HAL
open science

Demystifying the asymptotic behavior of global denoising

Antoine Houdard, Andrès Almansa, Julie Delon

To cite this version:

Antoine Houdard, Andrès Almansa, Julie Delon. Demystifying the asymptotic behavior of global denoising. 2016. hal-01340822v2

HAL Id: hal-01340822

https://hal.science/hal-01340822v2

Preprint submitted on 13 Jul 2016 (v2), last revised 21 Mar 2017 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Demystifying the asymptotic behavior of global denoising

Antoine Houdard, Andrés Almansa and Julie Delon

***Abstract*—In this work, we revisit the global denoising framework recently introduced by Talebi & Milanfar, with the classic formalism of diagonal estimation. We analyze the asymptotic behavior of its mean-squared error restoration performance 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 discuss open issues concerning the most challenging aspect, namely the extension of these results to the non-oracle case.**

***Index Terms*—Diagonal Estimation, Global Denoising, Wiener Filtering, Asymptotic study**

I. Introduction

The story of image denoising, and more generally of image restoration, is probably as old as the story of image processing. If classic image denoising methods have always gathered various mathematical tools, such as neighborhood filters [1], variationnal models [2], non linear Partial Differential Equations [3] or transform domain estimation [4], the last true revolution came with the introduction of patch-based methods in 2004. 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.

The idea behind patch-based approaches, also called non-local approaches, was mostly to exploit the self-similarity in images in order to significantly improve the performance of image denoising algorithms. The first denoising algorithms relying on this idea appear in 2004 with the successive introduction of the Discrete Universal Denoiser for binary images [5], [6], of the UINTA filters [7] and of the Non Local Means [8]. Authors of these approaches propose to reduce the noise variance, assumed additive and Gaussian, by averaging repeated structures in images, the average being weighted by the similarity between these structures. In practice, the success of these approaches was partly due to the resistance of patches to noise, and to the fact that they created less artifacts in images than the best concurrent methods at the time. These non local methods have inspired a considerable body of works ever since, under the form of variants and improvments [9], [10], extensions to other noise models [11], [12], or to more complex inverse problems [13], [14], [15]. Several principles underly the recent devlopments of these non local methods: the Bayesian paradigm, which necessitates good prior distributions on patches [16], [17], [18], [19], the use of sparse representations on well adapted dictionnaries [20], [21], and very lately the reintroduction in this framework of transform-domain approaches [22], [23], [24].

Leading methods nowadays are all patch-based. 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 optimality when the image size tends toward infinity. Consistency results, under stationnarity hypotheses, have been shown for instance for the DUDE algorithm [5], [6] and for the Non Local Means [8]. Now, despite their non-local nature, most of these algorithms limit the search area for similar patches to a medium-sized nighborhood around each pixel. Doing otherwise would confront them to a dilemma [25]. 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 by [26], 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 [27], [28] proposed a truly global denoising approach where each pixel is used to denoise every other pixel. They claimed in a subsequent paper [29] that this approach is asymptotically optimal, in the sense that *"the mean-squared error monotonically decays with increasing image size"*, regardless of image content. In this context, this paper raises again the question: can non-local 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) ? The debate is open by their algorithm that shows 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 natural 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

The authors would like to thank H. Talebi and P. Milanfar for encouraging and insightful discussions. This work has been partially funded by the French Research Agency (ANR) under grant nro ANR-14-CE27-001 (MIRIAM).

revisit the theory of diagonal estimation (refered to as Wiener filtering in Talebi's paper) that was first developed for wavelet bases, and reused in the context of non-local filters by Talebi and Milanfar.

The paper is organized as follows. In Section II we provide a short reminder on the theory of diagonal estimation in an orthonormal basis. Section III uses this framework to present the global denoising formalism and to put it into perspective relatively to classic diagonal estimation results. The main contribution of this paper is a novel asymptotic study of global denoising, presented in Section IV. Basically, we introduce precise conditions both on the image and the global filter to ensure that the mean-squared error (MSE) for global image denoising decays toward 0 for increasing image size. We also discuss and show experiments on several open issues, including the extension of these results to the non-oracle case.

II. Diagonal estimation : a short reminder

A. Principle

a) Notations.: In this paper we will always consider images as vectors of $\mathbf{R}^N$, where N is the number of pixels. We consider the following image formation model

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

where $\mathbf{u}$ is a deterministic unknown image and $\widetilde{\mathbf{u}}$ is the observed image which differs from $\mathbf{u}$ by a Gaussian white noise ϵ. Put another way, ϵ is the realization of a random vector $\mathcal{E} \sim \mathcal{N}(0, \sigma^2 I_N)$. Observe that 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.

b) Diagonal estimation.: Consider the inverse problem of recovering $\mathbf{u}$ from the observation $\widetilde{\mathbf{u}}$. In signal and image processing, to solve this estimation problem, it is very common to resort to so-called diagonal estimators. 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 $V = \{V_i\}_{i=1,\dots,N}$, which means that it can be written

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

where Λ is a diagonal matrix whose k^{th} coefficient λ_k 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 [30], [31], [4], or can for instance be built up as an orthonormal dictionnary from the image itself [32]. The success of these methods stems partly from the fact that if the image u is sparse in the orthonormal basis V, these "diagonal estimators are nearly optimal among all non linear estimators", as stated in [4].

B. Quadratic risk

The mean quadratic risk or mean squared error (MSE) of a diagonal estimator 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_i) 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,*

$$\mathrm{MSE}(\widehat{\mathbf{u}}|\mathbf{u}) = \frac{1}{N}\sum_{j=1}^{N}\left((1-\lambda_j)^2\mathbf{b}_j^2 + \sigma^2\lambda_j^2\right). \tag{3}$$

Proposition 1.

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

□

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

C. Oracle quadratic risk minimization

a) 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 is global minimum for

$$\lambda_i^\star = \frac{b_i^2}{\sigma^2 + b_i^2}. \tag{4}$$

The corresponding minimal value of the MSE is

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

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 $\mathbf{u}$. This image being unknown, 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, even if it can be shown that some well chosen thresholding estimators have a risk which is not too far from the oracle one [4].

b) 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 u. The quantity $\mathrm{MSE}^\star$ depends only on $\mathbf{b} = V^T\mathbf{u}$, the projection of the oracle image $\mathbf{u}$ on the basis V. The matrix V^T being

orthonormal, the quantity $\mathrm{MSE}^\star$ has to be minized under the constraint $\|\mathbf{b}\|_2 = \|\mathbf{u}\|_2$.

Proposition 2. *Minimizing* $\mathbf{b} \mapsto \mathrm{MSE}^\star(\mathbf{b})$ *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 or* $\mathbf{R}^N$ *basis.*

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

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

Again, even if this filter is not reachable since it depends on the unknown oracle u, this results strongly support the intuitive idea that ideal bases should provide a sparse representation of u. In practice, diagonal estimation should be applied in a well-adapted basis for each image, typically a basis V that provides a fast decreasing of the $\{b_j\}$. The principle of global filtering [27], described in Section III, is to rely on classic non linear filters from the denoising literature to choose V.

D. 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 classic solution is to consider thresholding estimators in a given orthonormal basis V, in order to discard unrelevant components. For instance, hard thresholding is implemented as

$$\forall j \in \{1, \ldots, N\}, \quad \lambda_j^T \stackrel{def}{=} \begin{cases} 1 & \text{if} \quad |\widetilde{b}_j| \geqslant T \\ 0 & \text{if} \quad |\widetilde{b}_j| < T, \end{cases} \tag{5}$$

and the corresponding estimator can be written

$$\widehat{\mathbf{u}} = \sum_{1 \leq k \leq N;\ \langle \widetilde{\mathbf{u}}, V_k \rangle \geq T} \langle \widetilde{\mathbf{u}}, V_k \rangle \cdot V_k. \tag{6}$$

As illustrated by Figure 1, the result of hard thresholding even for a good choice of T is far from being as satisfying as the oracle estimation (4), at least on a Discrete Cosine Basis. However, for a well chosen value of T, the mean squared error obtained with hard thresholding can still be controlled by the one of the oracle attenuation (4).

Theorem 1 (Donoho-Johnstone [30])**.** *Let* $T = \sigma\sqrt{2 \ln N}$. *The MSE provided by the thresholded eigenvalues* λ^T *satisfies for* $N \geq 4$

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

This theorem 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. For the same quantity of noise, 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 thresholding estimation. Observe that the threshold $T = \sigma\sqrt{2 \ln N}$ is not really optimal in practice. A good way to fix T is to resort to the SURE estimator of the MSE [4].

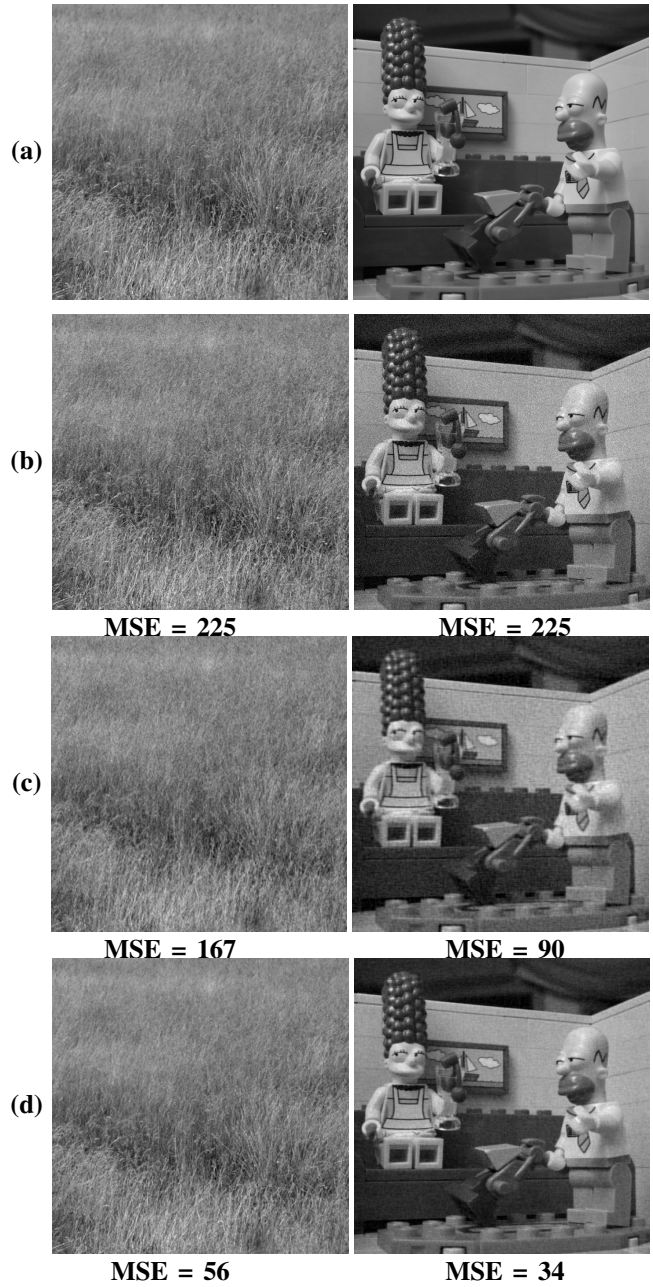


Figure 1. **(a)** Original images $\mathbf{u}$, **(b)** Noisy images $\widetilde{\mathbf{u}}$ with $\sigma = 15$, **(c)** Images $\widetilde{\mathbf{u}}$ denoised by hard-thresholding in a DCT basis, **(d)** Images $\widetilde{\mathbf{u}}$ denoised by diagonal estimation with the oracle λ_i^* in a DCT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

III. Global filtering

In 2014, Talebi and Milanfar [27], [28] introduce a formalism called *global denoising*, which 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. The idea of global denoising is to build V as an orthonormal basis that diagonalizes a classic non linear denoising filter (such as NLmeans [8]) computed on $\widetilde{\mathbf{u}}$. If the chosen denoising filter is efficient, it is hoped that the coefficients b_j will decrease relatively quickly and that the diagonal estimate will be all the more efficient.

A. 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 [8] type filters, can be written under the form

$$\widehat{\mathbf{u}} = W\widetilde{\mathbf{u}}, \tag{7}$$

where $W = D^{-1}K$, with K a positive definite kernel and D a diagonal matrix with entries $D_{ii} = \sum_j K_{ij}$, $i \in \{1, \dots, N\}$. Starting from a given denoising filter W, the idea of global denoising, made popular by Milanfar in [33], [34], 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 (4) of the eigenvalues can be applied to improve the filter.

B. Symmetrizing the filter W

To ensure the fact that W can be diagonalized, the authors of [27] propose to replace this filter by a doubly stochastic version W^s of W which minimizes the cross-entropy

$$\sum_{i,j} W^s_{ij} \log \frac{W^s_{ij}}{W_{ij}}. \tag{8}$$

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 [33]. In practice, the denoising results obtained with this symmetrized filter appear to be equivalent or slightly better than the ones obtained with W [33], [35]. In the following, we always consider the filter W in its symmetric and doubly stochastic version.

C. Deterministic filter

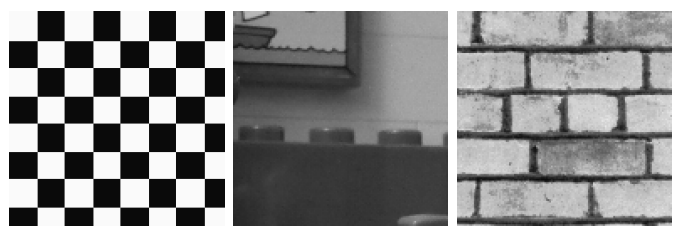
The mean-squared error formulation (3) is valid only if the filter $W = V\Lambda V^T$ is deterministic. This assumption is sensible when V is fixed, as a DCT or wavelet basis for instance. However when the filter W comes from the noisy image $\widetilde{\mathbf{u}}$, this hypothesis does not hold. To test this claim, we compare, for different choices of the filter W, the theoretical $\mathrm{MSE}_{\text{theo}}$ computed by formula (3) with the experimental mean-squared error

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

Figure 2 shows the relative error $\frac{|\,\mathrm{MSE}_{\text{eval}} - \mathrm{MSE}_{\text{theo}}\,|}{\mathrm{MSE}_{\text{theo}}}$ for the following filters W, computed on three different images:

1) a Non-Local Means filter [8] computed on the original image $\mathbf{u}$ (called NLM oracle);
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);

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 the mere fact to prefilter the image on which the NL-means kernel is computed permits to decouple the filter W from the noise, at least enough for the theoretical $\mathrm{MSE}_{\text{theo}}$ to be a good predictor of $\mathrm{MSE}_{\text{eval}}$.



(a) (b) (c)

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

Figure 2. Relative error $\frac{|\,\mathrm{MSE}_{\text{eval}} - \mathrm{MSE}_{\text{theo}}\,|}{\mathrm{MSE}_{\text{theo}}}$ between the theoretical MSE provided by formula (3) 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.

D. 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 oracle $\mathbf{u}$ is needed in practice to compute the vector $\mathbf{b}$ and 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 II-D that hard or soft thresholding could provide MSE results 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 [28]. 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.

IV. Asymptotic study

In this part, we study the asymptotic behavior of the MSE given by formula (3) when the image size increases. In [29], the authors claim that global denoising is asymptotically optimal. In order to explore the precise conditions of this convergence, we define in Section IV-B 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 IV-C different conditions of convergence for the MSE and its corresponding decay rate. Finally, in Section IV-D2, 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.

A. Upper bound on the optimal MSE

We have seen in Section II 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:

$$\begin{aligned}\mathrm{MSE}^\star &= \frac{\sigma^2}{N}\sum_{j=1}^{N}\frac{b_j^2}{\sigma^2+b_j^2}\\ &\leqslant \frac{\sigma^2}{N}\sum_{j=1}^{N}\frac{b_j^2}{2\sigma|b_j|}\\ &= \frac{\sigma}{2N}\|\mathbf{b}\|_1. \end{aligned} \tag{9}$$

The authors of [29] suggest that this upper bound might converge towards 0 when N grows to infinity. In order to prove this convergence, they assume that the coefficients b_j drop off at a given rate $\alpha > 0$

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

For a fixed image size, this hypothesis seems quite reasonable for different existing filters, as illustrated by Figure 3. When working with Fourier or space-frequency decompositions, the value of α was shown to be related to the regularity of the image [4], and values of α between 0.5 and 1 were shown to be in agreement to actual image data [36]. Such models

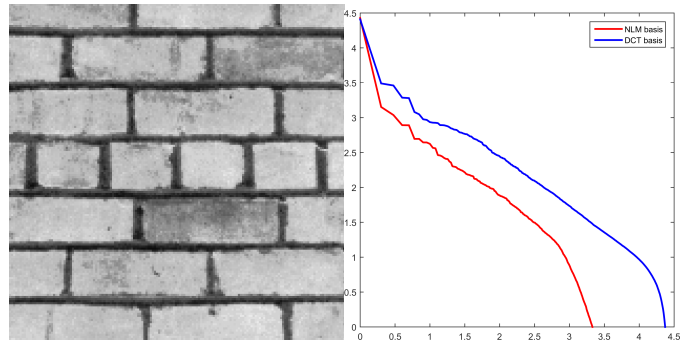


Figure 3. Decay rate of the coefficients b_j for the image on the left, in a loglog scale graph. In blue: coefficients in a DCT basis. In red: coefficients in the eigenbasis of the oracle NLM filter.

have also been correctly used in asymptotic studies where the image *resolution* tends to infinity, but here we are interested in the asymptotic behaviour 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 illustrate this claim, we show in Figure 4 the behavior of the largest [1] b_j^N (in magnitude) computed with a DCT basis for images of increasing size N. This largest coefficient clearly increases with N. In the following Section, we propose a model for an image whose size grows to infinity and a more complete parametric model for the decay of the coefficients b_j^N.

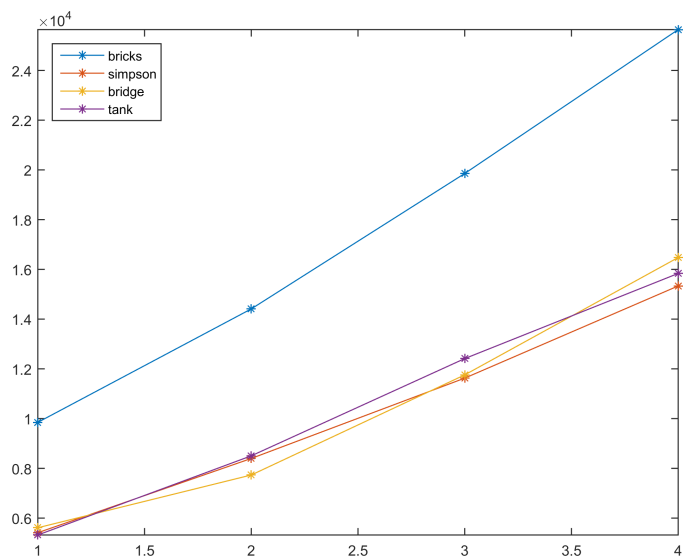


Figure 4. Behaviour of $\max_j(|b_j^N|)$ with increasing size N for three different images in loglog scale. Here $\mathbf{b}^N$ is the image $\mathbf{u}$ projected in the DCT basis.

B. Proposed models

a) Infinite image model.: Consider an image of infinite size

$$\mathbf{U} : \mathbb{Z}^2 \to \{m, \dots, M\}$$

[1] 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.

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

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

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

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

$$m\sqrt{N} \leqslant \|\mathbf{u}^N\|_2 \leqslant M\sqrt{N},$$

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 m is zero, the energy of the image may not grow as fast as $\sqrt{N}$. However we show that this case only occurs with images becoming sparse with increasing size. Indeed, let us assume that $\|u^N\|_2 = o(\sqrt{N})$. 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

$$c\frac{\#\{u_i^N \neq 0\}}{N} \leqslant \frac{\|u^N\|_2^2}{N} \underset{N\to\infty}{\longrightarrow} 0.$$

This shows that the ratio of non-zero pixels collapses when the image size goes to infinity.

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

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

$$m\sqrt{N} \leqslant \|\mathbf{u}^N\|_2 \leqslant M\sqrt{N}. \tag{10}$$

b) 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 $|b_j^N|$

In order to model it we use experimental results we obtained using the DCT basis for V. That is to say $\mathbf{b}^N$ being the discrete cosine transform of $\mathbf{u}$. From Figure 3 we notice that for a fixed N, the coefficients b_j are decaying at some rate $\frac{1}{j^\alpha}$ with $\alpha > 0$. This is coherent with the model proposed in [28] on the decay rate of the b_j, namely

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

with $\alpha > 0$. To reintroduce the parameter N into this model, we should consider that the constants C and α are depending on N, so we start with a model of the form

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

and we discuss how to simplify it based on information given by numerical experiments.

We start with the worst scenario, when the image $\mathbf{u}$ is composed of white noise. In this case, because the DCT is an isometry, $\mathbf{b}$ is also white noise, and the decay rate of the $|b_j|$ is very close to zero (see Figure 6). With increasing N, this decay rate should decrease towards zero. For a given natural image, presenting auto-similarity properties, we expect this decay rate to be faster than the white noise case. More precisely, if $\alpha = \min_N(\alpha_N)$, we can assume that $\alpha \neq 0$ and conservatively choose this slowest convergence rate to avoid the dependence of α on the image size N.

Still, we need to define how the "constant" C_N grows with the image size N. Figure 4 shows that the maximum of $\left(|b_j^N|\right)_j$ 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 $N, j \in \mathbb{N}$*

$$|b_j^N| \leqslant C\frac{N^\gamma}{j^\alpha}. \tag{11}$$

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

In the next section, we study the convergence of the upper-bound of the MSE under this hypothesis.

C. Conditions of convergence

In Section II 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,

$$\mathrm{MSE}(\lambda^\star) \leqslant \frac{\sigma^2}{N}\|\mathbf{b}\|_1 \leqslant C'\frac{1}{N^r}, \tag{9}$$

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

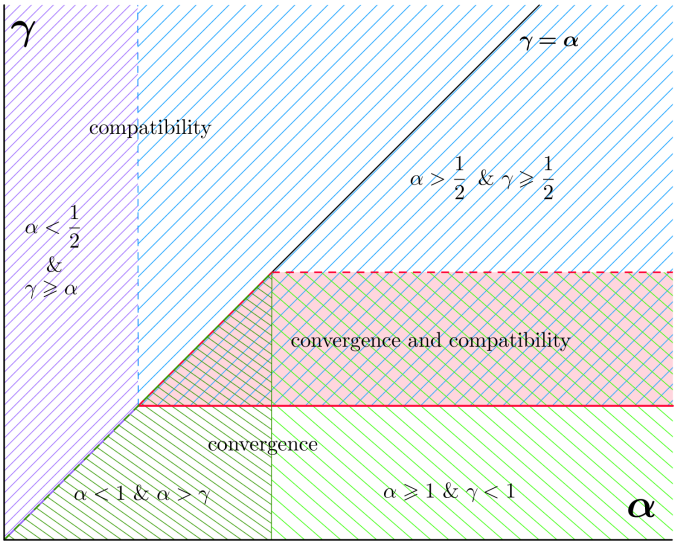


Figure 5. Illustration of the domain of compatibility and the domain of convergence provided by the two lemmas 1 and 2. The intersection in red represent the set of parameters that provides the result in Theorem 2.

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

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

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 (10) 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. Figure 5 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* $\mathbf{U}$ *satifies Hypothesis 1. If the projection of* $\mathbf{U}$ *on* $\mathbf{V}$ *satisfies the decay model of Hypothesis 2 with parameters* (C, α, γ)*, then either*

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

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 $\mathbf{b} \propto (1, \ldots, 1)$.

Proof of Lemma 1. Because $\|\mathbf{b}^N\|_2 = \|\mathbf{u}^N\|_2$, the model (10) 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 (11) we have convergence to zero of the bound (9) only if*

$$\alpha \geqslant 1 \quad \textit{and} \quad \gamma < 1$$
$$\textit{or}$$
$$\alpha < 1 \quad \textit{and} \quad \alpha > \gamma$$

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 [28] in $\frac{C}{j^\alpha}$ satisfies the convergence hypothesis. However, we saw with the previous lemma that this model is not compatible with Hypothesis 1 since it would require the energy of the infinite image to be concentrated on a finite support.

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

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

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

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

The second and the third cases lead to

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

and

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

□

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 can further use this result to prove 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*

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

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

D. 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 us to estimate these coefficients from a particular filter/basis based on a truncated image sequence.

The next Section IV-D1 explains how these model parameters are estimated from real data. Then in Section IV-D2 we analyze the asymptotic performance of several denoising algorithms based on the estimated parameters.

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:

$$|b_j^N| \approx C\frac{N^\gamma}{j^\alpha}. \tag{12}$$

Observing different curves $j \mapsto |b_j^N|$ for various images and sizes N in loglog scale (see the first column of Figure 8 for an example), we notice that the model (12) 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 6). 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.

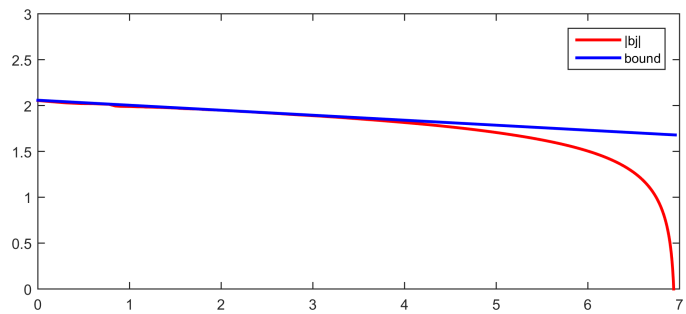


Figure 6. Decay of the coefficients $|b_j|$ for white noise in the DCT basis (red) in loglog scale. The slope of the bound is $\alpha_m \approx 0.05$ (blue).

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

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

with N from N_{min} to N_{max} and j from d to $\lfloor N^p \rfloor$.

2) Experimental results & concluding remarks: Table I shows the estimated model parameters for the 4 test images in Figure 7 and for 4 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;

Prefiltered NLM: The orthogonal basis which diagonalizes the (symmetrized) second iteration of a non-local means

bricks **simpson** **bridge** **tank**

Figure 7. The four images used for the experiments. The image size is 160×160 and for experiments we took sub-images of increasing size 64×64, 96×96, 128×128 and 160×160 from left-top corner.

Table I
FITTED PARAMETERS α AND γ FOR THE DIFFERENT IMAGES OF FIGURE 7 IN DIFFERENT BASES. THE PARAMETER r IS THE DECAY RATE OF COROLLARY 1.

image	basis	fitted α	fitted γ	r
bricks	DCT	0.699	0.606	0.094
	Wavelet	0.587	0.632	-
	O-NLM	0.756	0.500	0.256
	P-NLM	0.624	0.500	0.124
simpson	DCT	0.777	1.040	-
	Wavelet	0.811	1.020	-
	O-NLM	0.825	0.788	0.037
	P-NLM	0.568	0.663	-
bridge	DCT	0.640	0.583	0.057
	Wavelet	0.713	0.656	0.057
	O-NLM	0.759	0.500	0.259
	P-NLM	0.651	0.501	0.150
tank	DCT	0.599	0.578	0.021
	Wavelet	0.609	0.593	0.016
	O-NLM	0.725	0.500	0.225
	P-NLM	0.544	0.500	0.044

filter, *i.e.* with patch distances computed on the NLM-filtered noisy image, as explained in Section III-C.

In all cases the oracle NLM basis satisfies the conditions of Theorem 2 and provides the fastest asymptotic convergence rate (near $r = 0.25$ except for the cartoon-like simpson image). On the other hand the DCT and wavelet bases sometimes do not satisfy the conditions of Theorem 2, and when they do, the asymptotic convergence rate is extremely slow (always smaller than $r = 0.1$).

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 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 prefiltered NLM. Directly computing the NLM filter on the noisy image is not acceptable as explained in Section III-C. However applying it to a NLM filtered version of the image helps both *(a)* satisfy the requirement of independence of the filter and noise, and *(b)* make the filter closer to the oracle one. As shown in Table I, the asymptotic convergence rate we estimated for the prefiltered NLM basis is much better than that of the dct or wavelet bases, but still lags behind that of the oracle NLM basis.

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

- 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 (12) can not always be perfectly fit by all images and bases, particularly not by the prefiltered NLM basis as shown in the first column of Figure 8 and in Figure 10.
- Model (12) only gives a coarse upper bound for the actual MSE^*. The second column in Figure 8 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 9) 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 (0.15 for prefiltered NLM vs 0.06 for DCT and 0.26 for oracle 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 [37], 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 [38]. This shall be the subject of further research.

V. 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 [28], [27] 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 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 [31].

Our generalized theory provides less optimistic conclusions than those in [28] 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 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

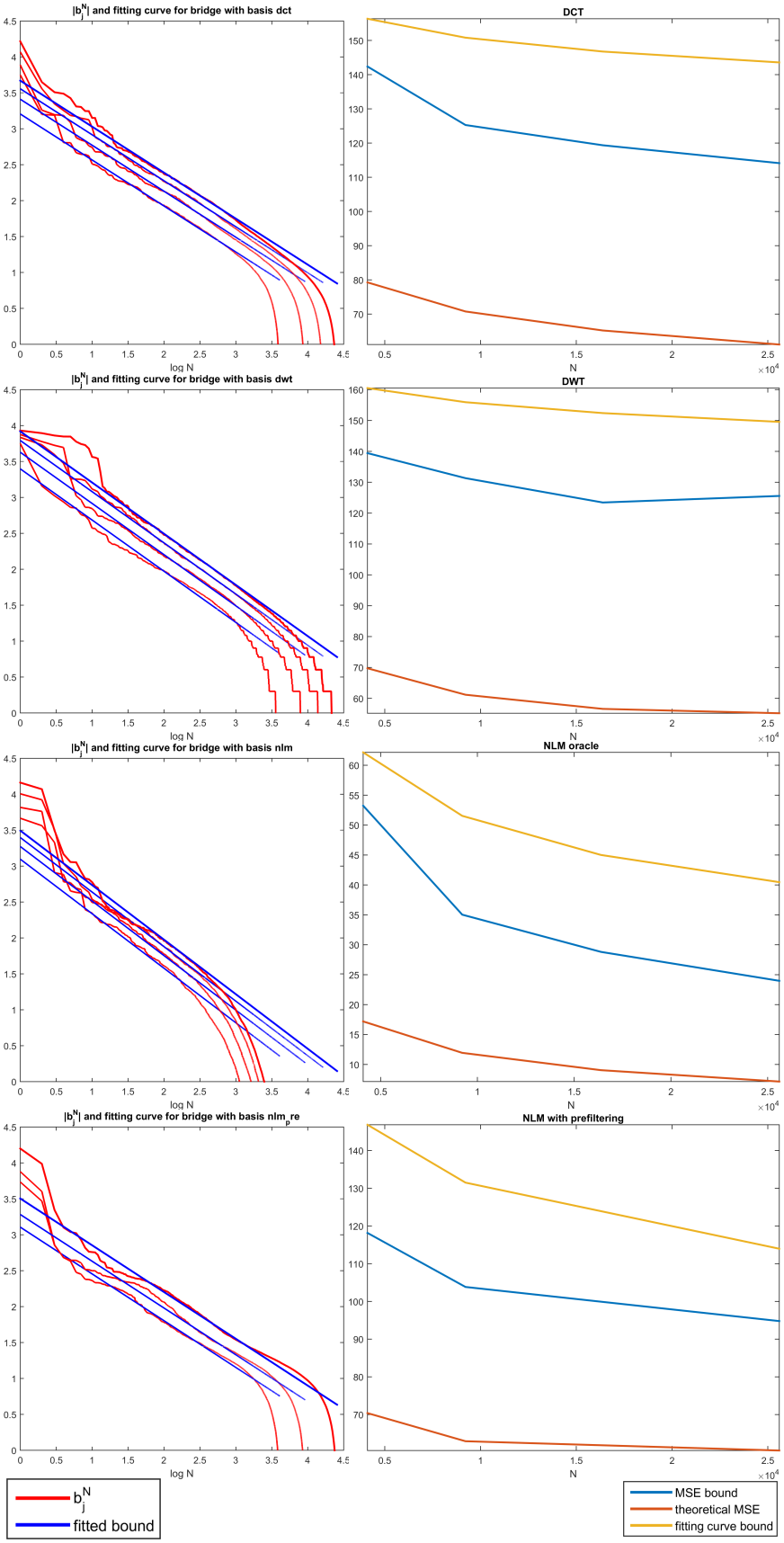


Figure 8. Left column: the decay of the $|b_j^N|$ (red) and the result of the model fitting (blue) for the bridge image for the different bases (from top to bottom) DCT, Wavelet, O-NLM and P-NLM. Right column: the decay of the theoretical $\mathrm{MSE}^\star$ (orange), the upper bound of the theoretical MSE (blue) and the decay provided by the fitting curve (yellow).

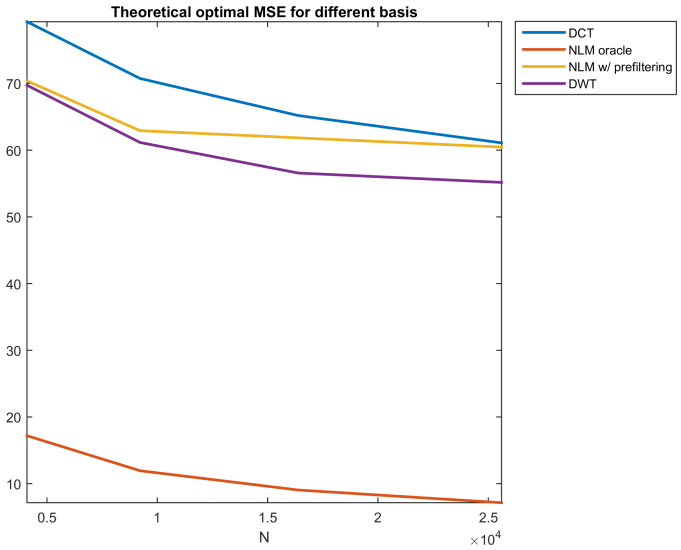


Figure 9. Comparison of the theoretical MSE* from the right column of the figure 8 on the image bridge for the different bases.

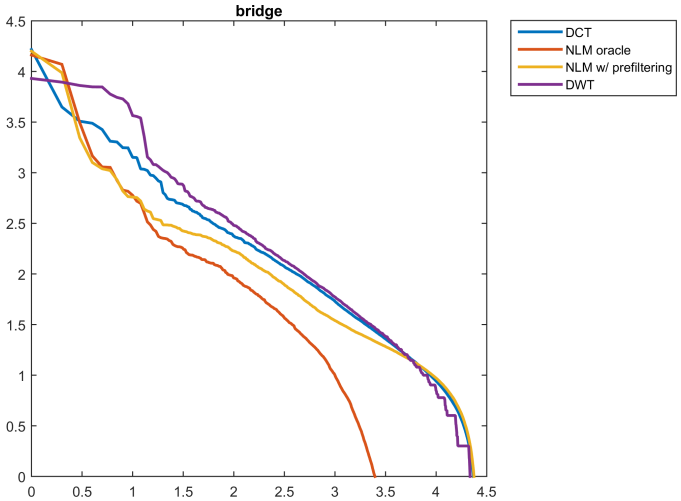


Figure 10. Comparison of the decay of the $|b_j^N|$ for the different bases on image bridge.

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.

References

[1] L. Yaroslavsky, *Digital picture processing: an introduction*. Springer Science & Business Media, 2012, vol. 9.

[2] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.

[3] F. Guichard, L. Moisan, and J.-M. Morel, "A review of pde models in image processing and image analysis," in *Journal de Physique IV (Proceedings)*, vol. 12, no. 1. EDP sciences, 2002, pp. 137–154.

[4] S. Mallat, *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.

[5] E. Ordentlich, G. Seroussi, S. Verdu, M. Weinberger, and T. Weissman, "A discrete universal denoiser and its application to binary images," in *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, vol. 1. IEEE, 2003, pp. I–117.

[6] T. Weissman, E. Ordentlich, G. Seroussi, S. Verdú, and M. J. Weinberger, "Universal discrete denoising: Known channel," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 5–28, 2005.

[7] S. Awate and R. Whitaker, "Image denoising with unsupervised information-theoretic adaptive filtering," in *International Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, 2004, pp. 44–51.

[8] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2. IEEE, 2005, pp. 60–65.

[9] C. Kervrann and J. Boulanger, "Optimal Spatial Adaptation for Patch-Based Image Denoising," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 2866–2878, Oct. 2006.

[10] ——, "Local Adaptivity to Variable Smoothness for Exemplar-Based Image Regularization and Representation," *Int. J. Comput. Vis.*, vol. 79, no. 1, pp. 45–69, Nov. 2007. [Online]. Available: http://link.springer.com/10.1007/s11263-007-0096-2

[11] C.-A. Deledalle, L. Denis, and F. Tupin, "Iterative weighted maximum likelihood denoising with probabilistic patch-based weights." *IEEE Trans. Image Process.*, vol. 18, no. 12, pp. 2661–72, Dec. 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19666338

[12] C.-A. Deledalle, F. Tupin, and L. Denis, "Poisson {NL} means: Unsupervised non local means for poisson noise," in *2010 IEEE Int. Conf. Image Process.*, ser. IEEE International Conference on Image Processing ICIP, IEEE; IEEE Signal Process Soc. 345 E 47TH ST, New York, NY 10017 USA: IEEE, 2010.

[13] G. Peyré, S. Bougleux, and L. Cohen, "Non-local Regularization of Inverse Problems," in *ECCV 2008*. Marseille, France: Springer-Verlag, 2008, pp. 57–68. [Online]. Available: http://portal.acm.org/citation.cfm?id=1478179http://www.springerlink.com/content/g748l32688pn4581

[14] A. Buades, B. Coll, J.-M. Morel, and C. Sbert, "Self-similarity driven color demosaicking." *IEEE Trans. Image Process.*, vol. 18, no. 6, pp. 1192–202, Jun. 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19403366

[15] P. Arias, V. Caselles, and G. Facciolo, "Analysis of a Variational Framework for Exemplar-Based Image Inpainting," *Multiscale Model. Simul.*, vol. 10, no. 2, pp. 473–514, Jan. 2012. [Online]. Available: http://epubs.siam.org/doi/abs/10.1137/110848281

[16] M. Lebrun, A. Buades, and J. M. Morel, "A Nonlocal Bayesian Image Denoising Algorithm," *SIAM J. Imaging Sci.*, vol. 6, no. 3, pp. 1665–1688, Sep. 2013. [Online]. Available: http://epubs.siam.org/doi/abs/10.1137/120874989

[17] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *2011 Int. Conf. Comput. Vis.* IEEE, Nov. 2011, pp. 479–486. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6126278http://eprints.pascal-network.org/archive/00009011/01/EPLLICCVCameraReady.pdf

[18] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: from Gaussian mixture models to structured sparsity." *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2481–99, May 2012. [Online]. Available: 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

[19] Y.-Q. Wang and J.-M. Morel, "SURE Guided Gaussian Mixture Image Denoising," *SIAM J. Imaging Sci.*, vol. 6, no. 2, pp. 999–1034, May 2013. [Online]. Available: http://epubs.siam.org/doi/abs/10.1137/120901131

[20] M. Elad and M. Aharon, "Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006. [Online]. Available: http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=4011956

[21] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009, pp. 2272–2279.

[22] C. Knaus and M. Zwicker, "Dual-domain image denoising." in *ICIP*, 2013, pp. 440–444.

[23] N. Pierazzo, M. Lebrun, M. Rais, J.-M. Morel, and G. Facciolo, "Non-local dual image denoising," in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 813–817.

[24] N. Pierazzo, M. Rais, J.-M. Morel, and G. Facciolo, "Da3d: Fast and data adaptive dual domain denoising," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 432–436.

[25] V. Duval, J.-F. Aujol, and Y. Gousseau, "A bias-variance approach for the nonlocal means," *SIAM Journal on Imaging Sciences*, vol. 4, no. 2, pp. 760–788, 2011.

[26] A. Levin and B. Nadler, "Natural image denoising: Optimality and inherent bounds," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 2833–2840.

[27] H. Talebi and P. Milanfar, "Global image denoising," *Image Processing, IEEE Transactions on*, vol. 23, no. 2, pp. 755–768, 2014.

[28] ——, "Global denoising is asymptotically optimal," *ICIP*, vol. 5, no. 3, 2014.

[29] ——, "Asymptotic performance of global denoising," *SIAM Journal on Imaging Sciences*, vol. 9, no. 2, pp. 665–683, 2016.

[30] D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[31] D. L. Donoho, I. M. Johnstone *et al.*, "Ideal denoising in an orthonormal basis chosen from a library of bases," *Comptes Rendus de l'Academie des Sciences-Serie I-Mathematique*, vol. 319, no. 12, pp. 1317–1322, 1994.

[32] Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad, "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*. IEEE, 1993, pp. 40–44.

[33] P. Milanfar, "Symmetrizing smoothing filters," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 263–284, 2013.

[34] ——, "A tour of modern image filtering: New insights and methods, both practical and theoretical," *Signal Processing Magazine, IEEE*, vol. 30, no. 1, pp. 106–128, 2013.

[35] S. H. Chan, T. Zickler, and Y. M. Lu, "Demystifying symmetric smoothing filters," *arXiv preprint arXiv:1601.00088*, 2016.

[36] G. Facciolo, A. Almansa, J.-F. Aujol, and V. Caselles, "Irregular to Regular Sampling, Denoising, and Deconvolution," *SIAM MMS*, vol. 7, no. 4, pp. 1574–1608, jan 2009. [Online]. Available: http://epubs.siam.org/doi/abs/10.1137/080719443

[37] N. Halko, P. G. Martinsson, and J. A. Tropp, "Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions," *SIAM Rev.*, vol. 53, no. 2, pp. 217–288, jan 2011. [Online]. Available: http://epubs.siam.org/doi/abs/10.1137/090771806

[38] M. Brand, "Fast low-rank modifications of the thin singular value decomposition," *Linear Algebra Appl.*, vol. 415, no. 1, pp. 20–30, may 2006. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0024379505003812