

1-1-2008

Data Denoising By Noise Invalidation

Nima Nikvand
Ryerson University

Follow this and additional works at: <http://digitalcommons.ryerson.ca/dissertations>



Part of the [Electrical and Computer Engineering Commons](#)

Recommended Citation

Nikvand, Nima, "Data Denoising By Noise Invalidation" (2008). *Theses and dissertations*. Paper 1172.

This Thesis is brought to you for free and open access by Digital Commons @ Ryerson. It has been accepted for inclusion in Theses and dissertations by an authorized administrator of Digital Commons @ Ryerson. For more information, please contact bcameron@ryerson.ca.

01001-12
TK
7867.5
N55
2008

DATA DENOISING BY NOISE INVALIDATION

by

Nima Nikvand

Bachelor of Science, Isfahan University of Technology, 2005

A thesis

presented to Ryerson University
in partial fulfillment of the
requirement for the degree of
Master of Applied Science

in the Program of
Electrical and Computer Engineering
Toronto, Ontario, Canada, 2008

© Nima Nikvand 2008

Author's Declaration

I hereby declare that I am the sole author of this thesis.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Abstract

Data Denoising by Noise Invalidation

© Nima Nikvand, 2008

Master of Applied Science (MASc)

Department of Electrical and Computer Engineering

Ryerson University

In this thesis, the problem of data denoising is studied, and two new denoising approaches are proposed. Using statistical properties of the additive noise, the methods provide adaptive data-dependent soft thresholding techniques to remove the additive noise. The proposed methods, Point-wise Noise Invalidating Soft Thresholding (PNIST) and Accumulative Noise Invalidation Soft Thresholding (ANIST), are based on *Noise Invalidation*. The invalidation exploits basic properties of the additive noise in order to remove the noise effects as much as possible. There are similarities and differences between ANIST and PNIST. While PNIST performs better in the case of additive white Gaussian noise, ANIST can be used with both Gaussian and non Gaussian additive noise. As part of a data denoising technique, a new noise variance estimation is also proposed. The thresholds proposed by NIST approaches are comparable to the shrinkage methods, and our simulation results promise that the new methods can outperform the existing approaches in various applications.

We also explore the area of image denoising as one of the main applications of data denoising and extend the proposed approaches to two dimensional applications. Simulations show that the proposed methods outperform common shrinkage methods and are comparable to the famous BayesShrink method in terms of Mean Square Error and visual quality.

Dedication

Not a single page of this thesis would have been possible without the help, support and mentorship of Dr Soosan Beheshti. She helped me when I needed help the most, and her humane and friendly behavior with her students will always be my model.

I chose this topic under her supervision; her knowledge and creative mind was an invaluable source of inspiration in solving various problems I faced during the research. I will always be proud to have worked under her supervision, and this thesis is dedicated to her.

Acknowledgments

This thesis would not have been possible without the help of my parents. My father, Jahansha, has always been the main source of support and confidence in my life and I have relied on his constant encouragement and love throughout my academic career. His love and sacrifice for me will always be a great source of inspiration in my life. I thank My mother, Forough, whose unconditional love, support and courage helped me withstand hardship and make the best decisions in my life. I will always be thankful to both of them, and I don't believe I will ever be able to love anyone as much as they love me.

Table of Contents

Abstract	iii
Acknowledgments	v
1 Introduction	1
2 Background	5
2.1 Wavelet Transform	6
2.1.1 Continuous Wavelet Transform (CWT)	6
2.1.2 Discrete Wavelet Transform (DWT)	7
2.2 Wavelet Shrinkage	10
2.3 Data Denoising	12
2.3.1 VisuShrink	12
2.3.2 SureShrink	12
2.3.3 BayesShrink	13
2.3.4 Minimum Noiseless Description Length (MNDL)	15
2.4 Concluding Remarks	17
3 Thresholding Method	18
3.1 Problem Statement	18
3.1.1 Quality of the Thresholding Approaches	19
3.2 Accumulative Noise Invalidating Soft Thresholding (ANIST) Method	20
3.2.1 Expected Value and Variance of Noise Effects in the Absence of Data	20
3.2.2 Noise Statistics in ANIST	21
3.2.3 Noise Invalidation in ANIST	22
3.2.4 Selection of a Proper λ	25
3.2.5 Simulation Results for ANIST	27
3.3 Point-wise Noise Invalidating Soft Thresholding (PNIST) Method	31
3.3.1 Expected Value and Variance of Noise Effects in the Absence of Data	31
3.3.2 Noise Statistics in PNIST	32
3.3.3 Noise Invalidation in PNIST	36
3.3.4 Simulation Results for PNIST	38
3.4 Concluding Remarks	40
4 Additive Noise in Denoising	43
4.1 Noise Estimation in MNDL Approach	44
4.2 Proposed Noise Variance Estimation Method	46
4.3 Noiseless Data with Gaussian Properties	49
4.3.1 Conditional Distributions and Data Denoising	51

4.4 Concluding Remarks	55
5 Image Denoising	56
6 Conclusion	61
Appendices	63
A Distribution of Order Statistics of a Random Variable	63
B Gaussian Data with Additive Noise	65
References	66

List of Tables

3.1	MSE by different λ values.	26
3.2	Proposed Thresholds by each method for Blocks test signal. Optimum threshold is the best threshold possible.	29
3.3	MSE for different Methods compared.	31
3.4	MSE for different Methods compared.(1)ANIST (2) PNIST.	38
3.5	MSE by Method for Blocks signal contaminated by uniform noise.	42
4.1	Estimated variance by the proposed method.	51
5.1	MSE for different Methods compared.	57

List of Figures

2.1	Short Time Fourier Transform windows, narrow in time domain means wide in frequency domain and the opposite. Image source:[27].	6
2.2	Block diagram of filters equivalent to wavelet decomposition. Image source:[42].	9
2.3	Block diagram of filters banks for multilevel wavelet decomposition. Image source:[42].	9
2.4	Four level Decomposition of Einstein.	10
2.5	Left: Hard thresholding, Right: Soft thresholding.	11
3.1	Histogram of noise for $\sigma^2 = 1$ at Top left: $m = 1$, Top right: $m = 2$, Bottom left $m = 3$, Bottom right $m = 4$.	22
3.2	Histogram of noise for $\sigma^2 = 1$ at Top left: $m = 10$, Top right: $m = 20$, Bottom left: $m = 100$, Bottom right: $m = 1000$.	23
3.3	Noise behavior for $\sigma^2 = 1$ Top left: Histogram at $m = 10$, Bottom left: Norm plot at $m = 10$, Top right: Histogram at $m = 20$, Bottom right: Norm plot at $m = 20$.	23
3.4	Noise behavior for $\sigma^2 = 1$ Top left: Histogram at $m = 100$, Bottom left: Norm plot at $m = 100$, Top right: Histogram at $m = 1000$, Bottom right: Norm plot at $m = 1000$.	24
3.5	$E(\Psi_m^v)$ and upper and lower bounds for $\lambda = 5$, $\sigma^2 = 1$.	25
3.6	ψ_m of contaminated Blocks signal crossing upper bound of Invalidation region @ $m^* = 956$ in decimal plot. Shaded area is Gaussian noise region with $\sigma^2 = 4$.	26
3.7	Test signals and their noisy forms. Top left: Blocks signal, Top right: Mishmash signal, Bottom left: Noisy Blocks with $\sigma^2 = 1$, Bottom right: Noisy Mishmash with $\sigma^2 = 1$.	28
3.8	Test signals and their wavelet coefficients. Top left: Blocks signal, Top right: Mishmash signal, Bottom right: Blocks coefficients, Bottom left: Mishmash coefficients.	28
3.9	ψ_m and $E(\psi_m^v)$ for Blocks, $\sigma^2 = 1$. The cross point is at $m^* = 12$.	29
3.10	ψ_m and $E(\psi_m^v)$ for Blocks, $\sigma^2 = 4$. The cross point is at $m^* = 986$.	30
3.11	Top: $E(\psi_m^v)$ and ψ_{sm} for Mishmash, $\sigma^2 = 1$, Bottom: zoomed version of top figure. The cross point is at $m^* = 22$.	30
3.12	Histogram of noise for $\sigma^2 = 1$ at Top left: $m = 1$, Top right: $m = 2$, Bottom left: $m = 3$, Bottom right: $m = 4$ for PNIST.	33
3.13	Histogram of noise for $\sigma^2 = 1$ at Top left: $m = 10$, Top right: $m = 20$, Bottom left: $m = 100$, Bottom right: $m = 1000$, for PNIST.	33
3.14	Histogram of noise for $\sigma^2 = 1$ at Top left: $m = 10$, Bottom left: Norm plot at $m = 10$, Top right: $m = 20$, Bottom left: Norm plot $m = 20$ for PNIST.	34

3.15	Histogram of noise for $\sigma^2 = 1$ at Top left: $m = 100$, Bottom left: Norm plot at $m = 100$, Top right: $m = 1000$, Bottom right: Norm plot $m = 1000$ for PNIST.	34
3.16	Probability distribution function of Γ_m^z at $m = 20$ compared with Gaussian probability distribution function of the same mean and variance.	35
3.17	Probability distribution function of Γ_m^z at $m = 100$ compared with Gaussian probability distribution function of the same mean and variance.	36
3.18	$E(\Gamma_m^z)$ and its upper and lower bounds for $\lambda = 5$	38
3.19	γ_m of contaminated Blocks signal crossing upper bound of Invalidation region @ $m^* = 700$ in decimal plot. Shaded area is Gaussian noise region with $\sigma^2 = 4$	39
3.20	PNIST: γ_m and upper bound for noise average in case of Blocks, $m^* = 862$, $\sigma^2 = 1$	39
3.21	PNIST: γ_m and upper bound for noise average in case of Mishmash, $m^* = 8$	40
3.22	Uniform noise with $\sigma^2 = 4$	41
3.23	ψ_m^v histogram and normplot for $\sigma^2 = 4$ at $m = 100$ and $m = 1000$. Note that sample data falls exactly on the straight line.	41
4.1	Data description length versus $\hat{\sigma}$ in MNDL method when the true variance is $\sigma^2 = 2$	45
4.2	Autocovariance of noise.	47
4.3	Autocovariance of noisy signal.	47
4.4	Autocovariance graph for different test variances, from left to right: (a) $\sigma^2 = 1$ (b) $\sigma^2 = 4$ (c) $\sigma^2 = 9$ (d) $\sigma^2 = 16$. True noise variance is $\sigma^2 = 9$	48
4.5	Area under autocovariance of discarded coefficients for true variance of $\sigma^2 = 4$	50
4.6	Area under autocovariance of discarded coefficients for true variance of $\sigma^2 = 8$	50
4.7	Histogram of Mishmash coefficients.	52
4.8	Noisy Gaussian signal, noisy Mishmash signal and Gaussian noise compared with 2^{10} samples.	52
4.9	Noisy Gaussian signal, noisy Mishmash signal and Gaussian noise compared with 2^{20} samples.	53
4.10	Distribution of $f(n y = a)$ & $f(x y = a)$	54
5.1	Barbara denoised by ANIST and PNIST, $\sigma_n^2 = 225$	57
5.2	Cameraman denoised, $\sigma_n^2 = 100$	58
5.3	Lena denoised, $\sigma_n^2 = 400$	59
5.4	Einstein denoised, $\sigma_n^2 = 900$	60

Chapter 1

Introduction

Data denoising is a well-known problem that is the concern of diverse application areas. For example, for generations, communication system designers have been trying to implement techniques to get rid of the natural nuisance and interference that corrupts the message signal; Thermal noise exists in all electronic elements, furthermore, channel introduces degradation to message signal, and defective instruments and data acquisition systems affect the quality of the observed data as well. Thus denoising is a necessary process in the receiver of all modern communication systems. Linear filters like *Wiener filters* are among the most simply implemented denoising methods [21]. More complicated methods may introduce nonlinearity to Wiener filters [44] or use other nonlinear techniques such as *Median filters* [44, 24]. A relatively new approach is called *Wavelet Shrinkage* and is based on rejecting those wavelet coefficients of the observed data that are smaller than a certain threshold value [10]. The principle behind this method is that wavelet transforms enable us to represent the signal with a high degree of sparsity [10]. Sparsity of wavelet coefficients, meaning that most of the signal information in wavelet expansion is conveyed by a relatively small number of large coefficients, helps us find an estimate of the signal after rejecting some of the coefficients considered as noise and yet keeping most of the information carried by the signal in the remaining coefficients [10, 5, 12]. Using wavelet transforms as a tool for denoising of signals has drawn a lot of research attention in the past decade leading to the famous shrinkage algorithms [10, 8, 3, 36, 2]. The first method proposed based on Wavelet Shrinkage was *Visu Shrink* [10]. VisuShrink proposed by

Donoho and Johnstone applies a universal threshold that depends only on noise variance and data length. SureShrink, proposed by the same authors, minimizes *Stein's Unbiased Risk Estimator* (SURE) to find a threshold [8]. In *BayesShrink*, an important image denoising approach, a threshold that minimizes *Bayesian Risk* is found based on the assumption that the image has the properties of a *Generalized Gaussian* signal [3]. Another recently introduced thresholding technique is called *Minimum Description Length* (MDL) [36]. MDL chooses a threshold that minimizes *Normalized Maximum Likelihood* (NML) of the noisy data in each subspace, and *Minimum Noiseless Description Length* (MNDL) suggests choosing a subspace which minimizes description length of the noiseless data [2].

In this thesis we propose two new thresholding methods based on *Noise Invalidation*. It is shown that a proper threshold can be found based on Gaussian invalidation of coefficients or a function of coefficients out of the noise region. In chapter two a brief review on one dimensional and two dimensional wavelet transforms and a survey of related shrinkage methods are provided. The Shrinkage methods, originally proposed by Donoho and Johnson [10, 8], play a key role in data denoising. Some of these methods are designed to apply a threshold based on limit theorem and some others assume a prior for the noiseless data and minimize the error using the prior such as bayeshrink[3]. BayesShrink, a thresholding method originally proposed for Image denoising applications, is applied to one dimensional signals in this research and it shows acceptable results for those signals that follow a Gaussian distribution. Working with noisy part of the signal in MNDL inspired us to examine the statistical properties of noise and noisy coefficients of a sorted signal, which led to the introduction of two novel soft thresholding methods in chapter three.

Chapter three introduces the two novel thresholding techniques and compares the results to other shrinkage methods. Based on statistical properties of noise and noisy parts of the signal, it is shown that a Gaussian approximation can be used to invalidate signal coefficients or a function of sum of square added coefficients of the observed signal at each point. Thanks to the fact that we are using a summation of weakly independent random variables in one of our methods, we can benefit from results of Central Limit Theorem for weakly dependent variables and apply the method for non Gaussian noise situations as well. Chapter four proposes a novel noise variance estimation method. Estimation of noise variance is crucial to any denoising algorithm and the literature on this subject is limited. While most methods pick the fine scale coefficients through an ad-hoc threshold to estimate noise, the novel method proposed in this research uses autocovariance properties of Gaussian noise combined with sparsity of wavelet coefficients to propose estimate of the noise variance. In chapter four, properties of Gaussian signals are also studied and it is shown that a large class of signals, like Mishmash signal follow Gaussian distribution and have similar behavior in the eye of the two proposed methods in chapter three. Using this fact, we propose a data and subband rejection criterion that ensures the share of noiseless signal is larger than the share of noise in the observed set of data with at least fifty percent probability. This gives us a data rejection criterion which is more conservative than that of BayesShrink. The proposed criterion ensures that a set of observed data or a subband of the data is worth going through the rest of the denoising procedure and can be used as a pre-processing condition. Chapter five applies the proposed thresholding techniques to images and compares the results with the mentioned methods. In this chapter, wavelet transform of noisy images are found and after pre-processing with the novel data rejection criterion proposed in chapter four, a global threshold is

found through either of the proposed thresholding methods. This threshold is then applied to soft thresholding function and the smoothed data is then taken for inverse wavelet transform and the denoised Image is reconstructed. The rest of the chapter compares other shrinkage methods in terms of mean square error and visual quality to the proposed methods and performance of the methods are compared for standard Images such as Barbara, Cameraman, Einstein and Lena.

Chapter 2

Background

The main idea of Data Denoising is to find a basis which represents the data better than the noisy data itself. One of the useful orthonormal bases is *Fourier Transform (FT)*. Fourier transform provides us with frequency components of a signal which prove to be invaluable in many applications including, linear time invariant analysis. However, Fourier transform is unable to provide information in time and frequency domain simultaneously. This means that Fourier transform provides us with the frequency components, but it does not tell us when they happen in time domain. This is not an issue in many applications where only frequency components are needed such as speech processing. However, in some applications such as Image processing, we need to have time and frequency information of the signal simultaneously. To solve this problem, a general idea would be to divide the signal into different parts and analyze frequency components of these parts separately. One of the early solutions to this problem was *Short Term Fourier Transform (STFT)* [15]. STFT proposes moving a window across the signal to analyze the frequency domain of the signal for each part. The problem still remains since STFT has a fixed resolution. This property is due to Heisenberg's uncertainty principle, which states the product of the standard deviation in time and frequency is limited. Thus measuring frequency and time can not be done simultaneously at a desired resolution. A wide window gives better frequency resolution but poor time resolution. A narrower window, on the other hand, gives good time resolution but poor frequency resolution.

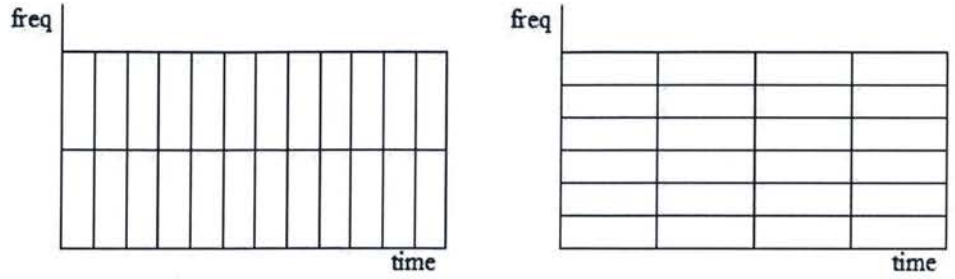


FIGURE 2.1. Short Time Fourier Transform windows, narrow in time domain means wide in frequency domain and the opposite. Image source:[27].

2.1 Wavelet Transform

Wavelet Transform (WT) was proposed as a solution to this problem. Wavelet transform uses a scalable modulated window to move across the signal. Using a small scale for high frequency parts and a big scale for low frequency parts enables wavelet transform to provide a good time and frequency resolution. This cycle is repeated and each time a slightly smaller or bigger window is selected. The result of the process will be a set of time frequency representations of the signal with different resolutions. These representations are called scale-time representations as opposed to frequency representations in Fourier transform.

2.1.1 Continuous Wavelet Transform (CWT)

Continuous wavelet transform is defined as:

$$\gamma(s, \tau) = \int_{-\infty}^{\infty} f(t) \Psi_{s,\tau}^*(t) dt \quad (2.1)$$

Where $\Psi_{s,\tau}$ is a dilated and translated wavelet function made of the mother wavelet function as the following:

$$\Psi_{s,\tau} = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \quad (2.2)$$

where $\psi(t)$ is the mother wavelet function. Finally the Inverse CWT is defined as:

$$f(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \gamma(s, \tau) \Psi_{s,\tau}^*(t) d\tau ds \quad (2.3)$$

The two main properties of wavelet functions are admissibility and regularity. It can be shown that a square integrable function such as $\Psi(t)$, can be used to decompose and then reconstruct a signal when $\Psi(t)$ meets the admissibility condition stated as follows[38]:

$$\int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty \quad (2.4)$$

admissibility condition implies:

$$|\Psi(\omega)|_{\omega=0}^2 = 0 \quad (2.5)$$

This also requires the mother wavelet signal to have a zero average across time. In other words¹:

$$\int_{-\infty}^{\infty} \Psi(t) dt = 0 \quad (2.6)$$

Regularity means the wavelet function should have smoothness and concentration in both frequency and time domains [27, 28].

2.1.2 Discrete Wavelet Transform (DWT)

To implement wavelet transform, we have to overcome some barriers. The first and the most important of these barriers is redundancy of CWT. Wavelet transform is calculated by constantly shifting a continuously scalable function over the signal and finding the correlation with different parts of the signal. It is shown that these scalable functions are not orthogonal [17, 11]. Thus the resulting coefficients are redundant. In practical applications this redundancy must be removed. Another problem is that for many signals wavelet transform has no analytical solution and must be calculated numerically. To overcome these problems, *Discrete Wavelet*

¹This is where the name wavelet is coming from as zero average across time domain means being oscillatory or a *wave*

Transform (DWT) was proposed. DWT is achieved by changing the wavelet of CWT in equation 2.2 to:

$$\Psi_{j,k}(t) = \frac{1}{\sqrt{s_0^j}} \Psi\left(\frac{t - k\tau_0 s_0^j}{s_0^j}\right) \quad (2.7)$$

DWT reduces the number of wavelets and ensures that under certain conditions, the resulting coefficients can be used to reconstruct the original data[6].

To remove redundancy, wavelets can be selected as orthogonal to their own scales and shifted versions. Thus we choose:

$$\int_{-\infty}^{\infty} \Psi_{j,k}(t) \Psi_{m,n}^*(t) dt = \begin{cases} 1 & \text{if } j = m \text{ \& } k = n, \\ 0 & \text{if } j \neq m \text{ or } k \neq n. \end{cases} \quad (2.8)$$

Discrete wavelet coefficients are found:

$$\gamma(j, k) = \int_{-\infty}^{\infty} f(t) \Psi_{j,k}^*(t) dt \quad (2.9)$$

and the original signal can be reconstructed:

$$f(t) = \sum_{j,k} \gamma(j, k) \Psi_{j,k}(t) \quad (2.10)$$

Wavelets have band pass spectrum in frequency domain, and from Fourier transform we know that dilation in time domain will act as compression in frequency domain. This gives us the idea to implement DWT as a bank of filters. DWT of a digital signal x can be found as the following:

$$y_{low}[n] = \sum_{k=-\infty}^{\infty} x[k] g[2n - k] \quad (2.11)$$

$$y_{high}[n] = \sum_{k=-\infty}^{\infty} x[k] h[2n - k] \quad (2.12)$$

Where h is impulse response of a high-pass filter and g is impulse response of a low-pass filter. These filters are quadrature mirror filters of the original bandpass

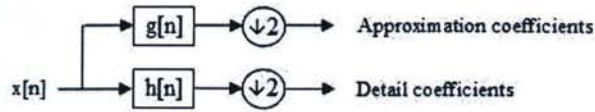


FIGURE 2.2. Block diagram of filters equivalent to wavelet decomposition. Image source:[42].

filter of the wavelet [27, 42]. Figure 2.2 shows the implementation of filters for one level decomposition.

For a multi level decomposition we can implement a filter bank shown in Figure 2.3. Signal is passed through each stage of low pass and high pass filters and the result is subsampled by two at the end of the stage. The high pass filter output, which is called detail subband, is the result of decomposition at each stage. Output of the low pass filter which is called approximate subband is then taken for further filtering stages.

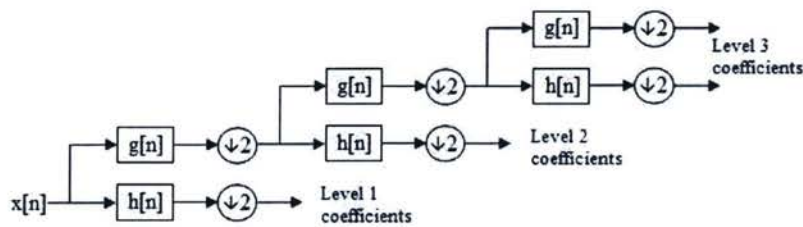


FIGURE 2.3. Block diagram of filters banks for multilevel wavelet decomposition. Image source:[42].

A similar process is used for images as for two dimensional data. It consists of alternating one decomposition by rows and another one by columns, repeating the procedure on the low-pass sub-image[27, 28, 30, 42]. Thus an image produces four subbands at every level. Figure 2.4 shows a four level decomposition of Einstein using "Haar" wavelet.

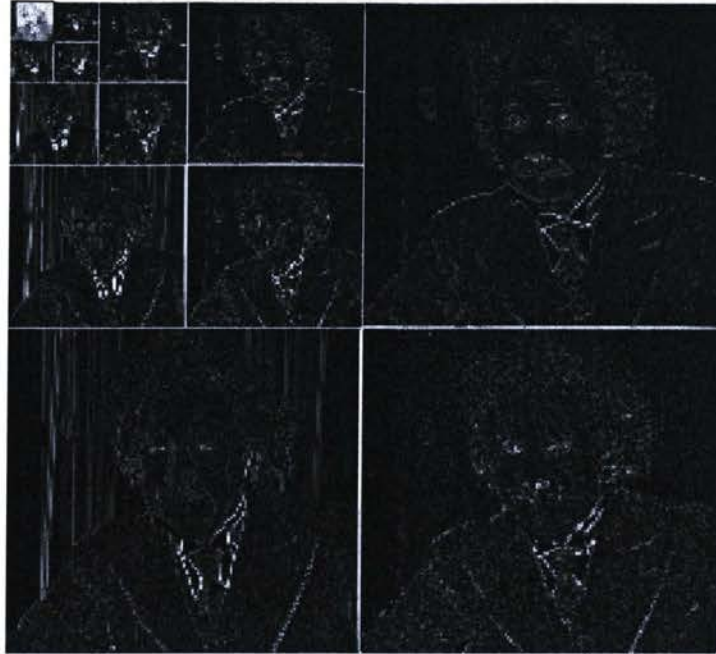


FIGURE 2.4. Four level Decomposition of Einstein.

In this thesis, we concentrate on using the wavelet transform as an orthonormal basis, however, the study can be generalized to any other orthonormal basis such as Fourier transform.

2.2 Wavelet Shrinkage

Wavelet Shrinkage (WS) is a nonlinear and nonparametric signal denoising technique[10]. Based on the sparsity of wavelet coefficients, we know that most of the information in a signal is carried by a small number of large coefficients, and the rest of the coefficients carry little or no information[10, 30]. Thus by selecting orthogonal wavelets, this method attempts to discard the small coefficients in the orthogonal transformation of data that tend to be attributed to noise.

There are two major thresholding techniques, namely Soft and Hard thresholding. Hard thresholding is simply setting to zero all the coefficients that are smaller than a certain value (threshold). Soft thresholding is an extension of hard thresholding which sets the coefficients smaller than the threshold to zero and reduces

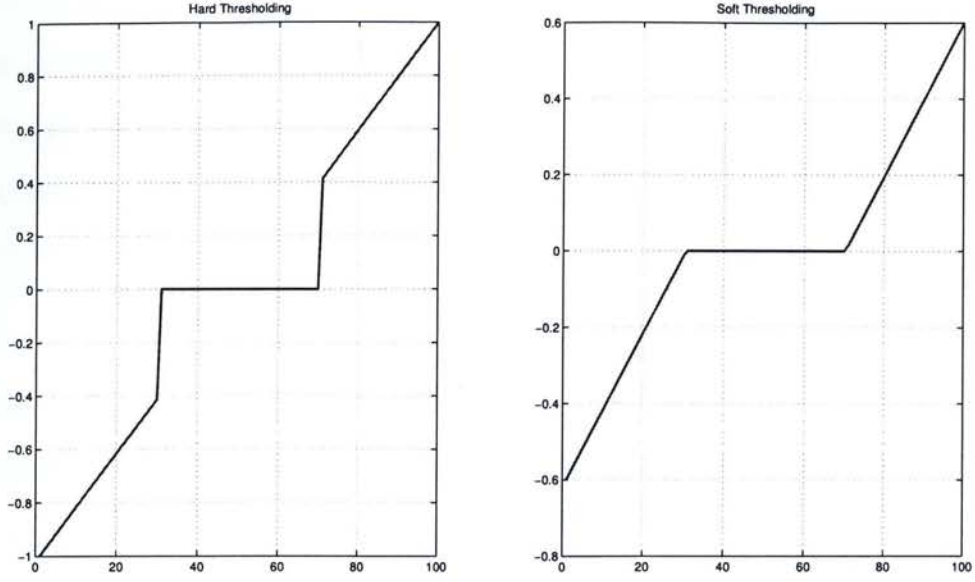


FIGURE 2.5. Left: Hard thresholding, Right: Soft thresholding.

the remaining coefficients by the threshold value. Figure 2.5 depicts output of hard and soft thresholding techniques.

Hard thresholding is defined as:

$$\eta_{T_h}(\theta) = \begin{cases} \theta & \text{if } |\theta| \geq T_h, \\ 0 & \text{if } |\theta| < T_h \end{cases} \quad (2.13)$$

and soft thresholding is:

$$\eta_{T_s}(\theta) = \begin{cases} \text{sgn}(\theta) ||\theta| - T_s| & \text{if } |\theta| \geq T_s, \\ 0 & \text{if } |\theta| < T_s \end{cases} \quad (2.14)$$

Some denoising methods use hard thresholding, and some are proper for soft thresholding. Hard thresholding, however, is not used for applications such as image denoising since it introduces artifacts to the result and is not appealing in terms of visual quality.

2.3 Data Denoising

Success of the thresholding technique in different applications depends on the choice of threshold. Thus choosing the proper threshold is crucial to a successful denoising. In the following sections some of the most famous threshold selection methods are reviewed.

2.3.1 VisuShrink

VisuShrink also known as *Universal Threshold* is the best known approach suggested by Donoho and Johnstone [10]. VisuShrink proposes a threshold that is dependent on the length of the data:

$$T_{visu} = \sigma_n \sqrt{2 \log N} \quad (2.15)$$

Where σ_n^2 is noise variance and N is the Data length.

However, VisuShrink is known to produce over smoothed results due to the fact that the threshold is dependent on data length and increases with number of samples under the same noise variance[18].

2.3.2 SureShrink

Another well known method proposed by Donoho is based on minimizing *Stein's Unbiased Risk Estimator (SURE)* and is called SureShrink[8]. SURE is an estimator that is used to find an unbiased estimate of $\|\hat{\mu} - \mu\|^2$ where $\mu = (\mu_i : i = 1, \dots, N)$ is a vector of length N and $Y = \{y_i\}$ ($y_i \sim N(0, 1)$) is a vector of multivariate normal observations with mean vector μ . In our case $\hat{\mu} = \hat{Y}$ and $\mu = Y_{orig}$. The risk for $E(\|\hat{Y} - Y_{orig}\|^2)$ is found to be:

$$SURE(T, y) = N - 2 \cdot \#\{i : |y_i| \leq T\} + \sum_{i=1}^N \min(|y_i|, T)^2 \quad (2.16)$$

and the threshold is selected as:

$$T_{Sure} = \arg \min_{0 < T < \sqrt{2 \log(N)}} SURE(T, y) \quad (2.17)$$

SureShrink is a subband dependent thresholding technique, which means it proposes a different threshold for each subband of data. Thus N in equation (2.17) is the number of coefficients in the corresponding subband. However, the threshold proposed by SURE has serious drawbacks in sparse subbands, thus a hybrid threshold of SURE and Visu is applied:

$$\hat{\theta} = \begin{cases} \eta_{T_{Sure}}(\theta) & \text{if } \zeta_N^2 \geq \gamma_N \\ \eta_{T_{Visu}}(\theta) & \text{if } \zeta_N^2 < \gamma_N \end{cases} \quad (2.18)$$

where $\zeta_N^2 = \frac{\sum_{j=1}^N (\theta_j^2 - 1)}{N}$ and $\gamma_N = \frac{\log_2(N)^{2/3}}{\sqrt{N}}$. Hybrid thresholding solves one problem to add another as SureShrink risks the over smoothing problem in VisuShrink due to hybrid thresholding.

2.3.3 BayesShrink

Different denoising methods assume different distributions for the noiseless data or its subbands[29, 30, 4, 3, 27]. The choice of a prior usually helps solving the optimum threshold problem through a Maximum A Posteriori (MAP) strategy or minimizing the risk function[19, 22, 20, 30, 4, 3]. In BayesShrink, proposed by Chang, we assume a *Generalized Gaussian Distribution (GGD)* prior for the noiseless data and find a threshold for each subband[3].

The GGD is given by:

$$GGD_{\sigma_X, \beta}(x) = C(\sigma_X, \beta) e^{-[\alpha(\sigma_X, \beta)|x|]^\beta} \quad (2.19)$$

where $-\infty < x < \infty$ and $\beta > 0$ with:

$$\alpha(\sigma_X, \beta) = \sigma_X^{-1} \left[\frac{\Gamma(3/\beta)}{\Gamma(1/\beta)} \right]^{\frac{1}{2}} \quad (2.20)$$

$$C(\sigma_X, \beta) = \frac{\beta \cdot \alpha(\sigma_X, \beta)}{2\Gamma(\frac{1}{\beta})} \quad (2.21)$$

and

$$\Gamma(t) = \int_0^\infty e^{-u} u^{t-1} du. \quad (2.22)$$

Assuming β (which is called Shape Parameter) has a range of 0.5 to 1, it has been observed that most noiseless images follow this distribution[3, 39, 40, 34]. Using a GGD prior, we can now minimize *Bayesian Risk* which is defined as the expected value of the mean square error[37, 43, 3]:

$$\tau(T) = E(\hat{X} - X)^2 = E_X E_{Y|X}(\hat{X} - X)^2 \quad (2.23)$$

where \hat{X} is estimate of noiseless data ($\hat{X} = \eta_T(y)$) and Y is the observed data ($Y|X \sim N(x, \sigma^2)$) with $X \sim GGD_{\sigma_X, \beta}$. The optimal threshold is found as:

$$T_{Bayes} = \arg \min_T \tau(T) \quad (2.24)$$

However, there is no close form solution for equation (2.24) thus; numerical solutions are used, and it is observed that:

$$T^* = \frac{\sigma_n^2}{\sigma_{\hat{\theta}}} \quad (2.25)$$

is closest to the optimum [3].

We need to have σ_n and $\sigma_{\hat{\theta}}$ to find T^* . Noise variance σ_n^2 is usually estimated using Robust Median Estimator:

$$\sigma_n(\theta) = \frac{\text{Median}|\Theta_D|}{0.675} \quad |\Theta_D| \in \text{HH1} \quad (2.26)$$

Variance of the observed data is estimated as:

$$\sigma_{\hat{\theta}}^2 = \frac{1}{N} \sum_{i,j=1}^N \theta(i,j)^2 \quad (2.27)$$

with N being data length or number of pixels in the considered subband. Since the problem assumes noiseless data, X , to be independent of noise we have:

$$\sigma_\theta^2 = \sigma_{\bar{\theta}}^2 + \sigma_n^2 \quad (2.28)$$

thus:

$$\sigma_{\bar{\theta}} = \sqrt{\sigma_\theta^2 - \sigma_n^2} \quad (2.29)$$

and by replacing the results from equations (2.29) and (2.26) in equation (2.25), we can find the closest proposed threshold to the optimum.

2.3.4 Minimum Noiseless Description Length (MNDL)

Minimum Noiseless Description Length (MNDL) is a new approach to subspace selection proposed by Beheshti and Dahleh[2]. MNDL's idea was inspired by Minimum Description Length(MDL)[36] However, in MNDL, length of noiseless data is estimated in every subspace, and the subspace with minimum noiseless data description length is selected as the optimum subspace.

When Gaussian noise is added to the noiseless data vector $\bar{y}^N = [\bar{y}(1), \bar{y}(2), \dots, \bar{y}(N)]^T$, the joint probability distribution function of noisy data $y^N = [y(1), y(2), \dots, y(n)]^T$ can be found as follows:

$$f_Y(y^N, \bar{y}^N) = \frac{1}{(\sqrt{2\pi\sigma_n^2})^N} e^{-\frac{\|y^N - \bar{y}^N\|_2^2}{2\sigma_n^2}} \quad (2.30)$$

and data description length is defined as:

$$DL(y^N; \bar{y}^N) = -\frac{1}{N} \log_2(f_Y(y^N; \bar{y}^N)) = \log_2 \sqrt{2\pi\sigma_n^2} + \frac{\|y^N - \bar{y}^N\|_2^2}{2\sigma_n^2 N} \log_2 e \quad (2.31)$$

However, equation (2.31) is not available in practical situations as noiseless data vector \bar{y}^N is unavailable, and we can attempt only to estimate noiseless data description length for different data representations.

$$DL(\bar{y}^N; \hat{y}_{S_m}^N) = \log_2 \sqrt{2\pi\sigma_n^2} + \frac{\|\hat{y}^N - \bar{y}_{S_m}^N\|_2^2}{2\sigma_n^2 N} \log_2 e \quad (2.32)$$

Thus it is $z_{S_m} = \frac{1}{N} \|\bar{y}^N - \hat{y}_{S_m}^N\|_2^2$ that we need to estimate to find the minimum description length and the optimum subspace. Note that z_{S_m} is a sample of random variable Z_{S_m} .

It is clear that if we find boundaries for Z_{S_m} we can choose the subspace which minimizes the upper bound of data description length as the optimum subspace. It is shown that Z_{S_m} follows Chi-square distribution[2] and:

$$P(|Z_{S_m} - E(Z_{S_m})| \leq D_{S_m}) = p_1 \quad (2.33)$$

obviously, D_{S_m} is a function of p_1 , and we can find an upper and lower bound on Z_{S_m} assuming $E(Z_{S_m})$ or its bounds are available.

$$\underline{Z_{S_m}(p_1)} \leq Z_{S_m} \leq \overline{Z_{S_m}(p_1)} \quad (2.34)$$

However, it is shown that $E(Z_{S_m})$ is unavailable and is dependent on another variable estimated by $x_{S_m} = \frac{1}{N} \|y^N - \hat{y}_{S_m}^N\|$ where x_{S_m} is a sample of random variable X_{S_m} . Through validation, we can find upper and lower bounds of the this random variable:

$$P(|X_{S_m} - E(X_{S_m})| \leq J_{S_m}) = p_2 \quad (2.35)$$

and finally the bounds for Z_{S_m} are found with relationships that are dependent on p_1 and p_2 probabilities:

$$\underline{Z_{S_m}(p_1, y^N, p_2)} \leq Z_{S_m} \leq \overline{Z_{S_m}(p_1, y^N, p_2)} \quad (2.36)$$

and the optimum subspace is selected as:

$$S_m^* = \arg \min_{S_m} \overline{Z_{S_m}(p_1, y^N, p_2)}. \quad (2.37)$$

The optimum threshold proposed by MNDL is found by comparing the nested subspaces of different orders. The absolute value of noisy coefficients are sorted in decreasing order. This provides a nested set of subspaces, and the optimum subspace provides the threshold:

$$T^* = \min_i |\theta(i)|, i \in j | s_j \in S_m^*. \quad (2.38)$$

However, MNDL-SS assumptions for the additive noise are not working in thresholding because sorting the coefficients provides subspaces that are no longer independent and the noise vectors of the nested subspaces are correlated. MNDL-SS assumes random choice of subspaces which leads to independent additive noise vectors. It is later shown in MNDL thresholding [13] that by estimating the noisy part of Z_{S_m} through square adding sorted coefficients of a certain number of generated noise vectors in ascending order, we can provide a better estimate of $E(Z_{S_m})$ in thresholding applications.

2.4 Concluding Remarks

In this chapter, wavelet transform, wavelet shrinkage and popular wavelet shrinkage methods were reviewed. Wavelet shrinkage is a non linear, non parametric, approach to data denoising. VisuShrink and MNDL are usually used for one dimensional data but can be extended for two dimensional data as well. SureShrink and BayesShrink are commonly applied to images. BayesShrink outperforms SureShrink and other methods in image denoising and is a suitable solution for daily image denoising applications.

Chapter 3

Thresholding Method

3.1 Problem Statement

Noiseless data $\bar{y} = [\bar{y}_1, \bar{y}_2, \dots, \bar{y}_N]^T$ of length N has been corrupted by an additive noise:

$$y[n] = \bar{y}[n] + w[n] \quad (3.1)$$

where $w^N = [w[1], w[2], \dots, w[N]]^T$ is an independent and identically distributed (i.i.d) Gaussian random process with zero mean and variance σ^2 . In the denoising process, we first project the noisy data into an orthogonal basis. The considered orthogonal basis s_i s are orthonormal and we have:

$$\langle s_i, s_j \rangle = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \quad (3.2)$$

and $\langle s_i, s_j \rangle$ is the inner product of vectors, s_i and s_j . The i th coefficient of y^N is:

$$\theta[i] = \langle s_i, y^N \rangle, \quad 1 \leq i \leq N \quad (3.3)$$

and we have:

$$\theta[i] = \theta^*[i] + \theta_w[i] \quad (3.4)$$

where θ^* is the desired coefficient of the noiseless data:

$$\bar{y} = \sum_{i=1}^N \theta^*(i) s_i, \quad y = \sum_{i=1}^N \theta(i) s_i \quad (3.5)$$

and noise coefficients vector, $\theta_w^N = [\theta_w[1], \theta_w[2], \dots, \theta_w[N]]$, is a white Gaussian random process with the same mean and variance of w^N due to the orthonormality

of the bases :

$$E(\theta_w) = 0, \text{var}(\theta_w) = \sigma_w^2 \quad (3.6)$$

In a denoising approach, the available coefficients of noisy data are thresholded with the purpose of providing the best estimate of the noiseless data coefficients. As was mentioned in chapter two, there are two main types of thresholding, hard thresholding and soft thresholding. Although we briefly studied the hard thresholding approach, our focus in this thesis is on soft thresholding approaches. We denote the thresholded coefficients with soft threshold T_s , as $\hat{\theta}_\zeta$:

$$\hat{\theta}_\zeta(i) = \begin{cases} \text{sgn}(\theta(i))(|\theta(i)| - T_s) & \text{if } |\theta(i)| \geq T_s, \\ 0 & \text{otherwise} \end{cases} \quad (3.7)$$

where ζ is the number of non-zero coefficients resulted from thresholding with T_s :

$$\zeta = \#\{\theta[i], 1 \leq i \leq N : |\theta[i]| \geq T_s\} \quad (3.8)$$

The estimate of the noiseless data with this threshold is denoted by:

$$\hat{y}_\zeta = \sum_{i=1}^N \hat{\theta}_\zeta[i] s_i \quad (3.9)$$

Based on selected coefficients by thresholding function, equation (3.9) provides an estimation of the noiseless data.

3.1.1 Quality of the Thresholding Approaches

For comparison and evaluation of the thresholding approaches we consider the measure of quality as mean squared error (MSE):

$$E(\|\bar{y} - \hat{y}_\zeta\|_2^2) = E(\|\theta^* - \hat{\theta}_\zeta\|_2^2) \quad (3.10)$$

(the equality of errors in data and bases spaces is a direct result of Parseval's theorem.) An estimator with smaller MSE is considered as a better estimator.

3.2 Accumulative Noise Invalidating Soft Thresholding (ANIST) Method

After finding the coefficients for the observed data and sorting them in ascending order, we define a new variable, ψ_m such that it is the sum of squared version of the first "m" values:

$$\psi_m = \sum_{n=1}^m |\theta_{sort}[n]|^2 = \sum_{n=1}^m |\theta_{sort}^*[n] + v[n]|^2 \quad (3.11)$$

where θ_{sort} is the set of ascending sorted version of coefficients based on the absolute value of the coefficients, and θ^* and v are the corresponding noiseless and noise parts of θ_{sort} . Here, $v[n]$ is a sample of random variable $V[n]$ and ψ_m is a sample of random variable Ψ_m . In the following sections we are looking for the proper values of m for soft thresholding and present a new approach by noise invalidation. Note that since the coefficients are sorted in an ascending order, the value of m in our method provides the ζ in equation (3.8) with the following relation:

$$\zeta = N - m \quad (3.12)$$

3.2.1 Expected Value and Variance of Noise Effects in the Absence of Data

Here we study the properties of the additive noise in ψ_m as a function of m . If the discarded coefficients in the thresholding process with T_s are only noise, then we denote ψ_m in (3.11) with ψ_m^v . In this case, ψ_m^v is a sample of random variable Ψ_m^v with the following expected value and variance:

$$E(\Psi_m^v) = E\left(\sum_{n=1}^m V^2[n]\right) \quad (3.13)$$

$$\text{var}(\Psi_m^v) = E\left(\sum_{n=1}^m V^2[n] - E(\Psi_m^v)\right)^2 \quad (3.14)$$

The expected value of noisy part of ψ_m and its variance play key roles in our analysis, and exact calculation of these variables can be cumbersome. However, in this scenario with $\theta^*[n] = 0$; $1 \leq n \leq m$, the ascending coefficients of noisy data are exactly the ascending coefficients of the noise. This will allow us to estimate the expected value and variance by generating L samples of Gaussian random vectors of length m and averaging them, assuming $v_i^\circ[n]$ denotes the sorted version of the i th set of the Gaussian noise:

$$\widehat{E(\Psi_m^v)} \approx \frac{1}{L} \sum_{i=1}^L \sum_{n=1}^m v_i^{\circ 2}[n] \quad (3.15)$$

$$\widehat{\text{var}(\Psi_m^v)} \approx \frac{1}{L-1} \sum_{i=1}^L \left(\sum_{n=1}^m v_i^{\circ 2}[n] - \widehat{E(\Psi_m^v)} \right)^2 \quad (3.16)$$

3.2.2 Noise Statistics in ANIST

As we sort and add squared coefficients of generated noise samples (ψ_m^v), we observe that as m increases, the random variable begins to behave similar to Gaussian distribution and can be well approximated by that. This can be seen by generating and plotting the histogram for a large number of samples of this random variable. Figure 3.1 and Figure 3.2 show that the histogram starts with a shape that is similar to chi-square distribution. However, as m grows, the distributions are very similar to that of Gaussian. Thus, the distribution of the sum of sorted squared noise tends to normality as degree of freedom approaches infinity or m increases. This property can be used to probabilistically invalidate the signal from noise and find the proper threshold.

To elaborate this property more precisely, Figures 3.3 and 3.4 display normal probability plot of the data. Normal probability plot determines if the sample data

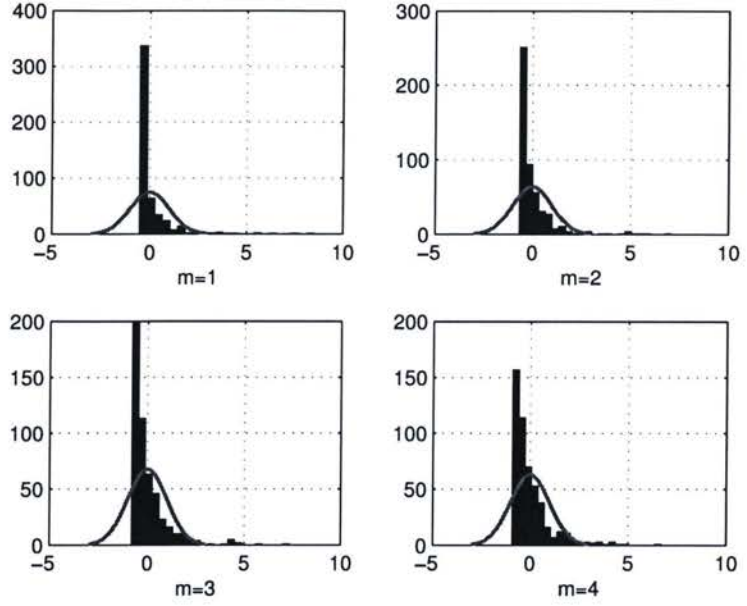


FIGURE 3.1. Histogram of noise for $\sigma^2 = 1$ at Top left: $m = 1$, Top right: $m = 2$, Bottom left $m = 3$, Bottom right $m = 4$.

follows Gaussian distribution. When the data is Gaussian, sample data displayed with (+) symbol follows the line that is connecting first and third quartiles of the data. Other distributions will cause curvature in the plot.

3.2.3 Noise Invalidation in ANIST

The observation of the behavior of ψ_m^v indicates that there is a weak dependence among the $v[n]$ s and therefore, for the summation of $\sum_{n=1}^m v^2[n]$ the use of the Central Limit Theorem (CLT) for large enough values of m is valid [14, 7]:

$$Pr\left\{\frac{|\psi_m^v - E(\Psi_m^v)|}{\lambda \sqrt{\text{var}(\Psi_m^v)}} \leq 1\right\} \approx \text{erf}\left(\frac{\lambda}{\sqrt{2}}\right) \quad (3.17)$$

where error function is:

$$\text{erf}(x) = \frac{1}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (3.18)$$

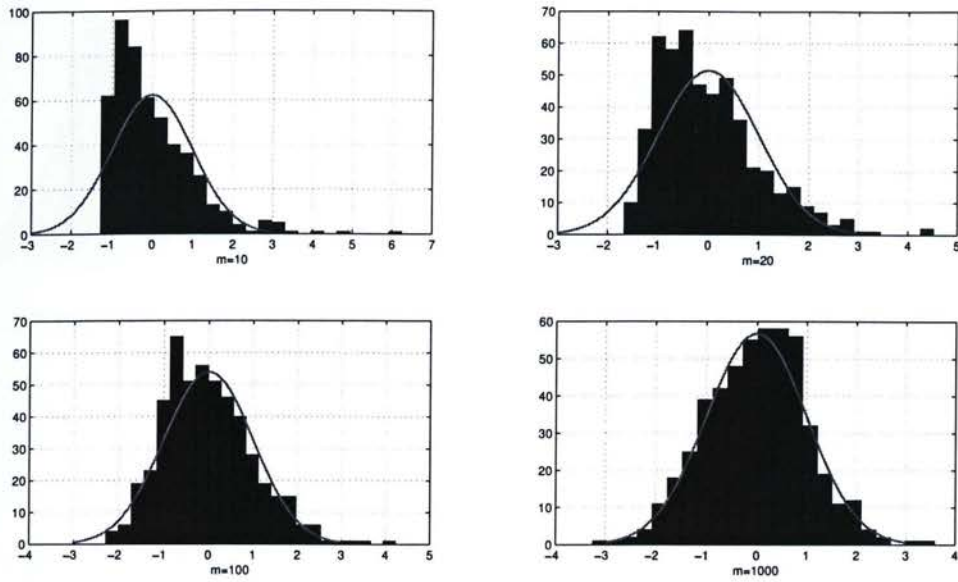


FIGURE 3.2. Histogram of noise for $\sigma^2 = 1$ at Top left: $m = 10$, Top right: $m = 20$, Bottom left: $m = 100$, Bottom right: $m = 1000$.

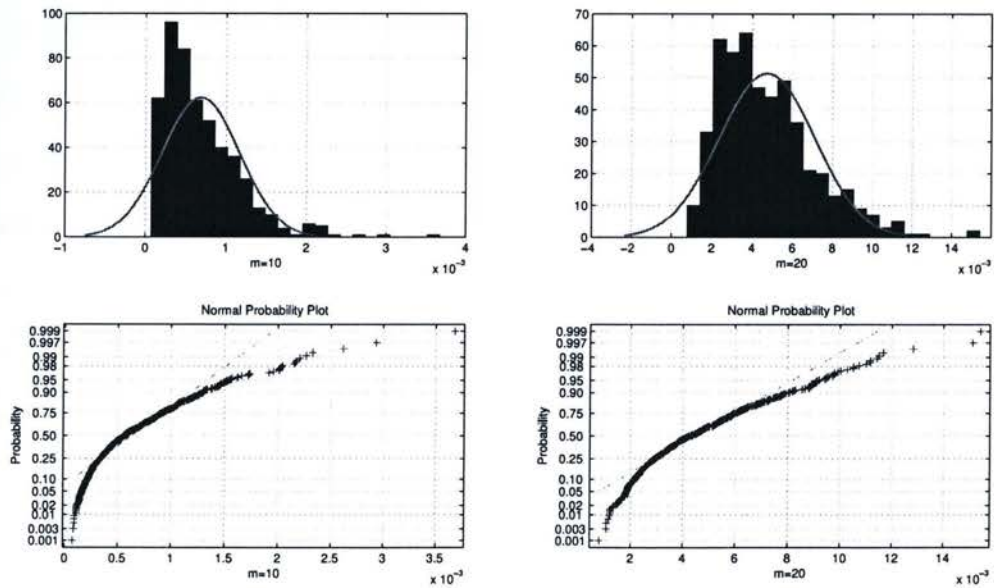


FIGURE 3.3. Noise behavior for $\sigma^2 = 1$ Top left: Histogram at $m = 10$, Bottom left: Norm plot at $m = 10$, Top right: Histogram at $m = 20$, Bottom right: Norm plot at $m = 20$.

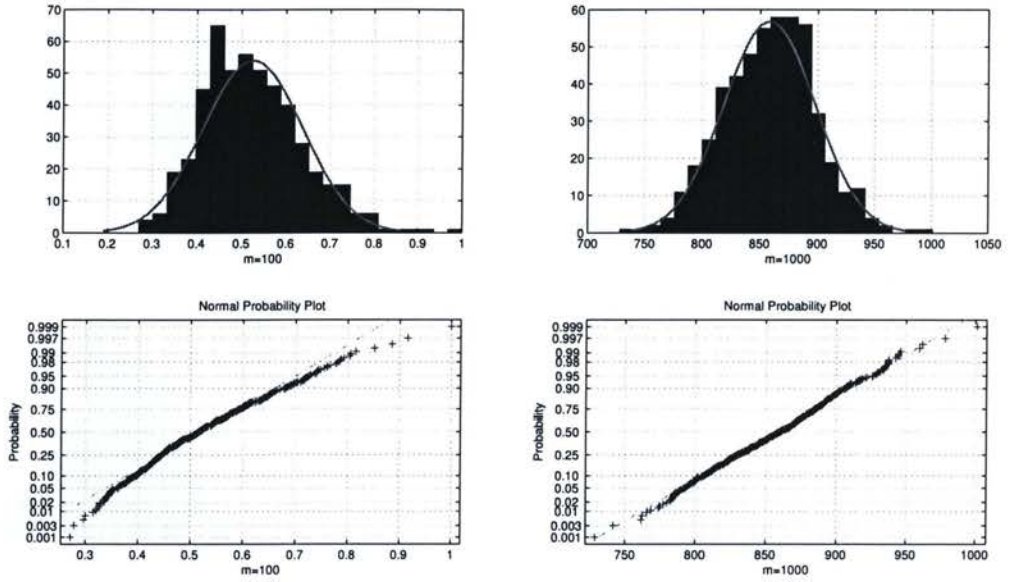


FIGURE 3.4. Noise behavior for $\sigma^2 = 1$ Top left: Histogram at $m = 100$, Bottom left: Norm plot at $m = 100$, Top right: Histogram at $m = 1000$, Bottom right: Norm plot at $m = 1000$.

Figure 3.5 Shows $E(\Psi_m^v)$ and its upper and lower bounds provided by standard deviation for Gaussian noise with unit variance. If λ is selected large enough, all noise cases fall between these two bounds with a close to one probability.

On the other hand, for a given data, the ψ_m in equation (3.11) is available. We can use the available property of ψ_m^v in Figure 3.5 to check for which m the available ψ_m is in the range of ψ_m^v with probability $\text{erf}(\lambda/\sqrt{2})$. As Figure 3.6 shows, the ψ_m will leave the boundaries of ψ_m^v once the presence of the noiseless coefficients is effective. Therefore, the invalidation criterion in this case is

$$\beta[m] = \frac{|\psi_m - \widehat{E(\Psi_m^v)}|}{\lambda \sqrt{\widehat{\text{var}(\Psi_m^v)}}} \quad (3.19)$$

and the desired m^* is found by checking the following condition:

$$m^* = \arg \min_m (\beta[m] \geq 1) \quad (3.20)$$

Thus m^* is the index where the ψ_m crosses the upper bound of ψ_m^v and the corresponding threshold with invalidation probability $\text{erf}(\lambda/\sqrt{2})$ is:

$$T_{ANIST} = |\theta_{\text{sort}}[m^*]| \quad (3.21)$$

which provides the number of coefficients $\zeta = N - m^*$ in equation (3.8).

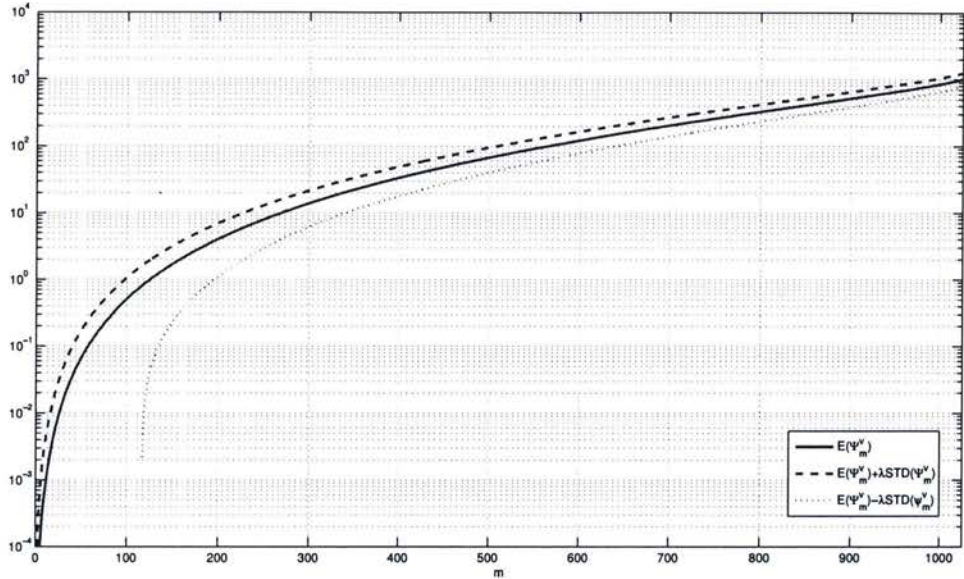


FIGURE 3.5. $E(\Psi_m^v)$ and upper and lower bounds for $\lambda = 5$, $\sigma^2 = 1$.

We denote this method as Accumulative Noise Invalidating Soft Thresholding and simply refer to it as ANIST [1].

3.2.4 Selection of a Proper λ

The value of λ is responsible for our measure of certainty in the invalidation process, and it should be selected carefully as it is a sensitivity factor in our method. From the normal curve table we know that usually $\lambda = 3.9$ is translated into probability 0.9. Thus, we expect that by choosing λ to be in a range of 3.5 to 7, we are ensuring that with a close to one probability, the condition of equation (3.20) is met out of

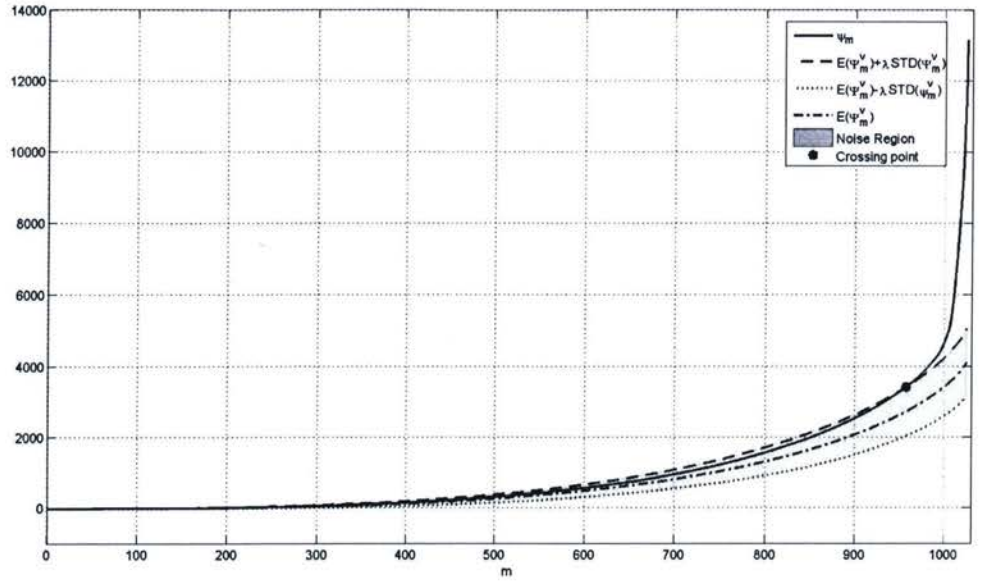


FIGURE 3.6. ψ_m of contaminated Blocks signal crossing upper bound of Invalidation region @ $m^* = 956$ in decimal plot. Shaded area is Gaussian noise region with $\sigma^2 = 4$.

noise invalidation process. For example, we have:

$$\text{erf}(3/\sqrt{2}) = 0.997300203937, \quad \text{erf}(5/\sqrt{2}) = 0.999999426697 \quad (3.22)$$

On the other hand, if the selected value for λ is too large, the invalidation region is larger and the sensitivity of the method is lost. To find the optimum value of λ , we compare the MSE results for a range of λ values in different noise variances.

For this study we consider a Blocks signal seen in Figure 3.7 with five level Haar wavelet decomposition.

λ	3.5	4	4.5	5	5.5	6
$\sigma^2 = 1$	0.49	0.37	0.26	0.20	0.26	0.32
$\sigma^2 = 4$	1.28	0.92	0.81	0.73	0.89	1.05
$\sigma^2 = 9$	1.78	1.29	1.12	0.91	0.97	1.31

TABLE 3.1. MSE by different λ values.

Table 3.1 shows that in most cases $\lambda = 5$ provides optimum results¹. A similar procedure can be taken to show that $\lambda = 5$ is proper for Mishmash signals.

3.2.5 Simulation Results for ANIST

We continue with the simulation for standard Blocks signal and mishmash signal, which are provided in MATLAB R2007b and were used as test signals in [10]. Wavelet transform is selected as the orthogonal basis for analysis and Blocks and Mishmash are selected as test signals. The two signals can represent two extreme cases. The Blocks signal has very small non-zero coefficients and Mishmash has no zero coefficients, thus they can present a very good test package between two

an Noise (AWGN) was added to original signals with data variances of $\sigma^2 = 1, 4, 9$ were tested. Figure 3.7 shows in noisy versions. Five level decomposition with "Haar" this experiment. Figure 3.8 shows test signals with their

result of one set of observed data. For this observation, the (3.20) is met in $m^* = 12$ and the desired subspace will be

the first 1012 ($\zeta = N - m^*$) thresholded coefficients of noisy signal in descending order.

Figure 3.10 shows noisy Blocks signal crossing the upper bound in $m^* = 986$. Proper threshold is selected at $m^* = 986$ and the value of this threshold is:

$$T_{ANIST} = |\theta_{sort}(986)|$$

¹The results are based on statistical average for $L = 1000$ generations.

#300 03-04-2010 12:50PM
Item(s) checked out to Abdullah, Alaa.
TITLE: Data denoising by noise invalidat
CALL #: TK7867.5 .N55 2008
BARCODE: 0010149751934
DUE DATE: 06-12-10

Check Your File:
<http://www.library.ryerson.ca/>

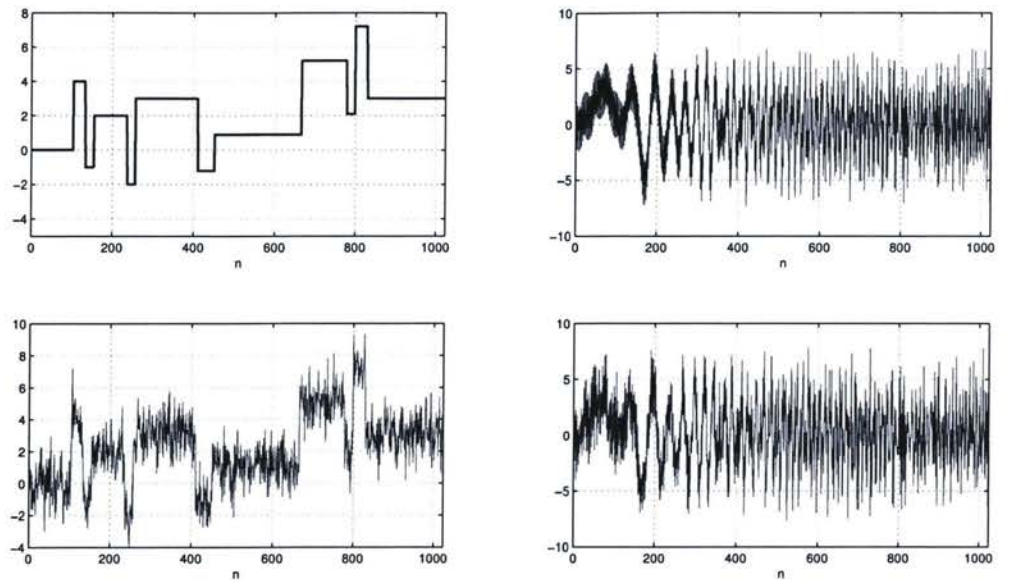


FIGURE 3.7. Test signals and their noisy forms. Top left: Blocks signal, Top right: Mishmash signal, Bottom left: Noisy Blocks with $\sigma^2 = 1$, Bottom right: Noisy Mishmash with $\sigma^2 = 1$.

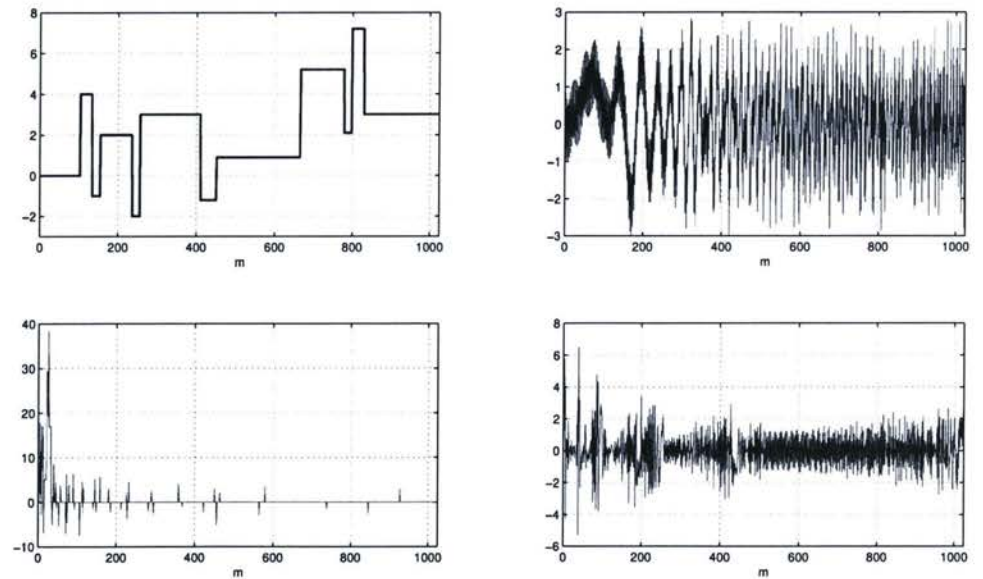


FIGURE 3.8. Test signals and their wavelet coefficients. Top left: Blocks signal, Top right: Mishmash signal, Bottom right: Blocks coefficients, Bottom left: Mishmash coefficients.

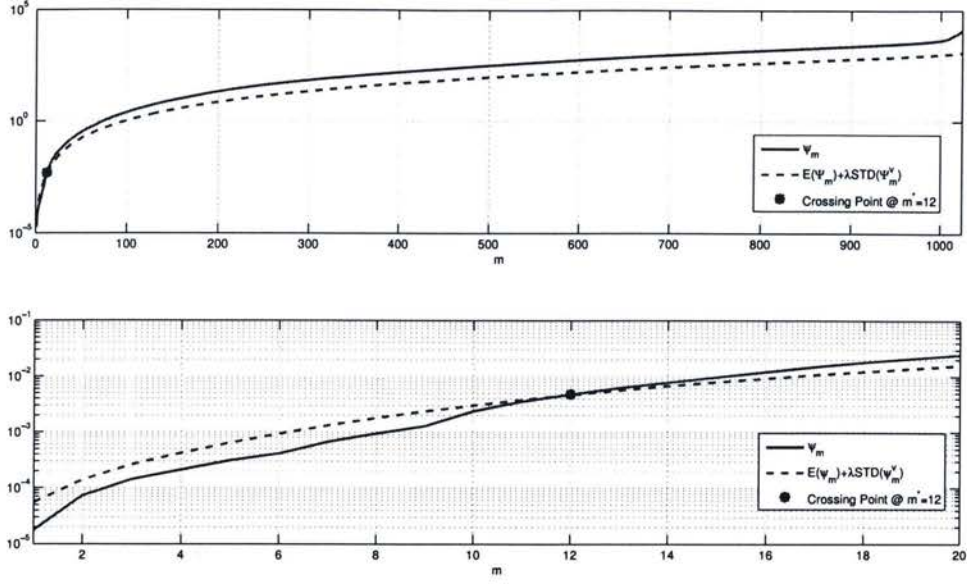


FIGURE 3.9. ψ_m and $E(\psi_m^v)$ for Blocks, $\sigma^2 = 1$. The cross point is at $m^* = 12$.

thus the desired subspace will be the first 38 thresholded coefficients in descending order.

Figure 3.11 shows noisy Mishmash signal crossing the upper bound in (a) and a zoom in of crossing point in (b).

We compare the performance of the proposed method with some of the existing methods, SureShrink[8], MNDL thresholding[2], BayesShrink[3] and VisuShrink[10]. BayesShrink, originally proposed as a soft thresholding Image denoising approach, is applied to one dimensional data in this simulation.

	OPTIMUM	SURE	MNDL	BAYES	VISU	ANIST
$\sigma^2 = 1$	2.69	2.91	2.7112	0.58	3.10	2.13
$\sigma^2 = 4$	6.57	7.08	6.02	2.27	6.78	2.31
$\sigma^2 = 9$	9.59	10.12	10.04	5.12	10.21	3.74

TABLE 3.2. Proposed Thresholds by each method for Blocks test signal. Optimum threshold is the best threshold possible.

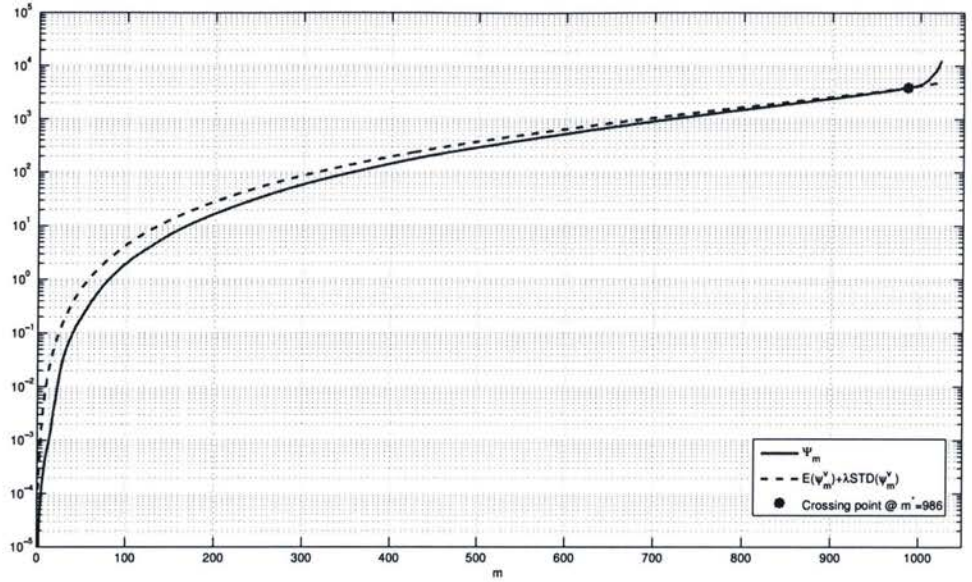


FIGURE 3.10. ψ_m and $E(\psi_m^v)$ for Blocks, $\sigma^2 = 4$. The cross point is at $m^* = 986$.

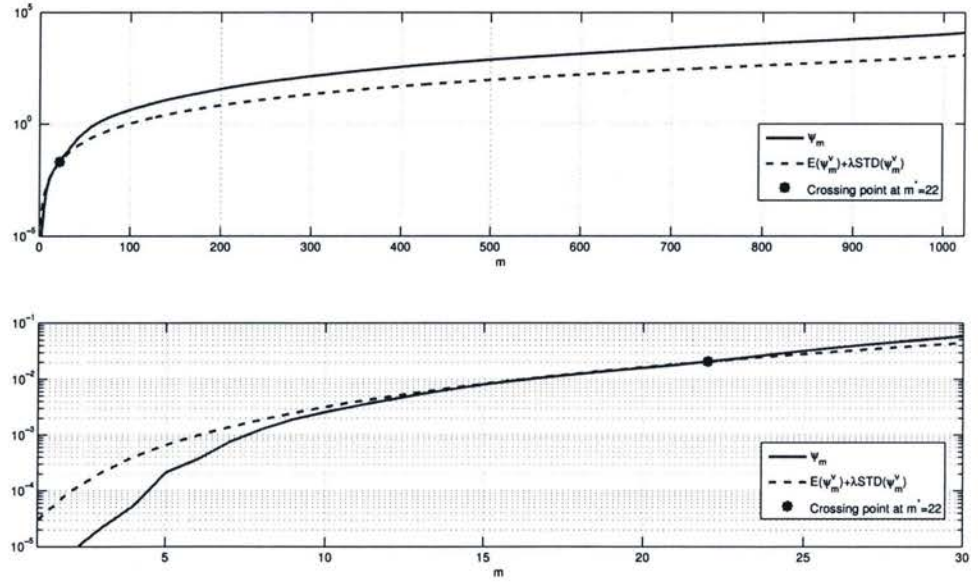


FIGURE 3.11. Top: $E(\psi_m^v)$ and $\psi_{s,m}$ for Mishmash, $\sigma^2 = 1$, Bottom: zoomed version of top figure. The cross point is at $m^* = 22$.

Table 3.3 provides the results with soft thresholding and compares ANIST with existing methods. As the table shows, ANIST outperforms other shrinkage methods and provides acceptable results in comparison with MNDL and BayesShrink.

Blocks	SURE	MNDL	BAYES	VISU	ANIST
$\sigma^2 = 1$	0.48	0.45	0.42	0.35	0.20
$\sigma^2 = 4$	0.91	0.71	0.86	0.96	0.73
$\sigma^2 = 9$	1.19	1.02	1.26	1.32	0.94
Mishmash					
$\sigma^2 = 1$	1.48	1.28	1.12	1.59	1.32
$\sigma^2 = 4$	4.02	3.99	3.54	4.31	3.34
$\sigma^2 = 9$	6.11	5.63	5.32	6.78	5.21

TABLE 3.3. MSE for different Methods compared.

3.3 Point-wise Noise Invalidating Soft Thresholding (PNIST) Method

A similar approach can be used to invalidate sorted coefficients directly instead of working with a summation of squared, sorted coefficients. To work with sorted coefficients directly, we define a new random variable, γ_m such that:

$$\gamma_m = |\theta_{\text{sort}}[m]| = |\theta_{\text{sort}}^*[m] + z[m]| \quad (3.23)$$

Where θ_{sort} is the set of ascending sorted version of coefficients and θ^* and z are noiseless and noise parts of θ_{sort} .

3.3.1 Expected Value and Variance of Noise Effects in the Absence of Data

If discarded coefficients in the thresholding process with T_s are only noise, we denote γ_m in equation (3.23) with γ_m^z . In this case γ_m^z is a sample of the random variable Γ_m^z with the following expected value:

$$E(\Gamma_m^z) = E(Z) \quad (3.24)$$

and its variance will be:

$$\text{var}(\Gamma_m^z) = E(\Gamma_m^z - E(\Gamma_m^z))^2 = E(Z - E(\Gamma_m^z))^2 \quad (3.25)$$

Similar to the situation in ANIST, the expected value of this noise, and its variance are crucial to our method and an exact calculation of these values can be quite cumbersome. However, the fact that ascending coefficients of noisy data are exactly the ascending coefficients of a Gaussian noise in the absence of noiseless data coefficients can be used to estimate the expected value and variance by generating J samples of sorted Gaussian random variable and averaging them. Consider z_i° , the i th sample of a sorted Gaussian noise, then we have:

$$\widehat{E(\Gamma_m^z)} \approx \frac{1}{J} \sum_{i=1}^J z_i^\circ[m] \quad (3.26)$$

$$\widehat{var(\Gamma_m^z)} \approx \frac{1}{J-1} \sum_{i=1}^J (z_i^\circ[m] - \widehat{E(\Gamma_m^z)})^2 \quad (3.27)$$

where Z represents the random variable with sample $z[n]$.

3.3.2 Noise Statistics in PNIST

To find the distribution of sorted noise coefficients (Γ_m^z) at each m we simulate J samples of Gaussian random vectors and look at the histograms at each m .

As it can be seen in Figures 3.12, 3.13, 3.14 and 3.15, the distribution behaves like Gaussian distribution as m increases.

In this case, order statistics can be used to analyze the probability distribution function of the Γ_m^z . Since the additive noise of coefficients (θ_w s in equation (3.4)) are independent, identically distributed (i.i.d) random variables which follow Gaussian probability distribution $\mu(0, \sigma^2)$ and are sorted in ascending order, thus the probability distribution function for Γ_m^z can be found as follows (Proof in Appendix A):

$$f_{z_m}(x) = n f(x) \binom{n-1}{m-1} F(x)^{m-1} (1 - F(x))^{n-m} \quad (3.28)$$

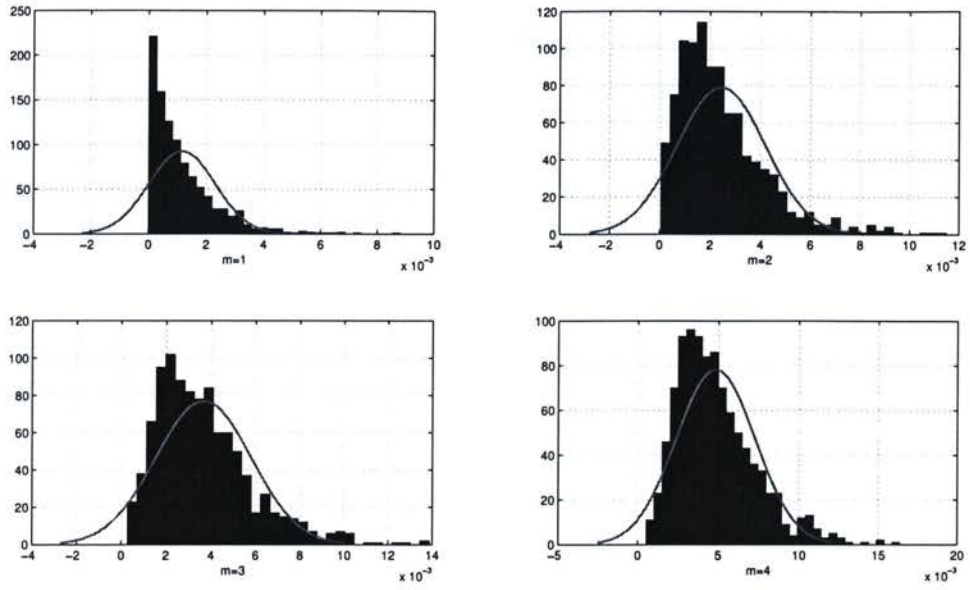


FIGURE 3.12. Histogram of noise for $\sigma^2 = 1$ at Top left: $m = 1$, Top right: $m = 2$, Bottom left: $m = 3$, Bottom right: $m = 4$ for PNIST.

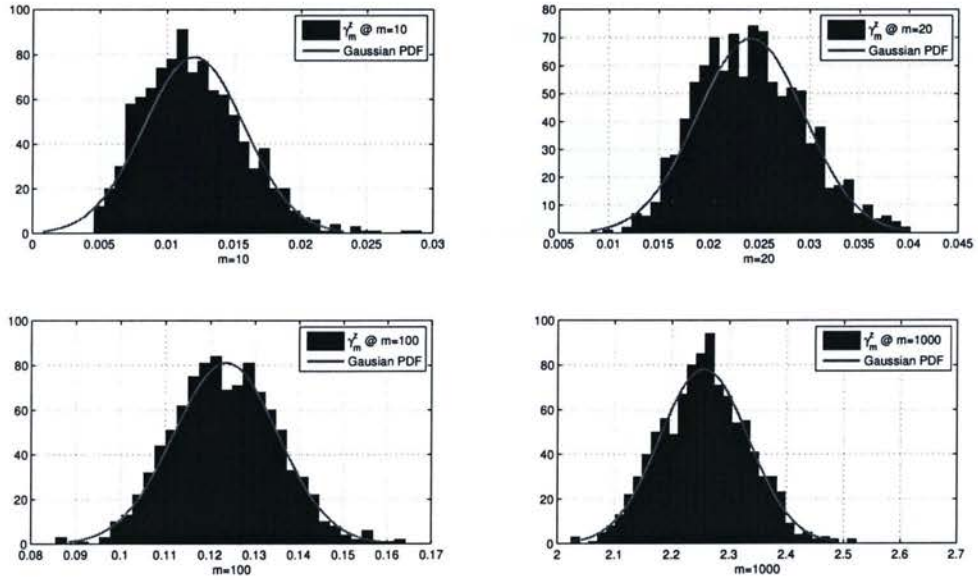


FIGURE 3.13. Histogram of noise for $\sigma^2 = 1$ at Top left: $m = 10$, Top right: $m = 20$, Bottom left: $m = 100$, Bottom right: $m = 1000$, for PNIST.

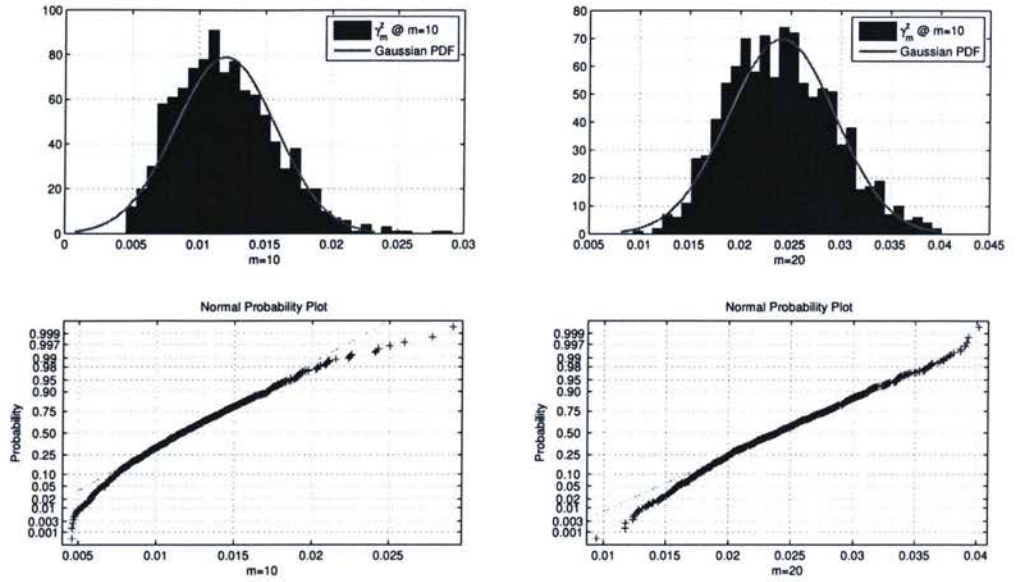


FIGURE 3.14. Histogram of noise for $\sigma^2 = 1$ at Top left: $m = 10$, Bottom left: Norm plot at $m = 10$, Top right: $m = 20$, Bottom left: Norm plot $m = 20$ for PNIST.

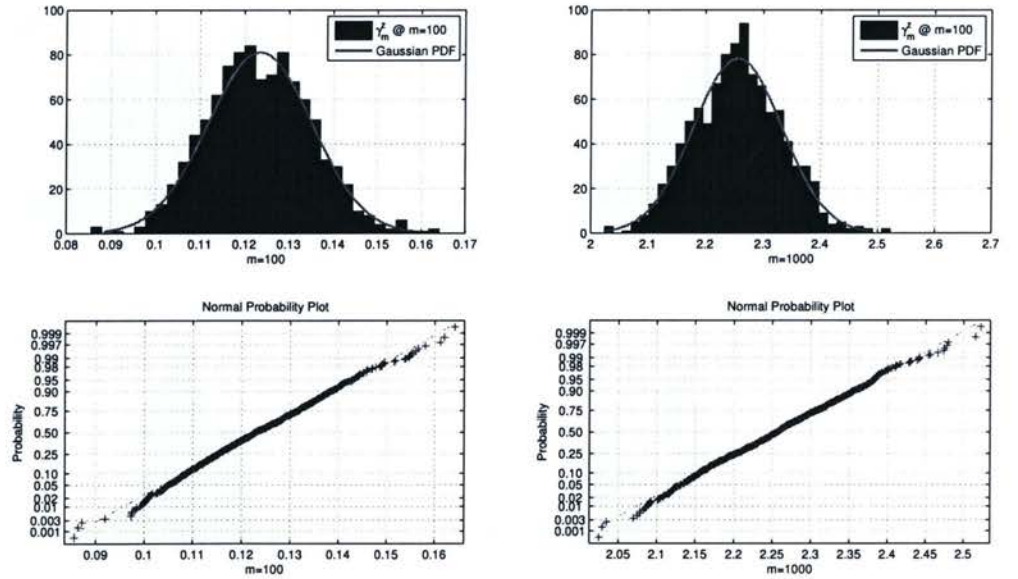


FIGURE 3.15. Histogram of noise for $\sigma^2 = 1$ at Top left: $m = 100$, Bottom left: Norm plot at $m = 100$, Top right: $m = 1000$, Bottom right: Norm plot $m = 1000$ for PNIST.

Where $f(x)$ is probability distribution function (pdf) of Gaussian distribution $\mu(0, \sigma^2)$ and $F(x)$ is the cumulative distribution function (CDF).

Using equation (3.28), we can compare probability distribution function of Γ_m^z with Gaussian probability distribution function at different m 's and ensure that the observed results in histograms are accurate.

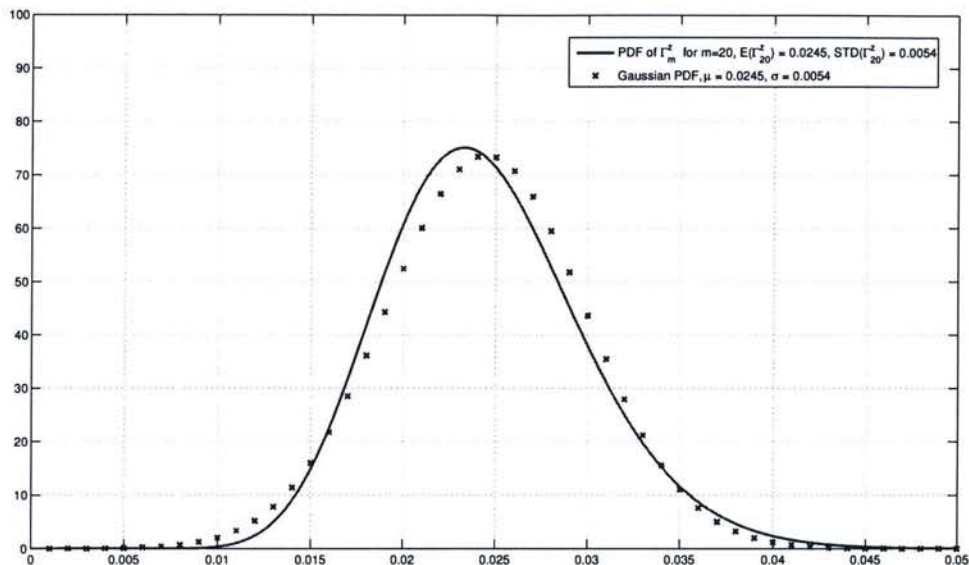


FIGURE 3.16. Probability distribution function of Γ_m^z at $m = 20$ compared with Gaussian probability distribution function of the same mean and variance.

Figure 3.16 shows probability distribution function of Γ_m^z at $m = 20$ for $\sigma^2 = 1$ compared to a Gaussian distribution function of the same mean and variance. The histogram of this distribution can be seen in Figure 3.12.

The small deviation from the Gaussian distribution for $m = 20$ is evident in the histogram and pdf shown in Figure 3.14. However, Figure 3.17 shows the probability distribution function acts closely like Gaussian as m increases to 100. Thus, as expected and seen in the histograms and probability distribution function, by increasing m , the distribution of Γ_m^z tends to behave similar to a Gaussian distri-

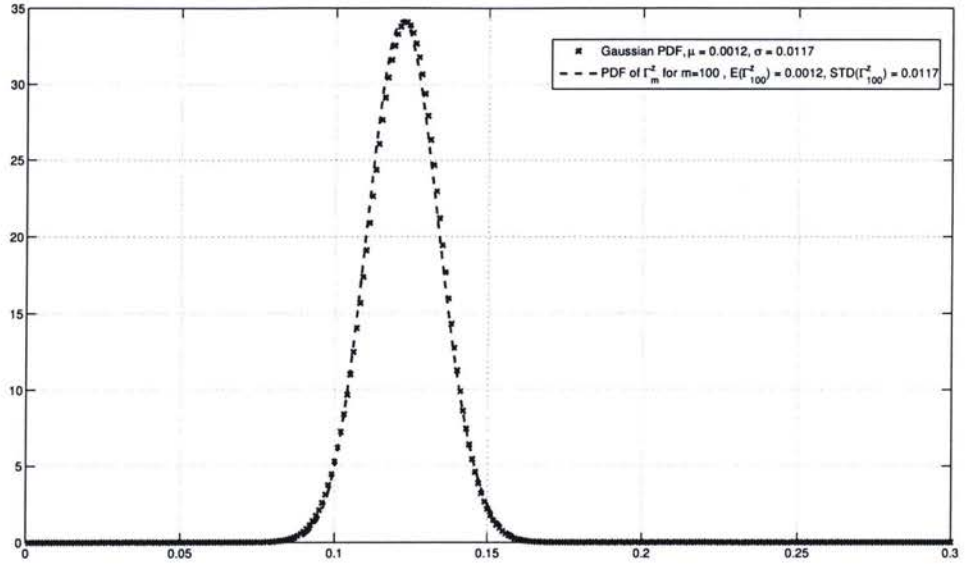


FIGURE 3.17. Probability distribution function of Γ_m^z at $m = 100$ compared with Gaussian probability distribution function of the same mean and variance.

bution. This behavior can be used to probabilistically invalidate signal coefficients out of noise region and find the proper threshold similar to the case in ANIST.

3.3.3 Noise Invalidation in PNIST

Figure 3.18 shows $E(\Gamma_m^z)$ and its upper and lower bounds provided by standard deviation for Gaussian noise with unit variance.

Since Γ_m^z approximately follows Gaussian distribution, the probability of a γ_m^z being bounded in a certain confidence interval such as Figure 3.18 can be approximated using Gaussian error function:

$$Pr\left\{\frac{|\gamma_m^z - E(\Gamma_m^z)|}{\lambda\sqrt{var(\Gamma_m^z)}} \leq 1\right\} \approx erf\left(\frac{\lambda}{\sqrt{2}}\right) \quad (3.29)$$

where the error function is:

$$erf(x) = \frac{1}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (3.30)$$

If λ is selected large enough, with close to one probability all noise cases fall between the two bounds. On the other hand, for a given data, the γ_m in the equation (3.23) is available. We can use the available property of γ_m^z shown in Figure 3.18 to check for which m the available γ_m is in the range of γ_m^z with probability $\text{erf}(\lambda/\sqrt{2})$. As Figure 3.19 shows, the γ_m will leave the boundaries of γ_m^z once the presence of the noiseless coefficients is effective. Therefore, the invalidation criterion in this case is:

$$\alpha[m] = \frac{|\gamma_m - \widehat{E(\Gamma_m^z)}|}{\lambda \sqrt{\widehat{\text{var}(\Gamma_m^z)}}} \quad (3.31)$$

To find the proper threshold using noise invalidating the proportion proposed in 3.31 is calculated at each m . Similar to ANIST, the desired m^* is found by checking the following condition:

$$m^* = \arg \min_m (\alpha[m] \geq 1) \quad (3.32)$$

Thus m^* is the index where γ_m crosses the upper bound of the noise mean for Gaussian invalidation, and the corresponding sorted coefficient at this point is the proper threshold.

$$T_{PNIST} = |\theta_{\text{sort}}[m^*]| \quad (3.33)$$

A table similar to table 3.1 can be produced to find the proper λ in this case. Simulation results show that $\lambda = 5$ provides acceptable results for this approach as well. Thus, we keep our selection of $\lambda = 5$.

We denote this approach as Point-wise Noise Invalidating Soft Thresholding (PNIST).

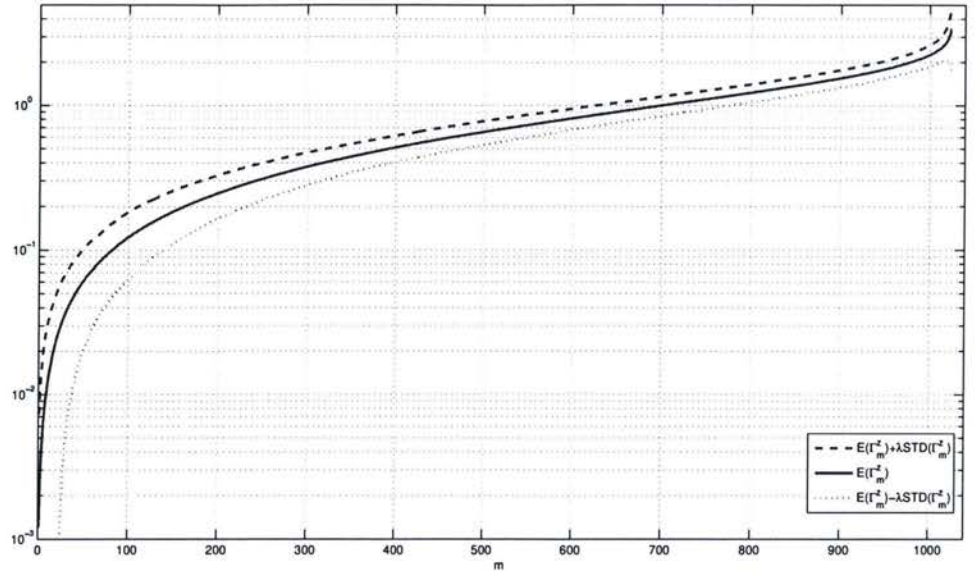


FIGURE 3.18. $E(\Gamma_m^z)$ and its upper and lower bounds for $\lambda = 5$.

3.3.4 Simulation Results for PNIST

We use the same test signals and add AWGN to original signals with a data length of 1024 and $\lambda = 5$ similar to the simulation for ANIST.

Figure 3.20 shows γ_m crossing the upper bound of noise average for $\sigma^2 = 1$ at $m^* = 862$. Figure 3.21 shows γ_m crossing the upper bound for $\sigma^2 = 1$.

Blocks	SURE	MNDL	BAYES	VISU	(1)	(2)
$\sigma^2 = 1$	0.48	0.45	0.42	0.35	0.20	0.18
$\sigma^2 = 4$	0.91	0.71	0.86	0.96	0.73	0.76
$\sigma^2 = 9$	1.19	1.02	1.26	1.32	0.94	0.90
Mishmash						
$\sigma^2 = 1$	1.48	1.28	1.12	1.59	1.32	1.24
$\sigma^2 = 4$	4.02	3.99	3.54	4.31	3.34	3.29
$\sigma^2 = 9$	6.11	5.63	5.32	6.78	5.21	5.27

TABLE 3.4. MSE for different Methods compared.(1)ANIST (2) PNIST.

Table 3.4 provides the results with soft thresholding and compares PNIST with ANIST and other existing methods. As the table shows, PNIST outperforms

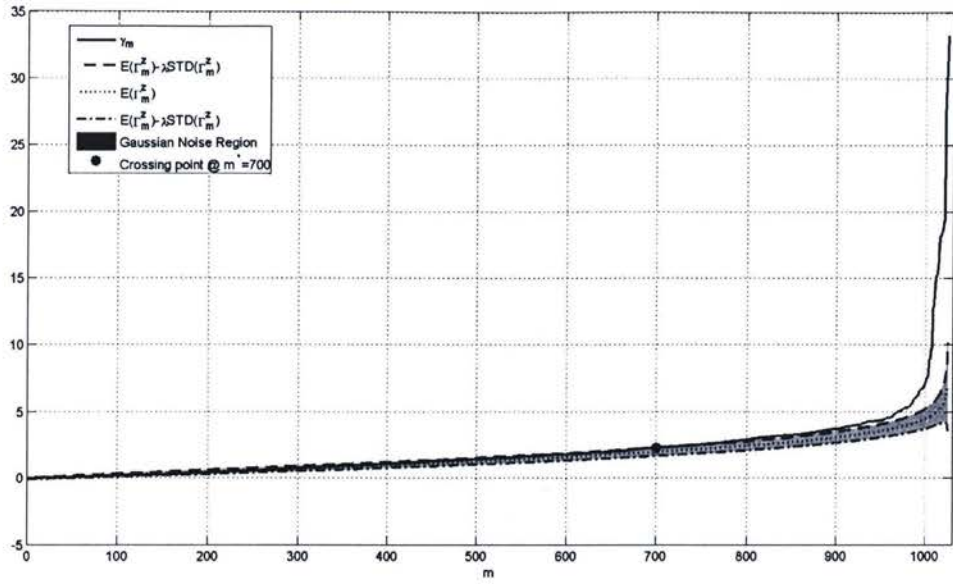


FIGURE 3.19. γ_m of contaminated Blocks signal crossing upper bound of Invalidation region @ $m^* = 700$ in decimal plot. Shaded area is Gaussian noise region with $\sigma^2 = 4$.

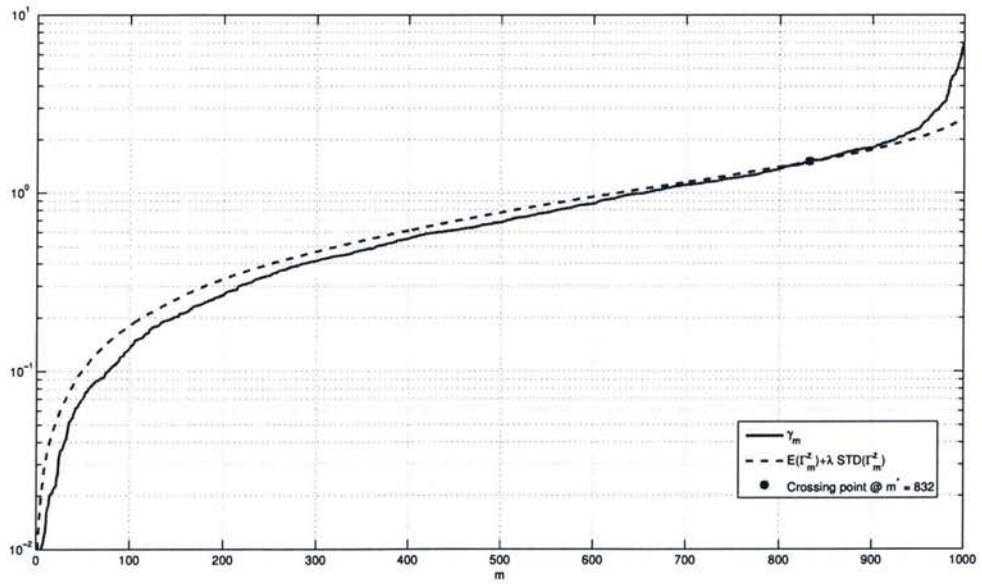


FIGURE 3.20. PNIST: γ_m and upper bound for noise average in case of Blocks, $m^* = 862$, $\sigma^2 = 1$.

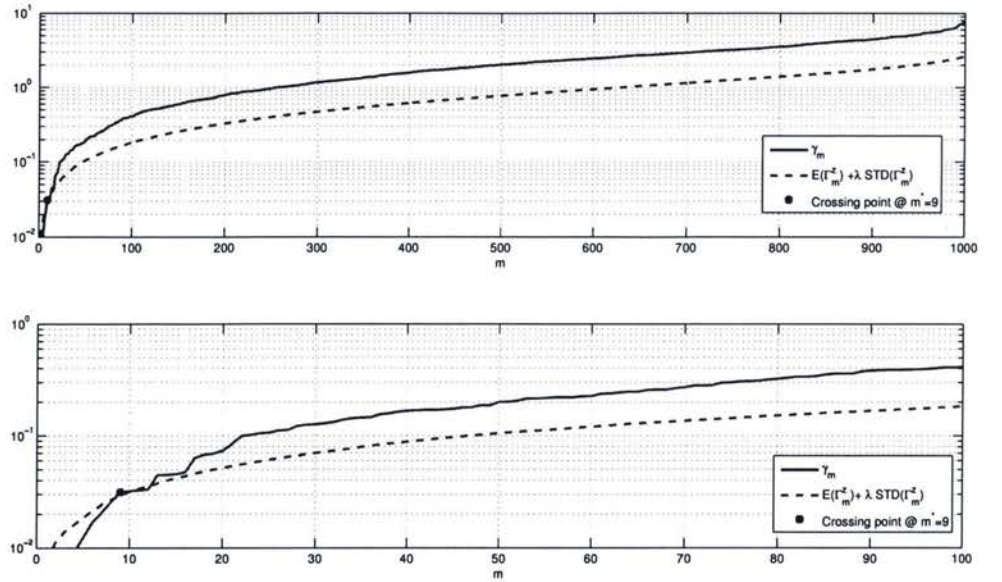


FIGURE 3.21. PNIST: γ_m and upper bound for noise average in case of Mishmash, $m^* = 8$.

ANIST and provides acceptable results in comparison with other shrinkage methods.

3.4 Concluding Remarks

As mentioned before, ANIST uses a function of square added sorted coefficients and its noise statistics to invalidate signal out of noise region. As a result of using accumulation in noise estimation, Central Limit Theorem (CLT) allows us to implement ANIST for additive noises other than Gaussian.

Our simulations also demonstrate this important property. Figure 3.22 shows zero mean uniform noise with $\sigma^2 = 4$. Figure 3.23 shows ψ_m^v for this noise with $M = 10000$ generations. It is evident that due to CLT, distribution of Ψ_m^v can be approximated by Gaussian distribution.

In another example with additive Laplacian noise, we compared the resulting MSE of BayesShrink and MNDL with that of ANIST for a Blocks signal con-

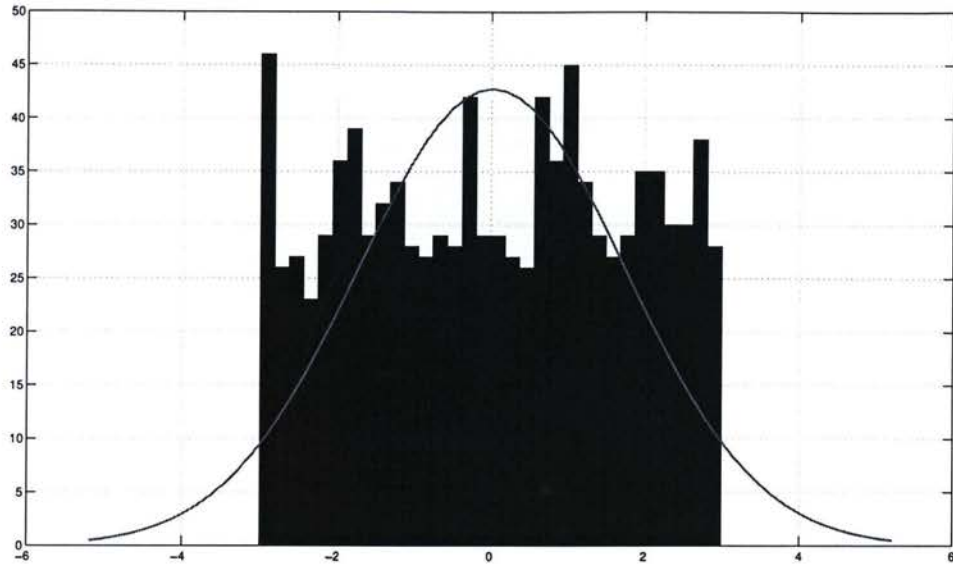


FIGURE 3.22. Uniform noise with $\sigma^2 = 4$.

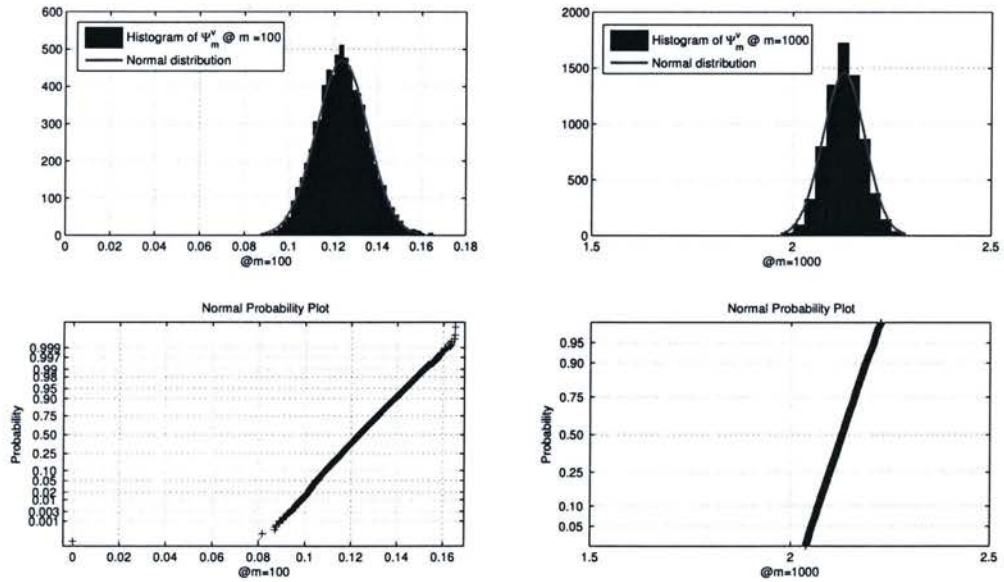


FIGURE 3.23. ψ_m^v histogram and normplot for $\sigma^2 = 4$ at $m = 100$ and $m = 1000$. Note that sample data falls exactly on the straight line.

taminated with Laplacian noise with variance of $\sigma^2 = 4$. The ANIST proposed threshold results in $MSE = 0.27$ while the MSE is 0.65 in BAYES and 0.39 in

MNDL. This shows the method's robustness towards various noise behavior. Table 3.5 shows MSE results for Blocks signal contaminated by uniform noise.

Uniform	MNDL	BAYES	VISU	ANIST	PNIST
$\sigma^2 = 1$	1.19	2.17	3.21	0.26	0.48
$\sigma^2 = 4$	2.89	3.87	5.12	0.89	1.29
$\sigma^2 = 9$	3.90	3.98	5.98	0.97	1.72

TABLE 3.5. MSE by Method for Blocks signal contaminated by uniform noise.

While this property does not hold true for PNIST method, PNIST provides better results in most cases when the additive noise is Gaussian, and the rate of convergence of noise distributions to Gaussian is faster in PNIST compared to ANIST. This can be seen graphically by comparing Figure 3.14 to Figure 3.3 where γ_m^z performs better in normal plots compared to ψ_m^v at $m = 10$ and $m = 100$. It can also be seen in Table 3.4 that Mean Square Error(MSE) results of PNIST outperform those of ANIST in most cases, when additive noise follows Gaussian distribution.

Chapter 4

Additive Noise in Denoising

The main goal of denoising is to recover the original noiseless data from the observed data. Many algorithms have been proposed for this purpose, and all of them require the value of noise variance. Thus, estimation of the noise variance is crucial to the denoising methods. Here, we propose a new noise variance estimation algorithm based on properties of the observed data. The literature on noise variance estimation is limited, and the existing algorithms use many assumptions on the data which makes their implementation limited to particular applications. For example, one of the first methods proposed to estimate variance of noise in speech processing assumes the uncontaminated signal to be Auto Regressive (AR) [33]. The method uses AR parameters to estimate noise variance. The drawback here is that not all the signals can be considered as AR, and thus the method is limited only to special applications. In another speech enhancement method, the white noise variance is estimated from preceding silent portions of speech [25]. However, this method has two major drawbacks. First, detection of silent portions is very difficult in low SNR situations. Second, the estimated variance can be used in other segments only if the noise is of stationary nature.

Another method that can be applied to both 1D and 2D data is based on estimation of the scatter of normally distributed data with high level of outliers [26]. The method is based on the shortest half sample method. In this method, the length of the shortest half sample is used as an estimation of standard deviation of the uncontaminated data. One disadvantage of this method is that the method

is applicable only to data with the majority of the data points having no signal present.

Another widely used method suggests estimating the variance using the median of absolute value of normalized, fine-scale wavelet coefficients [9, 8, 3]. Maximum Likelihood (ML) estimators provide another class of methods [23]. In other variance estimation methods, a primary threshold or filter is first chosen to pick the fine-scale wavelet coefficients and suppress the effect of unknown noiseless data [35, 32]. The noise variance is then estimated using the remaining coefficients. The sensitivity of these approaches to the choice of the first threshold or filter, which estimates the noise variance, is not known. These methods are successful only if enough of the coefficients can be assumed to be entirely noise dependent.

4.1 Noise Estimation in MNDL Approach

Minimum Noiseless Description Length (MNDL) proposes another variance estimation method that estimates the noise variance during the denoising process [2]. In MNDL, the noiseless data "Description Length" (DL) is a function of noise variance. To estimate noise variance, MNDL finds the optimum subspace as a function of the noise variances for a considered range of variances. For any noise variance, the bounds on the reconstruction error provide bounds on the description length. To estimate the noise variance, MNDL suggests choosing the noise variance that minimizes the noiseless data DL. From equation (2.34) we know that bounds for reconstruction error provide bounds for data description length:

$$\underline{DL_{\sigma_{\omega}^2}(\bar{y}^N; \hat{y}_{S_m}^N)} \leq DL_{\sigma_{\omega}^2}(\bar{y}^N; \hat{y}_{S_m}^N) \leq \overline{DL_{\sigma_{\omega}^2}(\bar{y}^N; \hat{y}_{S_m}^N)} \quad (4.1)$$

where the DL is defined in equation (2.32). Thus we have:

$$MNDL(y^N, \sigma_{\omega}^2) = \min_{S_m} \overline{DL_{\sigma_{\omega}^2}(\bar{y}^N; \hat{y}_{S_m}^N)} \quad (4.2)$$

and finally noise variance is estimated as the variance which minimizes MNDL in the optimum subspace:

$$\hat{\sigma}_{\omega}^2 = \arg \min_{\sigma_{\omega}^2} MNDL(y^N; \sigma_{\omega}^2) \quad (4.3)$$

However, it can be shown that in many cases, equation (4.3) fails to provide a global minimum. In most situations, it leads to a graph of several local minimums that do not provide us with the necessary information to estimate the variance. For example, Blocks signal contaminated with Gaussian noise of $\sigma^2 = 2$ was tested on MNDL noise estimation. As Figure 4.1 shows, minimum description length happens at $\sigma^2 = 5.5$, thus; the estimate of MNDL is 5.5. However, the true noise is much smaller, and the description length graph fails to give a global minimum.

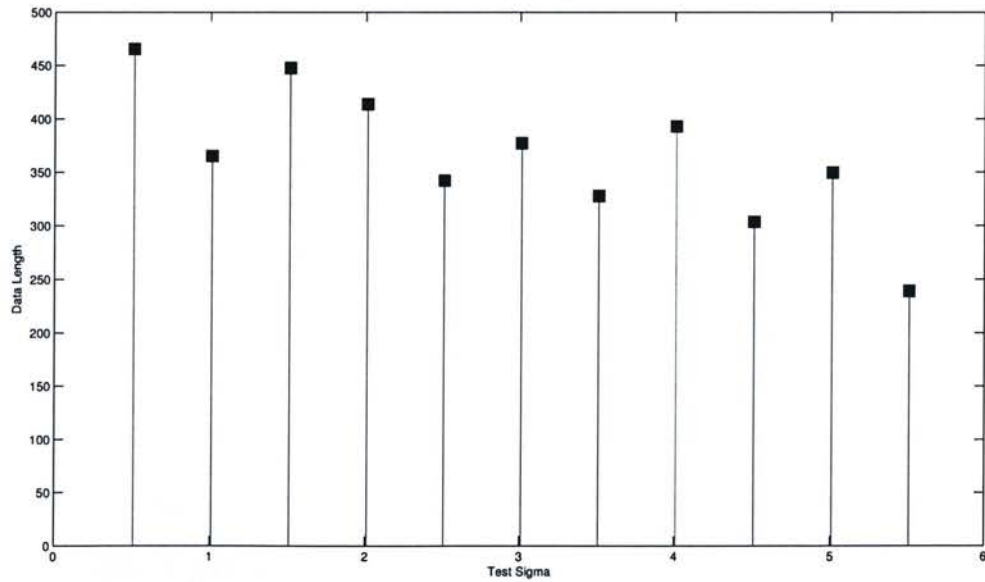


FIGURE 4.1. Data description length versus $\hat{\sigma}$ in MNDL method when the true variance is $\sigma^2 = 2$.

4.2 Proposed Noise Variance Estimation Method

To estimate the variance of noise, we propose a new approach that is based on the fact that autocovariance of a zero mean Gaussian noise vector is zero at any point other than the origin, whereas the autocovariance of the noiseless signal is impulse at the origin and non-zero at other points. This property is one of the main differences between the noiseless signal and noise part that has been considered in all denoising approaches. Thus, we expect the autocovariance of noise vector to have a spike at the origin and to be negligible at other points.

Autocovariance of a zero mean Wide Sense Stationary (W.S.S) random variable X is defined as

$$c(m) = E(X(n+m)X(n)) \quad (4.4)$$

This autocovariance can be estimated by a sample data of the random variable, x with length of N , as follows:

$$C(m) = \begin{cases} \frac{1}{N-m} \sum_{n=0}^{N-|m|-1} X(n+m)X(n) & m \geq 0 \\ C^*(-m) & m < 0 \end{cases} \quad (4.5)$$

Figure 4.2 shows the autocovariance of a Gaussian noise vector. As signal is added to the noise, we expect the random behavior of the autocovariance graph to change and the area under its curve to increase. Autocovariance of a noisy Blocks signal can be seen in Figure 4.3.

Let us demonstrate the new approach with an example. Consider the case that the true variance is $\sigma^2 = 9$. We want to estimate the variance using the mentioned autocovariance property. To do this, we start by selecting a range of values for the variance, hoping the true variance falls within this range. In this case we select our range to be $\sigma_{test} = [1 : 0.5 : 16]$, which means that starting from 1 we increase our test variance by a step of 0.5 to 16.

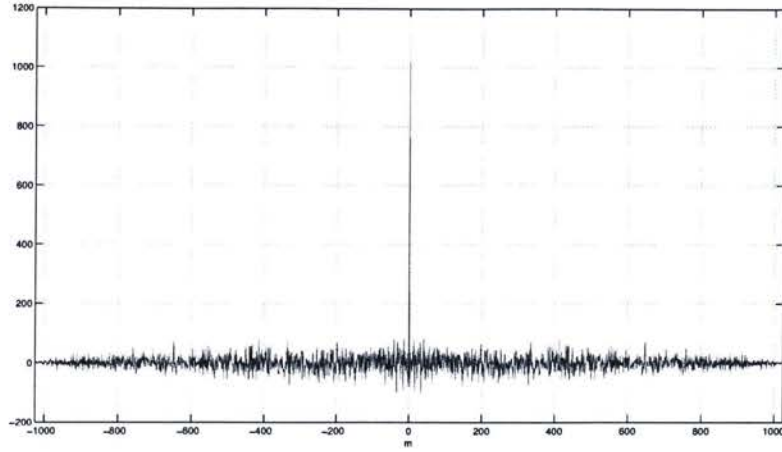


FIGURE 4.2. Autocovariance of noise.

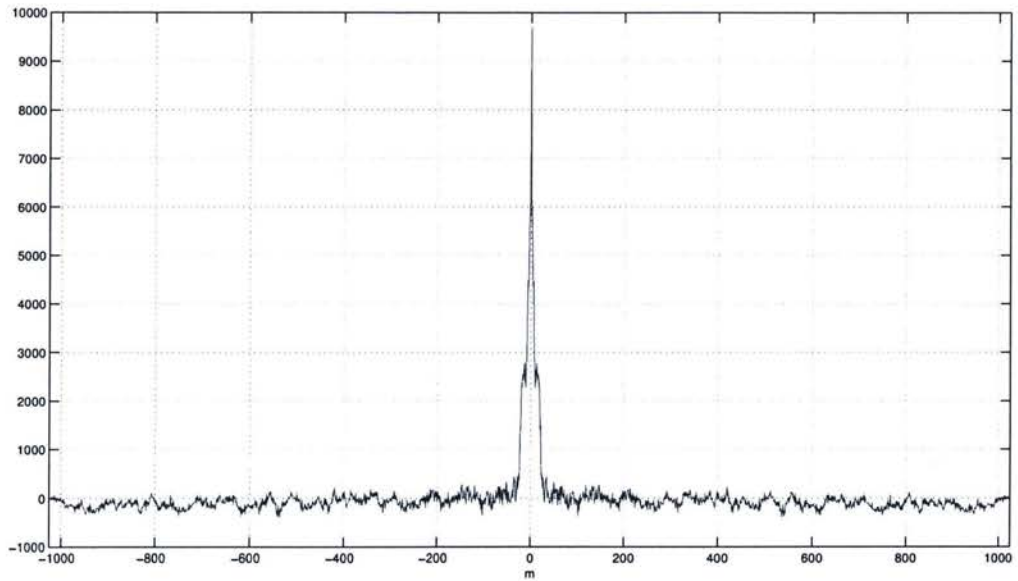


FIGURE 4.3. Autocovariance of noisy signal.

For each test variance, a proper threshold is calculated using the PNIST or ANIST (even MNDL) approaches, and the threshold is applied to acquire an estimate of the noiseless data. This estimate can be used to find an estimate of the noise vector:

$$\omega_{\sigma_{test}}[n] = y[n] - \hat{y}_{\sigma_{test}}[n] \quad (4.6)$$

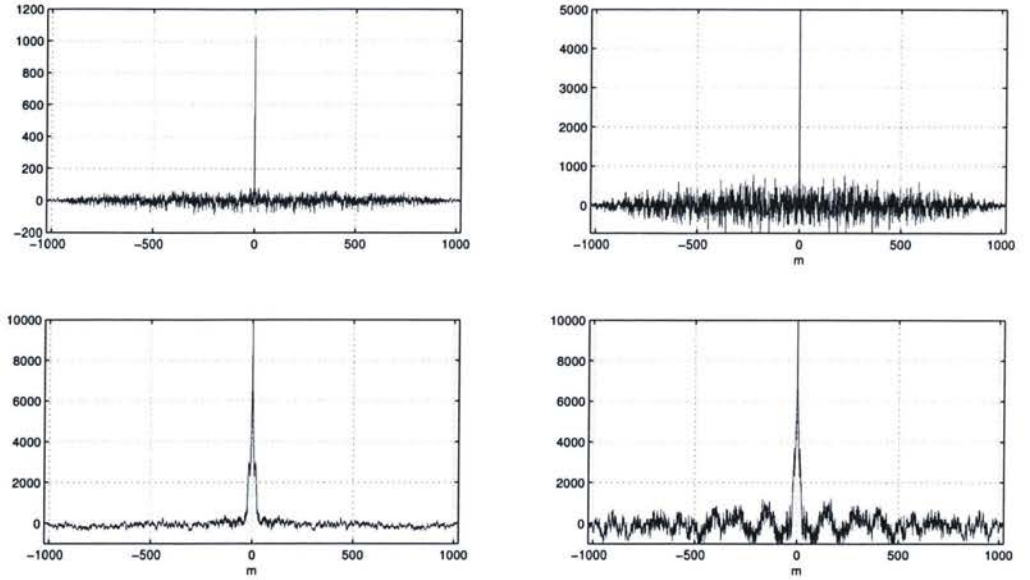


FIGURE 4.4. Autocovariance graph for different test variances, from left to right: (a) $\sigma^2 = 1$ (b) $\sigma^2 = 4$ (c) $\sigma^2 = 9$ (d) $\sigma^2 = 16$. True noise variance is $\sigma^2 = 9$.

Where y is the vector of observed data and \hat{y} is an estimate of noiseless data. The estimate of the autocovariance of this noise is

$$C_{\sigma_{test}}(m) = \begin{cases} \frac{1}{N-m} \sum_{n=0}^{N-|m|-1} \omega_{\sigma_{test}}(n+m) \omega_{\sigma_{test}}(n) & m \geq 0 \\ C_{\sigma_{test}}^*(-m) & m < 0 \end{cases} \quad (4.7)$$

Autocovariance of the estimated noise for some of the test variances is shown in Figure 4.4. The area under the autocovariance curve can be estimated using the following relation for different values of σ_{test} :

$$A_{\sigma_{test}}^2 = \int_{-\infty}^{\infty} C_{\sigma_{test}}^2 dt \quad (4.8)$$

In discrete cases such as our example, l_2 -norm of the estimated noise vector can be used as a measure of the area under the curve:

$$A_{\sigma_{test}}^2 \approx \sum_{k=1}^N C_{\sigma_{test}}^2[k] \quad (4.9)$$

As test variance increases, we observe an increase in the area under the curve of autocovariance of the $\omega_{\sigma_{test}}$, which is due to the presence of noiseless signal

coefficients. To find the best estimate of the noise vector, we need to eliminate the signal coefficients as much as possible. Therefore, the proper estimate of the noise vector, and the proper estimate of the noise variance happens where the area under the curve of the autocovariance of the noise estimate sees a jump in value due to presence of signal coefficients. This sudden increase in the area under the curve of the autocovariance can be detected with an experimental sensitivity condition

$$H[\sigma_{test}] = \frac{A(\sigma_{test} + 0.5)}{A(\sigma_{test})} \quad (4.10)$$

For example in our observations we noticed that the jumping ratio of 10 would provide a reasonable estimate of the noise variance and variance is estimated as:

$$\hat{\sigma}_\omega = \arg \min_{\sigma_{test}} \{ \sigma_{test} \mid H[\sigma_{test}] \geq 10 \} \quad (4.11)$$

Figure 4.5 shows a variance estimation scenario with Mishmash signal, where the true noise is $\sigma^2 = 4$, and the method of denoising is PNIST. The estimated variance is $\hat{\sigma}_\omega^2 = 3.5$.

Figure 4.6 shows an estimation case where a Blocks signal was contaminated by Gaussian noise with a variance of $\sigma^2 = 8$, and the denoising method is PNIST. It can be seen that the l_2 -norm value leaves zero axis at around $\sigma^2 = 7$, and the estimate in this case is $\sigma^2 = 7.5$. Table 4.1 shows the result of our variance estimation method in PNIST for a Blocks signal compared to MNDL variance estimation for different noise powers.

4.3 Noiseless Data with Gaussian Properties

In image denoising applications, usually a threshold is found for each subband assuming a Generalized Gaussian Distribution (GGD), and in general, a large class of noiseless signals show Gaussian behavior [3]. One example of this class

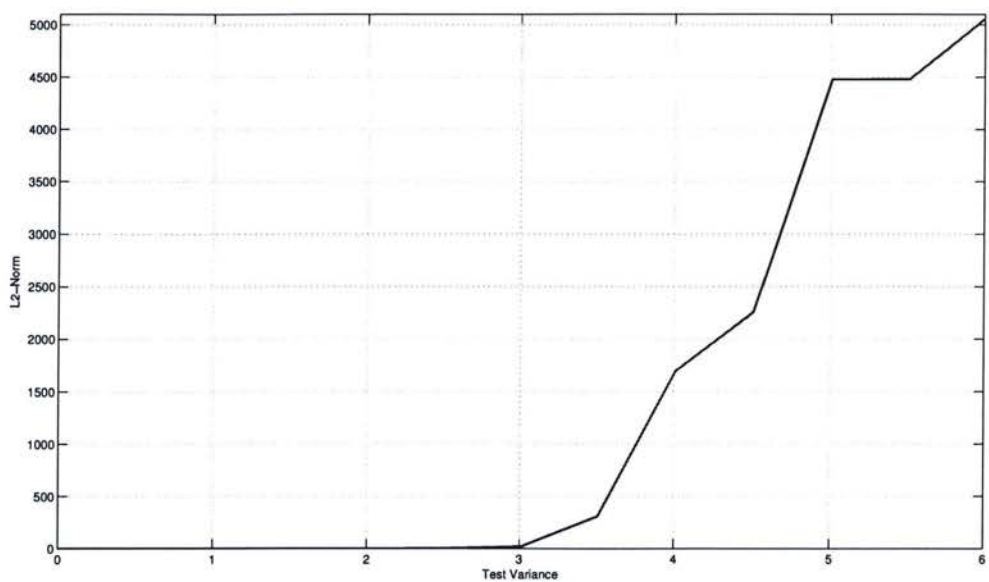


FIGURE 4.5. Area under autocovariance of discarded coefficients for true variance of $\sigma^2 = 4$.

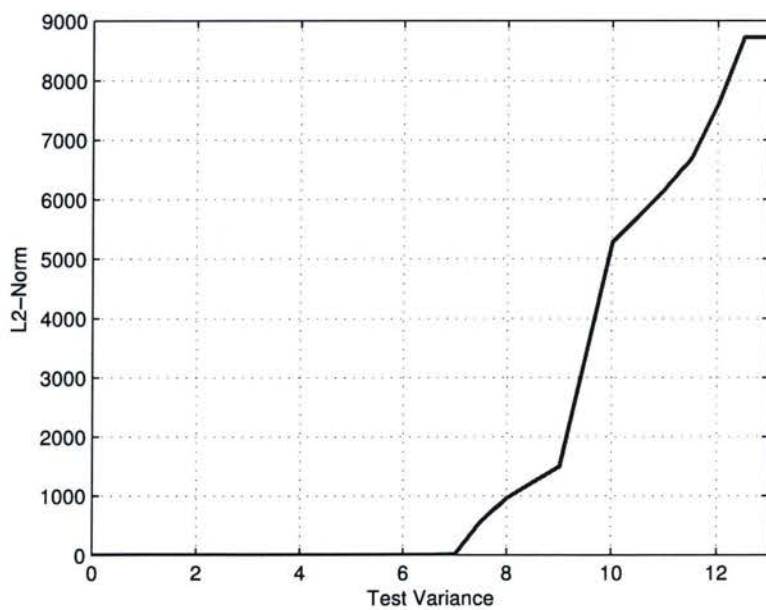


FIGURE 4.6. Area under autocovariance of discarded coefficients for true variance of $\sigma^2 = 8$.

True Variance	Proposed	MNDL
$\sigma^2 = 1$	2.5	5.5
$\sigma^2 = 3$	2	8
$\sigma^2 = 4$	3.5	12
$\sigma^2 = 5$	7.5	10
$\sigma^2 = 6$	9	13.5
$\sigma^2 = 9$	11	14

TABLE 4.1. Estimated variance by the proposed method.

is the Mishmash signal, which is zero mean and has a large number of non-zero coefficients that follow the behavior of the Gaussian distribution as shown in Figure 4.7. The figure shows the histogram of Mishmash signal coefficients. As the figure depicts, the distribution of the coefficients is very similar of that of a Gaussian signal of the same length and same power(variance). Figures 4.8 and 4.9 show sorted coefficients of a noisy Mishmash signal and sorted coefficients of a signal with Gaussian distribution with the same energy. As can be seen, the two sorted signals nearly overlap. The overlap becomes more evident when more data are used as sample data length is increased from 2^{10} in Figure 4.8 to 2^{20} in Figure 4.9.

This property inspires us to study in depth the behavior of a random Gaussian signal contaminated by a Gaussian noise; this can be a general problem of any signal of Gaussian class contaminated by noise. Note that the analysis is point wise as we know that the signal of image correlation makes it a well colored signal compared to the white additive noise.

4.3.1 Conditional Distributions and Data Denoising

Consider the following observed noisy signal

$$y = x + n \quad (4.12)$$

Where both the noiseless signal x and the noisy part n are zero mean Gaussian signals with variances σ_x^2 and σ_n^2 respectively. Therefore, the noisy signal y itself is

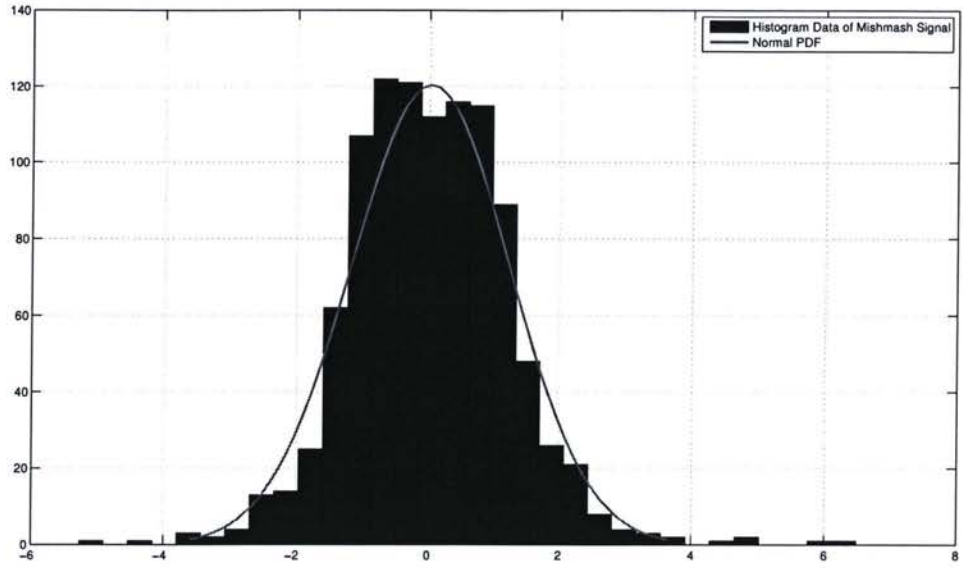


FIGURE 4.7. Histogram of Mishmach coefficients.

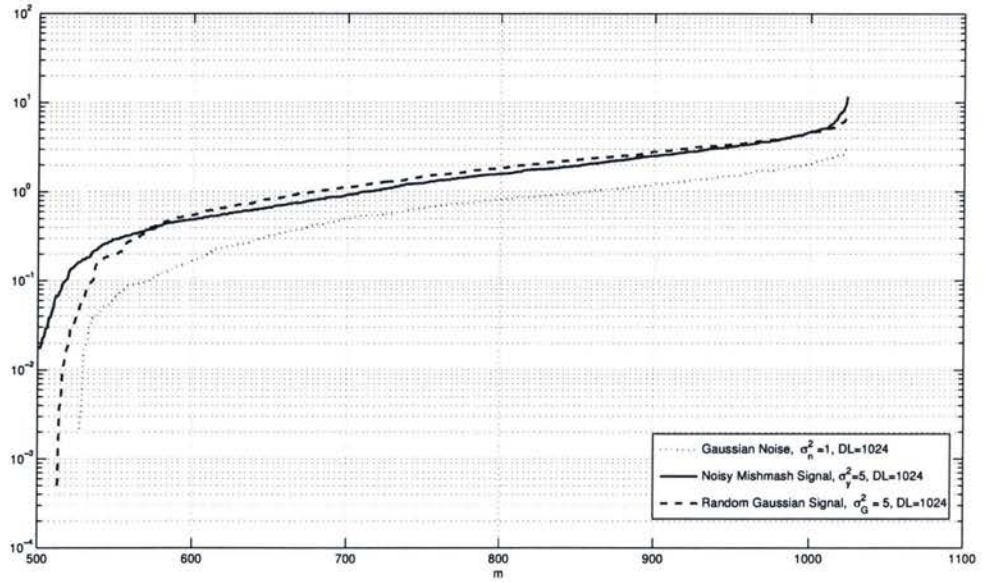


FIGURE 4.8. Noisy Gaussian signal, noisy Mishmach signal and Gaussian noise compared with 2^{10} samples.

a Gaussian signal with variance σ_y^2 .

$$\sigma_y^2 = \sigma_x^2 + \sigma_n^2 \quad (4.13)$$

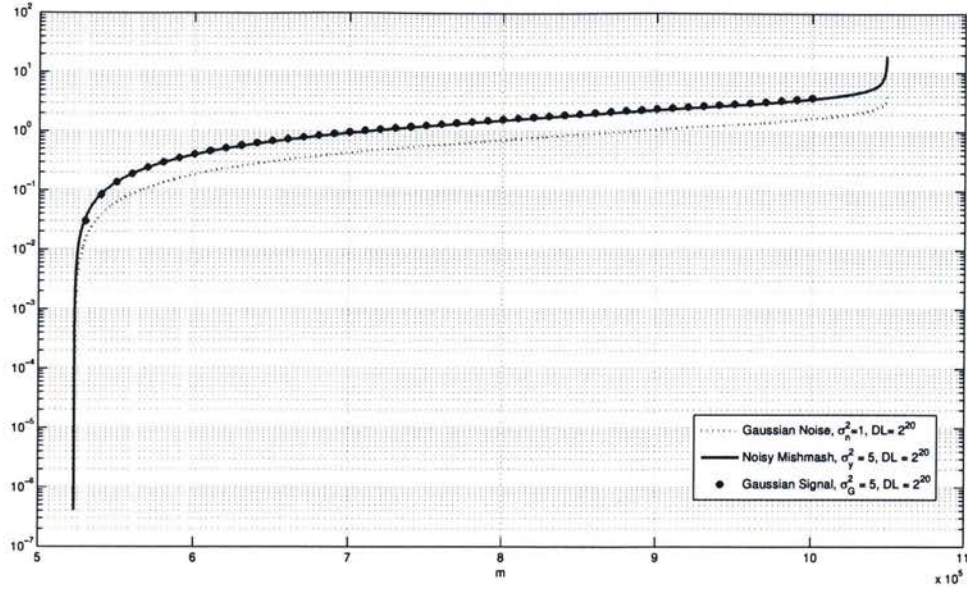


FIGURE 4.9. Noisy Gaussian signal, noisy Mishmash signal and Gaussian noise compared with 2^{20} samples.

In this case, given the observed $y = a$, we calculate the conditional density of the noiseless data, $f(x|y = a)$

$$f(x|y = a) = \frac{f(x, y = a)}{f(y = a)} = \frac{f(x)f(n = a - x)}{f(y = a)} \quad (4.14)$$

$$= \frac{\frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{x^2}{2\sigma_x^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(a-x)^2}{2\sigma_n^2}}}{\frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{a^2}{2\sigma_y^2}}} \quad (4.15)$$

The conditional distribution is

$$f(x|y = a) = \frac{1}{\sqrt{2\pi}} \frac{\sigma_y}{\sigma_x \sigma_n} e^{-\frac{(x - a \frac{\sigma_x^2}{\sigma_y^2})^2}{2 \frac{\sigma_x^2 \sigma_n^2}{\sigma_y^2}}} \quad (4.16)$$

which is a Gaussian distribution with the mean of $\mu_{x|y} = a \frac{\sigma_x^2}{\sigma_y^2}$ and the standard deviation of $\sigma_{x|y} = \frac{\sigma_n \sigma_x}{\sigma_y}$. A Complete proof may be found at Appendix B. Similarly

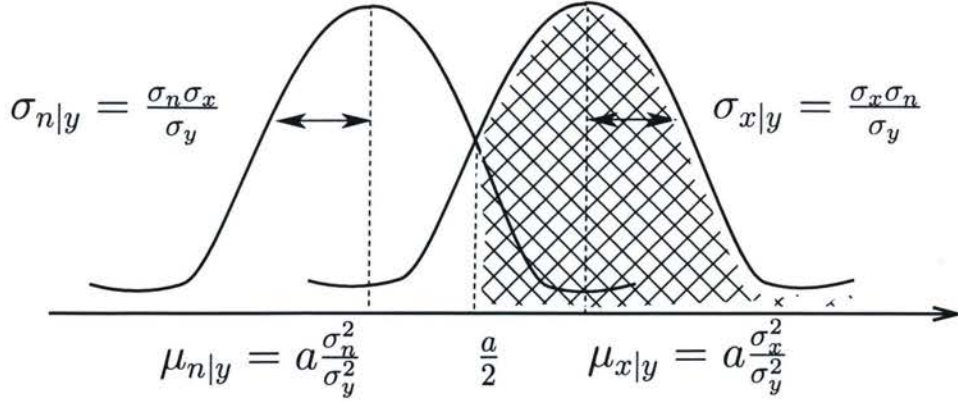


FIGURE 4.10. Distribution of $f(n|y=a)$ & $f(x|y=a)$.

the conditional distribution of the noise part is:

$$f(n|y=a) = \frac{1}{\sqrt{2\pi} \sigma_x \sigma_n} e^{-\frac{(n - a \frac{\sigma_n^2}{\sigma_y^2})^2}{2 \frac{\sigma_n^2 \sigma_x^2}{\sigma_y^2}}} \quad (4.17)$$

which is a Gaussian distribution with mean of $\mu_{n|y} = a \frac{\sigma_n^2}{\sigma_y^2}$ and std $\sigma_{n|y} = \frac{\sigma_n \sigma_x}{\sigma_y}$.

Therefore, we can check for which cases the probability that $x \leq n$ given an observed data $y = a$ is larger than 0.5.

$$P(x > n|y=a) \geq 50\% \quad (4.18)$$

$$P(x > n) = p(x > a - x) = p(2x > a) = p(x > \frac{a}{2}) \geq \frac{1}{2} \quad (4.19)$$

and we have

$$p(x > \frac{a}{2}) = p(x - \frac{a}{2} > 0) \geq \frac{1}{2} \Rightarrow \frac{a}{2} - \mu_{x|y} \leq 0 \quad (4.20)$$

and the criterion is found as:

$$\frac{a}{2} - a \frac{\sigma_x^2}{\sigma_y^2} \leq 0 \Rightarrow \sigma_y^2 \geq 2\sigma_n^2 = \sigma_x^2 \geq \sigma_n^2 \quad (4.21)$$

Note that the condition is independent of the observation value, and it means that if we want to be at least 50% sure that the original signal coefficient is larger

than the noise coefficient, we should check the condition proposed in equation (4.21). Therefore, this condition can be used to reject the data that is too noisy to be denoised.

4.4 Concluding Remarks

In this chapter a new variance estimation method was proposed. The method is applicable for the proposed NIST methods as well as for MNDL denoising. Also, the Gaussian noiseless signals were studied and based on an analytical approach, an important data rejection condition in data denoising was provided. The method can provide not only the condition to detect whether the data can be denoised but also provides the condition of subband rejection in wavelet denoising. This condition provides a more conservative criterion for subband rejection of noisy data than that of BayesShrink.

Chapter 5

Image Denoising

In this chapter we compare the performance of the proposed algorithms, PNIST and ANIST, with the existing algorithms for an example of two dimensional data in image denoising. Soft thresholding was used for all algorithms as soft thresholding has better results visually, and it is proved that soft thresholding achieves nearly Minimax error rate in most cases [9]. Similar to the simulation in previous chapters, *Haar* wavelet with five level of decomposition is used.

Some methods propose a different threshold for every subband. This is called subband dependent thresholding and is commonly used in BayesShrink and SureShrink. However, as we are dealing with noise invalidation in PNIST and ANIST and the properties of noise do not vary from one subband to another, these approaches provide a global threshold for the image and pre-process every subband of the denoised image using equation (4.21). The pre-processing rejects those subbands which are too noisy and contain little information about original data.

Figure 5.1 shows an image of Barbara, a noisy version of Barbara and Barbara denoised by ANIST and PNIST.

Figure 5.2 shows the famous Cameraman image denoised by BayesShrink, SureShrink, PNIST and ANIST. The visual quality of the denoised images by PNIST and ANIST clearly outperform the SureShrink image and are comparable with those of BayesShrink and MNDL.

MSE comparison of the proposed methods for Cameraman and Barbara can be found in Table 5.1. As the table shows both PNIST and ANIST provide acceptable results compared to other shrinkage methods.



Original



Noisy, $\sigma^2=225$



PNIST



ANIST

FIGURE 5.1. Barbara denoised by ANIST and PNIST, $\sigma_n^2 = 225$.

Cameraman	SURE	MNDL	BAYES	PNIST	ANIST
$\sigma_n^2 = 25$	18.3	15.9	17.1	16.2	17.9
$\sigma_n^2 = 100$	52.9	49.2	48.2	46.9	47.8
$\sigma_n^2 = 225$	93.8	95.1	101.9	90.2	107.3
Barbara					
$\sigma_n^2 = 25$	19.1	18.9	16.7	17.3	17.7
$\sigma_n^2 = 100$	59.9	51.6	48.3	49.5	47.6
$\sigma_n^2 = 225$	118.1	89.2	86.7	91.2	101.8

TABLE 5.1. MSE for different Methods compared.

Einstein and Lena denoised by common methods can be found in Figures 5.3 and 5.4.



Original



Noisy, $\sigma_n = 10$



BayesShrink



MNDL



PNIST



ANIST

FIGURE 5.2. Cameraman denoised, $\sigma_n^2 = 100$.

While PNIST outperforms ANIST in terms of MSE in most cases, it can be seen in Figures 5.1 and 5.3 that ANIST seems to have a more pleasing visual quality.



Original



Noisy, $\sigma_n = 20$



BayesShrink



MNDL

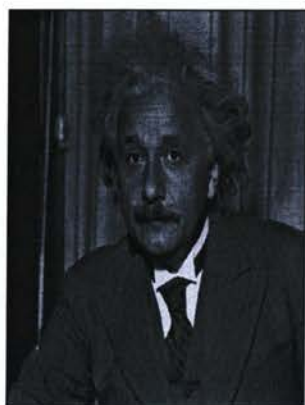


PNIST



ANIST

FIGURE 5.3. Lena denoised, $\sigma_n^2 = 400$.



Original



Noisy, $\sigma_n = 30$



BayesShrink



MNDL



PNIST



ANIST

FIGURE 5.4. Einstein denoised, $\sigma_n^2 = 900$.

Chapter 6

Conclusion

Two new thresholding techniques based on noise invalidation were proposed. The new methods, ANIST and PNIST, are based on Gaussian invalidation of coefficients of the observed data out of the noise region, where the coefficients represent the transformation of the data with a set of orthonormal bases. It was shown that the behavior of sorted absolute coefficients in PNIST and sorted square added coefficients in ANIST can be approximated with Gaussian distribution. This property was used to select a threshold as the first coefficient that is out of the noise region with high certainty. We studied ANIST and PNIST in terms of their performance for two classes of signals: Blocks signal as the representative of signals with a high number of zero coefficients and Mishmash signal as the representative of signals with a high number of non-zero coefficients. Our study shows that ANIST and PNIST can perform acceptably regardless of the original signal type in comparison with other shrinkage methods. We proposed a new noise variance estimation technique and showed that the method provides acceptable results and outperforms the recently proposed MNDL noise variance estimation method.

We extended the methods to two dimensional data, and the proposed image denoising methods were compared with other shrinkage methods. It was shown that ANIST and PNIST have comparable Mean Error Square (MSE) with BayesShrink and outperform other methods with a pleasing visual quality.

Future research could focus on extending ANIST for Magneto Resonance Images (MRI). It is a very well documented fact that MR images are corrupted by *Rician* noise due to defects in imaging instruments [31, 16]. As was shown ANIST can

be a solution to non Gaussian noise problems; thus ANIST can be extended as a general solution for this type of problems as well as any other applications where Gaussian noise solutions can not be applied.

Appendix

A Distribution of Order Statistics of a Random Variable

Let X_1, X_2, \dots, X_n be independent identically distributed random variables, and $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ be the corresponding order statistics. Assuming $f(x)$ is probability distribution function of random variable X_i and $F(x)$ is the corresponding cumulative distribution function, then probability distribution function of $X_{(k)}$ is:

$$f_{X_{(k)}} = \frac{d}{dx} F_{X_{(k)}} = \frac{d}{dx} P(X_{(k)} \leq x) = \frac{d}{dx} P(\text{at least } k \text{ of the } n \text{ } X\text{s are } \leq x) \quad (\text{A-1})$$

$$= \frac{d}{dx} P(\geq k \text{ success in } n \text{ trials}) \quad (\text{A-2})$$

$$= \frac{d}{dx} \sum_{j=k}^n \binom{n}{j} P(X_1 \leq x)^j (1 - P(X_i \leq x))^{(n-j)} \quad (\text{A-3})$$

$$= \frac{d}{dx} \sum_{j=k}^n \binom{n}{j} F(x)^j (1 - F(x))^{n-j} \quad (\text{A-4})$$

$$= \sum_{j=k}^n \left(n \binom{n-1}{j-1} F(x)^{j-1} (1 - F(x))^{n-j} - n \binom{n-1}{j} F(x)^j (1 - F(x))^{n-j-1} \right) f(x) \quad (\text{A-5})$$

$$= n f(x) \left(\sum_{j=k-1}^{n-1} \binom{n-1}{j} F(x)^j (1 - F(x))^{(n-1)-j} - \sum_{j=k}^{n-1} \binom{n-1}{j} F(x)^j (1 - F(x))^{(n-1)-j} \right) \quad (\text{A-6})$$

After simplifying the results, all terms cancel out each other except the first and the last, thus we have:

$$\begin{aligned}
&= nf(x) \left(\binom{n-1}{k-1} F(x)^{k-1} (1-F(x))^{(n-1)-(k-1)} \right. \\
&\quad \left. - \underbrace{\binom{n-1}{n} F(x)^n (1-F(x))^{(n-1)-n}} \right) \tag{A-7}
\end{aligned}$$

and the term over underbrace is zero thus:

$$f_{X_{(k)}}(x) = nf(x) \binom{n-1}{k-1} F(x)^{k-1} (1-F(x))^{n-k} \tag{A-8}$$

B Gaussian Data with Additive Noise

$$f(x|y=a) = \frac{f(x, y=a)}{f(y=a)} = \frac{f(x)f(n=a-x)}{f(y=a)} \quad (\text{B-1})$$

$$= \frac{\frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{x^2}{2\sigma_x^2}} \cdot \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{(a-x)^2}{2\sigma_n^2}}}{\frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{a^2}{2\sigma_y^2}}} \quad (\text{B-2})$$

$$= \frac{1}{\sqrt{2\pi} \frac{\sigma_n \sigma_x}{\sigma_y}} e^{-\frac{-\sigma_x^2 \sigma_y^2 (a-x)^2 - \sigma_n^2 \sigma_y^2 x^2 + a^2 \sigma_x^2 \sigma_n^2}{2\sigma_n^2 \sigma_x^2 \sigma_y^2}} \quad (\text{B-3})$$

$$= \frac{1}{\sqrt{2\pi} \frac{\sigma_x \sigma_n}{\sigma_y}} e^{-\frac{a^2 (\sigma_x^2 \sigma_n^2 - \sigma_x^2 \sigma_y^2) - x^2 (\sigma_n^2 \sigma_y^2 + \sigma_y^2 \sigma_x^2) + 2ax \sigma_x^2 \sigma_y^2}{2\sigma_n^2 \sigma_x^2 \sigma_y^2}} \quad (\text{B-4})$$

We have:

$$\sigma_y^2 = \sigma_x^2 + \sigma_n^2 \quad (\text{B-5})$$

thus:

$$(\sigma_x^2 \sigma_n^2 - \sigma_x^2 \sigma_y^2) = -\sigma_x^4 \quad (\text{B-6})$$

$$(\sigma_n^2 \sigma_y^2 + \sigma_y^2 \sigma_x^2) = \sigma_y^4 \quad (\text{B-7})$$

and by replacing equation (B-6) and equation (B-7) in equation (B-4) we have:

$$f(x|y=a) = \frac{1}{\sqrt{2\pi} \frac{\sigma_x \sigma_n}{\sigma_y}} e^{-\frac{(a\sigma_x^2 - x\sigma_y^2)^2}{2\sigma_n^2 \sigma_x^2 \sigma_y^2}} \quad (\text{B-8})$$

$$= \frac{1}{2\pi \frac{\sigma_x \sigma_n}{\sigma_y}} e^{-\frac{-(a\frac{\sigma_x^2}{\sigma_y^2} - x)^2}{2 \frac{\sigma_n^2 \sigma_x^2}{\sigma_y^2}}} \quad (\text{B-9})$$

and $f(x|y=a)$ follows Gaussian distribution with $\mu_{x|y=a} = a\frac{\sigma_x^2}{\sigma_y^2}$ and $\sigma_{x|y=a} = \frac{\sigma_x \sigma_n}{\sigma_y}$.

References

- [1] S. Beheshti, N. Nikvand, X. N. Fernando, "Soft thresholding by noise invalidation", Proceedings of the 24th Biennial Symposium on Communications, pp. 235 – 238, June 2008.
- [2] S. Beheshti and M.A. Dahleh, "A new information theoretic approach to signal denoising and best basis selection", *IEEE Transactions on Signal Processing*, vol. 53, No. 10, pp. 3613 - 3624, Oct. 2005.
- [3] S. G. Chang, Bin Yu, M. Vetterli, "Adaptive Wavelet Thresholding for Image Denoising and Compression", *IEEE Transactions on Image Processing*. vol. 9, Issue 9, pp. 1532 - 1546, Sep. 2000.
- [4] S. G. Chang, B. Yu, and M. Vetterli, "Image denoising via lossy compression and wavelet thresholding", *IEEE Transactions on Image Processing*, vol. 9, pp. 1532 – 1546, Sep. 2000.
- [5] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising,". In *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, Eds. Berlin, Germany: Springer-Verlag, 1995.
- [6] I. Daubechies, "Ten Lectures On Wavelets", Vol. 61 of Proc. CBMS-NSF Regional Conference Series in Applied Mathematics. Philadelphia, PA: SIAM, 1992.
- [7] J. Dedecker, P. Doukhan, G. Lang, R. Leon, J.R., S. Louhichi, C. Prieur, "Weak Dependence: With Examples and Applications", *Lecture Notes in Statistics*, Springer, vol. 190, 2007.
- [8] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage", *Journal of the American Statistical Association* vol. 90, pp.1200 - 1224, 1995.
- [9] D. L. Donoho, "De-noising by soft-thresholding", *IEEE Transactions on Information Theory*, vol. 41, Issue 3, pp. 613 - 627, May 1995.
- [10] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage", *Biometrika*, vol. 81, pp. 425 - 455, 1994.
- [11] D. L. Donoho and I. M. Johnstone, "Ideal denoising in an orthonormal basis chosen from a library of bases", *Compt. Rend. Acad. Sci. Paris Ser. A*, 319, pp. 13171322, 1994.
- [12] M. Elad and M. Aharon, "Image denoising via sparse and redundant representation over learned dictionaries", *IEEE Transactions on Image Processing*, vol. 15, No. 12, pp. 3736 – 3745, Dec. 2006.

- [13] A. Fakhrzadeh and S. Beheshti, "Minimum noiseless description length (MNDL) thresholding", *Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Image and Signal Processing*, 2007.
- [14] W. Feller, "The fundamental limit theorems in probability", *Bulletin of the American Mathematical Society*, vol. 51, pp. 800 – 832, 1945.
- [15] D. Gabor, "Theory of Communication", *In Journal of IEE*, vol. 93, pp. 429 – 441, 1946.
- [16] R. L. Gregg, R.D. Nowak, "Noise removal methods for high resolution MRI", *Proceedings of the 1997 IEEE Medical Imaging Conference*, vol. 2, pp. 1117 – 1121, Nov. 1997.
- [17] Hubbard, B. Burke, "*The world according to wavelets*", Wellesey, MA (USA), 1996.
- [18] I. Hussain, Hujun Yin, "A novel wavelet thresholding method for adaptive image denoising", *Proceedings of the 3rd International Symposium on Communications, Control and Signal Processing, ISCCSP 2008*, pp. 1252 – 1256, Mar. 2008.
- [19] M. Jansen and A. Bultheel, "Empirical bayes approach to improve wavelet thresholding for image noise reduction", *Journal of American Staistical Association* vol. 96, Issue 454, pp. 629639, Jun. 2001.
- [20] M. Jansen, M. Malfait, and A. Bultheel, "Generalized Cross Validation for Wavelet Thresholding", *In Signal Processing*, vol. 56, pp. 33 – 44, Jan. 1997.
- [21] Jingdong Chen, J. Benesty, Yiteng Huang, S. Doclo, "New insights into the noise reduction Wiener filter", *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, Issue 4, pp. 1218 – 1234, Jul. 2006.
- [22] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise", *Journal of the Royal Statistical Society. Series B (Methodological)* 59, pp. 319351, 1997.
- [23] A. Kosir, A. Mujcic, N. Suljanovic, J.F. Tasic, "Noise variance estimation based on measured maximums of sampled subsets" *Mathematics and Computers in Simulation*, vol. 65, issue 6, pp. 629-639, May 2004.
- [24] J. S. Lee, "Digital image enhancement and noise filtering by use of local statistics", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-2, pp. 165 – 168, Mar. 1980.
- [25] J. S. Lim, A.V. Oppenheim, "Enhancement and bandwidth compression of noisy speech", *Proceedings of the IEEE* vol. 67, Issue 12, pp. 1586 – 1604, Dec. 1979.

- [26] D. Makovoz, "Noise Variance Estimation In Signal Processing", Proceedings of the 2006 International Symposium on Signal Processing and Information Technology, pp. 364 – 369, Aug. 2006.
- [27] S. G. Mallat, "*A Wavelet Tour Of Signal Processing*", Elsevier Science & Technology Books, Sep. 1999.
- [28] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation", *IEEE Transactions on Pattern Analysis. Machine Intelligence*, vol. 11, pp. 674 – 693, Jul. 1989.
- [29] M. K. Mihcak, I. Kozintsev, K. Ramchandran, P. Moulin, "Low-Complexity Image Denoising Based on Statistical Modeling of Wavelet Coefficients", *IEEE Signal Processing Letters*, vol. 6, no. 12, Dec. 1999.
- [30] P. Moulin, J. Liu, "Analysis of multiresolution image denoising schemes using a generalized Gaussian and complexity priors", *IEEE Transactions on Information Theory*, vol. 45, pp. 909 – 919, 1999.
- [31] R. D. Nowak, "Wavelet-based Rician noise removal for magnetic resonance imaging", *IEEE Transactions on Image Processing*, vol. 8, Issue 10, pp. 1408 – 1419, Oct. 1999.
- [32] S.I. Olsen, "Estimation of Noise in Images: An Evaluation", *CVGIP: Graphical models and Image Processing*, vol. 55, No. 4, pp. 391-323, Jul. 1993.
- [33] K. K. Paliwal, "Estimation of noise variance from the noisy AR signal and its application in speech enhancement", *IEEE Transactions on Acoustics, Speech, and Signal Processing* vol. 36, Issue 2, pp. 292 – 294, Feb. 1988.
- [34] J. Portilla, Vasily Strela, Martin J. Wainwright, Eero P. Simoncelli, "Image Denoising using scale Mixtures of Gaussians in the Wavelet Domain", *IEEE Transactions on Image Processing*, vol. 12, No. 11, pp. 1338 – 1351, Nov. 2003.
- [35] K. Rank, M. Lendl, and R. Unbehauen, "Estimation of image noise variance", Proceedings of the IEE Vision, Image, and Signal Processing, vol. 146, Issue 2, pp. 80 – 84, Aug. 1999.
- [36] J. Rissanen, "Minimum description length denoising", *IEEE Transactions on Information Theory*, vol. 46, pp. 2537 – 2543, 2000.
- [37] F. Ruggeri and B. Vidakovic, "A Bayesian Decision Theoretic Approach to Wavelet Thresholding". *Journal of American Statistical Association*, vol. 93, pp. 183 – 197, 1998.
- [38] Y. Sheng, "Wavelet Transform", In: The transforms and applications handbook, Ed. by A. D. Poularikas, CRC Press, The Electrical Engineering Handbook Series, 1996, pp. 747 – 827.

- [39] E. P. Simoncelli, "Bayesian Denoising of Visual Images in the Wavelet Domain", In: Bayesian Inference in Wavelet Based Models, P. Muller and B. Vidakovic, Chapter 18, Lecture Notes in Statistics vol. 141, Sep. 1999, pp. 291 – 308.
- [40] E. P. Simoncelli, "Statistical Models for Images: Compression, Restoration and Synthesis", Proceedings of the 31st Asilomar Conference on Signals, Systems and Computers, Nov. 1997, pp. 673 – 678.
- [41] H. F. Trotter, "An Elementary Proof of the Central Limit Theorem", *Archiv der Mathematik*, vol. 10, pp. 226 – 334, 1959.
- [42] M. Vetterli and J. Kovacevic, "*Wavelets and Subband Coding*", in Englewood Cliffs, NJ Prentice-Hall, 1995.
- [43] B. Vidakovic, "Nonlinear Wavelet Shrinkage with Bayes Rules and Bayes Factors", *Journal of American Statistical Association*, vol. 93, no. 441, pp. 173 – 179, 1998.
- [44] Y. Washizawa, Y. Yamashita, "Non-linear Wiener filter in reproducing kernel Hilbert space", Proceedings of the 18th International Conference on Pattern Recognition, ICPR, vol. 1, pp. 967 – 970, 2006.