# CRITICAL ENCODING RATE IN COMBINED DENOISING AND COMPRESSION

*Sehoon Yea, William A. Pearlman*

Department of Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute
E-mail: {yeas,pearlw}@rpi.edu

**ABSTRACT**

In this paper, we elaborate on denoising schemes based on lossy compression. First, we provide an alternative interpretation of the so-called Occam filter and relate it with the complexity-regularized denoising schemes in the literature. Next, we discuss about the 'critical distortion' of a noisy source and argue that optimal denoising is achieved at the corresponding critical encoding rate rather than at the encoding rates suggested by other compression-based denoisers. Finally, we discuss the so-called 'indirect rate distortion problem.' We focus particularly on the high bit-rate encoding of noisy sources and show lossless compression of a denoised source is often very wasteful of bits, and suggest a simple way of determining an appropriate bit-rate for compressing a denoised source economically while retaining its initial denoised quality.

## 1. INTRODUCTION

An image is usually subject to noise in its acquisition or transmission. The purpose of denoising is to remove the noise while maintaining the important signal features.

It has long been well known that random noise is hard to compress, while information-bearing signal is not. Since a typical lossy compression scheme naturally prioritizes low frequency components (i.e., information-bearing part) of the source by taking advantage of energy compaction, it tends to filter out the high-frequency components usually dominated by noise. Natarajan's denoising technique called *Occam filter*[3] is one of the early attempts at denoising based on lossy compression capitalizing on such an observation. Also, Liu and P. Moulin[2] showed the *complexity-regularized denoising* based on lossy compression has an interesting interpretation on the operational rate-distortion curve of the compressor.

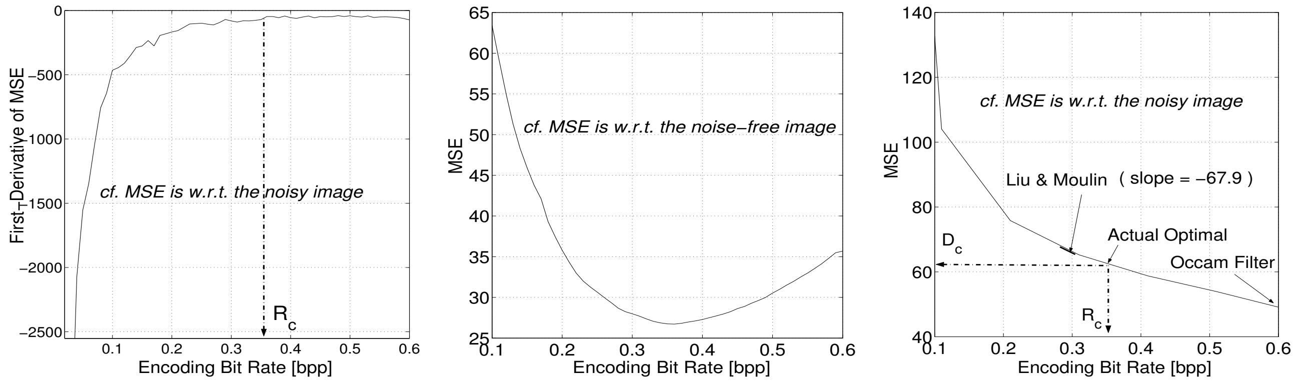
However, it is not clear how these two compression-based denoising schemes could be related each other, and furthermore, the denoised results based on these schemes seem to deviate from the optimality as measured by the mean-square-error with respect to the original noise-free source. In this paper, we investigate the combined denoising and compression problem based on a typical wavelet-based encoder such as SPIHT. We first investigate a connection between the Occam filter and the complexity-regularized denoising and show that the optimal compression rate for denoising is different from the ones suggested by the two aforementioned compression-based denoising schemes. We argue optimal denoising is achieved at the 'critical rate' of the noisy source.

While a considerable amount of research work has been done on the compression of noisy sources, its practical ramifications in the context of combined denoising and compression have by no means been fully explored. We turn our attention to the compression aspects and consider the high bit-rate compression of noisy sources. As evidenced both by theory and practice, compression preceded by a denoising step constructs an optimal way of compressing noisy sources especially at mid-to-high bit rates[4, 5]. However, since any practical denoiser has a limited denoising capability, it is shown that losslessly compressing the denoised source is often very wasteful of bits. In this regard, we show the notion of the 'critical rate' could be useful for determining an appropriate bit rate for compressing a denoised source economically while retaining its initial denoised quality.

The organization of this paper is as follows. Section 2 discusses the Occam filter and the complexity-regularized denoising. In Section 3, we argue the 'critical rate' of a noisy source is where the actual optimal denoising is achieved. Section 4 focuses on the high bit-rate compression of noisy sources. Section 5 concludes the paper.

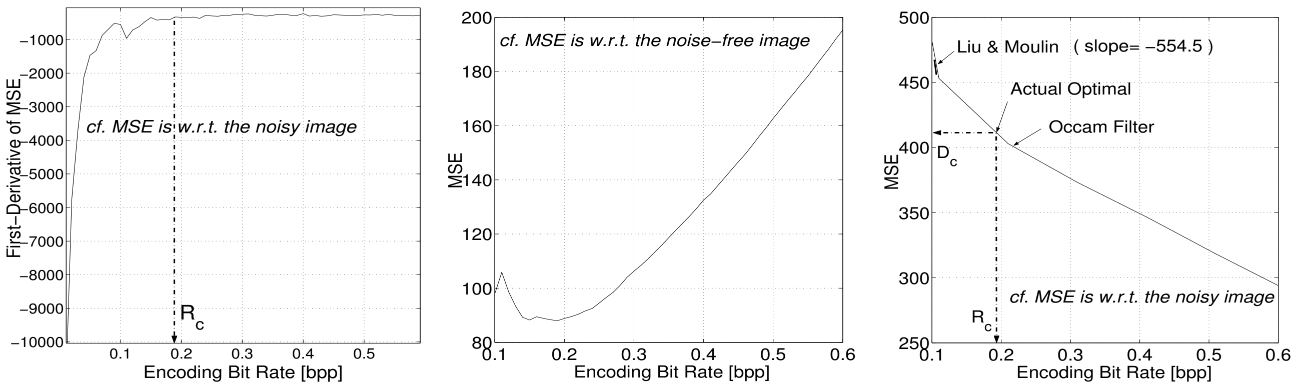
## 2. OCCAM FILTER AND COMPLEXITY-REGULARIZED DENOISING

The essence of the Occam filtering is that when a lossy compression algorithm is applied to a noisy signal with the allowed loss set equal to the noise variance $\sigma^2$, the loss and the noise tend to cancel each other asymptotically as the sample size goes to infinity[3]. The intuition is that where the distortion is smaller than $\sigma^2$, the encoder essentially attempts to compress the noise and it becomes notice-




(a) 1st-Order Derivative of MSE vs. Bit Rate (b) MSE with respect to the Original (c) Optimal Denoising Points Comparison

**Fig. 1**. Optimal Denoising Points for Lena with $\sigma$ = 7, image coder: SPIHT




(a) 1st-Order Derivative of MSE vs. Bit Rate (b) MSE with respect to the Original (c) Optimal Denoising Points Comparison

**Fig. 2**. Optimal Denoising Points for Lena with $\sigma$ = 20, image coder: SPIHT

ably harder to compress. On the other hand, J. Liu and P. Moulin[2] showed that an optimal complexity-regularized Maximum-Likelihood estimator (denoiser) can be constructed using a lossy compressor which operates at the point with the slope of $-1/(2\sigma^2 ln2)$ on the operational rate-distortion curve when the noise variance is $\sigma^2$. Here, the optimal trade-off between the estimation error and the complexity-penalty measured by compression bit rate is achieved as it becomes much harder (i.e., higher penalty) to compress the noisy source beyond the suggested point. It turns out that the suggested slope corresponds to that of the rate-distortion function when a backward test-channel holds with a Gaussian residual power equal to the noise variance $\sigma^2$. To see this, we need a few results from the rate-distortion theory.

**Lemma 2.1** *[1] The Shannon Lower Bound for the squared-error distortion measure is given as:*

$$R_L(D) = h(p) - \frac{1}{2}log(2\pi eD) = \frac{1}{2}log(\frac{Q}{D}) \quad (1)$$

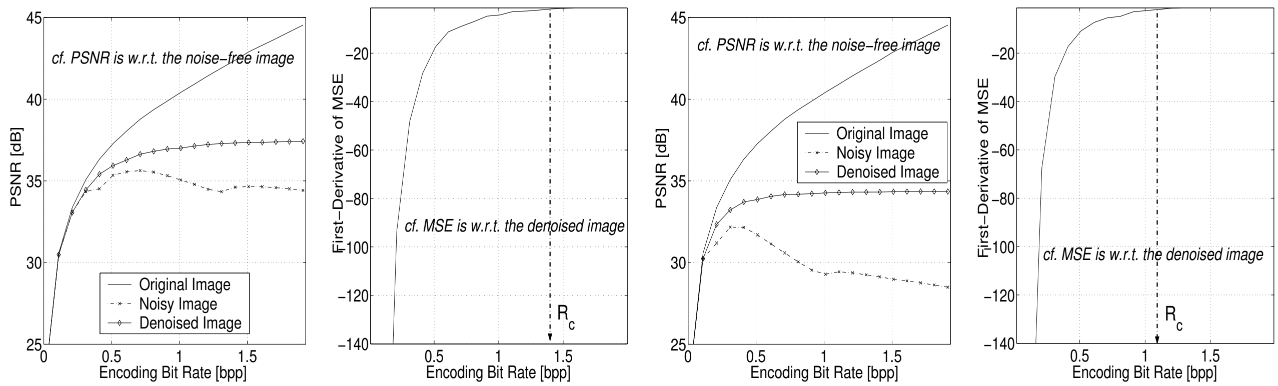
*,where the constant Q, called the entropy power of the source* $p(\cdot)$*, is the variance of a Gaussian density that has the same differential entropy as* $p(\cdot)$*, that is,* $h(p) = \frac{1}{2}log(2\pi eQ)$

**Theorem 2.2** *[1]* $R_L(D) = R(D)$ *iff the source r.v. X can be expressed as the sum of two statistically independent random variables one of which is distributed according to the probability density* $g_Z(z) = \frac{1}{\sqrt{2\pi D}}exp(\frac{-z^2}{2D})$ *in case the squared-error distortion measure is adopted.*

**Theorem 2.3** *Suppose random variables X,Y and Z are given such* $X = Y + Z$ *where* $Z \sim N(0, \sigma^2)$ *and is independent of Y. If* $d(x, y) = (x - y)^2$ *then the slope of the rate-distortion function at* $D = \sigma^2$ *is given as* $-1/(2\sigma^2 ln2)$

**Proof:** By Theorem 2.2, we have $R(D) = R_L(D)$ at $D = \sigma^2$. The derivative of $R_L(D)$ is $-1/(2Dln2)$ by Lemma 2.1. At $D = \sigma^2$, it reduces to $-1/(2\sigma^2 ln2)$. ∎

From Theorem 2.3, we can see the complexity-regularized denoising and the Occam filter are really equivalent assum-




(a) Encoding the original, noisy ( $\sigma^2$=25), and denoised Lena

(b) Derivative of MSE when encoding the denoised Lena in (a) )

(c) Encoding the original, noisy ($\sigma^2$=100), and denoised Lena

(d) Derivative of MSE when encoding the denoised Lena in (c) )

**Fig. 3**. Locating the saturation rate for high bit rate compression of denoised Lena, image coder: SPIHT

ing an *ideal* encoder, the former specifying the slope while the latter the distortion.

## 3. CRITICAL ENCODING RATE AND OPTIMAL DENOISING VIA COMPRESSION

Although the optimality of Occam filtering was proved for the *deterministic* source[3], it was observed that stopping the encoder at the residual power level equal to the noise variance does not, in general, lead to optimal denoising. This is because a typical *stochastic* source has a 'critical encoding rate'[1] beyond which its encoding residual is white. This 'incompressible' part of the source—essentially also AWGN (*Additive White Gaussian Noise*)—would interact with the noise and would be indiscernible from the encoder's viewpoint. Therefore, encoding the noisy source up to the point with distortion equal to the noise variance would mean an *over-encoding* in that not only the hard-to-compress part of the original source but also the noise would be encoded, thus leading to a suboptimal denoising performance. (See Figures 1(b),1(c) and 2(b), 2(c) .)

For the same reason, the suggested slope of the operational R-D curve by the complexity-regularization is usually steeper (meaning an underestimation of the amount of noise) than the slope at which the actual (optimal) denoising is achieved. Therefore it tends to be *under-encoded*, thus again leading to a suboptimal denoising. (Again see Figures 1(b),1(c) and 2(b), 2(c).)

Note that both the *under-encoding* and the *over-encoding* are due to the limited 'resolvability' of the source and the noise. Therefore, there seems to be an optimal (in the MSE sense) denoising bit rate determined both by the noise and the i.i.d. encoding residual of the original source where the encoder starts to see them as a combined i.i.d. noise. As an illustration, we consider the case of a first-order Gauss-Markov model with an AWGN. The noise-free source is given as

$$x_i = -ax_{i-1} + \sigma_n n_i \tag{2}$$

, where $0 < a < 1$ and $n_i$ is a unit-variance, zero-mean white Gaussian noise. The average power of the source $x$ is given as $\sigma_x^2 = \sigma_n^2/(1-a^2)$.

**Theorem 3.1** *Suppose a noisy source is given by $z_i = x_i + v_i$, where $x_i$ is a first-order Markov model defined by (2) and $v_i$ is an AWGN with zero-mean and variance $\sigma_v^2$. Then the rate distortion function of the noisy source coincides with the Shannon Lower Bound with the entropy power equal to $0.5\sigma_x^2(\alpha+\beta\gamma+\sqrt{\alpha^2+2\alpha\beta\gamma+\alpha^2\gamma^2})$ where $\alpha \equiv 1-a^2$, $\beta \equiv 1+a^2$ and $\gamma \equiv \sigma_x^2/\sigma_v^2$, if the distortion is no larger than the critical distortion $D_c = \frac{1-a}{1+a}\sigma_x^2 + \sigma_v^2$.*

**Proof:** A parametric representation of the distortion and the rate of this noisy source is given as follows:

$$D(\lambda) = \frac{1}{2\pi}\int_0^{2\pi} \min(\lambda, \Phi_{ZZ}(\omega))d\omega \tag{3}$$

$$R(\lambda) = \frac{1}{4\pi}\int_0^{2\pi} \max(0, \log\frac{\Phi_{ZZ}(\omega)}{\lambda})d\omega \tag{4}$$

, where $\Phi_{ZZ}(\omega) = \frac{\sigma_n^2}{1+a^2+2a\cos\omega} + \sigma_v^2$. Since $\Phi_{ZZ}(\omega)$ is symmetric around $\omega = \pi$ with the minimum at 0 and $2\pi$, we have $D(\lambda) = \lambda$ when $\lambda \le \Phi_{ZZ}(0) = \frac{\sigma_n^2}{(1+a)^2} + \sigma_v^2 = \frac{1-a}{1+a}\sigma_x^2 + \sigma_v^2$ ( $\equiv D_c$: *critical distortion* ). Under this condition, we have

$$\begin{aligned} R(\lambda) &= R(D) \\ &= \frac{1}{4\pi}\int_0^{2\pi}\max(0,\log\frac{\Phi_{ZZ}(\omega)}{D})d\omega \\ &= \frac{1}{2}\log\frac{\sigma_x^2(\alpha+\beta\gamma+\sqrt{\alpha^2+2\alpha\beta\gamma+\alpha^2\gamma^2})}{2D} \end{aligned} \tag{5}$$

, which is the Shannon Lower Bound of a source with the entropy power $0.5\sigma_x^2(\alpha+\beta\gamma+\sqrt{\alpha^2+2\alpha\beta\gamma+\alpha^2\gamma^2})$. The corresponding *critical rate* $R_c \equiv R(D_c)$ is given as

$$R_c = \frac{1}{2}\log\frac{0.5(\alpha+\beta\gamma+\sqrt{\alpha^2+2\alpha\beta\gamma+\alpha^2\gamma^2})}{(1-a)/(1+a)+\gamma} \quad (6)$$

■

While the author in [3] claims optimal denoising is achieved at the distortion level equal to the noise level, we argue the optimal distortion level and its corresponding optimal encoding rate are determined by both the noise and the source as illustrated above. Although the above theorem is about the the rate-distortion optimal encoder with a Gaussian source, we can use in practice a heuristic similar to the one proposed in [3] to locate the critical point by capitalizing on the difference in compressibility between the noise and the source. The critical rate ($R_c$) of the noisy source Lena indicated in each of the Figures 1(a) and 2(a) corresponds to the point where the derivative of MSE (vertical axis) begins to take a relatively constant value, or equivalently, its 2nd order derivative comes to zero. We can confirm the indicated $R_c$ values match very well the actual optimal denoising bit rate in Figures 1(b) and 2(b). Also note that $D_c$ for the noisy Lena was always 10 to 15 larger than the noise variance (see Figures 1(c) and 2(c)), which is thought to be the level of the white encoding residual inherent in the noise-free Lena.

## 4. HIGH-BIT RATE COMPRESSION OF NOISY SOURCES

Most of the literature in denoising focus on denoising aspects of the problem, treating the compression step as a separate, following stage in the processing chain. This is justified by, for example, the 'indirect RD problem' argument[5].

Alshayh et al.[4] performed a series of experiments based on the aforementioned setting of 'indirect RD problem.' They confirmed that, at mid-to-high bit rates, a denoising-then-compression approach results in large improvements in terms of PSNR over a direct compression of the noisy source. (See Figures 3(a), 3(c)) But they also noted that at mid-to-high bit rates the quality (as measure by the PSNR w.r.t. the noise-free image) of a denoised image did not improve beyond a certain point (call it the *saturation rate* hereafter). This tendency is caused by the fact that every denoiser has a limited denoising performance. In other words, since denoising inevitably leads to a somewhat limited reconstruction of the original noise-free image, encoding the fine-details of i.i.d. nature in the denoised image beyond the saturation rate to faithfully reproduce them would not contribute noticeably to improving PSNR w.r.t. the original noise-free image. However, since we do not have the original noise-free image available in real situations, it is not obvious how we can find the saturation rate for a given denoised image. It turned out that we could use the technique of detecting the 'critical rate' here to find the saturation rate. This is because the saturation rate is essentially nothing but the critical rate of the denoised image, where the fine details of i.i.d. nature start to 'appear' as we encode. Figure 3(a) and 3(c) compare the PSNR plots (w.r.t. the noise-free image) obtained by encoding the noisy and denoised Lena image when the noise variances are 25 and 100, respectively. We could detect the corresponding saturation rate ($R_c$) as in Figures 3(b) and 3(d). We can also confirm the benefits of using the saturation rate by looking at Table 1, where we achieved almost the same PSNR's at less than half of the lossless bit rates.

| Method | Rate[bpp] | PSNR[dB] | Rate | PSNR |
|---|---|---|---|---|
| Lossy | **1.41** | **37.32** | **1.11** | **34.29** |
| Lossy | 1.91 | 37.43 | 1.91 | 34.35 |
| Lossless | **3.31** | **37.60** | **2.97** | **34.40** |

**Table 1**. Compressing the denoised Lena, column 2 & 3 :Lena with noise $\sigma^2$=25 , columns 4 & 5:Lena with noise $\sigma^2$=100

## 5. CONCLUSION

In this paper, we introduced the notion of 'critical distortion' of a noisy source and argued that optimal denoising is achieved at the corresponding critical rate rather than at the encoding rates suggested by other compression-based denoisers in the literature. We also showed the usefulness of the notion of 'critical rate' in high bit-rate compression of noisy sources.

## 6. REFERENCES


[1] T. Berger, *Rate-Distortion Theory*, Prentice-Hall, 1971.

[2] Liu and J. and Moulin, P.,"Complexity-regularized image denoising,"*IEEE Transactions on Image Processing*, Vol.10, Iss.6, pp.841-851,Jun 2001.

[3] B. Natarajan.,"Filtering random noise from deterministic signals via data compression," *IEEE Transactions on Signal Processing*,Vol.43, No.11, pp.2595-2605,Novemeber, 1995.

[4] O.K. Al-Shaykh, R.M. Mersereau,"Lossy compression of noisy images,"*IEEE Transactions on Image Processing*,Vol. 7 , No.12, pp 1641-1652, December, 1998.

[5] J. Wolf and J. Ziv,"Transmission of noisy information to a noisy receiver with minimum distortion,"*IEEE Transactions on Information Theory*,Vol. 16, No 4, pp. 406-411, July, 1970.