# Fast Reconstruction of Piecewise Smooth Signals from Incoherent Projections

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

***Abstract*— The Compressed Sensing framework aims to recover a sparse signal from a small set of projections onto random vectors; the problem reduces to searching for a sparse approximation of this measurement vector. Conventional solutions involve linear programming or greedy algorithms and can be computationally expensive. These techniques are generic, however, and assume no structure in the signal aside from sparsity. In this paper, we design an algorithm that enables fast recovery of *piecewise smooth signals*, sparse signals that have a distinct "connected tree" structure in the wavelet domain. Our Tree Matching Pursuit (TMP) algorithm significantly reduces the search space of the traditional Matching Pursuit greedy algorithm, resulting in a substantial decrease in computational complexity for recovering piecewise smooth signals. An additional advantage of TMP is that it performs an implicit regularization to combat noise in the reconstruction. TMP also applies to the more general case of "incoherent" measurement vectors.**

## I. Introduction

The ongoing evolution and proliferation of sensing devices has sparked an explosion in the amount of information available to scientists in many disciplines and to consumers in general. This phenomenon has fueled new research in the fields of compression and coding, which enable compact representations and rapid 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 transmitted or 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 framework of *Compressed Sensing* (CS) [1, 2] enables a reduction in the communication and computation costs at the sensor. The key idea in CS is that for a signal that is compressible (i.e., has a *sparse* representation in an orthonormal basis or tight frame), a small number of *random projections* of that signal contains sufficient information for *exact* reconstruction. Thus, the process of determining the significant sparse basis vectors can be bypassed entirely at the sensor/encoder. Signal reconstruction in this framework consists of solving a sparse approximation problem: find the sparsest signal that explains the acquired measurements. Techniques originally proposed for finding sparse approximations from redundant dictionaries (such as Basis Pursuit (BP) [3], Matching Pursuit (MP) [4], and Orthogonal Matching Pursuit (OMP) [5]) have also been adopted for the CS framework [1, 2, 6]. BP employs linear programming and offers good performance but suffers from high computational complexity. MP provides a low-complexity alternative to BP but requires an unbounded number of iterations for convergence. OMP converges in a fixed number of iterations but requires the added complexity of the dictionary orthogonalization at each step.

These algorithms for reconstruction are generic, in the sense that they do not exploit any structure (aside from sparsity) that may exist in the sensed signals. An important subclass of sparse signals, however, is the class of *piecewise smooth* signals — many punctuated real-world phenomena give rise to such signals [7]. The wavelet transform of a piecewise smooth signal yields a sparse, *structured* representation of signals in this class: the largest coefficients tend to form a connected subtree of the wavelet coefficient tree. While other methods have been proposed for fast reconstruction of wavelet-sparse signals [8, 9], these methods do not fully exploit this connectedness property.

In this paper, we propose a new *Tree Matching Pursuit* (TMP) algorithm for fast reconstruction of piecewise smooth signals that refines MP to exploit the connectedness of the significant wavelet coefficients. TMP features a lower computational complexity than other generic algorithms. We study the performance of TMP and propose adaptations and extensions for applications such as denoising. Section II describes the necessary background in CS theory; Section III describes the basic framework for the TMP algorithm; and Section IV describes some important adaptations to enhance the performance. Section V suggests some extensions to the algorithm and offers conclusions.

## II. Compressed Sensing Background

Assume that we acquire an $N$-sample signal $x$ for which a basis $\Psi$ having columns $[\psi_1|\psi_2|\dots|\psi_N]$ provides a $K$-sparse representation; that is, we can write

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

where $\{n_i\}_i \subset \{1, \dots, N\}$ are the vector indices, $\{\alpha_i\}$ are the transform coefficients, and $K \ll N$. In matrix form, this is expressed as $x = \Psi\alpha$. (The theory is also easily extended to the case where $\Psi$ is a tight frame.)

The standard procedure to compress such signals is then to (*i*) acquire the full $N$-point signal $x$; (*ii*) compute the complete set of transform coefficients $\alpha$; (*iii*) locate the (few) largest, significant coefficients and discard the (many) small coefficients; (*iv*) encode the *values and locations* of the largest coefficients. This procedure is inherently inefficient, in that the encoder must compute *all* $N$ transform coefficients $\{\alpha_i\}$, even though it will discard most of them.

This raises a simple question: For a given signal, is it possible to directly estimate the set of large $\alpha_i$'s that will not be discarded? While this seems impossible, recent work in CS [1, 2] has shown that about $K$ random projections contain enough information to reconstruct sparse signals. In the CS theory we do not measure the large $\alpha_i$'s directly. Rather, we measure the projections $y_i = \langle x, \sigma_i \rangle$ of the signal onto a *second set* of $M$ vectors $\Sigma = \left[\sigma_1^T|\sigma_2^T|\dots|\sigma_M^T\right]^T$ which we call the *measurement vectors*. The CS theory tells us that when certain conditions hold, namely that the $\Psi$ basis set does not provide sparse representations of the elements of $\{\sigma_i\}$ (a condition known as *incoherence* of the two bases), then it is indeed possible to recover the set of large $\alpha_i$'s from a similarly sized set of measurements $\{y_i\}$. This incoherence property holds for many pairs of bases, including for example, delta spikes and the sine waves of the Fourier basis, or sine waves and wavelets; in particular a random basis tends to be incoherent with any fixed basis. The algorithms in this paper apply not only to random but also to more general incoherent measurement

This work was supported by NSF, ONR, AFOSR, and the Texas Instruments Leadership University Program.

The authors are with the Department of Electrical and Computer Engineering at Rice University.

Email: {duarte, wakin, richb}@rice.edu. Web: dsp.rice.edu/cs

matrices. The CS theory also applies to redundant representations such as curvelets [10], which are well-suited to match geometrical edge structures in 2D images.

Recovery of the set of significant (large) coefficients $\{\alpha_i\}$ is achieved using algorithms for sparse approximation. Given measurements $y \in \mathbb{R}^M$, we solve for the sparsest approximation of $y$ in the dictionary defined by the columns of $\Phi = \Sigma\Psi = [\phi_1|\dots|\phi_N]$. (Equivalently, we seek the sparsest $\alpha$ that explains the measurements $y$.) Because of the incoherence between the original ($\Psi$) and the measurement ($\Sigma$) bases, if the original signal $x$ is sparse in the original basis $\Psi$, then no other set of sparse coefficients can exist to explain the measurements $y$. There is no free lunch, however; according to the theory, the observation set must have size $M = |y| \geq CK$, where $C$ is dependent on the matrix $\Phi$. Commonly quoted as $C = O(\log(N))$ [2], we have found that $C \approx \log_2(1 + N/K)$ provides a useful rule-of-thumb, and so $M$ may still be much smaller than $N$.

Techniques originally proposed for finding sparse approximations from redundant dictionaries have been adopted for CS [1, 2, 6]. BP minimizes $\|\alpha\|_1$ with the constraint $y = \Phi\alpha$ that the coefficients must explain the signal. Finding this $\alpha$ involves solving a large linear program, however, which requires $O(N^3 \log(N))$ operations. MP is a computationally simple iterative greedy algorithm: at each iteration, it selects the atom that explains most of the energy in the signal. MP runs as follows:

1) Initialize the residual $r_0 = y$ and the approximation $\widehat{\alpha} = 0$, $\widehat{\alpha} \in \mathbb{R}^N$. Set $t = 1$.
2) Select the dictionary vector that maximizes the projection of the residual
$$n_t = \arg \max_{i \in \{1,\dots,N\}} |\langle r_{t-1}, \phi_i \rangle| / \|\phi_i\|.$$
3) Update the estimate of the coefficient for the selected vector and the residual
$$\begin{aligned} r_t &= r_{t-1} - \langle r_{t-1}, \phi_{n_t} \rangle \phi_{n_t} / \|\phi_{n_t}\|^2, \\ \widehat{\alpha}_{n_t} &= \widehat{\alpha}_{n_t} + \langle r_{t-1}, \phi_{n_t} \rangle / \|\phi_{n_t}\|^2. \end{aligned}$$
4) If $\|r_t\|_2 > \epsilon \|y\|_2$, then increment $t$ and go to Step 2; otherwise, terminate.

The convergence criterion is the minimum proportion $\epsilon$ of energy that can be left in the residual. MP has been proven to achieve an accurate decomposition of the signal as a linear combination of dictionary vectors, although the required number of iterations is unbounded [7]. Thus, the complexity of MP is approximately $O(CKNI)$, where $I$ corresponds to the unknown number of iterations. OMP limits the number of MP iterations by orthogonalizing the non-selected dictionary vectors against those already selected. This allows the algorithm to converge in at most $M$ iterations, but requires the added computational cost of the orthogonalization at each iteration; the total complexity is $O(CK^2N)$.

Though effective for generic CS reconstruction, BP, MP, and OMP do not exploit any structure (aside from sparsity) that may be present in the sensed signal. This can lead both to inefficiencies in the recovery algorithm and to artifacts in the reconstructed signals [9]. By considering the additional wavelet-domain structure of piecewise smooth signals, we hope both to improve the efficiency of these algorithms and to benefit from the regularization implicit in the reconstruction.

## III. Tree Matching Pursuit

### A. Multiscale Wavelet Structure

We aim to customize existing reconstruction algorithms for piecewise smooth signals, which have sparse, structured representations in the wavelet domain. Without loss of generality, we focus on 1D signals, though similar arguments apply for 2D images in the wavelet or curvelet [10] domains. In a typical 1D wavelet transform, each coefficient at scale $j \in \{1, \dots, \log_2(N)\}$ describes a portion of the signal of size $O(2^{-j})$. 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 $j - 1$. Notions of descendants and ancestors are propagated naturally at higher and lower scales, respectively.

Due to the analysis properties of wavelets, coefficient values tend to perpetuate 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. Wavelet coefficients also have decaying magnitudes as the scale decreases [7]. This causes the significant wavelet coefficients of piecewise smooth signals to form a connected subtree within the wavelet binary tree.

Not surprisingly, we observe for piecewise smooth signals that MP tends to select wavelet coefficients located near the top of the tree first and then continues selecting down the tree, effectively building a connected tree that contains the most significant coefficients from the top down. This suggests that it may not be necessary for the MP algorithm to check *all* possible coefficients at each stage. Rather, the next most important coefficient at each stage is likely to be among the children of the currently selected coefficients.

We must refine this heuristic, however, to obtain an effective algorithm. In particular, for real world piecewise smooth signals, the nonzero coefficients generally do not form a perfect connected subtree. The reasons for this are twofold. First, since wavelets are bandpass functions, wavelet coefficients oscillate positive and negative around singularities [11]. Second, due to the linearity of the wavelet transform, two or more singularities in the signal may cause destructive interference among large wavelet coefficients. 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. We can still define a connected subtree that contains all of the nonzero valued coefficients, however, which will contain some *gaps* consisting of sequences of small or zero values. Our proposed algorithm features a parameter designed to address this complication.

### B. Algorithms

Tree Matching Pursuit (TMP) considers only a subset of the basis vectors at each iteration, and then expands that set as significant coefficients are found. Define two sets of coefficients $S_t$ and $C_t$, which contain the set of *selected* 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 will be evaluated at the next iteration). These sets are initialized as
$$S_t = \emptyset, \quad C_t = \{1\} \cup D_b(1), \tag{1}$$
where the $b$-depth set of descendants $D_b(i)$ is the set of coefficients within $b$ levels below coefficient $i$ in the wavelet tree.

At each iteration, we search for the dictionary vector $\phi_i$ in $S_t \cup C_t$ that yields the maximum inner product with the current residual; if the selected vector comes from $C_t$, then that coefficient $i$ (and its ancestors, denoted $A(i)$) is moved to the set of selected coefficients $S_t$ and removed from $C_t$, and the descendant set $D_b(i)$ is added to $C_t$. For *b-Tree Matching Pursuit* ($b$-TMP) and *b-Tree Orthogonal Matching Pursuit* ($b$-TOMP) we adapt Step 2 of MP/OMP as follows:

TABLE I

*Computational complexity of CS algorithms. $N$ = signal length; $K$ = signal sparsity (typically $\ll N$); $I$ = convergence factor, $C$ = oversampling factor; $B$ = TMP band width.*

| Algorithm | **BP** | **MP** | **OMP** | $b$-**TMP** | $b$-**TOMP** |
|---|---|---|---|---|---|
| Complexity | $O(N^3 \log(N))$ | $O(CKNI)$ | $O(CK^2N)$ | $O(2^bCK^2I)$ | $O(2^bCK^3)$ |

2) Among the candidates, select the dictionary vector that maximizes the projection of the residual

$$n_t = \arg \max_{i \in S_{t-1} \cup C_{t-1}} |\langle r_{t-1}, \phi_i \rangle| / \|\phi_i\|.$$

If $n_t \in C_{t-1}$, then update the sets

$$\begin{aligned} S_t &= S_{t-1} \cup n_t \cup A(n_t), \\ C_t &= C_{t-1} \setminus (n_t \cup A(n_t)) \cup D_b(n_t). \end{aligned}$$

While the existence of gaps in the wavelet subtree containing the set of meaningful coefficients will hamper the ability to reach some nonzero coefficients, the parameter $b$ enables us to define a "lookahead" band of candidate coefficients wide enough that each possible gap is contained in the band. 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 $D_i$. However, the computational complexities of $b$-TMP and $b$-TOMP, given by $O(2^bCK^2I)$ and $O(2^bCK^3)$, respectively, will increase with $b$. For moderate $b$ both still represent a significant improvement over their generic counterparts, and $b$-TOMP improves upon BP by a factor of $O((N/K)^3)$; Table 1 summarizes the computational complexity of the various algorithms.

TMP is robust to noise (see Section V-A) and computationally attractive compared to the generic algorithms. Figure 1 (top right) shows that for $b = 2$ TMP fails due to a small coefficient inside the subtree containing the meaningful coefficients. However, when $b = 3$ (bottom left), the band is wide enough to bypass this gap and reconstruct the signal satisfactorily with approximately half as many inner products as required for MP (top middle). The $b$-TOMP algorithm performs comparably.

## IV. Adaptations of Tree Matching Pursuit

In the previous section we proposed a $b$-level lookahead band to compensate for possible gaps among the significant wavelet coefficients. In this section we propose alternative techniques to alleviate the problems caused by gaps.

### A. Complex Wavelet Transform

By using a *complex wavelet transform* ($\mathbb{C}$WT) [11], we can avoid some of the pitfalls of the standard real wavelet transform. The $\mathbb{C}$WT shares the same binary tree structure as the real wavelet transform, but the wavelet functions are complex-valued

$$\psi_c(t) = \psi_r(t) + j\psi_i(t).$$

The component $\psi_r(t)$ is real and even, while $j\psi_i(t)$ is imaginary and odd; they form an approximate Hilbert transform pair. The $\mathbb{C}$WT transform can be easily implemented using a dual-tree structure, where we simply compute two *real* wavelet transforms ($\psi_r(t)$ and $\psi_i(t)$) in parallel, obtaining the sequences of coefficients $\alpha_r$ and $\alpha_i$. The complex wavelet coefficients are then defined as $\alpha_c = \alpha_r + j\alpha_i$.

Note that either the real or the imaginary part of the wavelet coefficients would suffice to reconstruct the signal; however, the dual representation establishes a strong coherency among the complex magnitudes. Due to the Hilbert transform relationship between the real and imaginary wavelets, when a discontinuity is present and the real (or imaginary) wavelet coefficient is small, the imaginary (or real) wavelet coefficient is large [11]. Thus, the shift-sensitivity of the standard real-wavelet transform is alleviated. As such, when the $b$-TMP algorithm is implemented using the $\mathbb{C}$WT, a much smaller band will be necessary for efficient reconstruction. Figure 1 (bottom right) shows the approximate recovery of Blocks using a band of width 1. Unfortunately, complex coefficients can still interfere destructively, suggesting $b$ slightly greater than 1 as a conservative choice.

### B. Random Lookahead

We propose a second modification to TMP that can be applied to both the real and $\mathbb{C}$WT variants. The modification involves a probabilistic definition of the candidate set $C_t$ at each iteration, based on the Hidden Markov Tree wavelet model [12]. In this model, two states exist for each coefficient ($L$ and $S$ for large and small), and the probabilities of transition between states for pairs of parent and child coefficients ($\pi_{SS}$, $\pi_{SL}$, $\pi_{LS}$, and $\pi_{LL}$) are defined such that connected strings of $L$-state coefficients are likely. In our matching pursuit algorithm, we label the coefficients selected at each iteration as large ($L$), i.e., $P(n_t = L) = 1$ , and calculate the conditional probability that each of its descendants is in the $L$ state. During the candidate set selection, for each leaf $i$ in the subtree containing the set of selected coefficients, we select a random sample of descendants $D_{\mathrm{HMT}}(i)$ according to the probability that each descendant is in the large state, where for a coefficient $j$ that is $d$ levels below coefficient $i$, $P(j = L) = (\pi_{SS})^d$. Thus, coefficients with higher estimates of $P(j = L)$ are more likely to be selected in the candidate set.

We amend this formulation slightly for easier computation by choosing a constant $s$ and then constructing $D_{\mathrm{HMT}}(i)$ by randomly selecting $s$ descendant coefficients from each scale below $i$. We denote by Markov Tree Matching Pursuit (MTMP) the TMP algorithm that uses this descendant set in the updates. It is worth noting that by setting $s = 2^b$, the descendants selected by the $s$-MTMP algorithm contain the set of descendants selected by the original $b$-TMP algorithm. The algorithm should enable recovery of signals having large gaps inside the set of meaningful coefficients, while keeping the number of coefficients in the candidate sets relatively small. In Figure 1 (bottom middle), we see that by using the random lookahead with $s = 4$, the significant coefficients below the gap are recovered.

## V. Extensions and Conclusions

### A. Regularization and Denoising

Most current implementations of reconstruction in CS are not robust to the injection of noise in the measurements; the noise is either accounted for in the distortion estimation [13] or in the reconstruction constraints [8]. When the signal is sparse in the wavelet basis, we can effectively perform denoising by thresholding [14] by varying the convergence criterion $\epsilon$ as a function of the signal-to-noise ratio. We then identify only the most significant coefficients using the MP or $b$-TMP algorithm and effectively threshold their values at the reconstruction. An alternative EM-based algorithm for recovery from noisy measurements was described in [15] along with bounds for the distortion in the recovered signal.

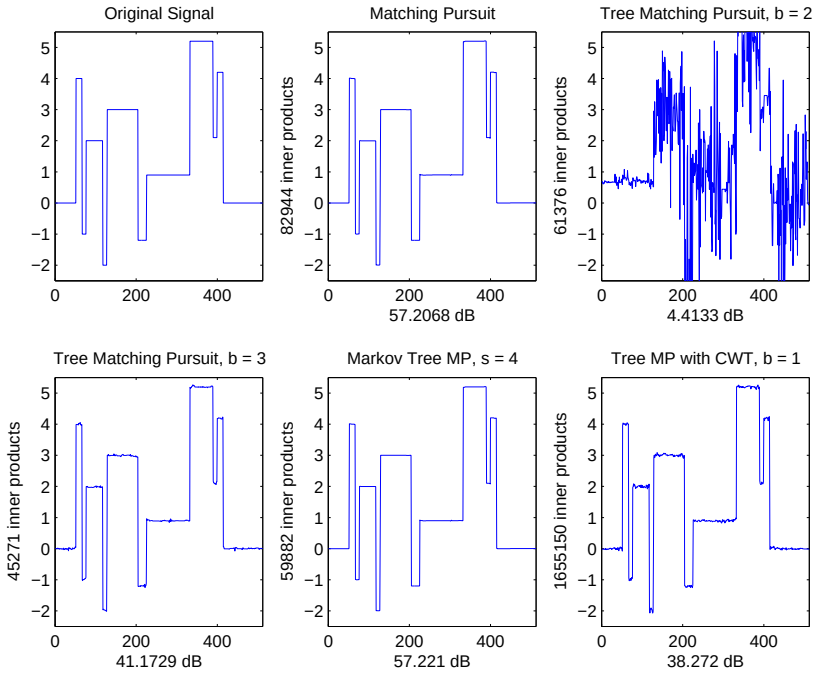


Fig. 1. *CS reconstructions using several different algorithms. Axis labels indicate reconstruction quality and computational complexity. Top left: original signal. Top middle: MP. Top right: $b$-TMP with $b = 2$; the band is too small to cover the gaps in the wavelet coefficients and reconstruction fails. Bottom left: $b$-TMP with $b = 3$; the band is large enough to bypass the gap, leading to correct reconstruction. Bottom middle: $s$-MTMP, $s = 4$. Bottom right: $b$-TMP with the $\mathbb{C}$WT, $b = 1$. Both of these modifications yield approximate reconstruction.*

CS reconstructions using the standard methods (BP and MP) also typically suffer from artifacts since the energy of the signal is not discriminated by band [9]; 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 that resemble small-scale noise. By giving preference to the coarsest coefficients over the finest, the TMP algorithms help mitigate this effect during reconstruction.

### *B. Matching Pursuit using Transforms*

Although MP enables computationally feasible reconstruction for higher-dimensional signals than BP, we still pay the price of storage of the matrix $\Phi = \Sigma\Psi$, which is of size $M \times N = O(KN \log N)$. For high-dimensional signals a large amount of memory is needed to perform MP using this matrix. We do have an alternative: if we can replace the projection of a vector using the matrix product with a transform, then we do not need to store these matrices in memory. Many useful transforms have computational complexity lower than multiplication by the corresponding matrix operator. However, since all coefficients are calculated at once, we lose the computational gain from the TMP algorithms. The denoising properties of these algorithms are still preserved, however.

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 propose a similar method, denoted the Permuted Fast Fourier Transform (PFFT), as $y = F(x) = \mathrm{FFT}_{1:M/2}(\mathcal{P}(x))$, where $\mathcal{P}(x)$ is a fixed permutation of the samples in the vector $x$, performed before the truncated transform $\mathrm{FFT}_{1:M/2}$ is applied, in which only the first $M/2$ Fourier coefficients are kept — giving us $M$ measurements from the signal by counting the real and imaginary parts of the coefficients as separate measurements. The PFFT measurement vectors tend to be incoherent with any standard basis, including the (nonpermuted) Fourier basis. Other transforms can be used together with the permutation technique, such as the Discrete Cosine Transform.

*Acknowledgments:* We thank Emmanuel Candès, Justin Romberg, Joel Tropp, Hyeokho Choi, Dror Baron, and Shriram Sarvotham for helpful discussions.

## REFERENCES

[1] D. Donoho, "Compressed sensing," 2004, Preprint.
[2] E. Candès and T. Tao, "Near optimal signal recovery from random projections and universal encoding strategies," *IEEE Trans. Inform. Theory*, 2004, Submitted.
[3] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, 1998.
[4] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. on Signal Proc.*, vol. 41, no. 12, 1993.
[5] Y. C. Pati, R. Rezaifar, and P.S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals, Systems and Computation*, Nov. 1993.
[6] J. Tropp and A. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," 2005, Preprint.
[7] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, USA, 1999.
[8] D. Donoho and Y. Tsaig, "Extensions of compressed sensing," 2004, Preprint.
[9] E. Candès and J. Romberg, "Practical signal recovery from random projections," *IEEE Trans. Signal Proc.*, 2005, Submitted.
[10] E. Candès and D. Donoho, "Curvelets - A surprisingly effective nonadaptive representation for objects with edges," *Curves and Surfaces*, 1999.
[11] R. Baraniuk, N. Kingsbury, and I. Selesnick, "The dual-tree complex wavelet transform - A coherent framework for multiscale signal and image processing," *IEEE Signal Proc. Mag.*, 2005, To appear.
[12] H. Choi, J. Romberg, R. Baraniuk, and N. Kingsbury, "Hidden Markov tree modeling of complex wavelet transforms," in *Proc. ICASSP.*, 2000.
[13] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Applied Math.*, 2005, Submitted.
[14] D. Donoho, "Denoising by soft thresholding," *IEEE Trans. on Inform. Theory*, vol. 41, no. 3, 1995.
[15] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. on Inform. Theory*, 2005, Submitted.