Fast Reconstruction from Random Incoherent Projections

Marco F. Duarte

Dept. of Electrical and Computer Engineering, Rice University, Houston, TX 77005

Abstract

The Compressed Sensing paradigm consists of recovering signals that can be sparsely represented in a given basis from a small set of projections into random vectors; the problem is typically solved using an adapted Basis Pursuit algorithm. We show that the recovery of the signal is equivalent to determining the sparsest representation of the measurement vector using the dictionary obtained by applying the projections to the basis elements, and therefore more efficient algorithms can be used during recovery; we explore the Matching Pursuit and Orthogonal Matching Pursuit algorithms. We also design an algorithm that allows for even faster recovery for piecewise smooth signals: the algorithm exploits the tree structure of the sparse coefficients of the signal in a wavelet representation to select subsets of coefficients to estimate at each iteration. We define a class of signals for which such a Tree Matching Pursuit algorithm performs successful recovery and present variations of the algorithms for different classes of signals. Other applications of TMP include effective denoising, fast approximation in overcomplete wavelet bases, and distributed compression.

I. Introduction

The present trend in evolution and proliferation of sensing devices has caused an explosion in the amount of information available to scientists in many disciplines and to consumers in general. Witness the success of digital cameras, digital music and video recorders, etc. This phenomenon has promoted research in the fields of compression and coding, which allow for compact representations, portability, and fast transmission of the gathered information. In many cases, the data is compressed through a *transform* that yields a sparse representation, which is then encoded and stored. However, the power consumption due to this sensing and compression process is high and currently limits the range of applications for many classes of sensing devices in emerging areas such as sensor networks.

The recently introduced paradigm of *Compressed Sensing* (CS) [4, 18] enables a reduction in the communication and computation costs at the sensor. This theory, inspired by the groundbreaking work of Candès, Tao and Romberg [3], shows that a small number of random projections of a compressible signal contains enough information for exact reconstruction. The key idea in CS is that many signals are compressible, i.e. *sparse* relative to orthonormal bases or tight frames [20]. In many settings we are aware of a basis or tight frame that models a class of signals well, in the sense that “most” signals can be represented sparsely. The existence of such bases for many classes of real world signals has enabled advances in compression; by specifying only the largest coefficients in the transform domain, a high-quality representation of the signal can be reconstructed. However, the determination of this subset of coefficients relies on the availability of the entire signal. In contrast, CS [4, 18] has shown that for a N-sample signal that is K-sparse — meaning that the sparse representation has K nonzero coefficients — the availability of $O(K\log(N))$ [7] projections of the signal along random directions suffices to obtain a reconstruction of the signal of the same quality with high probability.

The implications of Compressed Sensing are very promising. Instead of sampling the signal N times, only CK random measurements suffice, where K can be orders of magnitude less than N and C is an *oversampling factor*. Therefore, a node can transmit far fewer measurements to a processing center, which can reconstruct the signal and then process it in any manner. In low-power applications, such as surveillance

This is a Final Report for ELEC599 - First Year Graduate Student Project. Advisors: Richard Baraniuk and Hyeokho Choi. Committee: Richard Baraniuk, Hyeokho Choi and Yehia Massoud

and sensor networks, CS is advantageous due to the automatic compression and encryption of the sensed data and the universal quality of the encoding methodology. Its main drawback, however, is the computational complexity of the signal reconstruction, consisting of an ℓ_0 norm minimization which has been proven to be NP-hard [13]. An important result has shown the equivalence between the ℓ_0 norm minimization and the ℓ_1 norm minimization for cases of interest [17]; this new problem can be recast as a linear program and solved using interior point methods. Nonetheless, the computational complexity of this simpler method still remains a challenge for signals of normal length — for example, current digital cameras produce images of sizes in the millions of pixels, which renders reconstruction unfeasible with current computational power. Clearly, it is very important to find fast reconstruction methods if the CS framework is to be deployed in real-world applications.

In this paper we tackle this problem by changing the framework in which the reconstruction is performed. We show that the reconstruction problem can be recast as a search for the *sparsest representation of the measurements in an overcomplete basis* and review the existing methods applied to this second problem. We then propose a new *Tree Matching Pursuit* algorithm that allows for fast reconstruction of piecewise smooth signals, and define a class of signals for which this algorithm will successfully perform accurate reconstruction of signals compressed through CS. We mention some properties of this TMP algorithm and propose extensions for known piecewise smooth signal processing applications such as denoising. During the completion of this project, we became aware of new work by Joel Tropp and Anna Gilbert [26] that also explores the use of Matching Pursuits for CS reconstruction.

This paper is organized as follows. Section II describes the necessary background in Compressed Sensing Theory. Section III describes the formulation of CS recovery as a sparse approximation. Section IV describes our Tree Matching Pursuit algorithms, Section V describes some adaptations to improve its performance, and Section VI describes some extensions to the algorithm. Section VII lists some related work, and Section VIII offers conclusions.

II. Compressed Sensing Background

Assume that we acquire a N-sample signal x for which a basis or tight frame $\Psi = [\psi_1, \ldots, \psi_N]$ provides a K-sparse representation

$$x = \sum_{i=1}^{K} \alpha_{n_i} \psi_{n_i}, \tag{1}$$

where $\{n_i\}_i \subset \{1, \ldots, N\}$ are the vector indices, each ψ_{n_i} being one of the elements of the sparsity-inducing basis or tight frame, and $\{\alpha_i\}$ are the transform coefficients; in matrix form, this is expressed as $x = \Phi\alpha$. Such bases, including wavelets [20], Gabor bases [20], curvelets [2], wedgelets [16], etc., have been found for many classes of natural signals.

The standard procedure to achieve compression of such signals is then to (*i*) acquire the full N-point signal x; (*ii*) compute the complete set of transform coefficients α; (*iii*) locate the (few) largest, significant coefficients and discard the (many) small coefficients; (*iv*) encode the *values and locations* of the largest coefficients. Such a procedure will allow us to recover a high-quality representation of the original signal.

This procedure has three inherent inefficiencies: first, for a high-dimensional signal, we must start with a large number of values N. Second, the encoder must compute *all* of the N transform coefficients $\{\alpha_i\}$, even though it will discard most of them. Third, the encoder must encode the locations of the large coefficients, since the pattern of where they are located will change with each signal.

This begs a simple question: For a given signal, is it possible to directly estimate the set of large α_i's that will not be discarded? While this seems impossible, the groundbreaking work of Candès, Romberg and Tao [3] has shown that about K random projections contain enough information to reconstruct piecewise smooth signals. An offshoot of this work, which has been given the name of *Compressed Sensing* (CS) [4–8, 18, 19], has emerged that builds on this principle.

In the CS theory we do not measure the large α_i's directly. Rather, we measure the projections $v_i = \langle x, V_i \rangle$ of the signal onto a *second set* of basis functions $\{V_i\}_i$, which we will call the *measurement basis*. The CS theory tells us that when certain conditions hold, namely that the $\{V_i\}_i$ basis set does not provide sparse representations of the elements $\{\psi_i\}_i$ — a condition known as *incoherence* of the two bases, then it is indeed possible to recover the set of large α_i's from a similarly sized set of measurements $\{y_i\}_i$. This incoherence property holds for many pairs of bases, including for example, delta spikes and the sine waves of the Fourier basis, or the Fourier basis and wavelets. The CS theory also applies to redundant representations such as curvelets [2], which are well-suited to match geometrical edge structures in images.

The recovery of the set α of significant (large) coefficients α_i is achieved using *optimization*. Given a set $y \in \mathbb{R}^M$ of measurements $y_i = \langle x, V_i \rangle$ (where presumably the number of measurements $M \ll N$), we can solve for the signal having the ℓ_0-sparsest transform $\{\alpha_i\}$ that agrees with the observed coefficients y_i:

$$\begin{aligned} \widehat{\alpha} &= \arg\min \|\alpha\|_0 \\ & \text{s.t. } y = V\Psi\alpha \end{aligned} \tag{2}$$

where $V = [V_1^T, \ldots, V_M^T]^T$ is the matrix representation of the measurement basis. The columns of $\Phi = V\Psi = [\phi_1, \ldots, \phi_n]$ are called the *holographic basis*.

Because of the incoherence between the original (Ψ) and the measurement (V) bases, if the signal is sparse in the original basis Ψ, then no other set of sparse coefficients α' can yield the same projections y. Unfortunately, solving this ℓ_0 optimization problem is prohibitively complex, requiring combinatorial enumeration of the possible sparse $\{\psi_i\}$ subspaces.

The amazing revelation that enables this new theory is that it is not necessary to solve this ℓ_0-minimization problem to recover the set of largest α_i. In fact, a much simpler problem yields an equivalent solution: given some technical conditions on Φ, we need only solve for the ℓ_1-sparsest transform α that agrees with the observed coefficients y [3–8, 18, 19]:

$$\begin{aligned} \widehat{\alpha} &= \arg\min \|\alpha\|_1 \\ & \text{s.t. } y = \Phi\alpha. \end{aligned} \tag{3}$$

There is no free lunch, however; according to the theory, the observation set must have size $M = |y| = CK$, where C is dependent on the matrix Φ. Fortunately, in practice, it suffices to choose C between 3 and 6.

This optimization problem is significantly more approachable: traditional linear programming techniques can be used to solve the following equivalent problem:

$$\begin{aligned} \widehat{\beta} &= \arg\min \mathbf{1}^T\beta \\ & \text{s.t. } y = A\beta, \end{aligned} \tag{4}$$

where $A = [\Phi, -\Phi]$ is a $cK \times 2N$ matrix, and the solution $\widehat{\beta} \in \mathbb{R}^{2N}$ can be decomposed into $\widehat{\beta} = [\widehat{\alpha}_P^T, \widehat{\alpha}_N^T]^T$ to recover $\widehat{\alpha} = \widehat{\alpha}_P - \widehat{\alpha}_N$.

Succinct requirements on the matrix Φ for the CS machinery to work were introduced by Candès and Tao [5]:

Definition 1: For every integer $1 < K < N$, the K-restricted isometry constant δ_K for the matrix Φ is the smallest quantity such that for all $T = \{t_1, t_2, \ldots\} \subset \{1, 2, \ldots, n\}$, $|T| < K$,

$$(1 - \delta_K)\|c\|^2 \leq \|\Phi_T c\|^2 \leq (1 + \delta_K)\|c\|^2,$$

where $\Phi_T = [\phi_{t_1}, \phi_{t_2}, \ldots]$ and $c \in \mathbb{R}^n$.

The success of CS is dependent on these constants:

Theorem 1: Let K be such that $\delta_K + \delta_{2K} + \delta_{3K} < 1$ for Φ. Then for any vector α_0 supported on T_0 with $|T_0| \leq K$, α_0 is the unique minimizer of (3).

This property is essentially equivalent to having all small subsets of columns of Φ form almost orthogonal matrices. One of the interesting aspects of CS is that random Gaussian and Bernoulli matrices hold Theorem

1 with high probability for $K \lesssim M/\log(N)$; the incomplete Fourier Transform matrix also holds this property as well [4, 18].

For signals of large size, the calculation of the solution to the linear program (4) is still computationally unfeasible; some algorithms have been proposed for special cases of CS, such as noisy measurements or availability of side information [7, 22].

III. Reconstruction in Compressed Sensing as a Sparse Approximation Problem

The ℓ_1 minimization problem (3) was previously formulated by Chen, Donoho and Saunders [11] as *Basis Pursuit*, an algorithm to obtain a sparse representation of the vector y in a *dictionary* or overcomplete basis $\Phi = [\phi_1, \ldots, \phi_N]$, where $\phi_i = V\psi_i$; in other words, we are attempting to reconstruct the original coefficients α by finding a sparse representation of the measurements y in the projected sparse basis $\Phi = V\Psi$; notice that since

$$y = Vx = V\Psi\alpha = \Phi\alpha, \tag{5}$$

the sparsest representation of y in the basis Φ will be equivalent to the sparsest representation of the original signal x in the original basis Ψ, given that the matrix Φ holds Theorem 1. Notice that the basis Φ is an overcomplete basis for $\mathbb{R}^M$ containing N basis vectors. It can be seen intuitively that the conditions specified by Theorem 1 are necessary so that the sparse approximation of the measurement y in the basis Φ will coincide with the sparse decomposition α of the original signal.

The original purpose of the Basis Pursuit algorithm was to find a sparse representation of a signal in a union of bases, such as spikes and sines, that would allow for a compact representation of signals containing different classes features, where each class can be compactly expressed in a different basis. Other algorithms that find such sparse representations have been proposed and will be described in this section. The main characteristic of these *pursuit* algorithms is that they obtain efficient but suboptimal sparse representations of the original signal. Our goal is to evaluate the computational complexity and performance of these algorithms, since Basis Pursuit presents two computational challenges: the algorithm consists of a global optimization that requires large computational resources, and the computations — being in the order of $O(N^3)$ — become prohibitive for N in the thousands.

A. Matching Pursuit

Matching Pursuit [21] is an algorithm with reduced computational complexity via a greedy strategy: basis vectors are selected one by one from the dictionary, while optimizing the signal approximation at each step. The Matching Pursuit algorithm is as follows:

1) Initialize the residual $r_0 = y$ and the approximation $\widehat{\alpha} = 0$, $\widehat{\alpha} \in \mathbb{R}^N$.
2) Select the dictionary vector that maximizes the value of the projection of the residual:

$$n_t = \arg\max_{i=1,\ldots,N} \frac{\langle r_{t-1}, \phi_i \rangle}{\|\phi_i\|^2}. \tag{6}$$

3) Update the estimate of the coefficient for the selected vector and the residual:

$$r_t = r_{t-1} - \frac{\langle r_{t-1}, \phi_{n_t} \rangle}{\|\phi_{n_t}\|^2}\phi_{n_t}, \tag{7}$$

$$\widehat{\alpha}_{n_t} = \widehat{\alpha}_{n_t} + \frac{\langle r_{t-1}, \phi_{n_t} \rangle}{\|\phi_{n_t}\|^2}. \tag{8}$$

4) If $\|r_t\|_2 > \epsilon\|y\|_2$, repeat iteration; otherwise, terminate.

The parameter ϵ determines the target error level allowed for algorithm convergence. At each step, the relationship between the previous and current residual is given by

$$r_t = \frac{\langle r_t, \phi_{n_t} \rangle}{\|\phi_{n_t}\|^2}\phi_{n_t} + r_{t+1}; \tag{9}$$

since r_{t+1} is orthogonal to ϕ_{n_t},

$$\|r_t\|^2 = \|\langle r_t, \phi_{n_t}\rangle\|^2 + \|r_{t+1}\|^2. \tag{10}$$

Since these are all nonnegative quantities, the following theorem can be easily proven [20]:

Theorem 2: There exists $\lambda > 0$ such that for all $t \geq 0$,

$$\|r_t\| \leq 2^{-\lambda t}\|y\|. \tag{11}$$

This theorem states that the residual converges exponentially to zero, and thus the reconstruction converges to the original signal. The convergence rate λ decreases when the size M of the signal space increases. Some issues with Matching Pursuit include the unbounded number of iterations necessary for convergence and the calculation of inner products with the residual for each iteration. Some optimizations have been formulated for this last case [20].

B. *Orthogonal Matching Pursuit*

The approximations of a Matching Pursuit are improved by orthogonalizing the vectors in the overcomplete basis with a Gram-Schmidt procedure; this optimization is termed Orthogonal Matching Pursuit and was first proposed by Pati, Rezaifar and Krishnaprasad [24]. The resulting algorithm convereges with a finite number of iterations which is less than or equal to the dimensionality of the signal space. The price to be paid is the large computational cost of the Gram-Schmidt orthogonalization.

The vector ϕ_{n_t} selected at each step by Matching Pursuit is in general not orthogonal to the previously selected vectors $\{\phi_{n_i}\}_{1\leq i<t}$. When subtracting the projection of r_{t-1} over ϕ_{n_t}, the algorithm reintroduces new components in the directions of $\{\phi_{n_i}\}_{1\leq i<t}$. This is avoided by projecting the residues on an orthogonal family $\{\gamma_i\}_{1\leq i<t}$.

The Orthogonal Matching Pursuit algorithm is then as follows:

1) Initialize the residual $r_0 = y$, the set of picked indexes $I = \emptyset$ and the approximation $\widehat{\beta} = 0, \beta \in \mathbb{R}^M$.
2) Select the dictionary vector that maximizes the value of the projection of the residual, and add its index to the set of picked indexes:

$$\begin{aligned} n_t &= \arg\max_{i=1,\ldots,N} \frac{\langle r_{t-1}, \phi_i\rangle}{\|\phi_i\|^2}, & (12)\\ I &= [I, n_t]. & (13) \end{aligned}$$

3) Orthogonalize the picked basis vector against the orthogonalized set of previously picked dictionary vectors:

$$\gamma_t = \phi_{n_t} - \sum_{p=0}^{t-1} \frac{\langle \phi_{n_t}, \gamma_p\rangle}{\|\gamma_p\|^2}\gamma_p. \tag{14}$$

4) Update the estimate of the coefficient for the selected vector and the residual:

$$\begin{aligned} r_t &= r_{t-1} - \frac{\langle r_{t-1}, \gamma_t\rangle}{\|\gamma_t\|^2}\gamma_t, & (15)\\ \widehat{\beta}_t &= \frac{\langle r_{t-1}, \gamma_t\rangle}{\|\gamma_t\|^2}. & (16) \end{aligned}$$

5) If $\|r_t\|_2 > \epsilon\|y\|_2$, repeat iteration; otherwise, terminate.

The previous algorithms provides us with a decomposition of y into the orthogonalized dictionary vectors $\Gamma = [\gamma_1, \ldots, \gamma_M]$:

$$y = \sum_{p=1}^{M} \frac{\langle r_p, \gamma_p\rangle}{\|\gamma_p\|^2}\gamma_p. \tag{17}$$

To expand y over the original overcomplete basis Φ, we consider the relationship between Γ and Φ given by the QR factorization

$$\Phi_I = \Gamma R, \tag{18}$$

where $\Phi_I = [\phi_{n_1}, \ldots, \phi_{n_M}]$ is the so-called *mutilated basis*. Since $y = \Gamma\beta = \Phi_I \alpha_I = \Gamma R \alpha_I$, where α_I is the mutilated coefficient vector, then we can reconstruct as follows:

$$\begin{aligned} \alpha_I &= R^{-1}\beta, && (19) \\ \widehat{\alpha}_i &= \begin{cases} \alpha_{I,j} & \text{if } i = n_j \in I \\ 0 & \text{if } i \notin I. \end{cases} && (20) \end{aligned}$$

When the number of iteration increases and becomes closer to M, the residues of an orthogonal pursuit have norms that decrease faster than for a non-orthogonal pursuit.

C. *Matching Pursuits for Compressed Sensing*

We evaluate the performance of the Matching Pursuit algorithms for CS reconstruction purposes. Our intuition is that since these algorithms will yield the sparsest decomposition $\widehat{\alpha}$ of the measurement vector y in the projected basis Φ, the reconstruction can be used to recover the original signal x in the sparsity-inducing basis Ψ as $\widehat{x} = \Psi\widehat{\alpha}$. Since the decompositions are the same, and the MP and OMP algorithms act in a greedy fashion by picking the maximum projection, the algorithms tend to pick the coefficient in order from largest to smallest magnitudes; an error in the estimates is induced by the nonorthogonality of the dictionary vectors, whose magniutde is controlled by the δ_K constants described earlier.

The accuracy of the coefficient estimates is critical to the CS reconstruction; we derive the relationship between the coefficient estimate and the actual coefficient, given by

$$\begin{aligned} \widehat{\alpha}_{i,t} &= \widehat{\alpha}_{i,t-1} + \frac{\langle r_{t-1}, \phi_i \rangle}{\|\phi_i\|^2} = \widehat{\alpha}_{i,t-1} + \frac{\langle \sum_j (\alpha_{j,t-1} - \widehat{\alpha}_j)\phi_j, \phi_i \rangle}{\|\phi_{n_t}\|^2} \\ &= \widehat{\alpha}_{i,t-1} + \frac{(\alpha_i - \widehat{\alpha}_{i,t-1})\|\phi_i\|^2 + \sum_j [(\alpha_j - \widehat{\alpha}_{j,t-1})\langle \phi_j, \phi_i \rangle]}{\|\phi_i\|^2} \\ &= \alpha_i + \frac{\sum_{j \neq i} [(\alpha_j - \widehat{\alpha}_{j,t-1})\langle \phi_j, \phi_i \rangle]}{\|\phi_i\|^2}. \end{aligned} \tag{21}$$

Since Theorem 1 requires that all small subsets of columns of Φ (up to K in our case of interest) form almost orthogonal matrices, we expect that the error, proportional to $\langle \phi_j, \phi_i \rangle$, will be small.

Figure 1 show the reconstruction of standard test signals HeaviSine and Blocks, processed using CS and recovered using the Daubechies-8 and Haar wavelet bases, respectively. The first column shows the original signal; the second column shows the sparse approximation with the given number of coefficients K, which shows that the signals have a sparse representation. The signal was projected into CK random Gaussian vectors, where $C = 4$, to obtain the vector of CS measurements y. The third, fourth and fifth columns show the reconstruction of the signal from the CS measurements using Basis Pursuit, Matching Pursuit and Orthogonal Matching Pursuit, respectively. Each reconstruction shows the Mean Square Error and the computation time for the reconstruction.

These results show that all of the Matching Pursuit algorithms effectively reconstruct the signal, while the Matching Pursuit and Orthogonal Matching Pursuits perform much faster than the Basis Pursuit (40-350 and 3-13 times faster, respectively), with the overhead in the OMP algorithm caused by the orthogonalization of the basis vectors.

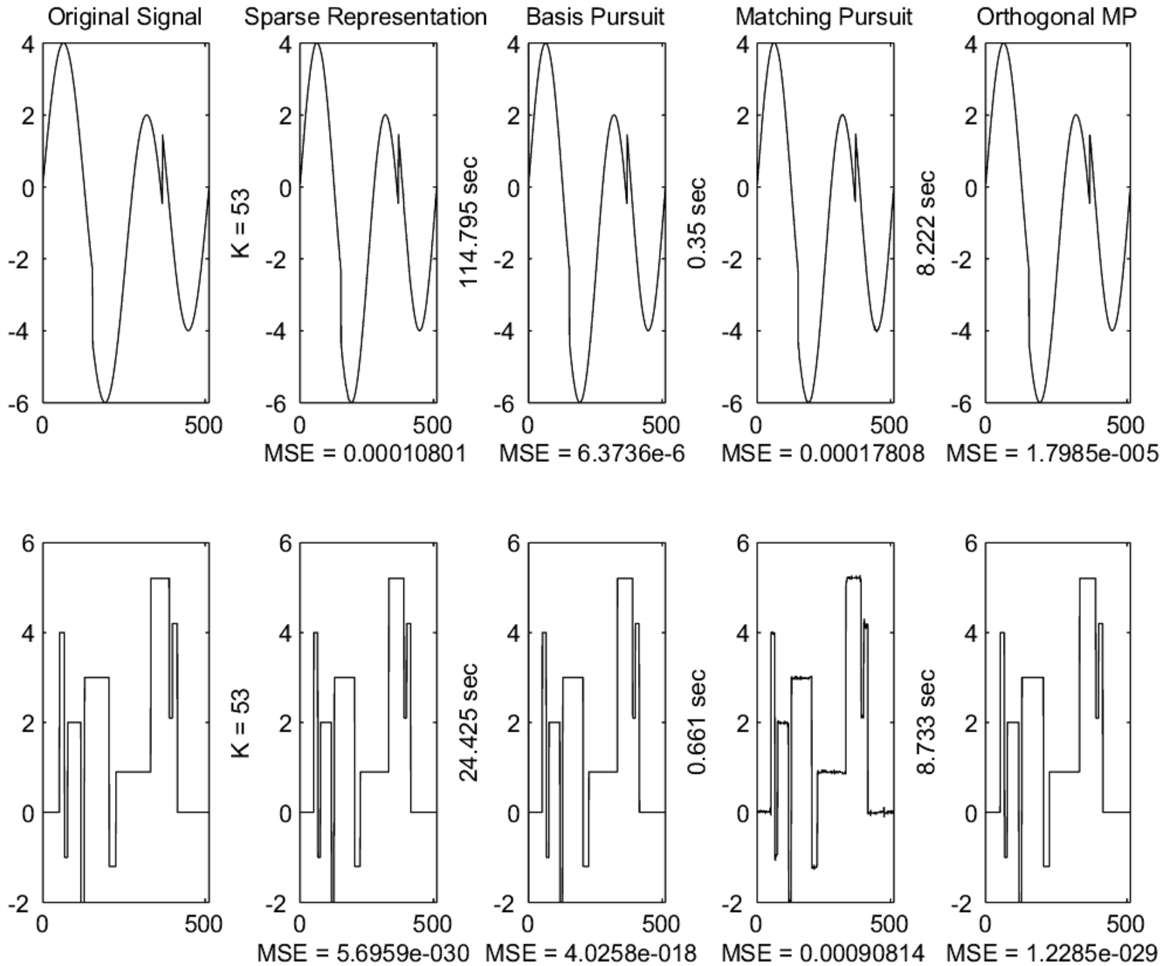

Fig. 1. Performance of pursuit algorithms for CS reconstruction. Top row: HeaviSine; bottom row: Blocks.

IV. Tree Matching Pursuits

A. Motivation

We wish to customize the previous reconstruction algorithms for a class of signals of interests: piecewise smooth signals, which have sparse representations in wavelet bases.

A wavelet basis for the space $\mathbb{R}^N$, where $N = 2^l$, $l \in \mathbb{Z}$, is defined by the scaling and wavelet functions ϕ and ψ, which in turn define the basis functions

$$\psi_{j,k}[n] := 2^{j/2}\psi(2^j t - k), \tag{22}$$

with j and k indexing the scale and position of the wavelet atom function, respectively. The decomposition of the signal into this wavelet basis is as follows:

$$x = u\phi + \sum_{j=0}^{l-1}\sum_{k=1}^{2^j} w_{j,k}\psi_{j,k}, \tag{23}$$

where $u = \langle x, \phi\rangle$ and $w_{j,k} = \langle x, \psi_{j,k}\rangle$. The transform can be expressed in matrix form as $x = \Psi\alpha$, where $\Psi = [\phi, \psi_{0,1}, \psi_{1,1}, \ldots, \psi_{l-1,2^{l-1}}]$ and $\alpha = [u, w_{0,1}, w_{1,1}, \ldots, w_{l-1,2^{l-1}}]$, both sets being ordered first by scale and then by offset.

The multiscale nesting structure of the wavelet atoms — the support of each $\psi_{j,k}$ contains the supports of $\psi_{j+1,2k-1}$ and $\psi_{j+1,2k}$ — induces a binary tree structure in the wavelet coefficients. We denote this

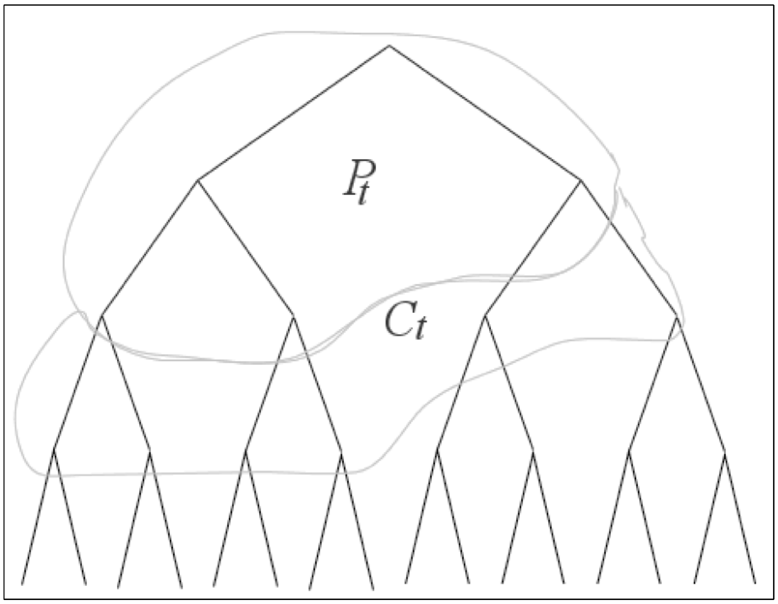

Fig. 2. Picked and candidate coefficient sets in a wavelet binary tree.

relationship between these coefficients by stating that the former is the *parent* coefficient and the two latter are the *children* coefficients.

It is commonly known that when the wavelet coefficients of a piecewise smooth signal are arranged in a binary tree structure, relationships are present between the coefficients of parent and child nodes. A large wavelet coefficient (in magnitude) generally indicates the presence of a discontinuity in the original signal inside its support; a small wavelet coefficient generally indicates a smooth region in the support of the given wavelet. Due to the nesting properties of child wavelets inside their parents, edges manifest themselves in the wavelet domain as chains of large coefficients propagating across scales in the wavelet tree. This causes the most significant wavelet coefficients of most piecewise smooth signals to form a connected subtree in the wavelet binary tree. Wavelet coefficients also have decaying magnitudes as the scale decreases [20].

We observe that for piecewise smooth signals, due to these characteristics of the wavelet coefficients, the Matching Pursuit algorithms tend to select coefficients located at the top of the tree and then continue the selection down the tree, effectively building the connected tree that contains the most significant coefficients from the top down. For this reason, we propose an adaptation of the Matching Pursuit algorithms to achieve faster reconstruction for piecewise smooth signals, which exploits the tree structure of the wavelet coefficients. The algorithms will be described in the next section.

B. Algorithms

We formulate a modification to the Matching Pursuit algorithms that will allow for lower computational complexity while maintaining good reconstruction quality. In this algorithm, we take only a subset of the basis vectors into consideration at each of the iterations, and expand that set at each iteration. We begin by defining two sets of coefficients P_t and C_t, which contain the set of picked vectors — those vectors that correspond to nonzero coefficients in the estimate $\widehat{\alpha}$ — and the candidate vectors — vectors with zero coefficients in $\widehat{\alpha}$ but whose projections are evaluated. These sets are initialized as $P_t = \emptyset$ and $C_t = \{1\} \cup D(1)$, where $D(i)$ denotes a select of descendent coefficients of coefficient i. We will initially set $D(i)$ to be the two children coefficients of coefficient i, a model that we will expand later. Figure 2 shows an example on the selection of the coefficients in the wavelet tree.

At each iteration of the algorithm, we will search for the dictionary vector in $P_t \cup C_t$ that yields the maximum projection to the current residual; if the picked vector belongs in C_t, that coefficient i (and the set of its ancestors, denoted by $A(i)$) will be moved to the set of picked coefficients P_t and removed from C_t, and the descendent set $D(i)$ will be added to C_t. The *Tree Matching Pursuit* algorithm follows:

1) Initialize the residual $r_0 = y$, the approximation $\widehat{\alpha} = 0$, $\widehat{\alpha} \in \mathbb{R}^N$, and the picked and candidate coefficient sets $P_0 = \emptyset$ and $C_0 = 1 \cup D(1)$.
2) Select the dictionary vector that maximizes the value of the projection of the residual among those available:
$$n_t = \arg \max_{i \in P_{t-1} \cup C_{t-1}} \frac{\langle r_{t-1}, \phi_i \rangle}{\|\phi_i\|^2}. \tag{24}$$
3) If $n_t \in C_{t-1}$, i.e. if the coefficient is not currently among those picked, move the coefficient and its ancestors to the set of picked coefficients, remove them from the set of candidates and add the descendants of n_t to the set of candidates:
$$P_t = P_{t-1} \cup n_t \cup A(n_t), \tag{25}$$
$$C_t = C_{t-1} \setminus (n_t \cup A(n_t)) \cup D(n_t). \tag{26}$$
4) Update the estimate of the coefficient for the selected vector and the residual:
$$r_t = r_{t-1} - \frac{\langle r_{t-1}, \phi_{n_t} \rangle}{\|\phi_i\|^2} \phi_{n_t}, \tag{27}$$
$$\widehat{\alpha}_{n_t} = \widehat{\alpha}_{n_t} + \frac{\langle r_{t-1}, \phi_{n_t} \rangle}{\|\phi_{n_t}\|^2}. \tag{28}$$
5) If $\|r_t\|_2 > \epsilon \|y\|_2$, repeat iteration; otherwise, terminate.

The *Tree Orthogonal Matching Pursuit* algorithm is defined in a similar way:

1) Initialize the residual $r_0 = y$, the set of picked indexes $I = \emptyset$, the approximation $\widehat{\beta} = 0, \beta \in \mathbb{R}^M$, and the picked and candidate coefficient sets $P_0 = \emptyset$ and $C_0 = 1 \cup D(1)$.
2) Select the dictionary vector that maximizes the value of the projection of the residual, and add its index to the set of picked indexes:
$$n_t = \arg \max_{i \in P_{t-1} \cup C_{t-1}} \frac{\langle r_{t-1}, \phi_i \rangle}{\|\phi_i\|^2}, \tag{29}$$
$$I = [I, n_t]. \tag{30}$$
3) If $n_t \in C_{t-1}$, move the coefficient and its ancestors to the set of picked coefficients, remove them from the set of candidates and add the descendants of n_t to the set of candidates:
$$P_t = P_{t-1} \cup n_t \cup A(n_t), \tag{31}$$
$$C_t = C_{t-1} \setminus (n_t \cup A(n_t)) \cup D(n_t). \tag{32}$$
4) Orthogonalize the picked basis vector against the orthogonalized set of previously picked dictionary vectors:
$$\gamma_t = \phi_{n_t} - \sum_{p=0}^{t-1} \frac{\langle \phi_{n_t}, \gamma_p \rangle}{\|\gamma_p\|^2} \gamma_p. \tag{33}$$
5) Update the estimate of the coefficient for the selected vector and the residual:
$$r_t = r_{t-1} - \frac{\langle r_{t-1}, \gamma_t \rangle}{\|\gamma_t\|^2} \gamma_t, \tag{34}$$
$$\widehat{\beta}_t = \frac{\langle r_{t-1}, \gamma_t \rangle}{\|\gamma_t\|^2}. \tag{35}$$
6) If $\|r_t\|_2 > \epsilon \|y\|_2$, repeat iteration; otherwise, terminate.

We note that since the number of coefficients that are considered at each iteration is much smaller than the number of available dictionary vectors, the reduction in computations of these algorithms is significant.

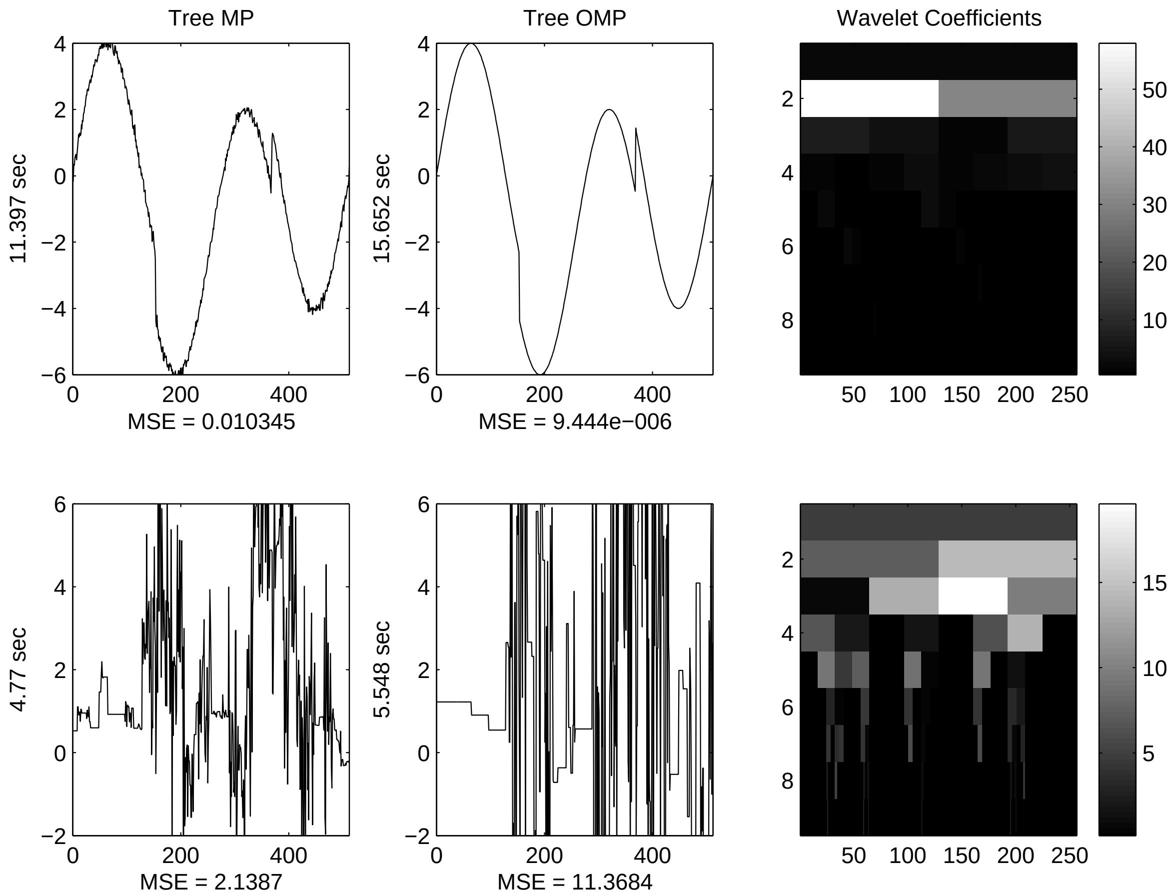

Fig. 3. Initial performance of tree pursuit algorithms for CS reconstruction. Top row: HeaviSine; bottom row: Blocks. Left column: TMP reconstruction; middle column: TOMP recontruction; right column: wavelet coefficient magnitudes — rows represent scales, largest to smallest.

V. Adaptations of Tree Matching Pursuit

A. Performance of TMP

Figure 3 shows the performance of the tree-based Matching Pursuit algorithms for the same signals shown in Figure 1. We see some minor artifacts on the reconstruction of the HeaviSine, some major artifacts on the TMP reconstruction of Blocks and an incorrect reconstruction from TOMP for Blocks. Notice that the first quarter of this last reconstruction is flat, implying all zero coefficients for that part of the signal. In looking for an explanation for this behavior, it is convenient to observe the wavelet coefficient tree for each one of the signals. We see that for the Blocks signal, the significant coefficients *actually do not* form a connected subtree; notice the small coefficient $\alpha_5 = w_{3,1}$. The Tree Matching Pursuit algorithms, due to their nature, will not immediately include the coefficient subtree that is rooted at this small coefficient, and thus the energy from these coefficients may be partially or completely reallocated to other coefficients in the rest of the tree, which is seen for the TMP and TOMP algorithms. This indicates a need to reevaluate our model for piecewise smooth signals in the wavelet coefficient domain.

It is also intuitive to see that the tree-based algorithms do not yield the same approximation as the original pursuit algorithms due to the fact that the maximum non-picked coefficient at some stage was not included among the set of candidate coefficients. This is noticeable especially in the Blocks signal, since the coefficients are not monotonically decreasing.

B. Smoothness

It is also worth pointing out that, save for the small coefficient already mentioned, the rest of the coefficients would indeed form a connected tree. If we were to include not only the children of a given coefficient

i in the set D_i, but the 'grandchildren' and even more descendants of the coefficient i, the algorithm would have correctly identified the next significant coefficients $w_{4,1}$ and its descendants. This modification has advantages and disadvantages; it is clear from the results that the reconstruction will be the same or better as we add more descendants into D_i, but the computational advantage against the original greedy algorithms will be reduced. We try to reach a balance between these two extremes by defining an adaptive set of descendants:

Definition 2: The l-depth set of descendants $D_l(i)$ is the set of coefficients that are within l levels below coefficient i in the wavelet tree.

We revisit the process of basis vector selection for the proposed algorithms with this new definition of the set D: our interest is to look for $l \in [1, log_2(N)]$ large enough such that the maximum non-picked coefficient is always inside the set of candidate coefficients:

1) Initialize

$$P_{l,0} = \emptyset \text{ and } C_{l,0} = \{1\} \cup D_l(1).$$

2) Expand these sets at iteration t as follows:

$$\begin{aligned} n_{l,t} &= \arg\max_{i \in C_{l,t}} \frac{\langle x, \psi_i \rangle}{\|\psi_i\|^2}, \\ P_{l,t} &= P_{l,t-1} \cup A(n_{l,t}) \cup \{n_{l,t}\}, \\ C_{l,t} &= C_{l,t-1} \setminus [A(n_{l,t}) \cup \{n_{l,t}\}] \cup D_l(n_{l,t}). \end{aligned}$$

3) The iterations end when $P_{l,t} = \{1, \ldots, N\}$ and $C_{l,t} = \emptyset$.

This allows us to define a class of signals for which the algorithm behaves well:

Definition 3: We say that a signal $x = \Psi\theta$ is *l-degree smooth* in the wavelet basis Ψ if for all $t_2 > t_1$, $\langle x, \psi_{n_{l,t_2}} \rangle \le \langle x, \psi_{n_{l,t_1}} \rangle$

We then can describe the modified l-Tree Matching Pursuit and l-Tree Orthogonal Matching Pursuit algorithms, where the sets of coefficients P and C are updated as described above; we can also postulate the following result on performance of these algorithms.

Theorem 3: For a signal x that is l-degree smooth in Ψ, the l-TMP and l-TOMP algorithm on the set of measurements $y = Vx$ using the dictionary $\Phi = V\Psi$ will recover the same approximation as the MP and OMP algorithms, respectively.

Proof. The original greedy algorithm selects the coefficient with maximum projection to the corresponding vector at each iteration. If at some iteration of the tree pursuit algorithm such coefficient is inside the set of picked or candidate coefficients, then the algorithms will behave in the same way. If it the coefficient is not a part of either set, then the picked coefficient will be smaller than the actual maximum, which will eventually be included in the set of candidates; therefore we would find that the maximum projection among candidates at a future time t_2 would be larger than the same maximum at an earlier time t_1. This in turn means that if this previous condition never occurs, then the algorithms will pick the same coefficients at each iteration and the reconstruction will be identical. □

We analyze the smoothness of our test signals by defining a measurement of l-smoothness:

Definition 4: The l-degree smoothness measurement of the signal $x \in \mathbb{R}^N$ in the basis $\Psi = [\psi_1, \ldots \psi_N]$ is given by

$$\delta_l = \max\left\{ \max_{1 \le t_1 < t_2} \left[\langle x, \psi_{n_{l,t_2}} \rangle - \langle x, \psi_{n_{l,t_1}} \rangle \right], 0 \right\}. \quad (36)$$

A signal will be l-degree smooth if $\delta_l = 0$. We also want to review the accuracy of the estimation and the computational complexity for different values of l for these test signals and see the correlation between these values.

Figure 4 shows the values of δ_l for the HeaviSine and Blocks signals of length 512 samples, as well as the reconstruction error and computational complexity (in number of inner products) of the TMP, TOMP, l-TMP and l-TOMP algorithms. We see that when $\delta_l \sim 0$, the reconstruction error is negligible and the computational complexity for the l-TMP algorithm is lower than that of the original algorithm, albeit increasing with

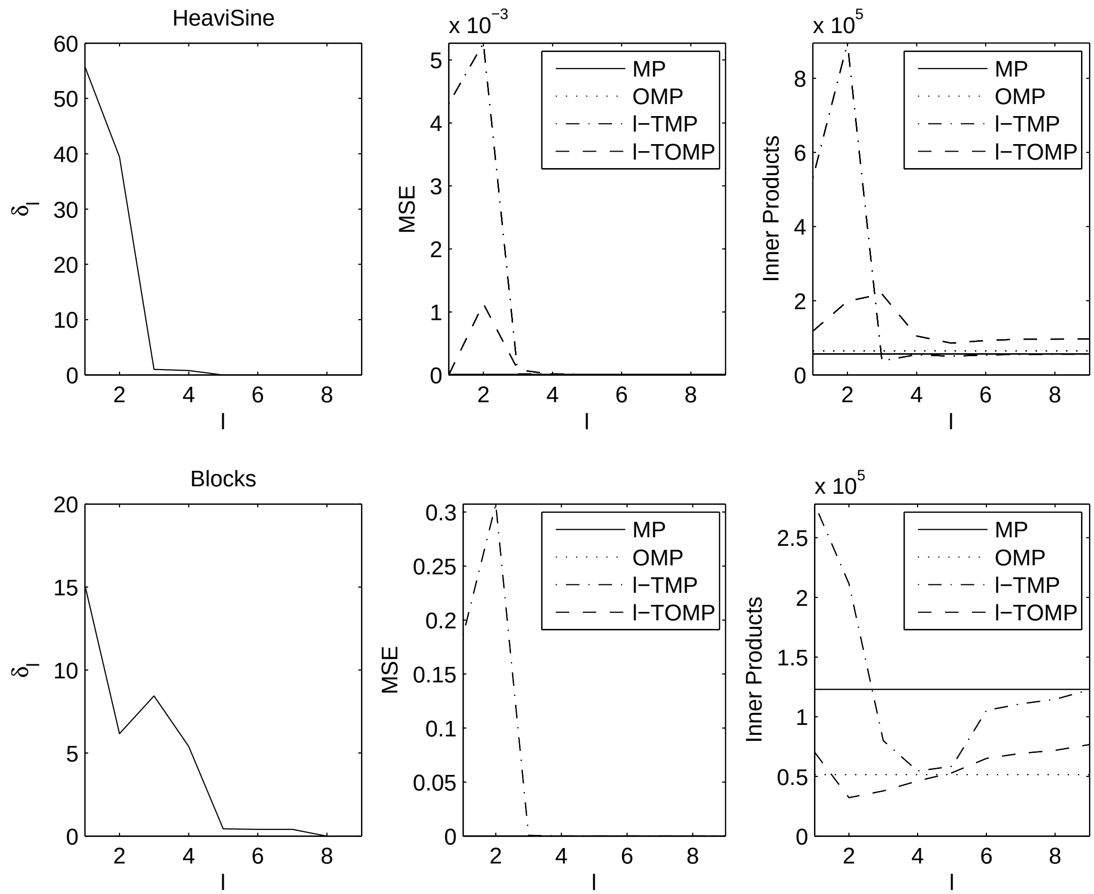

Fig. 4. Performance of l-TMP and l-TOMP algorithms. Top row: Heavisine; bottom row: Blocks. For l with small δ_l, the reconstructions have small error; the computational complexity is smallest for the lowest l with small δ_l in each signal.

l. The large computational complexity for small l are due to the corrections in the values of the coefficients made by the Tree Matching Pursuit Algorithms as few coefficients are included at each step. Notice that the figure shows that HeaviSine and Blocks have smoothness degrees of 5 and 8 in their respective wavelet bases.

C. Complex Wavelet Transform

We have discovered a few issues in our assumptions for the structure of wavelet coefficients in piecewise smooth signals. A problem of concern to us, as described in [1], is the *oscillations* of the wavelet function: since wavelets are bandpass functions, the wavelet coefficients can oscillate positive and negative around singularities. This may cause the wavelet coefficient corresponding to a discontinuity to be small, therefore overstating the assumption that for piecewise smooth signals, the wavelet coefficients form a connected tree.

In [1], a *complex wavelet transform* is proposed, inspired on the Fourier transform, which does not have some of the problems that the real-valued wavelet transform has. A complex wavelet transform has the same structure shown in (22) and (23), but the wavelet is complex-valued:

$$\psi_c(t) = \psi_r(t) + j\psi_i(t) \tag{37}$$

where $\psi_r(t)$ is real and even and $j\psi_i(t)$ is imaginary and odd, and they (approximately) form a Hilbert transform pair. The complex wavelet transform ($\mathbb{C}$WT for short) can be easily implemented using a doubletree structure, where we build two *real* wavelet transforms by using $\psi_r(t)$ and $\psi_i(t)$ separately, obtaining then sequences of coefficients α_r and α_i. The complex wavelet coefficients are then defined as

$$\alpha_c = \alpha_r - j\alpha_i \tag{38}$$

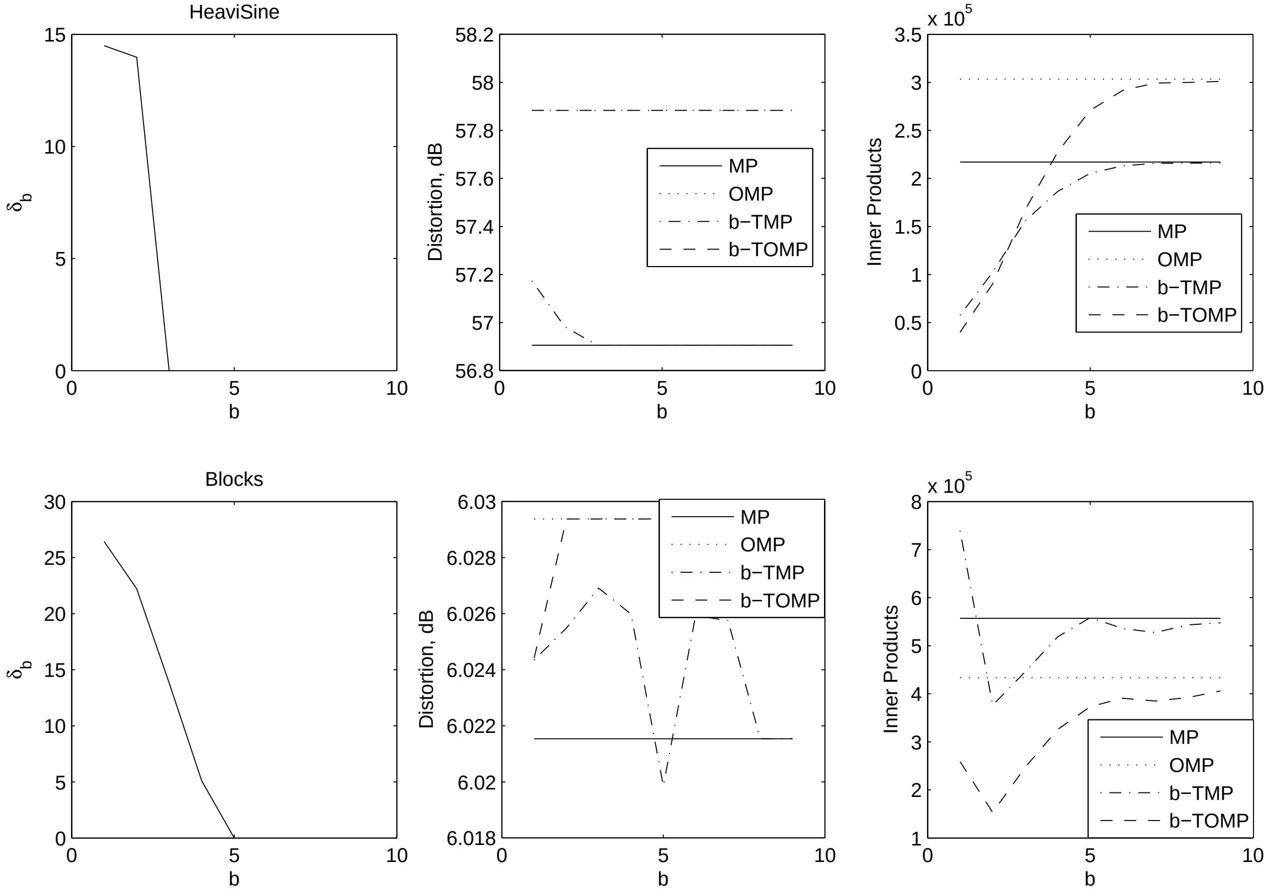

Fig. 5. Performance of l-TMP and l-TOMP algorithms using complex wavelet transform. Top row: Heavisine; bottom row: Blocks. The value of l for which $\delta_l \sim 0$ has decreased for both signals; the computational complexity is smallest for the lowest l with small δ_l in each signal.

Notice that either the real or the imaginary part of the wavelet coefficients would suffice to reconstruct the signal; however, a complex wavelet representation allows for more piecewise smooth signals to form a connected subtree of significant wavelet coefficients when the complex magnitude is evaluated, due to the Hilbert Pair quality of the real and imaginary components of the complex wavelet: when a discontinuity is present and the real (or imaginary) wavelet coefficient is small, the imaginary (or real) wavelet coefficient is large, yielding shift invariance on the coefficient magnitudes of the signal. As such, when the Tree Matching Pursuit algorithm is implemented using a Complex Wavelet basis, a smaller band will be necessary for efficient reconstruction. We choose to use $\mathbb{C}$WT as our sparsity-inducing basis, which will allow for more efficient solutions using the l-TMP algorithms than the original real wavelet transform implementation.

Figure 5 shows the values of δ_l for the HeaviSine and Blocks signals of length 512 samples when the $\mathbb{C}$WT is used, as well as the reconstruction error and computational complexity (in number of inner products) of the TMP, TOMP, l-TMP and l-TOMP algorithms. As the figure shows, the degree of smoothness of these two test signals is smaller when the $\mathbb{C}$WT is used; as such, the l-TMP algorithm that achieves good reconstruction will be computationally simpler.

For the $\mathbb{C}$WT based Matching Pursuit algorithms, we change our formulation for the coefficient estimates slightly:

$$\widehat{\alpha}_{c,i,t} = \widehat{\alpha}_{c,i,t-1} + \frac{2\langle r_{t-1}, \phi_{c,i}\rangle}{\|\phi_{c,i}\|^2}, \tag{39}$$

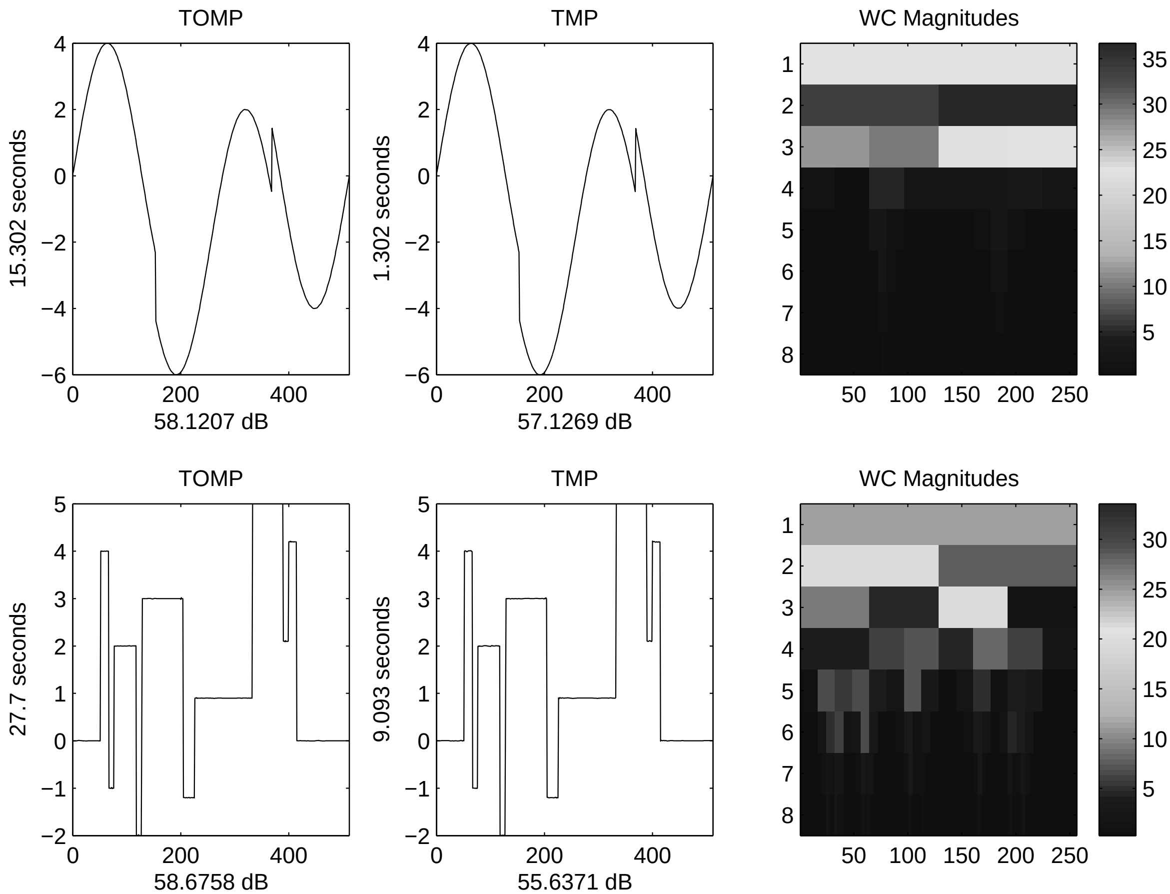

Fig. 6. Reconstructions using l-TMP and l-TOMP algorithms using complex wavelet transform.

which yields an approximation error as follows:

$$\widehat{\alpha}_{c,i,t} = \alpha_i + \frac{2\alpha_i^* \langle \phi_i^*, \phi_i \rangle + \sum_{j \neq i} [(\alpha_j - \widehat{\alpha}_{j,t-1}) \langle \phi_j, \phi_i \rangle + (\alpha_j - \widehat{\alpha}_{j,t-1})^* \langle \phi_j^*, \phi_i \rangle]}{\|\phi_i\|^2} \tag{40}$$

where both $\langle \phi_j, \phi_i \rangle$ and $\langle \phi_j^*, \phi_i \rangle$ are small if Theorem 1 holds.

Figure 6 shows the reconstruction using the tree Matching Pursuit algorithms, with $l = 1$ and $l = 2$ respectively. We choose to divide the contribution on the estimate of the real and imaginary parts of the complex wavelet transforms equally to exploit the inherent redundancy

VI. Extensions of Tree Matching Pursuit

We consider several extensions and applications of these algorithms, both in Compressed Sensing and in other fields.

A. Denoising

The current implementations of reconstruction in CS are not robust to the injection of noise in the measurements; in [19], the Basis Pursuit with Denoising algorithm is proposed to alleviate this problem, while in [7], bounds are given for the distortion of the recovered signal as a function of the noise power. Also, [22] recently proposed an EM-based algorithm to recover the signal from random projections, and also offer bounds for the distortion in the recovered signal; no examples were given for rate of convergence or distortion in the reconstruction.

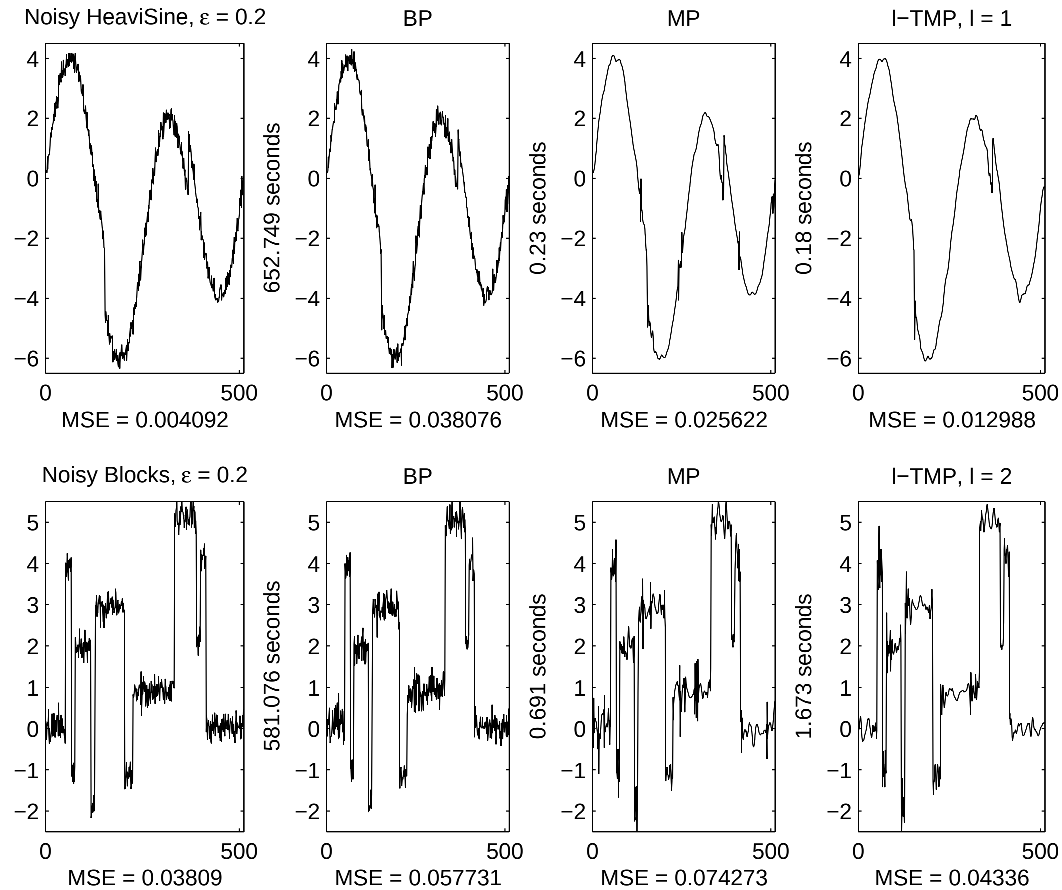

Fig. 7. Denoising performance results for several algorithms. The structure exploited by the Tree Matching Pursuit algorithm allows for better reconstruction of the signal from noisy measurements.

When the signal is sparse in the wavelet basis, we can effectively perform denoising by thresholding [15] by varying the convergence criterion as a function of the signal to noise ratio. In this fashion, we will only identify the most significant coefficients using the Matching Pursuit or Tree Matching Pursuit algorithm and effectively threshold the coefficient values at the reconstruction.

Noise artifacts are also observed with the standard methods of reconstruction (Basis Pursuit and Matching Pursuit) since the energy of the signal is not discriminated by band; in this case, a small amount of the energy from the coefficients in the coarsest scales 'leaks' to the finer scales and causes low-amplitude, high-frequency artifacts which resembles small-scale noise. By giving preference to the coarsest coefficients over the finest, the Tree Matching Pursuit algorithms help mitigate this effect during reconstruction.

Figure 7 show the performance of the algorithm in denoising. White Gaussian noise was added to the HeaviSine and Block signals with power 0.2; the results from the different algorithms are shown.

B. *Matching Pursuit using Transforms*

Although Matching Pursuit allows for computationally feasible reconstruction for higher-dimensional signals than Basis Pursuit, we still need to pay the price of storage of the original, measurement and/or holographic bases in memory. The original basis is of size $N \times N = O(N^2)$; the measurement and holographic bases are of size $M \times N = O(KN \log N)$ each. For large dimensional signals a large amount of memory is needed to perform Matching Pursuit using the matrices. As an example, for a image as small as 64×64 pixels, we have $N = 2^{12}$ and $K = 410$ with $M = 4K$, and we need about 240 Megabytes of memory to store these matrices.

We do have an alternative, however. Usually a wavelet transform exists to obtain the sparse representation of a signal. If we can replace the projection of a vector using the matrix product to a method using a transform, we do not need to store these matrices in memory. However, the computational complexity of the problem increases slightly, assuming that it is equally or more complex to calculate the transform than to apply the transform matrix. In this method, we also lose the computational gain from the Tree Matching Pursuit algorithms, since all of the coefficients are calculated at once using the transform. The denoising properties of the algorithm are still preserved, however.

Our insight is that the estimate of the coefficient can be estimated by using only transforms: if for $x = \Psi\alpha$, and $y = F(x)$ we have transforms $\alpha = W(x)$ and $y = Vx = \Phi\alpha$, as well as inverse transforms $x = W^{-1}(\alpha)$, and $x = F^{-1}(y)$, then the projections are obtained by

$$\langle r_t, \phi_i \rangle = \langle r_t, V\psi_i \rangle = \langle V^H r_t, \psi_i \rangle = \langle F^{-1}(r_t), \psi_i \rangle = W_i(F^{-1}(r_t)), \tag{41}$$

where $W_i(x)$ denotes the i^{th} coefficient of the W transform of x.

In their work, Candès and Romberg advocate the use of Partial or Mutilated Fourier Transform as their measurement basis, guided by the physical acquisition of tomography signals. We now propose a similar method, denoted the Permuted Fast Fourier Transform (PFFT), as $y = F(x) = \text{FFT}_{1:M/2}(P(x))$, where $P(x)$ is a fixed permutation of the samples in the vector x, performed before the truncated Fast Fourier Transform $\text{FFT}_{1:M/2}$ is applied, in which only the first $M/2$ Fourier Transform coefficients are kept - giving us M measurements from the signal by counting the real and imaginary parts of the Fourier Transform coefficients as separate measurements. It is easy to see that we can inverse transform from y to x is $x = F^{-1}(y) = P^{-1}(\text{IFFT}_N(y))$, where $\text{IFFT}_N(y)$ is the inverse Fast Fourier Transform of y zero-padded to length N, and $P^{-1}(x)$ is the inverse permutation of the samples in x, such that $F^{-1}(F(x)) = x$. Other transforms can be used together with the permutation technique, such as the Discrete Cosine Transform.

Figures 8 and 9 show the performance of the algorithm for several test cases. We use the test signals with length 8192 points and reconstruct noiseless and noisy versions of the signals using the complex wavelet basis as the original basis and using the Permuted Discrete Cosine Transform to obtain the measurements. We also apply the methodology to the 256×256 pixel images Lena and Wet-Paint, where the two-dimensional PFFT and a real wavelet transform (Daubechies-8) is used instead. Once again, a Basis Pursuit implementation is not feasible for these examples, due both to the computational complexity and storage requirements for the relevant bases — although we expect that these Projection-based methods can be implemented in BP as well.

C. Sparse Approximation in Overcomplete Wavelet Bases

The algorithms shown here can also be applied in the normal sparse approximation environment, in which we have signals that are sparse in an overcomplete wavelet-like basis but for which there is no unique representation. The Tree Matching Pursuit algorithms shown here should effectively and accurately find a sparse representation of the signal, something that cannot be achieved using Basis Pursuit - since all coefficients are evaluated at every iteration - while requiring less computational work due to the tree-search structure of the algorithm. Some bases that follow this model are complex wavelets [1] and curvelets [2].

D. Matching Pursuit as a Preprocessing Step

The Basis Pursuit algorithm is solved through interior point methods that admit an initial estimate of the solution; the distance between the initial estimate and the actual solution will affect the convergence time of the algorithm. Since the convergence rate of the Matching Pursuit algorithm decays exponentially, we can expect an ideal point in time where it is convenient to switch from Matching Pursuit to Basis Pursuit, where the estimate being given by the MP algorithm is used as the initial estimate for the BP algorithm, causing it to converge in a fast manner and obtaining an optimally sparse solution.

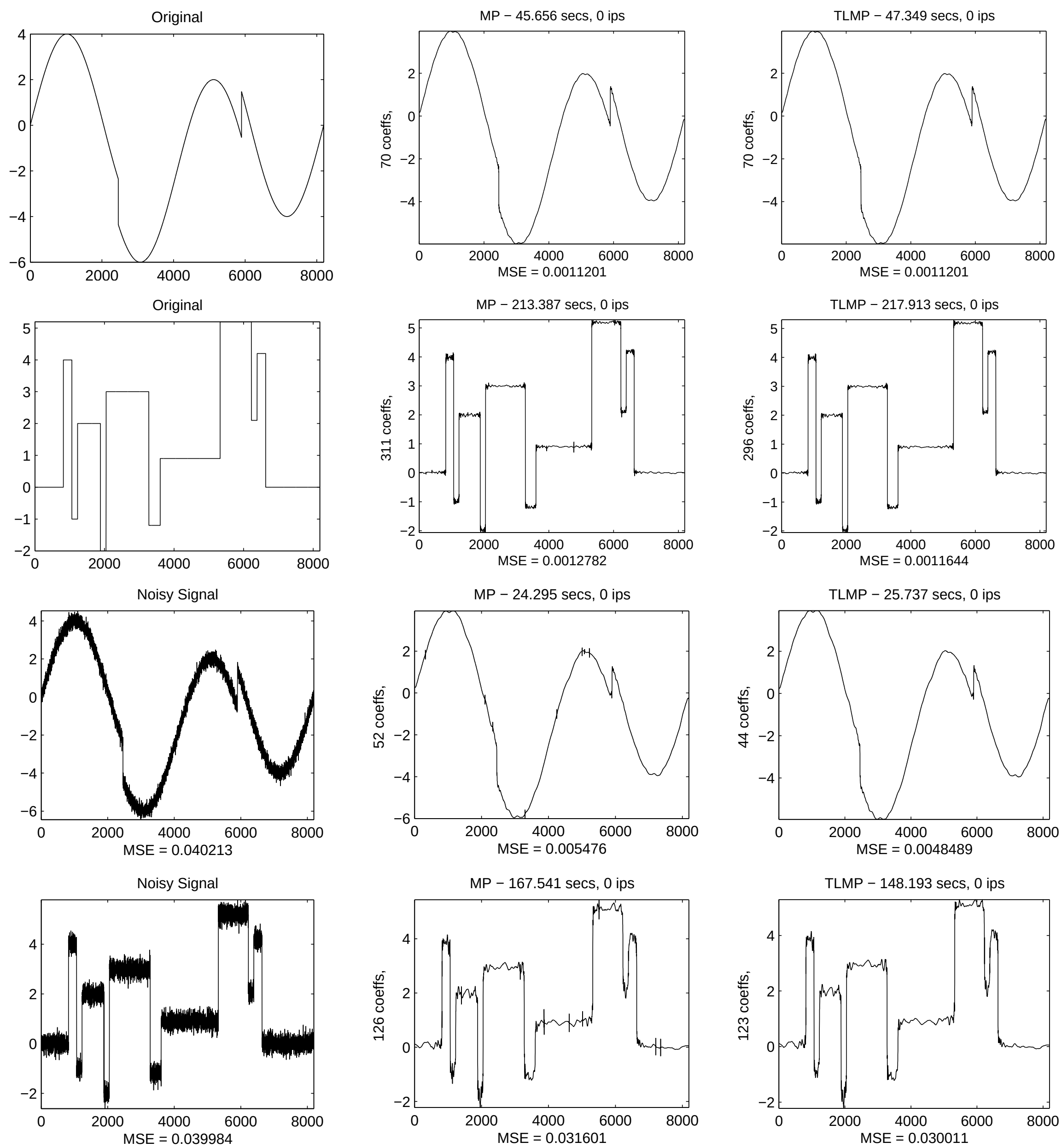

Fig. 8. Performance of matching pursuit algorithms on reconstruction of one-dimensional signals using transforms; no inner products appear in this methodology. Top two rows: reconstruction from noiseless measurements; bottom two rows: reconstruction from noisy measurements, $\epsilon = 0.2$.

E. Reservations on Matching Pursuit

Previous work on sparse approximations has shown that the Matching Pursuit algorithm has some flaws. DeVore and Temlyakov [14] show that the Matching Pursuit algorithm output may have a high level of error ($O(m^{-1/2})$) for a specific dictionary of vectors, Tropp [25] shows similar results, and establishes measures of goodness for dictionaries that are similar to those already established for CS matrices [7, 18].

VII. Related Work

In the published work on Compressed Sensing that has been released in the past year, all of the authors implement reconstruction through Basis Pursuit except for a few cases. Candès and Romberg [6] proposed a method based on projection onto convex sets which requires side information in the form of the ℓ_1 norm of either the compressed signal or its wavelet coefficient vector for each of the available scales. The authors

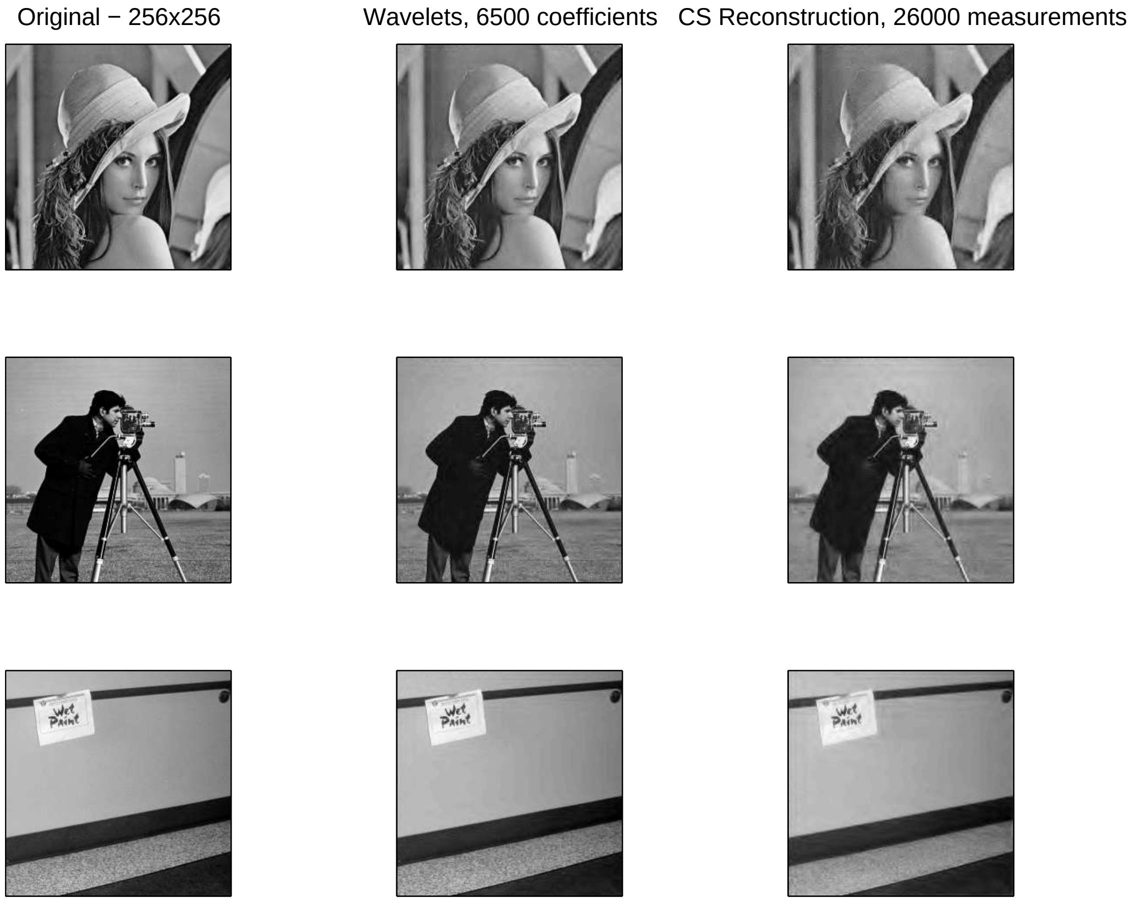

Fig. 9. Performance of matching pursuit algorithms on reconstruction of images using transforms. Left column: original images; middle column: sparse approximations; right column: reconstructions

recognize that this side information is not easily obtainable in real-world cases in which the signal is not known a-priori, and obtaining the norms would defeat the benefits of Compressed Sensing on actual sensors. The approach is still useful for universal coding, compression and encryption scenarios. For cases where this side information is not available, the authors reformulate their reconstruction algorithm using Lagrange Multipliers and test it on tomography image reconstruction with good results. We have found that by performing reconstruction using Matching Pursuit, a quality loss of about 3 dB in PSNR of the reconstructed images is observed.

Tsaig and Donoho [19] propose adaptations of the Basis Pursuit algorithm for signals that are sparse in the wavelet domain by performing two-gender hybrid CS by separately compressing the scaling and wavelet coefficients, or by performing multiscale CS by separately compressing the coefficients at each one of the scales. Once again, this method is not feasible in real-world cases, since the calculation of per-band or per-gender coefficients would also defeat the same purpose of CS as the method previously mentioned.

Haupt and Nowak [22] recently proposed an EM-based algorithm to recover the signal from random projections, and also offer bounds for the distortion in the recovered signal; no examples were given for rate of convergence or distortion in the reconstruction.

Finally, Tropp and Gilbert [26] demonstrated that Orthogonal Matching Pursuit can reliably recover a signal using an oversampling factor of 8 estimated empirically; one of the advantages of the OMP reconstruction is that when the signal is successfully recovered, at most K iterations of the algorithm are required for the recovery, and the residual becomes null at that time. This also gives a convenient indication of the success of the recovery algorithm.

Some previous work also exists on implementations of Matching Pursuit using tree structures. Cotter and

TABLE I

COMPUTATIONAL COMPLEXITY OF PURSUIT ALGORITHMS. K: NUMBER OF MEASUREMENTS; N: SIGNAL DIMENSION; I: CONVERGENCE FACTOR

Algorithm	**Complexity**
Basis Pursuit	$O(N^3)$
Matching Pursuit	$O(KNI)$
Orthogonal Matching Pursuit	$O(K^2N)$
Tree Matching Pursuit	$O(K^2I)$
Tree Orthogonal Matching Pursuit	$O(K^3)$

Rao [12] propose a Matching Pursuit algorithm in which for each iteration the best-matching k atoms are selected and a residual is generated for each atom; each one of these residuals is used, one at a time, for the next iteration. This process forms a tree in which the nodes at level l represent the best k^l approximations of the signal; the tree can be pruned to k nodes per level as the iterations progress.

De Vleeschouwer and Macq [10] propose a Matching Pursuit algorithm in which the dictionary atoms correspond to wavelet functions obtained from a filterbank to acquire sparse representations of displaced frame difference images, used in MPEG video coding; at each iteration, the remainder is first partitioned into blocks and the block with highest energy is used for the iteration. The computational advantage of the tree structure in this case is limited to the fast calculation of the coefficient estimates through the filterbank since all atoms are evaluated at every iteration of the algorithm.

Jost, Vandergheynst and Frossard [23] propose an algorithm similar in nature to ours: the atoms in the dictionary are clustered into groups of size k by similarity, such that atoms that would have large projection magnitudes simultaneously are clustered together, and atoms that do not hold this property are clustered apart. At each level, available clusters are clustered again until less than k clusters exist. Each cluster is assigned a representative cluster atom, which is a linear combination of the atoms in the cluster. This builds a k-children tree where the cluster atoms are represented by interior nodes and the dictionary atoms are represented by the leaves. The reconstruction then proceeds by selecting the cluster at the first level of the tree that yields maximum projection magnitude, and performs projections against the cluster atoms for the cluster's children. This procedure goes on until we select a leaf, or dictionary atom, at which point the corresponding coefficient and the residual are updated normally. This methodology speeds up the search for the best atom in the dictionary, but requires that these atoms can be clustered as described; since the atoms used in CS reconstruction are essentially random, they cannot be clustered, and such clustering will almost surely not follow the structure of the wavelet functions.

VIII. CONCLUSIONS AND FURTHER WORK

Each one of the algorithms presented in this paper has advantages and disadvantages; while the Basis Pursuit algorithm is reliable, it is computationally unfeasible for moderately-sized signals. On the other hand, the Matching Pursuit algorithm is computationally simple, but its unbounded number of iterations required for convergence might make it undesirable in certain applications. The Orthogonal Matching Pursuit algorithm, while being bounded in the number of iterations required, has a much higher computational complexity due to the orthogonalization of the basis vectors required at each iteration. The Tree Matching Pursuit algorithm here proposed is robust to noise, is computationally feasible for a larger domain of signals, and exploits the inherent structure in piecewise smooth signals to achieve faster, better reconstruction. Table I shows the computational complexity of the different algorithms. Our only reservation is once again on its rate of convergence.

Future research directions include finding measurement bases that will allow for faster reconstruction, as well as algorithms that will allow for scaled reconstruction as more CS measurements are received through a communications channel. It is worth mentioning that probabilistic formulations of the parent-child relationship in the wavelet coefficient tree, such as that proposed in [9] may benefit the performance of the algorithm, while requiring training of the model on the class of signals that the compression system will observe.

IX. Acknowledgements

The author would like to thank the project advisors, Richard Baraniuk and Hyeokho Choi, as well as Michael Wakin and Dror Baron, for the help and guidance throughout the duration of this project.

References

[1] R. Baraniuk, N. Kingsbury and I. Selesnick, "The Dual-Tree Complex Wavelet Transform - A Coherent Framework for Multiscale Signal and Image Processing," Submitted to *IEEE Signal Processing Magazine*, 2004.

[2] E. Candès and D. Donoho, "Curvelets - A Surprisingly Effective Nonadaptive Representation for Objects with Edges ," *Curves and Surfaces*, L. L. Schumaker et al. (eds), Vanderbilt University Press, Nashville, TN.

[3] E. Candès, J. Romberg and T. Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information," Submitted to *IEEE Transactions on Information Theory*, 2004.

[4] E. Candès and T. Tao, "Near Optimal Signal Recovery from Random Projections And Universal Encoding Strategies," Submitted to *IEEE Transactions on Information Theory*, 2004.

[5] E. Candès and T. Tao, "Decoding by Linear Programming," Preprint, 2004.

[6] E. Candès and J. Romberg, "Practical Signal Recovery from Random Projections," Submitted to *IEEE Transactions on Signal Processing*, 2005.

[7] E. Candès, J. Romberg and T. Tao, "Stable Signal Recovery from Incomplete and Inaccurate Measurements", Submitted to *Communications on Pure and Applied Mathematics*, 2005.

[8] E. Candès and J. Romberg, "Quantitative Robust Uncertainty Principles and Optimally Sparse Decompositions", Submitted to *Foundations of Computational Mathematics*, 2004.

[9] H. Choi, J. Romberg, R. Baraniuk and N. Kingsbury, "Hidden Markov Tree Modeling of Complex Wavelet Transforms," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, pp. 133–136, 2000.

[10] C. De Vleeschouwer and B. Macq, "Subband Dictionaries for Low-Cost Matching Pursuits of Video Residues", *IEEE Transactions on Circuits and Systems For Video Technology*, vol. 9, no. 7, pp. 984–993, 1999.

[11] S. Chen, D. Donoho and M. Saunders, "Atomic Decomposition by Basis Pursuit", *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33-61, 1998.

[12] S. Cotter and B. Rao, "Application of Tree-Based Searches to Matching Pursuit," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 6, pp. 3933–3936, 2001.

[13] G. Davis, S. Mallat and M. Avellaneda, "Adaptive Greedy Approximations," *Journal of Constructive Approximation*, vol. 13, pp. 57–98, 1997.

[14] R. DeVore and V. Temlyakov, "Some Remarks on Greedy Algorithms," *Advances on Computational Mathematics*, no. 5, pp. 173–187, 1996.

[15] D. Donoho, "Denoising by Soft Thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613-627, 1995

[16] D. Donoho, "Wedgelets: Nearly Minimax Estimation of Edges," Technical Report, Department of Statistics, Stanford University, 1997.

[17] D. Donoho, "For Most Large Underdetermined Systems Of Linear Equations, The Minimal L1-Norm Solution Is Also The Sparsest Solution," Preprint, 2004.

[18] D. Donoho, "Compressed Sensing," Preprint, 2004.

[19] D. Donoho and Y. Tsaig, "Extensions of Compressed Sensing," Preprint, 2004.

[20] S. Mallat, *A Wavelet Tour of Signal Processing.* San Diego, CA, USA: Academic Press, 1999.

[21] S. Mallat and Z. Zhang, "Matching Pursuits with Time-Frequency Dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[22] J. Haupt and R. Nowak, "Signal Reconstruction from Noisy Random Projections," Submitted to *IEEE Transactions on Information Theory*, 2005.

[23] P. Jost, P. Vandergheynst and P. Frossard, "Tree-Based Pursuit", EPFL Signal Processing Institute, Technical Report 003, 2004.

[24] Y. C. Pati, R. Rezaifar and P.S. Krishnaprasad, "Orthogonal Matching Pursuit: Recursive Function Approximation With Applications To Wavelet Decomposition," *Proceedings of the 27th Asilomar Conference on Signals, Systems and Computation*, November 2003.

[25] J. Tropp, "Greed is Good: Algorithmic Results for Sparse Approximation", *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[26] J. Tropp and A. Gilbert, "Signal Recovery from Partial Information Via Orthogonal Matching Pursuit", Preprint, 2005.