GLOBAL DENOISING IS ASYMPTOTICALLY OPTIMAL

Hossein Talebi and Peyman Milanfar

Department of Electrical Engineering
University of California, Santa Cruz
Email : {htalebi, milanfar}@soe.ucsc.edu

ABSTRACT

In this paper an upper bound on the decay rate of the mean-squared error for global image denoising is derived. As image size increases, this upper bound decays to zero; that is, the global denoising is asymptotically optimal. Unlike patch-based methods such as BM3D, this property only holds for global denoising schemes. In practice, and as demonstrated in this work, this performance gap between patch-based and global denoisers can grow rapidly with image size.

Index Terms— Global Denoising, Denoising Bound

1. INTRODUCTION

We begin with the model of additive noise problem:

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

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 $\mathbf{y}$. The following filtering scheme summarizes the existing restoration methods [1]:

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

where the $n \times n$ matrix $\mathbf{W}$ denotes the employed filter to obtain the estimated image $\widehat{\mathbf{z}}$. We highlight that even if the pixels are estimated locally, the effect of overlapped patches and the corresponding aggregations can be reflected in this global filtering scheme. The matrix $\mathbf{W}$ can be decomposed using the eigen-decomposition of the filter and the aggregation matrix $\mathbf{A}$:

$$\mathbf{W} = \mathbf{A}\mathbf{V}\mathbf{S}\mathbf{V}^{-1} \tag{3}$$

where $\mathbf{S} = \text{diag}[\lambda_1, ..., \lambda_n]$ represents the shrinkage factors and columns of the matrix $\mathbf{V} = [\mathbf{v}_1, ..., \mathbf{v}_n]$ contain the eigenvectors.

A vast number of denoising algorithms have been proposed to find the optimal basis, shrinkage and aggregation strategy. Typically, most of these methods use a set of locally sparse basis functions to effectively separate the noise from latent patches. Training based dictionaries [2, 3], fixed basis functions such as wavelet and DCT [4, 5] and data adapted functions obtained from principle component analysis (PCA) [6, 7] are a few examples of the commonly used bases for the purpose of denoising. Although the basis functions vary for different methods, the optimal shrinkage strategy has been shown to be the Wiener criterion [8].

Sparse image representations are facilitated by image redundancies. Henceforth, a good denoiser should demonstrate performance improvement as the number of samples increases. This fact was evidenced for the case of restoring binary images from context in [9, 10]. More recent, the Non-Local Means (NLM) method [11] was inspired in part by [9,10] and itself gave a proof of the asymptotic consistency of the NLM method.

Recently we proposed a truly global scheme [12] where *all* the existing pixels in the image were considered in developing a denosing filter, and the computational complexity was also notably lowered by using a subsampling strategy based on the Nyström extension. This current work is motivated by our observation in [12] that unlike the patch-based methods, performance of the global approach was consistently improved with image size. In the patch-based scenario, as the image size grows, the performance improvement is limited by the lower likelihood of finding closely similar patches (see Fig. 1). Increasing the size of the patches reduces the number of available similar patches (see Fig. 2), but is not enough to force the error to zero asymptotically. Consequently, performance stays almost constant with increasing image size (see Fig. 6).

To alleviate the limitations imposed by the patch-based denoisers, we can consider an alternative scenario where first, the patch matching is replaced by matching similar *pixels* and second, *all* the pixels in the image are contributing in denoising every single pixel (Equivalently an identity aggregation matrix $\mathbf{A} = \mathbf{I}$ and a global rather than local basis in (3)). In the current work, we show that the denoising performance of the global filter *always* improves as a function of the image size. The rate of this improvement is a function of the sparsity of the image in a naturally constructed adapted basis. More specifically, for an image of size n pixels, our oracle upper bound on the mean-squared-error of the estimated image has a decay rate of *at least* n^{-1}.

The paper is organized as follows. Section 2 describes our statistical analysis of the global filter and behavior of its MSE

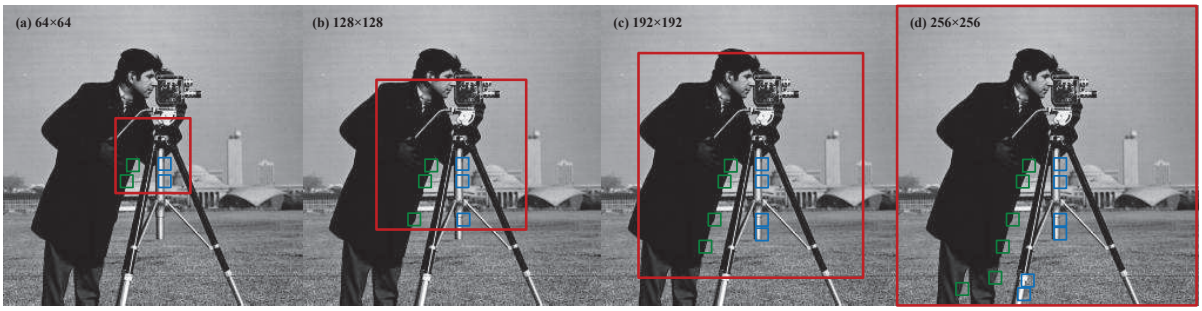


Fig. 1. Image size effect on patch matching. As the window size increases, patches with less similarities are matched.

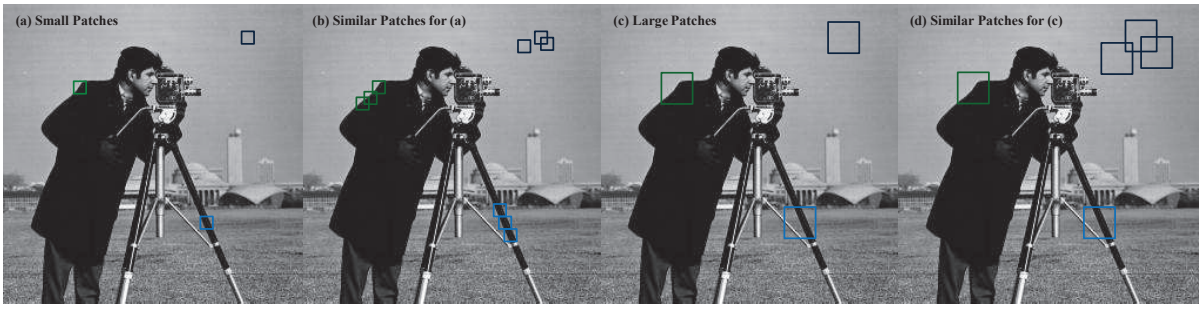


Fig. 2. Patch size effect on patch matching. As the patch size grows, fewer similar patches are available.

function. Section 3 represents the proposed MSE bound and Section 4 provides some experiments to evaluate performance of the global denosier and its bound. Finally we conclude this paper in Section 5.

2. MINIMUM MSE ESTIMATION

In general, the filter matrix $\mathbf{W}$ is defined by affinity weights between pixels, or patches around pixels [1]. While any valid kernel [1] could be employed, here for simplicity we use the NLM kernel weights [11] to measure the similarity between the samples y_i and y_j as:

$$K_{ij} = \exp\left\{\frac{-\|\mathbf{y}_i - \mathbf{y}_j\|^2}{h^2}\right\}, \tag{4}$$

where $\mathbf{y}_i$ and $\mathbf{y}_j$ are patches centered at y_i and y_j, respectively. Then, the denoiser performing normalized weighted averaging using this kernel is:

$$\widehat{\mathbf{z}} = \begin{bmatrix} \mathbf{w}_1^T \\ \mathbf{w}_2^T \\ \vdots \\ \mathbf{w}_n^T \end{bmatrix} \mathbf{y} = \mathbf{W}\mathbf{y}, \tag{5}$$

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

$$\mathbf{w}_i = \frac{1}{\sum_{j=1}^n K_{ij}}[K_{i1}, K_{i2}, \ldots, K_{in}]^T. \tag{6}$$

The filter matrix $\mathbf{W}$ can be closely approximated with a positive, doubly-stochastic and symmetric matrix [1, 13], and its eigen-decomposition can be expressed as:

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

in which $\mathbf{S} = \text{diag}[\lambda_1, ..., \lambda_n]$ denotes the shrinkage factors in decreasing order $0 \leq \lambda_n \leq ... < \lambda_1 = 1$ and the orthonormal basis is the columns of the matrix $\mathbf{V} = [\mathbf{v}_1, ..., \mathbf{v}_n]$. Since the filter $\mathbf{W}$ is affected by the noise in the given image, in practice the filter is never computed directly from the input noisy image. To have the filter stochastically decoupled from noise, the kernel weights are typically computed from a "pre-filtered" version of $\mathbf{y}$. In the current work, for the purpose of *oracle* performance evaluation, the filter is directly computed from the clean latent image $\mathbf{z}$ and as a result is deterministic.

As was shown in [1], the squared bias of the estimate given in (2) is:

$$\begin{aligned} \|\text{bias}(\widehat{\mathbf{z}})\|^2 &= \|\mathbf{z} - \text{E}(\widehat{\mathbf{z}})\|^2 \\ &= \|(\mathbf{I} - \mathbf{V}\mathbf{S}\mathbf{V}^T)\mathbf{z}\|^2 = \sum_{j=1}^n (1-\lambda_j)^2 b_j^2 \end{aligned} \tag{8}$$

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:

$$\text{var}(\widehat{\mathbf{z}}) = \text{tr}\left(\text{cov}(\widehat{\mathbf{z}})\right) = \sigma^2 \sum_{j=1}^n \lambda_j^2 \tag{9}$$

Therefore, the (oracle) MSE of the estimated image is:

$$\begin{aligned} \text{MSE} &= \frac{1}{n}\left(\|\text{bias}(\widehat{\mathbf{z}})\|^2 + \text{var}(\widehat{\mathbf{z}})\right) \\ &= \frac{1}{n}\sum_{j=1}^n (1-\lambda_j)^2 b_j^2 + \sigma^2 \lambda_j^2 \end{aligned} \tag{10}$$

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

$$\frac{\partial \text{MSE}(\lambda)}{\partial \lambda} = 0 \implies \lambda_j^* = \frac{1}{1 + \text{snr}_j^{-1}}, \tag{11}$$

where, somewhat unsurprisingly, the "optimal" eigenvalues $\{\lambda_j^*\}$ are the Wiener coefficients with $\text{snr}_j = \frac{b_j^2}{\sigma^2}$. This shrinkage strategy leads to the minimum value of the MSE[1]

$$\text{MSE}(\lambda^*) = \frac{\sigma^2}{n} \sum_{j=1}^{n} \lambda_j^* \tag{12}$$

3. ORACLE MSE BOUND

The minimum MSE given by (12) can be written as:

$$\text{MSE}(\lambda^*) = \frac{\sigma^2}{n} \sum_{j=1}^{n} \lambda_j^* = \frac{1}{n} \sum_{j=1}^{n} \frac{\sigma^2 b_j^2}{\sigma^2 + b_j^2} \tag{13}$$

This equality can be expressed as:

$$\text{MSE}(\lambda^*) = \frac{1}{2n} \sum_{j=1}^{n} \frac{\sigma |b_j|}{\frac{\sigma^2 + b_j^2}{2}} \sigma |b_j| \tag{14}$$

Using the Arithmetic-Geometric Means inequality [14] we have $\sigma |b_j| \leq \frac{\sigma^2 + b_j^2}{2}$, then:

$$\frac{1}{2n} \sum_{j=1}^{n} \frac{\sigma |b_j|}{\frac{\sigma^2 + b_j^2}{2}} \sigma |b_j| \leq \frac{1}{2n} \sum_{j=1}^{n} \sigma |b_j| \tag{15}$$

And this in turn means

$$\text{MSE}(\lambda^*) \leq \frac{\sigma}{2n} \|\mathbf{b}\|_1 \tag{16}$$

The l_1 norm of the projection coefficients $\mathbf{b}$ gives a bound on the oracle MSE. This indicates that the more sparse the signal is in the basis, the smaller the MSE error will be. Moreover, 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. In the worst case scenario wherein $|b_j| = c$ where c is constant (This essentially corresponds to the signal being "white noise" in the basis defined by the kernel), the MSE will be upper bounded by a constant $\frac{c\sigma}{2}$. In general, however, we expect the coefficients to decay at some $\alpha > 0$ rate where $|b_j| = \frac{c}{j^\alpha}$. This means that

$$\frac{\sigma}{2n} \sum_{j=1}^{n} |b_j| = \frac{\sigma}{2n} \sum_{j=1}^{n} \frac{c}{j^\alpha} \tag{17}$$

[1]Technically, λ_1^* have to be fixed at 1, but this would result in a very small increase in the minimum MSE, on the order of $1/n$. For sufficiently large n, this difference is negligible.

For all $\alpha > 0$, $\text{MSE}(\lambda^*)$ will tend to zero as $n \to \infty$. This describes the most general case of MSE convergence. On a more detailed analysis of the convergence rate, it is useful to consider the coefficients b_j as samples of the continuous function $|b(t)| = c/t^\alpha$ as $b_j = b(j)$.

The following lower and upper bounds are obtained using the integral test for convergence (Maclaurlin-Cauchy test) [15]:

$$\frac{1}{n} \int_1^{n+1} \frac{c}{t^\alpha}\, dt \leq \frac{1}{n} \sum_{j=1}^{n} |b_j| \leq \frac{1}{n} (c + \int_1^n \frac{c}{t^\alpha}\, dt) \tag{18}$$

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

$$c(\frac{(n+1)^{1-\alpha} - 1}{(1-\alpha)n}) \leq \frac{1}{n} \sum_{j=1}^{n} \frac{c}{j^\alpha} \leq c(\frac{n^{1-\alpha} - \alpha}{(1-\alpha)n}) \tag{19}$$

which is corresponding to a convergence rate of $\frac{1}{n^\alpha}$.

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

$$\frac{c \ln(n+1)}{n} \leq \frac{1}{n} \sum_{j=1}^{n} \frac{c}{j} \leq \frac{c(1 + \ln(n))}{n} \tag{20}$$

which means a rate of $\ln(n)/n$. Finally, the decay rate is at least $\frac{1}{n}$ for $\alpha > 1$ since the summation in (17) 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. These results are evaluated with some experiments in the following.

4. EXPERIMENTS

Some benchmark images used in our denoising evaluations 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 (256×256) are very small and some of them are even below round off error. It is worth mentioning that for practical purposes, our experiments are carried out using a truncated filter with only a small number of the leading eigenvectors of $\mathbf{W}$ [12].

The average MSE of the images in Fig. 4 is shown in Fig. 5 where the MSE and the bound given in (16) are plotted for different window sizes. Clearly the estimated bound captures the decay rate of the averaged MSEs.

(a) Parrot (b) House (c) Monarch (d) Cameraman

Fig. 3. Some benchmark images used in our experiments.

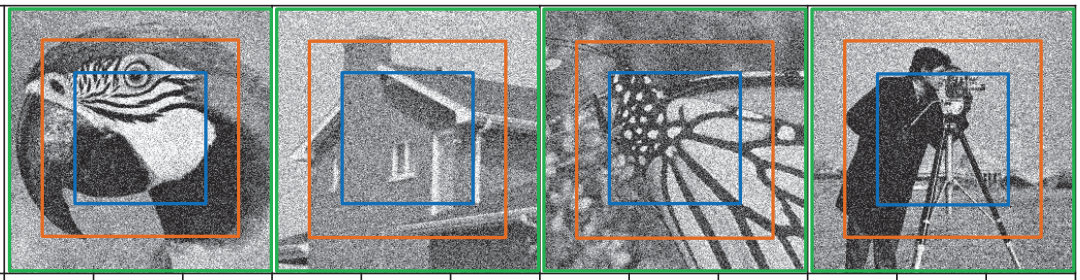
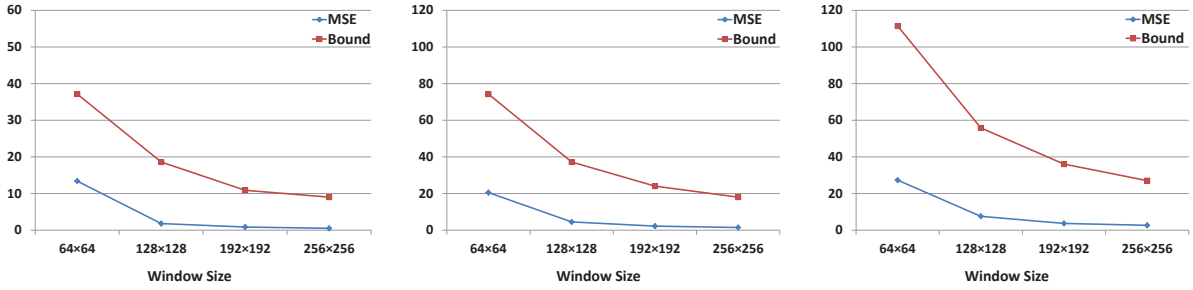
Selected Window												
Window Size	128×128	192×192	256×256	128×128	192×192	256×256	128×128	192×192	256×256	128×128	192×192	256×256
σ=20	2.68	1.06	**0.58**	0.74	0.45	**0.27**	1.73	0.72	**0.45**	2.12	1.07	**0.63**
σ=40	6.78	2.86	**1.62**	2.04	1.31	**0.76**	4.34	1.83	**1.19**	5.54	3.04	**1.74**
σ=60	11.62	5.02	**2.98**	3.63	2.37	**1.38**	7.19	3.10	**2.08**	9.35	5.42	**3.07**

Fig. 4. Oracle performance of the global denoising scheme for different window sizes. MSE values are averaged over 20 different white Gaussian noise realizations.



(a) $\sigma = 20$ (b) $\sigma = 40$ (c) $\sigma = 60$

Fig. 5. The estimated bound is evaluated for three noise levels and different window sizes. For each noise level, the bound in (16) and the oracle MSE values are computed and then averaged across images given in Fig. 3.

Image size effect on the BM3D [5] and the proposed method are compared in Fig. 6. The selected *Building* image has locally and globally repetitive patterns which fit both BM3D and our scheme to achieve the best results. As can be seen in Fig. 6, our method has a very large performance advantage over 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 consistent improvement by enlarging the image size[2].

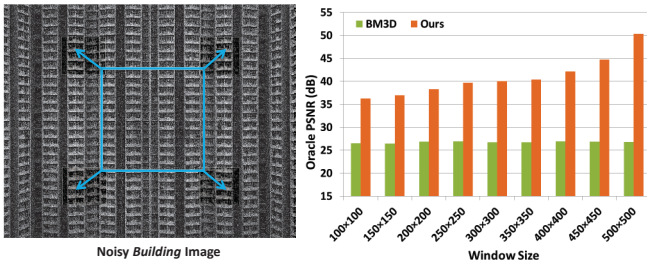


Fig. 6. Left: Building image corrupted by white Gaussian noise of standard deviation $\sigma = 40$. Right: Averaged oracle PSNR of the denoised window size for BM3D [5], and our method.

5. CONCLUSION

Although the oracle results do not correspond to practical denoising algorithms, interestingly, the existing patch-based bounds are improved upon by the asymptotic behavior of the global filtering. Unlike oracle versions of algorithms like BM3D, the oracle global filter is able to almost perfectly reconstruct the input image. This distinguishes global filtering as an interesting area of future work to further push the performance bound of practical algorithms.

[2]BM3D consists of two similar denoising stages 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. Consequently, feeding the clean image to the second stage can lead to the oracle BM3D.

6. REFERENCES

[1] P. Milanfar, "A tour of modern image filtering," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 106–128, 2013. 1, 2

[2] 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. 1

[3] 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. 1

[4] 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. 1

[5] 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, 4

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

[7] 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. 1

[8] P. Chatterjee and P. Milanfar, "Is denoising dead?," *IEEE Trans. on Image Proc.*, vol. 19, no. 4, pp. 895–911, April 2010. 1

[9] 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. 1

[10] 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. 1

[11] 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. 1, 2

[12] H. Talebi and P. Milanfar, "Global image denoising," *IEEE Trans. on Image Proc.*, vol. 23, no. 2, pp. 755–768, February 2014. 1, 4

[13] P. Milanfar, "Symmetrizing smoothing filters," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 263–284, 2013. 2

[14] J. M. Steele, *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities*, Cambridge University Press, 2004. 3

[15] K. Knopp, *Infinite Sequences and Series*, Dover Publication, 1956. 3