Asymptotic Performance of Global Denoising

Hossein Talebi and Peyman Milanfar

Department of Electrical Engineering

University of California, Santa Cruz

Email : {htalebi, milanfar}@soe.ucsc.edu

EDICS Category: TEC-RST

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 to zero with increasing image size. Hence, global denoising is asymptotically optimal. 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.

Index Terms

Image Denoising Bound, Non-Local Filters, Global Filter, Optimal Image Denoising.

I. 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 converges to zero 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 binary images from context in [7], [8]. More relevant still, the seminal paper on the Non-Local Means (NLM) method [9] 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], [10]. 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 Fig. 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 Fig. 2). As a result, performance flattens out with increasing image size (see Fig. 7).

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 basis adapted to the content of the image. More specifically, we give an oracle upper bound on the mean-squared-error for estimating each pixel using all the pixels in the image, and show that for typical images, it decays at the rate of *at least* n^{-1} for a $\sqrt{n} \times \sqrt{n}$ image.
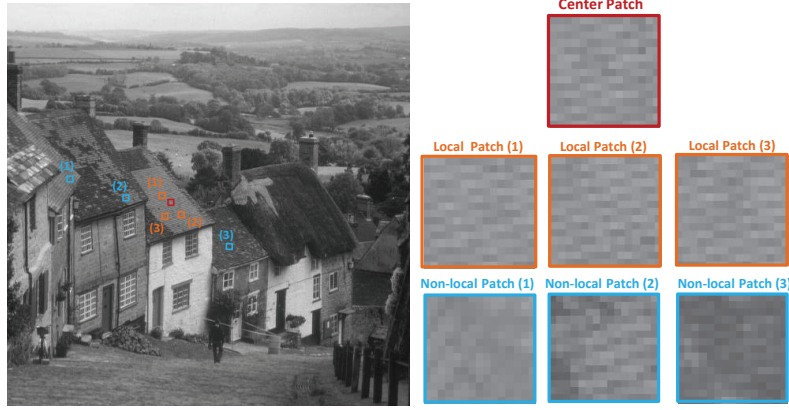
Fig. 1. Comparison of patch matching for local and non-local patches. Likelihood of finding *closely* similar patches drops as the size of the window search increases.
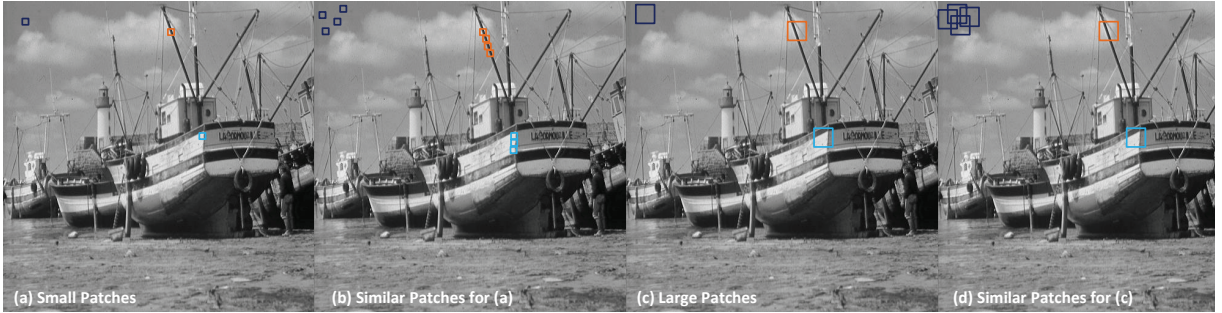

Fig. 2. Comparison of patch matching for different patch sizes. As the patch size grows, fewer similar patches are available.

II. BACKGROUND

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

$$\mathbf{y} = \mathbf{z} + \mathbf{e} \tag{1}$$

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 $\mathbf{y}$. Essentially all restoration approaches can be summarized into the following filtering scheme [11]:

$$\widehat{\mathbf{z}} = \mathbf{W}\mathbf{y} \tag{2}$$

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 [11]. 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:

$$K_{ij} = \exp\left\{\frac{-\|\mathbf{y}_i - \mathbf{y}_j\|^2}{h^2}\right\}, \tag{3}$$

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:

$$\widehat{\mathbf{z}} = \begin{bmatrix} \mathbf{w}_1^T \\ \mathbf{w}_2^T \\ \vdots \\ \mathbf{w}_n^T \end{bmatrix} \mathbf{y} = \mathbf{W}\mathbf{y}, \tag{4}$$

where the i-th row of the matrix $\mathbf{W}$ defined above contains the corresponding normalized weights as:

$$\mathbf{w}_i = \frac{1}{\sum_{j=1}^n K_{ij}} [K_{i1}, K_{i2}, \dots, K_{in}]^T. \tag{5}$$

The filter matrix $\mathbf{W}$ can be closely approximated with a positive-definite, doubly-stochastic and symmetric matrix [11], [12], and therefore its eigen-decomposition can be expressed as:

$$\mathbf{W} = \mathbf{V}\mathbf{S}\mathbf{V}^T, \tag{6}$$

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

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 the this global description. Having the eigen-decomposition of the filter and the aggregation matrix $\mathbf{A}$, the matrix $\mathbf{W}$ can be decomposed as:

$$\mathbf{W} = \mathbf{A}\mathbf{V}\mathbf{S}\mathbf{V}^T \tag{7}$$

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 sparse representation. Fixed basis functions such as wavelet and DCT [4], [13], data adapted functions obtained from principal component analysis (PCA) [14], [15] and, training based dictionaries [10], [16] 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 to minimize the mean-squared error (MSE) [4], [11], [17].

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 possible by patches,) and (2) *all* the pixels in the image are forced to contribute in denoising every single pixel. These conditions are equivalent to having an identity aggregation matrix $\mathbf{A} = \mathbf{I}$ and using a global rather than local basis in (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 [11], spatial domain filters such as Non-Local 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 can not be said for the aggregated collection of local eigenvectors provided by the patch-based methods. Furthermore, the spectral decomposition of the global filter makes it possible to have a relatively straight-forward 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 quality of the basis functions in representing the image (i.e. the sparsity of the image in the basis provided by the filter). Building on our previous work in [18], 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 approaches zero, whereas the same is not true of strictly patch-based methods.

III. 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 "pre-filter" is always applied to $\mathbf{y}$ first to reduce the effect of noise, 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:

$$\mathbf{w}_i^T = \sum_{j=1}^{n} \lambda_j \mathbf{v}_j(i) \mathbf{v}_j^T, \tag{8}$$

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

$$\widehat{z}_i = \sum_{j=1}^{n} \lambda_j \mathbf{v}_j(i) \mathbf{v}_j^T \mathbf{y}, \tag{9}$$

The bias of this estimate is:

$$\text{bias}(\widehat{z}_i) = z_i - \text{E}(\widehat{z}_i) = z_i - \sum_{j=1}^{n} \lambda_j \mathbf{v}_j(i) b_j \tag{10}$$

where $\mathbf{b} = \mathbf{V}^T \mathbf{z} = [b_1, ..., b_n]^T$ contains the coefficients representing the latent image in the global basis. The variance, for its part, has the following form:

$$
\begin{aligned}
\mathrm{var}(\widehat{z}_i) &= \sigma^2(\mathbf{w}_i^T \mathbf{w}_i) = \sigma^2 (\sum_{j=1}^n \lambda_j \mathbf{v}_j(i) \mathbf{v}_j^T)(\sum_{j'=1}^n \lambda_{j'} \mathbf{v}_{j'}(i) \mathbf{v}_{j'}) \\
&= \sigma^2 \sum_{j=1}^n \lambda_j^2 \mathbf{v}_j(i)^2
\end{aligned}
\tag{11}
$$

Therefore, the (oracle) MSE of the i-th estimated pixel is:

$$
\begin{aligned}
\mathrm{MSE}_i &= \mathrm{bias}(\widehat{z}_i)^2 + \mathrm{var}(\widehat{z}_i) \\
&= (z_i - \sum_{j=1}^n \lambda_j \mathbf{v}_j(i) b_j)^2 + \sigma^2 \sum_{j=1}^n \lambda_j^2 \mathbf{v}_j(i)^2 \\
&= z_i^2 + \sum_{j=1}^n (\lambda_j^2(b_j^2 + \sigma^2)\mathbf{v}_j(i)^2 - 2z_i \lambda_j \mathbf{v}_j(i) b_j)
\end{aligned}
\tag{12}
$$

Having MSE_i, the overall MSE for the whole image is:

$$
\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^n \mathrm{MSE}_i = \frac{1}{n}\sum_{i=1}^n z_i^2 + \frac{1}{n}\sum_{j=1}^n \left((\lambda_j^2 - 2\lambda_j)b_j^2 + \sigma^2 \lambda_j^2\right)
\tag{13}
$$

where $\|\mathrm{bias}(\widehat{\mathbf{z}})\|^2 = \sum_{i=1}^n z_i^2 + \sum_{j=1}^n (\lambda_j^2 - 2\lambda_j)b_j^2$ and $\mathrm{var}(\widehat{\mathbf{z}}) = \mathrm{tr}(\mathrm{cov}(\widehat{\mathbf{z}})) = \sigma^2 \sum_{j=1}^n \lambda_j^2$. 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:

$$
\mathrm{MSE} = \frac{1}{n}\sum_{j=1}^n (\lambda_j - 1)^2 b_j^2 + \sigma^2 \lambda_j^2
\tag{14}
$$

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

$$
\frac{\partial \mathrm{MSE}(\lambda)}{\partial \lambda} = 0 \implies \lambda_j^* = \frac{1}{1 + \mathrm{snr}_j^{-1}},
\tag{15}
$$

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[1]:

$$
\mathrm{MMSE} = \mathrm{MSE}(\lambda^*) = \frac{\sigma^2}{n}\sum_{j=1}^n \lambda_j^*
\tag{16}
$$

---

[1]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.

A. Bounding the Oracle MSE

Expanding the minimum MSE given by (16):

$$\text{MMSE} = \frac{\sigma^2}{n} \sum_{j=1}^{n} \lambda_j^* = \frac{1}{n} \sum_{j=1}^{n} \frac{\sigma^2 b_j^2}{\sigma^2 + b_j^2} \tag{17}$$

The last equality can be expressed as:

$$\text{MMSE} = \frac{1}{2n} \sum_{j=1}^{n} \frac{\sigma |b_j|}{\frac{\sigma^2 + b_j^2}{2}} \sigma |b_j| \tag{18}$$

Using the Arithmetic-Geometric means inequality [19] we have $\sigma |b_j| \leq \frac{\sigma^2 + b_j^2}{2}$, which implies:

$$\text{MMSE} = \frac{1}{2n} \sum_{j=1}^{n} \frac{\sigma |b_j|}{\frac{\sigma^2 + b_j^2}{2}} \sigma |b_j| \leq \frac{1}{2n} \sum_{j=1}^{n} \sigma |b_j| \tag{19}$$

And this in turn means

$$\text{MMSE} \leq \frac{\sigma}{2n} \|\mathbf{b}\|_1 \tag{20}$$

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 $\mathbf{b}$ can not 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 grown 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

$$\frac{\sigma}{2n} \sum_{j=1}^{n} |b_j| = \frac{\sigma}{2n} \sum_{j=1}^{n} \frac{c}{j^\alpha} \tag{21}$$

As $n \to \infty$, MMSE will tend to zero for all $\alpha > 0$, so this establishes the most general 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 (Maclaurlin-Cauchy test) [20], we have the following lower and upper bound:

$$\frac{1}{n} \int_1^{n+1} \frac{c}{t^\alpha} \, dt \leq \frac{1}{n} \sum_{j=1}^{n} |b_j| \leq \frac{1}{n} \left(c + \int_1^{n} \frac{c}{t^\alpha} \, dt \right) \tag{22}$$

For $0 < \alpha < 1$ we have:

$$c \left(\frac{(n+1)^{1-\alpha} - 1}{(1-\alpha)n} \right) \leq \frac{1}{n} \sum_{j=1}^{n} \frac{c}{j^\alpha} \leq c \left(\frac{n^{1-\alpha} - \alpha}{(1-\alpha)n} \right) \tag{23}$$

Fig. 3. Some benchmark images used to evaluate performance of our denoising method.

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

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

$$\frac{c\ln(n+1)}{n} \leq \frac{1}{n}\sum_{j=1}^{n}\frac{c}{j} \leq \frac{c(1+\ln(n))}{n} \tag{24}$$

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 (21) 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. Next, we illustrate these results with some experiments.

IV. EXPERIMENTS

Some benchmark images used in this section are shown in Fig. 3. Effect of the image size on the denoising performance is explored in the first set of experiments in Fig. 4. For this experiment we denoised different image windows 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.

Fig. 5 illustrates the image coefficients $|b_j|$ for some images in Fig. 4. These curves can be used to determine relative denoising limit of each image. For example, the *Goldhill* image coefficients are the lowest compared to the other images which predicts the least MSE values in Fig. 4.

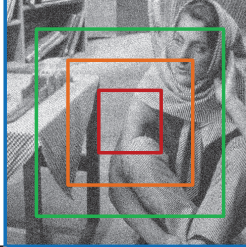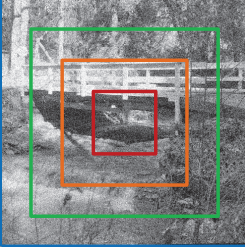The oracle MSE for the images in Fig. 4 are shown in Fig. 6 where for each noise level, the bound
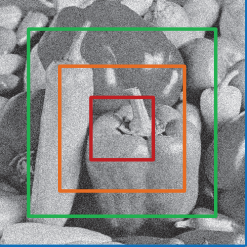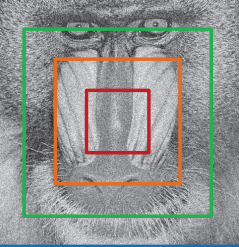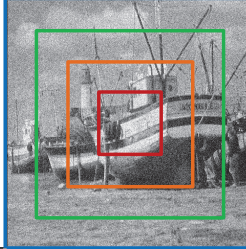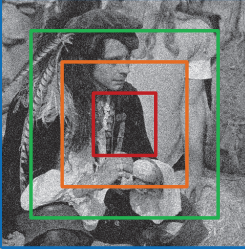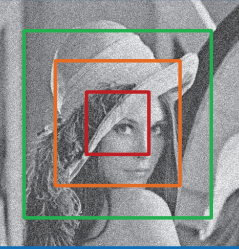
Subimage Size	128×128	256×256	384×384	512×512	128×128	256×256	384×384	512×512	128×128	256×256	384×384	512×512	128×128	256×256	384×384	512×512
$\sigma = 20$	2.96	0.40	0.14	**0.07**	12.76	0.74	0.14	**0.07**	5.30	0.20	0.09	**0.06**	0.56	0.20	0.11	**0.05**
$\sigma = 40$	3.73	0.82	0.39	**0.21**	13.55	1.32	0.38	**0.18**	6.15	0.51	0.26	**0.19**	1.24	0.53	0.29	**0.16**
$\sigma = 60$	4.84	1.34	0.70	**0.40**	17.78	2.13	0.66	**0.32**	7.42	0.91	0.49	**0.36**	2.07	0.95	0.51	**0.30**

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
$\sigma = 20$	12.74	0.45	0.15	**0.07**	17.54	0.73	0.16	**0.08**	4.59	0.21	0.07	**0.04**	5.09	0.30	0.11	**0.02**
$\sigma = 40$	13.69	0.96	0.40	**0.21**	18.44	1.27	0.42	**0.20**	5.26	0.45	0.21	**0.13**	5.73	0.71	0.31	**0.16**
$\sigma = 60$	15.11	1.60	0.73	**0.42**	19.78	1.96	0.78	**0.37**	6.13	0.72	0.39	**0.26**	6.60	1.21	0.57	**0.29**

Fig. 4. Performance of the global denoising scheme for different window sizes. MSE values are averaged over 20 different WGN realizations.

in (35) and the oracle MSE values are computed and then averaged across images given in Fig. 3[2].

Our next experiment in Fig. 7 compares the effect of image size on BM3D [4] (which is close to optimal among patch-based denoisers) and on the proposed method[3]. The selected *Building* image has patterns that are both locally and globally repetitive, so it fits both BM3D and our scheme to achieve the best results. As can be seen in Fig. 7, our oracle method has a very large performance advantage over

[2]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 Fig. 6 capture the decay rate for the tested images as hypothesized. See Appendix A.

[3]BM3D is a two-stage image denoising scheme in which the first stage, as a pre-filter, provides a "pilot" estimate of the noise free image. The second stage uses the pre-filtered 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 output of the first stage is assumed to be the clean image.
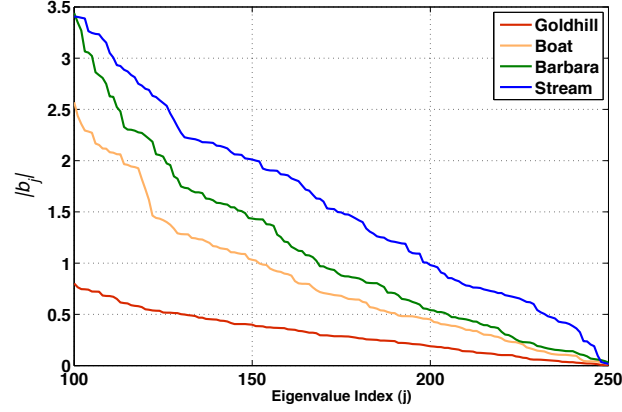
Fig. 5. Projected image coefficients onto the eigenvectors for the images of size 512×512. The coefficients correspond to the higher eigenvector indices are shown to depict the decay rate of each tail.
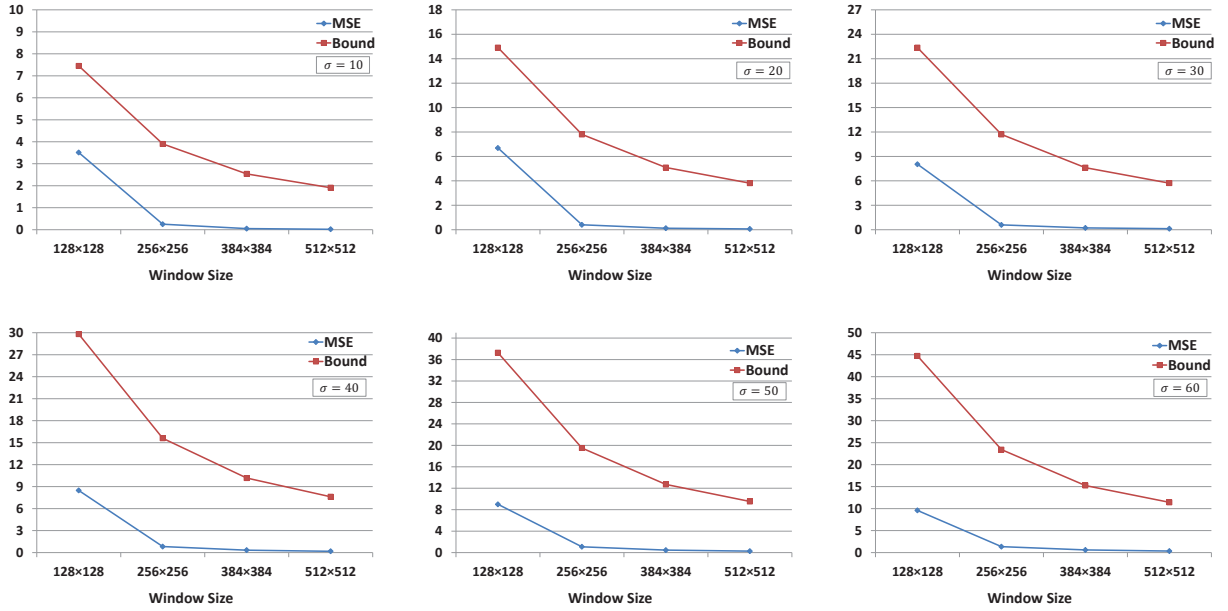
Fig. 6. Averaged MSE of denoising images given in Fig. 3 for different noise levels. The estimated bound given in (35) is averaged for all the images.

oracle BM3D even for small window sizes, and as the window size increases, the advantage grows. Also as predicted, BM3D results are not showing any regular improvement by enlarging the image size.

The oracle performance of NLM [9], BM3D and our approach our compared in Table I. As the noise level increases, the gap between our oracle scheme and oracle BM3D monotonically grows.

Fig. 7. Top left: Building image. Bottom left: Building image corrupted by WGN of standard deviation $\sigma = 40$. Right: Averaged oracle PSNR of the denoised window size for BM3D [4], and our method.
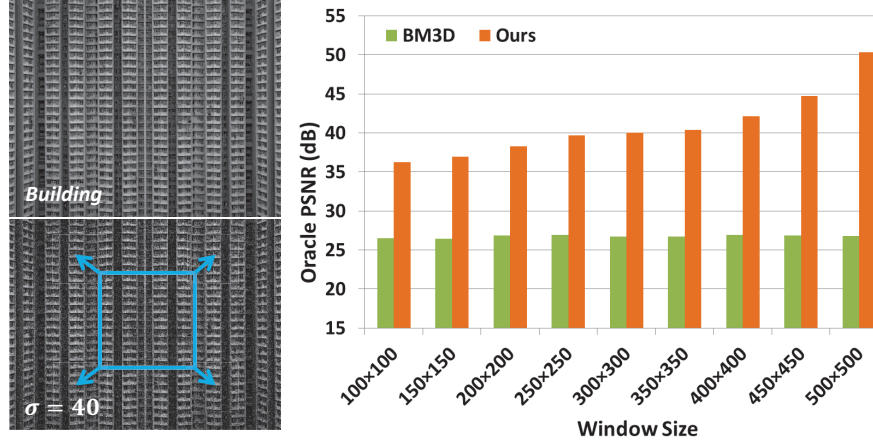
V. CONCLUSION

We emphasize that the oracle results do not correspond to practical denoising algorithms yet, and 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 like NLM and 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

As shown in [3], it is possible to efficiently compute and use only the first m leading eigenvectors of the filter matrix (this is facilitated in practice by using the Nyström extension [21].) This approximation is employed to illustrate the validity of our analysis in a somewhat more practical scenario.

Keeping m leading eigenvectors of the filter $\mathbf{W}$, the truncated filter can be expressed as:

$$\widehat{\mathbf{W}} = \mathbf{V}_m \mathbf{S}_m \mathbf{V}_m^T, \tag{25}$$

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:

$$\widehat{\mathbf{w}}_i^T = \sum_{j=1}^{m} \lambda_j \mathbf{v}_j(i) \mathbf{v}_j^T, \tag{26}$$

TABLE I

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		
	NLM	BM3D	Ours	NLM	BM3D	Ours	NLM	BM3D	Ours	NLM	BM3D	Ours
10	29.36	11.34	**0.03**	62.95	26.93	**0.03**	21.87	11.41	**0.02**	70.97	31.32	**0.02**
20	42.15	22.98	**0.07**	120.77	63.42	**0.07**	33.43	20.81	**0.06**	166.23	76.30	**0.06**
30	61.91	34.11	**0.14**	171.91	98.39	**0.12**	45.18	28.53	**0.12**	196.61	121.54	**0.10**
40	91.74	45.14	**0.21**	233.03	130.44	**0.18**	58.84	35.83	**0.19**	275.49	164.95	**0.16**
50	126.43	59.80	**0.3**	285.12	158.55	**0.25**	75.23	48.33	**0.27**	391.81	188.16	**0.22**
60	163.33	71.26	**0.4**	329.89	184.09	**0.32**	93.52	56.87	**0.36**	431.36	220.80	**0.3**
σ	Boat			Man			Goldhill			Lena		
	NLM	BM3D	Ours	NLM	BM3D	Ours	NLM	BM3D	Ours	NLM	BM3D	Ours
10	31.87	14.30	**0.02**	36.25	16.26	**0.03**	31.16	14.93	**0.01**	19.77	9.71	**0.02**
20	56.96	29.42	**0.07**	67.85	35.59	**0.07**	53.47	29.72	**0.04**	31.07	18.34	**0.05**
30	80.44	43.16	**0.13**	94.89	53.72	**0.13**	78.67	42.06	**0.08**	44.68	26.28	**0.1**
40	108.72	56.16	**0.21**	125.95	70.61	**0.2**	105.51	52.91	**0.13**	60.46	34.13	**0.16**
50	138.42	75.05	**0.31**	158.01	93.27	**0.28**	131.17	69.78	**0.19**	76.56	45.22	**0.22**
60	168.69	88.32	**0.42**	188.98	109.14	**0.37**	155.07	80.54	**0.26**	92.94	53.70	**0.29**

In a fashion similar to the full-space filter $\mathbf{W}$, the truncated filter's MSE for the i-the pixel is:

$$\text{MSE}_i^{(m)} = z_i^2 + \sum_{j=1}^{m}(\lambda_j^2(b_j^2 + \sigma^2)\mathbf{v}_j(i)^2 - 2z_i\lambda_j\mathbf{v}_j(i)b_j) \tag{27}$$

The total MSE for the whole image is:

$$\text{MSE}^{(m)} = \frac{1}{n}\sum_{i=1}^{n}\text{MSE}_i^{(m)} = \frac{1}{n}\sum_{i=1}^{n}z_i^2 + \frac{1}{n}\sum_{j=1}^{m}\left((\lambda_j^2 - 2\lambda_j)b_j^2 + \sigma^2\lambda_j^2\right) \tag{28}$$

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 as:

$$\text{MMSE}^{(m)} = \frac{1}{n}\sum_{i=1}^{n}z_i^2 - \frac{1}{n}\sum_{j=1}^{m}b_j^2\lambda_j^* \tag{29}$$

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:

$$\begin{aligned}
\text{MMSE}^{(m)} &= \frac{1}{n}\sum_{j=1}^{n} b_j^2 - \frac{1}{n}\sum_{j=1}^{m} b_j^2 \lambda_j^* \\
&= \frac{1}{n}\sum_{j=1}^{n} b_j^2 - \frac{1}{n}\sum_{j=1}^{n} b_j^2 \lambda_j^* + \frac{1}{n}\sum_{j=m+1}^{n} b_j^2 \lambda_j^* \\
&= \underbrace{\frac{1}{n}\sum_{j=1}^{n} \frac{\sigma^2 b_j^2}{\sigma^2 + b_j^2}}_{\text{MMSE}} + \underbrace{\frac{1}{n}\sum_{j=m+1}^{n} \frac{b_j^4}{b_j^2 + \sigma^2}}_{\Delta\text{MSE}^{(m)}}
\end{aligned} \tag{30}$$

where the first term is the same as the minimum MSE given in (17) 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 $\Delta\text{MSE}^{(m)}$:

$$\Delta\text{MSE}^{(m)} = \frac{1}{n}\sum_{j=m+1}^{n} \frac{b_j^4}{b_j^2 + \sigma^2} \leq \frac{1}{\sigma^2 n}\sum_{j=m+1}^{n} b_j^4 = \frac{1}{\sigma^2 n}(\|\mathbf{b}\|_4^4 - \|\mathbf{b}_m\|_4^4) \tag{31}$$

where [4] $\|\mathbf{b}_m\|_4^4 = \sum_{j=1}^{m} b_j^4$. Again, assuming a decay rate of $\alpha > 0$ for the coefficients $|b(t)| = c/t^\alpha$ and using the integral test for convergence [20], we obtain the following lower and upper bound:

$$\frac{1}{\sigma^2 n}\int_{m+1}^{n+1} \frac{c^4}{t^{4\alpha}}\,dt \leq \frac{1}{\sigma^2 n}\sum_{j=m+1}^{n} b_j^4 \leq \frac{1}{\sigma^2 n}\left(\frac{c^4}{(m+1)^{4\alpha}} + \int_{m+1}^{n} \frac{c^4}{t^{4\alpha}}\,dt\right) \tag{32}$$

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

$$\frac{c^4}{\sigma^2}\left(\frac{(n+1)^{1-4\alpha} - (m+1)^{1-4\alpha}}{(1-4\alpha)n}\right) \leq \frac{1}{\sigma^2 n}\sum_{j=m+1}^{n} \frac{c^4}{j^{4\alpha}} \leq \frac{c^4}{\sigma^2}\left(\frac{1}{n(m+1)^{4\alpha}} + \left(\frac{n^{1-4\alpha} - (m+1)^{1-4\alpha}}{(1-4\alpha)n}\right)\right) \tag{33}$$

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

$$\frac{c^4}{\sigma^2}\left(\frac{\ln(\frac{n+1}{m+1})}{n}\right) \leq \frac{1}{\sigma^2 n}\sum_{j=m+1}^{n} \frac{c^4}{j^4} \leq \frac{c^4}{\sigma^2}\left(\frac{1}{n(m+1)} + \frac{\ln(\frac{n}{m+1})}{n}\right) \tag{34}$$

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:

$$\text{MMSE}^{(m)} \leq \frac{\sigma}{2n}\|\mathbf{b}\|_1 + \frac{1}{\sigma^2 n}(\|\mathbf{b}\|_4^4 - \|\mathbf{b}_m\|_4^4) \tag{35}$$

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.

---

[4]In practice, for $j > m$ we have $b_j^2 \ll \sigma^2$ (see Fig. 5). This means that $\frac{b_j^4}{b_j^2 + \sigma^2} \leq \frac{b_j^4}{\sigma^2} \leq b_j^2$.

REFERENCES

[1] I. Ram, M. Elad, and I. Cohen, "Image processing using smooth ordering of its patches," *to appear in IEEE Trans. Image Proc.* 1

[2] M. Zontak, I. Mosseri, and M. Irani, "Separating signal from noise using patch recurrence across scales," *IEEE Conf. on Computer Vision and Pattern Recognition*, 2013. 1

[3] H. Talebi and P. Milanfar, "Global image denoising," *IEEE Trans. on Image Proc.*, vol. 23, no. 2, pp. 755–768, February 2014. 1, 2, 5, 11

[4] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. on Image Proc.*, vol. 16, no. 8, pp. 2080–2095, August 2007. 1, 2, 4, 9, 11, 12

[5] P. Chatterjee and P. Milanfar, "Practical bounds on image denoising: From estimation to information," *IEEE Trans. on Image Proc.*, vol. 20, no. 5, pp. 1221–1233, May 2011. 1

[6] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948. 2

[7] E. Ordentlich, G. Seroussi, S. Verdu, M. Weinberger, and T. Weissman, "A discrete universal denoiser and its application to binary images," *Proceedings of ICIP*, 2003. 2

[8] T. Weissman, E. Ordentlich, G. Seroussi, S. Verdu, and M. Weinberger, "Universal discrete denoising: Known channel," *IEEE Trans. Info. Theory*, vol. 51, no. 1, pp. 5–28, 2005. 2

[9] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling and Simulation (SIAM interdisciplinary journal)*, vol. 4, no. 2, pp. 490–530, 2005. 2, 5, 10, 12

[10] P. Chatterjee and P. Milanfar, "Clustering-based denoising with locally learned dictionaries," *IEEE Trans. on Image Proc.*, vol. 18, no. 7, pp. 1438–1451, July 2009. 2, 4

[11] P. Milanfar, "A tour of modern image filtering," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 106–128, 2013. 3, 4, 5

[12] ——, "Symmetrizing smoothing filters," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 263–284, 2013. 4

[13] J. Portilla, V. Strela, M. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. on Image Proc.*, vol. 12, no. 11, pp. 1338–1351, November 2003. 4

[14] D. D. Muresan and T. W. Parks, "Adaptive principal components and image denoising," *Proceedings of ICIP*, vol. 1, pp. 101–104, Sep. 2003. 4

[15] L. Zhang, W. Dong, D. Zhang, and G. Shi, "Two-stage image denoising by principal component analysis with local pixel grouping," *Pattern Recognition*, vol. 43, pp. 1531–1549, Apr. 2010. 4

[16] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. on Image Proc.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006. 4

[17] P. Chatterjee and P. Milanfar, "Is denoising dead?" *IEEE Trans. on Image Proc.*, vol. 19, no. 4, pp. 895–911, April 2010. 4

[18] H. Talebi and P. Milanfar, "Global denosing is asymptotically optimal," *submitted to ICIP*, 2014. 5

[19] J. M. Steele, *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities*. Cambridge University Press, 2004. 7

[20] K. Knopp, *Infinite Sequences and Series*. Dover Publication, 1956. 7, 13

[21] E. Nyström, "Über die praktische auflösung von linearn ingtegraglechungen mit anwendungen auf randwertaufgaben der potentialtheorie," *commentationes Physico-Mathematicae*, vol. 4, no. 15, pp. 1–52, April 1928. 11