

# Practical Bounds on Image Denoising: From Estimation to Information

Priyam Chatterjee, *Student Member, IEEE*, and Peyman Milanfar, *Fellow, IEEE*

**Abstract**—Recently, in a previous work, we proposed a way to bound how well any given image can be denoised. The bound was computed directly from the noise-free image that was assumed to be available. In this work, we extend the formulation to the more practical case where no ground truth is available. We show that the parameters of the bounds, namely the cluster covariances and level of redundancy for patches in the image, can be estimated directly from the noise corrupted image. Further, we analyze the bounds formulation to show that these two parameters are interdependent and they, along with the bounds formulation as a whole, have a nice information-theoretic interpretation as well. The results are verified through a variety of well-motivated experiments.

**Index Terms**—Bayesian Cramér-Rao lower bound, image clustering, image denoising, image patch model, mutual information, Rényi entropy, Shannon entropy.

## I. INTRODUCTION

WITH recent advances in imaging technology, image denoising has found renewed interest among both researchers and camera manufacturers. Faster shutter speeds and higher density of image sensors (pixels) result in higher levels of noise in the captured image, which must then be processed by denoising algorithms to yield an image of acceptable quality. This is especially true when images are captured in unfavorable lighting conditions. The goal of such image denoising algorithms is to reduce noise artifacts, at the same time retaining details such as edges and texture in the image. Considerable research has been devoted towards achieving these contradictory goals leading to the varied collection of current state-of-the-art denoising methods [2]–[7]. Despite their differences, the best of these methods perform quite comparably. This led us to study the performance limits of denoising to understand how close the current state-of-the-art is to the fundamental limits for this problem [1]. There we showed that modern denoising methods perform quite close to the fundamental limits for a certain class of textured images, whereas relatively smoother images show room for performance improvement. Even though image denoising is a well-studied problem, not much is known about its statistical performance bounds. Voloshynovskiy *et*

*al.* [8] provided a brief analysis of maximum *a posteriori* (MAP) based denoising methods. Recently, Treibitz *et al.* [9] studied the limits of denoising, among other ill-posed image processing problems, as limits to recovering particular objects or image regions. In both the studies, the images are assumed to be denoised point-wise. That is to say, the data model for the observed image is assumed to be

$$y_i = z_i + \eta_i, \quad i = 1, 2, \dots, M \quad (1)$$

where  $M$  is the number of pixels in the image. Here  $z_i$  is assumed to be the actual pixel intensity which is corrupted by noise  $\eta_i$ . Simplifying assumptions are often made about the noise and/or the underlying image data, the most common of which is to assume the noise to be independent and identically distributed (iid). The denoising performance is then studied as estimation of  $z_i$  given the noisy observations  $y_i$ . Our analysis in [1] studies the performance limits for patch-based methods, since these have recently proven to be quite successful [3]–[7]. For our patch-based bounds formulation we, thus, model the image patches as

$$\mathbf{y}_i = \mathbf{z}_i + \boldsymbol{\eta}_i, \quad i = 1, 2, \dots, M \quad (2)$$

where  $\mathbf{z}_i$  denotes a vectorized form of a group of image pixels in a neighborhood with  $z_i$  at its center, and  $\boldsymbol{\eta}_i$  denotes the corrupting noise patch, with known noise statistics. The denoising bounds in [1], thus, reflect the performance limits of estimating the  $\mathbf{z}_i$  vectors of the image. To the best of our knowledge no other such study exists to date in the literature.

The bounds formulation in [1] showed that the denoising bound is a function of the corrupting noise characteristics (strength and density function) as well as the complexity of the underlying geometric structure of the image patches. Further, the bound is also a function of the amount of redundancy that exists among image patches. In computing the bounds for denoising, we estimated these factors from the underlying noise-free image. As a result, the bounds computation method in [1] cannot be directly applied to the case when only the noisy observation is available. Extending these bounds, as we do in this paper, to the case where only a noisy image is given, has significant practical advantages. Namely, we can predict how well a captured image can be denoised even before any denoising algorithm is applied to it. Such a method will be useful to photographers who can then adjust camera parameters to capture images that will be of acceptable quality, once denoised. In this paper, we present a way of learning the patch-based bounds for noisy images. We treat the problem as that of estimating the parameters of the bounds (namely, the

Manuscript received May 12, 2010; revised September 18, 2010; accepted November 01, 2010. Date of publication November 15, 2010; date of current version April 15, 2011. This work was supported in part by the AFOSR Grant FA9550-07-1-0365 and NSF Grant CCF-1016018. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yongyi Yang.

The authors are with the Department of Electrical Engineering, University of California, Santa Cruz, Santa Cruz, CA 95064 USA (e-mail: priyam@soe.ucsc.edu; milanfar@soe.ucsc.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2010.2092440

patch redundancy and the geometric complexity of the underlying image patches) from the noisy image. Experimentally, we show that these parameters can be estimated quite accurately even from considerably noisy images, allowing us to predict the bounds for such noisy images.

Comparisons of the bounds to the performance of state-of-the-art denoising methods for various images in [1] showed that a class of simple images could be denoised better than the current state-of-the-art by taking further advantage of patch redundancy. On the other hand, images containing complex semi-stochastic textures typically tend to have fewer similar patches. For such images, modern denoising methods seem to perform quite close to the theoretical denoising limits. Intuitively, one can then expect a relationship between existence of similar patches and patch complexity in an image. In the present work, we study the relation between these two parameters and illustrate that the bounds formulation developed in our previous work has an information-theoretic interpretation as well. We show that the patch redundancy and complexity are directly related through the Shannon entropy of the image. In the noisy case, we show that the bounds formulation is related to the mutual information between the noisy and the unknown noise-free image. Using entropy estimation methods, we show that the relationship between the bounds and the information content can be used as an indicator of relative denoising difficulty between images.

In what follows, in the interest of completeness, we first provide a brief overview of the bounds formulation [1] in Section II. We then propose a method of estimating the bounds for any given *noisy* image in Section III. This is followed by an information-theoretic interpretation of the bounds formulation in Section IV. Finally, in Section V, we summarize our findings along with a few remarks on future work.

## II. LOWER BOUND ON THE MSE

In [1], the denoising problem is modeled as that of estimating the patch intensities  $\mathbf{z}_i$  given the observation model of (2). There we assumed that the noise patches  $\boldsymbol{\eta}_i$  are iid with a known probability density function (pdf) and that they are independent of the clean image patches  $\mathbf{z}_i$ . To study this estimation problem, the  $\mathbf{z}_i$  vectors were considered to be instances of some random variable  $\mathbf{z}$  sampled from some (unknown) probability density function  $p(\mathbf{z})$ . The performance limit of estimating the  $\mathbf{z}_i$  vectors is then derived as the Bayesian Cramér-Rao lower bound [10] on the MSE of biased estimators. This analysis was made possible assuming image patches were geometrically similar. Thus, images containing varying geometric structure were segmented into  $k = 1, \dots, K$  groups of geometrically similar patches, as is shown in Fig. 1. There it can be seen that geometric clustering groups together the smooth patches, while separating patches containing horizontal and vertical edges, and corner regions into clusters of their own. A bound on the MSE was obtained for *each patch* within a cluster  $\Omega_k$  as

$$E [\|\mathbf{z}_i - \hat{\mathbf{z}}_i\|^2] \geq \text{Tr} \left[ (\mathbf{J}_i + \mathbf{C}_z^{-1})^{-1} \right] \quad (3)$$

where  $\mathbf{J}_i$  is the Fisher information matrix (FIM) and  $\mathbf{C}_z$  is the covariance of  $\mathbf{z}$  from pdf  $p_k(\mathbf{z})$ .  $\mathbf{C}_z$  captures the cluster complexity in terms of variation between member patches while

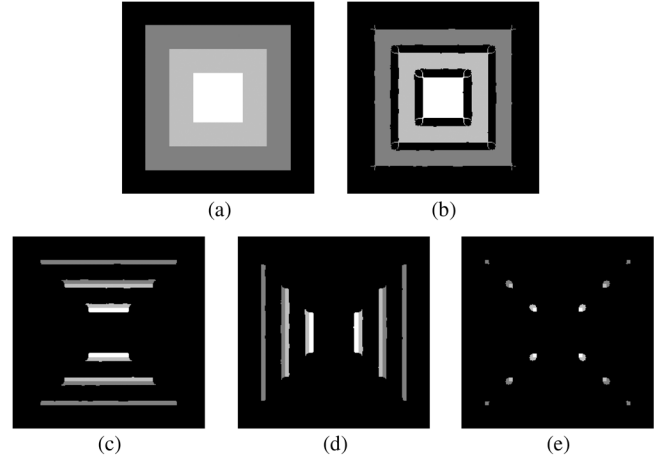


Fig. 1. Example of geometric clustering: (a) Noise-free image, (b)–(e) few clusters based upon geometric structure of patches.

the FIM depends upon the noise characteristics and the number ( $N_i$ ) of similar patches that exist for each patch ( $\mathbf{z}_i$ ). Thus, the bound for each cluster is a function of the cluster complexity as well as the noise characteristics. For zero mean Gaussian noise  $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ , the (conditional) FIM takes the simple form

$$\mathbf{J}_i = -E \left[ \frac{\partial^2 \ln p(\mathbf{y}|\mathbf{z})}{\partial \mathbf{z}_i \partial \mathbf{z}_i^T} \right] = N_i \frac{\mathbf{I}}{\sigma^2} \quad (4)$$

where  $\mathbf{I}$  denotes the identity matrix, and  $N_i$  is the number of patches within the image that are *photometrically* similar to the patch  $\mathbf{z}_i$ . Note that this form of the FIM is obtained only when the noise patches are assumed to be iid and independent of the patch intensity.<sup>1</sup> Photometric similarity and, therefore,  $N_i$  are determined patchwise for each  $\mathbf{z}_i$  by searching for the number of patches  $\mathbf{z}_j$  in the entire image that satisfy the condition

$$\mathbf{z}_j = \mathbf{z}_i + \boldsymbol{\varepsilon}_{ij} \quad \text{such that } \|\boldsymbol{\varepsilon}_{ij}\| \leq \gamma \quad (5)$$

where  $\gamma$  is a threshold dependent on the patch size [1]. The bound for a cluster  $\Omega_k$  can then be computed as

$$E [\|\mathbf{z}_i - \hat{\mathbf{z}}_i\|^2]_{\Omega_k} \geq \frac{1}{M_k} \sum_{i=1}^{M_k} \text{Tr} \left[ (\mathbf{J}_i + \mathbf{C}_z^{-1})^{-1} \right] \quad (6)$$

for all  $\mathbf{z}_i \in \Omega_k$  with cluster cardinality  $M_k$ . It is interesting to note that the bound does not require us to know the entire pdf  $p_k(\mathbf{z})$  (or the joint distribution of  $p_k(\mathbf{y}, \mathbf{z})$ ) as is the case for some other related studies [13], [14]. To estimate the bounds, we need to estimate only the first two moments of  $p_k(\mathbf{z})$  which is done using a bootstrapping mechanism [15]. The lower bound on the MSE for the entire image is then calculated as a weighted sum of the bounds for each cluster.

In [1], it was shown that the expression for the bound corresponds to the performance of the Bayesian minimum mean

<sup>1</sup>In practice, this does not hold true. Modern cameras have nonlinear response functions and the sensitivity to noise is a function of the pixel intensities. In fact, researchers have proposed denoising methods [11], [12] that account for such intensity dependent noise. In our work, we assume the noise to be independent of pixel intensities in deriving a denoising bound. However, the framework in [1] is general enough to be able to account for intensity dependent noise as well. In that case, the FIM will take a different form than (4). We consider that case to be outside the scope of this paper.

squared error (B-MMSE) estimator when  $p(\mathbf{z})$  is Gaussian. More generally, the linear MMSE estimator can be shown to achieve the performance bound in each cluster for the Gaussian noise case. However, such an estimator requires oracle clustering and perfect knowledge of  $\mathbf{C}_z$ , making it apparently unachievable. Yet, comparisons with state-of-the-art denoising methods have shown that for certain class of textured images, current methods perform remarkably close to the theoretical limits. Meanwhile, images lacking such finer details provide considerable room for performance improvement. For such comparisons in [1], the bounds were calculated from the noise-free images. In this paper, we propose a method of estimating the denoising bounds from noisy images where the noise statistics are assumed to be known.

### III. LEARNING THE BOUND WITHOUT GROUND TRUTH

In the previous section, we provided a brief overview of the bounds formulation introduced in [1] where the bounds are calculated directly from the noise-free image. However, in practical applications, it may be necessary to estimate the bounds from a noisy image without any ground-truth. Such an estimate can be useful in automatically deciding parameters of a denoising method when only a single noisy image is available. If embedded within a camera, such information will also be useful to camera manufacturers and photographers, who can then tune camera parameters (that control the amount of noise that appears in the captured image) to ensure that the captured noisy image can be denoised well enough to yield a resultant image of acceptable quality. In this section, we propose a method of estimating the MSE bounds through accurate estimation of the parameters  $N_i$  for each patch and  $\mathbf{C}_z$  for each cluster.

#### A. Estimating the Patch Covariance Matrix

We first discuss a method of estimating the covariance matrix  $\mathbf{C}_z$  given the noisy image patches. Since the covariance of the underlying patch is computed using all the patches in the cluster, we first need to cluster the image based upon underlying geometry. In [16], we presented a method of learning the covariance matrix without performing any explicit clustering on the noisy image. However, that method is restrictive as it depends on selecting a very large and relevant database of clean image patches. Instead, here we compute the local steering kernels [5], [7] for each patch in the noisy image and then perform clustering using K-Means [17], much in the same way as done for the noise-free case in [1]. In [7], we have shown that such normalized kernels can be quite robust to the presence of noise, leading to relatively robust clustering performance. This is illustrated in Fig. 3 where we show (color-coded) clustering of the noise-free and noisy Barbara images, where it can be seen that the clustering is similar for the two images. However, inaccuracies do appear, especially when dealing with strong noise. In such cases, one can prefilter the noisy image to reduce the effect of noise and perform clustering on the denoised image. As will be apparent from experimental results shown in Section III-C, the covariance estimate obtained using such a clustering leads to quite accurate estimates of the denoising bounds, even when substantial noise ( $\sigma = 25$ ) is considered.

Once the image has been clustered, we proceed to compute the covariance of the noisy patches in each cluster. For this, we employ the bootstrapping method of Efron [15], although other stable and computationally efficient methods (such as [18] and references therein) are equally applicable. This allows us to estimate the covariance  $\mathbf{C}_y$  of the noisy patches within a given cluster, from which we need to estimate the covariance matrix  $\mathbf{C}_z$ . From the data model of (2), it is easy to see that

$$\mathbf{C}_z = \mathbf{C}_y - \mathbf{C}_\eta \quad (7)$$

where  $\mathbf{C}_\eta = \sigma^2\mathbf{I}$  is the covariance of the iid noise, which is assumed to be independent of the patch intensity. However, directly using (7) can lead to an estimate of  $\mathbf{C}_z$  that may not be positive semidefinite, a necessary property of covariance matrices. To avoid such problems in the estimation of the covariance, we use a modified plug-in estimator [19]–[21]

$$\hat{\mathbf{C}}_z = [\hat{\mathbf{C}}_y - \sigma^2\mathbf{I}]_+ \quad (8)$$

where  $\hat{\mathbf{C}}_y$  is the covariance estimated from the noisy image patches and  $[\mathbf{X}]_+$  denotes a matrix with the negative eigenvalues of  $\mathbf{X}$  replaced by 0. Such an estimate is, thus, always positive semidefinite. Note that it may still be the case that  $\hat{\mathbf{C}}_z$  is rank deficient and, hence, not invertible. Therefore, we compute the bounds using an alternate formulation based upon the matrix inversion lemma [22] as

$$\begin{aligned} E[\|\mathbf{z}_i - \hat{\mathbf{z}}_i\|^2] &\geq \text{Tr} \left[ \left( \mathbf{J}_i + \hat{\mathbf{C}}_z^{-1} \right)^{-1} \right] \\ &= \text{Tr} \left[ \mathbf{J}_i^{-1} - \mathbf{J}_i^{-1} \left( \mathbf{J}_i^{-1} + \hat{\mathbf{C}}_z \right)^{-1} \mathbf{J}_i^{-1} \right] \end{aligned} \quad (9)$$

where the covariance estimate  $\hat{\mathbf{C}}_z$  need not necessarily be invertible.

Another point to note is that we assume knowledge of the noise variance in our estimation of the covariance matrix in (8). However, in practice, this needs to be estimated from the given noisy image. In this case, one can employ methods outlined in [11], [23] where it is shown that noise variance can be quite accurately estimated from a single noisy image. Later, in Section III-C, we show that such an estimate of  $\mathbf{C}_z$  is quite robust to minor inaccuracies in noise variance estimation, as well as to the presence of outliers that appear due to errors in clustering a noisy image. As such, the estimated covariances are sufficiently accurate for us to estimate the bounds from any given noisy image.

#### B. Estimation of $N_i$

Next, we need to calculate the FIM from the noisy image. Considering the corrupting noise to be additive white Gaussian with known (or estimated [11], [20], [23]) variance and zero mean, estimating the FIM reduces to estimating the redundancy factor  $N_i$  for each patch. We obtain a  $k$ -nearest neighbor based estimate for  $N_i$  from the noisy input image, similar to the case where the MSE bounds were estimated from noise-free images [1]. However, the similarity measure of (5) needs to be modified to account for the effects of the corrupting noise. In the present context, given any noisy patch  $\mathbf{y}_i$ , we wish to identify patches  $\mathbf{y}_j$  in the noisy image, such that their corresponding noise-free

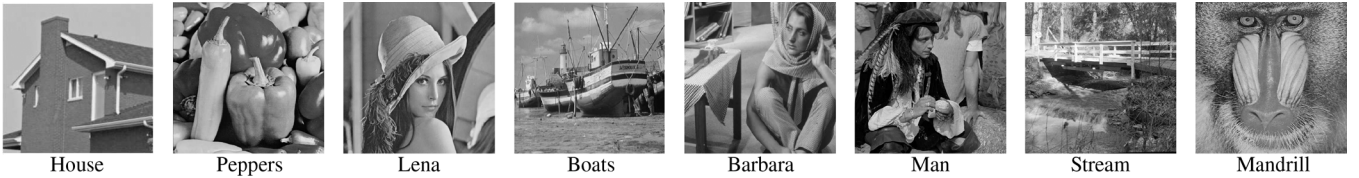


Fig. 2. Some popularly used images on which we perform our analysis.

counterparts  $\mathbf{z}_i$  and  $\mathbf{z}_j$  satisfy the similarity condition defined in (5). Thus, we define a measure of similarity between *noisy* patches as

$$\begin{aligned} \mathbf{z}_j &= \mathbf{z}_i + \boldsymbol{\varepsilon}_{ij} \\ \Rightarrow \mathbf{y}_j - \boldsymbol{\eta}_j &= \mathbf{y}_i - \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_{ij} \quad [\text{from Eq. 2}] \\ \Rightarrow \mathbf{y}_j &= \mathbf{y}_i + \underbrace{(\boldsymbol{\eta}_j - \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_{ij})}_{\tilde{\boldsymbol{\varepsilon}}_{ij}} \end{aligned} \quad (10)$$

$$\begin{aligned} \text{where } \|\tilde{\boldsymbol{\varepsilon}}_{ij}\|^2 &= \|\boldsymbol{\varepsilon}_{ij}\|^2 + \|\boldsymbol{\eta}_j - \boldsymbol{\eta}_i\|^2 + 2\boldsymbol{\varepsilon}_{ij}^T(\boldsymbol{\eta}_j - \boldsymbol{\eta}_i) \\ \Rightarrow E[\|\tilde{\boldsymbol{\varepsilon}}_{ij}\|^2] &= E[\|\boldsymbol{\varepsilon}_{ij}\|^2] + 2\sigma^2 n \end{aligned} \quad (11)$$

where  $n$  is the dimensionality of the image patches. The expression of (11) is obtained assuming the noise patches are iid. A noisy patch  $\mathbf{y}_j$  can be considered photometrically similar to  $\mathbf{y}_i$  if it satisfies the condition

$$\mathbf{y}_j = \mathbf{y}_i + \tilde{\boldsymbol{\varepsilon}}_{ij} \quad \text{such that} \quad \|\tilde{\boldsymbol{\varepsilon}}_{ij}\|^2 \leq \gamma^2 + 2\sigma^2 n \quad (12)$$

where  $\gamma$  is the threshold defined in (5). Note that, as with the estimation of the covariance matrix, we make use of the known (or estimated [11], [20], [23]) noise variance in identifying similar patches. With a similarity measure defined, we can now estimate  $N_i$  values for each patch from the given noisy image.

Once an estimate of the  $N_i$  values for each patch (denoted by  $\hat{N}_i$ ) and its associated covariance matrix ( $\hat{\mathbf{C}}_{\mathbf{z}}$ ) are obtained, we can estimate the MSE bound for denoising from the input noisy image as

$$E[\|\mathbf{z}_i - \hat{\mathbf{z}}_i\|^2] \geq \frac{1}{M} \sum_{i=1}^M \text{Tr} \left[ \hat{\mathbf{J}}_i^{-1} - \hat{\mathbf{J}}_i^{-1} (\hat{\mathbf{J}}_i^{-1} + \hat{\mathbf{C}}_{\mathbf{z}})^{-1} \hat{\mathbf{J}}_i^{-1} \right] \quad (13)$$

with  $\hat{\mathbf{J}}_i = \hat{N}_i \mathbf{I} / \sigma^2$  and  $M = \sum_k M_k$  is the total number of patches in the image. This proposed estimation method can be used to accurately predict the denoising bounds for images corrupted by considerable levels of noise, as we show in the next section. However, as expected, it degrades when the input signal-to-noise ratio is severely low. In our experiments with different images (Fig. 2), this breaking point occurs when the corrupting noise has a standard deviation  $\sigma$  greater than 15. In such cases, it is useful to prefilter the noisy image to reduce the effects of noise. The  $N_i$  values can then be estimated directly from the noise-suppressed version of the given image.

### C. Experimental Verification

In this section, we provide experimental validation of the bounds estimation method that we described previously. As a first step, we consider the accuracy of the covariance estimates from a given noisy image. The covariance estimates also depend on the clustering performance, which in turn is also in-

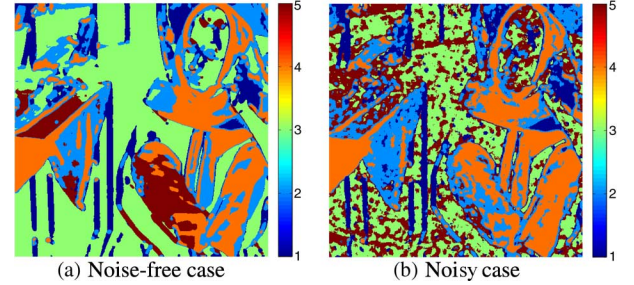


Fig. 3. Clustering of Barbara image into five clusters based upon geometric structure of patches. Clustering is performed with features calculated from (a) clean image, and (b) noisy image of noise standard deviation 15. Note how the kernel features can capture structural information and thereby properly cluster majority of patches even in the presence of noise.

fluenced by the presence of noise, as can be seen from Fig. 3. However, our experiments reveal that the covariance estimation process is quite robust to the presence of outliers within each cluster. This can be seen in Fig. 4 where we plot the bounds for the *covariance test* case where the  $N_i$  values are computed from the clean images. There it can be seen that even in the presence of strong noise ( $\sigma = 25$ ) the estimated bounds are quite close to the ground truth computed from clean images. The small error bars representing the standard deviations about the mean for the bounds estimates over five different realizations of noise illustrate the fact that the covariance estimation process is quite robust to the presence of outliers that occur due to errors in clustering.

Next, we move on to the case where the bounds are calculated entirely from the noisy image. That is to say that both  $N_i$  and  $\mathbf{C}_{\mathbf{z}}$  are estimated from the noisy image. The mean of the bounds estimates obtained for various images over five different realizations of noise are shown in Fig. 4. We observe that when the noise standard deviation  $\sigma \leq 15$ , the bounds are estimated quite accurately from the noisy image. However, when stronger noise is considered, our experiments indicate that the bounds estimates (not included in Fig. 4) can be quite inaccurate. In particular, we noted that estimation of  $N_i$  is severely affected when strong noise corrupts the image. However, the same is not the case for the estimation of  $\mathbf{C}_{\mathbf{z}}$ . This is not surprising since the  $N_i$  values are estimated pointwise, whereas the covariance matrices are computed from a much larger number of patches within each cluster. In fact, one of the most popular denoising algorithms, BM3D [6], that relies on identification of similar patches within the image, performs an initial filtering of highly noisy images to reduce the effects of noise before comparing patches to detect similarities. Along similar lines, we can denoise for our  $N_i$  estimation. However, strong denoising leads to considerable over-estimation of  $N_i$  values, especially for patches con-

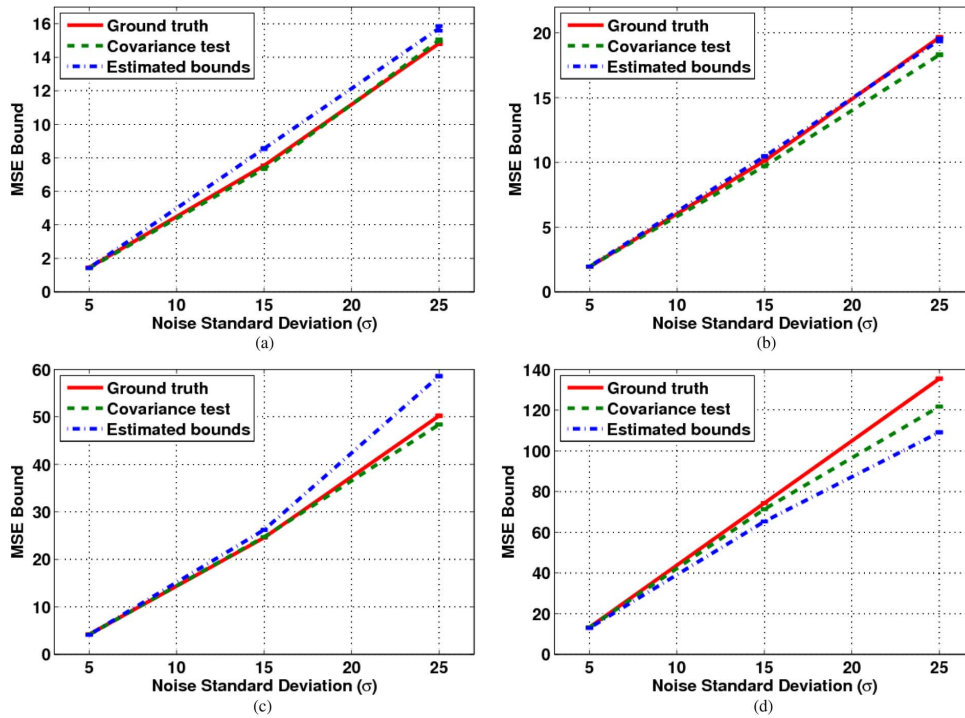


Fig. 4. MSE bounds estimated from noisy (a) house; (b) Lena; (c) Barbara; and (d) stream images (labeled *estimated bounds*) compared to the ground truth [1] where the bounds are calculated from clean images. We also test the accuracy of covariance estimation (labeled *covariance test*) by calculating the bounds using  $N_i$  values estimated from the clean image. For all the images, the  $N_i$  estimates used to compute the *estimated bounds* are obtained directly from the noisy images for noise standard deviation  $\sigma \leq 15$ , and from the prefiltered images for  $\sigma = 25$ .

taining fine texture, resulting in considerable under-estimation of the bounds. To avoid this, we perform only mild prefiltering in such a way so as to retain the texture in the image. For this preprocessing step we make use of the successful BM3D [6] algorithm, setting the parameter (input noise variance) of the algorithm such that the denoising process leaves behind sufficient noise so as to bring the prefiltered image to within effective range of the bound estimate. In particular, using the residual of the estimate, we set the BM3D parameter so as to ensure that a noise-suppressed image is obtained for which the estimated noise standard deviation  $\sigma \approx 5$  in the smoother regions of the image. Using such a method, the bounds are estimated more accurately even for images corrupted by strong noise ( $\sigma = 25$ ). This can be seen from the bounds estimates shown in Fig. 4, where we compare the bounds estimated from noisy images to those computed from their corresponding noise-free versions. For the case where the noise standard deviation  $\sigma \leq 15$ , we compute the bounds parameters directly from the noisy images. However, for the strong noise case ( $\sigma = 25$ ) the prefiltered images are used in estimating the  $N_i$  values. The patch covariance matrices, however, are still computed from the noisy images. It can be seen from the results that using a prefiltering step, we are able to estimate the bounds quite accurately even in the presence of strong noise.

#### IV. ANALYSIS OF THE BOUNDS FORMULATION

In this section, we explore the relationship between the denoising bounds and the mutual information between the noise-free and noisy image patches. This provides an interesting information-theoretic interpretation of the bounds, which, until

TABLE I  
COMPARISON OF BOUNDS FROM NOISY AND NOISE-FREE IMAGES CONSIDERED TO BE GROUND TRUTH. THE NOISE IS AWGN WITH STANDARD DEVIATION 15. THE MEAN BOUNDS FROM FIVE DIFFERENT REALIZATIONS OF NOISE ARE SHOWN, ALONG WITH THE STANDARD DEVIATIONS ABOUT THE MEANS SHOWN IN BRACES

Image	Bounds from		Error Percentage
	Ground truth	Noisy image	
House	7.54	8.55 (0.042)	13.40
Peppers	9.93	9.53 (0.052)	4.03
Lena	10.13	10.55 (0.042)	4.15
Boats	19.68	19.41 (0.069)	1.37
Barbara	24.58	26.22 (0.041)	6.67
Man	33.56	28.32 (0.035)	15.61
Stream	74.30	65.25 (0.037)	12.18
Mandrill	92.56	83.78 (0.123)	9.49
<b>Mean</b>	<b>34.04</b>	<b>31.45</b>	<b>7.61</b>

now, we have studied in an estimation theoretic setting. From Table I, it can be seen that relatively smoother images can be denoised much better (in terms of MSE) than those with more complicated structure (texture). Typically, the latter class of images tend to have lower patch redundancy along with a higher patch complexity, both of which influence the quality of the denoised image. In this section, we explore information-theoretic relationships between these parameters. In particular, we show that  $N_i$  and  $C_z$  are related through the cluster-wise Shannon entropy [24]. When only a noisy image is available, this relationship is manifested through the mutual entropy between the noisy and the (unknown) noise-free image patches. Here we study this relationship in the context of Gaussian noise.

### A. Relationship Between Denoising Bounds and Mutual Information

In [1], we showed that when  $\mathbf{z}$  is Gaussian, the lower bound coincides with the performance of the Bayesian minimum MSE estimator (assuming oracle knowledge of clustering and  $\mathbf{C}_z$ .) When no such assumption is made about  $\mathbf{z}$ , the bounds formulation is the performance of the linear MMSE estimator, considering the noise to be Gaussian. In Gaussian noise, the LMMSE estimator can be derived individually for each image patch  $\mathbf{z}_i$ , given  $N_i$  similar patches as

$$\underline{\mathbf{y}}_i = \mathbf{A}_i \mathbf{z}_i + \underline{\boldsymbol{\eta}}_i$$

where

$$\begin{aligned} \underline{\mathbf{y}}_i &= [\mathbf{y}_1^T \dots \mathbf{y}_j^T \dots \mathbf{y}_{N_i}^T]^T \in \mathbb{R}^{nN_i \times 1} \\ \underline{\boldsymbol{\eta}}_i &= [\boldsymbol{\eta}_1^T \dots \boldsymbol{\eta}_j^T \dots \boldsymbol{\eta}_{N_i}^T]^T \in \mathbb{R}^{nN_i \times 1} \\ \mathbf{A}_i &= [\mathbf{I} \dots \mathbf{I}]^T \in \mathbb{R}^{nN_i \times n} \end{aligned} \quad (14)$$

with  $\mathbf{I}$  denoting the  $n \times n$  identity matrix. The previous data model, written for each underlying  $\mathbf{z}_i$  patch, accounts for the  $N_i$  similar patches that exist for any given  $\mathbf{z}_i$ . The  $\underline{\mathbf{y}}_i$  vector is, thus, formed by concatenating all  $\mathbf{y}_j$  vectors that are similar to any given  $\mathbf{y}_i$  where similarity is defined in (12). The corresponding noise patch  $\underline{\boldsymbol{\eta}}_i$  formed from independent  $\boldsymbol{\eta}_j$  vectors then has a covariance of  $\mathbf{C}_{\boldsymbol{\eta}} = \sigma^2 \mathbf{I}_{nN_i}$  where  $\mathbf{I}_{nN_i}$  is the  $nN_i \times nN_i$  identity matrix. The LMMSE estimator for each  $\mathbf{z}_i$  then has the form  $\hat{\mathbf{z}}_i = E[\mathbf{z}_i | \underline{\mathbf{y}}_i]$  with the corresponding error covariance [10]

$$\begin{aligned} \mathbf{Q}_i &= (\mathbf{C}_z^{-1} + \sigma^{-2} \mathbf{A}_i^T \mathbf{I}_{nN_i} \mathbf{A}_i)^{-1} \\ &= \left( \mathbf{C}_z^{-1} + N_i \frac{\mathbf{I}}{\sigma^2} \right)^{-1}. \end{aligned} \quad (15)$$

The previous  $\mathbf{Q}_i$  matrix is, thus, our MMSE covariance matrix for the estimation of  $\mathbf{z}_i$ . Comparing to (3) & (4), it can be seen that the trace of  $\mathbf{Q}_i$  is in fact the lower bound derived for the denoising problem when the corrupting noise is Gaussian.

Although derived purely from an estimation theoretic point of view, the MSE bounds for denoising can be shown to be related to information-theoretic measures such as the mutual information of the noisy  $\mathbf{y}$  and noise-free  $\mathbf{z}$  patches. The mutual information (MI) of the random variables  $\mathbf{y}$  and  $\mathbf{z}$  is a measure of the information that one variable contains about the other and can be mathematically expressed as [25]

$$\begin{aligned} I(\mathbf{y}; \mathbf{z}) &= H(\mathbf{y}) - H(\mathbf{y} | \mathbf{z}) \\ &= H(\mathbf{y}) - H(\mathbf{z} + \boldsymbol{\eta} | \mathbf{z}) \quad (\text{from Eq. 2}) \\ &= H(\mathbf{y}) - H(\boldsymbol{\eta} | \mathbf{z}) \\ &= H(\mathbf{y}) - H(\boldsymbol{\eta}) \end{aligned} \quad (16)$$

where  $H(\mathbf{y})$  and  $H(\boldsymbol{\eta})$  denote the entropy of  $\mathbf{y}$  and the noise  $\boldsymbol{\eta}$  respectively.<sup>2</sup> The entropy of a random variable  $\mathbf{y}$  (or equivalently its pdf  $p(\mathbf{y})$ ) is defined as

$$H(\mathbf{y}) = -E[\ln p(\mathbf{y})] = - \int p(\mathbf{y}) \ln p(\mathbf{y}) d\mathbf{y}. \quad (17)$$

<sup>2</sup>We will alternately denote the entropy of a random variable  $\mathbf{x}$  with a pdf  $p(\mathbf{x})$  as  $H(p)$  or  $H(\mathbf{x})$ , as necessary for clarity of presentation. The notation used will be clear from context.

While  $H(\mathbf{y})$  can be estimated from the observed noisy image patches (see Appendix A), the noise entropy may be analytically calculated if the noise statistics are known. Specifically, for Gaussian noise with given covariance  $\mathbf{C}_{\boldsymbol{\eta}}$ , the entropy is given by

$$H(\boldsymbol{\eta}) = \ln \left[ (2\pi e)^{n/2} |\mathbf{C}_{\boldsymbol{\eta}}|^{1/2} \right] = \frac{1}{2} \ln (|\mathbf{C}_{\boldsymbol{\eta}}|) + \frac{n}{2} [1 + \ln(2\pi)] \quad (18)$$

where  $|\cdot|$  denotes the determinant.

In [26], Palomar *et al.* studied the relationship between the mutual information between the noisy and noise-free image patches and the minimum mean squared error (MMSE) on the estimation of the input given the output of a Gaussian channel. For the multivariate case (14), the authors show that the gradients of the MI with respect to the signal and noise covariance can be written in terms of the MMSE matrix of (15) as

$$\frac{d}{d \mathbf{C}_z} I(\mathbf{y}; \mathbf{z}) \mathbf{C}_z = \mathbf{A}_i^T \mathbf{C}_{\boldsymbol{\eta}}^{-1} \mathbf{A}_i \mathbf{Q}_i \quad \text{and} \quad (19)$$

$$\frac{d}{d \mathbf{C}_{\boldsymbol{\eta}}} I(\mathbf{y}; \mathbf{z}) = -\mathbf{C}_{\boldsymbol{\eta}}^{-1} \mathbf{A}_i \mathbf{Q}_i \mathbf{A}_i^T \mathbf{C}_{\boldsymbol{\eta}}^{-1}. \quad (20)$$

When dealing with iid noise, where  $\mathbf{C}_{\boldsymbol{\eta}} = \sigma^2 \mathbf{I}_{nN_i}$ , the previous relations can be written for each cluster as (see Appendix B for derivation)

$$\begin{aligned} \frac{d}{d \mathbf{C}_z} I(\mathbf{y}; \mathbf{z}) \mathbf{C}_z &= \frac{1}{M_k} \sum_{i=1}^{M_k} \frac{d}{d \mathbf{C}_z} I(\mathbf{y}_i; \mathbf{z}_i) \mathbf{C}_z \\ &= \frac{1}{M_k} \sum_{i=1}^{M_k} \frac{N_i}{\sigma^2} \mathbf{Q}_i \quad \text{and} \end{aligned} \quad (21)$$

$$\begin{aligned} \frac{d}{d \sigma^2} I(\mathbf{y}; \mathbf{z}) &= \frac{1}{M_k} \sum_{i=1}^{M_k} \frac{d}{d \sigma^2} I(\mathbf{y}_i; \mathbf{z}_i) \\ &= -\frac{1}{n M_k \sigma^4} \sum_{i=1}^{M_k} \text{Tr}(\mathbf{Q}_i). \end{aligned} \quad (22)$$

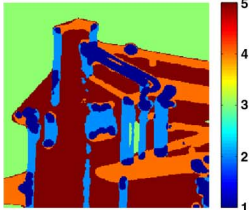
This establishes a direct relationship between the denoising bounds of (3) and the MI (16). As with the bounds formulation, the MI too is a function of both the input signal characteristics and the noise. This can be seen by further expansion of (21) as

$$\begin{aligned} \frac{d}{d \mathbf{C}_z} I(\mathbf{y}; \mathbf{z}) \mathbf{C}_z &= \frac{1}{M_k} \sum_{i=1}^{M_k} \frac{N_i}{\sigma^2} \left( \mathbf{C}_z^{-1} + N_i \frac{\mathbf{I}}{\sigma^2} \right)^{-1} \\ \Rightarrow \frac{d}{d \mathbf{C}_z} I(\mathbf{y}; \mathbf{z}) &= \frac{1}{M_k \sigma^2} \sum_{i=1}^{M_k} N_i \left( \mathbf{C}_z^{-1} + N_i \frac{\mathbf{I}}{\sigma^2} \right)^{-1} \mathbf{C}_z^{-1} \\ &= \frac{1}{M_k \sigma^2} \sum_{i=1}^{M_k} \left( \frac{\mathbf{I}}{N_i} + \frac{\mathbf{C}_z}{\sigma^2} \right)^{-1} \end{aligned} \quad (23)$$

assuming  $\mathbf{C}_z$  to be invertible. A positive definite gradient with respect to the covariance here implies that the MI is an increasing function of patch complexity. Further, it can be seen that as  $\mathbf{C}_z$  increases (with a corresponding drop in the expected  $N_i$ , as we shall see later in Section IV-B), the magnitude of the gradient decreases. This implies that the rate of increase of MI drops as the underlying patch complexity increases. However, with increase in noise strength, the MI can be expected to



TABLE II  
CLUSTERING OF THE HOUSE IMAGE AND THE CLUSTER-WISE MUTUAL INFORMATION ESTIMATES  
WHEN CORRUPTED BY VARIOUS LEVELS OF ADDITIVE WHITE GAUSSIAN NOISE



$\sigma$	Noise Entropy	Estimated MI $\hat{I}(\mathbf{y}; \mathbf{z})$					Overall MI
		$\Omega_1$	$\Omega_2$	$\Omega_3$	$\Omega_4$	$\Omega_5$	
0	-	430.76	365.15	212.71	375.09	345.91	322.49
5	366.43	119.25	80.53	27.88	87.00	60.52	63.93
15	499.37	62.94	43.17	21.63	47.31	30.59	37.64
25	561.18	45.77	33.50	20.35	36.46	25.03	33.05
35	601.89	35.58	27.92	17.81	30.36	22.31	23.97
45	632.30	27.80	23.42	13.47	25.18	19.55	19.74
55	656.58	21.07	19.06	8.59	19.97	16.13	15.23

TABLE III  
RANKING OF IMAGES BASED UPON DENOISING DIFFICULTY AS INDICATED BY THE MI, COMPARED TO THE ENTROPY,  
THE DENOISING BOUND AND MSE OF BM3D DENOISING ALGORITHM FOR ADDITIVE WHITE GAUSSIAN NOISE

Images	Size	Noise-free Entropy	Mutual Information $I(\mathbf{y}; \mathbf{z})$			Denoising Bounds [1]	BM3D MSE [6]
			$\sigma = 5$	$\sigma = 15$	$\sigma = 25$		
House	256 <sup>2</sup>	322.49	63.92	37.65	33.06	14.82	33.57
Lena	512 <sup>2</sup>	350.17	67.39	38.55	31.88	19.66	40.46
Peppers	512 <sup>2</sup>	374.29	72.56	38.37	30.53	19.21	42.96
Barbara	512 <sup>2</sup>	376.74	89.32	49.95	37.81	50.24	55.62
Boats	512 <sup>2</sup>	398.36	89.75	45.75	35.04	38.70	67.17
Man	512 <sup>2</sup>	407.16	94.28	43.49	29.25	62.97	96.46
Stream	512 <sup>2</sup>	473.65	136.18	63.67	43.52	135.46	158.26
Mandrill	512 <sup>2</sup>	498.75	153.67	74.50	51.59	181.61	185.60

decrease, as is implied by the negative gradient of the MI with respect to the noise variance in (22). Using (15) to rewrite (22) as

$$\frac{d}{d\sigma^2} I(\mathbf{y}; \mathbf{z}) = -\frac{1}{nM_k} \sum_{i=1}^{M_k} \text{Tr} \left[ (\sigma^4 \mathbf{C}_z^{-1} + \sigma^2 N_i \mathbf{I})^{-1} \right] \quad (24)$$

we can see that the rate of such decrease is also expected to drop as the noise strength increases.

We study this behavior of the mutual entropy as a function of the noise strength and patch complexity through a simple experiment. For this, we make use of the House image and estimate the MI  $\hat{I}(\mathbf{y}; \mathbf{z})$  for each cluster containing geometrically similar patches (color-coded in Table II) for various levels of AWG noise. For meaningful comparisons, we perform clustering on the noise-free image and use the same cluster membership in computing the MI estimates for the noisy cases. In Table II it can be seen that cluster  $\Omega_3$  consisting of the much smoother background patches has a much lower complexity than that of clusters  $\Omega_2$  and  $\Omega_4$  which capture the edge regions. This relative complexity is also captured by the MI estimates for the clusters (see table) where it can be seen that clusters with higher complexity exhibit higher MI. This can be seen to be in keeping with (23) which implies an increase of MI with increase in patch complexity. Further, the MI of any given cluster decreases as the noise increases, and the rate of such a decrease also drops with higher noise. This is in keeping with the relationship between the MI and the noise variance shown in (24).

Although the MI are clearly related to the parameters of the bounds, it is important to note that one cannot be used directly to predict the other. The formulation of (3) predicts an increase in the MSE bound for denoising as the image complexity and noise variance increases. However, the MI, which quantifies the relative information between a noisy patch  $\mathbf{y}$  and its noise-free counterpart  $\mathbf{z}$ , can be seen to increase with increasing image

complexity, but has quite the opposite effect as noise variance increases. This is in keeping with intuition that as noise increases the noisy patches look more like noise, resulting in a reduction of information that  $\mathbf{y}$  conveys about  $\mathbf{z}$  (and vice versa). However, as the complexity of the noise-free patch increases, stronger noise is needed for the noise to overwhelm the patch characteristics, thus, justifying an increase in MI. These relations are captured analytically in (23) and (24) and experimentally in Table II. It is also important to note that (23) and (24) relate the *rate of change* of MI as a function of changing image complexity and noise variance respectively. Thus, it is the magnitude of the rate of change of MI (and not the MI itself) that is inversely related to the bounds. Consequently, with only a single noisy observation the MI cannot be used to predict the denoising bound. However, the MI can be used to study the relative denoising difficulty of different images that are corrupted by similar levels of noise.

As before, we consider additive Gaussian noise to illustrate the effectiveness of the MI measure in studying relative complexity of images containing patches of diverse geometric structure. For this we need to first estimate the entropy of the entire noisy image from its cluster-wise entropy estimates as (see Appendix C)

$$H(\mathbf{y}) = \sum_{k=1}^K \omega_k H(\mathbf{y} \in \Omega_k) - \sum_{k=1}^K \omega_k \ln \omega_k \quad (25)$$

where  $\omega_k = M_k/M$  is the fraction of total patches that belong to cluster  $\Omega_k$ . For Gaussian noise, the noise entropy is calculated analytically (18) using the known noise covariance matrix. The overall MI can then be estimated using (16). In Table III, we show the estimated mutual information obtained for some images (Fig. 2) when corrupted by additive white Gaussian noise of different strength. There it can be seen that the MI is indicative of the relative denoising difficulty between images. This

can be seen by comparing it to the relative ranking obtained by the MSE of one of the best performing denoising methods (namely, BM3D [6]) for the images. In fact, this ranking can be seen to be more in keeping with the relative denoising difficulty exhibited by the practical methods than that obtained from the denoising bounds [1] calculated from the clean images. In the limiting case when the image is considered to be noise-free, the mutual information becomes the same as the Shannon entropy of the noise-free image.<sup>3</sup> In Table III we show that the relative denoising difficulty prediction of the entropy in that case is also in keeping with those obtained by the MI and the MSE of BM3D. This indicates that the entropy of the image is also related to the denoising bounds. In the next section, we explore this relationship further.

### B. Relationship Between Denoising Bounds and Entropy

The bounds formulation of (3) depends on two parameters, namely the FIM  $\mathbf{J}_i$  and the covariance matrix  $\mathbf{C}_z$  that corresponds to the cluster of which patch  $\mathbf{z}_i$  is a member. For additive white Gaussian noise, estimating the FIM amounts to estimating the number ( $N_i$ ) of similar patches that exist for each patch  $\mathbf{z}_i$ . In general, one can expect to find fewer similar patches in any given image if the variability between patches within a cluster is high. In [1] both these parameters are estimated from the noise-free image, in which case the MI of (16) reduces to the Shannon entropy of the noise-free image. In this section, we establish how each of these two parameters of the bounds formulation are related to the Shannon entropy, and as a result, to each other. In this paper our interest lies solely in analyzing the information-theoretic interpretations of the parameters and relating the two. As such, this relationship between the two does not translate to one being estimated from the other in practice, as will be apparent from our following discussions.

In [1],  $N_i$  is estimated for each patch  $\mathbf{z}_i$  by searching over the entire image. Assuming oracle clustering, one can expect to obtain a good estimate of  $N_i$  by limiting the search for patches similar to any given  $\mathbf{z}_i \in \Omega_k$  to patches within the same cluster  $\Omega_k$ . Let  $N_i$  then denote the number of similar patches that lie within the cluster  $\Omega_k$ , where similarity is defined in (5). We then approximate  $N_i$  by performing a nearest neighbor search within patches in  $\Omega_k$  with a search radius of  $\gamma$ . Considering  $\mathbf{z}_i \in \mathbb{R}^n$ , an estimate of the  $N_i$ -nearest neighbor probability density function can then be written as [27]

$$\begin{aligned} p_k(\mathbf{z}_i) &\approx \frac{N_i/(M_k - 1)}{V_i(\gamma)} = \frac{N_i}{(M_k - 1) \nu_n \gamma^n} \\ &= \frac{N_i \Gamma(1 + n/2)}{(M_k - 1) \pi^{n/2} \gamma^n} \end{aligned} \quad (26)$$

where  $V_i(\gamma)$  is the volume of the ball centered at  $\mathbf{z}_i$  with radius  $\gamma$  and  $\nu_n$  is the volume of the unit ball in  $\mathbb{R}^n$ . Solving for  $N_i$  we have

$$N_i \approx \frac{(M_k - 1) \pi^{n/2} \gamma^n}{\Gamma(n/2 + 1)} p_k(\mathbf{z}_i) \quad (27)$$

<sup>3</sup>Images considered to be “noise-free” can often contain noise as well [23]. However, the noise in such images is typically quite small and, hence, we consider them to be noise-free in our study.

where  $\Gamma(\cdot)$  denotes the Gamma function. Unfortunately, the relation of (26) is accurate only when a considerably large number of patches are present [27]. This is especially true when considering high dimensions (e.g.  $n = 121$  arising from choosing patch sizes of  $11 \times 11$  which have been shown in [1] to be a good choice for obtaining meaningful bounds.) Moreover, this requires us to know or estimate the multivariate pdf  $p_k(\mathbf{z})$ . However, (26) is still useful as it establishes a relation between the pdf  $p_k(\mathbf{z})$  and the number of similar patches that exist within the cluster  $\Omega_k$ . We now extend this relationship by considering the average patch redundancy level within each cluster. Let  $\bar{N}(k) = E[N_i \in \Omega_k | \gamma]$  be the conditional expected value of  $N_i$  for patches within the  $k$ th cluster for a given value of  $\gamma$ , with the expectation taken over  $\mathbf{z} \in \Omega_k$ . From (27), we can then express  $\bar{N}(k)$  as

$$\begin{aligned} \bar{N}(k) &= E \left[ \frac{(M_k - 1) \pi^{n/2} \gamma^n}{\Gamma(n/2 + 1)} p_k(\mathbf{z}_i) \right] \\ &= \frac{(M_k - 1) \pi^{n/2} \gamma^n}{\Gamma(n/2 + 1)} \int [p_k(\mathbf{z})]^2 d\mathbf{z}. \end{aligned} \quad (28)$$

It is interesting to note that  $\bar{N}(k)$  is related to the Rényi  $\alpha$ -entropy [28] which is defined as

$$R_\alpha(p_k) = \frac{1}{1 - \alpha} \ln \left( \int p_k(\mathbf{z})^\alpha d\mathbf{z} \right). \quad (29)$$

Choosing  $\alpha = 2$ , we can then express (28) as

$$\ln(\bar{N}(k)) = \ln \left( \frac{(M_k - 1) \pi^{n/2} \gamma^n}{\Gamma(n/2 + 1)} \right) - R_2(p_k). \quad (30)$$

This expression provides a relationship between  $\bar{N}(k)$  and the Rényi entropy. Namely, as the Rényi entropy increases, the expected number of similar patches within a cluster decreases. The Rényi entropy being a measure of uncertainty of a random variable, (30) then fits with the intuition of lower patch redundancy in clusters with more complicated structure.

Alternately, we can think of the level of redundancy within any cluster to be measured by the mean distance from any patch to its most similar patch (nearest neighbor). An overall smaller distance would then indicate the presence of a larger number of similar patches. Generalizing this alternate measure by considering the distance to the  $\tilde{N}(k)$ -most similar patch, one can then expect a smaller average distance for clusters exhibiting higher levels of redundancy for any fixed  $\tilde{N}(k)$ . Denoting  $\gamma_{i, \tilde{N}(k)}$  as the distance from  $\mathbf{z}_i$  to its  $\tilde{N}(k)$ th nearest neighbor in  $\Omega_k$ , we express the (conditional) mean distance to the  $\tilde{N}(k)$ th nearest neighbor using (27) as

$$\begin{aligned} E \left[ \gamma_{i, \tilde{N}(k)} | \tilde{N}(k) \right] &= E \left[ \left\{ \frac{\tilde{N}(k) \Gamma(1 + n/2)}{(M_k - 1) \pi^{n/2} p_k(\mathbf{z}_i)} \right\}^{1/n} \right] \\ &= \left( \frac{\tilde{N}(k) \Gamma(1 + n/2)}{(M_k - 1) \pi^{n/2}} \right)^{1/n} \\ &\quad \times \int p_k(\mathbf{z})^{(1 - \frac{1}{n})} d\mathbf{z} \end{aligned} \quad (31)$$



where, as before, the expectation is taken over  $\mathbf{z} \in \Omega_k$ . Evans *et al.* [29] derived a more general expression for the distance to the  $\tilde{N}(k)$ th nearest neighbor as

$$E[\gamma_{i,\tilde{N}(k)}|\tilde{N}(k)] = \frac{\Gamma\left(\tilde{N}(k) + \frac{1}{n}\right)}{[\nu_n(M_k - 1)]^{1/n} \Gamma\left(\tilde{N}(k)\right)} \times \int p_k(\mathbf{z})^{(1-\frac{1}{n})} d\mathbf{z} \quad (32)$$

where using the approximation

$$\frac{\Gamma(X + \frac{1}{n})}{\Gamma(X)} \approx X^{\frac{1}{n}} \quad (33)$$

one obtains the same relation as in (31). Note that in our case, we consider a fixed search radius of  $\gamma$  which is chosen independent of the image patches. Hence, we set  $E[\gamma_{i,\tilde{N}(k)}|\tilde{N}(k)] = \gamma$  and evaluate the corresponding  $\tilde{N}(k)$  for which the mean  $\tilde{N}(k)$ -nearest neighbor distance is  $\gamma$ . We are, thus, interested in determining the value of  $\tilde{N}(k)$  for which the mean distance to the  $\tilde{N}(k)$ -nearest patch is  $\gamma$ . Intuitively, we can then expect a larger  $\tilde{N}(k)$  for clusters with relatively simpler patches that are known to exhibit higher redundancy levels. Denoting

$$I_n(p_k) = \int p_k(\mathbf{z})^{(1-\frac{1}{n})} d\mathbf{z} \quad (34)$$

we can then rewrite (31) as

$$\begin{aligned} \gamma &= \left( \frac{\tilde{N}(k) \Gamma\left(\frac{n}{2} + 1\right)}{(M_k - 1)} \right)^{1/n} \frac{I_n(p_k)}{\sqrt{\pi}} \\ \Rightarrow \tilde{N}(k) &= \frac{(M_k - 1)}{\Gamma\left(\frac{n}{2} + 1\right)} \left( \frac{\sqrt{\pi}\gamma}{I_n(p_k)} \right)^n. \end{aligned} \quad (35)$$

From (35), we see that the expected number of similar patches that exist within the given cluster is directly proportional to the total number of patches in the cluster and the radius of the ball of similarity; and inversely proportional to the  $n$ th power of the integral  $I_n(p_k)$ . As can be seen from (29),  $I_n(p_k)$  is directly related to the Rényi entropy for the pdf  $p_k(\mathbf{z})$ , where now  $\alpha = (1 - (1/n)) < 1$ . Denoting the Rényi entropy for this choice of  $\alpha$  as  $R_n(p_k)$ , we obtain

$$\begin{aligned} R_n(p_k) &= n \ln(I_n(p_k)) \\ \Rightarrow \ln(\tilde{N}(k)) &= \ln\left(\frac{M_k - 1}{\Gamma\left(\frac{n}{2} + 1\right)}\right) \\ &\quad + n \ln(\sqrt{\pi}\gamma) - R_n(p_k). \end{aligned} \quad (36)$$

Equation (36), thus, provides a direct relationship between the number of  $\gamma$ -similar patches that can be expected for patches within any given cluster, and the Rényi entropy for that cluster. We can then relate  $\tilde{N}(k)$  to the Shannon entropy [24] by using the fact that as  $\alpha \rightarrow 1$ , the Rényi entropy closely approximates the Shannon entropy. For large  $n$  (such as  $n = 121$ ), we obtain a value of  $\alpha = (1 - (1/n)) \approx 0.992$  which is quite close to 1. Substituting the Shannon entropy,  $H(p_k)$ , for the Rényi entropy, we obtain a relation between  $\tilde{N}(k)$  and  $H(p_k)$  as

$$\ln(\tilde{N}(k)) \approx \ln\left(\frac{M_k - 1}{\Gamma\left(\frac{n}{2} + 1\right)}\right) + n \ln(\sqrt{\pi}\gamma) - H(p_k). \quad (37)$$

The higher the variability of patches within a cluster, the higher is its entropy. Keeping with intuition, (37) predicts an inverse relationship between the number of similar patches and the entropy of the cluster being considered. That is to say, when the entropy of  $\mathbf{z}$  within a particular cluster is high, a lower level of redundancy can be expected from the image patches. Moreover, the entropy of a pdf is dependent on the second-order moment which also captures the variability between patches within a cluster. This relationship has been documented for many of the most popularly used multivariate density functions by Zografos *et al.* [30]. Specifically, for the entropy maximizing  $n$ -dimensional Gaussian density function  $\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ , the entropy can be expressed as a function of the covariance  $\mathbf{C}$  (18). As can be seen from (37), an increase in entropy corresponds to the existence of fewer similar patches (lower  $\tilde{N}(k)$ ). Plugging this (maximum) entropy for the Gaussian pdf into (35), thus, provides us with an estimate of the minimum number of similar patches within an expected distance  $\gamma$  as

$$\tilde{N}_{\min}(k) = \frac{(M_k - 1)\gamma^n}{(2e)^{\frac{n}{2}} \Gamma\left(\frac{n}{2} + 1\right) |\mathbf{C}|^{\frac{1}{2}}}. \quad (38)$$

This  $\tilde{N}_{\min}(k)$  can then be taken to be the lower bound on  $\tilde{N}(k)$  that can be expected for any cluster with a covariance  $\mathbf{C}$ . Equation (38) also indicates that as the variance of the Gaussian increases in any of the  $n$  dimensions, the minimum number of similar patches that can be expected decreases. Further note that the redundancy measure is not dependent on the mean  $\boldsymbol{\mu}$ , which implies independence of the  $\tilde{N}_{\min}(k)$  value from the mean intensity of the patches within the cluster.

For the case of any general (unknown) pdf  $p_k(\mathbf{z})$ , (37) establishes a relation between the number of similar patches that one can expect in a cluster and the corresponding covariance matrix  $\mathbf{C}_z$  that captures the cluster complexity. Equation (37) provides the useful insight that the bounds formulation of (3) and the entropy are similarly related to patch redundancy and cluster complexity. Thus, the entropy can serve as a measure of denoising complexity when noise-free images are considered.

## V. CONCLUSIONS

In this paper we have extended our previous work [1] on analyzing the performance limits for image denoising. We presented a method of estimating the denoising bounds directly from the noisy image. We showed that even for considerably noisy images, the denoising bounds can be estimated quite accurately. We also presented further theoretical analysis of the bounds formulation by relating it to the overall entropy of the image. For noisy images, we established that the bounds formulation is related to the mutual information between the noisy and the corresponding noise-free image. In the limiting case where there is no noise, the mutual information reduces to the Shannon entropy of the image. We showed how the entropy is related to the parameters of the bounds. Although predominantly devised to support the theoretical analysis of the bounds formulation, our experiments point to some useful practical applications of the entropy measures by exploiting the relationship between the

denoising bounds and the mutual information. Namely, the mutual information can be used as an indicator of the relative performance of denoising that one can hope to obtain for noisy images. This can then be used to automatically set parameters in a denoising framework to control the level of smoothing required based upon image content and the level of noise corruption. In general, such tasks can also be performed using the bounds estimated from the noisy image. However, the entropy based approach can be computed faster and has been shown to be better representative of the practical difficulties in denoising any given noisy image.

Our experiments in this paper are restricted to the case where the noise is assumed to be Gaussian. Although the Gaussian pdf is popularly used to model the noise, it would be interesting to analyze the bounds for other noise distributions such as Poisson (noise pdf in low light photography) and Rician distributions (noise pdf in intensity component of magnetic resonance imaging). Since the noise in such cases are dependent on the image intensity, the FIM will also be a function of the image content. As such the relationship between the FIM and information-theoretic measures is well studied [31]. This, therefore, can give deeper insights into information-theoretic relations of the bounds. We consider this to be an interesting and practical direction where the present work can be extended.

#### APPENDIX A ENTROPY ESTIMATION

As mentioned previously, the entropy for any given cluster is related to its complexity, and can, therefore, serve as a measure of denoising difficulty for that cluster. The entropy could be calculated if the prior pdf  $p_k(\mathbf{z})$  could be ascertained or modeled accurately at all  $\mathbf{z}_i \in \Omega_k$ . Although many have proposed various models for natural images [32]–[34], they are not directly applicable to our case since we consider the patch vectors to be *geometrically* similar within each cluster. We, thus, *estimate* the entropy in each cluster from the available  $\mathbf{z}_i$  vectors. For this we make use of order statistics of the nearest-neighbor distances. Let  $\gamma_{i,N}$  denote the distance between the patch  $\mathbf{z}_i$  and its  $N$ -most similar (“nearest”) patch. We then obtain a set of  $\gamma_{i,N}$  measures for  $i = 1, \dots, M_k$  patches in the  $k$ th cluster. An estimator for the entropy can then be obtained by using (37) for a fixed  $N$  as

$$\begin{aligned} \hat{H}_N(p_k) &= -\frac{1}{M_k} \sum_{i=1}^{M_k} \ln(\hat{p}_k(\mathbf{z}_i)) \\ &= \frac{1}{M_k} \sum_{i=1}^{M_k} \ln \left[ \frac{(M_k - 1) \pi^{n/2} \gamma_{i,N}^n}{N \Gamma(1 + n/2)} \right]. \end{aligned} \quad (39)$$

Such an estimator based upon the nearest neighbor distance ( $N = 1$ ) with added bias correction terms was proposed by Kozachenko *et al.* [35] as

$$\hat{H}_1(p_k) = \frac{1}{M_k} \sum_{i=1}^{M_k} \ln \left[ \frac{(M_k - 1) \pi^{n/2} \gamma_{i,1}^n}{\Gamma(1 + n/2)} \right] + \psi$$

where  $\psi \approx 0.5772$  is the Euler constant. This was later extended by considering  $N$ -nearest neighbor distances in [36]–[38] (see also [39]). In general, such an estimator has the form

$$\begin{aligned} \hat{H}_N(p_k) &= \frac{1}{M_k} \sum_{i=1}^{M_k} \ln [(M_k - 1) \nu_n \gamma_{i,N}^n] - \Psi(N) \\ &= \frac{1}{M_k} \sum_{i=1}^{M_k} \ln \left[ \frac{(M_k - 1) \pi^{n/2} \gamma_{i,N}^n}{\Gamma(1 + n/2)} \right] \\ &\quad - \Psi(N) \end{aligned} \quad (40)$$

where  $\Psi(N) = \frac{d}{dN} \ln \Gamma(N)$  is the digamma function. Note that (40) provides us with an estimate of the entropy based upon the  $N$ th most similar patch and, thus, varies with the choice of  $N$ . Estimators of the Shannon (and Rényi) entropy based upon a combination of such estimates obtained using multiple values of  $N$  have been proposed in [40] and [41]. However, such estimators require computation of distances to the  $N$ -most similar patches for each patch in the cluster, a process that can be quite time consuming. Instead, we make use of only the most similar patch, that is  $N = 1$ . In that case, the digamma function  $\Psi(1) = -\psi$ . The entropy estimate obtained using only the distance to the most similar patch is very similar to that obtained using larger  $N$  for the high dimensional case. This can be seen from Fig. 5(a) where we show the entropy estimates obtained with different values of  $N$  for samples from Gaussian density functions  $\mathcal{N}(\mathbf{0}, \mathbf{I})$  of various dimensions. There it can be seen that most estimates are quite accurate when the data is relatively low dimensional. Moreover, for higher dimensions, increasing the number of samples results in better entropy estimation. This can be seen in Fig. 5(b) where for  $n = 121$ , we plot density estimates using (40) with  $N = 1$  as a function of sample size. It can be seen that as the number of samples increases, the estimate comes closer to the actual entropy value<sup>4</sup> of 171.64. However, as Fig. 5(b) illustrates, this convergence as a function of sample size is quite slow. This behavior is to be expected as these estimators are essentially *asymptotically* unbiased, converging to the true value of the entropy as  $M_k \rightarrow \infty$ .

#### APPENDIX B RELATION BETWEEN MI AND MMSE MATRIX

In [26] the authors derive expressions for the gradient of the mutual information between the input and output of a general multivariate Gaussian channel. The gradients are derived with respect to the model parameters specifically for the Gaussian channel model of the form

$$\mathbf{y}_i = \mathbf{A}_i \mathbf{z}_i + \boldsymbol{\eta}_i \quad (41)$$

where  $\mathbf{z}_i \in \mathbb{R}^n$  and  $\mathbf{y}_i \in \mathbb{R}^q$  are the input and output of the Gaussian channel respectively,  $\mathbf{A}_i$  is a  $q \times n$  deterministic matrix, and  $\boldsymbol{\eta}_i$  is iid Gaussian noise. The authors in [26] show that

<sup>4</sup>Our experiments with iid samples drawn from Gaussian pdfs with different covariance matrices indicate that the bias of the entropy estimator of (40) is a function of the dimensionality and the number of samples present, and is independent of the covariance matrix of the Gaussian pdf.

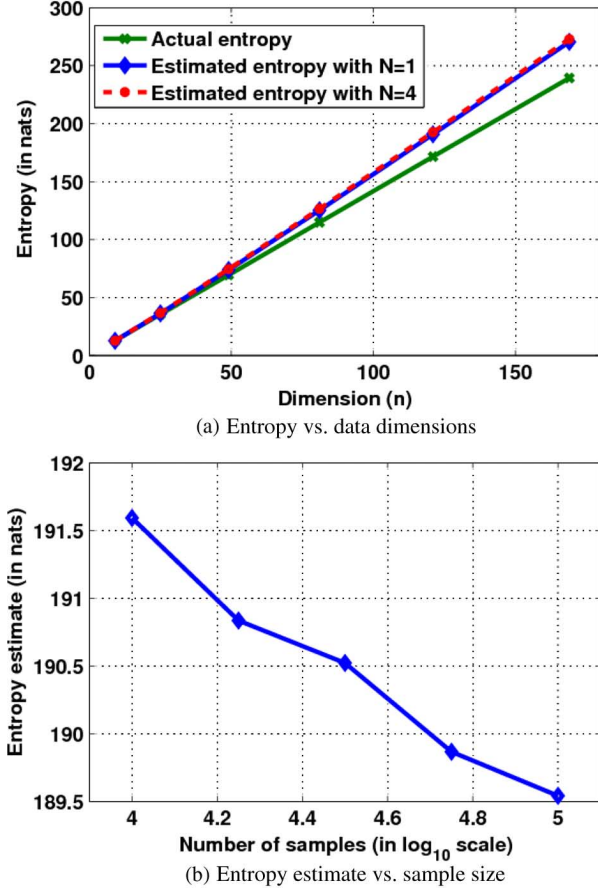


Fig. 5. Estimation of entropy (40) for data sampled from a multidimensional Gaussian density function  $\mathcal{N}(0, \mathbf{I})$  as a function of: (a) dimensions with 20000 samples, and (b) number of samples with  $n = 121$ , where actual entropy is 171.64. These show that the nearest neighbor entropy estimate ( $N = 1$ ) achieves a slightly better estimate of the entropy than using  $N = 4$ . Moreover, the entropy estimates are more accurate for lower dimensions. However, the estimate gets better as the number of samples increases.

the gradients of the MI with respect to the signal and noise covariances  $\mathbf{C}_z$  and  $\mathbf{C}_\eta$  respectively is related to the MMSE matrix  $\mathbf{Q}_i$  as

$$\frac{d}{d\mathbf{C}_z} I(\underline{\mathbf{y}}_i; \mathbf{z}_i) \mathbf{C}_z = \mathbf{A}_i^T \mathbf{C}_\eta^{-1} \mathbf{A}_i \mathbf{Q}_i \quad \text{and} \quad (42)$$

$$\frac{d}{d\mathbf{C}_\eta} I(\underline{\mathbf{y}}_i; \mathbf{z}_i) = -\mathbf{C}_\eta^{-1} \mathbf{A}_i \mathbf{Q}_i \mathbf{A}_i^T \mathbf{C}_\eta^{-1}. \quad (43)$$

The Gaussian channel model of (41) can be thought of as a generalization of the patchwise data model in (2), where in (41) we account for the number (say  $N_i$ ) of similar patches that exist in the cluster for each  $\mathbf{z}_i$ . The vector  $\underline{\mathbf{y}}_i$  is then formed by concatenating all  $\mathbf{y}_j$  patches that are similar to any given  $\mathbf{y}_i$ , and  $\mathbf{A}_i$  takes the form of  $N_i$  identity matrices stacked together, as shown in (14). Thus, in our case,  $q = nN_i$ . Assuming iid noise, we have the  $nN_i \times nN_i$  noise covariance  $\mathbf{C}_\eta = \sigma^{-2} \mathbf{I}_{nN_i}$ , and

$$\begin{aligned} \mathbf{A}_i^T \mathbf{C}_\eta^{-1} &= \sigma^{-2} [\mathbf{I} \dots \mathbf{I}] \mathbf{I}_{nN_i} \\ &= \sigma^{-2} [\mathbf{I} \dots \mathbf{I}] \end{aligned} \quad (44)$$

$$\Rightarrow \mathbf{Q}_i \mathbf{A}_i^T \mathbf{C}_\eta^{-1} = \sigma^{-2} [\mathbf{Q}_i \dots \mathbf{Q}_i] \quad (45)$$

$$\begin{aligned} \Rightarrow \mathbf{C}_\eta^{-1} \mathbf{A}_i \mathbf{Q}_i \mathbf{A}_i^T \mathbf{C}_\eta^{-1} &= \sigma^{-4} \begin{bmatrix} \mathbf{I} \\ \vdots \\ \mathbf{I} \end{bmatrix} [\mathbf{Q}_i \dots \mathbf{Q}_i] \\ &= \sigma^{-4} \begin{bmatrix} \mathbf{Q}_i & \mathbf{Q}_i & \dots \\ \mathbf{Q}_i & \ddots & \\ \vdots & & \end{bmatrix}. \end{aligned} \quad (46)$$

Now, we can rewrite (42) as

$$\begin{aligned} \frac{d}{d\mathbf{C}_z} I(\underline{\mathbf{y}}_i; \mathbf{z}_i) \mathbf{C}_z &= \mathbf{A}_i^T \mathbf{C}_\eta^{-1} \mathbf{A}_i \mathbf{Q}_i = [\mathbf{I} \dots \mathbf{I}] [\mathbf{I} \dots \mathbf{I}]^T \frac{\mathbf{Q}_i}{\sigma^2} \\ &= \frac{N_i}{\sigma^2} \mathbf{Q}_i \end{aligned} \quad (47)$$

and (43) as

$$\begin{aligned} \frac{d}{d\mathbf{C}_\eta} I(\underline{\mathbf{y}}_i; \mathbf{z}_i) &= -\sigma^{-4} \begin{bmatrix} \mathbf{Q}_i & \mathbf{Q}_i & \dots \\ \mathbf{Q}_i & \ddots & \\ \vdots & & \end{bmatrix} \\ \Rightarrow \text{Tr} \left[ \frac{d}{d\mathbf{C}_\eta} I(\underline{\mathbf{y}}_i; \mathbf{z}_i) \right] &= \sum_{j=1}^{nN_i} \frac{d}{d\sigma^2} I(\underline{\mathbf{y}}_i; \mathbf{z}_i) = -\frac{N_i}{\sigma^4} \text{Tr}(\mathbf{Q}_i) \\ \Rightarrow \frac{d}{d\sigma^2} I(\underline{\mathbf{y}}_i; \mathbf{z}_i) &= -\frac{1}{n\sigma^4} \text{Tr}(\mathbf{Q}_i). \end{aligned} \quad (48)$$

Until now we have shown how the MMSE matrix is related to the MI between  $\mathbf{z}_i$  and the vector  $\underline{\mathbf{y}}_i$  that contains all patches similar to  $\mathbf{y}_i$ . To relate the MMSE matrix to  $I(\underline{\mathbf{y}}_i; \mathbf{z}_i)$ , we derive a relation between  $I(\underline{\mathbf{y}}_i; \mathbf{z}_i)$  and  $I(\mathbf{y}_i; \mathbf{z}_i)$ . This is obtained by writing

$$\begin{aligned} I(\underline{\mathbf{y}}_i; \mathbf{z}_i) &= H(\underline{\mathbf{y}}_i) - H(\underline{\boldsymbol{\eta}}_i) \quad [\text{from Equations 16 \& 41}] \\ &= H([\mathbf{y}_1 \dots \mathbf{y}_{N_i}]) - H([\boldsymbol{\eta}_1 \dots \boldsymbol{\eta}_{N_i}]) \\ &= H(\mathbf{y}_1) + \sum_{j=2}^{N_i} H(\mathbf{y}_j | \mathbf{y}_{j-1} \dots \mathbf{y}_1) \\ &\quad - \left[ H(\boldsymbol{\eta}_1) + \sum_{j=2}^{N_i} H(\boldsymbol{\eta}_j | \boldsymbol{\eta}_{j-1} \dots \boldsymbol{\eta}_1) \right]. \end{aligned} \quad (49)$$

Now, for every  $\mathbf{y}_j$  similar to  $\mathbf{y}_i$ , we can relate their corresponding noise-free patches as  $\mathbf{z}_j = \mathbf{z}_i + \boldsymbol{\varepsilon}_{ij}$ . Using this relation and the data model of (2), we get

$$\begin{aligned} H(\mathbf{y}_j | \mathbf{y}_{j-1} \dots \mathbf{y}_1) &= H((\mathbf{z}_j + \boldsymbol{\eta}_j) | (\mathbf{z}_{j-1} + \boldsymbol{\eta}_{j-1}), \dots, (\mathbf{z}_1 + \boldsymbol{\eta}_1)) \\ &= H((\mathbf{z}_j + \boldsymbol{\eta}_j) | (\mathbf{z}_j + \boldsymbol{\varepsilon}_{j,j-1} + \boldsymbol{\eta}_{j-1}), \dots, (\mathbf{z}_j + \boldsymbol{\varepsilon}_{j,1} + \boldsymbol{\eta}_1)) \\ &= H(\boldsymbol{\eta}_j | (\boldsymbol{\varepsilon}_{j,j-1} + \boldsymbol{\eta}_{j-1}), \dots, (\boldsymbol{\varepsilon}_{j,1} + \boldsymbol{\eta}_1)) \\ &= H(\boldsymbol{\eta}_j) \end{aligned} \quad (50)$$

where the last step is a result of  $\boldsymbol{\eta}_j$  vectors being independent of  $\boldsymbol{\varepsilon}_{ij}$  and from each other. From the latter property, we also obtain

$$H(\boldsymbol{\eta}_j | \boldsymbol{\eta}_{j-1} \dots \boldsymbol{\eta}_1) = H(\boldsymbol{\eta}_j). \quad (52)$$

Plugging the previous relations into (49), and replacing  $\mathbf{y}_1$  and  $\boldsymbol{\eta}_1$  with  $\mathbf{y}_i$  and  $\boldsymbol{\eta}_i$  respectively (without loss of generality), we get

$$I(\underline{\mathbf{y}}_i; \mathbf{z}_i) = H(\mathbf{y}_i) - H(\boldsymbol{\eta}_i) = I(\mathbf{y}_i; \boldsymbol{\eta}_i). \quad (53)$$

Equations (47) and (48) can then be written as

$$\frac{d}{d\mathbf{C}_z} I(\mathbf{y}_i; \mathbf{z}_i) \mathbf{C}_z = \frac{N_i}{\sigma^2} \mathbf{Q}_i \quad \text{and} \quad (54)$$

$$\frac{d}{d\sigma^2} I(\mathbf{y}_i; \mathbf{z}_i) = -\frac{1}{n\sigma^4} \text{Tr}(\mathbf{Q}_i). \quad (55)$$

Note that the MMSE matrix is a function of  $N_i$  which can vary across patches within a cluster, where the  $\mathbf{z}_i$  patches are considered to be realizations of the random variable  $\mathbf{z}$ . We write the previous relations in terms of the MI of the random variables  $\mathbf{y}$  and  $\mathbf{z}$  and the MMSE matrix  $\mathbf{Q}_i$  as

$$\frac{d}{d\mathbf{C}_z} I(\mathbf{y}; \mathbf{z}) \mathbf{C}_z = \frac{1}{M_k \sigma^2} \sum_{i=1}^{M_k} N_i \mathbf{Q}_i \quad \text{and} \quad (56)$$

$$\frac{d}{d\sigma^2} I(\mathbf{y}; \mathbf{z}) = -\frac{1}{nM_k \sigma^4} \sum_{i=1}^{M_k} \text{Tr}(\mathbf{Q}_i) \quad (57)$$

by taking the average over all the patches within the cluster.

### APPENDIX C OVERALL ENTROPY DERIVATION

In this section, we derive an expression for the overall entropy from the clusterwise entropy. Our choice of features lead to patches in any given cluster being geometrically similar, thus, allowing us to assume that such patches are realizations of some random variable  $\mathbf{z}$  sampled from some unknown pdf  $p_k(\mathbf{z})$  in each cluster. To estimate the entropy of the entire image we, thus, need to derive an expression relating the entropy of the clusters with that of the entire image. For this, without loss of generality, we assume that the image consists of  $K = 2$  disjoint clusters  $\Omega_1$  and  $\Omega_2$ . The overall pdf of  $\mathbf{z}$  can then be written as

$$p(\mathbf{z}) = \omega_1 p_1 + \omega_2 p_2 \quad (58)$$

where  $\omega_1 + \omega_2 = 1$ . The overall entropy can be derived as

$$\begin{aligned} H(p) &= - \int_{\Omega} p(\mathbf{z}) \ln p(\mathbf{z}) d\mathbf{z} \\ &= - \left[ \int_{\Omega_1} p(\mathbf{z}) \ln p(\mathbf{z}) d\mathbf{z} + \int_{\Omega_2} p(\mathbf{z}) \ln p(\mathbf{z}) d\mathbf{z} \right] \\ &= - \left[ \int_{\Omega_1} \omega_1 p_1 \ln(\omega_1 p_1) d\mathbf{z} + \int_{\Omega_2} \omega_2 p_2 \ln(\omega_2 p_2) d\mathbf{z} \right]. \quad (59) \end{aligned}$$

The previous expression can be simplified by writing

$$\begin{aligned} & - \int_{\Omega_1} \omega_1 p_1 \ln(\omega_1 p_1) d\mathbf{z} \\ &= \omega_1 \left[ - \int_{\Omega_1} p_1 \ln p_1 d\mathbf{z} - \int_{\Omega_1} p_1 \ln \omega_1 d\mathbf{z} \right] \\ &= \omega_1 H(p_1) - \omega_1 \ln \omega_1. \quad (60) \end{aligned}$$

Thus, the overall entropy can be derived as

$$\begin{aligned} H(p) &= \omega_1 H(p_1) + \omega_2 H(p_2) - [\omega_1 \ln \omega_1 + \omega_2 \ln \omega_2] \\ &= \sum_{k=1}^K \omega_k H(p_k) - \sum_{k=1}^K \omega_k \ln \omega_k \quad (61) \end{aligned}$$

where  $\sum_k \omega_k = 1$ . In our derivation, the only assumption we have made is that of the clusters being disjoint, which is true in our case where the clustering is based upon geometric similarity of patches. Equation (61), thus, provides a general expression for calculating the overall entropy of an image from its  $K$  cluster-wise entropy estimates.

### REFERENCES

- [1] P. Chatterjee and P. Milanfar, "Is denoising dead?," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 895–911, Apr. 2010.
- [2] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using a scale mixture of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.
- [3] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 2866–2878, Oct. 2006.
- [4] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
- [5] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 349–366, Feb. 2007.
- [6] K. Dabov, A. Foi, V. Katkovnik, and K. O. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
- [7] P. Chatterjee and P. Milanfar, "Clustering-based denoising with locally learned dictionaries," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1438–1451, Jul. 2009.
- [8] S. Voloshynovskiy, O. Koval, and T. Pun, "Image denoising based on the edge-process model," *Signal Process.*, vol. 85, no. 10, pp. 1950–1969, Oct. 2005.
- [9] T. Treibitz and Y. Y. Schechner, "Recovery limits in pointwise degradation," in *Proc. IEEE Int. Conf. Comput. Photogr.*, San Francisco, CA, Apr. 2009, pp. 1–8.
- [10] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, ser. Signal Processing. Upper Saddle River, N.J.: Prentice-Hall, 1993, vol. 1.
- [11] C. Liu, R. Szeliski, S. B. Kang, C. L. Zitnick, and W. T. Freeman, "Automatic estimation and removal of noise from a single image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 299–314, Feb. 2008.
- [12] N. Joshi, C. Zitnick, R. Szeliski, and D. Kriegman, "Image deblurring and denoising using color priors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, Jun. 2009, pp. 1550–1557.
- [13] Z. Ben-Haim and Y. C. Eldar, "A lower bound on the bayesian MSE based on the optimal bias function," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5179–5196, Nov. 2009.
- [14] H. L. van Trees, *Detection, Estimation, and Modulation Theory*. Hoboken, NJ: Wiley, 1968.
- [15] B. Efron, "Bootstrap methods: Another look at the jackknife," *Ann. Statist.*, vol. 7, no. 1, pp. 1–26, 1979.
- [16] P. Chatterjee and P. Milanfar, "Learning denoising bounds for noisy images," in *Proc. IEEE Int. Conf. Image Process.*, Hong Kong, Sep. 2010, pp. 1157–1160.
- [17] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.
- [18] Y. Chen, A. Wiesel, Y. C. Eldar, and A. O. Hero, III, "Shrinkage algorithms for MMSE covariance estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5016–5029, Oct. 2010.
- [19] Y. C. Eldar and N. Merhav, "Minimax MSE-ratio estimation with signal covariance uncertainties," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1335–1347, Apr. 2005.
- [20] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1522–1531, Sep. 2000.

- [21] D. D. Muresan and T. W. Parks, "Adaptive principal components and image denoising," in *Proc. IEEE Int. Conf. Image Process.*, Barcelona, Spain, Sep. 2003, pp. 101–104.
- [22] M. A. Woodbury, "Inverting modified matrices," Statistical Research Group, Princeton Univ., Princeton, NJ, Memorandum Rep. 42, 1950.
- [23] D. Zoran and Y. Weiss, "Scale invariance and noise in natural images," in *Proc. IEEE Int. Conf. Comput. Vis.*, Kyoto, Japan, Sep.–Oct. 2009, pp. 2209–2216.
- [24] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Jul. 1948.
- [25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, ser. Wiley Series in Telecommunications and Signal Processing, 2nd ed. Hoboken, NJ: Wiley, 2006.
- [26] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141–154, Jan. 2006.
- [27] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, ser. Monographs on Statistics and Applied Probability. London, U.K.: Chapman & Hall, 1986.
- [28] A. Rényi, "On measures of information and entropy," in *Proc. 4th Berkeley Symp. Math. Statist. Probability*, Berkeley, CA, 1961, vol. 1, pp. 547–561.
- [29] D. Evans, "A law of large numbers for nearest neighbor statistics," *Proc. Roy. Soc. A: Math., Phys. Eng. Sci.*, vol. 464, no. 2100, pp. 3175–3192, Dec. 2008.
- [30] K. Zografos and S. Nadarajah, "Expressions for Rényi and Shannon entropies for multivariate distributions," *Statist. Prob. Lett.*, vol. 71, no. 1, pp. 71–84, Jan. 2005.
- [31] B. S. Clarke and A. R. Barron, "Information-theoretic asymptotics of Bayes' methods," *IEEE Trans. Inf. Theory*, vol. 36, no. 3, pp. 453–471, May 1990.
- [32] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.*, vol. 24, pp. 1193–1216, May 2001.
- [33] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu, "On advances in statistical modeling of natural images," *J. Math. Imag. Vis.*, vol. 18, pp. 17–33, 2003.
- [34] Y. Weiss and W. Freeman, "What makes a good model of natural images?," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Minneapolis, MN, Jun. 2007, pp. 1–8.
- [35] L. F. Kozachenko and N. N. Leonenko, "On statistical estimation of entropy of random vector," *Probl. Inf. Trans.*, vol. 23, no. 2, pp. 95–101, 1987.
- [36] H. Singh, N. Misra, V. Hnizdo, A. Fedorowicz, and E. Demchuk, "Nearest neighbor estimates of entropy," *Amer. J. Math. Manage. Sci.*, vol. 23, no. 3 & 4, pp. 301–321, 2003.
- [37] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Phys. Rev. E*, vol. 69, no. 6, p. 066138, Jun. 2004.
- [38] M. N. Gorla, N. N. Leonenko, V. Mergel, and P. L. N. Inverardi, "A new class of random vector entropy estimators and its applications in testing statistical hypotheses," *J. Nonparametric Statist.*, vol. 17, no. 3, pp. 277–297, Apr.–May 2005.
- [39] N. N. Leonenko, L. Pronzato, and V. Savani, "A class of Rényi information estimators for multidimensional densities," *Ann. Statist.*, vol. 36, no. 5, pp. 2153–2182, 2008.
- [40] A. Kaltchenko and N. Timofeeva, "Bias reduction via linear combination of nearest neighbor entropy estimators," *Int. J. Inf. Coding Theory*, vol. 1, no. 1, pp. 39–56, 2009.
- [41] E. Liitiainen, A. Lendasse, and F. Corona, "On the statistical estimation of Rényi entropies," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, Grenoble, France, Sep. 2009, pp. 1–6.



**Priyam Chatterjee** (S'07) received the B.Tech. degree in information technology from the University of Kalyani, India, in 2003, the M.Tech. degree in electrical engineering from the Indian Institute of Technology (IIT), Bombay, India, in 2006, and is currently pursuing the Ph.D. degree in electrical engineering at the University of California, Santa Cruz.

His research interests are in image and video processing (denoising, interpolation, deblurring and super-resolution).



**Peyman Milanfar** (F'10) received the B.S. degree in electrical engineering and mathematics from the University of California, Berkeley, and the M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge.

Until 1999, he was a Senior Research Engineer at SRI International, Menlo Park, CA. He is currently a Professor of Electrical Engineering and associate Dean for research at the University of California, Santa Cruz. He was a Consulting Assistant Professor of computer science at Stanford University, Stanford, CA, from 1998 to 2000, where he was also a Visiting Associate Professor in 2002. His technical interests are in statistical signal and image processing and inverse problems.

Dr. Milanfar won a National Science Foundation CAREER award. He was an Associate Editor for the IEEE TRANSACTION ON IMAGE PROCESSING and was an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS from 1998 to 2001. He won a best paper award from IEEE Signal Processing Society in 2010. He is a member of the Signal Processing Society's Image, Video, and Multidimensional Signal Processing (IVMSP) Technical Committee.