WAVELET-DOMAIN COMPRESSIVE SIGNAL RECONSTRUCTION USING A HIDDEN MARKOV TREE MODEL

Marco F. Duarte, Michael B. Wakin,* and Richard G. Baraniuk†

† Department of Electrical and Computer Engineering, Rice University
* Department of Electrical Engineering and Computer Science, The University of Michigan at Ann Arbor

ABSTRACT

Compressive sensing aims to recover a sparse or compressible signal from a small set of projections onto random vectors; conventional solutions involve linear programming or greedy algorithms that can be computationally expensive. Moreover, these recovery techniques are generic and assume no particular structure in the signal aside from sparsity. In this paper, we propose a new algorithm that enables fast recovery of piecewise smooth signals, a large and useful class of signals whose sparse wavelet expansions feature a distinct “connected tree” structure. Our algorithm fuses recent results on iterative reweighted ℓ_1-norm minimization with the wavelet Hidden Markov Tree model. The resulting optimization-based solver outperforms the standard compressive recovery algorithms as well as previously proposed wavelet-based recovery algorithms. As a bonus, the algorithm reduces the number of measurements necessary to achieve low-distortion reconstruction.

Index Terms— Compressive sensing, wavelet transforms, data compression, signal reconstruction, Hidden Markov Models.

1. INTRODUCTION

Compressive sensing (CS) is a new approach to simultaneous sensing and compression that enables a potentially large reduction in the sampling and computation costs at a sensor for signals having a sparse representation in some basis. CS builds on the work of Candès, Romberg, and Tao [1] and Donoho [2], who showed that a signal having a sparse representation in one basis can be reconstructed from a small set of projections onto a second, measurement basis that is incoherent with the first. Random projections play a central role as a universal measurement basis in the sense that they are incoherent with any fixed basis with high probability. The CS measurement process is nonadaptive, and the reconstruction process is nonlinear. A variety of reconstruction algorithms have been proposed [1–4].

These recovery algorithms are generic in the sense that they do not exploit any particular structure in the signal besides its sparsity in some basis. However, for some signals we have additional a priori information that we should be able to exploit for improving performance. For example, the piecewise smooth signals [5] that feature prominently in a wide range of applications are not only sparse in the wavelet domain, but also sport wavelet coefficients that cluster around a connected subtree in the wavelet domain [6, 7]. The goal of this paper is to design a new CS signal recovery algorithm that takes full advantage of this structure.

Previous work in this vein has developed tailored combinations of measurement/reconstruction algorithms [8]. Additionally, modified greedy algorithms have been proposed [9, 10] to exploit the parent-child wavelet coefficient relationships of piecewise smooth signals. To date, these algorithms suffer from mixed, signal-dependent performance. Furthermore, they do not exploit additional structures, such as the magnitude decay of the wavelet coefficients across scale.

In this paper, we endow a recently introduced iterative reweighted ℓ_1-norm minimization algorithm [11] with a weighting scheme based on the Hidden Markov Tree model (HMT) [6, 7] to enforce the additional tree-structured dependencies exhibited by piecewise smooth signals. The combination of these two approaches provides better reconstruction performance over standard algorithms; this improvement is significant when a limited number of measurements is available. The resulting scheme consists of a initial training stage, in which the model parameters are estimated, followed by alternating iterations of a reconstruction stage based on a linear and a weight update stage based on the Viterbi algorithm.

This paper is organized as follows. Section 2 provides the necessary background on CS, and Section 3 overviews work on CS for wavelet-sparse signals. Section 4 reviews the HMT and its properties, and Section 5 introduces our new algorithm. Section 6 contains experimental results, and Section 7 concludes.

2. COMPRESSIVE SENSING BACKGROUND

Let \( x \in \mathbb{R}^N \) be a signal and let the matrix \( \Psi := [\psi_1, \psi_2, \ldots, \psi_N] \) be a basis for \( \mathbb{R}^N \). We say that \( x \) is \( K \)-sparse if it can be expressed as a linear combination of \( K \) vectors from \( \Psi \); that is, \( x = \sum_{i=1}^{K} \theta_i \psi_i \). A basis that provides a sparse representation for a class of signals effectively captures the structure inherent in the class; for example, the Fourier basis sparsifies smooth signals, and wavelet bases sparsify piecewise smooth signals.

2.1. Incoherent measurements

Consider a signal \( x \) that is \( K \)-sparse in \( \Psi \). Consider also an \( M \times N \) measurement matrix \( \Phi \), \( M \ll N \), where the rows of \( \Phi \) are incoherent with the columns of \( \Psi \). For example, let \( \Phi \) contain i.i.d. Gaussian entries; with high probability, such a matrix is incoherent with any fixed \( \Psi \) (universality). Compute the measurements \( y = \Phi x \) and note that \( y \in \mathbb{R}^M \) with \( M \ll N \). The CS theory states that there exists an overmeasuring factor \( c > 1 \) such that only \( M := cK \) incoherent measurements \( y \) are required to reconstruct \( x \) with high...
probability \([1,2]\). That is, just \(cK\) incoherent measurements encode all of the salient information about the \(K\)-sparse signal \(x\).

### 2.2. Reconstruction from incoherent projections

The overmeasuring factor \(c\) required depends on the (nonlinear) reconstruction algorithm. Under the sparsity assumption, one can search for the sparsest signal that agrees with the obtained measurements; by using the \(\ell_0\)-norm \(|\theta|_0 = \#\{n: \theta_n \neq 0\}\), the reconstruction algorithm can be expressed as

\[
\hat{\theta} = \arg \min_{\theta} \|\theta\|_0 \quad \text{subject to} \quad \Phi \Psi \theta = y.
\]

While this algorithm demands the smallest possible overmeasuring factor \((c = 2)\), its computational complexity renders it unfeasible. Most of the existing literature on CS \([1,2,4]\) has concentrated on optimization-based methods for signal recovery, in particular \(\ell_1\)-norm minimization. The \(\ell_1\)-norm approach seeks a set of sparse coefficients \(\hat{\theta}\) by solving the linear program

\[
\hat{\theta} = \arg \min_{\theta} \|\theta\|_1 \quad \text{subject to} \quad \Phi \Psi \theta = y;
\]

the reconstruction of sparse signals via \(\ell_1\)-norm minimization is typically exact, provided that \(c = O(\log(N/K))\). Other algorithms, including greedy algorithms \([3]\), have also been proposed for CS reconstruction and require similar oversampling factors. However, all of these algorithms are generic in the sense that, aside from sparsity, they assume no particular structure within the signal coefficients.

### 2.3. Iterative reweighted \(\ell_1\)-norm minimization

When the complexity of the signal is measured using the \(\ell_1\)-norm, individual signal coefficients are penalized according to their magnitude; in contrast, when the \(\ell_0\)-norm is used to measure the signal complexity, the penalty for a nonzero coefficient is independent of its magnitude. The effect of this disparity is reflected in the increase of the oversampling factor between the two algorithms.

A small variation to the \(\ell_1\)-norm penalty function has been suggested to rectify the imbalance between the \(\ell_0\)-norm and \(\ell_1\)-norm penalty functions \([11]\). The basic goal is to minimize a weighted \(\ell_1\)-norm penalty function \(\|W\theta\|_1\), where \(W\) is a diagonal "weighting" matrix with entries \(W_{n,n}\) approximately proportional to \(1/|\theta_n|\). This creates a penalty function that achieves higher magnitude-independence. Since the true values of \(\theta_n\) are unknown (indeed they are sought), however, an iterative reweighted \(\ell_1\)-norm minimization (IRWL1) algorithm is suggested.

The algorithm starts with the solution to the unweighted \(\ell_1\)-norm minimization algorithm \((1)\), which we name \(\hat{\theta}^{(0)}\). The algorithm then proceeds iteratively: on iteration \(i > 0\), it solves the optimization problem

\[
\hat{\theta}^{(i)} = \arg \min_{\theta} \|W^{(i)}\theta\|_1 \quad \text{subject to} \quad \Phi \Psi \theta = y,
\]

where \(W^{(i)}\) is a diagonal reweighting matrix with entries

\[
W_{n,n}^{(i)} = \left(\frac{|\hat{\theta}_n^{(i-1)}|}{\hat{\theta}_n^{(i-1)}}\right)^{-1},
\]

and \(\epsilon\) is a small regularization constant. The algorithm can be terminated when the change between consecutive solutions is smaller than an established threshold or after a fixed number of iterations. Each iteration of this algorithm can be posed as a linear program, for which there exist efficient solvers.

### 3. CS FOR WAVELET-SPARSE SIGNALS

A widely used sparse representation in signal and image processing is the wavelet transform. Since piecewise polynomial signals have sparse wavelet expansions \([5]\) and since many real-world signals describe punctuated, piecewise smooth phenomena, it follows that many real-world signals have sparse or compressible wavelet expansions. However, the significant wavelet coefficients in general do not occur in arbitrary positions. Instead they exhibit a characteristic signal-dependent structure.

Without loss of generality, we focus on 1D signals, although similar arguments apply for 2D and higher dimensional data in the wavelet or curvelet domains. In a typical 1D wavelet transform, each coefficient at scale \(j \in \{1, \ldots, J := \log_2(N)\}\) describes a portion of the signal of size \(O(2^{j-1})\). With \(2^{j-1}\) such coefficients at each scale, a binary tree provides a natural organization for the coefficients. Each coefficient at scale \(j < \log_2(N)\) has 2 children at scale \(j + 1\), and each coefficient at scale \(j > 1\) has one parent at scale \(j - 1\).

Due to the analysis properties of wavelets, coefficient values tend to persist through scale. A large wavelet coefficient (in magnitude) generally indicates the presence of a singularity inside its support; a small wavelet coefficient generally indicates a smooth region. Thanks to the nesting of child wavelets inside their parents, edges in general manifest themselves in the wavelet domain as chains of large coefficients propagating across scales in the wavelet tree; we call this phenomenon the persistence property. Additionally, wavelet coefficients also have exponentially decaying magnitudes at finer scales \([5]\). This causes the significant wavelet coefficients of piecewise smooth signals to concentrate within a connected subtree of the wavelet binary tree.

This connected subtree structure was exploited by previous CS reconstruction algorithms known as Tree Matching Pursuit (TMP) and Tree Orthogonal Matching Pursuit (TOMP) \([9,10]\). TMP and TOMP are modifications to the standard greedy algorithms matching pursuit and orthogonal matching pursuit, with the proviso that each selection made by the greedy algorithm build upon a connected tree. Such techniques enable faster algorithms due to the restriction in the greedy search and lower reconstruction distortion due to the inherent regularization.

However, for real-world piecewise smooth signals, the nonzero coefficients generally do not form a perfectly connected subtree. The reasons for this are twofold. First, since wavelets are band-pass functions, wavelet coefficients oscillate between positive and negative values around singularities. Second, due to the linearity of the wavelet transform, two or more singularities in the signal may cause destructive interference among coarse scale wavelet coefficients; that is, the persistence of the wavelets across scales is weaker at coarser scales. Either of these factors may cause the wavelet coefficient corresponding to a discontinuity to be small yet have large children, yielding a non-connected set of meaningful wavelet coefficients. TMP and TOMP used heuristic rules to ameliorate the effect of this phenomenon. However, this considerably increases the computational complexity, and the success of such heuristics varies markedly between different signals in the proposed class.

In summary, we have identified several properties of wavelet expansions:

- large/small values of wavelet coefficients generally persist across the scales of the wavelet tree;
- persistence becomes stronger as we move to finer scales; and
the magnitude of the wavelet coefficients decreases exponentially as we move to finer scales.

We also note that the sparsity of the wavelet transform causes the coefficients to have a peaky, non-Gaussian distribution.

4. HIDDEN MARKOV TREE MODELS

The properties identified in Section 3 induce a joint statistical structure among the wavelet coefficients that is far stronger than simple sparsity or the simple connectivity models used in the TMP and TOMP algorithms. HMT [6] offers one modeling framework that succinctly and accurately captures this joint structure. HMT modeling has been used successfully to improve performance of denoising, classification, and segmentation algorithms for wavelet-sparse signals.

The HMT models the probability density function of each wavelet as a Gaussian mixture density with a hidden binary state that determines whether the coefficient is large or small. The persistence across scale is captured by a tree-based Markov model that correlates the states of parent and children coefficients. The following properties are captured by the HMT.

Non-Gaussianity: Sparse coefficients can be modeled probabilistically using a mixture of Gaussians: one component features a large variance that models large nonzero coefficients and receives a small weight (to encourage few such coefficients), while a second component features a small variance that models small and zero-valued coefficients and receives a large weight. We distinguish these two components by associating to each wavelet coefficient \( \theta_n \) an unobserved hidden state \( S_n \in \{S, L\} \); the value of \( S_n \) determines which of the two components of the mixture model is used to generate \( \theta_n \). Thus we have

\[
\begin{align*}
    f(\theta_n | S_n = S) &= N(0, \sigma^2_{S,n}), \\
    f(\theta_n | S_n = L) &= N(0, \sigma^2_{L,n}),
\end{align*}
\]

with \( \sigma^2_{S,n} > \sigma^2_{L,n} \). To generate the mixture, we apply a probability distribution to the available states: \( p(S_n = S) = p^n_S \) and \( p(S_n = L) = p^n_L \), with \( p^n_S + p^n_L = 1 \).

Persistence: The perpetuation of large and small coefficients from parent to child is well-modeled by a Markov model that links the hidden state \( S_n \) to that of its parent \( P(n) \). The Markov model is then completely determined by the set of state transition matrices for the different coefficients \( \theta_n \) at wavelet scales \( 1 < j \leq J \):

\[
A_n = \begin{bmatrix} p^n_{L-S} & p^n_{L-L} \\ p^n_{S-S} & p^n_{S-L} \end{bmatrix}.
\]

The persistence property implies that the values of \( p^n_{L-S} \) and \( p^n_{S-S} \) are significantly larger than their complements. If we are provided the hidden state probabilities for the wavelet coefficients in the coarsest scale \( p^1_L \) and \( p^1_S \), then the probability distribution for any hidden state can be obtained recursively:

\[
p(S_n = L) = p^n_L p^n_{L-S} + p^n_S p^n_{S-L}.
\]

As posed, the HMT parameters include the probabilities for the hidden state \( \{p^1_S, p^1_L\} \), the state transition matrices \( A_n \), and Gaussian distribution variances \( \{\sigma^2_{S,n}, \sigma^2_{L,n}\} \) for each of the wavelet coefficients \( \theta_n \). To simplify the model, the coefficient-dependent parameters are made equal for all coefficients within a scale; that is, the new model has parameters \( A_j \) for \( 1 < j \leq J \) and \( \{\sigma^2_{L,j}, \sigma^2_{S,j}\} \) for \( 1 \leq j \leq J \).

Magnitude decay: To enforce the decay of the coefficient magnitudes across scale, the variances \( \sigma^2_{L,j} \) and \( \sigma^2_{S,j} \) are modeled so that they decay exponentially as the scale becomes finer [7]:

\[
\begin{align*}
    \sigma^2_{L,j} &= C_{L,j} 2^{-j \cdot \alpha_L}, \\
    \sigma^2_{S,j} &= C_{S,j} 2^{-j \cdot \alpha_S}.
\end{align*}
\]

Since the wavelet coefficients that correspond to signal discontinuities decay slower than those representing smooth regions, the model sets \( \alpha_S > \alpha_L \).

Scale-dependent persistence: To capture the weaker persistence present in the coarsest scales, the values of the state transition matrices \( A_j \) follow a model that strengthens the persistence at finer scales [7]. Additionally, the model must reflect that in general, any large parent generally implies only one large child (that which is aligned with the discontinuity). This implies that the probability that \( S_n = L \), given that \( S_{p(n)} = L \), should be roughly 1/2. HMT accounts for both factors by setting

\[
\begin{align*}
    p^j_{L-L} &= \frac{1}{2} + C_{LL} 2^{-\gamma_L j}, \\
    p^j_{L-S} &= \frac{1}{2} - C_{LL} 2^{-\gamma_L j}, \\
    p^j_{S-S} &= 1 - C_{SS} 2^{-\gamma_S j}, \quad \text{and} \quad p^j_{S-L} = C_{SS} 2^{-\gamma_S j}.
\end{align*}
\]

Estimation: We can obtain estimates of all parameters

\[
\Theta = \{p_S, p_L, \alpha_S, \alpha_L, C_{LL}, C_{SS}, \gamma_S, C_{LL}, C_{SS}\}
\]

for a set of coefficients \( \theta \) using maximum likelihood estimation:

\[
\Theta_{M.L.} = \arg \max_{\theta} f(\theta | \Theta).
\]

The expectation-maximization (EM) algorithm in [6] efficiently performs this estimation. Similarly, one can obtain the state probabilities \( p(S_n = S | \theta, \Theta) \) using the Viterbi algorithm; the state probabilities for a given coefficient will be dependent on the states and coefficient values of all of its predecessors in the wavelet tree.

5. HMT-BASED WEIGHTS FOR IRWL1

The IRWL1 algorithm described in Sec. 2.3 provides an opportunity to implement flexible signal penalizations while retaining the favorable computational complexity of \( l_1 \)-norm minimizations.

We now pose a new weight rule for the IRWL1 algorithm that integrates the HMT model to enforce the wavelet coefficient structure during CS reconstruction. Our weighting scheme, dubbed HMT+IRWL1, employs the following weighting scheme:

\[
W^{(i)}_{m,n} = \left( p \left( S_n = L | \tilde{\theta}^{(i-1)}, \Theta \right) + \delta \right)^{-q}.
\]

In words, for each wavelet coefficient in the current estimate we obtain the probability that the coefficient’s hidden state is large; in the next iteration, we apply to that coefficient a weight that is inversely proportional to that probability. The parameter \( \delta \) is a regularization parameter for cases where \( p(S_n = L | \tilde{\theta}^{(i-1)}) \) is very small, and the exponent \( q \) is a parameter that regulates the strength of the penalization for small coefficients. The goal of this weighting scheme is to penalize coefficients with large magnitudes that have low likelihood of being generated by a wavelet sparse signal; these coefficients are often the largest contributors to the reconstruction error.

The first step of HMT+IRWL1 consists of an initial training stage in which an EM algorithm solves (3) to estimate the values of the parameters for a representative signal; additionally, the solution \( \tilde{\theta}^{(0)} \) for the standard formulation (1) is obtained. Subsequently, we proceed iteratively with two alternating steps: a weight update step in which the Viterbi algorithm for state probability calculations is executed for the previous solution \( \tilde{\theta}^{(i-1)} \), and a reconstruction
The HMT+IRWL1 scheme reduces the artifacts of the \( \ell_1 \)-norm minimization reconstruction over fewer iterations than the standard iterative reweighted \( \ell_1 \) minimization algorithm. Our scheme can be used as well in the extensions of the iterative reweighted \( \ell_1 \)-norm minimization scheme, such as reconstruction from noisy measurements.

8. REFERENCES