SIAM J. Imaging Sciences
Vol. 9, No. 2, pp. 665–683

Asymptotic Performance of Global Denoising*

Hossein Talebi† and Peyman Milanfar†

Abstract. We provide an upper bound on the rate of convergence of the mean-squared error for global image denoising and illustrate that this upper bound decays with increasing image size. Hence, global denoising is asymptotically optimal. At least in an oracle scenario this property does not hold for patch-based methods such as BM3D, thereby limiting their performance for large images. As observed in practice and shown in this work, this gap in performance is small for moderate size images, but it can grow quickly with image size.

Key words. image denoising bound, nonlocal filters, global filter, optimal image denoising

AMS subject classifications. 62G08, 93E14, 93E11, 15B51, 15B48, 60J60

DOI. 10.1137/15M1020708

1. Introduction.

When it comes to denoising, there is no shortage of algorithms. Patch-based methods have been the front-runners in performance; and as the race continues, newer methods have largely been variations on this theme. Leading patch-based methods have been modestly improved upon recently [1, 2] by innovations in the way patches are selected, how they are clustered, etc. Recently, we advocated abandoning the explicit use of patches [3] (as done in leading methods such as BM3D) in favor of a global approach where every pixel contributes to the denoising of every other pixel in the image. The similarity of pixels in this approach *can* still be measured using patches, but the application of the filter is truly global. The advantage of this approach is that it is asymptotically optimal in the sense that its mean-squared error (MSE) monotonically decays with increasing image size—a property that does not hold for any of the leading patch-based methods [4]—even if the size of the image grows infinitely large, and the range of search for similar patches is allowed to grow as well [5].

It is by now beyond dispute that images (or natural signals generally) contain many redundancies. This notion has been cleverly exploited to design high-performance image denoisers with great success. It stands to reason then that a good denoiser should exhibit improved performance as the number of samples (i.e., image size) grows. This concept is not new. In fact, we can go back as far as Shannon who pointed out [6] more than 60 years ago that "*If the source already has a certain redundancy and no attempt is made to eliminate it, a sizable fraction of the letters can be received incorrectly and still [perfectly] reconstructed by the context.*" More recently, this fundamental result was in fact shown for the case of restoring

*Received by the editors May 11, 2015; accepted for publication (in revised form) February 24, 2016; published electronically May 12, 2016. This work was supported in part by AFOSR grant FA9550-07-1-0365 and NSF grant CCF-1016018.

http://www.siam.org/journals/siims/9-2/M102070.html

†Department of Electrical Engineering, University of California, Santa Cruz, CA 95064 (htalebi@soe.ucsc.edu, milanfar@soe.ucsc.edu).

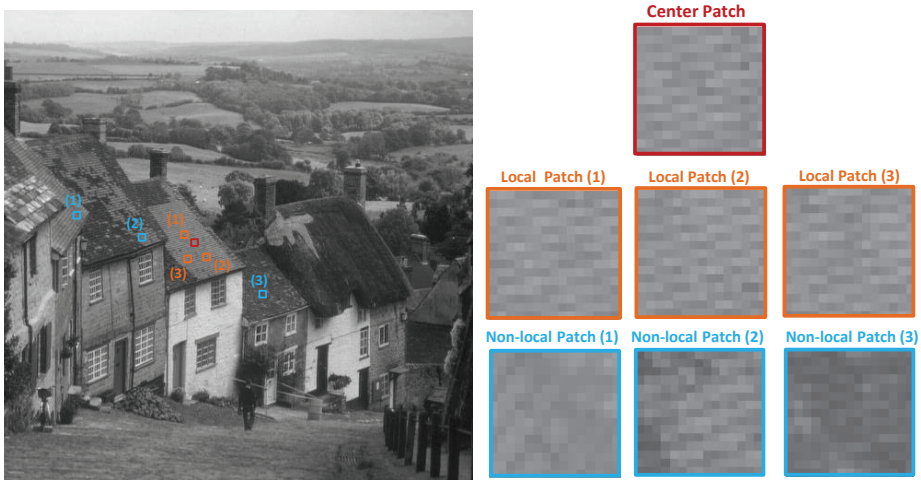


Figure 1. *Comparison of patch matching for local and nonlocal patches. Likelihood of finding* closely *similar patches drops as the size of the search window increases.*

binary images from context in [7, 8]. More relevant still, the seminal paper on the nonlocal means (NLM) method [9, 10, 11] was inspired in part by [7, 8] and itself gave a proof of the asymptotic consistency of the NLM method.

Over time, however, the idea of globally considering the denoising problem was abandoned in favor of more computationally friendly methods that treat patches (or groups of patches) together [4, 12]. Our approach in [3] relied on a truly global methodology where the effect of every pixel was taken into account to develop a denoiser, and the significant questions of computational complexity were also dealt with by using a subsampling strategy based on the Nyström extension. In the course of that work, we noted that the performance of the global approach consistently improved with image size, but the same was not observed for patch-based methods, hence motivating the work presented here. Intuitively, in larger images, as the total number of patches grows, the expected performance improvement due to availability of more overall patches is offset by the lower likelihood of finding closely matching patches (see Figure 1).[1] Increasing the size of the patches reduces the number of available patches, but increases the dimension of the space in which these patches live, hence sometimes providing a helpful effect, but never enough to drive the error to zero asymptotically (see Figure 2). As a result, performance flattens out with increasing image size.

The story is very different (and much more favorable) with global filters. These use *all* the pixels in the input image to denoise every single pixel. Here, we take the analysis a step deeper to prove that the performance of the global approach *always* improves as a function of image size, regardless of image content. Furthermore, we provide a rate for this improvement and show that this rate is a function of the sparsity of the image in a naturally constructed

[1]We are not claiming that one should only use the internal image similarities. Mosseri, Zontak, and Irani propose a method [13] to use "external" and "internal" image patches to improve the denoising performance. The global filter could include external similarities incorporated into it such that the growing size of this extra (and of course relevantly similar) information should essentially decrease the denoising error.

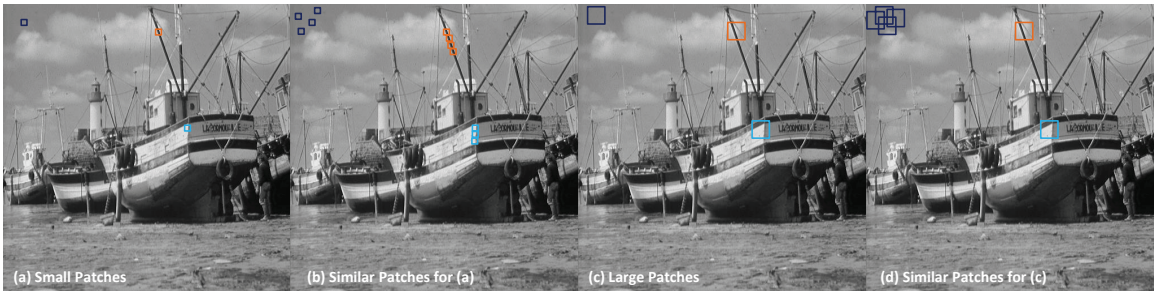


Figure 2. *Comparison of patch matching for different patch sizes. As the patch size grows, fewer similar patches are available.*

basis adapted to the content of the image. More specifically, we give an oracle upper bound on the MSE for estimating each pixel using all the pixels in the image, and we show that for typical images, it decays at the rate of *at least* $n^{-\alpha}$ for a $\sqrt{n} \times \sqrt{n}$ image where $\alpha > 0$ depends on the content of the input image.

2. Background. Let's begin with the model of the problem. The additive noise model for measurement of a corrupted image is

$$\mathbf{y} = \mathbf{z} + \mathbf{e}, \tag{2.1}$$

where the zero-mean white noise vector $\mathbf{e}$ with variance σ^2 is added to the latent signal vector $\mathbf{z}$ of length n to get the noisy observation $\mathbf{y}$. Most popular restoration approaches can be summarized into the following filtering scheme [14]:

$$\widehat{\mathbf{z}} = \mathbf{W}\mathbf{y}, \tag{2.2}$$

where the $n \times n$ matrix $\mathbf{W}$ represents the employed filter to obtain the estimated image $\widehat{\mathbf{z}}$. More specifically, we construct the filtering matrix by first defining affinities between pixels, or patches around pixels. This can be done very generally [14]. But here we can, for instance, use the NLM definition of weights to measure the similarity between the samples y_i and y_j as

$$K_{ij} = \exp\left\{\frac{-\|\mathbf{y}_i - \mathbf{y}_j\|^2}{h^2}\right\}, \tag{2.3}$$

where $\mathbf{y}_i$ and $\mathbf{y}_j$ are patches centered at y_i and y_j, respectively. The denoiser is then described as follows:

$$\widehat{\mathbf{z}} = \begin{bmatrix} \mathbf{w}_1^T \\ \mathbf{w}_2^T \\ \vdots \\ \mathbf{w}_n^T \end{bmatrix} \mathbf{y} = \mathbf{W}\mathbf{y}, \tag{2.4}$$

where the ith row of the matrix $\mathbf{W}$ defined above contains the corresponding normalized weights as

$$\mathbf{w}_i = \frac{1}{\sum_{j=1}^{n} K_{ij}} [K_{i1}, K_{i2}, \ldots, K_{in}]^T. \tag{2.5}$$

The filter matrix $\mathbf{W}$ can be closely approximated with a positive-definite, doubly stochastic and symmetric matrix [14, 15], and therefore its eigendecomposition can be expressed as

$$\mathbf{W} = \mathbf{V}\mathbf{S}\mathbf{V}^T, \tag{2.6}$$

in which the columns of the matrix $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ form an orthonormal basis, and where $\mathbf{S} = \text{diag}[\lambda_1, \ldots, \lambda_n]$ denotes the eigenvalues (i.e., shrinkage factors) in decreasing order $0 \leq \lambda_n \leq \cdots < \lambda_1 = 1$.

It is worth pointing out that this global description subsumes the local filter descriptions because even if the pixels are estimated locally, the effect of overlapped patches and the corresponding aggregations can be reflected with a simple modification to this global description. Having the eigendecomposition of the filter and the aggregation matrix $\mathbf{A}$, the matrix $\mathbf{W}$ can be decomposed as

$$\mathbf{W} = \mathbf{A}\mathbf{V}\mathbf{S}\mathbf{V}^T. \tag{2.7}$$

Numerous denoising algorithms have been proposed to find the optimal basis, shrinkage, and aggregation strategy. In general, most of these methods try to use a set of local basis functions in which the patches have a compact representation. Fixed basis functions such as wavelet and DCT [4, 16], data adapted functions obtained from principal component analysis (PCA) [17, 18], and training-based dictionaries [12, 19] are a few examples of the commonly used bases. While the basis selection strategies vary widely, the Wiener shrinkage has been established as the optimal strategy for minimizing the mean-squared error (MSE) for a given basis [4, 14, 20].

To overcome the limitations of this wide class of denoising filters, we can consider a somewhat different scenario where (1) the patch matching and filtering procedure is replaced by matching similar *pixels* (with some appropriate context provided possibly by patches), and (2) *all* the pixels in the image are forced to contribute in denoising every single pixel.[2] These conditions are equivalent to having an identity aggregation matrix $\mathbf{A} = \mathbf{I}$ and using a global rather than local basis in (2.7). This is what we advocate, and as we will show, a key consequence of this global approach is to get continuously better denoising performance for increasingly larger images.

As discussed in [14], spatial domain filters such as nonlocal means (NLM) [9] can be interpreted as transform domain filters, where eigenvectors of the filter matrix $\mathbf{W}$ form the orthonormal basis and the eigenvalues are the shrinkage coefficients, respectively. In practice, the global eigenvectors corresponding to the leading eigenvalues encode the latent image contents [3] well, whereas the same cannot be said for the aggregated collection of local eigenvectors provided by the patch-based methods. Furthermore, the spectral decomposition of

[2]To clarify, patch-based denoising includes collaborative filtering of similar patches in a limited size search window. Typically, in this filtering scheme, a small number of local patches are matched and filtered together. As a result, only a restricted number of pixels are incorporated in denoising each pixel. In contrast, the proposed global denoising scheme can be interpreted as a pixel averaging scheme in a search window as large as the underlying image. In other words, unlike patch-based methods, which make a decision a priori about how many patches to compare, the proposed global filter does not have this restriction and allows all patches to be compared to each other. Furthermore, by using NLM affinities in the global filter, the problem of finding similar pixels is alleviated.

the global filter makes it possible to have a relatively straightforward estimation of the MSE, leading to global performance analysis. As we will explain in the next section, the obtained MSE function is in fact inversely dependent on the number of pixels, and directly related to the efficiency of the basis functions in representing the image. Building on our previous work in [21], our current results suggest that (1) there is a huge performance gap between "oracle" versions of patch-based denoising and the proposed global scheme, and (2) as the image size grows, the MSE of the global scheme asymptotically decays, whereas the same is not essentially true of strictly patch-based methods.

The global denoising bound for stationary signals was previously introduced in [21]. In the present work, the denoising bound is extended to generic images beyond stationary signals. More explicitly, by employing an image clustering scheme, each group of similar pixels can be treated as stationary, leading to an estimation of the overall MSE bound. Furthermore, this work also discusses the effect of the Nyström filter approximation [3] on the estimated bound.

3. Computing and bounding the oracle global MSE. The filter $\mathbf{W}$, being data dependent, is of course impacted by the noise in the given image. In practice the filter is never computed directly from the raw, noisy input pixels. Instead, a "prefilter" is always applied to $\mathbf{y}$ first to reduce the effect of noise,[3] and then the filter weights are computed from this result. When it comes time to the actual filtering, however, this is done using the filter coefficients on the original noisy pixels. In the present discussion, since we are interested in the *oracle* performance, we consider the case where the filter is directly computed from the clean latent image $\mathbf{z}$ and is therefore deterministic.

Recall that each row of the filter can be expressed as

$$\mathbf{w}_i^T = \sum_{j=1}^{n} \lambda_j \mathbf{v}_j(i) \mathbf{v}_j^T, \tag{3.1}$$

where $\mathbf{v}_j(i)$ denotes the ith entry of the jth eigenvector. Then each estimated pixel $\widehat{z}_i$ has the following form:

$$\widehat{z}_i = \sum_{j=1}^{n} \lambda_j \mathbf{v}_j(i) \mathbf{v}_j^T \mathbf{y}. \tag{3.2}$$

The bias of this estimate is

$$\operatorname{bias}(\widehat{z}_i) = z_i - \mathrm{E}(\widehat{z}_i) = z_i - \sum_{j=1}^{n} \lambda_j \mathbf{v}_j(i) b_j, \tag{3.3}$$

where $\mathbf{b} = \mathbf{V}^T \mathbf{z} = [b_1, \ldots, b_n]^T$ contains the coefficients representing the latent image in the

[3]The prefilter can be a simple denoiser such as NLM. By applying the prefilter, the filter matrix $\mathbf{W}$ is mostly dependent on the latent image rather than the noisy input image. We have previously shown that applying the global filter on the prefiltered image can effectively improve the denoising performance [3].

global basis. The variance, for its part, has the following form:

$$\begin{aligned}\operatorname{var}(\widehat{z}_i) &= \sigma^2(\mathbf{w}_i^T\mathbf{w}_i) = \sigma^2\left(\sum_{j=1}^{n}\lambda_j\mathbf{v}_j(i)\mathbf{v}_j^T\right)\left(\sum_{j'=1}^{n}\lambda_{j'}\mathbf{v}_{j'}(i)\mathbf{v}_{j'}\right)\\ &= \sigma^2\sum_{j=1}^{n}\lambda_j^2\mathbf{v}_j(i)^2. \qquad (3.4)\end{aligned}$$

Therefore, the (oracle) expected squared error of the ith estimated pixel is

$$\begin{aligned}\mathrm{E}[(\widehat{z}_i - z_i)^2] &= \operatorname{bias}(\widehat{z}_i)^2 + \operatorname{var}(\widehat{z}_i)\\ &= \left(z_i - \sum_{j=1}^{n}\lambda_j\mathbf{v}_j(i)b_j\right)^2 + \sigma^2\sum_{j=1}^{n}\lambda_j^2\mathbf{v}_j(i)^2\\ &= z_i^2 + \sum_{j=1}^{n}(\lambda_j^2(b_j^2+\sigma^2)\mathbf{v}_j(i)^2 - 2z_i\lambda_j\mathbf{v}_j(i)b_j). \qquad (3.5)\end{aligned}$$

Having $\mathrm{E}[(\widehat{z}_i - z_i)^2]$, the overall MSE for the whole image is

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[(\widehat{z}_i - z_i)^2] = \frac{1}{n}\sum_{i=1}^{n}z_i^2 + \frac{1}{n}\sum_{j=1}^{n}\left((\lambda_j^2 - 2\lambda_j)b_j^2 + \sigma^2\lambda_j^2\right), \qquad (3.6)$$

where $\|\operatorname{bias}(\widehat{\mathbf{z}})\|^2 = \sum_{i=1}^{n}z_i^2 + \sum_{j=1}^{n}(\lambda_j^2 - 2\lambda_j)b_j^2$ and $\operatorname{var}(\widehat{\mathbf{z}}) = \operatorname{tr}(\operatorname{cov}(\widehat{\mathbf{z}})) = \sigma^2\sum_{j=1}^{n}\lambda_j^2$. Since $\mathbf{V}$ is orthonormal, we can replace $\sum_{i=1}^{n}z_i^2$ with $\sum_{j=1}^{n}b_j^2$ to get

$$\mathrm{MSE} = \frac{1}{n}\sum_{j=1}^{n}(\lambda_j - 1)^2b_j^2 + \sigma^2\lambda_j^2. \qquad (3.7)$$

Minimizing the MSE with respect to the eigenvalues λ_i requires a simple differentiation:

$$\frac{\partial\mathrm{MSE}(\lambda)}{\partial\lambda} = 0 \implies \lambda_j^* = \frac{1}{1+\mathrm{snr}_j^{-1}}, \qquad (3.8)$$

where, somewhat unsurprisingly, the "optimal" eigenvalues $\{\lambda_j^*\}$ are the Wiener coefficients with $\mathrm{snr}_j = \frac{b_j^2}{\sigma^2}$. This shrinkage strategy leads to the minimum value of the MSE:[4]

$$\mathrm{MMSE} = \mathrm{MSE}(\lambda^*) = \frac{\sigma^2}{n}\sum_{j=1}^{n}\lambda_j^*. \qquad (3.9)$$

[4]Since the equivalent shrunk filter should be kept doubly stochastic, λ_1^* should in theory be 1, a constraint which increases the minimum MSE. This MSE increment is $\Delta\mathrm{MSE} = \frac{\sigma^2}{n}(1-\lambda_1^*) = \frac{\sigma^4}{n(b_1^2+\sigma^2)}$. Having the first eigenvector $\mathbf{v}_1 = \frac{1}{\sqrt{n}}\mathbf{1}_n$, the squared signal projection coefficient can be expressed as $b_1^2 = \frac{1}{n}(\sum_{i=1}^{n}z_i)^2$. Practically, for a moderate size image in the range of [0,255], $\Delta\mathrm{MSE}$ is very small and can be neglected (or, equivalently, $\lambda_1^* \approx 1$). Consequently, it is not necessary to impose $\lambda_1^* = 1$ as a constraint in our analysis.

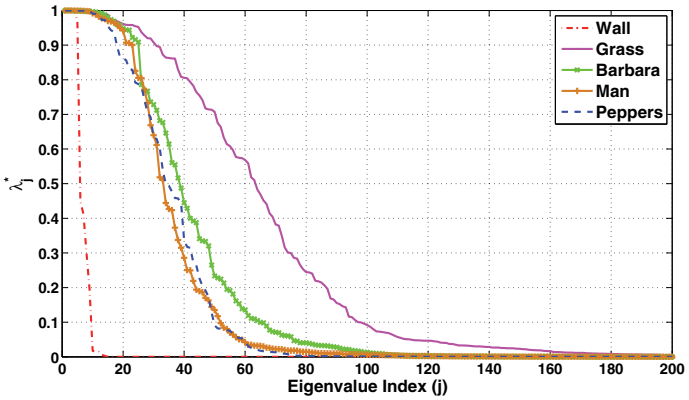


Figure 3. *Wiener shrinkage eigenvalues (λ_j^*) computed for some test images shown in Figure 7. Images with repetitive patterns such as Wall represent fast decaying Wiener coefficients. On the contrary, the optimal shrinkage factors of nonstationary images (e.g., Grass) drop off in a slow fashion.*

Figure 3 depicts Wiener shrinkage factors of some test images shown in Figure 7. Evidently, stationary images with repetitive patterns such as Wall show a faster decay rate of λ_j^*. This indicates that sparse signals (in the basis defined by the filter) are easier to recover from additive white noise. In the case of stationary signals, as the image size grows the decay rate of the Wiener shrinkage factors increases. Assuming that the shrinkage factors decay in some fashion, the minimum MSE given in (3.9) can be bounded (see section 3.1). Furthermore, by clustering pixels into relatively stationary subimages, this condition can be eased and a more generic denoising bound will be obtained (see section 3.3).

3.1. Bounding the oracle MSE of stationary images. Expanding the minimum MSE given by (3.9), we obtain

$$\text{MMSE} = \frac{\sigma^2}{n}\sum_{j=1}^{n}\lambda_j^* = \frac{1}{n}\sum_{j=1}^{n}\frac{\sigma^2 b_j^2}{\sigma^2 + b_j^2}. \tag{3.10}$$

The last equality can be expressed as

$$\text{MMSE} = \frac{1}{2n}\sum_{j=1}^{n}\frac{\sigma|b_j|}{\frac{\sigma^2+b_j^2}{2}}\sigma|b_j|. \tag{3.11}$$

Using the arithmetic-geometric means inequality [22], we have $\sigma|b_j| \le \frac{\sigma^2+b_j^2}{2}$, which implies

$$\text{MMSE} = \frac{1}{2n}\sum_{j=1}^{n}\frac{\sigma|b_j|}{\frac{\sigma^2+b_j^2}{2}}\sigma|b_j| \le \frac{1}{2n}\sum_{j=1}^{n}\sigma|b_j|, \tag{3.12}$$

and this in turn means

$$\text{MMSE} \le \frac{\sigma}{2n}\|\mathbf{b}\|_1. \tag{3.13}$$

The oracle MSE is evidently bounded by the l_1-norm of the projection coefficients $\mathbf{b}$. This implies that for a given n, the more sparse the signal is in the basis given by the filter kernel, the smaller the MSE error will be. Furthermore, for a signal (image) with finite energy, the 1-norm of $\mathbf{b}$ cannot grow faster than n with increasing dimension, so the upper bound must collapse to zero asymptotically. Let's consider the worst-case pathology wherein $|b_j| = c$ (a constant), resulting in linear growth of $\|\mathbf{b}\|_1$ with n. This essentially corresponds to the signal being "white noise" in the basis defined by the kernel. In this worst-case scenario, the MMSE is upper bounded by a constant $\frac{c\sigma}{2}$. In general, however, we expect the coefficients to drop off at some rate, say $\alpha > 0$. That is, $|b_j| = \frac{c}{j^\alpha}$, which implies that

$$\frac{\sigma}{2n}\sum_{j=1}^{n}|b_j| = \frac{\sigma}{2n}\sum_{j=1}^{n}\frac{c}{j^\alpha}. \tag{3.14}$$

As $n \to \infty$, MMSE will tend to zero for all $\alpha > 0$, so this establishes the most general (stationary) case of MSE convergence. Now let's have a look at the rate of convergence in more detail. For this purpose, it is useful to consider the coefficients b_j as samples of the function $|b(t)| = c/t^\alpha$. That is, define $b_j = b(j)$.

Using the integral test for convergence (Maclaurin–Cauchy test) [23], we have the following lower and upper bounds:

$$\frac{1}{n}\int_1^{n+1}\frac{c}{t^\alpha}\,dt \le \frac{1}{n}\sum_{j=1}^{n}|b_j| \le \frac{1}{n}\left(c + \int_1^{n}\frac{c}{t^\alpha}\,dt\right). \tag{3.15}$$

For $0 < \alpha < 1$ we have

$$c\left(\frac{(n+1)^{1-\alpha}-1}{(1-\alpha)n}\right) \le \frac{1}{n}\sum_{j=1}^{n}\frac{c}{j^\alpha} \le c\left(\frac{n^{1-\alpha}-\alpha}{(1-\alpha)n}\right), \tag{3.16}$$

which means a convergence rate of $O(n^{-\alpha})$.

On the other hand, for $\alpha = 1$ we have

$$\frac{c\ln(n+1)}{n} \le \frac{1}{n}\sum_{j=1}^{n}\frac{c}{j} \le \frac{c(1+\ln(n))}{n}, \tag{3.17}$$

which indicates a rate of $O(n^{-1}\ln(n))$. Finally, the decay rate is $O(n^{-1})$ for $\alpha > 1$ since the summation in (3.14) converges to a finite constant. In summary, as long as the coefficients decay at all, *at whatever rate*, the minimum MSE is guaranteed to approach zero. Of course, in the case of stationary images this decay rate is guaranteed to be fast. Yet, in the case of natural images, one might argue that the drop off rate could be hamstrung by image size increment. In other words, as the image size grows, the MSE bound computed for nonstationary images could be increasing. This is due to the evolving filter eigenvectors, which affect the projection coefficients b_j (or, equivalently, the Wiener shrinkage factors λ_j^*). We address this issue next.

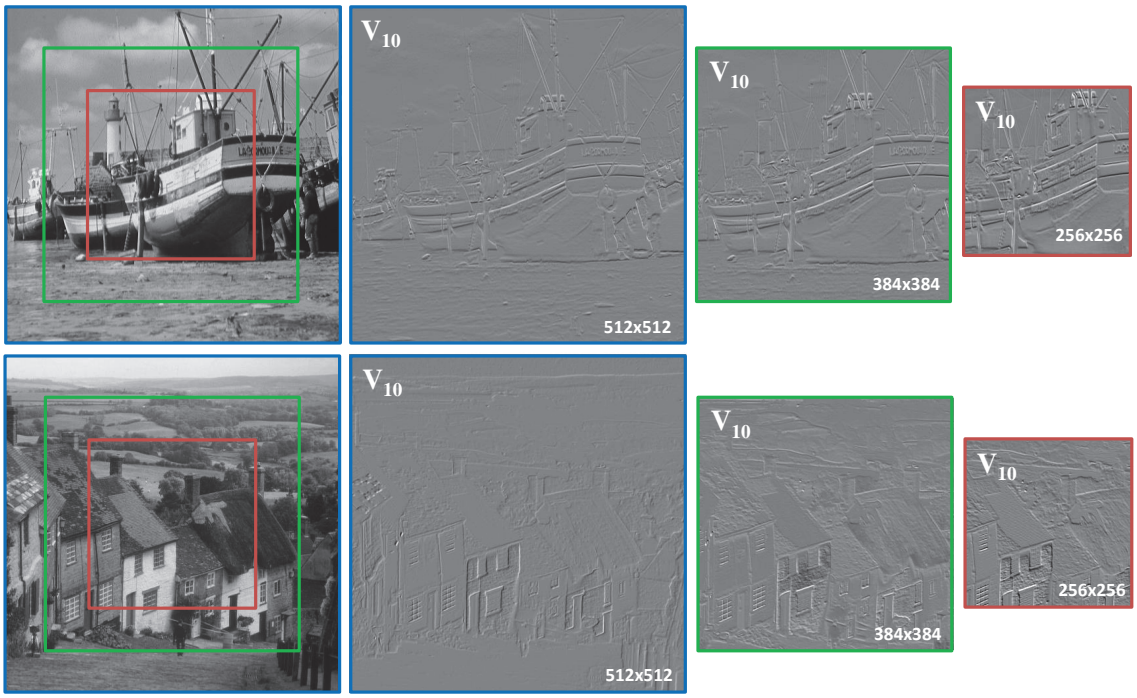


Figure 4. *Sample eigenvectors computed from image windows of different sizes. Top: Boat image. Bottom: Goldhill image. The 10th eigenvectors ($\boldsymbol{v}_{10}$) of the three subimages are illustrated.*

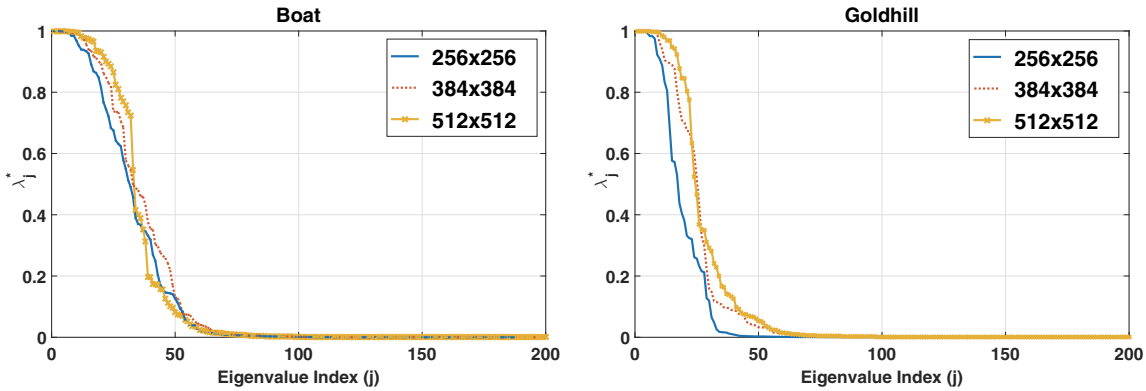


Figure 5. *Wiener shrinkage factors (λ_j^*) of the global filter computed for image windows of different sizes shown in Figure 4.*

3.2. Filter eigenvectors. Some eigenvector examples computed from image windows of sizes 256×256, 384×384, and 512×512 are shown in Figure 4. As the image size grows, the 256×256 eigenvector window may remain visually unchanged (such as the Boat image in Figure 4) or may alter across the three eigenvectors (such as the Goldhill image in Figure 4). Increasing the window size may introduce new content to the image, which may lead to changing eigenvectors. Figure 5 shows that the image size growth could affect the drop-off

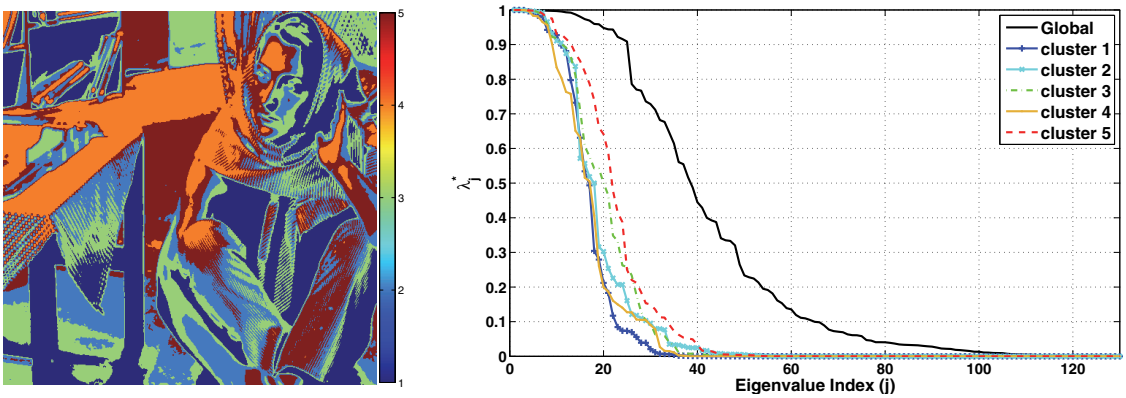


Figure 6. *Left: clustering map. Right: Wiener shrinkage factors (λ_j^*) of the global and clustered filters. The shrinkage coefficients of the clustered pixels show a faster decay rate compared to the global filter.*

rate of the Wiener shrinkage factors (λ_j^*). The shrinkage factors illustrated in Figure 5 are computed for the image windows given in Figure 4. This change in the decay rate of the λ_j^* coefficients has direct impact on the estimated bound. To tackle the problem of evolving eigenvectors, we propose clustering pixels, such that each cluster filter deals with relatively stationary subimages. We demonstrate this in the following section.

3.3. Bounding the oracle MSE of generic images. For pixel grouping, the concept of diffusion maps [24] is used, wherein each pixel located at position $\mathbf{x}_i$ is mapped into a manifold defined by the weighted eigenvectors as

$$\mathbf{\Psi}_k(\mathbf{x}_i) = (\lambda_2^k v_{i2}, \lambda_3^k v_{i3}, \ldots, \lambda_m^k v_{im}), \tag{3.18}$$

where v_{im} denotes the ith entry of the mth eigenvector and k represents the diffusion parameter. These descriptors are fed into the k-means classifier to obtain the clustering map of p clusters shown in Figure 6 (in our experiments k and p are set to 1 and 5, respectively). Our descriptors are then computed from the filter eigendecomposition, leading to clustering similar pixels together.

Now we are ready to apply a denoising filter to each cluster separately and compute the overall MSE bound. Figure 6 shows the benefit of grouping similar pixels over the global filter. As can be seen, in comparison to the global filter, the Wiener shrinkage factors decay more rapidly for the similar pixels in each cluster.

The overall minimum MSE bound can be expressed as

$$\begin{aligned} \text{MMSE} &\leq \frac{\sigma}{2n}\|\mathbf{b}_1\|_1 + \cdots + \frac{\sigma}{2n}\|\mathbf{b}_p\|_1 \\ &\leq \frac{\sigma}{2n}\sum_{l=1}^{p}\|\mathbf{b}_l\|_1, \end{aligned} \tag{3.19}$$

where p denotes the number of clusters and $\mathbf{b}_l$ represents the projection coefficients of the lth cluster onto the respective basis. With $p = 1$, the bound will be given by the expression

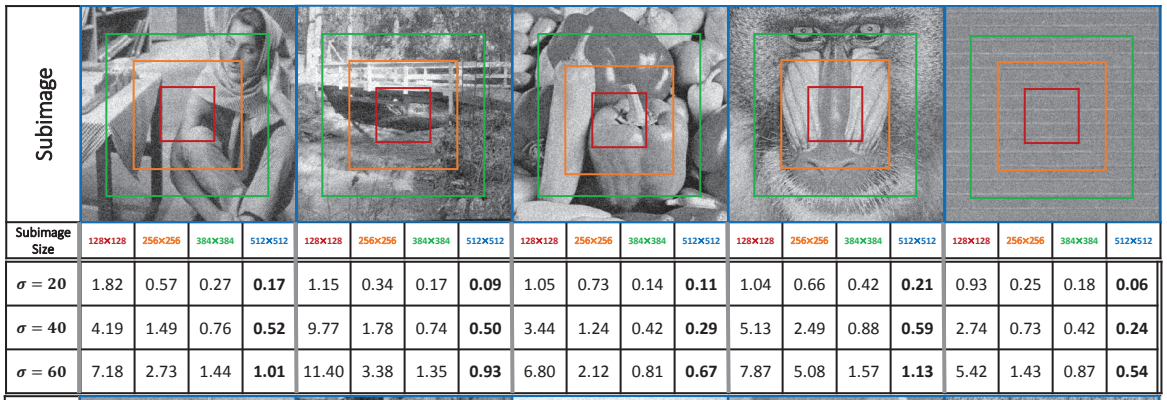
Figure 7. *Some benchmark images used to evaluate performance of our denoising method.*

in (3.13). As the image size grows, assuming that each newly added pixel falls, at worst, into one unique cluster of size 1 (meaning that $p = n$), the MMSE will be bounded by $\frac{\sigma max(\|\mathbf{b}_1\|_1,\ldots,\|\mathbf{b}_n\|_1)}{2}$. In practice, however, the number of clusters is much smaller than the number of pixels ($p \ll n$). When each subimage is relatively stationary, the overall decay rate of the MSE bound is determined by the minimum drop-off rate of $|\mathbf{b}_l|$ for $l = 1, \ldots, p$. In other words, as the image size grows, and assuming that no new cluster is added, the overall decay rate of the MSE bound is governed by the slowest term in (3.19). Next, we illustrate these results with some experiments.

4. Experiments. Some benchmark images used in this section are shown in Figure 7. The effect of image size on denoising performance is explored in the first set of experiments in Figure 8. For this experiment we denoised the central part of an image with increasing size. As can be seen, the increment in the number of pixels consistently leads to lower MSE. It is also important to highlight that the MSE values of the full size image (512×512) are very small and below round-off error.

The oracle MSE for the images in Figure 8 are shown in Figure 9, where for each noise level, the bound in (3.19) and the oracle MSE values are computed and then averaged across images given in Figure 7.[5] Each experiment is repeated and averaged for 20 independent noise realizations and 20 different initializations of the k-means clustering. In this example, the function $\frac{\gamma}{n^\alpha}$ is fitted on both bound and MSE data points using a least squares approach where n denotes the image size and γ is a constant. Table 1 represents the estimated α values for our test images. The estimated α_{MSE} values indicate the decay rate of MSE with respect to image size. As can be seen, stochastic textures such as Mandrill and Grass show smaller α_{MSE}. Although the estimated α_{Bound} does not match α_{MSE} (because (3.19) is an

[5]We note that for practical purposes, our experiments are carried out using a truncated filter with only a small percentage of the leading eigenvectors of $\mathbf{W}$. Still, the averaged MSEs in Figure 9 capture the decay rate for the tested images as hypothesized. See Appendix A.



Subimage Size	128×128	256×256	384×384	512×512	128×128	256×256	384×384	512×512	128×128	256×256	384×384	512×512	128×128	256×256	384×384	512×512	128×128	256×256	384×384	512×512
$\sigma = 20$	1.82	0.57	0.27	**0.17**	1.15	0.34	0.17	**0.09**	1.05	0.73	0.14	**0.11**	1.04	0.66	0.42	**0.21**	0.93	0.25	0.18	**0.06**
$\sigma = 40$	4.19	1.49	0.76	**0.52**	9.77	1.78	0.74	**0.50**	3.44	1.24	0.42	**0.29**	5.13	2.49	0.88	**0.59**	2.74	0.73	0.42	**0.24**
$\sigma = 60$	7.18	2.73	1.44	**1.01**	11.40	3.38	1.35	**0.93**	6.80	2.12	0.81	**0.67**	7.87	5.08	1.57	**1.13**	5.42	1.43	0.87	**0.54**

Subimage Size	128×128	256×256	384×384	512×512	128×128	256×256	384×384	512×512	128×128	256×256	384×384	512×512	128×128	256×256	384×384	512×512	128×128	256×256	384×384	512×512
$\sigma = 20$	1.39	0.56	0.33	**0.22**	6.49	0.70	0.28	**0.12**	3.32	0.64	0.29	**0.17**	1.57	0.69	0.23	**0.09**	4.28	1.73	0.98	**0.71**
$\sigma = 40$	3.79	1.58	0.91	**0.47**	5.46	1.51	0.77	**0.33**	4.64	0.99	0.49	**0.28**	3.85	1.72	0.72	**0.21**	8.19	3.25	1.66	**1.41**
$\sigma = 60$	6.84	2.94	1.70	**1.04**	9.90	2.56	1.43	**0.65**	5.12	1.93	0.93	**0.48**	6.56	3.32	1.22	**0.51**	14.24	4.22	2.73	**2.61**

Figure 8. *Oracle performance of the global denoising scheme for different window sizes. MSE values are averaged over* 20 *independent WGN (white Gaussian noise) realizations and k-means initialization points.*

upper bound), relative ordering of the decay rates is preserved. The estimated bound is also depicted individually for each test image in Figure 10. As can be seen, the proposed bound nearly captures the decay rate of the MSE function.

The oracle performance of NLM [9], BM3D, and our approach are compared in Table 2. BM3D is a two-stage image denoising scheme in which the first stage, as a prefilter, provides a "pilot" estimate of the noise-free image. The second stage uses the prefiltered image to obtain the *near* optimal Wiener shrinkage using an estimate of the SNR and also to perform a more accurate patch matching. In other words, in the oracle BM3D the input of the second stage is the clean image, but the Wiener filter *assumes* it is a prefiltered image. As the noise level increases, results in Table 2 suggest that the gap between our oracle scheme and oracle BM3D grows monotonically. Figure 11 illustrates results from oracle BM3D and the proposed global filter. As can be seen, the global filter can produce results very close to the clean image. In terms of computational complexity, test images of size 512×512 with an unoptimized implementation of our method in MATLAB take, on average, about 80 seconds. The optimized C and MATLAB implementation of BM3D takes about 3 seconds for these images.

Next, the difference between the patch-based and pixel-based denoising is explored. In at

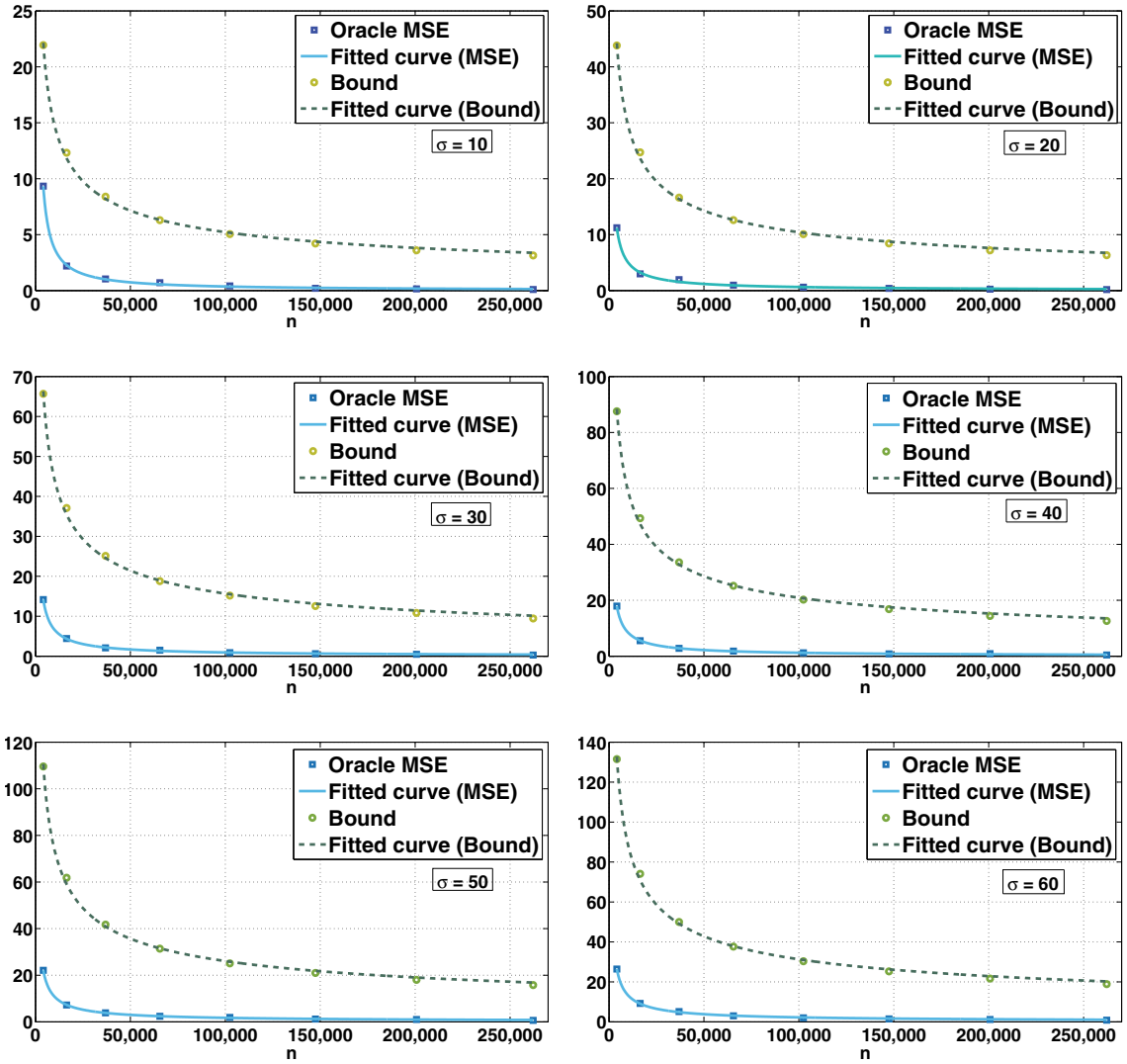


Figure 9. *Averaged MSE of denoising images given in Figure 7 for different noise levels. The estimated bound given in (3.19) is averaged across all the images.*

least the oracle scenario (the scope of this paper), in contrast to the patch-based denoisers, the pixel-based approaches benefit more from globalizing. We have observed that denoising performance of the patch-based methods will not substantially improve by increasing the size of the patch matching search window, unless the patch size is decreased as well. This is shown in Figure 12, where average MSE of the oracle BM3D is plotted for various image sizes and different patch sizes. In the left plot, the patch search window is enlarged to the size of the underlying image and all the overlapped patches are used in the collaborative filter. This means that all the available image patches are being used to denoise every single patch. Even though we are feeding noise-free patches to BM3D, its performance does not essentially

Table 1

Estimated decay rate of the oracle MSE and estimated bound obtained from test images corrupted by WGN with $\sigma = 30$. By using a least squares approach, $\frac{\gamma}{n^{\alpha}}$ is fitted on the data points of the MSE and estimated bound.

Test images	**Peppers**	**Man**	**Stream**	**Lena**	**Wall**
α_{MSE}	1.32 ($\pm$0.19)	0.98 ($\pm$0.15)	0.96 ($\pm$0.11)	0.94 ($\pm$0.08)	0.89 ($\pm$0.12)
α_{Bound}	0.57 ($\pm$0.05)	0.55 ($\pm$0.04)	0.52 ($\pm$0.05)	0.51 ($\pm$0.03)	0.50 ($\pm$0.01)
Test images	**Barbara**	**Goldhill**	**Boat**	**Mandrill**	**Grass**
α_{MSE}	0.79 ($\pm$0.04)	0.75 ($\pm$0.07)	0.71 ($\pm$0.13)	0.58 ($\pm$0.10)	0.55 ($\pm$0.16)
α_{Bound}	0.50 ($\pm$0.05)	0.48 ($\pm$0.03)	0.49 ($\pm$0.05)	0.44 ($\pm$0.02)	0.45 ($\pm$0.07)

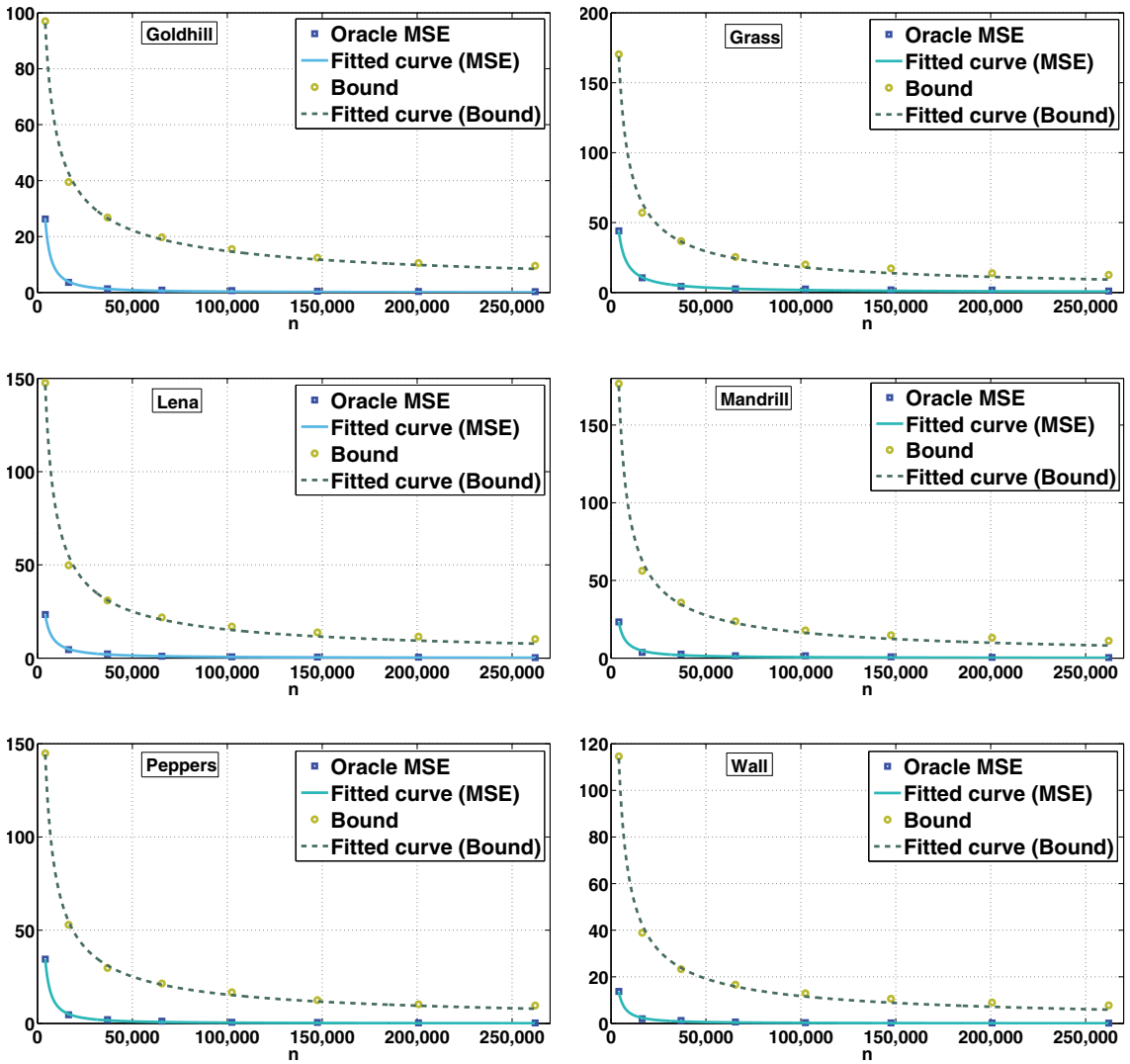


Figure 10. *Fitted curves of MSE and the estimated bound for some test images corrupted by WGN of $\sigma = 30$. The decay rates of the fitted curves are given in Table 1.*

Table 2

Oracle MSE values of NLM [9] (1st column), oracle BM3D [4] (2nd column), and Ours (3rd column). The MSE values are averaged over 20 independent noise realizations for each σ.

σ	Barbara			Stream			Peppers			Mandrill			Wall		
	NLM	BM3D	Ours	NLM	BM3D	Ours	NLM	BM3D	Ours	NLM	BM3D	Ours	NLM	BM3D	Ours
10	29.36	11.34	**0.08**	62.95	26.93	**0.05**	21.87	11.41	**0.06**	70.97	31.32	**0.02**	18.91	8.02	**0.03**
20	42.15	22.98	**0.17**	120.77	63.42	**0.09**	33.43	20.81	**0.11**	166.23	76.30	**0.06**	28.72	16.23	**0.05**
30	61.91	34.11	**0.29**	171.91	98.39	**0.29**	45.18	28.53	**0.38**	196.61	121.54	**0.16**	39.33	24.52	**0.09**
40	91.74	45.14	**0.52**	233.03	130.44	**0.50**	58.84	35.83	**0.59**	275.49	164.95	**0.24**	56.27	31.26	**0.15**
50	126.43	59.80	**0.73**	285.12	158.55	**0.81**	75.23	48.33	**0.86**	391.81	188.16	**0.42**	70.42	41.91	**0.20**
60	163.33	71.26	**1.01**	329.89	184.09	**0.93**	93.52	56.87	**1.13**	431.36	220.80	**0.54**	87.43	48.92	**0.27**
σ	Boat			Man			Goldhill			Lena			Grass		
	NLM	BM3D	Ours	NLM	BM3D	Ours	NLM	BM3D	Ours	NLM	BM3D	Ours	NLM	BM3D	Ours
10	31.87	14.30	**0.09**	36.25	16.26	**0.06**	31.16	14.93	**0.10**	19.77	9.71	**0.02**	68.35	30.04	**0.02**
20	56.96	29.42	**0.22**	67.85	35.59	**0.12**	53.47	29.72	**0.17**	31.07	18.34	**0.09**	164.01	75.72	**0.05**
30	80.44	43.16	**0.31**	94.89	53.72	**0.19**	78.67	42.06	**0.23**	44.68	26.28	**0.16**	191.16	120.56	**0.15**
40	108.72	56.16	**0.47**	125.95	70.61	**0.33**	105.51	52.91	**0.28**	60.46	34.13	**0.21**	270.90	160.11	**0.24**
50	138.42	75.05	**0.74**	158.01	93.27	**0.51**	131.17	69.78	**0.39**	76.56	45.22	**0.39**	390.17	186.94	**0.41**
60	168.69	88.32	**1.04**	188.98	109.14	**0.65**	155.07	80.54	**0.48**	92.94	53.70	**0.51**	420.11	211.56	**0.59**

Figure 11. *Visual comparison of the proposed method and oracle BM3D for the Man image.*

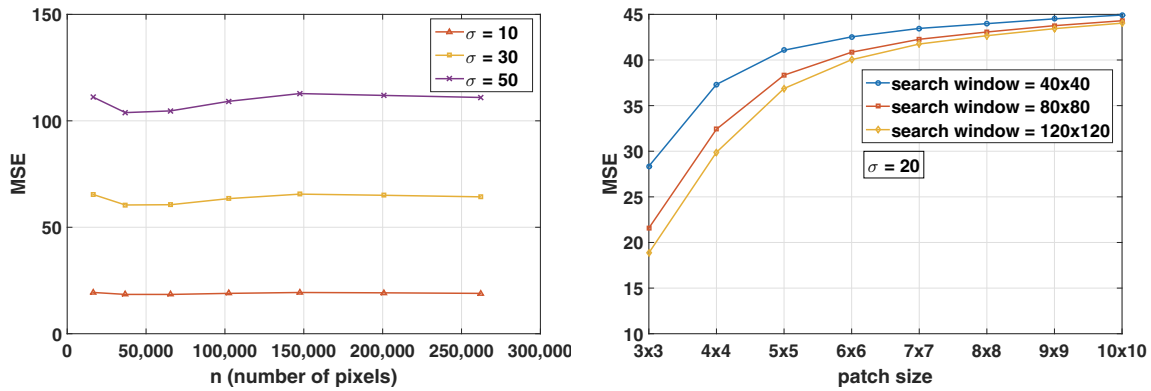


Figure 12. *Averaged MSE of denoising images given in Figure 7 using oracle BM3D. Left: varying image size and setting BM3D search window as large as the underlying image. Right: varying patch size and search window size.*

improve as the image size grows. However, by decreasing the patch size (right plot shown in Figure 12), the average MSE drops significantly. This is due to the more efficient patch matching of BM3D when dealing with smaller patches. This observation shows potential for future works. In practice, this might be useful in a denoising scenario where noise is first suppressed by a patch-based prefilter, and then further improvement is made by a second pass of the filter, which employs smaller patch sizes and larger search windows (e.g., global filter).

5. Conclusion. We emphasize that the oracle results do not correspond to practical denoising algorithms yet, and a practical realization of the global scheme remains to be studied. However, global filtering has an interesting asymptotic behavior that surpasses the existing patch-based bounds by a large margin. The oracle MSE values for the global filter converge to perfect reconstruction of the clean image, which is apparently impossible to achieve for oracle versions of algorithms such as BM3D. This implies that global filtering is promising as a way to see how much farther practical algorithms can be pushed.

Appendix A. The truncated filter and its MSE analysis. Keeping m leading eigenvectors of the filter $\mathbf{W}$, the truncated filter can be expressed as

$$\widehat{\mathbf{W}} = \mathbf{V}_m \mathbf{S}_m \mathbf{V}_m^T, \tag{A.1}$$

where $\mathbf{V}_m$ contains the first m eigenvectors as its columns and the $m \times m$ matrix $\mathbf{S}_m$ contains the m corresponding leading eigenvalues. Each row of the truncated filter can be expressed as

$$\widehat{\mathbf{w}}_i^T = \sum_{j=1}^{m} \lambda_j \mathbf{v}_j(i) \mathbf{v}_j^T. \tag{A.2}$$

In a fashion similar to the full-space filter $\mathbf{W}$, the truncated filter's expected squared error for the ith pixel is

$$\mathrm{E}^{(m)}[(\widehat{z}_i - z_i)^2] = z_i^2 + \sum_{j=1}^{m} (\lambda_j^2 (b_j^2 + \sigma^2) \mathbf{v}_j(i)^2 - 2 z_i \lambda_j \mathbf{v}_j(i) b_j). \tag{A.3}$$

The total MSE for the whole image is

$$\text{MSE}^{(m)} = \frac{1}{n}\sum_{i=1}^{n} \text{E}^{(m)}[(\widehat{z}_i - z_i)^2] = \frac{1}{n}\sum_{i=1}^{n} z_i^2 + \frac{1}{n}\sum_{j=1}^{m}\left((\lambda_j^2 - 2\lambda_j)b_j^2 + \sigma^2\lambda_j^2\right), \tag{A.4}$$

where $\|\text{bias}(\widehat{\mathbf{z}})\|^2 = \sum_{i=1}^{n} z_i^2 + \sum_{j=1}^{m}(\lambda_j^2 - 2\lambda_j)b_j^2$ and $\text{var}(\widehat{\mathbf{z}}) = \sigma^2\sum_{j=1}^{m}\lambda_j^2$. Truncation results in larger bias and smaller variance, as shown by the expressions.

The optimal eigenvalues are again the Wiener shrinkage coefficients (λ_j^*) which lead to the minimum MSE

$$\text{MMSE}^{(m)} = \frac{1}{n}\sum_{i=1}^{n} z_i^2 - \frac{1}{n}\sum_{j=1}^{m} b_j^2\lambda_j^*. \tag{A.5}$$

Of course, this MSE will necessary be larger than the case where all the eigenvectors are used. Replacing $\frac{1}{n}\sum_{i=1}^{n} z_i^2$ with $\frac{1}{n}\sum_{j=1}^{n} b_j^2$ and after some simplifications, we obtain

$$\begin{aligned}
\text{MMSE}^{(m)} &= \frac{1}{n}\sum_{j=1}^{n} b_j^2 - \frac{1}{n}\sum_{j=1}^{m} b_j^2\lambda_j^* \\
&= \frac{1}{n}\sum_{j=1}^{n} b_j^2 - \frac{1}{n}\sum_{j=1}^{n} b_j^2\lambda_j^* + \frac{1}{n}\sum_{j=m+1}^{n} b_j^2\lambda_j^* \\
&= \underbrace{\frac{1}{n}\sum_{j=1}^{n}\frac{\sigma^2 b_j^2}{\sigma^2 + b_j^2}}_{\text{MMSE}} + \underbrace{\frac{1}{n}\sum_{j=m+1}^{n}\frac{b_j^4}{b_j^2 + \sigma^2}}_{\Delta\text{MSE}^{(m)}},
\end{aligned} \tag{A.6}$$

where the first term is the same as the minimum MSE given in (3.10) and $\Delta\text{MSE}^{(m)}$ denotes the filter truncation effect on the MSE. We can show that the added MSE term is bounded and, consequently, the $\text{MMSE}^{(m)}$ will be upper bounded.

We start with an upper bound on $\Delta\text{MSE}^{(m)}$:

$$\Delta\text{MSE}^{(m)} = \frac{1}{n}\sum_{j=m+1}^{n}\frac{b_j^4}{b_j^2 + \sigma^2} \le \frac{1}{\sigma^2 n}\sum_{j=m+1}^{n} b_j^4 = \frac{1}{\sigma^2 n}(\|\mathbf{b}\|_4^4 - \|\mathbf{b}_m\|_4^4), \tag{A.7}$$

where[6] $\|\mathbf{b}_m\|_4^4 = \sum_{j=1}^{m} b_j^4$. Again, assuming a decay rate of $\alpha > 0$ for the coefficients $|b(t)| = c/t^\alpha$ and using the integral test for convergence [23], we obtain the following lower and upper bounds:

$$\frac{1}{\sigma^2 n}\int_{m+1}^{n+1}\frac{c^4}{t^{4\alpha}}\,dt \le \frac{1}{\sigma^2 n}\sum_{j=m+1}^{n} b_j^4 \le \frac{1}{\sigma^2 n}\left(\frac{c^4}{(m+1)^{4\alpha}} + \int_{m+1}^{n}\frac{c^4}{t^{4\alpha}}\,dt\right). \tag{A.8}$$

[6] In practice, for $j > m$ we have $b_j^2 \ll \sigma^2$. This means that $\frac{b_j^4}{b_j^2+\sigma^2} \le \frac{b_j^4}{\sigma^2} \le b_j^2$.

For $0 < \alpha < \frac{1}{4}$ we have

(A.9)
$$\frac{c^4}{\sigma^2}\left(\frac{(n+1)^{1-4\alpha}-(m+1)^{1-4\alpha}}{(1-4\alpha)n}\right) \leq \frac{1}{\sigma^2 n}\sum_{j=m+1}^{n}\frac{c^4}{j^{4\alpha}} \leq \frac{c^4}{\sigma^2}\left(\frac{1}{n(m+1)^{4\alpha}}+\left(\frac{n^{1-4\alpha}-(m+1)^{1-4\alpha}}{(1-4\alpha)n}\right)\right),$$

where a convergence rate of $O(n^{-4\alpha})$ is guaranteed. For $\alpha = \frac{1}{4}$ we have

$$\frac{c^4}{\sigma^2}\left(\frac{\ln(\frac{n+1}{m+1})}{n}\right) \leq \frac{1}{\sigma^2 n}\sum_{j=m+1}^{n}\frac{c^4}{j^4} \leq \frac{c^4}{\sigma^2}\left(\frac{1}{n(m+1)}+\frac{\ln(\frac{n}{m+1})}{n}\right), \tag{A.10}$$

which means a decay rate of $O(n^{-1}\ln(n))$. Similar to our previous analysis, for $\alpha > \frac{1}{4}$ the decay rate is $O(n^{-1})$. Overall, the minimum MSE of denoising by the truncated filter has the following bound:

$$\mathrm{MMSE}^{(m)} \leq \frac{\sigma}{2n}\|\mathbf{b}\|_1 + \frac{1}{\sigma^2 n}(\|\mathbf{b}\|_4^4 - \|\mathbf{b}_m\|_4^4). \tag{A.11}$$

It is not hard to see that even when m is kept fixed and n tends to infinity, a sufficiently fast decay of the coefficients $\mathbf{b}$ will still yield an asymptotic MSE of zero. In our experiments the ratio between m and n is kept fixed as $\frac{m}{n} = \frac{1}{1000}$, which leads to $m \approx 250$ for a test image of size 512×512.

REFERENCES

[1] I. Ram, M. Elad, and I. Cohen, *Image processing using smooth ordering of its patches*, IEEE Trans. Image Process., 22 (2013), pp. 2764–2774.

[2] M. Zontak, I. Mosseri, and M. Irani, *Separating signal from noise using patch recurrence across scales*, in IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 1195–1202.

[3] H. Talebi and P. Milanfar, *Global image denoising*, IEEE Trans. Image Process., 23 (2014), pp. 755–768.

[4] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, *Image denoising by sparse 3-D transform-domain collaborative filtering*, IEEE Trans. Image Process., 16 (2007), pp. 2080–2095.

[5] P. Chatterjee and P. Milanfar, *Practical bounds on image denoising: From estimation to information*, IEEE Trans. Image Process., 20 (2011), pp. 1221–1233.

[6] C. E. Shannon, *A mathematical theory of communication*, Bell System Tech. J., 27 (1948), pp. 379–423.

[7] E. Ordentlich, G. Seroussi, S. Verdu, M. Weinberger, and T. Weissman, *A discrete universal denoiser and its application to binary images*, in Proceedings of the 2003 International Conference on Image Processing (ICIP 2003), 2003.

[8] T. Weissman, E. Ordentlich, G. Seroussi, S. Verdu, and M. Weinberger, *A discrete universal denoiser and its application to binary images*, IEEE Trans. Inform. Theory, 51 (2003), pp. 5–28.

[9] A. Buades, B. Coll, and J. M. Morel, *A review of image denoising algorithms, with a new one*, Multiscale Model. Simul., 4 (2005), pp. 490–530, http://dx.doi.org/10.1137/040616024.

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

[11] C. Kervrann and J. Boulanger, *Optimal spatial adaptation for patch-based image denoising*, IEEE Trans. Image Process., 15 (2006), pp. 2866–2878.

[12] P. Chatterjee and P. Milanfar, *Clustering-based denoising with locally learned dictionaries*, IEEE Trans. Image Process., 18 (2009), pp. 1438–1451.

[13] I. Mosseri, M. Zontak, and M. Irani, *Combining the power of internal and external denoising*, in 2013 IEEE International Conference on Computational Photography, 2013, pp. 1–9.

[14] P. Milanfar, *A tour of modern image filtering*, IEEE Signal Process. Mag., 30 (2013), pp. 106–128.
[15] P. Milanfar, *Symmetrizing smoothing filters*, SIAM J. Imaging Sci., 6 (2013), pp. 263–284, http://dx.doi.org/10.1137/120875843.
[16] J. Portilla, V. Strela, M. Wainwright, and E. P. Simoncelli, *Image denoising using scale mixtures of Gaussians in the wavelet domain*, IEEE Trans. Image Process., 12 (2003), pp. 1338–1351.
[17] D. D. Muresan and T. W. Parks, *Adaptive principal components and image denoising*, in Proceedings of the 2003 International Conference on Image Processing (ICIP 2003), 2003, pp. 101–104.
[18] L. Zhang, W. Dong, D. Zhang, and G. Shi, *Two-stage image denoising by principal component analysis with local pixel grouping*, Pattern Recognition, 43 (2010), pp. 1531–1549.
[19] M. Elad and M. Aharon, *Image denoising via sparse and redundant representations over learned dictionaries*, IEEE Trans. Image Process., 15 (2006), pp. 3736–3745.
[20] P. Chatterjee and P. Milanfar, *Is denoising dead?*, IEEE Trans. Image Process., 19 (2010), pp. 895–911.
[21] H. Talebi and P. Milanfar, *Global denoising is asymptotically optimal*, in Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP 2014), 2014, pp. 818–822.
[22] J. M. Steele, *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities*, Cambridge University Press, Cambridge, UK, 2004.
[23] K. Knopp, *Infinite Sequences and Series*, Dover, New York, 1956.
[24] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, F. Warner, and S. Zucker, *Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps*, Proc. Natl. Acad. Sci. USA, 102 (2005), pp. 7426–7431.