Image Denoising Using Wavelets

— Wavelets & Time Frequency —

Raghuram Rangarajan
Ramji Venkataramanan
Siddharth Shah

December 16, 2002

Abstract

Wavelet transforms enable us to represent signals with a high degree of sparsity. This is the principle behind a non-linear wavelet based signal estimation technique known as wavelet denoising. In this report we explore wavelet denoising of images using several thresholding techniques such as *SUREShrink*, *VisuShrink* and *BayesShrink*. Further, we use a Gaussian based model to perform combined denoising and compression for natural images and compare the performance of these methods.

Contents

"If you painted a picture with a sky, clouds, trees, and flowers, you would use a different size brush depending on the size of the features.Wavelets are like those brushes."

-Ingrid Daubechies

1 Background and Motivation

1.1 Introduction

From a historical point of view, wavelet analysis is a new method, though its mathematical underpinnings date back to the work of Joseph Fourier in the nineteenth century. Fourier laid the foundations with his theories of frequency analysis, which proved to be enormously important and influential. The attention of researchers gradually turned from frequency-based analysis to scale-based analysis when it started to become clear that an approach measuring average fluctuations at different scales might prove less sensitive to noise. The first recorded mention of what we now call a "wavelet" seems to be in 1909, in a thesis by Alfred Haar.

In the late nineteen-eighties, when Daubechies and Mallat first explored and popularized the ideas of wavelet transforms, skeptics described this new field as contributing additional useful tools to a growing toolbox of transforms. One particular wavelet technique, wavelet denoising, has been hailed as "offering all that we may desire of a technique from optimality to generality" [6]. The inquiring skeptic, however maybe reluctant to accept these claims based on asymptotic theory without looking at real-world evidence. Fortunately, there is an increasing amount of literature now addressing these concerns that help us appraise of the utility of wavelet shrinkage more realistically.

Wavelet denoising attempts to remove the noise present in the signal while preserving the signal characteristics, regardless of its frequency content. It involves three steps: a linear forward wavelet transform, nonlinear thresholding step and a linear inverse wavelet transform.Wavelet denoising must not be confused with smoothing; smoothing only removes the high frequencies and retains the lower ones.

Wavelet shrinkage is a non-linear process and is what distinguishes it from entire linear denoising technique such as least squares. As will be explained later, wavelet shrinkage depends heavily on the choice of a thresholding parameter and the choice of this threshold determines, to a great extent the efficacy of denoising. Researchers have developed various techniques for choosing denoising parameters and so far there is no "best" universal threshold determination technique.

The aim of this project was to study various thresholding techniques such as *SUREShrink*[1], *VisuShrink*[3] and *BayeShrink*[5] and determine the best one for image denoising. In the course of the project, we also aimed to use wavelet denoising as a means of compression and were successfully able to implement a compression technique based on a unified denoising and compression principle.

1.2 The concept of denoising

A more precise explanation of the wavelet denoising procedure can be given as follows. Assume that the observed data is

$$X(t) = S(t) + N(t)$$

where S(t) is the uncorrupted signal with additive noise N(t). Let $W(\cdot)$ and $W^{-1}(\cdot)$ denote the forward and inverse wavelet transform operators.. Let $D(\cdot, \lambda)$ denote the denoising operator with threshold λ. We intend to denoise X(t) to recover $\hat{S}(t)$ as an estimate of S(t). The procedure can be summarized in three steps

$$Y = W(X)$$
$$Z = D(Y, \lambda)$$
$$\hat{S} = W^{-1}(Z)$$

$D(\cdot, \lambda)$ being the thresholding operator and λ being the threshold.
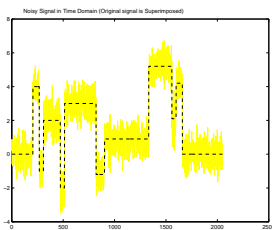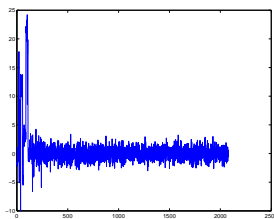
Figure 1: A noisy signal in time domain.

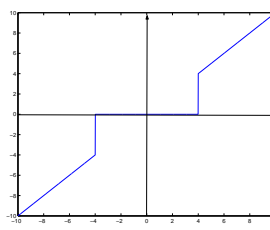Figure 2: The same signal in wavelet domain. Note the sparsity of coefficients.
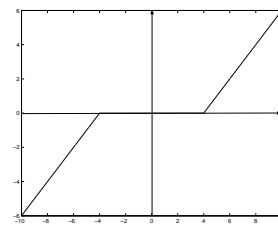

Figure 3: Hard Thresholding.

Figure 4: Soft Thresholding.

2 Thresholding

2.1 Motivation for Wavelet thresholding

The plot of wavelet coefficients in Fig 2 suggests that small coefficients are dominated by noise, while coefficients with a large absolute value carry more signal information than noise. Replacing noisy coefficients (small coefficients below a certain *threshold* value) by zero and an inverse wavelet transform may lead to a reconstruction that has lesser noise. Stated more precisely, we are motivated to this thresholding idea based on the following assumptions:

- The decorrelating property of a wavelet transform creates a sparse signal: most untouched coefficients are zero or close to zero.

- Noise is spread out equally along all coefficients.

- The noise level is not too high so that we can distinguish the signal wavelet coefficients from the noisy ones.

As it turns out, this method is indeed effective and thresholding is a simple and efficient method for noise reduction. Further, inserting zeros creates more sparsity in the wavelet domain and here we see a link between wavelet denoising and compression which has been described in sources such as [5].

2.2 Hard and soft thresholding

Hard and soft thresholding with threshold λ are defined as follows

The hard thresholding operator is defined as

$$D(U, \lambda) = U \text{ for all} |U| > \lambda$$
$$= 0 \text{ otherwise}$$

The soft thresholding operator on the other hand is defined as

$$D(U, \lambda) = sgn(U)max(0, |U| - \lambda)$$

Hard threshold is a "keep or kill" procedure and is more intuitively appealing. The transfer function of the same is shown in Fig 3. The alternative, soft thresholding (whose transfer function is shown in Fig 4), shrinks coefficients above the threshold in absolute value. While at first sight hard thresholding may seem to be natural, the continuity of soft thresholding has some advantages. It makes algorithms mathematically more tractable [3]. Moreover, hard thresholding does not even work with some algorithms such as the GCV procedure [4]. Sometimes, pure noise coefficients may pass the hard threshold and appear as annoying 'blips' in the output. Soft thesholding shrinks these false structures.

2.3 Threshold determination

As one may observe, threshold determination is an important question when denoising. A small threshold may yield a result close to the input, but the result may still be noisy. A large threshold on the

other hand, produces a signal with a large number of zero coefficients. This leads to a smooth signal. Paying too much attention to smoothness, however, destroys details and in image processing may cause blur and artifacts.

To investigate the effect of threshold selection, we performed wavelet denoising using hard and soft thresholds on four signals popular in wavelet literature: Blocks, Bumps, Doppler and Heavisine[2].

The setup is as follows:

- The original signals have length 2048.

- We step through the thresholds from 0 to 5 with steps of 0.2 and at each step denoise the four noisy signals by both hard and soft thresholding with that threshold.

- For each threshold, the MSE of the denoised signal is calculated.

- Repeat the above steps for different orthogonal bases, namely, Haar, Daubechies 2,4 and 8.

The results are tabulated in the table 1

Blocks				Bumps	
	Hard	Soft		Hard	Soft
Haar	1.2	1.6	Haar	1.2	1.6
Db2	1.2	1.6	Db2	1.4	1.6
Db4	1.2	1.6	Db4	1.4	1.6
Db8	1.2	1.8	Db8	1.4	1.8

Heavy Sine				Doppler	
	Hard	Soft		Hard	Soft
Haar	1.4	1.6	Haar	1.6	2.2
Db2	1.4	1.6	Db2	1.6	1.6
Db4	1.4	1.6	Db4	1.6	2.0
Db8	1.4	1.6	Db8	1.6	2.2

Table 1: Best thresholds, empirically found with different denoising schemes, in terms of MSE

2.4 Comparison with Universal threshold

The threshold $\lambda_{UNIV} = \sqrt{2lnN}\sigma$ (N being the signal length, σ^2 being the noise variance) is well known in wavelet literature as the *Universal threshold*. It is the optimal threshold in the asymptotic sense and minimises the cost function of the difference between the function and the soft thresholded version of the same in the L_2 norm sense(i.e. it minimizes $E \parallel Y_{Thresh} - Y_{Orig}. \parallel^2$). In our case, N=2048, $\sigma = 1$, therefore theoretically,

$$\lambda_{UNIV} = \sqrt{2ln(2048)}(1) = 3.905 \qquad (1)$$

As seen from the table, the best empirical thresholds for both hard and soft thresholding are much lower than this value, independent of the wavelet used. It therefore seems that the universal threshold is not useful to determine a threshold. However, it is useful for obtain a starting value when nothing is known of the signal condition. One can surmise that the universal threshold may give a better estimate for the soft threshold if the number of samples are larger (since the threshold is optimal in the asymptotic sense).

3 Image Denoising using Thresholding

3.1 Introduction: Revisiting the underlying principle

An image is often corrupted by noise in its acquisition or transmission. The underlying concept of denoising in images is similar to the 1D case. The goal is to remove the noise while retaining the important signal features as much as possible.

The noisy image is represented as a two-dimensional matrix $\{x_{ij}\}, i,j = 1, \cdots, N$. The noisy version of the image is modelled as

$$y_{ij} = x_{ij} + n_{ij} \qquad i,j = 1, \cdots, N.$$

where $\{n_{ij}\}$ are *iid* as N$(0,\sigma^2)$. We can use the same principles of thresholding and shrinkage to achieve denoising as in 1-D signals. The problem again boils down to finding an optimal threshold such that the
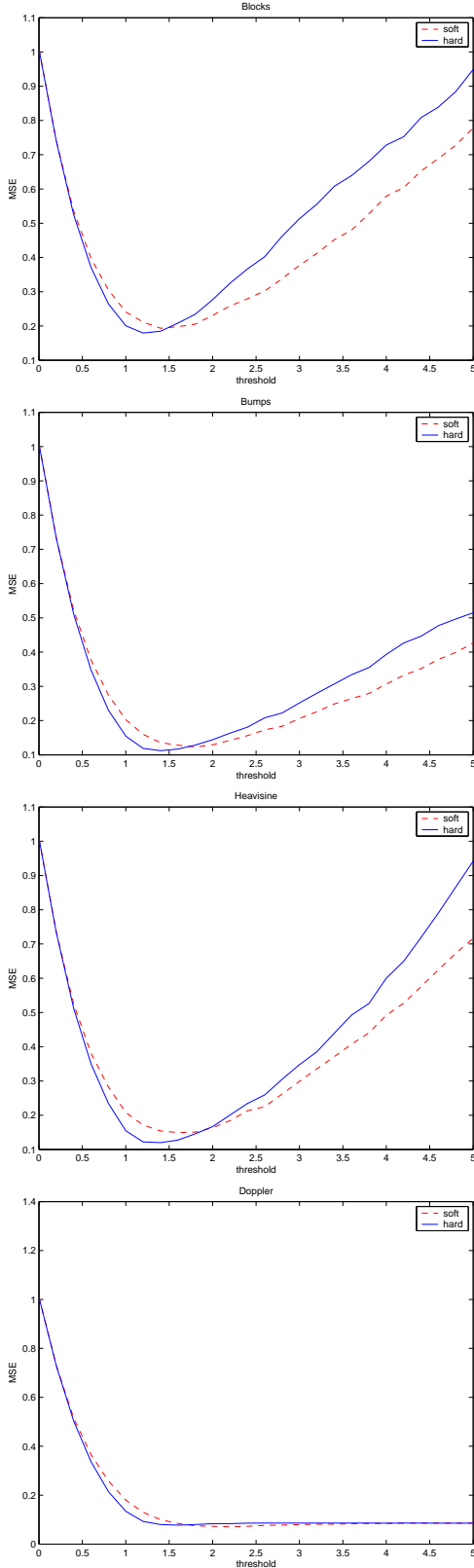
Figure 5: MSE V/s Threshold values for the four test signals.

mean squared error between the signal and its estimate is minimized.

The wavelet decomposition of an image is done as follows: In the first level of decomposition, the image is split into 4 subbands,namely the HH,HL,LH and LL subbands. The HH subband gives the diagonal details of the image;the HL subband gives the horizontal features while the LH subband represent the vertical structures. The LL subband is the low resolution residual consisiting of low frequency components and it is this subband which is further split at higher levels of decomposition.

The different methods for denoising we investigate differ only in the selection of the threshold. The basic procedure remains the same :

- Calculate the DWT of the image.

- Threshold the wavelet coefficients.(Threshold may be universal or subband adaptive)

- Compute the IDWT to get the denoised estimate.

Soft thresholding is used for all the algorithms due to the following reasons: Soft thresholding has been shown to achieve near minimax rate over a large number of Besov spaces[3]. Moreover, it is also found to yield visually more pleasing images. Hard thresholding is found to introduce artifacts in the recovered images.

We now study three thresholding techniques- *VisuShrink,SureShrink* and *BayesShrink* and investigate their performance for denoising various standard images.

3.2   VisuShrink

Visushrink is thresholding by applying the Universal threshold proposed by Donoho and Johnstone [2]. This threshold is given by $\sigma\sqrt{2logM}$ where σ is the noise variance and M is the number of pixels in the image.It is proved in [2] that the maximum of any M values *iid* as $N(0,\sigma^2)$will be smaller than the universal threshold with high probability, with the probability approaching 1 as M increases.Thus, with high

(a) 512×512 image of 'Lena'

(b) Noisy version of 'Lena'

(c) Denoised using Hard Thresholding

(d) Denoised using Soft Thresholding

Figure 6: Denoising using *VisuShrink*

probabilty, a pure noise signal is estimated as being identically zero.

However, for denoising images, Visushrink is found to yield an overly smoothed estimate as seen in Figure 6. This is because the universal threshold(UT) is derived under the constraint that with high probability the estimate should be at least as smooth as the signal. So the UT tends to be high for large values of M, killing many signal coefficients along with the noise. Thus, the threshold does not adapt well to discontinuities in the signal.

3.3 SureShrink

3.3.1 What is SURE ?

Let $\mu = (\mu_i : i = 1, \ldots d)$ be a length-d vector, and let $x = \{x_i\}$ (with x_i distributed as N(μ_i,1)) be multivariate normal observations with mean vector μ. Let $\widehat{\mu} = \widehat{\mu}(x)$ be an fixed estimate of μ based on the observations x. *SURE* (Stein's unbiased Risk Estimator) is a method for estimating the loss $\|\widehat{\mu} - \mu\|^2$ in an unbiased fashion.

In our case $\widehat{\mu}$ is the soft threshold estimator $\widehat{\mu}_i^{(t)}(x) = \eta_t(x_i)$. We apply Stein's result[1] to get an unbiased estimate of the risk $E\|\widehat{\mu}^{(t)}(x) - \mu\|^2$:

$$SURE(t;x) = d - 2 \cdot \#\{i : |x_i| < T\} + \sum_{i=1}^{d} min(|x_i|, t)^2.$$
(2)

For an observed vector x(in our problem, x is the set of noisy wavelet coefficients in a subband), we want to find the threshold t^S that minimizes SURE(t;x),i.e

$$t^S = \operatorname{argmin}_t SURE(t;x).$$
(3)

The above optimization problem is computationally straightforward. Without loss of generality, we can reorder x in order of increasing $|x_i|$. Then on intervals of t that lie between two values of $|x_i|$, SURE(t) is strictly increasing. Therefore the minimum value of t^S is one of the data values $|x_i|$. There are only d values and the threshold can be obtained using $O(d \log(d))$ computations.

3.3.2 Threshold Selection in Sparse Cases

The SURE principle has a drawback in situations of extreme sparsity of the wavelet coefficients. In such cases the noise contributed to the SURE profile by the many coordinates at which the signal is zero swamps the information contributed to the SURE profile by the few coordinates where the signal is nonzero. Consequently, *SureShrink* uses a Hybrid scheme.

The idea behind this hybrid scheme is that the losses while using an universal threshold, $t_d^F = \sqrt{2\log d}$, tend to be larger than SURE for dense situations, but much smaller for sparse cases.So the threshold is set to t_d^F in dense situations and to t^S in sparse situations. Thus the estimator in the hybrid method works as follows

$$\widehat{\mu}^x(x)_i = \begin{cases} \eta_{t_d^F}(x_i) & s_d^2 \le \gamma_d \\ \eta_{t^S}(x_i) & s_d^2 > \gamma_d, \end{cases} \quad (4)$$

where

$$s_d^2 = \frac{\sum_i(x_i^2 - 1)}{d} \qquad \gamma_d = \frac{\log_2^{3/2}(d)}{\sqrt{d}} \quad (5)$$

η being the thresholding operator.

3.3.3 SURE applied to image denoising

We first obtain the wavelet decomposition of the noisy image. The SURE threshold is determined for each subband using (2) and (3). We choose between this threshold and the universal threshold using (4).The expressions s_d^2 and γ_d in (5), given for $\sigma = 1$ have to suitably modified according to the noise variance σ and the variance of the coefficients in the subband.

The results obtained for the image 'Lena' (512×512 pixels) using *SureShrink* are shown in Figure 7(c). The 'Db4' wavelet was used with 4 levels of decomposition. Clearly, the results are much better than *VisuShrink*. The sharp features of the image are retained and the MSE is considerably lower. This because *SureShrink* is subband adaptive- a separate threshold is computed for each detail subband.

3.4 BayesShrink

In *BayesShrink*[5] we determine the threshold for each subband assuming a **Generalized Gaussian Distribution(GGD)** . The GGD is given by

$$GG_{\sigma_X,\beta}(x) = C(\sigma_X, \beta)exp-[\alpha(\sigma_X,\beta)|x|]^\beta \quad (6)$$

$-\infty < x < \infty, \beta > 0$, where

$$\alpha(\sigma_X,\beta) = \sigma_X^{-1}\left[\frac{\Gamma(3/\beta)}{\Gamma(1/\beta)}\right]^{1/2}$$

and

$$C(\sigma_X,\beta) = \frac{\beta \cdot \alpha(\sigma_X,\beta)}{2\Gamma(\frac{1}{\beta})}$$

and $\Gamma(t) = \int_0^\infty e^{-u}u^{t-1}du$.

The parameter σ_X is the standard deviation and β is the shape parameter It has been observed[5] that with a shape parameter β ranging from 0.5 to 1, we can describe the the distribution of coefficients in a subband for a large set of natural images.Assuming such a distribution for the wavelet coefficients, we empirically estimate β and σ_X for each subband and try to find the threshold T which minimizes the *Bayesian Risk*, i.e, the expected value of the mean square error.

$$\tau(T) = E(\widehat{X} - X)^2 = E_X E_{Y|X}(\widehat{X} - X)^2 \quad (7)$$

where $\widehat{X} = \eta_T(Y), Y|X \sim N(x,\sigma^2)$ and $X \sim GG_{x,\beta}$. The optimal threshold T^* is then given by

$$T^*(\sigma_x,\beta) = arg\min_T \tau(T) \quad (8)$$

This is a function of the parameters σ_X and β. Since there is no closed form solution for T^*, numerical calculation is used to find its value.

It is observed that the threshold value set by

$$T_B(\sigma_X) = \frac{\sigma^2}{\sigma_X} \quad (9)$$

is very close to T^*.

The estimated threshold $T_B = \sigma^2/\sigma_X$ is not only nearly optimal but also has an intuitive appeal. The

normalized threshold, T_B/σ. is inversely proportional to σ, the standard deviation of X, and proportional to σ_X, the noise standard deviation. When $\sigma/\sigma_X \ll 1$, the signal is much stronger than the noise, T_b/σ is chosen to be small in order to preserve most of the signal and remove some of the noise; when $\sigma/\sigma_X \gg 1$, the noise dominates and the normalized threshold is chosen to be large to remove the noise which has overwhelmed the signal. Thus, this threshold choice adapts to both the signal and the noise characteristics as reflected in the parameters σ and σ_X.

3.4.1 Parameter Estimation to determine the Threshold

The GGD parameters, σ_X and β, need to be estimated to compute $T_B(\sigma_X)$. The noise variance σ^2 is estimated from the subband HH_1 by the robust median estimator[5],

$$\widehat{\sigma} = \frac{Median(|Y_{ij}|)}{0.6745}, \qquad Y_{ij} \in subband HH_1 \quad (10)$$

The parameter β does not explicitly enter into the expression of $T_B(\sigma_X)$. Therefore it suffices to estimate directly the signal standard deviation σ_X. The observation model is $Y = X + V$, with X and V independent of each other, hence

$$\sigma_Y^2 = \sigma_X^2 + \sigma^2 \qquad (11)$$

where σ_Y^2 is the variance of Y. Since Y is modelled as zero-mean, σ_Y^2 can be found empirically by

$$\widehat{\sigma}_Y^2 = \frac{1}{n} \sum_{i,j=1}^{n} Y_{ij}^2 \qquad (12)$$

where $n \times n$ is the size of the subband under consideration. Thus

$$\widehat{T}_B(\widehat{\sigma}_X) = \frac{\widehat{\sigma}^2}{\widehat{\sigma}_X} \qquad (13)$$

where

$$\widehat{\sigma}_X = \sqrt{max(\widehat{\sigma}_Y^2 - \widehat{\sigma}^2, 0)} \qquad (14)$$

In the case that $\widehat{\sigma}^2 \geq \widehat{\sigma}_Y^2$, $\widehat{\sigma}_X$ is taken to be zero, i.e, $\widehat{T}_B(\widehat{\sigma}_X)$ is ∞, or, in practice, $\widehat{T}_B(\widehat{\sigma}_X) = max(|Y_{ij}|)$, and all coefficients are set to zero.

To summarize,Bayes Shrink performs soft-thresholding, with the data-driven, subband-dependent threshold,

$$\widehat{T}_B(\widehat{\sigma}_X) = \frac{\widehat{\sigma}^2}{\widehat{\sigma}_X}.$$

The results obtained by *BayesShrink* for the image 'Lena' (512×512 pixels) is shown in figure 7(d).The 'Db4' wavelet was used with four levels of decomposition. We found that *BayesShrink* performs better than *SureShrink* in terms of MSE. The reconstruction using *BayesShrink* is smoother and more visually appealing than the one obtained using *SureShrink*. This not only validates the approximation of the wavelet coefficients to the GGD but also justifies the approximation to the threshold to a value independent of β.

4 Denoising and Compression using Gaussian-based MMSE Estimation

4.1 Introduction

The philosophy of compression is that a signal typically has structural redundancies that can be exploited to yield a concise representation.White noise, however does not have correlation and is not easily compressible. Hence, a good compression method can provide a suitable method for distinguishing between signal and noise.So far,we have investigated wavelet thresholding techniques such as *SureShrink* and *BayesShrink* for denoising.We now use MMSE estimation based on a Gaussian prior and show that significant denoising can be achieved using this method. We then perform compression of the denoised coefficients based on their distribution and find that this can be done without introducing significant quantization error. Thus, we achieve simultaneous denoising and compression.

4.2 Denoising using MMSE estimation

As explained in the previous section, the *Generalized Gaussian distribution* (GGD) is a good model for the distribution of wavelet coefficients in each detail subband of the image. However, for most images, a Gaussian distribution is found to be a satisfatory approximation. Therefore, the model for the ith detail subband becomes

$$Y_j^i = X_j^i + N_j^i \qquad j = 1, 2, \cdots, M_i. \qquad (15)$$

where M_i is the number of wavelet coefficients in the ith detail subband. The coefficients $\{X_j^i\}$ are independent and identically distributed as $N(0, \sigma_{X^i}^2)$ and are independent of $\{N_j^i\}$, which are *iid* draws from $N(0, \sigma^2)$. We want to get the best estimate of $\{X_j^i\}$ based on the noisy observations $\{Y_j^i\}$. This is done through the following steps:

1. The noise variance σ^2 is estimated as described in the previous section.

2. The variance $\sigma_{Y^i}^2$ is calculated as

$$\hat{\sigma}_{Y^i}^2 = \frac{1}{n^2} \sum_{j=1}^{M_i} Y_j^{i\,2}$$

3. $\hat{\sigma}_X$ for the subband i is estimated as before as

$$\hat{\sigma}_{X^i} = \sqrt{\max(\hat{\sigma}_{Y^i}^2 - \hat{\sigma}^2, 0)}.$$

This comes about because

$$\hat{\sigma}_{Y^i}^2 = \hat{\sigma}_{X^i}^2 + \hat{\sigma}^2$$

and in the case that $\hat{\sigma}^2 \geq \hat{\sigma}_{Y^i}^2$, $\hat{\sigma}_{X^i}$ is taken to be zero. This means that the noise is more dominant than the signal in the subband and so the signal cannot be estimated with the noisy observations.

4. Based on (15), the MMSE estimate of X_j^i based on observing Y_j^i is

$$\hat{X}_j^i = E[X/Y] = \frac{\hat{\sigma}_{X^i}^2}{\hat{\sigma}_{Y^i}^2} \cdot Y_j^i \qquad (16)$$

(a) 512 × 512 image of 'Lena'

(b) Noisy version of 'Lena'

(c) Denoised using *SureShrink*

(d) Denoised using *BayesShrink*

Figure 7: Denoising by BayesShrink and SureShrink($\sigma = 30$)

We observe the similarity of this step to wavelet shrinkage, since each coefficient Y_j^i is brought closer to zero in absolute value by multiplying with $\frac{\hat{\sigma}_{X^i}^2}{\hat{\sigma}_{Y^i}^2}(< 1)$. This effect is similar to that of wavelet shrinkage in soft thresholding.

Steps 2 through 4 are repeated for each detail subband i. Note that the coefficients in the low resolution LL subband are kept unaltered.

The results obtained using this method for the 'Elaine' image with a `Db4` wavelet with 4 levels are shown in the first three parts of Figure 8.The MSE comparison plot in Figure 9 shows that denoising by Gaussian estimation performs slightly better than *SureShrink* for the 'Clock' image. The slightly inferior performance to *BayesShrink* is to be expected since a GGD prior is a more exact representation of the wavelet coefficients in a subband than the Gaussian prior.

4.3   Compression

We now introduce a quantization scheme for a concise representation of the denoised coefficients $\{\hat{X}_j^i\}$. From (16), the $\{\hat{X}_j^i\}$ are *iid* with distribution $N(0, \frac{\hat{\sigma}_{X^i}^4}{\hat{\sigma}_{Y^i}^2})$. The number of bits used to encode each coefficient $\hat{X}_j^i$ is determined as follows. For simplicity of notation , we denote $\hat{X}_j^i$ as A_j, keeping in mind that A_j is a part of subband i

1. We first fix the maximum allowable distortion, say D, for each coefficient.

2. The variance of each coefficient A_j is found empirically by calculating the variance of a 3×3 block of coefficients centered at A_j.

   It is assumed that we have available a finite set of optimal Lloyd Max quantizers for the $N(0, 1)$ distribution. In our experiments, we took 5 quantizers with number of quantization levels $M = 2, 4, 8, 16$ and 32.

3. Each coefficient A_j is encoded using the quantizer with the least M so that $(A_j - \hat{A}_j)^2 \leq D$. Note that both D and the quantizer levels, defined for $N(0, 1)$ have to scaled by σ_{A_j} for each coefficient A_j.

4. Steps 2 and 3 are repeated for all the coefficients A_j in a subband and for all the detail subbands.

5. The coeffients in the low resolution subband are quantized assuming a uniform distribution [5]. This is motivated by the fact that the LL coefficients are essentially local averages of the image and are not characterized by a Gaussian distribution.

4.4   Results

Figure 8 shows the results obtained when this denoising and compression scheme is applied to the image 'Elaine' with $\sigma = 30$.We used `Db-4` discrete wavelet series with 4 levels of decomposition.We see the denoised version has much lower MSE (143.7 vs $\sigma^2 = 900$)and better visual quality too. The compressed version looks very similar to the denoised image with an additional MSE of around 20. It has been encoded using 1.52 bpp (distortion value D set at=0.1). The rate can be controlled by changing the distortion level D. If we fix a large distortion level D, we get a low encoding rate, but have a price to pay-larger quantization error. We choose to operate at a particular point on the 'Rate v Distortion' curve based on the distortion we are prepared to tolerate.

The performance of the different denoising schemes is compared in Figure 9. A 200×200 image 'Clock' is considered and the MSEs for different values of σ are compared. Clearly, *VisuShrink* is the least effective among the methods compared. This is due to the fact that it is based on a Universal threshold and not subband adaptive unlike the other schemes. Among these, *BayesShrink* clearly performs the best. This is expected since the GGD models the distribution of coefficients in a subband well. MMSE estimation based on a Gaussian distribution performs slightly worse than *BayesShrink*. We also see that a quantization error(approximately constant) is introduced due to compression. Among the subband

(a) 200 × 200 image of 'Elaine'

(b) Noisy version of 'Elaine'

(c) Denoised version of 'Elaine'

(d) Quantized image of 'Elaine'

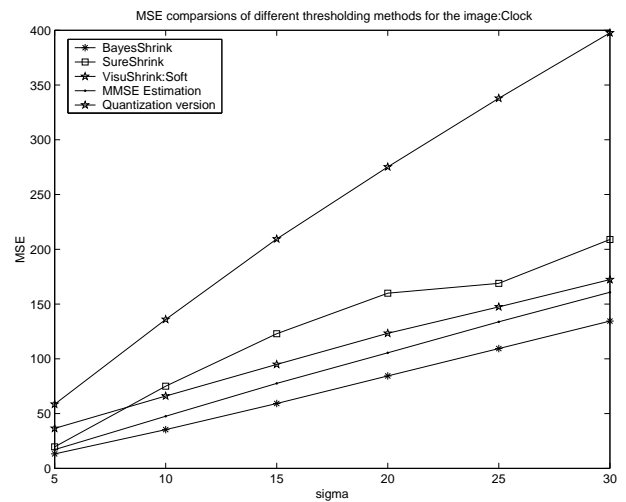Figure 8: MMSE Denoising and Quantization

Figure 9: Comparison of MSE of various denoising schemes

adaptive schemes, *SureShrink* has the highest MSE. But it should be noted that *SureShrink* has the desirable property of adapting to the discontinuities in the signal. This is more evident in 1-D signals such as 'Blocks' than in images.

5 Conclusions

We have seen that wavelet thresholding is an effective method of denoising noisy signals. We first tested hard and soft on noisy versions of the standard 1-D signals and found the best threshold. We then investigated many soft thresholding schemes viz. *VisuShrink, SureShrink* and *BayesShrink* for denoising images. We found that subband adaptive thresholding performs better than a universal thresholding. Among these, *BayesShrink* gave the best results. This validates the assumption that the GGD is a very good model for the wavelet coefficient distribution in a subband. By weakening the GGD assumption and taking the coefficients to be Gaussian distributed, we obtained a simple model that facilitated both denoising and compression.

An important point to note is that although

SureShrink performed worse than *BayesShrink* and Gaussian based MMSE denoising, it adapts well to sharp discontinuities in the signal. This was not evident in the natural images we used for testing. It would be instructive to compare the performance of these algorithms on artificial images with discontinuities (such as medical images). It would also be interesting to try denoising (and compression) using other special cases of the GGD such as the Laplacian (GGD with $\beta = 1$).Most images can be described with a GGD with shape parameter β ranging from 0.5 to 1. So a Laplacian prior may give better results than a Gaussian prior ($\beta = 2$) although it may not be as easy to work with.

References

[1] Iain M.Johnstone David L Donoho. Adapting to smoothness via wavelet shrinkage. *Journal of the Statistical Association*, 90(432):1200–1224, Dec 1995.

[2] David L Donoho. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, August 1994.

[3] David L Donoho. De-noising by soft thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, May 1995.

[4] Maarten Jansen. *Noise Reduction by Wavelet Thresholding*, volume 161. Springer Verlag, United States of America, 1 edition, 2001.

[5] Martin Vetterli S Grace Chang, Bin Yu. Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on Image Processing*, 9(9):1532–1546, Sep 2000.

[6] Carl Taswell. The what, how and why of wavelet shrinkage denoising. *Computing in Science and Engneering*, pages 12–19, May/June 2000.