
Fig. 8 Decay of the coefficients $|b_j|$ for white noise in the DCT basis (red) in loglog scale. The slope of the bound is $\alpha_m \approx 0.05$ (blue).

Table 1 Fitted parameters α and γ for the different images of figure 7 in the three bases DCT, DWT and O-NLM. The parameter r is the decay rate of corollary 1.

image	basis	fitted α	fitted γ	r
lena	DCT	0.827	0.767	0.060
	DWT	0.858	0.752	0.10
	O-NLM	0.806	0.570	**0.236**
simpson	DCT	0.941	0.741	0.201
	DWT	1.106	0.753	0.247
	O-NLM	1.025	0.572	**0.428**
bricks	DCT	0.796	0.755	0.041
	DWT	0.913	0.877	0.035
	O-NLM	0.902	0.688	**0.213**
sparse	DCT	0.842	0.375	0.467
	DWT	0.909	0	0.909
	O-NLM	1.099	0.021	**0.979**
synthetic	DCT	0.708	0.727	-
	DWT	0.554	0.570	-
	O-NLM	0.646	0.540	**0.106**
man	DCT	0.720	0.520	0.200
	DWT	0.802	0.578	0.224
	O-NLM	0.759	0.366	**0.393**

4.2 Experimental results

Table 1 shows the estimated model parameters for the test images from Figure 7 and for three orthonormal bases, namely:

DCT: The DCT basis which diagonalizes convolution filters;
Wavelet: The orthogonal Haar basis, implemented via the discrete wavelet transform;
Oracle NLM: The orthogonal basis which diagonalizes the oracle (symmetrized) non-local means filter, *i.e.* with patch distances computed on the oracle clean image.

In all cases the oracle NLM basis seems to satisfy the conditions of Theorem 2 and to provide the fastest asymptotic convergence rate. On the other hand, on these experiments, the DCT and wavelet bases some-

times seem to not satisfy the conditions of Theorem 2, and when they do, the asymptotic convergence rate is extremely slow (always smaller than $r = 0.1$) except for sparse images that do not verify Hypothesis 1 and trivially converge in many common bases.

This means that if the oracle NLM basis was known for an arbitrarily large noisy image, then we could use hard thresholding as in Corollary 2 to obtain a denoised image with arbitrarily small MSE. Of course the same conclusion was known (since Donoho-Johnstone) for the non-adaptive wavelet and DCT bases, but convergence does not hold for all natural images, and when it does it may be too slow for the procedure to be practical. For oracle NLM asymptotic convergence seems to be faster with respect to image size but we are confronted to two difficulties:

1. the W-oracle is in principle unknown; and
2. diagonalizing an NLM filter is extremely expensive computationally ($\mathcal{O}(N^3)$) with respect to the number N of pixels).

In order to address the first difficulty we included in our tests the asymptotic performance of the pre-filtered NLM. Directly computing the NLM filter on the noisy image is not acceptable as explained in Section 2.2.3. However, applying it to a pre-filtered version of the image helps both *(a)* to satisfy the requirement of independence of the filter and noise, and *(b)* to make the filter closer to the oracle one. We show in Table 2 the asymptotic convergence rate we estimated for the pre-filtered NLM basis and for the image **lena**. For these experiments we used the denoising algorithm NL-Bayes [14] to obtain the pre-filtered image. Moreover, we tuned the parameters of NL-Bayes in order to slightly over-denoise the image. This trick allows to ensure that the filter is almost independent of the noise realization (at the expense of the potential loss of some subtle image structures). The experiment shows that for the P-NLM basis not only do we achieve asymptotic convergence, but the convergence rate is surprisingly close to the convergence rate for the oracle NLM basis. However, as it can be seen in Figure 11 the actual MSE and the bound are always larger for Prefiltered NLM than for Oracle NLM.

Unfortunately this asymptotic behaviour in the non-oracle case cannot be generalized to all natural images. Indeed under certain texture models a lower bound has been established for all possible image denoising algorithms as recalled in section 4.3.3.

However all these model estimates should be taken with a grain of salt, for several reasons:

Table 2 Fitted parameters α and γ for lena image in the three bases DCT, O-NLM and P-NLM. The parameter r is the decay rate of corollary 1.

image	basis	fitted α	fitted γ	r
lena	DCT	0.827	0.767	0.060
	DWT	0.858	0.752	0.10
	O-NLM	0.806	0.570	**0.236**
	P-NLM	0.7822	0.54846	**0.234**

- The cubic computational cost of exactly computing the eigenbasis of the NLM filters obliged us to limit our evaluation to relatively small image sizes.
- Model (20) can not always be perfectly fit by all images and bases. The model seems to hold for "stationary" images or for images that contain a relatively small number of stationary components. Otherwise the task of fitting this model is particularly difficult.
- Model (20) only gives a coarse upper bound for the actual MSE*. The second column in Figures 12 through 16 shows that even though this upper bound is relatively coarse, the actual MSE* does follow the same kind of decay with N as the upper bound. Nevertheless, when comparing the actual MSE* of all four bases (Figure 11) we observe that the real performance of the prefiltered NLM is actually comparable to that of DCT or wavelet bases; even though the convergence rate r estimated on this model (0.234 for P-NLM vs 0.06 for DCT, 0.10 for DWT and 0.236 for O-NLM) seemed to indicate that the prefiltered NLM was much superior to DCT and rather close to the oracle NLM performance.

Clearly more experiments on larger images are required to confirm or infirm the conclusions of this initial experimental study. Doing so will require the use of more sophisticated and numerically efficient ways to compute the eigenbasis of the NLM filter on medium to large-size images. This could be achieved by means of randomized numerical linear algebra [13], but such techniques do assume a low rank structure of the filtering matrix, so they cannot be used to estimate the full spectrum of eigenvalues of W. Rather they should be used in conjunction with incremental schemes like in [3]. This shall be the subject of further research.

4.3 Discussion about specific cases

In the previous section, we mainly discuss about experiments for images satisfying our main hypotheses. Now let's analyze what happens in two pathological cases:
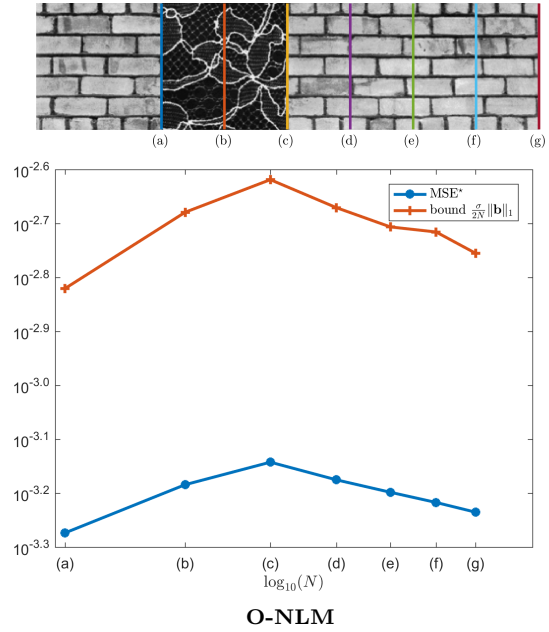

Fig. 9 Top: Image **mixed** with two different textures. Bottom: the corresponding MSE for diagonal estimation using O-NLM basis when the size grows from (a) to (g) in $\log_{10}\log_{10}$ scale.

4.3.1 Sparse Image

When the image is not widespread (like in the image **sparse** in Figure 7), it shows trivial internal redundancy that can be exploited by both the DWT and the NLM bases. Hence the behavior of the MSE is dominated by the black part of the image, and for this reason we obtain a very fast decay of the MSE, hence almost matching the theoretical rate $r = 1$ that is achieved when we use not only optimal eigenvalues λ but also an optimal basis V.

4.3.2 Texture change

In the previous section, we saw experiments for images that do not change drastically with increasing size. But we can wonder what happens when the image suddenly changes with increasing size.

We show in Figure 9 an image composed of two textures and the corresponding curve presenting the behavior of the MSE for the O-NLM case. When we add the second texture, the MSE increases, but when the first texture reappears the MSE starts to decay again. This behavior can explain the fact that there is no need for a stationary hypothesis on the image to obtain convergence. For a sufficiently good basis, able to capture the self-similarity of images, such as NLM-O, we can hope for an asymptotically optimal denoising. This relies on the fact that when the scene size tends toward
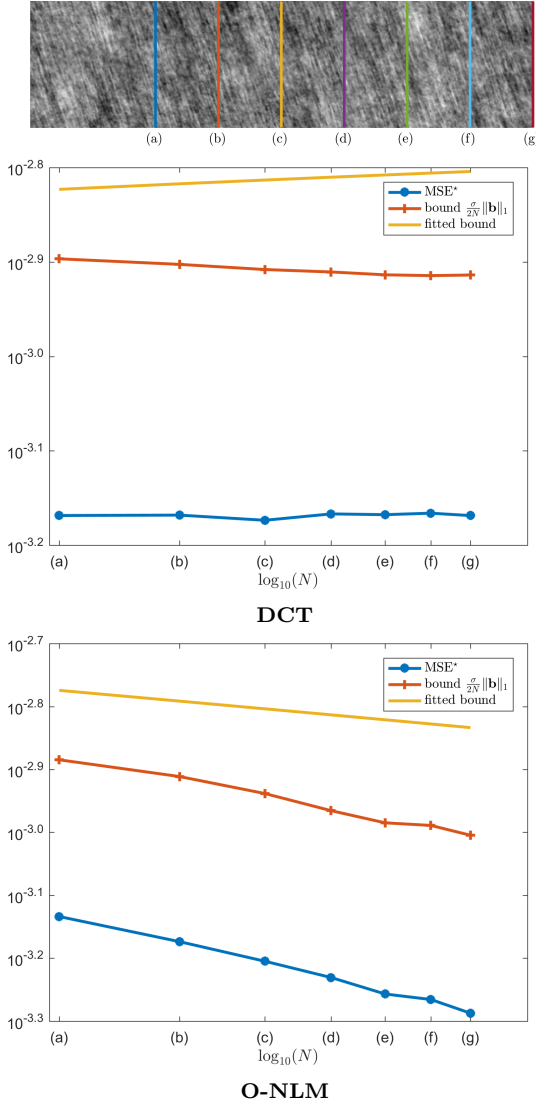
Fig. 10 Top: Image **synthetic**. Bottom: the corresponding MSE for diagonal estimation using DCT basis and O-NLM basis when the size grows from (a) to (g) in $\log_{10}\log_{10}$ scale.

infinity, we can expect similar structures to reappear again and again in the image.

4.3.3 Gaussian & dead-leaves texture models

To emphasize the importance of the basis we provide a numerical experiment with a synthetic Gaussian texture in Figure 10. We proved in Section 3.4.2 that for such a texture the convergence does not hold for a DCT basis even if this texture presents a lot of self-similarity. The numerical experiment confirms that result and provides experimental evidence for the asymptotic decay of the MSE in the Oracle-NLM basis.

Unfortunately, this positive result for the O-NLM case cannot possibly be extended to the non-oracle case. Indeed Levin *et al.* [15] established a strictly positive

lower bound for any image denoising algorithm. This result hols for infinite images that do not present long-distance statistical dependencies. In that case the optimal denoiser for a given pixel x uses the values of the noisy image in pixels y within a neighborhood of x which does not exceed a certain maximal distance D. For y beyond that neighborhood, $\mathbf{u}(x)$ and $\mathbf{u}(y)$ are independent, so the values of $\widetilde{\mathbf{u}}(y)$ provide no useful information to estimate $\mathbf{u}(x)$.

This is the case for Gaussian textures generated by a compactly supported kernel $\mathbf{h}$, and for the dead-leaves model [1]. For images of this kind, Levin's positive lower bound implies that asymptotically zero MSE is impossible to achieve by any non-oracle denoising algorithm. Our experimental result on the Gaussian texture suggests asymptotic convergence of global denoising towards zero MSE, only in the W-oracle case, as seen in section 3.4.3 which explains this mechanism. But this result does not extend to the case where global denoising does not use an oracle to define the filter W.

5 Conclusion

In this paper, we analyzed the following question:

> Can an image denoising algorithm attain asymptotically zero estimation error when the image size tends to infinity?

This question was recently raised in [23,24] in the context of oracle-optimized non-local filtering schemes. That work suggests a positive answer but their reasoning is based on conditions on the infinite image that we show incompatible with reasonable assumptions. We refine these conditions to better account for natural images, and provide a more general theory of optimal asymptotic denoising performance. In particular our theory explores how to partially avoid the use of an oracle, it does not restrict itself to global image denoising, and establishes links to the older diagonal estimation theory, as well as with the optimality results of Donoho and Johnstone [8].

More specifically, our work highlights the central role played by the oracle in the work of [23], and makes a clear distinction between two different ways in which the oracle is used, namely: First a W-*oracle* is used to construct the entries in the non-local filter W whose diagonalization provides a basis V. Then a λ-*oracle* is used to find optimal weights $\{\lambda_j^*\}$ for a given basis V. The link we established with diagonal estimation theory means that the λ-oracle can be avoided using Donoho and Johnstone's theorem, meaning that we can study the convergence of a denoising algorithm that uses a λ-oracle, in order to predict the asymptotic convergence
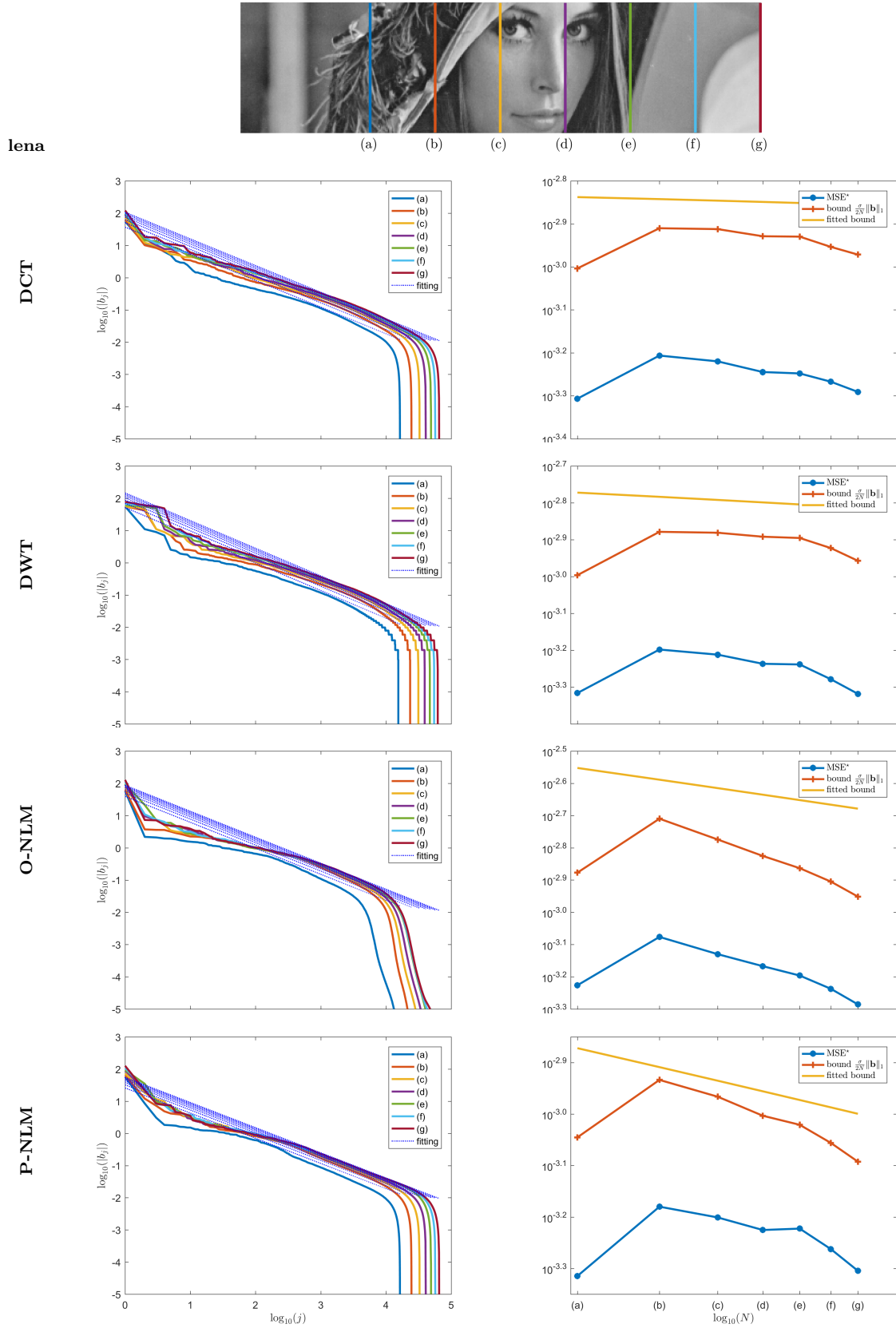
Fig. 11 Left column: the decay of the $|b_j^N|$ for each size and the result of the model fitting (dotted lines) for the image **lena** for the different bases (from top to bottom) DCT, Wavelet, O-NLM and P-NLM. Right column: the decay of the MSE* (blue), the upper bound from (14) (orange) and the fitted bound (yellow).

(at a slower rate) of an algorithm that does not use such a λ-oracle.

The W-oracle, however, is more difficult to avoid, since we do not have a tool equivalent to Donoho and Johnstone's theorem in this setting. Hence non-oracle convergence properties need to be directly tested on a version of the algorithm that does not use the W-oracle. And this is quite problematic because, without an oracle, special care is required to ensure that the filter W and the noise n are independent. And this independence is required for our asymptotic analysis of the MSE to be valid. The quest for more general ways to define non-local and non-oracle filters W, in a way that their independence from image noise is ensured, is still an open subject for future research.

As a whole our generalized analysis of the asymptotic behaviour of global image denoising provides less optimistic conclusions than those in [23] but still leaves the door open for asymptotically zero denoising error. Our experimental study on small images seems to indicate that the oracle non-local means filter can be optimized to attain asymptotically zero error, and that a non-oracle version (i.e. without W-oracle) of that filter may have a similar behaviour, even though at a much slower convergence rate and on a more restricted number of examples. Clearly, more extensive experimentation on a wider variety of larger-sized images is required to determine whether these conclusions may have any practical interest. However, performing such an experimental evaluation requires huge amounts of computation, and can only be addressed if faster and more incremental matrix decomposition algorithms are developed.

These conclusions and the prospect of asymptotically zero MSE may appear to be in contradiction with the strictly positive lower bounds for image denoising established by Levin *et al.* [15]. A careful inspection reveals that there is no such contradiction, rather different models and complementary viewpoints that we shall try to clarify below:

The positive lower bound of Levin *et al.* is valid for certain statistical image models such as Gaussian and dead-leaves textures For images of this kind, Levin's positive lower bound implies that asymptotically zero MSE is impossible to achieve by any non-oracle denoising algorithm. Our experimental result on the Gaussian texture suggests asymptotic convergence of global denoising towards zero MSE, only in the W-oracle case. But this result does not extend to the case where global denoising does not use an oracle to define the filter W.

For more decidedly self-similar images like **lena** or **bricks** our experiments indicate that even the $P - NLM$ filter that does not use a W-oracle is compat-

ible with asymptotically zero MSE. This shows that for this image Levin's assumption of absence of long-distance dependencies does not hold, otherwise there would be a contradiction.

The quest for a statistical model for natural images that takes self-similarity into account in a realistic way is still a very active area of research. Extending such a model for image sizes tending to infinity poses yet an additional challenge. Future research in that direction would hopefully allow to unify Levin's and Talebi's views on asymptotic behaviour of image denoising.

A Global denoising for Gaussian textures on Fourier basis

Proposition 3 *Let* $\mathbf{u} = \mathbf{h} * \mathbf{m}$ *be a Gaussian texture where* $m_i \sim \mathcal{N}(0, \tau^2)$ *iid, and the kernel* $\mathbf{h} \in \ell^1(\mathbb{Z}^2)$ *has a smooth Fourier transform* $\widehat{\mathbf{h}} \in L^1([-\frac{\pi}{2}, \frac{\pi}{2}]^2) \cap C^\infty$. *Then when choosing* V *as a Fourier or DCT basis, and* λ *as the optimal oracle eigenvalues for that basis, then the MSE for global denoising is upper bounded by*

$$\mathrm{MSE}^*_{bound} = \frac{\sigma^2}{2N}\|\mathbf{b}\|_1 = \frac{1}{N^2}\frac{\sigma}{2}\sum_k |\widehat{\mathbf{h}_N}(k)||\widehat{\mathbf{m}_N}(k)| =: A_N$$

Asymptotically we have a strictly positive MSE bound

$$\mathrm{MSE}^*_{bound} = A_N \xrightarrow[N \to \infty]{} A_\infty = \frac{\sigma\tau\|\hat{\mathbf{h}}\|_1}{\sqrt{2\pi}} > 0.$$

Proof Consider for simplicity images of square size $N = n^2$ with $n \in \mathbb{Z}^2$.

Now consider $\mathbf{m}^N$ the restriction of m to $I_N = [-\frac{N}{2}, \frac{N}{2}^2) \cap \mathbb{Z}^2$.

In the truncated case the convolution is understood in the periodic sense, so that we can write

$$\widehat{\mathbf{u}_N}(k) = \widehat{\mathbf{h}_N}(k) \cdot \widehat{\mathbf{m}_N}(k) \quad \forall k \in I_N$$

with $\widehat{\mathbf{h}_N}(k) = \hat{h}(\frac{2\pi k}{n})$.

Now if we decompose $\mathbf{u}_N$ in the Fourier basis $V = F^*$ then

$$\mathbf{b}^N_k = (F\mathbf{u}_N)_k = \widehat{\mathbf{h}_N}(k) \cdot \widehat{\mathbf{m}_N}(k)$$

So the upper bound given in equation (14) for the optimal MSE is

$$\mathrm{MSE}^*_N \leq \frac{1}{N^2}\frac{\sigma\|\mathbf{b}_N\|_1}{2} = \frac{1}{N^2}\frac{\sigma}{2}\sum_{k \in I_N}|\widehat{\mathbf{h}_N}(k)||\widehat{\mathbf{m}_N}(k)| =: A_N$$

In the asymptotic case when $N \to \infty$, a simple Riemann sum argument based on the regularity of $\hat{h}$ and on the known value of $\int_{dI}|\hat{m}| = |dI|\mathbb{E}(\widehat{m_N}(k)) = \tau\sqrt{\frac{2}{\pi}}|dI|$ over a small interval (mean absolute deviation of a gaussian), leads to the conclusion that the upper bound tends to a strictly positive constant

$$\mathrm{MSE}^*_\infty \leq \lim_{N \to \infty} A_N = \frac{\sigma\tau\|\hat{h}\|_1}{\sqrt{2\pi}} = A_\infty > 0.$$

$\square$

B An Oracle filter that provides asymptotically zero MSE

Proposition 4 *Consider an oracle filter which consists in denoising $\tilde{\mathbf{u}}$ by averaging at pixel i all values $\tilde{\mathbf{u}}_j$ such that $|u_i - u_j| \leq \varepsilon$ for a given threshold $\varepsilon > 0$ and an infinite image $\mathbf{U}$ bounded with values in $[0, 1[$. Then the value $\mathrm{MSE}(\widehat{\mathbf{u}^N}|\mathbf{u}^N)$ converges to a limit smaller than ε^2.*

Proof Indeed, let $n = \lceil \frac{1}{\varepsilon} \rceil$, for $0 \leq k < n$ let $C_k = \{i \in \Omega;\ u_i^N \in [\frac{k}{n}, \frac{k+1}{n}[\}$, and for every pixel i let $A_i^\varepsilon = \{j;\ |u_i^N - u_j^N| \leq \varepsilon\}$. It is clear that if $i \in C_k$, then $C_k \subset A_i^\varepsilon$. We can write

$$\mathrm{MSE}(\widehat{\mathbf{u}^N}|\mathbf{u}^N) = \frac{1}{N}\sum_{k=0}^{n-1}\sum_{i \in C_k} \mathbb{E}[|u_i^N - \widehat{u_i^N}|^2]$$

$$= \frac{1}{N}\sum_{k=0}^{n-1}\sum_{i \in C_k} \mathbb{E}[|u_i^N - \frac{1}{\#A_i^\varepsilon}\sum_{j \in A_i^\varepsilon}\tilde{u}_j^N|^2]$$

$$= \frac{1}{N}\left[\sum_{k=0}^{n-1}\sum_{i \in C_k} |u_i^N - \frac{1}{\#A_i^\varepsilon}\sum_{j \in A_i^\varepsilon}u_j^N|^2 \right.$$

$$\left. + \sum_{k=0}^{n-1}\sum_{i \in C_k} \frac{\sigma^2}{\#A_i^\varepsilon} \right]$$

$$\leq \varepsilon^2 + \frac{\sigma^2}{N}\sum_{k=0}^{n-1}\sum_{i \in C_k} \frac{1}{\#A_i^\varepsilon}$$

$$\leq \varepsilon^2 + \frac{\sigma^2}{N}\sum_{k=0}^{n-1}\sum_{i \in C_k} \frac{1}{\#C_k} \leq \varepsilon^2 + \frac{n\sigma^2}{N}.$$

For a given value of ε, this term converges to ε^2 when $N \to \infty$. $\square$

C Additionnal experiments

Figures 12 through 16 show the detailed asymptotic convergence results for the images in Figure 7 and Table 1.

References

1. Alvarez, L., Gousseau, Y., Morel, J.M.: The Size of Objects in Natural and Artificial Images, vol. 111, pp. 167–242. Elsevier (1999). DOI 10.1016/S1076-5670(08)70218-0. URL http://dx.doi.org/10.1016/S1076-5670(08)70218-0

2. Awate, S., Whitaker, R.: Image denoising with unsupervised information-theoretic adaptive filtering. In: International Conference on Computer Vision and Pattern Recognition (CVPR 2005), pp. 44–51 (2004)

3. Brand, M.: Fast low-rank modifications of the thin singular value decomposition. Linear Algebra Appl. **415**(1), 20–30 (2006). DOI 10.1016/j.laa.2005.07.021. URL http://linkinghub.elsevier.com/retrieve/pii/S0024379505003812

4. Brodatz, P.: Textures: a photographic album for artists and designers. Dover Pubns (1966)

5. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 2, pp. 60–65. IEEE (2005)

6. Chan, S.H., Zickler, T., Lu, Y.M.: Demystifying symmetric smoothing filters. arXiv preprint arXiv:1601.00088 (2016)

7. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-d transform-domain collaborative filtering. Image Processing, IEEE Transactions on **16**(8), 2080–2095 (2007)

8. Donoho, D.L., Johnstone, I.M., et al.: Ideal denoising in an orthonormal basis chosen from a library of bases. Comptes Rendus de l'Academie des Sciences-Serie I-Mathematique **319**(12), 1317–1322 (1994)

9. Donoho, D.L., Johnstone, J.M.: Ideal spatial adaptation by wavelet shrinkage. Biometrika **81**(3), 425–455 (1994)

10. Duval, V., Aujol, J.F., Gousseau, Y.: A bias-variance approach for the nonlocal means. SIAM Journal on Imaging Sciences **4**(2), 760–788 (2011)

11. Facciolo, G., Almansa, A., Aujol, J.F., Caselles, V.: Irregular to Regular Sampling, Denoising, and Deconvolution. SIAM MMS **7**(4), 1574–1608 (2009). DOI 10.1137/080719443. URL http://epubs.siam.org/doi/abs/10.1137/080719443

12. Guichard, F., Moisan, L., Morel, J.M.: A review of pde models in image processing and image analysis. In: Journal de Physique IV (Proceedings), vol. 12, pp. 137–154. EDP sciences (2002)

13. Halko, N., Martinsson, P.G., Tropp, J.A.: Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions. SIAM Rev. **53**(2), 217–288 (2011). DOI 10.1137/090771806. URL http://epubs.siam.org/doi/abs/10.1137/090771806

14. Lebrun, M., Buades, A., Morel, J.M.: A Nonlocal Bayesian Image Denoising Algorithm. SIAM J. Imaging Sci. **6**(3), 1665–1688 (2013). DOI 10.1137/120874989. URL http://epubs.siam.org/doi/abs/10.1137/120874989

15. Levin, A., Nadler, B., Durand, F., Freeman, W.T.: Patch complexity, finite pixel correlations and optimal denoising. (ECCV 2012) LNCS **7576 LNCS**(PART 5), 73–86 (2012). DOI 10.1007/978-3-642-33715-4_6

16. Mallat, S.: A wavelet tour of signal processing: the sparse way. Academic press (2008)

17. Milanfar, P.: A tour of modern image filtering: New insights and methods, both practical and theoretical. Signal Processing Magazine, IEEE **30**(1), 106–128 (2013)

18. Milanfar, P.: Symmetrizing smoothing filters. SIAM Journal on Imaging Sciences **6**(1), 263–284 (2013)

19. Ordentlich, E., Seroussi, G., Verdu, S., Weinberger, M., Weissman, T.: A discrete universal denoiser and its application to binary images. In: Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on, vol. 1, pp. I–117. IEEE (2003)

20. Pati, Y.C., Rezaiifar, R., Krishnaprasad, P.: Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In: Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on, pp. 40–44. IEEE (1993)

21. Pierazzo, N., Rais, M., Morel, J.M., Facciolo, G.: Da3d: Fast and data adaptive dual domain denoising. In: Image Processing (ICIP), 2015 IEEE International Conference on, pp. 432–436. IEEE (2015)

22. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena **60**(1), 259–268 (1992)

23. Talebi, H., Milanfar, P.: Global denoising is asymptotically optimal. ICIP **5**(3) (2014)

24. Talebi, H., Milanfar, P.: Global image denoising. Image Processing, IEEE Transactions on **23**(2), 755–768 (2014)
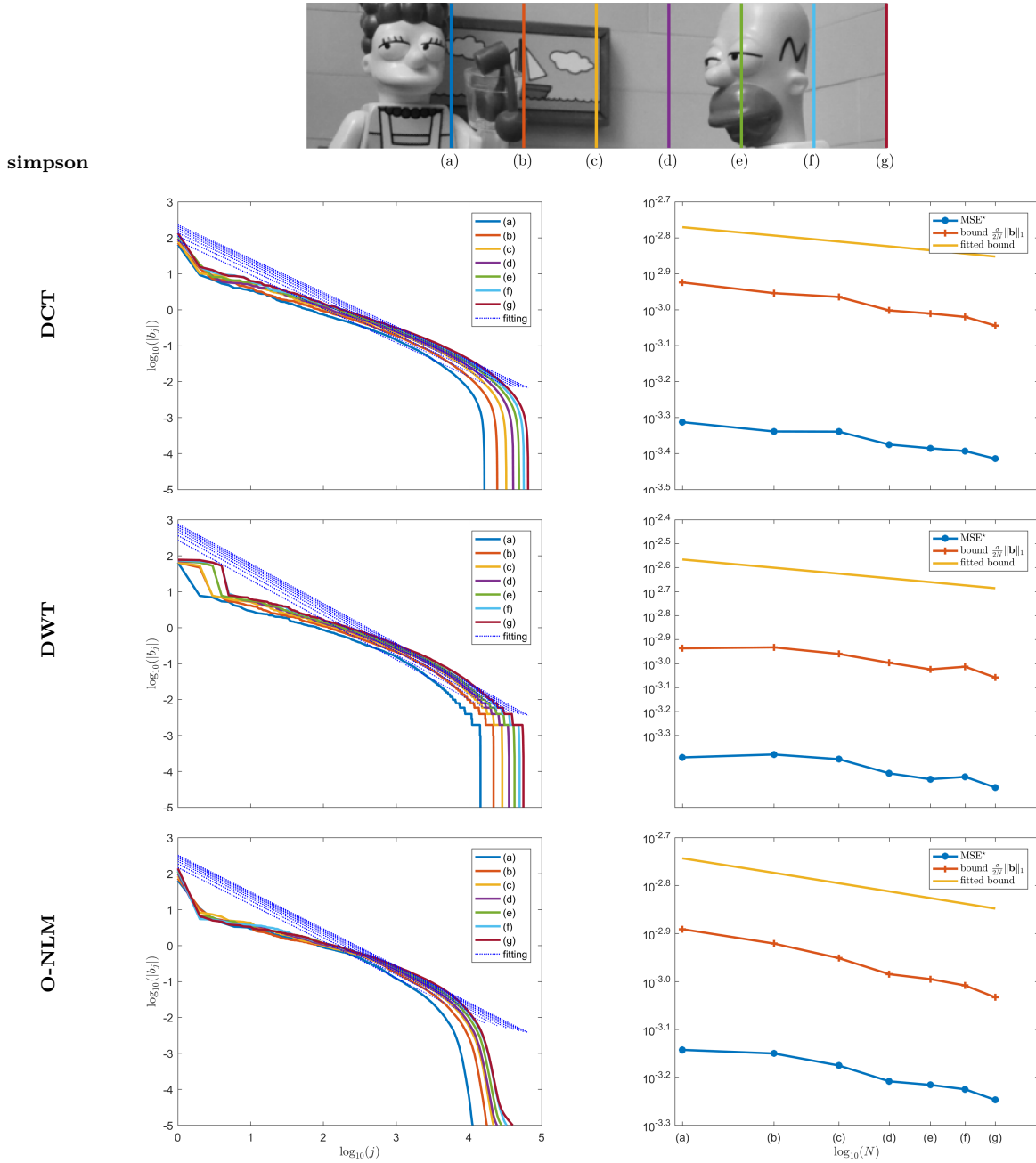
Fig. 12 Left column: the decay of the $|b_j^N|$ for each size and the result of the model fitting (dotted lines) for the image **simpson** for the different bases (from top to bottom) DCT, Wavelet, and O-NLM. Right column: the decay of the MSE* (blue), the upper bound from (14) (orange) and the fitted bound (yellow).

25. Talebi, H., Milanfar, P.: Asymptotic performance of global denoising. SIAM Journal on Imaging Sciences **9**(2), 665–683 (2016)
26. Weissman, T., Ordentlich, E., Seroussi, G., Verdú, S., Weinberger, M.J.: Universal discrete denoising: Known channel. IEEE Transactions on Information Theory **51**(1), 5–28 (2005)
27. Yaroslavsky, L.: Digital picture processing: an introduction, vol. 9. Springer Science & Business Media (2012)
28. Yu, G., Sapiro, G., Mallat, S.: Solving inverse problems with piecewise linear estimators: from Gaussian mixture models to structured sparsity. IEEE Trans. Image Process. **21**(5), 2481–99 (2012). DOI 10.1109/TIP.2011.2176743. URL 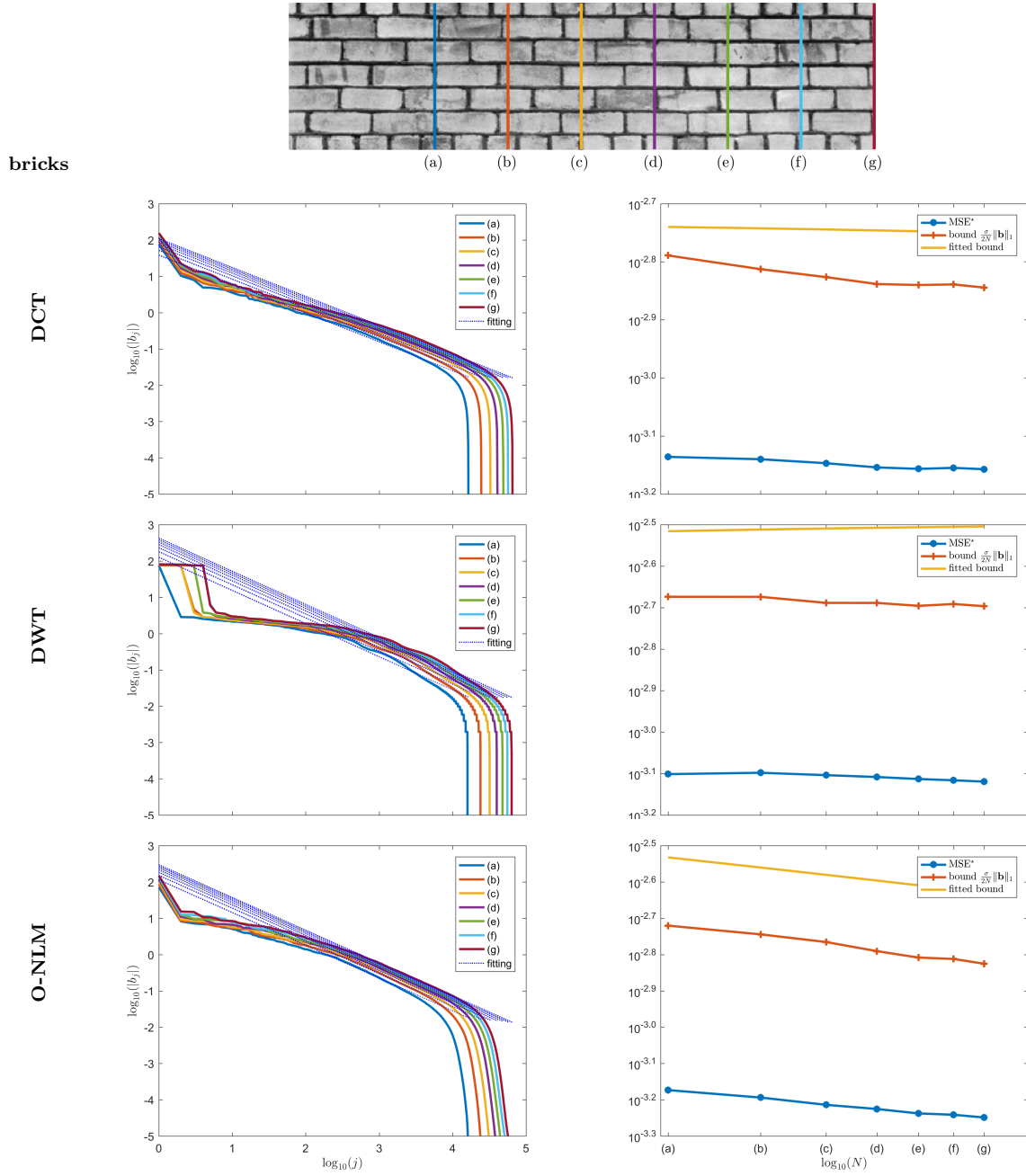http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6104390http://www.di.ens.fr/~mallat/papiers/SSMS-journal-subm.pdfhttp://www.ncbi.nlm.nih.gov/pubmed/22180506
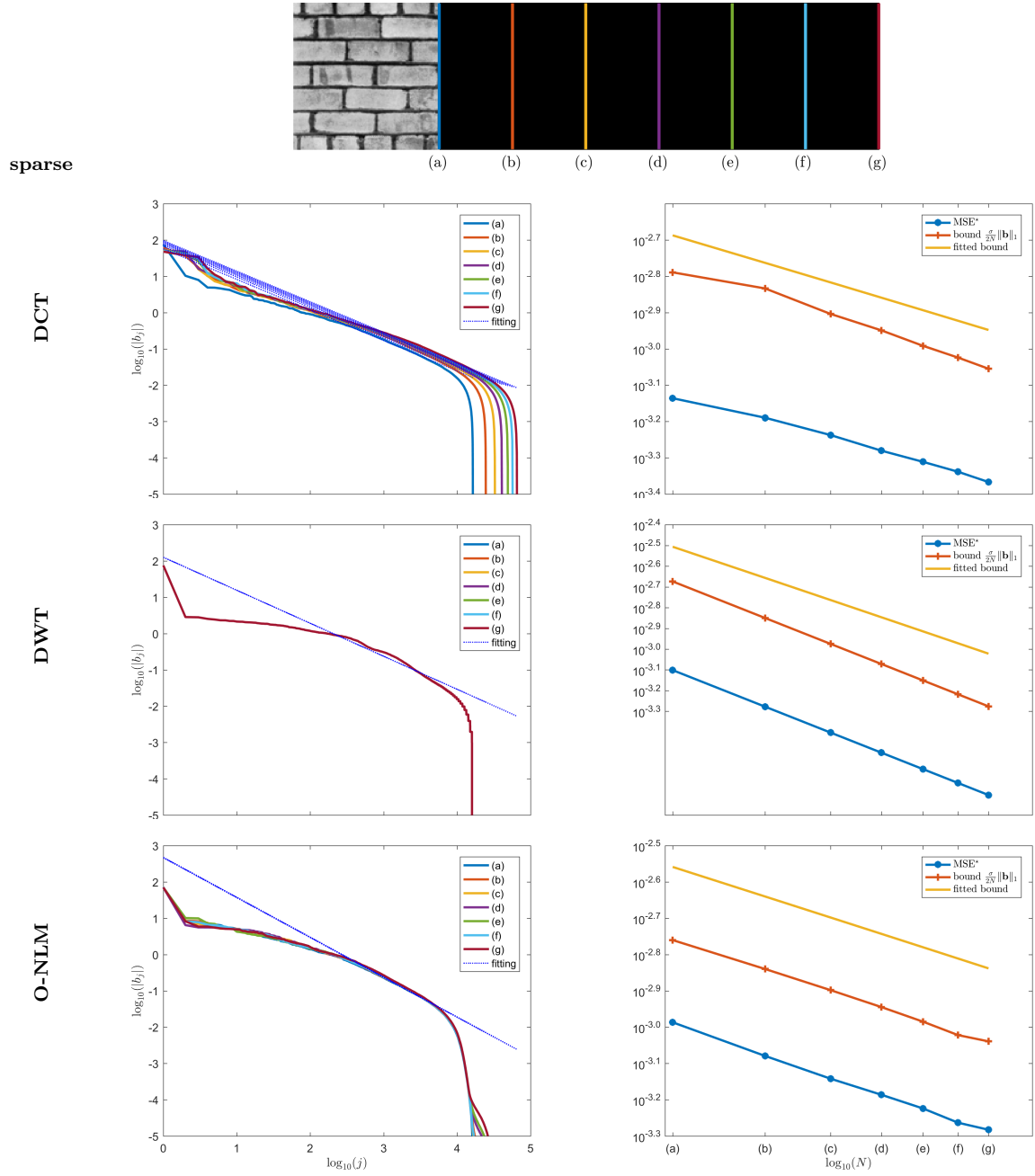
Fig. 13 Left column: the decay of the $|b_j^N|$ for each size and the result of the model fitting (dotted lines) for the image **bricks** for the different bases (from top to bottom) DCT, Wavelet, and O-NLM. Right column: the decay of the MSE* (blue), the upper bound from (14) (orange) and the fitted bound (yellow).
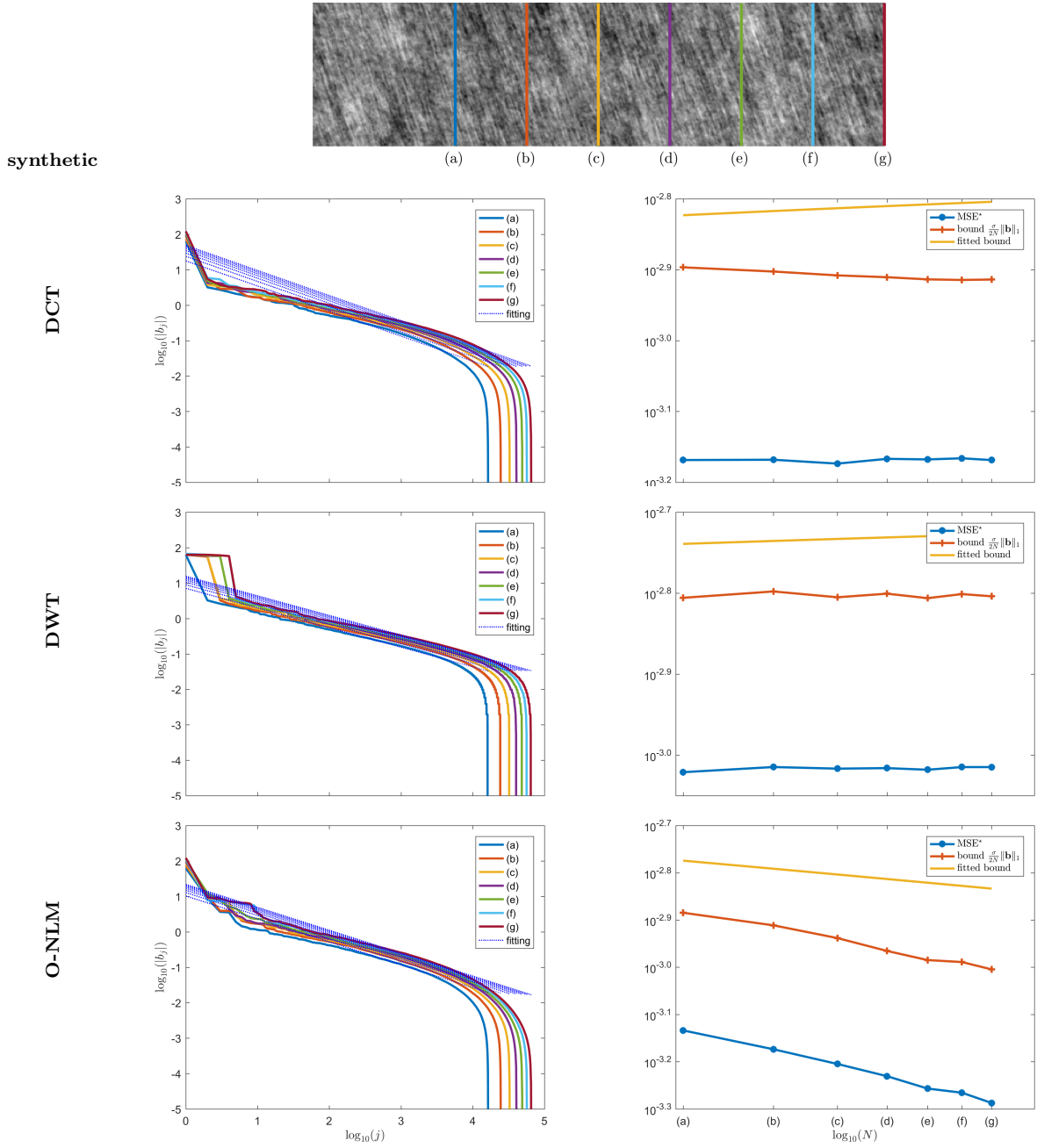
Fig. 14 Left column: the decay of the $|b_j^N|$ for each size and the result of the model fitting (dotted lines) for the image **sparse** for the different bases (from top to bottom) DCT, Wavelet, and O-NLM. Right column: the decay of the MSE* (blue), the upper bound from (14) (orange) and the fitted bound (yellow).

Fig. 15 Left column: the decay of the $|b_j^N|$ for each size and the result of the model fitting (dotted lines) for the image **synthetic** for the different bases (from top to bottom) DCT, Wavelet, and O-NLM. Right column: the decay of the MSE* (blue), the upper bound from (14) (orange) and the fitted bound (yellow).
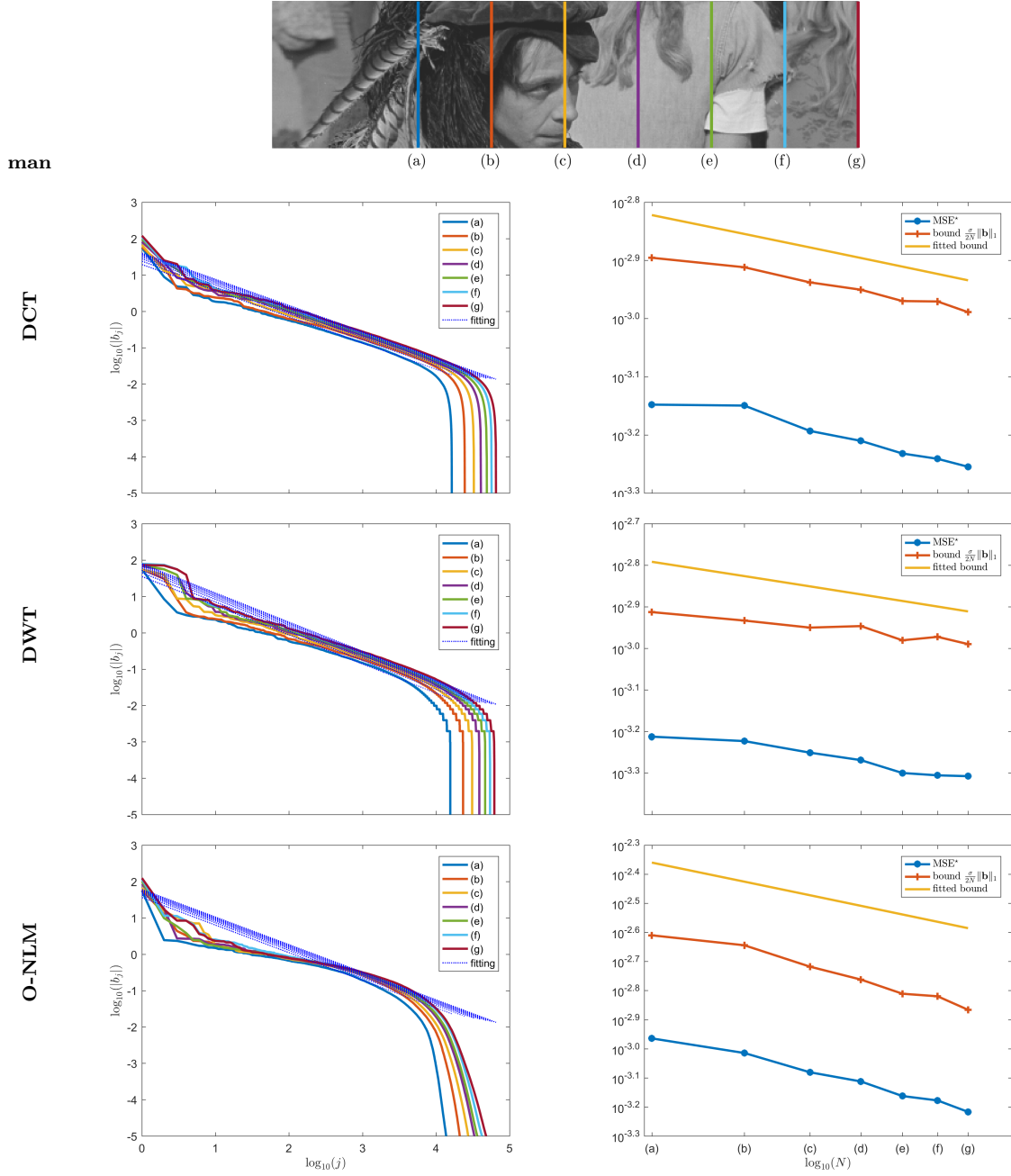
Fig. 16 Left column: the decay of the $|b_j^N|$ for each size and the result of the model fitting (dotted lines) for the image **man** for the different bases (from top to bottom) DCT, Wavelet, and O-NLM. Right column: the decay of the MSE* (blue), the upper bound from (14) (orange) and the fitted bound (yellow).