BEST BASIS ALGORITHM FOR SIGNAL ENHANCEMENT

Hamid Krim S. Mallat† D. Donoho‡ and A.S. Willsky

Stochastic Systems Group, LIDS, Room 35-431,
MIT, Cambridge, MA 02139,
e-mail: ahk@mit.edu, tel: 617-258-3370, fax: 617-258-8553,
†Courant Institute, NY. NY,
‡Stanford Univ., CA

ABSTRACT

We propose a Best Basis Algorithm for Signal Enhancement in white gaussian noise. We base our search of best basis on a criterion of minimal reconstruction error of the underlying signal. We subsequently compare our simple error criterion to the Stein unbiased risk estimator, and provide a subtantiating example to demonstrate its performance.

1. Introduction

A universal wavelet basis is more than one could achieve given the plethora of classes of signals. Adapted wavelet bases have consequently been proposed [1,7,9] to alleviate this problem. In a sense, an adapted (best) basis search is intimately tied to noise removal (or signal enhancement).

To address an inherent variability of the entropy-based basis search [9] in noisy scenarios, a new class of algorithms have recently been studied in [3] and also in [6]. Donoho and Johnstone [3] base their algorithm on the removal of additive noise from deterministic signals. The signal is discriminated from the noise by choosing an orthogonal basis which efficiently approximates the signal (with few non-zero coefficients). Signal enhancement is achieved by discarding components below a predetermined threshold. Wavelet orthonormal bases are particularly well adapted to approximate piece-wise smooth functions. The non-zero wavelet coefficients are typically located in the neighborhood of sharp signal transitions. It was shown in [4] that thresholding at a specific level a noisy piece-wise smooth signal in a wavelet basis, provides a quasi optimal min-max estimator of the signal values.

When a signal includes more complex structures and in particular high frequency oscillations, wavelet bases can approximate it with only a few non-zero coefficients. It then becomes necessary to adaptively select an appropriate "*best basis*" which provides the best signal estimate by discarding (thresholding) the noisy coefficients. This approach was first proposed by Johnstone and Donoho [3] who proceed to a best basis search in families of orthonormal bases constructed with wavepackets or local cosine bases.

The work of the first and last authors was supported in part by the Army Research Office (DAAL-03-92-G-115), Air Force Office of Scientific Research (F49620-92-J-2002) and National Science Foundation (MIP-9015281). The second author is supported by AFOSR grant F49620-93-1-0102 and ONR grant N00014-91-J-1967

This paper focuses on a more precise estimate of the mean-square error introduced by the thresholding algorithm, which leads to a new strategy of searching for a best basis. The precision of the estimate is analyzed by calculating the exact mean square error with the help of the Stein unbiased risk estimator in a Gaussian setting.

Next section gives a brief review of noise removal by thresholding and of wavepacket orthonormal bases. In Section 3, we derive the error estimate of our signal enhancement procedure. In Section 4, we compare the optimal best basis search criterion to a suboptimal one, and provide a numerical example of the resulting algorithm to enhance underwater signals (whale signals) in Section 5.

2. Noise Removal by Thresholding

Let $s[m]$ be a deterministic discrete unknown signal embedded in a discrete white Gaussian noise,

$$x[m] = s[m] + n[m] \quad \text{with} \quad n[m] \sim N(0, \sigma^2), \qquad (1)$$

and $m = 0, \cdots, N$. Let $\mathcal{B} = \{W_p\}_{1 \leq p \leq N}$ be a basis of our observation space. The thresholding procedure consists of discarding all inner products $< x, W_p >$ above T, in order to reconstruct an estimate $\hat{s}$ of s. Let K be the number of inner products such that $| < x, W_p > | > T$. Suppose that $< x, W_p >$ are sorted so that $| < x, W_p > |$ is decreasing for $1 \leq p \leq N$,

$$\hat{s}[m] = \sum_{n=1}^{K} < x, W_p > W_p[m].$$

The threshold T is set such that it is unlikely that $| < n, W_p > | > T$. For a Gaussian white noise of variance σ^2, $\{< n, W_p >\}_{1 \leq p \leq N}$ are N independent Gaussian random variables with the same variance. The value assumed by the maximum of $\{| <$

0-7803-2431-5/95 $4.00

$n, W_p > |^2\}_{1 \leq p \leq N}$ is "$2\sigma^2 \log N$" w.p.1 [5]. To guarantee that the thresholded coefficients always include some signal information, one chooses $T = 2\sigma^2 \log N$, which was shown to be the optimal threshold from a number of perspectives [4,6]. The vectors W_p for $p \leq K$ generally have for weights non-zero signal coefficient $| < s, W_p > |$.

If the energy of $s(n)$ is concentrated in high amplitude coefficients, such a representation can provide an accurate estimate of $s(m)$. Wavelet bases are known to concentrate the energy of piece-wise smooth signals into a few high energy coefficients [2].

When the signal possesses more complex features, one has to search for the basis which would result in its best compressed representation. In [3] a proposed to search for a basis, amounted to that which resulted in the best noise suppression among all wavepacket or local cosine bases.

We thus have a collection of orthonormal bases $\left(\mathcal{B}^i = \{W_p^i\}_{1 \leq p \leq N}\right)_{i \in I}$ among which we want to choose the basis which leads to the best thresholded estimate $\hat{s}$. We consider two particular classes of families of orthonormal bases. Trees of wavepacket bases studied by Coifman and Wickerhauser [9], are constructed by quadrature mirror filter banks and are composed of signals that are well localized in time and frequency. This family of orthonormal bases divides the frequency axis in intervals of different sizes, varying with the selected wavepacket basis. Another family of orthonormal bases studied by Malvar [7] and Coifman and Meyer [1] can be constructed with a tree of window cosine functions. These orthonormal bases correspond to a division of the time axis in intervals of varying sizes. For a discrete signal of size N, one can show that a tree of wavepacket bases or local cosine bases include more than 2^N bases, and the signal expansion in these bases is computed with algorithms that require $O(N \log N)$ operations, since these bases include many vectors in common. Wickerhauser and Coifman [9] proved that for any signal f and any function $\mathcal{C}(x)$, finding the best basis $\mathcal{B}^{i_0}$ which minimizes an "additive" cost function over all bases

$$Cost(f, \mathcal{B}^i) = \sum_{n=1}^{N} \mathcal{C}(| < f, W_p^i > |^2)$$

requires $O(N \log N)$ operations.

In this paper we derive an expression of $\mathcal{C}(x)$ so that $Cost(x, \mathcal{B}^i)$ approximates the mean-square error $\|s - \hat{s}\|^2$ of the noise removal algorithm. The best noise removal basis is then obtained by minimizing this cost function.

3. Error Estimation

We first derive with qualitative arguments an estimate of the error $\|s - \hat{s}\|$ and then prove that this estimate is a lower bound of the true error. Since $x[m] = s[m] + n[m]$, following the thresholding and assuming that K terms are above T, the error is

$$\|s - \hat{s}\|^2 = \sum_{n=1}^{K} | < n, W_p > |^2 + \sum_{n=K+1}^{N} | < s, W_p > |^2. \quad (2)$$

Since $n[m]$ is a random process, K is a random variable dependent upon $n(\cdot)$. Given that $n[m]$ is a Gaussian white noise of variance σ^2, $< n, W_p >$ has a variance σ^2. To estimate $\|s - \hat{s}\|^2$, we obtain for a given K,

$$\mathrm{E}\{\sum_{n=1}^{K} (< n, W_p >)^2 \mid K\} \approx K\sigma^2,$$

and

$$\mathrm{E}\{ \sum_{n=K+1}^{N} | < x, W_p > |^2 \mid K\} =$$

$$\sum_{n=K+1}^{N} | < s, W_p > |^2 + \sum_{n=K+1}^{N} \mathrm{E}\{(< n, W_p >)^2 \mid K\}$$

Since the basis includes N vectors

$$\mathrm{E}\{|s - \hat{s}|^2 \mid K\} \approx \epsilon_x^2 = -N\sigma^2 + 2K\sigma^2 + \sum_{n=K+1}^{N} \mathrm{E}\{| < x, W_p > |^2 \mid K\} \quad (3)$$

The estimator ϵ_x^2 can be written as an additive cost function (?). Let us define

$$\mathcal{C}(u) = \begin{cases} u & \text{if } |u| > T \\ +2\sigma^2 & \text{if } |u| \leq T, \end{cases}$$

$$\epsilon_x^2 = -N\sigma^2 + \mathrm{E}\{\sum_{n=1}^{N} \mathcal{C}(| < x, W_p > |^2)\}.$$

Among a collection $\{\mathcal{B}^i\}_i$ of orthonormal bases, we propose to use this estimated error to search for the best basis $\mathcal{B}^{i_0}$ which minimizes $\sum_{n=1}^{N} \mathcal{C}(| < x, W_p > |^2)$, i.e., minimize the conditional error estimate. With this cost function, we are able to efficiently compute the best basis in wavepacket and local cosine tree bases, with $O(N \log N)$ operations.

In the following theorem we show that the conditional estimator ϵ_x^2 of $\|s - \hat{s}\|^2$ is biased and compute the biais by using the Stein unbiased risk estimator [8]. As will be shown in Section 3, the bias will not, however, have a bearing on the optimality of our search, if a threshold T is judiciously chosen.

Theorem 1 *Let $\{W_p\}_{1 \leq p \leq N}$ be an orthonormal basis of our observations space. If $n[p]$ is a discrete Gaussian white noise of variance σ^2, the biais of the estimator is*

$$E[\|s - \hat{s}\|^2] - E[\epsilon_x^2] = 2T\sigma^2 \sum_{p=1}^{N} (\phi(T - < s, W_p >) + \phi(-T - < s, W_p >)) \quad (4)$$

with

$$\phi(u) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{u^2}{2\sigma^2}}.$$

Proof: First define,

$$\begin{aligned} \gamma_t(\eta_p) &= \eta_p \mathcal{I}_{\{|\eta_p|>|T\}} \\ g(\eta_p) &= -\eta_p \mathcal{I}_{\{|\eta_p|\leq T\}}, \end{aligned} \quad (5)$$

with $\eta_p \sim N(\eta_{sp}, \sigma^2)$, with η_{sp} being the true signal coefficient, in $\eta_p = \eta_{sp} + \eta_{np}$ and $\mathcal{I}_{\{.\}}$ is an indicator function constrained by its argument. We then use

$$\gamma(\eta_p) = \eta_p + g(\eta_p),$$

to obtain the following,

$$\begin{aligned} \mathrm{E}\{\sum_{p=1}^{N}(\gamma(\eta_p) - \eta_{sp})^2\} &= \sum_{p=1}^{N} \mathrm{E}\{(\eta_p - \eta_{sp}) + (g(\eta_p))^2\} \\ &= \sum_{p=1}^{N} \mathrm{E}\{\eta_{np}^2\} + 2\mathrm{E}\{\eta_{np} g(\eta_p)\} \\ &\quad + \mathrm{E}\{g^2(\eta_p)\}. \end{aligned} \quad (6)$$

Using the property [8],

$$\begin{aligned} \mathrm{E}\{\eta_{np} g(\eta_p)\} &= \int \eta_{np} g(\eta_{np} + \eta_{sp}) \phi(\eta_{np}) d\eta_{np} \\ &= -\sigma^2 \int g(\eta_{np} + \eta_{sp}) \phi'(\eta_{np}) d\eta_{np} \\ &= \sigma^2 \int g'(\eta_{np} + \eta_{sp}) \phi(\eta_{np}) d\eta_{np}, \end{aligned}$$

and calling upon derivatives in the generalized (distributions) sense, one can write,

$$\frac{d}{d\eta_p} \mathcal{I}_{\{\eta_p \geq 0\}} = \delta_0,$$

with δ_0 denoting the Dirac impulse, and which we can in turn use to derive

$$\begin{aligned} &\int g'(\eta_{np} + \eta_{sp}) \phi(\eta_{np}) d\eta_{np} = \\ &-\int \mathcal{I}_{\{|\eta_p| \leq T\}} \phi(\eta_{np}) d\eta_{np} + T(\phi(T - \eta_{sp}) + \phi(-T - \eta_{sp})) \end{aligned}$$

After susbtitution of the above expressions back into Eq. 6, we obtain,

$$\begin{aligned} \mathrm{E}\{\sum_{p=1}^{N}(\gamma(\eta_p) - \eta_{sp})^2\} &= \epsilon_x^2 + 2T\sigma^2 \sum_{p=1}^{N}(\phi(T - \eta_{sp}) + \\ &\quad \phi(-T - \eta_{sp})). \end{aligned}$$

■

This theorem proves that the expected value of our conditional estimator ϵ_x is a lower bound of the mean-square error. The biais of the estimator is explained by the assumpotion that all signal components are always above $| T |$ in (2) and (3). We thus did not take into account the errors due to an erroneous decision (i.e. signal + noise component may indeed fall below the threshold)

4. Evaluation and Numerical Results

To analyze the performance of our error-based criterion for a best basis selection, we first numerically compute biais for different classes of signals by using Eq. (4). We then apply this criterion to choose a best basis among wavepacket bases, and subsequently to enhance signals of biological underwater sounds.

Let us consider a class of signals $s[p]$ that are well approximated by K coefficients of the orthonormal basis $\{W_p\}_{1\leq p\leq N}$. We associate to the inner products $< s, W_p >$ a distribution density given by

$$p(\theta) = \frac{N-K}{N}\delta(\theta) + \frac{K}{N}h(\theta).$$

This may be interpreted as on average, out of N coefficients there are $N - K$ zero-coefficients and K non-zero coefficients whose values are specified by $h(\theta)$. The smaller the proportion K/N the better the performance of the noise removal algorithm. Fig. 1 shows the mean-square error $E[\|s - \hat{s}\|^2]$ as a function of T, for different values of K/N. In these numerical computations, we suppose that the signal to noise ratio is of $0dB$. The parameters of $h(\theta)$ are adjusted so that the total signal energy is equal to the total noise energy. The minimum expected value of the unbiased risk is obtained for a value of T which is close to $2\sigma^2 \log N$[1]. The value of this optimal T does not remain invariant and is a function of K/L. Fig. 1 also gives the expected error $E[\epsilon_x^2]$ computed with our estimator. The precision of this lower bound increases when the poportion of non-zero coefficients K/L decreases. For small values of T the biais is very large but it is considerably reduced at $T = 2\sigma^2 \log N$ which corresponds to the typical threshold we choose in our pratical algorithm. For this threshold, the suboptimal error estimator provides a resonable estimate of the mean-square error.

Note that the bias term in Eq. (4 assumes a prior knowledge of the signal coefficients, which is clearly not the case. This difficulty can be partially lifted by obtaining rather an upper bound of the bias. This is carried out by picking the Maximum Likelihood Estimate, namely the observation itself (i.e. the noisy coefficient) at the given point. We show in Fig. 2 that the effect on the risk can be rather drastic in comparison to the optimal risk.

We show in Fig. 3 a noisy whale sound[2] with its best enhanced representation, using the proposed algorithm.

5. References

[1] R. Coifman and Y. Meyer. Remarques sur l'analyse de Fourier à fenêtre. *C. R. Acad. Sci. Série I*, pages 259–261, 1991.

[2] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Com. Pure and Appl. Math.*, XLI:909–996, 1988.

[1] N=100.

[2] This data originated at NUSC.

[3] D. Donoho and I. Johnstone. Ideal denoising in an orthogonal basis chosen from a library of bases. Oct. 1994. preprint Dept. Stat., Stanford Univ.

[4] D. L. Donoho and I. M Johnstone. Ideal spatial adaptation by wavelet shrinkage. preprint Dept. Stat., Stanford Univ., Jun. 1992.

[5] B. Gnedenko. Sur la Distribution Limite du Terme Maximum d'une Serie Aleatoire. *Annals of Mathematics*, 44(3):423–453, July 1943.

[6] H. Krim and J.-C. Pesquet. Robust multiscale representation of processes and optimal signal reconstruction. In *Time Freq./Time Scale Symp.*, Philadelphia, PA, 1994.

[7] H. Malvar. Lapped transforms for efficient transform subband coding. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-38:969–978, Jun. 90.

[8] C.M. Stein. Estimation of the Mean of a Multivariate Normal Distribution. *The annals of Statistics*, 9(6):1135–1151, 1981.

[9] M. V. Wickerhauser. INRIA lectures on wavelet packet algorithms. In *Ondelettes et paquets d'ondelettes*, pages 31–99, Roquencourt, Jun. 17-21 1991.

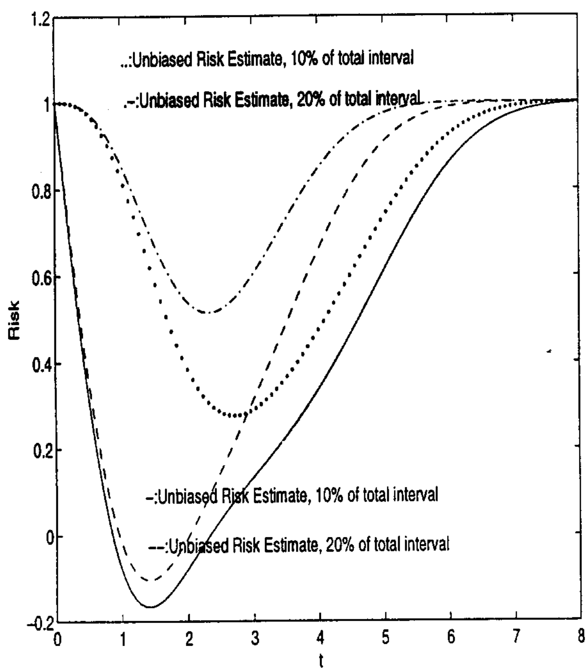

Figure 1: Risk estimates for various compression ratios

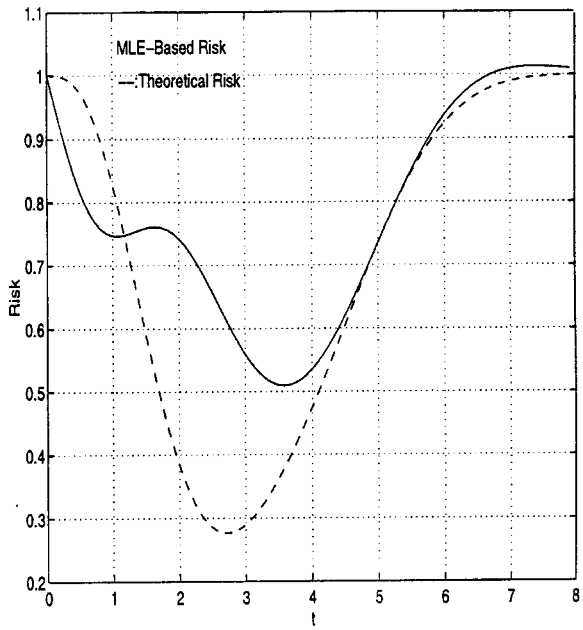

Figure 2: Comparison of MLE-based Risk and suboptimal Risk

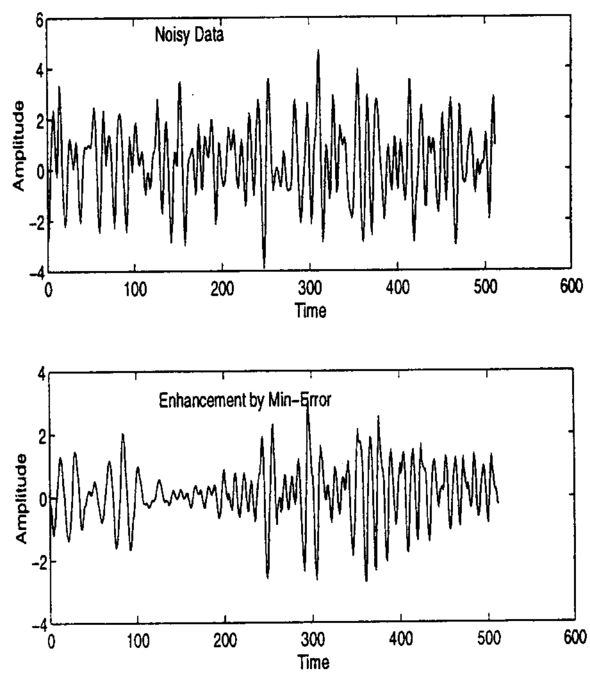

Figure 3: Biological Sounds of a whale