Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Image and Signal Processing (CIISP 2007)

MINIMUM NOISELESS DESCRIPTION LENGTH (MNDL) THRESHOLDING

Azadeh Fakhrzadeh, Soosan Beheshti

Electrical and Computer Engineering Department, Ryerson University, Toronto, Canada
afakhrza@ryerson.ca soosan@ee.ryerson.ca

1. ABSTRACT

In this paper, a new thresholding approach for data denoising is presented. The approach is based minimum noiseless description length (MNDL), a new method for optimum subspace selection in data representation. By using the observed noisy data, this information theoretic approach provides the optimum threshold that minimizes the description length of the noiseless signal. Comparison of the new method with the existing thresholding methods is provided.

2. INTRODUCTION

The goal of a data denoising approach is to discriminate between noise and data and to remove the unwanted noise from data. All denoising methods attempt to restore the noiseless data. Over the years, various data denoising methods have been proposed. Traditionally, these methods, such as Wiener filters, were linear. However, recently researchers have focused mainly on nonlinear approaches, since they always converge better than the linear ones.

One of the well known approaches to data denoising is Wavelet thresholding which was first introduced by Donoho and Johnstone [1]. When orthogonal wavelet basis is used, the coefficients with small absolute values tend to be attributed to the additive noise. Taking advantage of this property, finding a proper threshold and setting all absolute value of coefficients smaller than the threshold to zero can suppress the noise. Thus, the most crucial issue in these approaches is defining a trustable threshold which is usually obtained by solving a minmax problem. Different methods have been proposed to determine the desired threshold. Well-known thresholding methods are VisuShrink [1], Minimum Description Length (MDL) denoising [2] and SureShrink [3]. VisuShrink proposed by Donoho and Johnstone [1], uses universal threshold, $\sigma\sqrt{2\log(N)}$, for denoising where σ^2 is the variance of the additive noise and N is the data length. MDL denoising is a method recommended by Rissanen [2]. In this method the normalized maximum likelihood (NML) of the noisy data is defined as the description length of data. This method suggests choosing the subspace that minimizes the description length of the noisy data and the proposed threshold is $\sigma\sqrt{\log(N)}$. Sureshrink is another approach introduced by Donoho and Johnstone that provides an optimum soft threshold for data denoising.

Minimum Noiseless Description Length (MNDL) is a new approach to signal denoising which has recently been proposed by Beheshti and Dahleh [4]. This method suggests choosing the subspace that minimizes the description length of the "noiseless" data. For this purpose, MNDL provides bounds on the reconstruction error or MSE (minimum square error). In this approach, the subset of order m represents the bases with only m nonzero coefficients of the estimated denoised signal. The bounds on the reconstruction error are estimated in each subset and the subset that minimizes the upper bound of this error is chosen. With a proper choice of the competing subsets, the method not only chooses the optimum subset, but also provides the optimum threshold simultaneously. The resulted optimum threshold is a function of the noise variance σ, the data length N, and the noisy data itself.

Details of the MNDL as a subset selection approach is provided in [4] and the use of this approach as a thresholding method is discussed briefly. In this paper, we explore application of MNDL as a thresholding method in detail. As in any thresholding approach, first the sorted version of the basis coefficients of the noisy signal is calculated. In this paper, we show that in MNDL thresholding, the effects of the additive noise are no more the chi-square random variables that are used in random selection of competing subsets in the existing MNDL approach in [4]. By consideration of the structure of the involved distributions in this case, we improve the performance of MNDL for threshoding purposes.

The paper is arranged as follows. Section 3 described the considered thresholding problem. Section 4 briefly described the fundamentals of MNDL subset selection. Section 5 introduces the MNDL thresholding approach. Section 6 provides the simulation results and Section 7 is the conclusion.

3. PROBLEM STATEMENT

The noiseless data $\{\bar{y}(n), \quad n = 1, ..., N\}$ of length N has been corrupted by an additive noise:

$$y(n) = \bar{y}(n) + w(n) \qquad (1)$$

where $w(n)$ is an independent and identically distributed (iid) Gaussian random process with zero mean and variance σ^2. In the denoising process, we deﬁne the noisy data by using an orthogonal basis. The goal is to provide the optimum threshold for the resulted coefﬁcients. Thresholding the coefﬁcients provides the best estimate of the noiseless data.

Assume that the noiseless data belongs to space S_N, $\bar{y}(n) \in S_N$. The space S_N can be expanded by some orthogonal basis vectors

$$< s_i, s_j >= \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \quad (2)$$

where $< s_i, s_j >$ is the inner product of vectors, s_i and s_j. The noiseless, using this basis, is as follows:

$$\bar{y}^N = \sum_{i=1}^{N} \theta^*(i) s_i \quad (3)$$

where $\theta^*(i)$ is i-th coefﬁcient of the noiseless data. The noisy data from (1) in space S_N is:

$$y^N = \sum_{i=1}^{N} \theta(i) s_i. \quad (4)$$

In denoising approaches based on thresholding, some of the coefﬁcients of the noisy signal are ignored. In general two thresholding methods exist; Hard and Soft thresholding. Hard thresholding kills or keeps the coefﬁcient by comparing them with the threshold

$$\hat{\theta}(i) = \begin{cases} \theta(i) & \text{if } |\theta(i)| \geq T_h, \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

Where T_h is hard threshold. Soft thresholding kills the coefﬁcients below T_s and reduces the absolute of the rest of coefﬁcients

$$\hat{\theta}(i)_{S_m} = \begin{cases} \text{sgn}(\theta(i))(\mid \theta(i) \mid - T_s) & \text{if } |\theta(i)| \geq T_s, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

In this paper, we use the theory of MNDL to provide a new thresholding denoising approach.

4. MNDL SUBSPACE SELECTION

In MNDL method, some of $\theta(i)$s are chosen as the noiseless data coefﬁcients. Therefore, the noiseless data $\hat{y}^N_{S_m}$, can be deﬁned as

$$\hat{y}^N_{S_m} = \sum_{i=1}^{N} \hat{\theta}_{s_m}(i) s_i \quad (7)$$

Where S_m, is a subspace of S_N that is spanned by m elements of basis and $\hat{\theta}_{S_m}$ deﬁned as follows

$$\hat{\theta}(i) = \begin{cases} \theta(i) & \text{if } s_i \in S_m, \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

MNDL compares subspaces of different order and chooses the optimum subspace by minimizing the reconstruction error among the competing subspaces.

The quality of denoised data is generally evaluated by the minimum square error (MSE) between the denoised data and the original one. This quantity is not accessible due to its dependance on original data. Different denoising methods have tried to estimate this quantity. With minimizing this estimation, they suggest an approximation for noiseless data. The more closer the estimation of MSE to its true value, much better denoised data will be obtained.

MNDL provides a novel method to calculate the bounds on MSE error. MNDL is based on a new information-theoretic criterion, the description length of "noiseless" data. It suggests choosing a subspace in which the length of noiseless data is minimum. The calculation of length of noiseless data leads to the calculation of MSE. Hence, its minimization corresponds to the minimization of MSE. Here, probabilistic bounds on reconstruction error are computed as a function of noisy data and noise variance. The subspace that minimizes the upper bound of this error is chosen. After selecting the optimum subspace or estimating all noiseless coefﬁcients, the threshold will be computed.

The desired unavailable MSE error z_{S_m} in subspace S_m is

$$z_{S_m} = 1/N \parallel \bar{y}^N - \hat{y}^N_{S_m} \parallel_2^2, \quad (9)$$

The novelty of MNDL is that the method uses the available data error x_{S_m} of the following form

$$x_{S_m} = 1/N \parallel y^N - \hat{y}^N_{S_m} \parallel_2^2, \quad (10)$$

to provide probabilistic bounds on the desired MSE, z_{S_m}. Note that based on Parseval's Theorem these error cab be written in the following form

$$x_{S_m} = 1/N \parallel \theta - \hat{\theta}_{S_m} \parallel_2^2, \quad (11)$$

$$z_{S_m} = 1/N \parallel \theta^* - \hat{\theta}_{S_m} \parallel_2^2 . \quad (12)$$

The process of providing bounds on z_{S_m} based on observation of x_{S_m} is explained in [4] in detail. These two values are samples of random variables Z_{S_m} and X_{S_m}. MNDL studies the structure of these two random variables and uses the connection between these two random variables. By using one sample of X_{S_m}, MDL validates the expected value and variance of this random variables. The validated values are then used in estimation of second order statistics of Z_{S_m} and enables us to provide probabilistic bounds on the desired z_{S_m}. The optimum subspace is chosen based on the estimated upperbounds for subspaces with different order.

5. MNDL THRESHOLDING

The main difference between MMNDL thresholding and MNDL optimum subspace selection is in forming the competing subsets. While in MNDL, we assume that the subspaces are chosen a priori and are not functions of the observed data, in MNDL thresholding, the subspaces are chosen based on the observed data. In the thresholding MNDL approach, the competing subsets have to be chosen as nested subsets based on this sorted version of coefficients. For example the first subset represents the the basis associated with the largest absolute value of the coefficients. The subset with two coefficients includes this basis and the basis with the second largest value from the sorted coefficients. The thresholding question is then answered by providing the optimum subset. In this case the threshold is the smallest absolute value of sorted coefficients in this subset.

In MNDL, the structure of second order statistics of the two random variables are the main ingredients of the analysis. Since in subspace selection, the additive noise effects is independent from the data, it is shown that these two random variables are Chi-square random variables and the additive noise parts in both expected values and variances of these random variables are in linear form as a function of subspace order. For example, the general form of the expected values are

$$E(X_{S_m}) = E(noise_{xsm}) + \frac{1}{N} \parallel \Delta_{S_m} \parallel_2^2 \qquad (13)$$

$$E(Z_{s_m}) = E(noise_{zsm}) + \frac{1}{N} \parallel \Delta_{S_m} \parallel_2^2 \qquad (14)$$

where $\frac{1}{N} \parallel \Delta_{S_m} \parallel_2^2$ is the the effect of the discarded noiseless part of the data in subspace S_m. In other words, Δ_{S_m} is a vector of length $N - m$, corresponding to the coefficients of the basis that are not in S_m). In the MNDL subspace selection the noise effects derived from the chi-square structure are

$$E(noise_{xsm}) = (1 - \frac{m}{N})\sigma^2. \qquad (15$$

$$E(noise_{zsm}) = \frac{m}{N}\sigma^2. \qquad (16$$

However, in MNDL thresholding, as the subspaces are chose based on the noisy data, the effects of the additive noise is nc simply derived from a chi-square random variable. We esti mate the noise part effects in this case by finding the estimat of sorted version of the additive noise effects in the coeffi cients. We generate the noisy part M times and calculate it mean. In each trial, an additive white Gaussian noise wit variance σ_w^2 is generated and the associated coefficients ar sorted. Denote the sorted noise coefficients of length N wit $v_i[n]$ where i represents the ith trial. Therefore, the noisy pa of z_{S_m} is estimated as follows:

$$E(noise_{zsm}) = \frac{1}{M}\sum_{i=1}^{M}\sum_{n=1}^{m} v_i^2[n] \qquad (17)$$

and noisy part of x_{S_m} is sum of v_is from m to M divided by M:

$$E(noise_{xsm}) = \frac{1}{M}\sum_{i=1}^{M}\sum_{n=m}^{N} v_i^2[n] \qquad (18)$$

Next, we approximate x_{S_m} with its expected value as follows. From equation (13) we have

$$\frac{1}{N} \parallel \hat{\Delta}_{S_m} \parallel_2^2 \simeq x_{S_m} - E(noise_{zsm}). \qquad (19)$$

Using this value we can estimate Z_{S_m} as follows:

$$\hat{z}_{sm} \simeq \frac{1}{N} \parallel \hat{\Delta}_{S_m} \parallel_2^2 + E(noise_{xsm}) \qquad (20)$$

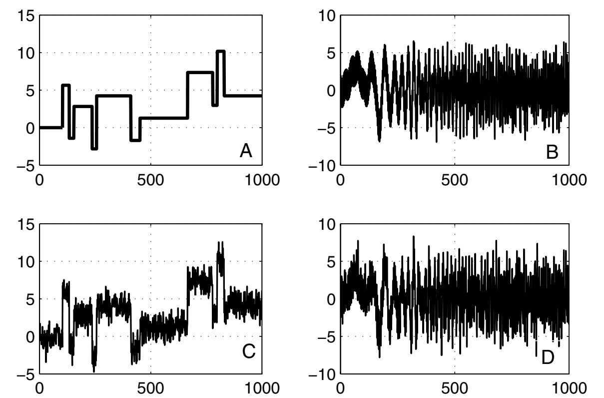

Fig. 1. Noiseless signal $\bar{y}[n]$ and Noisy signal $y[n]$. (A)noiseless block. (B)noiseless mishmash. (C)noisy block. (D)noisy mishmash

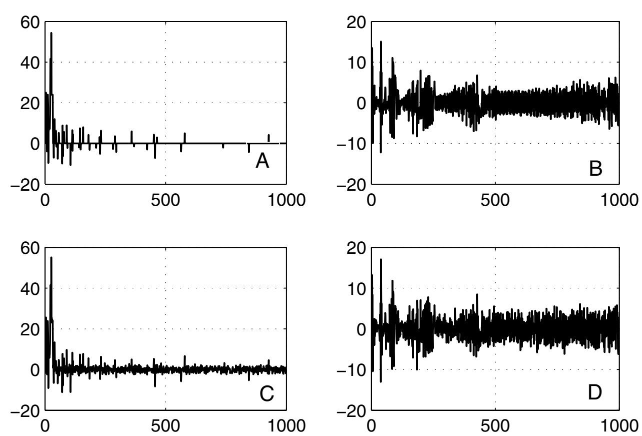

Fig. 2. Noiseless signal coefficients θ^* and noisy coefficients. (A) noiseless block . (B) noiseless mishmash. (C) noisy block. (D) noisy mishmash

6. SIMULATION RESULTS

Standard block and mishmash signal were used as tested signals with length 1024. The block signal was introduced in [1], and mishmash was provided in Matlab R12. We chose these two signals as two extreme cases; the block signal is a signal with very small number of nonzero coefficients and the mishmash is the one with 1024 number of nonzero coefficients. We added noise with different levels $\sigma^2 = 1, 3, 5$ to the original signals. Figure 1 shows the noiseless and noisy signals. The wavelet transform employs Haar wavelet with five scales of orthogonal decomposition. Figure 2 shows the wavelet coefficients of both noisy and noiseless signals. MNDL uses the available data error (x_{S_m}), the error between the noisy signal and thresholded signal with the mth sorted coefficient, to estimate unavailable MSE error (z_{S_m}). In this method, the noisy part of these two errors are estimated by using Chi-square random variable. In MNDL thresholding method, the subspaces are chosen based on the noisy data. Therefore, the noisy part of errors is not chi-square random variable any more.

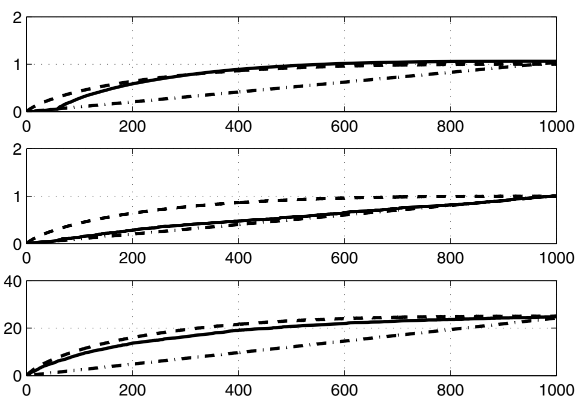

Fig. 3. Noisy part of true reconstruction error for (A) noisy part of block signal when σ^2=1, (B) noisy part of mishmash when σ^2=1, and (C) noisy part of mishmash when σ^2=5. Dashed line is MNDL sorted estimate, solid line is the unavailable desired noisy part and dash-dotted line is MNDL unsorted estimate $\frac{m}{N}\sigma^2$.

Figure 3 shows the true noisy part of reconstruction error, z_{S_m}, the noise estimate $\frac{N}{m}\sigma^2$ used by MNDL and the new estimate proposed in this paper for both signals. The noise variance in this case is one. As the figure shows, for block signal with almost 70 nonzero coefficients, for smaller values of m, the noisy part coincides with $\frac{N}{m}\sigma^2$. However, for higher values of m, the new approximation from (17), performs better than $\frac{N}{m}\sigma^2$. For noise variance larger than one, the similarity between the true noise effect and the new approximation is valid for larger values of m. As we increase m, the new estimation outperforms the existing MNDL approach. For mishmash signal, with almost 1024 nonzero coefficients and unit noise variance, the true noisy part of reconstruction error is very close to $\frac{N}{m}\sigma^2$. For this signal with smaller noise variances, since the noiseless signal coefficients are larger than that of noise, we sort the signal coefficients rather than the noise. Therefore, to estimate the noise effects, we do not use the sorted noise and the noise estimate is the same as that of the existing MNDL.

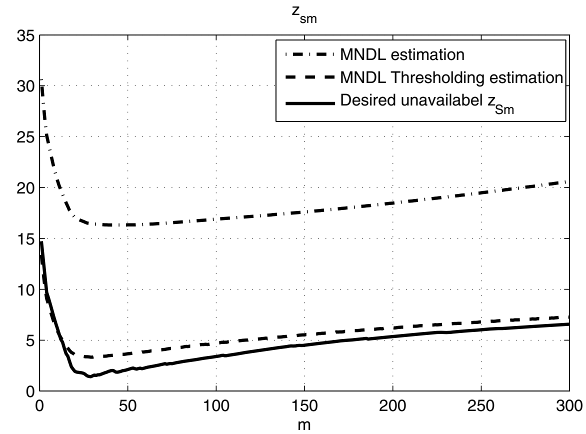

Fig. 4. Desired unavailable reconstruction error z_{S_m} and its estimates using MNDL thresholding and existing MNDL methods for block signal with noise variance $\sigma^2 = 3$.

	(1)	(2)	(3)
$\sigma^2 = 1$	72	77	91
$\sigma^2 = 3$	34	29	40
$\sigma^2 = 5$	22	18	37

Table 1. Optimum order m_{opt} for block signal with different noise variances using (1) true z_{S_m}; (2) estimate of z_{S_m} with the proposed MNDL thresholding; and (3) estimate of z_{S_m} with existing MNDL subspace selection method.

The true z_{S_m}, its estimates using existing MNDL, and MNDL thresholding methods have been plotted in Figure 4. In this simulation, the noise variance is 3. To have a better comparison between these two methods, the optimum subspace order m_{opt} from different approaches are provided in Table 1 for different noise variances. This optimum order corresponds to an m that minimizes z_{S_m} or its estimates. This value directly provides the optimum threshold. A method that its m_{opt}, and consequently, its threshold is closer to true z_{S_m}'s is a better method. Table 1 shows that m_{opt} of MNDL thresholding is closer to true m_{opt} than that the existing MDL for all levels of noise variance. In addition to MNDL, MNDL thresholding has also been compared with other existing thresholding methods VisuShrink and MDL thresholding. Table 2 provides the resulted thresholds of these methods. The MSE of these methods are displayed in Table 3. As the table shows, in the case of block signal with small nonzero coefficients, MNDL thresholding outperforms all the existing methods except for unit noise variance. For this noise variance MNDL thresholding and MDL thresholding provide similar results. However, for mishmash signal with large nonzero coefficients, when the level of noise is small, MNDL thresholding works much better than existing methods.

	(1)	(2)	(3)	(4)
$\sigma^2 = 1$	2.14	2.7	2.6	2.6
$\sigma^2 = 3$	6.5	11	7.8	9.6
$\sigma^2 = 5$	14	18.6	13	17.8

Table 2. Thresholds of different thresholding methods for different noise variance levels:(1) existing MNDL; (2) Visushrink; (3) MDL thresholding; (4) proposed MNDL thresholding .

Block	(1)	(2)	(3)	(4)
$\sigma^2 = 1$	0.2	0.19	**0.18**	**0.18**
$\sigma^2 = 3$	2.18	1.70	1.76	**1.58**
$\sigma^2 = 5$	4.38	3.8	4.1	**3.7**
Mishmash				
$\sigma^2 = 1$	1.2	3.27	2.1	**1**
$\sigma^2 = 3$	7.4	**7.34**	7.38	8.4
$\sigma^2 = 5$	15	7.9	10.25	**7.86**

Table 3. Mean square error (MSE) of different thresholding approaches for different noise variance levels: (1) existing MNDL; (2) Visushrink; (3) MDL thresholding; (4) proposed MNDL thresholding .

7. CONCLUSION

A new thresholding method based on MNDL basis selection approach was proposed. MNDL is a new subspace selection method that provides bounds on the desired reconstruction mean square error (MSE) for subspaces of different order. The approach uses the available data error to provide estimate of the desired MSE for comparison of competing subspaces. In this approach, the structures of the desired MSE and the data error play important roles. These two quantities are samples of two random variables and the approach heavily relies on the second order statistics of these two random variables. In this paper, we developed MNDL thresholding by providing the exact required statistics for when MNDL is used for thresholding. In this case, the subspace selection is based on the observed data and unlike the exiting MNDL, the statistics of the involved random variables are data dependent. We provided a new approach for calculation of the desired statistics of MNDL in thresholding approach. We also compared the proposed MNDL thresholding with the existing thresholding methods. It was shown in the simulation results that the new proposed method outperforms all the existing approaches. It is important to mention that although the main focus of this paper was on hard thresholding approaches, the fundamental arguments in this paper can be generalized for soft thresholding using MNDL and need to be developed in future research studies.

8. REFERENCES

[1] D. L. Donoho and I. M. Johnstone, “Ideal spatial adaption via wavelet shrinkage,” *Biometrika*, vol. 81, pp. 425-455, 1994.

[2] J. Rissanen, “Minimum description length denoising,” *IEEE Trans. Inform. Theory*, vol. 46, pp. 2537-2543, 2000.

[3] D. L. Donoho and I. M. Johnstone, “Adapting to unknown smoothness via wavelet shrinkage”, *Journal of the American Statistical Assos.*Volume 90, pp.1200-1224, 1995.

[4] S. Beheshti and M.A. Dahleh, “A new information theoretic approach to signal denoising and best basis selection” *IEEE Trans. on Signal Processing,* Vol 53, No. 10, pp. 3613-3624, October 2005.