Top-Down and Bottom-Up Tree Search Algorithms for Selecting Bases in Wavelet Packet Transforms

Carl Taswell⋆

Scientific Computing and Computational Mathematics,
Bldg 460 Room 314, Stanford University, Stanford, CA 94305-2140

Abstract

Search algorithms for finding signal decompositions called near-best bases using decision criteria called non-additive information costs have recently been proposed by Taswell [12] for selecting bases in wavelet packet transforms represented as binary trees. These methods are extended here to distinguish between top-down and bottom-up tree searches. Other new non-additive information cost functions are also proposed. In particular, the near-best basis with the non-additive cost of the Shannon entropy on probabilities is compared against the best basis with the additive cost of the Coifman-Wickerhauser entropy on energies [3]. All wavelet packet basis decompositions are also compared with the nonorthogonal matching pursuit decomposition of Mallat and Zhang [7] and the orthogonal matching pursuit decomposition of Pati *et al* [8]. Monte Carlo experiments using a constant-bit-rate variable-distortion paradigm for lossy compression suggest that the statistical performance of top-down near-best bases with non-additive costs is superior to that of bottom-up best bases with additive costs. Top-down near-best bases provide a significant increase in computational efficiency with reductions in memory, flops, and time while nevertheless maintaining similar coding efficiency with comparable reconstruction errors measured by ℓ^p-norms. Finally, a new compression scheme called parameterized model coding is introduced and demonstrated with results showing better compression than standard scalar quantization coding at comparable levels of distortion.

1 Introduction

Much of the statistically oriented wavelet literature focuses on theoretical models of stochastic processes and/or asymptotic properties of statistical estimators related to wavelet analysis and methods. But according to Berkson [1], "Statistics, however you define it, is very much earthbound and deals with real observable data; what is statistically true must be literally verifiably true for such data." Referring to theorems of

⋆Internet email: taswell@sccm.stanford.edu; Telephone: 415-723-4101; Facsimile: 415-723-2411.

aymptotic analysis, he elaborates that "if these theorems were valid for large samples, they must refer to *infinitely* large samples, which is to say, samples so large that no statistician ever gets them, at least not on this unpleasant earth." He then advocates the use of actual experiments to evaluate the performance of statistical methods on small samples. It is this pragmatic empirical approach of Berkson that is adopted as the foundation for the work presented in this report. In particular, the experimental statistical performance of wavelet packet decomposition methods are investigated with regard to three separate issues: 1) the search method – various complete basis searches versus matching pursuit searches, 2) the decision criterion – various information cost functions such as entropy, and 3) the coding method – a novel method based on parameterized regression modelling versus standard scalar quantization.

Coifman and Wickerhauser [3] presented an algorithm for the selection of the best basis representation of a signal within a library of orthonormal basis representations generated by wavelet packet transforms which can be searched as balanced binary trees. They defined the best basis to be that which minimized an information cost function $\mathcal{C}$ and chose the $-\ell^2 \ln \ell^2$ functional (related to the Shannon entropy) as their archetype for $\mathcal{C}$. The computational cost of the best basis algorithm is $O(LN)$ where $L = \lfloor \log_2 N \rfloor$ is the number of levels or depth of the transform or tree and N is the length of the signal. Mallat and Zhang [7] presented a greedy algorithm for the selection of the best matching pursuit decomposition of a signal into time-frequency packets from a large dictionary of such packet waveforms. The computational cost of the matching pursuit algorithm is $O(MLN)$ where $M <= N$ is the number of packets selected. As discussed by Mallat and Zhang [7], the matching pursuit algorithm with its local optimization properties guarantees a more compact signal decomposition (typically $M << N$) than that of the best basis algorithm with its global optimization properties. However, this more compact signal coding is achieved at the expense of greater computational cost in which although $M << N$, nevertheless $1 << M$. So the additional computational cost of the matching pursuit algorithm is significant relative to that of the best basis algorithm.

This trade-off between computational cost efficiency versus signal coding efficiency for the two algorithms, best basis versus matching pursuit, raises several questions: Can "compromise" algorithms be developed with intermediate or adjustable rates of coding and computational efficiency as required by the application? Under what circumstances is it relevant and necessary to perform additional computations in order to obtain more efficient coding? In other words, what is the point of "diminishing returns"? In an initial attempt to explore these questions, Taswell proposed near-best bases with non-additive costs [12] as an alternative to best bases with additive costs [3]. In this report, I propose several more basis selection algorithms and decision criteria to be added to the list of those already presented in [12]. In particular, top-down and bottom-up tree searches are distinguished. Furthermore, the Shannon entropy on probabilities and the Coifman-Wickerhauser entropy on energies are distinguished. The statistical performance of the various basis and pursuit decompositions are investigated with Monte Carlo experiments on test signals with additive white noise. In lossy compression experiments with a uniform-quantization and fixed-bit-rate paradigm as well as with a new compression scheme called parameterized model coding, top-down bases are shown to provide performance superior to bottom-up bases with significant savings in computational requirements for memory, flops, and time.

2 Information Cost Functions

We consider data vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$ and $\mathbf{z} \in \mathbb{R}^{2N}$ for which we wish to compare their information costs by some measure.

2.1 Additive Costs and Comparisons

Additive costs were originally intended for use with the best bases of Coifman and Wickerhauser [3].

Definition: A cost functional $\mathcal{C}^{\rm add}$ from vectors $\mathbf{x} \in \mathbb{R}^N$ to $\mathbb{R}$ is called an *additive information cost function* if $\mathcal{C}^{\rm add}(0) = 0$ and $\mathcal{C}^{\rm add}(\mathbf{x}) = \sum_i \mathcal{C}^{\rm add}(x_i)$.

Definition: The inequality $\mathcal{C}^{\rm add}(\mathbf{z}) \leq \mathcal{C}^{\rm add}(\mathbf{x}, \mathbf{y})$ between vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$ and $\mathbf{z} \in \mathbb{R}^{2N}$ is called an *additive information cost comparison* if $\mathcal{C}^{\rm add}(\mathbf{x}, \mathbf{y}) \equiv \mathcal{C}^{\rm add}(\mathbf{x} \oplus \mathbf{y}) = \mathcal{C}^{\rm add}(\mathbf{x}) + \mathcal{C}^{\rm add}(\mathbf{y})$.

We can define several additive information cost functions as

$$\begin{aligned}
\mathcal{C}_1^{\rm add}(\mathbf{y}) &= \mathcal{E}^p(\mathbf{y}) = \sum_i |y_i|^p \\
\mathcal{C}_2^{\rm add}(\mathbf{y}) &= \mathcal{F}(\mathbf{y}) = -\sum_{i:y_i \neq 0} y_i^2 \ln y_i^2 \\
\mathcal{C}_3^{\rm add}(\mathbf{y}) &= \mathcal{G}(\mathbf{y}) = \sum_{i:y_i \neq 0} \ln y_i^2
\end{aligned}$$

which are respectively the ℓ^p functional related to energy and the ℓ^p norm, the $-\ell^2 \ln \ell^2$ functional related to Shannon entropy, and the $\ln \ell^2$ functional related to Gauss-Markov entropy[2] (*cf.* [15]).

2.2 Non-Additive Costs and Comparisons

Non-additive costs were proposed for use with the near-best bases of Taswell [12].

Definition: A cost functional $\mathcal{C}^{\rm non}$ from vectors $\mathbf{x} \in \mathbb{R}^N$ to $\mathbb{R}$ is called a *non-additive information cost function* if it serves as a decision criterion for a basis selection algorithm and it is not an additive information cost function $\mathcal{C}^{\rm add}$.

Definition: The inequality $\mathcal{C}^{\rm non}(\mathbf{z}) \leq \mathcal{C}^{\rm non}(\mathbf{x}, \mathbf{y})$ between vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$ and $\mathbf{z} \in \mathbb{R}^{2N}$ is called a *non-additive information cost comparison* if $\mathcal{C}^{\rm non}(\mathbf{x}, \mathbf{y}) \equiv \mathcal{C}^{\rm non}(\mathbf{x} \oplus \mathbf{y}) \neq \mathcal{C}^{\rm non}(\mathbf{x}) + \mathcal{C}^{\rm non}(\mathbf{y})$.

We can construct several examples of non-additive cost functions from the probability density function for the data coefficients. In the discrete vector context, a probability mass function (pmf) can be estimated with simple histogram binning methods in conjunction with various rules for the number of bins. Thus let

$$\begin{aligned}
J_{\rm S} &= 1 + \log_2 N \\
J_{\rm D} &= 1 + \log_2 N + \log_2(1 + \hat{\gamma}\sqrt{N/6}) \\
J_{\rm TS} &= \sqrt[3]{2N}
\end{aligned}$$

[2] More precisely, the Shannon entropy of a Gauss-Markov process.

be the number of bins J according, respectively, to the Sturges', Doane's, and Terrell-Scott's rules [11, pages 48 and 73], where $\hat{\gamma}$ is an estimate of the standardized skewness coefficient. Given the number of bins J and the sample data interval $[a, b]$ where $a = \min_i y_i$ and $b = \max_i y_i$, then the bin width is $w = (b - a)/J$. Using the bin width w, the frequency f_j for the j^{th} bin is defined as

$$f_j = \#\{y_i \mid y_i \leq a + jw\} - \sum_{k=1}^{j-1} f_k$$

and the probabilities p_j are calculated from the frequencies f_j simply as $p_j = f_j/N$. Let $\mathbf{p}_{\text{S}}$, $\mathbf{p}_{\text{D}}$, and $\mathbf{p}_{\text{TS}}$ denote the pmf vectors $\mathbf{p}$ when estimated with J_{S}, J_{D}, and J_{TS}, respectively.

Now the Shannon entropy $\mathcal{H}_{\text{S}}$ [6] for a finite scheme $\{(A_j, p_j) \mid 1 \leq j \leq J\}$ of events A_j with probabilities p_j is defined as

$$\mathcal{H}_{\text{S}}(\mathbf{p}) = -\sum_{j=1}^{J} p_j \log_2 p_j$$

where the probabilistic events (A_j, p_j) are identified with the fractions of coefficients located within the histogram bin intervals. Therefore, three non-additive cost functions can be defined as

$$\begin{aligned}
\mathcal{C}_1^{\text{non}}(\mathbf{y}) &= \mathcal{H}_{\text{S}}(\mathbf{p}_{\text{S}}(\mathbf{y})) \\
\mathcal{C}_2^{\text{non}}(\mathbf{y}) &= \mathcal{H}_{\text{S}}(\mathbf{p}_{\text{D}}(\mathbf{y})) \\
\mathcal{C}_3^{\text{non}}(\mathbf{y}) &= \mathcal{H}_{\text{S}}(\mathbf{p}_{\text{TS}}(\mathbf{y})).
\end{aligned}$$

Another non-additive cost function is the Coifman-Wickerhauser entropy $\mathcal{H}_{\text{CW}}$ [3]. This functional is also the Shannon entropy of a finite scheme but one where the probabilistic events (A_j, p_j) are identified with the normalized energies rather than probabilities of the data coefficients:

$$\mathcal{C}_4^{\text{non}}(\mathbf{y}) = \mathcal{H}_{\text{CW}}(\mathbf{y}) = -\sum_{i=1}^{N} \frac{|y_i|^2}{\|\mathbf{y}\|_2^2} \ln \frac{|y_i|^2}{\|\mathbf{y}\|_2^2}.$$

We can construct additional examples of $\mathcal{C}^{\text{non}}$ with the sorted vector $[y_{(k)}]$ where

$$y_{(1)} = |y_{i_1}| \geq \cdots \geq y_{(N)} = |y_{i_N}|$$

so that $y_{(k)} = |y_{i_k}|$ is the k^{th} largest absolute value element of the vector $[y_i]$. The decreasing-absolute-value sorted vector $[y_{(k)}]$ suffices to define the weak-ℓ^p norm (*cf.* [5]). However, constructing the decreasingly sorted, powered, and cumulatively summed vector $[v_k(\mathbf{y}, p)]$, and renormalized vector $[u_k(\mathbf{y}, p)]$ where

$$u_k(\mathbf{y}, p) = \frac{v_k(\mathbf{y}, p)}{v_N(\mathbf{y}, p)} \quad \text{with} \quad v_k(\mathbf{y}, p) = \sum_{i=1}^{k} y_{(i)}^p$$

makes it convenient to define several other $\mathcal{C}^{\text{non}}$. (Note that $0 \leq u_k(\mathbf{y}, p) \leq 1$ because of the normalization.) Thus, with $[y_{(k)}]$ and $[u_k(\mathbf{y}, p)]$ obtained from $[y_i]$, define the non-additive information cost functions

$$\begin{aligned}
\mathcal{C}_5^{\text{non}}(\mathbf{y}) &= \mathcal{W}\ell^p(\mathbf{y}) = \max_k \; k^{(1/p)} y_{(k)} \\
\mathcal{C}_6^{\text{non}}(\mathbf{y}) &= \mathcal{N}_f^p(\mathbf{y}) = \arg\min_k |u_k(\mathbf{y}, p) - f| \\
\mathcal{C}_7^{\text{non}}(\mathbf{y}) &= \mathcal{A}^p(\mathbf{y}) = N - \sum_k u_k(\mathbf{y}, p)
\end{aligned}$$

which are respectively the weak-ℓ^p norm, data compression number, and data compression area [12].

Here the power p and fraction f are parameters[3] chosen from the intervals $0 < p \leq 2$ and $0 < f < 1$. The functions $\mathcal{N}_f^p$ and $\mathcal{A}^p$ were designed to yield scalar values that could be meaningfully minimized in a basis search algorithm and were named according to their natural or geometric interpretation. For example, choosing $p = 2$ and $f = .99$ and then using $\mathcal{N}_{.99}^2$ yields the minimum number of vector coefficients containing 99% of the energy of the entire vector. The data compression number $\mathcal{N}_f^p$ and area $\mathcal{A}^p$ can be contrasted by observing that the number $\mathcal{N}_f^p$ is a local measure with varying "sensitivity" to different intervals of the u_k versus k curve whereas the area $\mathcal{A}^p$ is a global measure of the entire curve. The minimum values attainable represent maximum compression. They are readily computed for a Kroniker delta vector δ with unit energy: $\mathcal{N}_f^p(\delta) = 1$ and $\mathcal{A}^p(\delta) = 0$.

3 Basis Selection Methods

We consider vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$ and orthonormal transformation matrix $\mathbf{B} \in \mathbb{R}^{N \times N}$. Then $\mathbf{y} = \mathbf{B}\mathbf{x}$ and $\mathcal{C}(\mathbf{y})$ are respectively the coefficient vector and information cost scalar for $\mathbf{x}$ in the coordinate system represented by the basis $\mathbf{B}$. We wish to find a basis $\mathbf{B}$ for which $\mathcal{C}(\mathbf{y})$ is minimal, subject possibly to some constraint on the search.

3.1 Bottom-Up Tree Searches

Both the best basis of Coifman and Wickerhauser [3] and the near-best basis of Taswell [12] were originally defined for bottom-up tree searches rather than top-down tree searches. We consider the top and bottom of the tree, respectively, to be the root node and terminal nodes of the full balanced tree corresponding to the finest and coarsest resolutions of the data.

3.1.1 Best Basis Search The best basis was defined for additive costs by Coifman and Wickerhauser [3].

Definition: The *best basis* relative to $\mathcal{C}^{\text{add}}$ for a vector $\mathbf{x}$ in a library $\mathcal{B}$ of bases is that $\mathbf{B}$ for which $\mathcal{C}^{\text{add}}(\mathbf{B}\mathbf{x})$ is minimal.

Wickerhauser [15] provided notes for an implementation of the best basis search. This search algorithm is presented below with some changes in terminology and notation and with an emphasis on data structure implementation using matrices and vectors. A discrete packet transform is considered to be any multiresolution transform (such as a wavelet packet transform or local trigonometric transform) that yields a table of transform coefficients which can be organized as a balanced binary tree. The table is

[3]The use of the parameter p for power and f for fraction should not be confused with the use of the vectors $\mathbf{p}$ and $[p_j]$ for probabilities and $\mathbf{f}$ and $[f_j]$ for frequencies.

called a discrete packet table $\mathbf{P}$ with levels l and blocks b of the table corresponding to levels l and branches b of the tree.[4] For both tables and trees, the finest and coarsest resolution scales are indexed levels 0 and L respectively. There are 2^l blocks on each level and thus $K = 2^{(L+1)} - 1$ blocks in the entire table. Within each block b on level l, there are $2^{-l}N$ cells c where N is the length of the original signal $\mathbf{x}$. Thus each coefficient in the packet table $\mathbf{P}$ can be specified as the 4-vector $[a, l, b, c]$ where a is the packet's amplitude and l, b, and c are its level, block, and cell indices.[5]

To exploit modularity, it is necessary to build two trees for each packet table $\mathbf{P}$: the additive information cost tree $\mathbf{C}^{\text{add}}$ and the basis selection tree $\mathbf{S}$. In WavBox 4.1 © 1994 Carl Taswell [13], the functions *dpt2ict* and *ict2bst* perform these mappings from discrete packet table to information cost tree and from information cost tree to basis selection tree, respectively, as

$$\begin{aligned}\mathbf{C}^{\text{add}} &= \text{dpt2ict}(\mathbf{P}, \mathcal{C}^{\text{add}}) \\ \mathbf{S} &= \text{ict2bst}(\mathbf{C}^{\text{add}}, \mathcal{S})\end{aligned}$$

with the notational convention that cost functions $\mathcal{C}$ and selection methods $\mathcal{S}$ are denoted in script font while cost trees $\mathbf{C}$ and selection trees $\mathbf{S}$ are denoted in bold font. This modularity permits 1) the output of various cost trees $\mathbf{C}^{\text{add}}$ for the same packet table $\mathbf{P}$ input to *dpt2ict* with various choices of cost functions $\mathcal{C}^{\text{add}}$ as second argument, and 2) the output of various selection trees $\mathbf{S}$ for the same cost tree $\mathbf{C}^{\text{add}}$ input to *ict2bst* with various choices of selection methods $\mathcal{S}$ as second argument.

To compare various decompositions, it is convenient to convert discrete packet tables $\mathbf{P}^{\text{table}}$ to discrete packet lists $\mathbf{P}^{\text{list}}$ representing the selected bases. Each list contains M packets specified as row 4-vectors $[a_i, l_i, b_i, c_i]$ with rows $i = 1, ..., M$ ordered so that $|a_1| \geq \cdots \geq |a_M|$. In WavBox 4.1, the function *dpt2dpl* performs this restructuring of the data via the mapping $\mathbf{P}^{\text{list}} = \text{dpt2dpl}(\mathbf{P}^{\text{table}}, \mathbf{S})$. To study the complete basis decomposition, we must examine the entire list where $M = N$. However, we may also study subsets of the list where $M < N$, for example, where we choose $M = \mathcal{N}^2_{.99} < N$. And as noted already in Section 1, matching pursuit decompositions generate $\mathbf{P}^{\text{list}}$ directly with $M << N$ so that lists with N packets are simply not available for them.

Thus, there are four data structures: $\mathbf{P}^{\text{table}} \in \mathbb{R}^{N\times(L+1)}$, $\mathbf{C}^{\text{add}} \in \mathbb{R}^K$, $\mathbf{S} \in \chi^K$ where $\chi = \{0, 1\}$, and $\mathbf{P}^{\text{list}} \in \mathbb{R}^{M\times 4}$. Since tables and trees are implemented respectively as matrices and vectors, table blocks and corresponding tree branches indexed by (l, b) are respectively vectors and scalars; they are denoted $P^{\text{table}}_{lb} \equiv P^{\text{table}}_{i_{lb},j_{lb}}$, $C^{\text{add}}_{lb} \equiv C^{\text{add}}_{k_{lb}}$, and $S_{lb} \equiv S_{k_{lb}}$ where for $l \in \{0, 1, \ldots, L\}$ and $b \in \{0, 1, \ldots, 2^l - 1\}$, the row and column vector indices i_{lb}, j_{lb} are for level l block b in a table matrix, and the scalar index k_{lb} is for level l branch b in a tree vector. Since the i^{th} packet in $\mathbf{P}^{\text{list}}$ will be denoted $P^{\text{list}}_i \equiv [a_i, l_i, b_i, c_i]$, it should be clear from context that P_i is from the list $\mathbf{P}^{\text{list}}$ while P_{lb} is from the table $\mathbf{P}^{\text{table}}$.

Now with $C^{\text{add}}_{lb} = \mathcal{C}^{\text{add}}(P_{lb})$ already computed for all l and b, and S_{lb} initialized to 1 for all b on level L and to 0 elsewhere, then the comparison and selection step of the best basis search can be expressed as

[4] For the sake of mnemonics, the term branch is used here in place of the more customary term node.

[5] These level, block, and cell indices also correspond to scale, frequency, and position indices.

if $C_{lb}^{\text{add}} \leq C_{l+1,2b}^{\text{add}} + C_{l+1,2b+1}^{\text{add}}$
then $S_{lb} = 1$
else $C_{lb}^{\text{add}} = C_{l+1,2b}^{\text{add}} + C_{l+1,2b+1}^{\text{add}}$

and the search is performed breadth-first and bottom-up through the tree. Retaining only the top-most selected branches of $\mathbf{S}$ by resetting any lower selected branches to 0 (*ie.,*pruning descendant lines) yields the best basis selection tree $\mathbf{S}$ with $S_{lb} = 1$ indicating a selected branch.

3.1.2 Near-Best Basis Search The near-best basis was defined for non-additive costs by Taswell [12]. It is modified here to be more general, allowing for inclusion of both additive and non-additive costs.

Definition: The *near-best basis* relative to $\mathcal{C}$ (either $\mathcal{C}^{\text{non}}$ or $\mathcal{C}^{\text{add}}$) for a vector $\mathbf{x}$ in a library $\mathcal{B}$ of bases is that $\mathbf{B} \in \mathcal{B}^* \subset \mathcal{B}$ for which $\mathcal{C}(\mathbf{Bx})$ is minimal subject to the constraints of the search within the subset $\mathcal{B}^*$ defined by the search type.

Here $\mathcal{B}^*$ is the proper subset of library bases that are searched by the selection algorithm. Searching the subset $\mathcal{B}^*$ defined by the Coifman-Wickerhauser bottom-up tree search yields the optimal or best basis within the entire library $\mathcal{B}$ for an *additive* information cost function $\mathcal{C}^{\text{add}}$ (*cf.* proof [3, page 717]). However, since $\mathcal{B}^* \neq \mathcal{B}$, this search is not exhaustive and cannot guarantee the selection of a *best* basis for a *non-additive* information cost function $\mathcal{C}^{\text{non}}$. Moreover, there are many other search types, including top-down tree searches, for which neither additive nor non-additive costs can guarantee the selection of a best basis. For this reason as well as empirical evidence suggesting nearly equivalent performance, a basis selected by either a non-additive or additive cost subject to the constraint of a search within a proper subset $\mathcal{B}^*$ of the library $\mathcal{B}$ is called a *near-best* basis.

The same sequence of comparisons of basis blocks' information costs are performed for the bottom-up near-best basis search as for the bottom-up best basis search. However, $\mathcal{C}^{\text{add}}$ is replaced by $\mathcal{C}^{\text{non}}$. This substitution invalidates the modular independence separating computation of costs from selection of bases described in Section 3.1.1. It is therefore necessary to combine the basis selection with the cost computation. So with $C_{lb}^{\text{non}} = \mathcal{C}^{\text{non}}(P_{lb})$ already computed for all b on level L, and S_{lb} initialized to 1 for all b on level L and to 0 elsewhere, then the comparison and selection step of the near-best basis search can be expressed as

if $\mathcal{C}^{\text{non}}(P_{lb}) \leq \mathcal{C}^{\text{non}}(P_{l+1,2b} \oplus P_{l+1,2b+1})$
then $S_{lb} = 1$
else $P_{lb} = P_{l+1,2b} \oplus P_{l+1,2b+1}$

and the search is performed breadth-first and bottom-up through the tree with pruning of descendant lines as described in Section 3.1.1. In WavBox 4.1, the function *dpt2bst* performs this mapping from discrete packet table to basis selection tree as $[\mathbf{S}, \mathbf{C}] = \text{dpt2bst}(\mathbf{P}, \mathcal{S}, \mathcal{C})$. The additional computational cost of *dpt2bst* with $\mathcal{C} = \mathcal{C}^{\text{non}}$ relative to *dpt2ict* and *ict2bst* with $\mathcal{C}^{\text{add}}$ is essentially just the cost of the sorting for those examples ($\mathcal{W}\ell^p$, $\mathcal{N}_f^p$, and $\mathcal{A}^p$) of $\mathcal{C}^{\text{non}}$ which require it as described in Section 2.2. Although not detailed here, it is possible to implement this algorithm *without* repeating for the same coefficients the required sorts and powers.

3.2 Top-Down Tree Searches

The best and near-best bases as described above are selected by breadth-first bottom-up searches through the table or tree. These searches can be implemented as the additive or non-additive cost comparison and basis selection step inside an inner for-loop for the table blocks or tree branches and an outer for-loop for the levels. Therefore, they can also be named bottom-up additive best and non-additive near-best bases with selection method $\mathcal{S} = \mathcal{U}$ to distinguish them from top-down additive near-best and non-additive near-best bases with $\mathcal{S} = \mathcal{D}$. These top-down bases are selected in the opposite direction by depth-first top-down searches with the search terminated as soon as the cost of the children blocks or branches is greater than the cost of the parent block or branch. They can be implemented as the cost comparison and basis selection step within a recursion controlled by a last-in first-out stack. Table 1 provides a summary of these alternative selection algorithms.

Table 1: Search algorithms for selecting bases in wavelet packet decompositions.

Wavelet Packet Decomposition	Notation
(Bottom-Up Additive) Best	WPDB($\mathcal{U}, \mathcal{C}^{\text{add}}$)
Bottom-Up Non-Additive Near-Best	WPDB($\mathcal{U}, \mathcal{C}^{\text{non}}$)
Top-Down Additive Near-Best	WPDB($\mathcal{D}, \mathcal{C}^{\text{add}}$)
Top-Down Non-Additive Near-Best	WPDB($\mathcal{D}, \mathcal{C}^{\text{non}}$)

Since top-down searches do not necessarily examine the entire table or tree, they cannot guarantee finding an optimal basis. However, they enable the possibility of performing the cost computation and basis selection simultaneously with generation of the packet table transform coefficients. Because the algorithm runs unidirectionally downward in the tree, it can be performed essentially "in place" thus significantly reducing memory storage requirements. Furthermore, because the algorithm does not necessarily require that the entire table and tree be generated and searched, it can be performed with significant savings in machine operations and computing time. This reduction in computational cost corresponds to a number $\hat{L}$ representing the number of levels of the transform that need to be computed. This number is estimated by totalling the number of coefficients in all parental blocks above and in the two children blocks below each selected block. Thus, $\hat{L}$ is a rational (not necessarily integer) number that ranges between 1 and L.

4 Compression and Distortion

For experiments investigating lossy compression of signals, we wish to minimize the distortion D resulting between the reconstruction $\hat{\mathbf{x}}$ and the original signal $\mathbf{x}$ following compression and coding of the wavelet packet decomposition $\mathbf{P}^{\text{list}}$. Compression can be achieved by truncating the N packets in the list to the $M < N$ largest absolute-value packets and then quantizing and coding the remaining M packets. Standard methods of coding data include scalar and vector quantization. The quantization and coding of the M packets remaining after truncation of the list applies only to the amplitudes a and not to the level-l, block-b, and cell-c indices which must be coded without loss of information. Since the (l, b, c)-index information is retained for each packet retained in this compression scheme, it is possible to consider other coding schemes for the packet amplitudes a. I propose a new method incorporating parameter estimation by statistical regression modeling with $Y_i = f(X_i, \boldsymbol{\theta}) + \epsilon_i$ for some function f, parameter $\boldsymbol{\theta}$, and noise ϵ_i. Applied in this context, I propose estimating the parameters of the smooth curve traced by a plot of v_k (defined in Section 2.2) as a function of the index k. This curve can be modeled by a low order polynomial with coefficients estimated by (possibly weighted) linear regression. Alternatively, it can be approximated by a variety of well-known nonlinear models with a small number of parameters [10]. Note that this coding scheme requires that a sign bit be retained for each packet amplitude because the sign information is lost in the conversion to v_k. However, the improvement in compression can still be significant: If the coding rate is R_{PA} bits per scalar quantized packet amplitude and R_{MC} bits per parameterized model coefficient with R_{CM} coefficients per model, then the total numbers of bits required for encoding the M packets' amplitudes are $R_{SQ} = MR_{PA}$ for scalar quantized coding and $R_{PM} = M + R_{MC}R_{CM}$ for parameterized model coding. Finally, the total number of bits required for encoding the M packets' (l, b, c) indices is denoted R_{lbc}.

4.1 Experimental Methods

Wavelet packet decompositions by basis search denoted WPDB$(\mathcal{S}, \mathcal{C})$ were compared with each other and with wavelet packet decompositions by matching pursuit denoted WPDP$(\mathcal{O})$ where $\mathcal{O}$ is a Boolean flag for orthogonality. Although not reviewed in detail here, both the nonorthogonal matching pursuit method of Mallat and Zhang [7] denoted WPDP(0) and the orthogonal matching pursuit method of Pati *et al* [8] denoted WPDP(1) decompose an N-coefficient signal $\mathbf{x}$ into an M-packet list $\mathbf{P}^{\text{list}}$ usually for which $M << N$ and for which $M_{WPDP(1)} \leq M_{WPDP(0)}$. Therefore, to compare the various decompositions, M was chosen to be that obtained from WPDP(1) which was then held fixed for all the other decompositions. Thus, the fundamental experimental paradigm investigated was the comparison of distortions resulting from decompositions compressed at equal bit-rates.

More precisely, for each Monte Carlo experiment, a test signal and wavelet packet analysis were chosen, and then for each trial t of the Monte Carlo experiment, the following steps were performed: 1) white noise was added to the test signal at signal-to-noise ratio 7; 2) the noisy signal was normalized to energy 1 and considered the t^{th}-trial signal $\mathbf{x}$ to be analyzed; 3) both WPDP$(\mathcal{O})$ were computed with a stopping criterion of 0.005 for the fraction residual signal energy; 4) all WPDB$(\mathcal{S}, \mathcal{C})$ and their R_{lbc} and $\hat{L}$ statistics were computed; 5) all WPDB$(\mathcal{S}, \mathcal{C})$ were converted to packet lists; 6) packet lists from WPDP(0) and all WPDB$(\mathcal{S}, \mathcal{C})$ were truncated to the $M_{WPDP(1)}$ largest absolute-value packets; 7) packet amplitudes from the packet lists were coded by uniform scalar quantization at the rates $R_{PA} = \{5, 8, 48\}$ bits per symbol code and by parameterized model coding at the rates $R_{CM} = \{4, 7\}$ coefficients per linear polynomial model; 8) signal estimates $\hat{\mathbf{x}}$ were reconstructed from the compressed packet lists; and 9) relative ℓ^p-norms of reconstruction errors were computed as measures of distortion

$D_p = \|\mathbf{x} - \hat{\mathbf{x}}\|_p / \|\mathbf{x}\|_p$ for $p = \{1, 2, \infty\}$. After completion of T trials in the experiment, means $\hat{\mu}$, standard deviations $\hat{\sigma}$, and coefficients of variation $\widehat{CV}$ of R_{lbc}, $\hat{L}$, and D_p were estimated.

Experiments were performed on the test signals (artificial "transients" with $N = 512$ and the spoken word "greasy" with $N = 512$ segments uniformly sampled from the original signal with total length 5632) studied in [7]; these signals were kindly provided by S. Mallat. They were analyzed by WPDB and WPDP using wavelet packet libraries constructed from boundary-adjusted wavelets [2] of order 2–4 (with interior wavelet filters of length 4–8) and from circular-periodized wavelets [14] of orders 2–8 (with lengths 4–16) where all of these wavelets were derived from Daubechies' orthogonal least asymmetric family [4]. The test signals were analyzed with $L = 5$ levels. Monte Carlo experiments were performed with $T = 50$ trials on a Pentium class machine with MATLAB 4.2c. All tables of results are shown for the test signal "transients" analyzed with circular-periodized wavelets of order 8.

4.2 Experimental Results

Table 2 presents means $\hat{\mu}$ and coefficients of variations $\widehat{CV}$ of distortions D_2 of reconstructions for each of the named decompositions with identification numbers listed from 1 to 44. All standard deviations were smaller than the means with coefficients of variation on the order of 0.02–0.2; this experiment was therefore considered reliable for making inferences concerning the means to levels of significance determined by corresponding confidence intervals.

4.2.1 Coding by Parameterized Model versus Scalar Quantization For scalar quantized coding, a rate of $R_{PA} = 8$ was sufficient to reduce distortion to the approximate amount obtained with the rate of $R_{PA} = 48$. Thus, quantization of the retained M packets at rates of $R_{PA} \geq 8$ did not add significant distortion beyond that already produced by truncation of the packet list to $M < N$ packets. Moreover, reducing the rate further to $R_{PA} = 5$ increased the distortion by a relative amount on the order of only 10%. For parameterized model coding, a rate of $R_{CM} = 7$ was sufficient to reduce distortions for many (but not all) of the decompositions to the amounts produced by scalar quantized coding. For these approximately equal distortions D, estimates for bit rates were $\hat{\mu}(R_{SQ}) = 211 \times 5 = 1055$ and $\hat{\mu}(R_{PM}) = 211 + 7 \times 18 = 337$ using $\hat{\mu}(M) = 211$. Taking for example WPDB$(\mathcal{U}, \mathcal{A}^1)$ with a value of $\hat{\mu}(R_{lbc}) = 1227$ (*cf.* Table 3), then $\hat{\mu}(R_{SQ}) + \hat{\mu}(R_{lbc}) = 2282$ and $\hat{\mu}(R_{PM}) + \hat{\mu}(R_{lbc}) = 1564$. Thus, parameterized model coding provided 30% better compression for comparable distortion.

4.2.2 WPDB versus WPDP Table 3 presents results for distortions listed in rank order $1 \leq i \leq 44$ with identification numbers and $\hat{\mu} \pm \hat{\sigma}$ for compression at rate $R_{PA} = 48$. For distortions D_p measured by all relative ℓ^p-norms, WPDP$(\mathcal{O})$ rank first and second with no appreciable difference between WPDP(0) and WPDP(1). Since the mean numbers of flops and mean times in seconds were 1.34×10^7 and 172 for WPDP(0) and 6.08×10^7 and 379 for WPDP(1), there was no advantage gained by the use of WPDP(1) instead of WPDP(0). And since the range of computational cost for the WPDB$(\mathcal{S}, \mathcal{C})$ varied from 1.54×10^5 flops and 2.45 seconds to 2.50×10^5 flops and 8.42 seconds, all significantly less than the cost for WPDP$(\mathcal{O})$, there was strong incentive to focus attention on WPDB($\mathcal{S}$,$\mathcal{C}$) that approximate WPDP($\mathcal{O}$) in minimizing reconstruction error. Of these, WPDB($\mathcal{U}, \mathcal{A}^1$) and WPDB($\mathcal{U}, \mathcal{W}\ell^{.5}$) ranked highest for the D_1 and D_2 measures while WPDB($\mathcal{U}, \mathcal{W}\ell^{.5}$) and WPDB($\mathcal{U}, \mathcal{E}^{.5}$) ranked highest for the D_∞ measure.

4.2.3 WPDB($\mathcal{U}, \mathcal{C}^{\text{non}}$) versus WPDB($\mathcal{U}, \mathcal{C}^{\text{add}}$) The top-ranking bases WPDB($\mathcal{U}, \mathcal{C}^{\text{non}}$) selected with non-additive costs $\mathcal{C}^{\text{non}}$ ranked higher than the top-ranking bases WPDB($\mathcal{U}, \mathcal{C}^{\text{add}}$) selected with additive costs $\mathcal{C}^{\text{add}}$. In particular, the two highest-ranking non-additive WPDB($\mathcal{U}, \mathcal{A}^1$) and WPDB($\mathcal{U}, \mathcal{W}\ell^{.5}$) yielded smaller reconstruction error than did the two highest-ranking additive WPDB($\mathcal{U}, \mathcal{E}^{.5}$) and WPDB($\mathcal{U}, \mathcal{E}^1$). However, the differences were *not* statistically significant. Furthermore, WPDB($\mathcal{U}, \mathcal{H}_{\text{S}}(\mathbf{p}_{\text{D}})$) and WPDB($\mathcal{U}, \mathcal{F}$), which are non-additive and additive respectively, also yielded comparable distortion.[6] Thus, there was evidence for choosing between bottom-up bases selected by non-additive costs (such as the Shannon entropy on probabilities) and bottom-up bases selected by additive costs (such as the Coifman-Wickerhauser entropy on normalized energies) by considering computational issues rather than differences in distortion for fixed compression rates.

4.2.4 WPDB($\mathcal{D}, \mathcal{C}$) versus WPDB($\mathcal{U}, \mathcal{C}$) Thus, restricting attention now to top-down bases (WPDB with identification numbers $\{2, 4, 6, \ldots, 42\}$), WPDB($\mathcal{D}, \mathcal{A}^1$) and WPDB($\mathcal{D}, \mathcal{W}\ell^{.5}$) performed best overall with distortion less than or comparable to that of WPDB($\mathcal{U}, \mathcal{F}$) which is the conventional bottom-up Coifman-Wickerhauser best basis. Moreover, for WPDB($\mathcal{D}, \mathcal{A}^1$), $\hat{\mu}(\hat{L}) = 4.46$ was observed, while for WPDB($\mathcal{U}, \mathcal{F}$), $L = 5$ was required. This result provided evidence for choosing to use a top-down near-best basis rather than a bottom-up best basis. However, the savings with WPDB($\mathcal{D}, \mathcal{A}^1$) and its $\hat{L} = 4.46$ might not have been considered sufficient. Especially if reduction in computational cost was viewed as the major issue and differences in distortion viewed as insignificant within a certain range, then it would have been reasonable to choose one of the bases with a lower value of $\hat{L}$. In this situation, a good compromise choice would have been, say, WPDB($\mathcal{D}, \mathcal{N}^1_{.9}$) with $\hat{L} = 3.84$ which still would have maintained lower distortion than WPDB($\mathcal{U}, \mathcal{F}$).

5 Discussion

Although tables of results have been shown consistently throughout this report for the same example (the "transients" signal analyzed to level 5 with circularly-periodized wavelets of order 8), analogous results were obtained for all examples (both the "transients" and "greasy" signals with wavelets of various orders) that were investigated. These results can be summarized as follows: 1) Orthogonal matching pursuit decompositions did not provide any advantage over nonorthogonal matching pursuit decompositions despite the large increase in computational cost. 2) Nonorthogonal matching pursuit decompositions provided lower distortion than did wavelet packet decompositions, albeit at a much higher computational cost. 3) When minimizing distortion of the complete basis decomposition is viewed as the primary concern, bottom-up near-best bases with non-additive costs performed better than or comparably to bottom-up best

[6] Both the additive WPDB($\mathcal{U}, \mathcal{F}$) and the non-additive WPDB($\mathcal{U}, \mathcal{H}_{\text{CW}}$) yielded identical results.

Table 2: $\hat{\mu}(D_2)$ and $\widehat{CV}(D_2)$ distortions for reconstructions from decompositions compressed at rate R.

Decomposition	ID#	$R_{CM}=4$	$R_{CM}=7$	$R_{PA}=5$	$R_{PA}=8$	$R_{PA}=48$
WPDB($\mathcal{U},\mathcal{F}$)	1	.3890 .046	.2344 .046	.1956 .036	.1765 .042	.1762 .042
WPDB($\mathcal{D},\mathcal{F}$)	2	.4074 .027	.2662 .035	.2093 .030	.1910 .037	.1907 .037
WPDB($\mathcal{U},\mathcal{G}$)	3	.2540 .103	.1773 .054	.1761 .035	.1559 .046	.1556 .046
WPDB($\mathcal{D},\mathcal{G}$)	4	.2907 .124	.1887 .102	.1905 .073	.1710 .086	.1707 .087
WPDB($\mathcal{U},\mathcal{E}^{0.5}$)	5	.2655 .075	.1760 .045	.1733 .034	.1533 .045	.1530 .045
WPDB($\mathcal{D},\mathcal{E}^{0.5}$)	6	.2859 .073	.1804 .059	.1826 .057	.1628 .070	.1625 .070
WPDB($\mathcal{U},\mathcal{E}^{1.0}$)	7	.2674 .050	.1799 .043	.1750 .032	.1554 .042	.1551 .042
WPDB($\mathcal{D},\mathcal{E}^{1.0}$)	8	.2919 .087	.1861 .080	.1866 .041	.1674 .051	.1671 .051
WPDB($\mathcal{U},\mathcal{E}^{1.5}$)	9	.2751 .104	.1828 .068	.1782 .037	.1591 .047	.1588 .047
WPDB($\mathcal{D},\mathcal{E}^{1.5}$)	10	.4014 .068	.2596 .064	.2071 .030	.1886 .036	.1883 .037
WPDB($\mathcal{U},\mathcal{W}\ell^{0.5}$)	11	.2588 .086	.1744 .048	.1724 .035	.1524 .045	.1520 .045
WPDB($\mathcal{D},\mathcal{W}\ell^{0.5}$)	12	.2753 .080	.1759 .057	.1780 .050	.1581 .058	.1578 .059
WPDB($\mathcal{U},\mathcal{W}\ell^{1.0}$)	13	.2707 .045	.1807 .051	.1847 .032	.1660 .040	.1657 .040
WPDB($\mathcal{D},\mathcal{W}\ell^{1.0}$)	14	.2902 .103	.1916 .094	.1981 .079	.1798 .091	.1796 .092
WPDB($\mathcal{U},\mathcal{W}\ell^{1.5}$)	15	.3773 .108	.2384 .111	.2037 .046	.1851 .054	.1848 .054
WPDB($\mathcal{D},\mathcal{W}\ell^{1.5}$)	16	.3823 .115	.2603 .113	.2200 .092	.2024 .109	.2022 .109
WPDB($\mathcal{U},\mathcal{A}^{0.5}$)	17	.2484 .097	.1763 .050	.1765 .036	.1563 .047	.1560 .047
WPDB($\mathcal{D},\mathcal{A}^{0.5}$)	18	.2804 .109	.1842 .080	.1881 .070	.1680 .082	.1677 .083
WPDB($\mathcal{U},\mathcal{A}^{1.0}$)	19	.2537 .096	.1741 .048	.1721 .035	.1519 .044	.1516 .044
WPDB($\mathcal{D},\mathcal{A}^{1.0}$)	20	.2719 .098	.1752 .059	.1776 .056	.1576 .067	.1572 .067
WPDB($\mathcal{U},\mathcal{A}^{2.0}$)	21	.2715 .046	.1811 .043	.1783 .033	.1593 .042	.1590 .042
WPDB($\mathcal{D},\mathcal{A}^{2.0}$)	22	.2907 .037	.1889 .052	.1910 .040	.1723 .049	.1720 .050
WPDB($\mathcal{U},\mathcal{N}^{1.0}_{0.900}$)	23	.2534 .083	.1755 .054	.1759 .041	.1552 .051	.1549 .051
WPDB($\mathcal{D},\mathcal{N}^{1.0}_{0.900}$)	24	.2892 .123	.1882 .085	.1915 .077	.1715 .090	.1712 .091
WPDB($\mathcal{U},\mathcal{N}^{1.0}_{0.990}$)	25	.2769 .165	.1962 .134	.1917 .072	.1721 .087	.1718 .087
WPDB($\mathcal{D},\mathcal{N}^{1.0}_{0.990}$)	26	.3638 .122	.2448 .123	.2360 .113	.2193 .131	.2190 .132
WPDB($\mathcal{U},\mathcal{N}^{1.0}_{0.999}$)	27	.2900 .177	.2038 .139	.1977 .081	.1785 .095	.1782 .095
WPDB($\mathcal{D},\mathcal{N}^{1.0}_{0.999}$)	28	.3741 .096	.2602 .069	.2496 .094	.2340 .109	.2338 .110
WPDB($\mathcal{U},\mathcal{N}^{2.0}_{0.900}$)	29	.2670 .094	.1776 .041	.1749 .032	.1559 .042	.1556 .042
WPDB($\mathcal{D},\mathcal{N}^{2.0}_{0.900}$)	30	.2868 .080	.1827 .083	.1836 .084	.1646 .100	.1643 .100
WPDB($\mathcal{U},\mathcal{N}^{2.0}_{0.990}$)	31	.2520 .074	.1764 .054	.1733 .041	.1526 .052	.1523 .052
WPDB($\mathcal{D},\mathcal{N}^{2.0}_{0.990}$)	32	.2799 .107	.1847 .070	.1859 .071	.1656 .084	.1653 .084
WPDB($\mathcal{U},\mathcal{N}^{2.0}_{0.999}$)	33	.2553 .148	.1831 .094	.1825 .046	.1625 .057	.1622 .057
WPDB($\mathcal{D},\mathcal{N}^{2.0}_{0.999}$)	34	.3189 .162	.2122 .142	.2073 .106	.1886 .125	.1883 .125
WPDB($\mathcal{U},\mathcal{H}_{\mathrm{S}}(\mathbf{p}_{\mathrm{S}})$)	35	.3582 .152	.2253 .119	.1942 .054	.1754 .063	.1751 .063
WPDB($\mathcal{D},\mathcal{H}_{\mathrm{S}}(\mathbf{p}_{\mathrm{S}})$)	36	.4021 .052	.2651 .049	.2093 .032	.1908 .038	.1905 .038
WPDB($\mathcal{U},\mathcal{H}_{\mathrm{S}}(\mathbf{p}_{\mathrm{D}})$)	37	.2774 .048	.1863 .048	.1857 .037	.1669 .045	.1666 .045
WPDB($\mathcal{D},\mathcal{H}_{\mathrm{S}}(\mathbf{p}_{\mathrm{D}})$)	38	.3332 .096	.2553 .129	.2566 .126	.2421 .143	.2419 .143
WPDB($\mathcal{U},\mathcal{H}_{\mathrm{S}}(\mathbf{p}_{\mathrm{TS}})$)	39	.3651 .181	.2410 .151	.2059 .045	.1874 .051	.1871 .051
WPDB($\mathcal{D},\mathcal{H}_{\mathrm{S}}(\mathbf{p}_{\mathrm{TS}})$)	40	.3398 .177	.2258 .166	.2071 .069	.1888 .080	.1885 .081
WPDB($\mathcal{U},\mathcal{H}_{\mathrm{CW}}$)	41	.3890 .046	.2344 .046	.1956 .036	.1765 .042	.1762 .042
WPDB($\mathcal{D},\mathcal{H}_{\mathrm{CW}}$)	42	.4074 .027	.2662 .035	.2093 .030	.1910 .037	.1907 .037
WPDP(0)	43	.3661 .055	.1920 .078	.1178 .029	.0707 .006	.0699 .002
WPDP(1)	44	.3581 .099	.1638 .127	.1095 .041	.0718 .067	.0695 .004

Table 3: Rank-ordered $\hat{\mu} \pm \hat{\sigma}$ statistics for reconstructions from decompositions with indicated ID number.

Rank	D_1	D_2	D_∞	R_{lbc}	$\hat{L}$
1	44) .0762±.0011	44) .0695±.0003	43) .0360±.0015	7) 1212± 29	38) 1.595±1.206
2	43) .0786±.0009	43) .0699±.0002	44) .0384±.0021	5) 1217± 34	28) 1.980±0.885
3	19) .1692±.0080	19) .1516±.0067	11) .0940±.0106	9) 1221± 31	26) 2.512±0.906
4	11) .1696±.0080	11) .1520±.0068	5) .0940±.0091	21) 1224± 28	16) 3.148±0.769
5	31) .1700±.0089	31) .1523±.0079	29) .0945±.0085	19) 1227± 34	34) 3.418±0.772
6	5) .1707±.0082	5) .1530±.0069	7) .0951±.0109	3) 1228± 34	40) 3.535±0.583
7	23) .1724±.0094	23) .1549±.0079	19) .0951±.0101	11) 1231± 28	39) 3.558±0.263
8	7) .1729±.0076	7) .1551±.0065	9) .0961±.0096	29) 1232± 29	2) 3.592±0.195
9	29) .1731±.0080	29) .1556±.0066	12) .0974±.0100	17) 1233± 40	42) 3.592±0.195
10	3) .1734±.0080	3) .1556±.0072	31) .0975±.0131	13) 1240± 33	36) 3.675±0.173
11	17) .1737±.0080	17) .1560±.0074	21) .0984±.0118	31) 1252± 33	10) 3.680±0.201
12	20) .1747±.0116	20) .1572±.0106	17) .0985±.0121	33) 1264± 49	24) 3.838±0.547
13	12) .1753±.0105	12) .1578±.0093	3) .0989±.0116	23) 1265± 40	14) 3.882±0.656
14	9) .1772±.0086	9) .1588±.0075	20) .0993±.0115	20) 1287± 68	22) 4.032±0.292
15	21) .1773±.0081	21) .1590±.0067	23) .0994±.0107	12) 1290± 69	32) 4.052±0.486
16	6) .1794±.0127	33) .1622±.0092	30) .1015±.0133	1) 1309± 28	4) 4.070±0.583
17	33) .1805±.0099	6) .1625±.0114	13) .1020±.0113	41) 1309± 28	15) 4.075±0.287
18	30) .1810±.0125	30) .1643±.0165	32) .1023±.0135	25) 1315± 85	18) 4.120±0.569
19	32) .1823±.0145	32) .1653±.0140	33) .1033±.0104	6) 1315± 86	27) 4.132±0.491
20	13) .1842±.0079	13) .1657±.0066	37) .1036±.0094	37) 1320± 76	35) 4.190±0.295
21	8) .1848±.0098	37) .1666±.0075	6) .1040±.0105	30) 1321±108	37) 4.215±0.264
22	37) .1851±.0084	8) .1671±.0085	24) .1063±.0147	35) 1321± 61	8) 4.228±0.335
23	18) .1851±.0139	18) .1677±.0139	18) .1068±.0151	8) 1338± 83	30) 4.228±0.602
24	4) .1880±.0152	4) .1707±.0148	8) .1072±.0122	15) 1349± 65	1) 4.312±0.102
25	24) .1886±.0157	24) .1712±.0155	14) .1080±.0149	27) 1356± 96	41) 4.312±0.102
26	22) .1896±.0094	25) .1718±.0150	25) .1086±.0162	18) 1360±122	6) 4.320±0.393
27	25) .1901±.0158	22) .1720±.0085	35) .1088±.0113	32) 1367±103	25) 4.360±0.415
28	35) .1937±.0120	35) .1751±.0111	22) .1089±.0099	4) 1368±121	12) 4.435±0.329
29	1) .1956±.0077	1) .1762±.0073	1) .1091±.0147	22) 1372± 73	20) 4.462±0.339
30	41) .1956±.0077	41) .1762±.0073	41) .1091±.0147	14) 1388±133	23) 4.600±0.156
31	27) .1969±.0180	27) .1782±.0170	4) .1111±.0162	24) 1408±116	33) 4.610±0.200
32	14) .1974±.0134	14) .1796±.0165	27) .1126±.0165	36) 1448± 38	31) 4.640±0.142
33	34) .2033±.0175	15) .1848±.0100	15) .1164±.0156	10) 1448± 41	13) 4.650±0.162
34	15) .2047±.0104	39) .1871±.0096	38) .1202±.0079	2) 1458± 37	29) 4.678±0.145
35	40) .2054±.0110	10) .1883±.0069	34) .1202±.0174	42) 1458± 37	21) 4.680±0.105
36	39) .2060±.0106	34) .1883±.0236	40) .1202±.0147	39) 1459± 52	9) 4.698±0.113
37	10) .2066±.0082	40) .1885±.0152	39) .1212±.0140	40) 1468±117	11) 4.708±0.153
38	2) .2093±.0085	36) .1905±.0072	10) .1216±.0142	34) 1488±142	17) 4.745±0.151
39	42) .2093±.0085	2) .1907±.0071	36) .1245±.0130	16) 1521±136	19) 4.750±0.118
40	36) .2094±.0086	42) .1907±.0071	2) .1245±.0156	26) 1640±163	5) 4.755±0.126
41	16) .2168±.0131	16) .2022±.0220	42) .1245±.0156	28) 1733±166	3) 4.758±0.142
42	38) .2233±.0176	26) .2190±.0288	16) .1273±.0169	38) 1809±222	7) 4.762±0.095
43	26) .2273±.0218	28) .2338±.0256	26) .1329±.0226	44) 1856± 40	43) not applic.
44	28) .2322±.0131	38) .2419±.0346	28) .1338±.0190	43) 1865± 39	44) not applic.

bases with additive costs, although differences were not significant. 4) When minimizing computational cost (memory, flops, and time) is viewed as the primary concern, top-down near-best bases performed better than bottom-up best bases. 5) For comparable distortion, parameterized model coding provided better compression than scalar quantization coding.

These results were obtained with an experimental paradigm intended to provide not only a valid comparison between the different complete basis decompositions but also between them and the matching pursuit decompositions. Thus, it was necessary to truncate packet lists to a fixed number of packets and then to quantize uniformly at a fixed bit rate for the remaining packets. This simple paradigm was intended to reveal some intuition about how various information cost functions might perform in selecting bases in wavelet packet decompositions relative to the choice function of always selecting the current largest absolute value packet in each iterative step of the matching pursuit decompositions. However, a new coding scheme incorporating parameterized model regression was also introduced. This new coding scheme proved quite successful and will be explored in future research. In particular, the nonlinear models mentioned in Section 4 should be investigated. In addition, hybrid schemes could be considered. For example, using scalar quantization on a small number of the largest absolute-value packets together with parameterized modeling of the remaining smaller absolute-value packets might yield better results.

The performance of the different complete basis decompositions in general as well as in particular in terms of their explicit ordered rankings relative to each other (rather than to the matching pursuit decompositions) undoubtedly depends upon the application including the signal and filter classes as well as the compression and coding scheme investigated. Therefore, further research should extend these Monte Carlo experiments to investigate the statistical performance of these methods for the complete basis list of N wavelet packets (*ie.,* not truncated to $M << N$ with M determined by matching pursuit decompositions). This approach would eliminate the need to code each packet's indices in addition to its amplitude and would therefore enable a full analysis of compression and distortion in the context of more sophisticated bit-allocation methods such as that of Ramchandran and Vetterli [9]. These methods could also be extended from 1-D signals to 2-D images and other higher-dimensional signals where the advantages of top-down bases over bottom-up bases become even more important because of the tremendous savings in computational requirements for memory, flops, and time. Also, recalling that best bases with additive costs require the use of orthogonal wavelets, near-best bases with non-additive costs permit the use of biorthogonal wavelets thereby gaining advantages such as the linear phase response considered important in image processing applications.

As a final general comment and conclusion, the relative statistical performance of the different search methods and decision criteria for selecting complete basis decompositions in wavelet packet transforms did not appear to yield significant differences in compression and distortion. Thus, in this sense, the top-down tree searches with additive or non-additive information cost functions were just as good as bottom-up tree searches with additive information cost functions. However, top-down tree searches required significantly less computation than bottom-up tree searches. If given these results and the desire to minimize computational costs, then top-down tree searches should be performed instead of bottom-up tree searches when selecting wavelet packet decompositions, except possibly when characterizing unknown signal classes.

References

[1] Joseph Berkson. Relative precision of minimum chi-square and maximum likelihood estimates of regression coefficients. In Jerzy Neyman, editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 471–479. University of California Press, 1951.

[2] Albert Cohen, Ingrid Daubechies, and Pierre Vial. Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, 1(1):54–81, December 1993.

[3] Ronald R. Coifman and Mladen Victor Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38(2):713–718, March 1992.

[4] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets: Ii. variations on a theme. *SIAM Journal on Mathematical Analysis*, 24(2):499–519, March 1993.

[5] David L. Donoho. Unconditional bases are optimal bases for data compression and for statistical estimation. *Applied and Computational Harmonic Analysis*, 1(1):100–115, December 1993.

[6] A. I. Khinchin. *Mathematical Foundations of Information Theory.* Dover Publications, Inc., New York, 1957.

[7] Stephane G. Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, December 1993.

[8] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with application to wavelet decomposition. In *Proceedings of the 27th Annual Asilomar Conference on Signals Systems and Computers*, pages 40–44, Pacific Grove, CA, November 1993.

[9] Kannan Ramchandran and Martin Vetterli. Best wavelet packet bases in a rate-distortion sense. *IEEE Transactions on Image Processing*, 2(2):160–175, April 1993.

[10] David A. Ratkowsky. *Handbook of Nonlinear Regression Models*, volume 107 of *Statistics: Textbooks and Monographs.* Marcel Dekker, Inc., New York, NY, 1990.

[11] David W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization.* Wiley Series in Applied Probability and Statistics. John Wiley & Sons, Inc., New York, 1992.

[12] Carl Taswell. Near-best basis selection algorithms with non-additive information cost functions. In Moeness G. Amin, editor, *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, pages 13–16, Philadelphia, PA, October 1994. IEEE Press 94TH8007.

[13] Carl Taswell. WavBox 4: A software toolbox for wavelet transforms and adaptive wavelet packet decompositions. In Anestis Antoniadis and Georges Oppenheim, editors, *Wavelets and Statistics*, Lecture Notes in Statistics. Springer Verlag, 1995. Proceedings of the Villard de Lans Conference November 1994.

[14] Carl Taswell and Kevin C. McGill. Algorithm 735: Wavelet transform algorithms for finite-duration discrete-time signals. *ACM Transactions on Mathematical Software*, 20(3):398–412, September 1994.

[15] Mladen Victor Wickerhauser. INRIA lectures on wavelet packet algorithms. Technical report, INRIA, Roquencourt, France, 1991. minicourse lecture notes.