NOISE VARIANCE IN SIGNAL DENOISING

Soosan Beheshti, Munther A. Dahleh

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.
soosan@mit.edu,dahleh@mit.edu

ABSTRACT

In thresholding method of denoising optimum threshold is obtained as a function of additive noise variance. In practical problems, where the variance of the noise is unknown, the £rst step is to estimate the noise variance. The estimated noise variance is then implemented in calculation of the optimum threshold. The current available methods of variance estimation are heuristic. Here, we provide a new method for estimation of the additive noise variance. The method is derived from a new denoising method which is proposed in [2]. Unlike thresholding approaches the denoising method in [2] is based on comparison of subspaces of the basis. It compares a de£ned description length(DL) of the noisy data in the subspaces. We show how the estimation of the noise variance and the denoising process can be done simultaneously.

1. INTRODUCTION

Well-known methods of signal denoising are thresholding methods. The thresholding method removes the additive noise by eliminating the basis coef£cients with small absolute value which tend to be attributed to the noise. The pioneer method of thresholding is formalized by Donoho and Johnstone in wavelet denoising [4]. They provide an upperbound for the mean square error by solving a min-max problem. In this calculation the variance of the additive white Gaussian noise is assumed to be known. It is shown that the optimum threshold for wavelet denoising of a piecewise smooth signal, asymptomatically, is $\sigma_w\sqrt{2\log N/N}$, where σ_w^2 is the variance of the additive noise [4]. In [5] an estimate of the mean square denoising error as a function of a given threshold is provided heuristically. The estimate is for any class of bases. It demonstrates that, for a class of signals, $\sigma_w\sqrt{2\log N/N}$ may not necessarily provide the optimal threshold. A different denoising approach is recommended by Rissanen in [6]. The method provides a threshold which is almost half of the suggested wavelet threshold in [4].

A new method of denoising is presented in [2]. The method is based on comparison of an information theoretic criterion which is the description length of the data. The description length of data is calculated for different subspaces of the basis. The method suggests to choose the subspace for which the description length is minimum. Since the method aims to extract the most information from the noisy data, it does not provide a threshold before estimating all the basis vector coef£cients. The advantages of this method, both theoretically and practically, are discussed in [2] and [1].

The discussed thresholding methods provide thresholds which are function of the variance of the additive noise. In practical problems the variance of the noise is unknown. Heuristic methods are used to estimate the variance of additive noise([3], [6]). The estimate of the variance is then implemented to provide the optimum threshold. In this paper we provide a new method of estimation of noise variance. The method is interwoven with the new subspace comparison method of denoising in [2]. Unlike the thresholding methods, for any given noise variance the criterion, which is the description length of the data, in each subspace can be calculated. We suggest to choose the noise variance and the subspace for which the description length of the data is minimum. Therefore, the estimation of the variance and the denoising are not two separate procedures.

2. PROBLEM STATEMENT

Consider noisy data y of length N,

$$y(n) \quad = \quad \bar{y}(n) + w(n), \tag{1}$$

where $\bar{y}$ is the noiseless data and w is the additive white Gaussian noise with zero mean and variance σ_w^2. Data denoising is achieved by choosing an orthogonal basis which approximates the data with fewer nonzero coef£cients than the length of data. Consider the orthogonal basis of order N, S_N. The basis vectors $s_1, s_2, \cdots, s_N$ are such that $||s_i||_2^2 = N$. Any vector of length N can be represented with such basis, therefore there exist h_i's such that $\bar{y}(n) =$

$\sum_{i=1}^{N} s_i(n)h_i$. As a result the noisy data is

$$y(n) = \sum_{i=1}^{N} s_i(n)h_i + w(n). \tag{2}$$

The least square estimate of each basis coef£cient is

$$\hat{h}_i = \frac{1}{N} s_i^T y^N = h_i + \frac{1}{N} s_i^T w \tag{3}$$

where $y^N = [y(1), y(2), \cdots, y(N)]$, the observed noisy data, is a sample of random variable Y^N. The bene£t of using a proper basis is that $\frac{1}{N} s_i^T w$ is almost zero as N is assumed to be large enough and we hope that there exist large number of basis vectors for which $h_i = 0$. Therefore the estimation of the noisy signal on this basis has the advantage of noise elimination. For such reason conventional basis denoising methods suggest choosing a threshold, τ, for the coef£cient estimates $\hat{h}_i$'s. The denoising process is to ignore the coef£cient estimates smaller than the threshold

$$\begin{aligned} \hat{h}_i &= \frac{1}{N} s_i^T y^N, \text{ if } |\frac{1}{N} s_i^T y^N| \geq \tau \\ \hat{h}_i &= 0, \quad \text{ if } |\frac{1}{N} s_i^T y^N| < \tau \end{aligned} \tag{4}$$

and the estimate of the noiseless signal is

$$\hat{y}^N(n) = \sum_{i=1}^{N} s_i(n)\hat{h}_i. \tag{5}$$

A very important factor in solving the denoising problem and choosing the proper threshold in [4] and [5] is the behavior of the mean square reconstruction error

$$\frac{1}{N} E(||\bar{y}^N - \hat{Y}^N||_2^2). \tag{6}$$

Instead of focusing on £nding a threshold one can compare the signal estimate in different subspaces of the basis. Choosing a subspace to estimate the data is equivalent to setting the coef£cients of the basis vectors out of that subspace to zero without thresholding. [2] investigates on estimation of a criterion which is de£ned for the subspaces of the basis. For each subspace S_m, $\hat{h}_{S_m}$ denotes the estimate of the coef£cients in that subspace. The goal is to £nd an estimate of the coef£cient estimation error in each subspace, $||h - \hat{h}_{S_m}||_2^2$. Note that, as a result of the Parseval's theorem, this error is the same as the *reconstruction error* for each subspace

$$||h - \hat{h}_{S_m}||_2^2 = \frac{1}{N} ||\bar{y}^N - \hat{y}_{S_m}^N||_2^2. \tag{7}$$

The objection in [2] is to compare the worst case behavior of this error in different subspaces probabilistically. The best representative of the signal is then the signal estimate of the subspace which minimizes such criterion.

2.1. The New Denoising Method

Consider a subspace of order m of the orthogonal basis, S_m. For the subspace S_m, matrix A_{S_m} separates the basis vectors as follows

$$\begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix} = \begin{bmatrix} A_{S_m} & B_{S_m} \end{bmatrix} \begin{bmatrix} h_{S_m} \\ \Delta_{S_m} \end{bmatrix} + w \tag{8}$$

where columns of A_{S_m} are $s_i \in S_m$, columns of B_{S_m} are basis vectors which are not in S_m, $s_i \in \bar{S}_m$, and h_{S_m} is the coef£cients of the noiseless data $\bar{y}^N = [\bar{y}(1), \cdots \bar{y}(N)]^T$ in S_m. The least square estimate of coef£cients in each subspace using the noisy data is

$$\begin{aligned} \hat{h}_{S_m} &= \frac{1}{N} [A_{S_m} \; B0_{S_m}]^T y^N & (9) \\ &= h_{S_m} + \frac{1}{N} [A_{S_m} \; B0_{S_m}]^T w. & (10) \end{aligned}$$

where $B0_{S_m}$ is a matrix with zero elements and with dimension of B_{S_m}. Therefore for the subspace error we have

$$||\hat{h}_{S_m} - h||_2^2 = \frac{1}{N} ||A_{S_m}^T w||^2|| + ||\Delta_{S_m}||_2^2, \tag{11}$$

were $||\Delta_{S_m}||$ is the norm of the discarded coef£cients vector in each subspace. The estimate of the data using this subspace is $\hat{y}_{S_m} = [A_{S_m} \; B0_{S_m}]\hat{h}_{S_m}$. In [2] the description length of $\hat{y}_{S_m}$ is de£ned as

$$\text{DL}_h(\hat{y}_{S_m}, \sigma_w) = -\log \frac{1}{\sqrt{2\pi\sigma_w^2}} + \frac{||\hat{h}_{S_m} - h||^2}{2\sigma_w^2}. \tag{12}$$

The subspace comparison method is to £nd S_m for which the description length of $\hat{y}_{S_m}$ is minimum. To £nd the solution we have to estimate the reconstruction error. [2] suggests the following approach to estimate the reconstruction error.

The coef£cient error in (11) is a Chi-square random variables. Expected value and variance of coef£cient error $Z_{S_m} = ||\hat{H}_{S_m} - h||_2^2$ are

$$\text{E}(Z_{S_m}) = \text{E}||\hat{H}_{S_m} - h||_2^2 = \frac{m}{N}\sigma_w^2 + ||\Delta_{S_m}||^2 \tag{13}$$

$$\text{var}(Z_{S_m}) = \text{var}||\hat{H}_{S_m} - h||_2^2 = \frac{2m}{N^2}(\sigma_w^2)^2. \tag{14}$$

In [2] £rst $||\Delta_{S_m}||$ is validated probabilistically using the observed sample x_{S_m} from the random variable $X_{S_m} = \frac{1}{N}||Y - \hat{Y}_{S_m}||_2^2$. With validation probability of $Q(\alpha)$, where $Q(\alpha) = \int_{-\alpha}^{\alpha} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$, we have

$$L_{S_m} \leq ||\Delta_{S_m}||_2^2 \leq U_{S_m}. \tag{15}$$

The upper bound U_{S_m} is

$$U_{S_m} = x_{S_m} - m_w + \frac{2\alpha^2\sigma_w^2}{N} + K_{S_m}(\alpha). \qquad (16)$$

where $m_w = (1 - \frac{m}{N})\sigma_w^2$ and

$$K_{S_m}(\alpha) = 2\alpha\frac{\sigma_w}{\sqrt{N}}\sqrt{\frac{\alpha^2\sigma_w^2}{N} + x_{S_m} - \frac{1}{2}m_w}. \qquad (17)$$

If $(m_w - \alpha\sqrt{v_m}) \leq x_{S_m} \leq (m_w + \alpha\sqrt{v_m})$,where $v_m = \frac{2}{N}(1 - \frac{m}{N})\sigma_w^4$, the lower bound L_{S_m} is zero and if $(m_w + \alpha\sqrt{v_m}) \leq x_{S_m}$ then

$$L_{S_m} = x_{S_m} - m_w + \frac{2\alpha^2\sigma_w^2}{N} - K_{S_m}(\alpha) \qquad (18)$$

Next the probabilistic bounds on the reconstruction error are provided as following. With probability $Q(\beta)$ we have

$$|Z_{S_m} - E(Z_{S_m})| \leq \beta\sqrt{\mathrm{var}Z_{S_m}}. \qquad (19)$$

Therefore, for the choice of optimum subspace, choose m^* for which [2],

$$S_m^* = \arg\min_{S_m} \max_{||\Delta_{S_m}|| \in (L_{S_m}, U_{S_m})} \{E(Z_{S_m}) + \beta\sqrt{\mathrm{var}Z_{S_m}}\}$$

$$= \arg\min_{S_m}\{\frac{m}{N}\sigma_w^2 + U_{S_m} + \beta\frac{\sqrt{2m}}{N}\sigma_w^2\}, \qquad (20)$$

which is the bound valid with probability $Q(\beta)$ and validation probability of $Q(\alpha)$.

Note that for the validation of $||\Delta_{S_m}||$ and to obtain lower and upper bounds for $||\Delta_{S_m}||$, α has to be large enough and such that [2, 1]

$$\alpha \geq \frac{N}{\sqrt{2(N-m)}}\left(1 - \frac{m}{N} - \frac{x_{S_m}}{\sigma_w^2}\right). \qquad (21)$$

Parameters α and β can be chosen large enough such that $Q(\alpha)$ and $Q(\beta)$ are close to one. However, to have tight bounds on $||\Delta_{S_m}||$, α has to be chosen such that $\alpha/\sqrt{N}$ is small. Also β/N has to be chosen small enough so that the upper and lower bound for the con£dence region found by (19) are close. If these conditions are satis£ed, when the length of data is large enough, we can provide tight bounds on the reconstruction error with probability close to one.

3. UNKNOWN NOISE VARIANCE

The solution of the existing thresholding methods and the new method are based on the knowledge of the additive noise variance. However, in practical problems the variance of the additive noise is not known. In [3] and for wavelet thresholding it is suggested to estimate the variance with $\hat{\sigma}_w = MAD/.6745$, where MAD is the median of absolute value of normalized £ne scale wavelet coef£cients. This estimation method is a heuristic method.

We suggest a method of variance estimation which is obtained from the new MDL method for when the unknown variance is £nite number. Calculate the MDL of the data as a function of σ_w

$$\mathrm{MDL}(y, \sigma_w) = \min_{S_m} \mathrm{DL}_h(\hat{y}_{S_m}, \sigma_w^2). \qquad (22)$$

Choose the optimal noise variance such that

$$\hat{\sigma}_w = \arg\min_{\sigma_w} \mathrm{MDL}(y, \sigma_w). \qquad (23)$$

3.1. Simulation

The unit-power signal shown in £gure (1) is used to illustrate the performance of the proposed method. Figure (2) shows the absolute value of the discrete Fourier transform of the signal. In this example we use the new MDL in

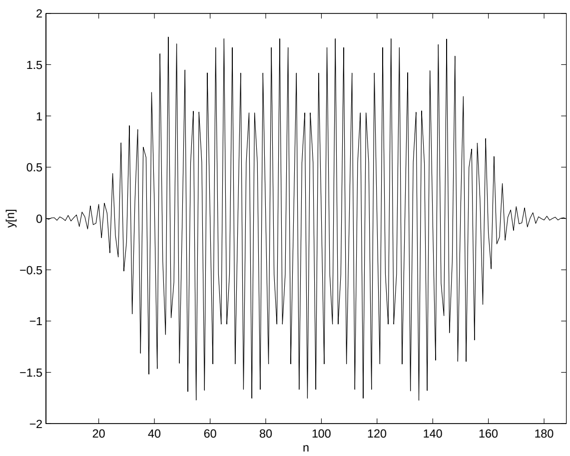

Fig. 1. Noiseless unit-power signal of length 188.

(12) to estimate the noise variance. The data is corrupted with AWGN which has variance of 0.25. First step is to £nd the valid α's for which the upper bound can be calculated. When the variance of noise is known the lower bound for α can be found by using the condition in (21). If the noise variance is $.25$ the available data shows that any α greater than $.5$ is valid. We now check for a proper choice of α, when variance is not known. Figure (3) shows the MDL for variable variances when α varies. The minimum valid α is the one for which the MDL still is a positive number. In this case, for $\beta = 1$, as the simulation shows, the lower bound for α is $.64$. The lower bound is obtained through validation of the MDL in (12). Next we choose a valid α and choose the variance for which the MDL is minimized for that α. As £gure(3) shows for $\alpha = 1$ the optimum variance is $.27$. Figure(4) shows the description length of the

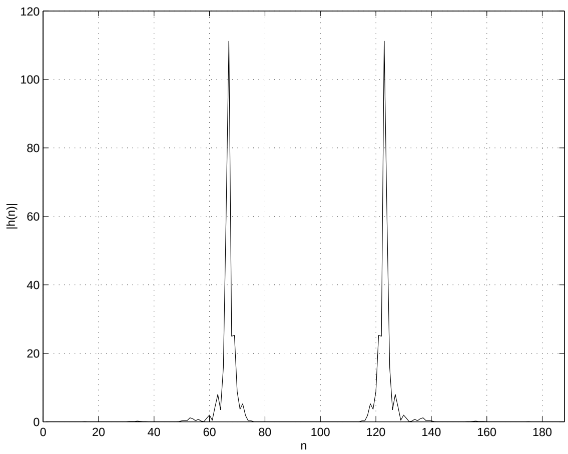

Fig. 2. 188 points discrete Fourier transform of the signal.

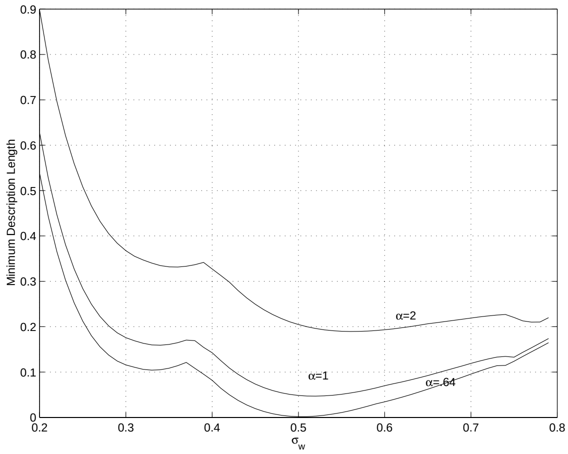

Fig. 3. MDL for standard deviations from .2 to .8 and for different αs with $\beta = 1$.

data with variance .27 as a function of m. The corresponding S_m is the one which minimizes the DL for a £xed m. In this case the new MDL denoising method chooses S_8 as the best subspace to represent the data. The same £gure also shows the true description length of the data with the known variance. The optimum subspace if the variance is known is also S_8. In this case the validation probability and the con£dence probability are both $Q(1) = .68$. Note that the simulations shows that the algorithm is robust on the choice of α. For example for $\alpha = 2$ (Q(2)=.95) the optimum variance as £gure (3) shows is $\hat{\sigma}_w^2 = .36$ ($\hat{\sigma}_w = .6$) and in this case S_8 is still the optimum subspace.

4. CONCLUSION

In this paper we presented a new method for noise variance estimation in signal denoising. The method was de-

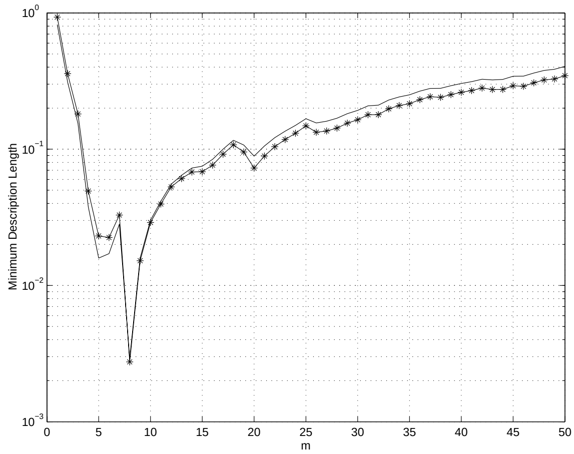

Fig. 4. Denoising with $\beta = 1$ and $\alpha = 1$: Solid line is the description length with variance $\sigma_w^2 = .25$ ($\sigma_w = .5$) for subspace S_m. Line with '*' is the description length which is provided with the estimated variance $\hat{\sigma}_w^2 = .27$.

rived based on the denoising method in [2]. The advantage of this method is that the denoising and estimation of the noise variance are provided simultaneously. The consistent theory of this method promises to overcome several practical problems with the existing noise variance estimation and thresholding methods, which are now widely used for denoising and are heuristic methods.

5. REFERENCES

[1] S. Beheshti. "Minimum Description Complexity", Ph.D. thesis, MIT, 2002.

[2] S. Beheshti and M.A. Dahleh. "On Denoising and Signal Representation", *Proceedings of the 10th Mediterranean Conference on Control and Automation*, July 2002.

[3] D. Donoho. "De-noising by Soft Thresholding", *IEEE Trans. on Information Theory*, vol.41, pp. 613-627, 1995.

[4] D. Donoho, I. M. Johnstone. "Ideal Spatial Adaptation by Wavelet Shrinkage", *Biometrika*, pp. 425-455, (1994).

[5] H. Krim, D. Tucker, S. Mallat, D. Donoho. "On Denoising and Best Signal Representation", *IEEE Trans. on Information Theory*, vol.45, pp. 2225-2238, 1999

[6] J. Rissanen. "Minimum Description Length Denoising", *IEEE Trans. on Information Theory*, vol.46, pp. 2537-2543, 2000.