FUNDAMENTAL LIMITS OF IMAGE DENOISING: ARE WE THERE YET?

Priyam Chatterjee and Peyman Milanfar

Dept. of Electrical Engineering
University of California, Santa Cruz, CA 95064, USA

ABSTRACT

In this paper, we study the fundamental performance limits of image denoising where the aim is to recover the original image from its noisy observation. Our study is based on a general class of estimators whose bias can be modeled to be affine. A bound on the performance in terms of mean squared error (MSE) of the recovered image is derived in a Bayesian framework. In this work, we assume that the original image is available, from which we learn the image statistics. Performances of some current state-of-the-art methods are compared to our MSE bounds for some commonly used experimental images. These show that some gain in denoising performance is yet to be achieved.

***Index Terms*—** Image denoising, estimation, Bayesian Cramér-Rao lower bound, mean squared error.

1. INTRODUCTION

With the advent of affordable hardware, it is now possible for anyone to capture images of considerably high resolution with devices like cell phones, webcams, digital cameras, etc. However, the observed images are often noisy and further enhancement is required. Such post-processing is usually carried out by denoising algorithms where the objective is to infer the actual image from its degraded observation. Most algorithms perform such a task by decomposing the image into small blocks or *patches* and estimating the image patches $\mathbf{z}_i$ from the noisy observations

$$\mathbf{y}_i = \mathbf{z}_i + \boldsymbol{\eta}_i, \tag{1}$$

where $\boldsymbol{\eta}_i$ is the noise that arises in the image acquisition process. This estimation problem is highly ill-posed and most denoising algorithms use some sort of prior information about the original image and the noise to converge to an acceptable recovered image. In this work, we assume that the noise is zero mean independent identically distributed (IID) Gaussian with patch-wise covariance $\sigma^2\mathbf{I}$ (although, our framework can be used for other noise distributions just as well.)

Here, we are interested in studying the fundamental performance limits for this estimation problem. Studying such performance bounds will allow us to understand the difficulties involved in image denoising, and to compare the performance of the current state-of-the-art algorithms to the theoretical limits. More importantly, it will allow us to ascertain if further improvements are possible and even identify ways of achieving the same. Recently, Treibitz *et al.* [1] studied the recovery limits for particular objects or regions in an image under pointwise degradation. Voloshynovskiy *et al.* [2] briefly analyzed the performance of MAP estimators for the denoising problem. However, our bounds are developed in a much more general setting and to the best of our knowledge, no such study exists for a general image denoising framework.

In the next section, we develop the bounds in a Bayesian Cramér-Rao lower bound (B-CRLB) framework assuming that the image patches are geometrically similar, although, they may vary in patch intensity (Fig. 1). Later in that section, we extend our bounds formulation for general images using a clustering mechanism based on geometric similarity of patches [3]. In Sec. 3, the formulation is experimentally validated and compared to the current state-of-the-art denoising methods [3–6]. Finally, in Sec. 4, we conclude with a few words on further research on this topic.

2. MSE BOUND

In this section we develop an expression for the bound on the MSE for biased estimators, since such estimators can yield a lower MSE than an unbiased one [7]. Moreover, unbiased estimators for an inverse problem like denoising will invariably lead to an unpleasant denoised image with unacceptably large variance. We model the bias function $\mathbf{b}(\mathbf{z})$ using a first order approximation of the Taylor expansion around $\mathbf{z}_i = \mathbf{0}$:

$$\mathbf{b}(\mathbf{z}) = \mathbf{b}(\mathbf{z}_i) + \{\nabla\mathbf{b}(\mathbf{z}_i)\}^T(\mathbf{z} - \mathbf{z}_i) + \textit{higher order terms}$$
$$\Rightarrow \mathbf{b}(\mathbf{z}) \approx \{\nabla\mathbf{b}(\mathbf{0})\}^T\mathbf{z} + \mathbf{b}(\mathbf{0}) = \mathbf{M}\mathbf{z} + \mathbf{u} \tag{2}$$

where ∇ denotes the gradient operator, $\mathbf{M} = \nabla\mathbf{b}(\mathbf{0})$ is the (symmetric) gradient matrix of the bias and $\mathbf{u}$ is the bias when $\mathbf{z}_i = \mathbf{0}$. Thus, we model the bias function to be *locally*[1] affine

This work was supported in part by the U.S. Air Force under Grant F49620-03-1-0387.

[1]By *locality* here, we mean within a group of geometrically similar patches. The actual patch intensities, though, may be quite different. Later in this section, we extend our bounds formulation to general images containing diverse patch geometry. There, the bias for geometrically similar patches can be modeled to be *locally* affine within a cluster, as can be seen in Fig. 1.

with model parameters $\mathbf{M}$ and $\mathbf{u}$. Such a model has been shown to be effective when the noise-free image patches are geometrically similar [8, 9].

An effective performance limit can then be formulated by studying the lower bound on the MSE for such *affine- biased* estimators. To derive this, we first express the conditional MSE for biased estimators as

$$\begin{aligned} E\left[\|\mathbf{z}-\widehat{\mathbf{z}}\|^2 \mid \mathbf{z}\right] &= \mathrm{Tr}\left\{\mathrm{cov}(\widehat{\mathbf{z}}|\mathbf{z})\right\} + \|\mathbf{b}(\mathbf{z})\|^2 \\ &\geq \mathrm{Tr}\left\{(\mathbf{I}+\mathbf{M})\,\mathbf{J}^{-1}\,(\mathbf{I}+\mathbf{M})^T\right\} + \|\mathbf{M}\mathbf{z}+\mathbf{u}\|^2 \quad (3) \end{aligned}$$

where $\mathrm{cov}(\widehat{\mathbf{z}}|\mathbf{z})$ is the conditional covariance of the estimate $\widehat{\mathbf{z}}$, $\mathbf{b}(\mathbf{z})$ is the bias function, $\mathbf{J}$ is the *Fisher information matrix* (FIM), $\mathbf{I}$ is the identity matrix, and $\mathrm{Tr}\{.\}$ denotes the trace of a matrix. The inequality of (3) is obtained using the generalized B-CRLB formulation[2] for the covariance of any given biased estimator [11]. It can be seen that the expression on the right hand side of (3) can be minimized by optimizing the bias function [7, 11, 12]. We obtain the lower bound on the overall Bayesian MSE using the law of total expectation as

$$E\left[\|\mathbf{z}-\widehat{\mathbf{z}}\|^2\right] = \int_{\mathbf{z}} E\left[\|\mathbf{z}-\widehat{\mathbf{z}}\|^2 \mid \mathbf{z}\right] p(\mathbf{z}) d\mathbf{z} \quad (4)$$

$$\geq \int_{\mathbf{z}} \Big[\mathrm{Tr}\left\{(\mathbf{I}+\mathbf{M})\,\mathbf{J}^{-1}\,(\mathbf{I}+\mathbf{M})^T\right\} + (\mathbf{M}\mathbf{z}+\mathbf{u})^T\,(\mathbf{M}\mathbf{z}+\mathbf{u})\Big]\, p(\mathbf{z}) d\mathbf{z} = Q(\mathbf{M},\mathbf{u}). \quad (5)$$

The lower bound on the MSE can now be calculated by finding the optimal bias parameters, namely $\mathbf{M}$ and $\mathbf{u}$, that minimize the expression on the right hand side of (4). Taking derivatives of Q with respect to $\mathbf{M}$ and $\mathbf{u}$ and equating them to zeros, we get expressions for the optimal bias parameters $\mathbf{M}^*$ and $\mathbf{u}^*$ as follows:

$$\mathbf{u}^* = -\mathbf{M}\int_{\mathbf{z}} \mathbf{z}\, p(\mathbf{z})\, d\mathbf{z} = -\mathbf{M}E\left[\mathbf{z}\right] \quad (6)$$

$$\mathbf{M}^* = -\mathbf{J}^{-1}\left[\mathbf{J}^{-1} + \mathrm{cov}(\mathbf{z})\right]^{-1}. \quad (7)$$

Equations 6 & 7 thus give us expressions for the optimal bias parameters that minimize the MSE bound Q. The minimum Q can then be obtained as

$$\begin{aligned} Q_{\min} &= \int_{\mathbf{z}} \mathrm{Tr}\left\{(\mathbf{I}+\mathbf{M}^*)\mathbf{J}^{-1}(\mathbf{I}+\mathbf{M}^*)^T + \mathbf{M}^*\left(\mathbf{z}-E\left[\mathbf{z}\right]\right)\left(\mathbf{z}-E\left[\mathbf{z}\right]\right)^T\mathbf{M}^{*T}\right\} p(\mathbf{z})\, d\mathbf{z} \\ &= \mathrm{Tr}\left\{(\mathbf{I}+\mathbf{M}^*)\mathbf{J}^{-1}(\mathbf{I}+\mathbf{M}^*)^T\right\} + \mathrm{Tr}\left\{\mathbf{M}^*\int_{\mathbf{z}}\left(\mathbf{z}-E\left[\mathbf{z}\right]\right)\left(\mathbf{z}-E\left[\mathbf{z}\right]\right)^T p(\mathbf{z})\, d\mathbf{z}\, \mathbf{M}^{*T}\right\} \\ &= \mathrm{Tr}\left\{\mathbf{M}^*\mathbf{J}^{-1}\mathbf{M}^{*T} + 2\mathbf{M}^*\mathbf{J}^{-1} + \mathbf{J}^{-1} + \mathbf{M}^*\mathrm{cov}(\mathbf{z})\,\mathbf{M}^{*T}\right\} \\ &= \mathrm{Tr}\left\{\mathbf{M}^*\left(\mathbf{J}^{-1}+\mathrm{cov}(\mathbf{z})\right)\mathbf{M}^{*T} + 2\mathbf{M}^*\mathbf{J}^{-1} + \mathbf{J}^{-1}\right\} \\ &= \mathrm{Tr}\left\{\mathbf{J}^{-1} - \mathbf{J}^{-1}\left(\mathbf{J}^{-1}+\mathrm{cov}(\mathbf{z})\right)^{-1}\mathbf{J}^{-1}\right\}. \quad (8) \end{aligned}$$

Eq. 8 can be further simplified using the matrix inversion lemma [13] to express the lower bound on the MSE as

$$E\left[\|\mathbf{z}-\widehat{\mathbf{z}}\|^2\right] \geq \mathrm{Tr}\left[\left(\mathbf{J}+\mathrm{cov}^{-1}(\mathbf{z})\right)^{-1}\right]. \quad (9)$$

Before we can compute an expression for the bound, we need to calculate the FIM $\mathbf{J}$. In many denoising methods, inference on each $\mathbf{z}_i$ vector is drawn from multiple (say N_i) similar observations. For our estimation problem, where the noise patches $\boldsymbol{\eta}_i$ are assumed to be sampled from a Gaussian pdf, we can derive the FIM $\mathbf{J}$ as

$$p(\mathbf{y}|\mathbf{z}) = \frac{1}{(\sqrt{2\pi}\sigma)^{n|\Omega|}}\exp\left\{\sum_j \frac{-\|\mathbf{y}_j-\mathbf{z}_j\|^2}{2\sigma^2}\right\} \quad (10)$$

$$\Rightarrow \mathbf{J} = -E\left[\frac{\partial^2 \ln p(\mathbf{y}|\mathbf{z})}{\partial \mathbf{z}_i\, \partial \mathbf{z}_i^T}\right] = N_i\frac{\mathbf{I}}{\sigma^2} \quad (11)$$

where $|\Omega|$ is the total number of patches[3] in the image, n denotes the number of pixels in each patch, and the FIM takes this form from considering N_i radiometrically similar patches to infer the patch $\mathbf{z}_i$. The N_i value for each patch is estimated from the noise-free image using a nearest neighbor approach. The bound is thus computed for each patch within the cluster. The bound for the entire cluster is then obtained as the average bound of all its member patches.

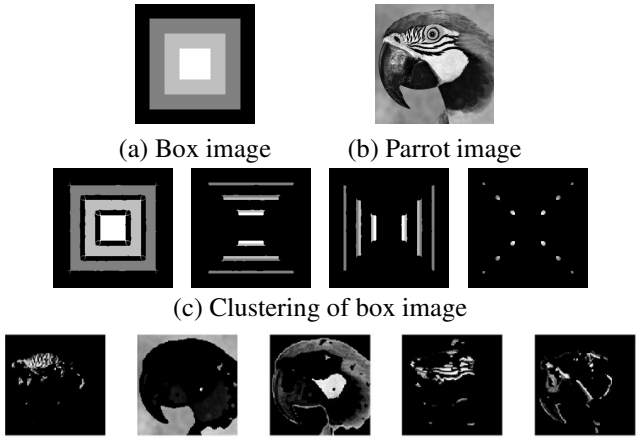
Note that the expression for the lower bound is a function of the noise variance as well as the variability in image patches in each cluster (captured by $\mathrm{cov}(\mathbf{z})$). This is in keeping with the intuition that for a given noise level, images that are mostly smooth are easier to denoise than ones rich in texture and sharp edges. The bound also indicates that when many radiometrically similar patches are present (i.e. large N_i), a better denoising performance can be expected. This is, in fact, exploited by several recent methods [3, 4, 6].

Another interesting implication of (9) is that the right hand side is the Bayesian minimum MSE estimate of $\mathbf{z}$ for Gaussian noise, when the prior pdf $p(\mathbf{z})$ is also assumed to be Gaussian [13]. However, we make no such assumption on $p(\mathbf{z})$. In fact, our formulation does not even require complete knowledge of the prior $p(\mathbf{z})$. Moreover, the bound is an expression for the Linear Minimum MSE that seems to indicate that, in theory, a linear estimator exists that achieves the bound. However, such an estimator needs perfect prior knowledge of the first and second order moments of $p(\mathbf{z})$.

In deriving the MSE bound, we have assumed that the image patches are geometrically similar and, hence, all instances

[2] Although the formulation in (3) is similar in form to the B-CRLB of van Trees [10], it is not the same since the covariance and the FIM here are both based on the conditional pdf $p(\mathbf{y}|\mathbf{z})$ as opposed to the joint pdf in [10].

[3] For the noise patches $\boldsymbol{\eta}_i$ to be independent, the patches in the image need to be non-overlapping. Our FIM (and hence the bounds) are thus obtained for non-overlapping patches. For the overlapping case, $\mathbf{J}$ may be singular. We consider that case to be outside the scope of this paper.



(a) Box image (b) Parrot image

(c) Clustering of box image

(d) Clustering of parrot image

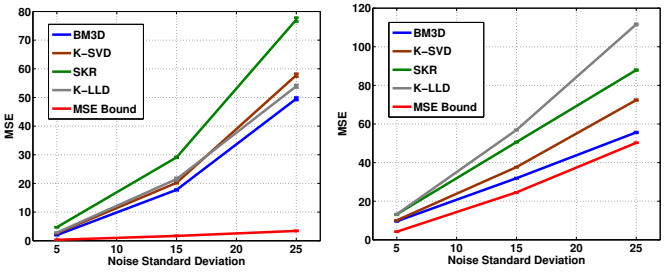
Fig. 1. Clustering based on geometric structure of patches in an image. Note how patches that are mainly smooth or containing edges in a particular direction are clustered together.

of $\mathbf{z}_i$ can be thought to be realizations of a random variable $\mathbf{z}$ sampled from some unknown pdf $p(\mathbf{z})$. Note that, in the formulation of (9), we do not need to learn or assume prior knowledge of the entire pdf, despite the fact that the derivation is done using $p(\mathbf{z})$. Specifically, only the first and second order statistics of $\mathbf{z}$ need to be known or estimated. For the purposes of this study, we assume that the ground truth (noise-free image) is available, and we estimate $\mathrm{cov}(\mathbf{z})$ from the image patches using a bootstrapping method [14].

Most general images consist of flat, edge and texture regions and thus geometric homogeneity is not exhibited among all image patches for such images. To apply our bounds to such diverse images, we first segment the given noise-free image into (not necessarily contiguous) clusters where each cluster consists of patches of similar geometric structure (Fig. 1). For this, we make use of the clustering technique proposed in [3]. Since the image patches are structurally similar in each cluster, an MSE bound can be calculated independently for each cluster. This allows us to study how well each such structurally similar region can be denoised. However, to obtain a bound on the MSE of the entire image, these cluster-wise bounds need to be combined as

$$Q_{\min} = \sum_{k=1}^{K} \frac{|\Omega_k|}{|\Omega|} Q_{\min}^{(k)} \tag{12}$$

where $|\Omega_k|$ is the number of patches in the k-th cluster, K is the number of clusters, and $Q_{\min}^{(k)}$ is the bound on the MSE for the k-th cluster. Although K is a user input, our experiments suggest that using a value of K within the range of 5 to 10 is typically optimal for most general images. Fig. 1(d) shows an example of geometric clustering of a general image where patches containing edges of roughly similar orientation are grouped together, irrespective of the pixel intensities.



(a) Box image (b) Barbara image

Fig. 2. Comparison of some state-of-the-art methods with our bounds formulation with varying levels of additive white Gaussian noise for (a) box image, where $K = 4$; and (b) Barbara image with $K = 5$. The patch size used in each case is 11×11.

3. RESULTS

In this section we experimentally demonstrate the bound by comparing it to some state-of-the-art denoising methods[4] [3–6]. Such methods identify (radiometrically) similar patches in the image to perform denoising. The effect of considering similar patches is captured in the FIM.

We begin our comparison with the simulated box image (Fig. 1(a)) where the number of clusters (K) that the image must be divided into is known for certain. It is clear that the patches in this image can be grouped together based on whether they contain smooth regions, horizontal or vertical edges, or corner regions. The clustering based on such geometric similarity is shown in Fig. 1. Also, it can be seen from Fig. 1(a), several duplicates exist for most patches. We first compute the bounds separately for each cluster using the expression in (9). The bound on the overall MSE is then calculated from Eq. 12. In Fig. 2(a), we show the bounds obtained for different noise standard deviations. It can be seen there that the fundamental limits are much lower than the MSE obtained by the BM3D algorithm [4], which is considered the state-of-the-art technique at present. This is especially true for lower signal to noise ratios (SNR).

For our experiments with general images, we choose $K = 5$. As before, the bounds are calculated independently for each cluster and combined to produce a single lower bound on the overall MSE for denoising. The bounds are compared to the MSE obtained by different recently proposed denoising methods in Fig. 2(b). As opposed to the case of the box image, the bounds can be seen to be only slightly lower than the MSEs obtained by the denoising methods. It can be further seen from the comparisons presented in Table 1 that for most complex images the best performing method in each case produces MSE values that are comparable to the predicted bounds. These seem to imply that images rich in texture and edges are already denoised quite well, especially

[4]Supporting software to compute the bounds, and additional results are available at `http://users.soe.ucsc.edu/~priyam/bounds/`

Table 1. MSE bounds obtained on some frequently used images, ranked according to denoising difficulty as predicted by our bounds for noise standard deviation 25.

Image	**Size**[5]	**BM3D** [4]	**K-SVD** [5]	**SKR** [6]	**K-LLD** [3]	**Bound**
Box	200	49.56	57.78	77.17	53.93	3.42
House	256	33.57	40.05	47.57	42.82	14.82
Peppers	512	42.96	54.07	50.81	48.53	19.21
Lena	512	40.46	48.09	44.09	46.02	19.66
Boat	512	67.17	78.39	78.44	77.45	38.70
Barbara	512	55.62	72.39	87.91	111.58	50.24
Parrot	256	86.98	100.29	98.54	93.64	83.82
Stream	512	158.26	167.29	168.58	157.70	135.46
Mandrill	512	185.60	196.20	195.75	188.84	181.61

[5] Size denotes the number of pixels in each direction.

for higher noise levels.

Another interesting observation that can be made by comparing the MSE bounds of Figures 2(a) & (b) is that the bounds are considerably lower for the structurally simpler simulated box image than those predicted for the Barbara image. This correctly reflects the fact that the box image lacks much of the texture and structural variability that appears in the patches of the Barbara image. Difference in denoising performance is also reflected by the MSEs obtained by the various methods in the two cases. Hence, as an added application, our bounds formulation can be used to compare images for their denoising difficulty. In Table 1, we rank some popular test images based on their denoising difficulty as predicted by our MSE bounds. It can be seen that, in general, our bounds can predict the level of difficulty in denoising any given image in comparison to the other test images without specific reference to a particular denoising algorithm.

Our experimental comparisons over a vast range of images, some of which are presented in this paper, indicate that not much room for improvement exists for the class of much textured images at high noise levels. At lower noise, however, there is still some performance to be gained. On the other hand simpler images can still be better denoised, even for higher noise levels. On the whole, the results illustrate that denoising as a research area is not yet dead.

4. CONCLUSIONS

In this paper, we studied the performance limits of image denoising. Our formulation accounts for the statistics of the image patches and the noise. We showed that the formulation makes intuitive sense. Experimental verification on various images illustrate that denoising is not yet a solved problem.

In our framework, the noise-free image is used for clustering and estimation of intra-cluster variability of the image structure, as well as in searching for patch redundancy. However, these information are not accurately estimated when only the noisy image is available, especially at low SNR. As a result, the predicted bounds may be unattainable. Thus, a more interesting problem will be to consider computing the bounds given only a noisy image to predict how well a given *noisy* image can be denoised. Such information will be useful to camera manufacturers who can use the predicted bounds to determine if further increase in the noise level can be effectively handled by denoising algorithms to generate visually acceptable results. We leave this as a direction of future research.

5. REFERENCES

[1] T. Treibitz and Y. Y. Schechner, "Recovery Limits in Pointwise Degradation," in *Proc. of IEEE Intl. Conf. on Computational Photography*, San Francisco, USA, April 2009.

[2] S. Voloshynovskiy, O. Koval, and T. Pun, "Image Denoising Based on the Edge-Process Model," *Signal Processing*, vol. 85, no. 10, pp. 1950 – 1969, October 2005.

[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.

[4] K. Dabov, A. Foi, V. Katkovnik, and K. O. 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.

[5] 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, December 2006.

[6] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel Regression for Image Processing and Reconstruction," *IEEE Trans. on Image Proc.*, vol. 16, no. 2, pp. 349–66, February 2007.

[7] Y. C. Eldar, "MSE Bound with Affine Bias Dominating the Cramér-Rao Bound," *IEEE Trans. on Signal Proc.*, vol. 56, no. 8, pp. 3824–3836, August 2008.

[8] P. Chatterjee and P. Milanfar, "Bias Modeling for Image Denoising," in *Proc. of 43rd Asilomar Conf. on Signals, Systems and Comp.* Pacific Grove, CA: IEEE Signal Processing Society, November 2009.

[9] ——, "Is Denoising Dead?" *IEEE Trans. on Image Proc.*, to appear.

[10] H. L. van Trees and K. L. Bell, Eds., *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering / Tracking*, 1st ed. Piscataway, NJ: Wiley-IEEE Press, August 2007.

[11] Z. Ben-Haim and Y. C. Eldar, "A Lower Bound on the Bayesian MSE Based on the Optimal Bias Function," *IEEE Trans. on Info. Theory*, vol. 55, no. 11, pp. 5179–5196, November 2009.

[12] T. Y. Young and R. A. Westerberg, "Error Bounds for Stochastic Estimation of Signal Parameters," *IEEE Trans. on Info. Theory*, vol. 17, no. 5, September 1971.

[13] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, ser. Signal Processing. Upper Saddle River, N.J.: Prentice-Hall, Inc., 1993, vol. 1.

[14] B. Efron, "Bootstrap Methods: Another Look at the Jacknife," *The Annals of Statistics*, vol. 7, no. 1, pp. 1–26, 1979.