Stochastic Thresholding: An approach to Estimator Optimization via Fisher Information Maximization

S. Dasgupta
Department of Electronics and Electrical Communication Engineering,
Indian Institute of Technology,
Kharagpur 721302, India.
e-mail: sdg_kgp@yahoo.co.in

***Abstract*— In stochastic thresholding, the threshold for quantization of a signal is randomized. An estimator based on quantized signal data can be optimized through stochastic thresholding. By controlling certain parameters of the probability distribution function of the threshold, we can achieve a gain in Fisher information, a measure of efficacy of any unbiased estimator. Thus any conceivable estimator based on quantized signal will perform better than an estimator operating on original signal provided the stochastic thresholding scheme is followed. Both bi-level and tri-level quantization cases are discussed. The optimization is illustrated by a Maximum Likelihood Estimator to estimate the amplitude of a sinusoid drowned in heavy noise. Stochastic thresholding can also be used to maximize output SNR. This is illustrated by applying 3-level quantization with stochastic thresholding on a digital image. Contrary to intuition, it is seen that loss of information through quantization can be minimized via randomizing the threshold of comparison.**

***Index Terms*— Estimator Optimization, Fisher Information, Stochastic thresholding, Maximim Likelihood Estimation, Non-linear Signal Processing.**

I. Introduction

Quantization is an essential step in digital signal processing. Issues like storage, noise immunity and ease in communication demand a small number of quantization levels, often two or three. Hence, the quantizer has to be designed in such a way that the output signal preserves as much information as possible. Recent approaches to this end have employed threshold variation and/or Stochastic Resonance [8] where Fisher Information has been the criterion for quantizer optimization. However, these methods did not yield a gain in Fisher information for Gaussian noise which is the most common type of noise encountered in nature. In this paper we show how Stochastic Thresholding achieves a Fisher Information gain greater than unity for Gaussian noise. We study both bi-level and tri-level quantization. To illustrate the method we optimize a Maximum Likelihood estimator. We study in detail the performance of the ML estimator — how it depends on the parameter for optimization and its robustness against noise after optimization.

When quantization becomes necessary, we can minimize the loss of information through Stochastic Thresholding. This is illustrated by quantizing a digital image with intensities in the range 0 to 255 to just three levels. We show the existence of an optimal threshold distribution belonging to the Rayleigh family that maximizes output SNR.

The main contributions of the paper are as follows:

- We propose a stochastic thresholding scheme for quantization and show how it achieves a Fisher Information gain greater than unity when the input noise is Gaussian.
- We devise a Maximum Likelihood estimator and optimize it by adjusting the parameter associated with the threshold distribution. We evaluate its performance for an extensive low SNR range.
- We show mathematically and graphically via a digital image how output SNR varies non-monotonically with the threshold distribution.

The paper is organized as follows: Section II introduces the estimation problem at hand. Section III describes the stochastic thresholding scheme proposed. Section IV shows how Fisher Information maximization is possible via bi-level and tri-level quatization. Section V demonstrates the concepts dicussed in Sec-

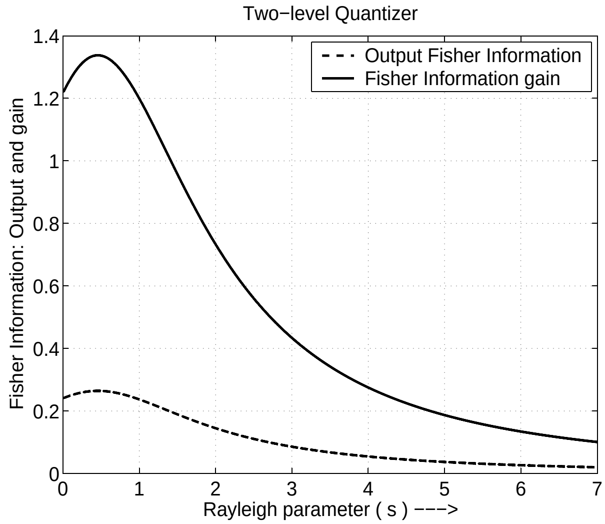

Fig. 1. The graph corresponds to the two-level quantizer. It shows the variation of output Fisher Information J_{out} and Fisher Information gain J_{gain} with the Rayleigh parameter 's'. The gain is seen to be greater than unity. Thus the estimator based on quantized data is expected to perform better.

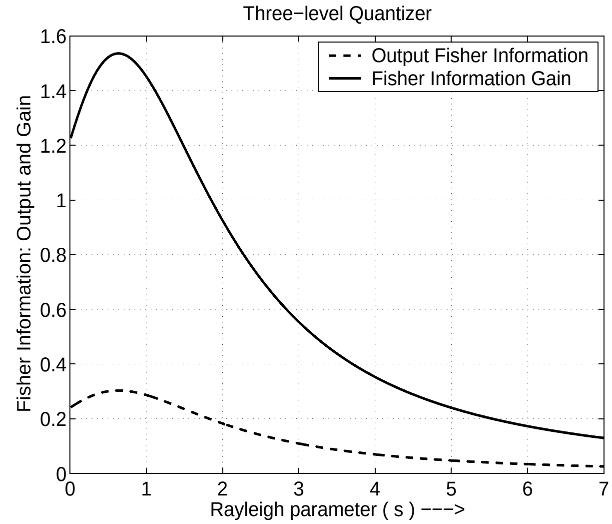

Fig. 2. This graph corresponds to the 3-level quantizer. J_{gain} is again greater than unity. Thus the estimator based on randomly thresholded signal will perform better.

tion IV by optimizing the Maximum Likelihood estimator whose performance after optimization is studied extensively. Concluding remarks are given in Section VII.

II. Problem Definition

We have a signal $s(n)$ corrupted by noise $\eta(n)$. The noise corrupted sequence $x(n)$ is given by:

$$x(n) = s(n) + \eta(n) \tag{1}$$

$\eta(n)$ is an independent and identically distributed zero-mean noise sequence with a gaussian probability distribution function.

$$f_\eta(u) = \frac{1}{\sqrt{2\pi\sigma^2}}\, exp(-\frac{u^2}{2\sigma^2}) \tag{2}$$

The aim is to design an optimal estimator to estimate a parameter 'a' associated with the signal $s(n)$ after quatization. 'a' can be such features as amplitude, frequency, etc.

III. Stochastic Thresholding

The threshold for quantization (γ) is chosen to be a random variable. It follows a Rayleigh Distribution with probability distribution function given by:

$$\begin{aligned} f_\gamma(u) &= \frac{u}{s^2}\, exp(-\frac{u^2}{2s^2}) && \text{if } u \geq 0 \\ &= 0 && \text{otherwise} \end{aligned} \tag{3}$$

Thus the threshold γ becomes a stochastic sequence $\gamma(n)$ with mean $\sqrt{\frac{\pi}{2}}\ s$ and variance $\sqrt{\frac{4-\pi}{2}}\ s^2$. The two-level quantizer is defined as:

$$\begin{aligned} y(n) &= +1 && \text{if s(n)+}\eta(n) \geq \gamma(n) \\ &= -1 && \text{if s(n)+}\eta(n) < \gamma(n) \end{aligned} \tag{4}$$

The three-level quantizer is defined as:

$$\begin{aligned} y(n) &= +1 && \text{if s(n)+}\eta(n) \geq \gamma(n) \\ &= -1 && \text{if s(n)+}\eta(n) \leq -\gamma(n) \\ &= 0 && \text{otherwise} \end{aligned} \tag{5}$$

IV. Fisher Information maximization

The Fisher Information contained in a random variable X with probability distribution f_{θ_0} (which belongs to the family of distributions $\{f_\theta : \theta\ \epsilon\ \Theta\}$) is defined by:

$$J(\theta_0) = E_{\theta_0}\left(\frac{\partial}{\partial\theta} log f(X|\theta_0)\right)^2 \tag{6}$$

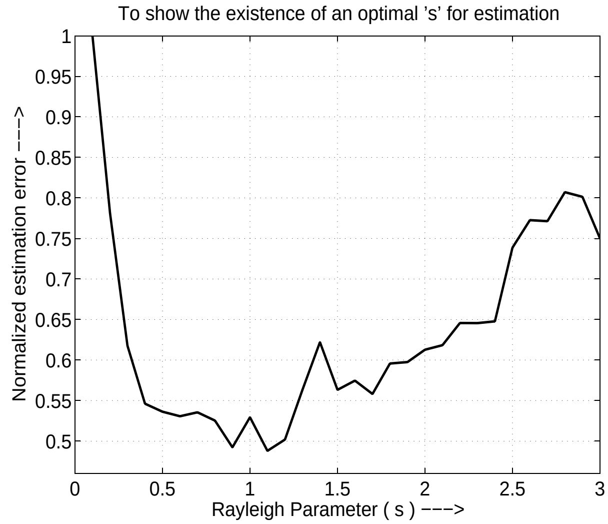

Fig. 3. This graph shows the variation of normalized estimation error with the rayleigh parameter. It shows that an optimal 's' exists where the error hits a bottom after which it rises again.

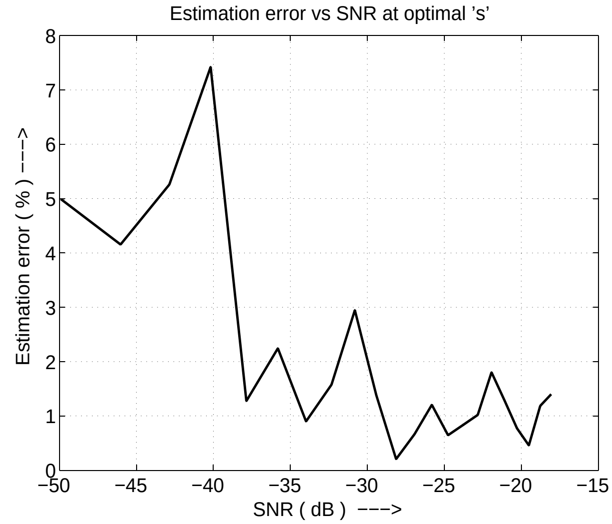

Fig. 4. This graph shows the variation of percentage estimation error with input snr. The Rayleigh parameter has been set to the optimal value. As snr increases the error drops sharply. Even at $-50dB$ it is seen that the error is as small as 5% which shows that the ML estimator has very good performance.

It will be shown that the Fisher Information of the quantized output can be maximized using Stochastic Thresholding by varying the parameter 's' of the threshold p.d.f. which has the effect of simultaneoulsy varying the mean and variance of the threshold. We shall discuss both two level and three level quantizer.

A. Bi-level Quantization

In this case, we take the input signal to be a constant d.c. $s(n) = a$ corrupted with noise. Let q represent the probability that $y(n) = 1$. Thus $1 - q$ is the probability that $y(n) = -1$. Hence, the Fisher Information of the quantized output J_{out} is given by:

$$
\begin{aligned}
J_{out} &= \sum_{z=-1,1} \frac{1}{Pr\{y=z\}} \left(\frac{\partial}{\partial a} Pr\{y=z\} \right)^2 \\
&= \left(\frac{\partial q}{\partial a} \right)^2 \left(\frac{1}{q} + \frac{1}{1-q} \right) \qquad (7)
\end{aligned}
$$

Now, q can be found as:

$$
\begin{aligned}
q &= Pr\{y = +1\} \\
&= Pr\{s(n) + \eta(n) \geq \gamma(n)\} \\
&= Pr\{\eta(n) - \gamma(n) \geq -a\} \qquad (8)
\end{aligned}
$$

To determine the probability distribution function of the new random variable $\alpha = \eta - \gamma$, let us define $\beta = \eta + \gamma$. Thus, $\eta = \frac{\alpha+\beta}{2}$ and $\gamma = \frac{\beta-\alpha}{2}$. The Jacobian $J(\alpha, \beta)$ is given by:

$$
\begin{aligned}
J(\alpha,\beta) &= \begin{vmatrix} \frac{\partial\eta}{\partial\alpha} & \frac{\partial\eta}{\partial\beta} \\ \frac{\partial\gamma}{\partial\alpha} & \frac{\partial\gamma}{\partial\beta} \end{vmatrix} \\
&= \frac{1}{2} \qquad (9)
\end{aligned}
$$

Let $f_{\alpha\beta}(\alpha, \beta)$ denote the joint p.d.f. of α and β. It is given by:

$$
f_{\alpha\beta}(\alpha,\beta) = f_\eta\{\eta(\alpha,\beta)\} f_\gamma\{\gamma(\alpha,\beta)\} |J(\alpha,\beta)| \qquad (10)
$$

Since γ is always greater than zero, $f_{\alpha\beta}(\alpha, \beta) = 0$ if $\beta < \alpha$. Otherwise,

$$
f_{\alpha\beta}(\alpha,\beta) = \frac{\beta-\alpha}{4s^2\sqrt{2\pi\sigma^2}}\, exp\left[-\left\{ \frac{(\alpha+\beta)^2}{8\sigma^2} + \frac{(\beta-\alpha)^2}{8s^2} \right\} \right]
$$

$f_\alpha(\alpha)$ can be found by integration $f_{\alpha\beta}(\alpha, \beta)$ over the range of β:

$$
\begin{aligned}
f_\alpha(\alpha) &= \int_\alpha^\infty f_{\alpha\beta}(\alpha,\beta) d\beta \\
&= \frac{\sigma}{\sqrt{2\pi}(s^2+\sigma^2)}\, exp\left(-\frac{\alpha^2}{2\sigma^2} \right) - \\
&\frac{\alpha s}{2(s^2+\sigma^2)^{3/2}}\, exp\left(\frac{-\alpha^2}{2(\sigma^2+s^2)} \right) erfc\left(\frac{\alpha s}{\sigma\sqrt{2(s^2+\sigma^2)}} \right) \qquad (11)
\end{aligned}
$$

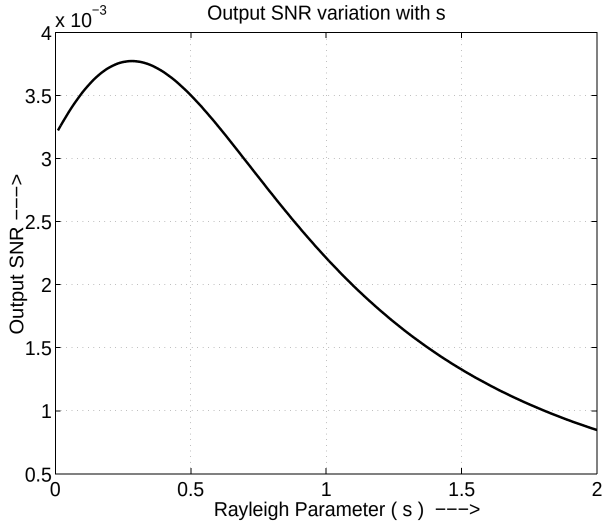

Fig. 5. This graph shows the variation of the output snr with the rayleigh parameter. The variation is non-monotonic which shows that the loss of information due to quantization can be minimized by proper choice of 's'.

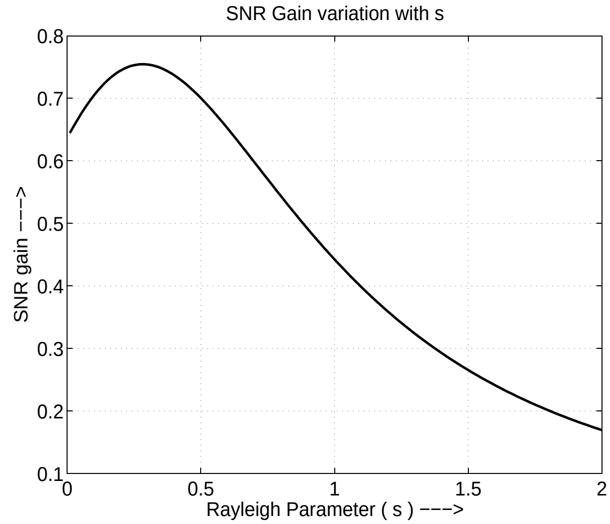

Fig. 6. This graph shows the variation of the snr gain with the rayleigh parameter. Though the dependance is non-monotonic, a gain greater than unity is not achievable.

The cumulative distribution function of α can be obtained as:

$$F_\alpha(\alpha) = \int_{-\infty}^{\alpha} f_\alpha(u)du$$

$$= \frac{1}{2}(1+erf(\frac{\alpha}{\sqrt{2}\sigma}))+\frac{s}{2\sqrt{s^2+\sigma^2}}exp(-\frac{\alpha^2}{2(s^2+\sigma^2)}) (1-erf\frac{\alpha s}{\sigma\sqrt{2(s^2+\sigma^2)}}) \quad (12)$$

The probability that $y=1$, i.e. q, is seen to equal $F_\alpha(-a)$. Also,

$$\frac{\partial q}{\partial a} = -F'_\alpha(-a) \quad (13)$$

Hence,

$$J_{out} = \frac{f_\alpha^2(-a)}{F_\alpha(-a)[1-F_\alpha(-a)]} \quad (14)$$

For the input signal, J_{in} is given by [8]:

$$J_{in} = \frac{1}{\sigma^4} \quad (15)$$

The Fisher Information gain is simply:

$$J_{gain} = \frac{J_{out}}{J_{in}} \quad (16)$$

Figure 1 shows the variation of J_{out} and J_{gain} with the Rayleigh parameter 's'. We see that a gain greater than unity has been achieved. Thus the estimator based on quantized data operating at optimal 's' can be expected to perform better.

B. Tri-level Quantization

In this case, we take the signal $s(n)$ as a sinusoid $a\cos(2\pi fn)$ which is corrupted by zero-mean additive white Gaussian noise sequence $\eta(n)$. The quantized signal is given by:

$$\begin{aligned} y(n) &= +1 && \text{if s(n)+}\eta(n) \geq \gamma(n) \\ &= -1 && \text{if s(n)+}\eta(n) \leq -\gamma(n) \\ &= 0 && \text{otherwise} \end{aligned} \quad (17)$$

The probability distribution function of β can be obtained as:

$$f_\beta(\beta) = \int_{-\infty}^{\beta} f_{\alpha\beta}(\alpha,\beta)d\alpha \quad (18)$$

It can be easily proved that $f_\beta(\beta) = f_\alpha(-\beta)$ from which we obtain

$$f_\beta(\beta) = \frac{\sigma}{\sqrt{2\pi}(s^2+\sigma^2)}\ exp\left(-\frac{\beta^2}{2\sigma^2}\right)\ +$$

$$\frac{\beta s}{2(s^2+\sigma^2)^{3/2}}\ exp(\frac{-\beta^2}{2(\sigma^2+s^2)})\{1+erf(\frac{\beta s}{\sigma\sqrt{2(s^2+\sigma^2)}})\} \tag{19}$$

Using the property $f_\beta(\beta) = f_\alpha(-\beta)$, $F_\beta(\beta)$ can be obtained as:

$$\begin{aligned} F_\beta(\beta) &= \int_{-\infty}^{\beta} f_\beta d\beta \\ &= 1 - F_\alpha(-\beta) \end{aligned} \tag{20}$$

The Fisher Information for the three-level quantizer can be obtained as:

$$\begin{aligned} J_{out} &= \frac{1}{Pr\{y=1\}}\left(\frac{\partial}{\partial a}Pr\{y=1\}\right)^2 + \\ &\quad \frac{1}{Pr\{y=0\}}\left(\frac{\partial}{\partial a}Pr\{y=0\}\right)^2 + \\ &\quad \frac{1}{Pr\{y=-1\}}\left(\frac{\partial}{\partial a}Pr\{y=-1\}\right)^2 \\ &= [\frac{(f_\alpha^+)^2}{1-F_\alpha^+} + \frac{(-f_\alpha^+ + f_\alpha^-)^2}{F_\alpha^+ + F_\alpha^- - 1} + \frac{(f_\alpha^-)^2}{1-F_\alpha^-} \\ &\quad]\cos^2(2\pi fn) \end{aligned} \tag{21}$$

Figure 2 shows the variation of J_{out} and J_{gain} with the Rayleigh parameter 's'. Again we see that a gain greater than unity is possible by choosing 's' properly. Hence, the estimation through 3-level quantized data is expected to give better results than the unquantized data.

V. Maximum Likelihood Estimator

Here we optimize the ML estimator for the amplitude a. Let m be the number of '1's and n the number of '-1's in the output sequence $y(n)$. The log likelihood function is given by,

$$\begin{aligned} f(Y|a) &= m\log<Pr\{y=1\}> + n\log<Pr\{y=-1\}> + \\ &\quad (N-m-n)\log<Pr\{y=0\}> \\ &= m\log(1- <F_\alpha(-a\cos 2\pi fn)>) + \\ &\quad n\log(<F_\beta(-a\cos 2\pi fn)>) + (N-m- \\ &\quad n)\log(<F_\alpha(-a\cos 2\pi fn) - F_\beta(-a\cos 2\pi fn)>) \end{aligned}$$

Here $<>$ denotes average over time. The time-averaged probability $<F_\alpha(-a\cos 2\pi fn)>$ for the case where the input snr is very low i.e. $\frac{a}{\sigma} < 0.1$, can be found as:

$$\begin{aligned} &<F_\alpha(-a\cos 2\pi fn)> \\ &\approx \frac{1}{2} + \frac{s}{2\sqrt{s^2+\sigma^2}}(1 - \frac{a^2}{4(s^2+\sigma^2)}) \end{aligned} \tag{22}$$

Also,

$$\begin{aligned} <F_\beta(-a\cos 2\pi fn)> &= <1 - F_\alpha(a\cos 2\pi fn)> \\ &= 1- <F_\alpha(-a\cos 2\pi fn)> \end{aligned}$$

Setting $\frac{\partial f}{\partial a} = 0$, we get

$$1- <F_\alpha(-a\cos 2\pi fn)> = \frac{m+n}{2N} \tag{23}$$

from which the amplitude can be obtained as:

$$a = \sqrt{\frac{8(s^2+\sigma^2)^{3/2}}{s}[\frac{m+n}{2N} + \frac{s}{2\sqrt{s^2+\sigma^2}} - \frac{1}{2}]} \tag{24}$$

Figure 3 shows how the estimation error varies with s. Since the sole purpose of this graph is to show the existence of an optimal 's' where the error gets minimized, we have normalized the error axis. Figure 4 shows the performance of the ML estimator at optimal 's' for varying snr. We see that as the snr increases, the error decreases as expected. Also the performance of the estimator is very good since even at $-50dB$ the error is only 5%. A logical question is how to set the 's' parameter when the amplitude is unknown (which is why we are using the estimator in the first place). For this we refer the reader to the adaptive ML scheme given in [8] which can be easily adopted for this situation.

VI. Output SNR improvement

The SNR of the quantized signal also varies non-monotonically with 's'. The output SNR is given by [3]:

$$SNR_{out} = \frac{<E^2[y(n)]> - <E[y(n)]>^2}{<\sigma_y^2(n)>} \tag{25}$$

where

$$\sigma_y^2(n) = E[y^2(n)] - E^2[y(n)] \tag{26}$$

E denotes the expectation operator and $<>$ is the time-average operator. We now derive a simple expression for SNR_{out} when input SNR is low i.e. $\frac{A}{\sigma} < 0.1$. We have,

Fig. 7. The Lena image quantized into three levels with s=1. Clearly, the information loss is considerable.

Fig. 8. The Lena image quantized into three levels with s=10. The output snr is seen to improve.

$$\begin{aligned} E[y(n)] &= 1-F_\alpha(-a\cos 2\pi fn)-F_\beta(-a\cos 2\pi fn) \\ &= 1-F_\alpha(-a\cos 2\pi fn)-\{1- \\ &\quad F_\alpha(a\cos 2\pi fn)\} \\ &\approx \sqrt{\frac{2}{\pi}}\frac{a\sigma}{s^2+\sigma^2}\cos 2\pi fn \end{aligned} \tag{27}$$

$$\begin{aligned} \sigma_y^2(n) &= 1-F_\alpha(-a\cos 2\pi fn)+F_\beta(-a\cos 2\pi fn)- \\ &\quad E^2[y(n)] \end{aligned}$$

Fig. 9. The Lena image quantized into three levels with s=21. The output snr reaches a maximum.

Fig. 10. The Lena image quantized into three levels with s=50. The output snr deteriorates again.

$$\begin{aligned} &= 1-F_\alpha(-a\cos 2\pi fn)+1-F_\alpha(a\cos 2\pi fn) \\ &\quad -E^2[y(n)] \end{aligned}$$

So,

$$< \sigma_y^2(n) > \;\approx\; 1-\frac{s}{\sqrt{s^2+\sigma^2}} \tag{28}$$

From Equation (25),

$$SNR_{out} = \frac{a^2\sigma^2}{\pi(s^2+\sigma^2)^2\left[1-\frac{s}{\sqrt{s^2+\sigma^2}}\right]} \tag{29}$$

Figure 5 shows the non-monotonic output SNR variation with the Rayleigh parameter. An expression for SNR_{gain} can be found as:

$$\begin{aligned} SNR_{gain} &= \frac{SNR_{out}}{SNR_{in}} \\ &= \frac{2}{\pi\left[1+(\frac{s}{\sigma}^2\right]^2\left[1-\frac{s}{\sqrt{s^2+\sigma^2}}\right]} \end{aligned} \quad (30)$$

Figure 6 shows the variation of SNR gain with the Rayleigh parameter 's'. It shows that at an optimal 's' the loss of information due to quantization is minimum. There is no imrovement in SNR for Gaussian noise. It can be proved that snr gain canot greater than unity. We proceed by the method of contradiction.

$$\begin{aligned} SNR_{gain} &> 1 \\ \Rightarrow \frac{2}{\pi\left[1+(\frac{s}{\sigma})^2\right]^2\left[1-\frac{s}{\sqrt{s^2+\sigma^2}}\right]} &> 1 \\ \Rightarrow \pi(1+(\frac{s}{\sigma})^2)^2[\pi(1+(\frac{s}{\sigma})^2)-4]+4 &< 0 \end{aligned} \quad (31)$$

which is clearly impossible as the minimum value of the function on right is positive. Hence, SNR gain cannot be greater than 1. However, in applications like data compression where quatization is neccessary, Stochastic Thresholding provides a means to minimize loss of information. To visualize this cocept, take a look at the sequence of quantized images in Figures 7, 8, 9 and 10. Only three levels of quantization have been used. We can clearly *see* that there is an optimal 's' at which maximum transmission of information takes place.

VII. Conclusion

In this paper we have shown how Stochastic Thresholding achieves a gain in Fisher Information with both bi-level and tri-level quantization. The ML estimator based on the data quantized by random thresholding is seen to have excellent performance down to $-50dB$ provided after optimization. We have shown the existence of an optimal 's' associated with the Rayleigh distributed threshold where the error hits a minimum. The output SNR also has been shown to follow a non-monotonic relationship with 's'. This gives us a method to minimize the loss of information due to quantization. This concept was illustrated by quantizing a digital image with intensities varying in the range 0 to 255 to just three levels. However, SNR gain could not be achieved greater than unity when input is corrupted by gaussian noise. It remains to be investigated whether SNR gain greater than unity is possible for non-gaussian noise.

Acknowledgement

I am thankful to Dr. G.V. Anand, Department of Electrical Communications, Indian Institute of Science, Bangalore for giving valuable advice. I am also grateful to Dr. R.M. Vasu, Department of Instrumentation, Indian Institute of Science, Bangalore who gave me the first opportunity to work in Stochastic Resonance.

References

[1] S. Dasgupta, J. Nair and S. Sarkar "Estimation of amplitude of a sinusoidal signal buried in noise using Stochastic Resonance with the help of a 3-level quantizer," *Proceedings of the IEEE INDICON 2004 Conference, Kharagpur, India*, pp. 41-46, December 2004.

[2] S. Dasgupta, J. Nair and S. Sarkar, Ch. I-8: "Imaging using Stochastic Resonance," *SPCCN0.1*, Edited by: K.H. Abhyankar, V.A. Deshmukh, V.J. Kulkarni and S.S. Agashe, Publisher: Tata McGraw-Hill, 2005.

[3] V.G. Guha "Detection of Weak Signals in Non Gaussian Noise using Stochastic Resonance," *M.S. Thesis, Indian Institute of Science, Bangalore,* October 2002.

[4] L. Gammaitoni, P. Hanggi, P.Jung and F. Marchesoni "Stochastic Resonance," *Reviews of Modern Physics*, Vol.70, pp. 223-287, 1998.

[5] B. McNamara, K. Wiesenfield "Theory of Stochastic Resonance," *Physics Review A39*, pages 4854-4869, May 1984.

[6] S.M. Kay "Fundamentals of Statistical Signal Processing," *Englewood Cliffs, NJ, Prentice-Hall Inc.*, 1998.

[7] D. Panchenko, "Statistics for applications," *MIT OpenCourseWare,* http://ocw.mit.edu/

[8] D. Rousseau, G.V. Anand and F. Chapeau-Blondeau, "Nonlinear Estimation From Quantized Signals: Quantizer Optimization and Stochastic Resonance," *3rd International Symposium on Physics in Signal & Image Processing,*, pp. 89-92, Grenoble, France, 29-3 January 2003.

[9] R.M. Vasu, G.V. Anand, S. Dasgupta, J. Nair and S. Sarkar, "Weak signal estimation using Stochastic Resonance for imaging via Polarisation Discrimination," *to be submitted,* www.geocities.com/sdg_kgp/SRpaper.pdf