Best Basis Algorithm for Signal Enhancement* 

Krim, H.  
Mallat, S.  
Donoho, D.  
Willsky, A.S.
BEST BASIS ALGORITHM FOR SIGNAL ENHANCEMENT

H. 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

This paper appeared in the proceedings of ICASSP’95

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 substantiating 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.

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,

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

and \( m = 0, \ldots, N \). Let \( B = \{W_p\}_{1 \leq p \leq N} \) 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
local cosine bases. To search for a basis, amounted to that which resulted in its best compressed representation. In [3] a proposed and one has to search for the basis which would result in a few high energy coefficients [2].

An accurate estimate of its best compressed representation can provide an estimate of 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 resulted in the best noise suppression among all wavepacket or local cosine bases.

We thus have a collection of orthonormal bases \( \{ B^t \} = \{ \{ W_p \} \}_{1 \leq p \leq N} \), 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 varying 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 \( C(x) \), finding the basis \( B^0 \) which minimizes an “additive” cost function over all bases

\[
\text{Cost}(f, B^t) = \sum_{n=1}^{N} C(\| f, W^t_p \| ^2)
\]

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

In this paper we derive an expression of \( C(x) \) so that \( \text{Cost}(x, B^t) \) 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 \( \| s - \hat{s} \| ^2 \) 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

\[
\| s - \hat{s} \| ^2 = \sum_{n=1}^{K} \| s, W^t_p \| ^2 + \sum_{n=K+1}^{N} \| s, W^t_p \| ^2.
\]

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 \( \sigma^2 \), \( \| s, W^t_p \| \) has a variance \( \sigma^2 \). To estimate \( \| s - \hat{s} \| ^2 \), we obtain for a given \( K \),

\[
E\{ \sum_{n=1}^{K} \| s, W^t_p \| ^2 | K \} \approx K \sigma^2,
\]

and

\[
E\{ \sum_{n=K+1}^{N} \| s, W^t_p \| ^2 | K \} = \sum_{n=K+1}^{N} \| s, W^t_p \| ^2 + \sum_{n=K+1}^{N} E\{ \| s, W^t_p \| ^2 | K \}
\]

Since the basis includes \( N \) vectors

\[
E\{ \| s - \hat{s} \| ^2 | K \} \approx \epsilon_k^2 = -\sigma^2 N + 2K \sigma^2 + \sum_{n=K+1}^{N} E\{ \| s, W^t_p \| ^2 | K \}
\]

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

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

\[
\epsilon_k^2 = -\sigma^2 N + \sum_{n=1}^{N} C(\| s, W^t_p \| ^2).
\]

Among a collection \( \{ B^t \} \), of orthonormal bases, we propose to use this estimated error to search for the best basis \( B^0 \) which minimizes \( \sum_{n=1}^{N} C(\| s, W^t_p \| ^2) \), 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 \( \epsilon_k^2 \) of \( \| s - \hat{s} \| ^2 \) is biased and compute the bias 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 \( \{ W^t_p \} \) be an orthonormal basis of our observations space. If \( n[p] \) is a discrete Gaussian white noise of variance \( \sigma^2 \), the bias of the estimator is

\[
E[\| s - \hat{s} \| ^2] - E[\epsilon_k^2] = 2T \sigma^2 \sum_{p=1}^{N} (\phi(T - \| s, W^t_p \| ) + \phi(-T - \| s, W^t_p \| ))
\]
with
\[ \phi(u) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{u^2}{2\sigma^2}}. \]

**Proof:** First define,
\[ \gamma_\ell(\eta_p) = \eta_p \mathcal{I}_{\{|\eta_p| > \ell\}} \]
\[ g(\eta_p) = -\eta_p \mathcal{I}_{\{|\eta_p| \leq T\}}, \]
with \( \eta_p \sim N(\eta_{sp}, \sigma^2) \), with \( \eta_{sp} \) being the true signal coefficient, in \( \eta_p = \eta_{sp} + \eta_n \) and \( \mathcal{I}_{\cdot} \) is an indicator function constrained by its argument. We then use
\[ \gamma(\eta_p) = \eta_p + g(\eta_p), \]
to obtain the following,
\[ \mathbb{E}\left\{ \sum_{p=1}^{N}(\gamma(\eta_p) - \eta_{sp})^2 \right\} = \sum_{p=1}^{N} \mathbb{E}\{ (\eta_p - \eta_{sp}) + (g(\eta_p))^2 \} = \sum_{p=1}^{N} \mathbb{E}\{ \eta_{sp}^2 \} + 2\mathbb{E}\{ \eta_{np} g(\eta_p) \} + \mathbb{E}\{ g^2(\eta_p) \}. \] (6)

Using the property \([8]\),
\[ \mathbb{E}\{ \eta_{np} g(\eta_p) \} = \int \eta_{np} g(\eta_{np} + \eta_{sp}) \phi(\eta_{np}) d\eta_{np} = -\sigma^2 \int g(\eta_{np} + \eta_{sp}) \phi(\eta_{np}) d\eta_{np} = \sigma^2 \int g'(\eta_{np} + \eta_{sp}) \phi(\eta_{np}) d\eta_{np}, \]
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 \( \delta_0 \) denoting the Dirac impulse, and which we can in turn use to derive
\[ \int g'(\eta_{np} + \eta_{sp}) \phi(\eta_{np}) d\eta_{np} = -\int \mathcal{I}_{\{|\eta_p| \leq T\}} \phi(\eta_{np}) d\eta_{np} + T(\phi(T - \eta_{sp}) + \phi(-T - \eta_{sp})). \]
After substitution of the above expressions back into Eq. 6, we obtain,
\[ \mathbb{E}\left\{ \sum_{p=1}^{N}(\gamma(\eta_p) - \eta_{sp})^2 \right\} = \sigma^2 + 2T \sigma^2 \sum_{p=1}^{N} (\phi(T - \eta_{sp}) + \phi(-T - \eta_{sp})). \]

4. **Evaluation and Numerical Results**

To analyze the performance of our error-based criterion for a best basis selection, we first numerically compute bias 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 \( \{W_p\}_{p=1}^{N} \). We associate to the inner products \( \langle s, W_p \rangle \) 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 \( \mathbb{E}\{\|s - \hat{s}\|^2\} \) 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 \( 2\sigma^2 \log N \).

5. **References**


\(^{1}N = 100.\)

\(^{2}\)This data originated at NUSC.


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