

TREE-BASED ORTHOGONAL MATCHING PURSUIT ALGORITHM FOR SIGNAL RECONSTRUCTION

Chinh La and Minh N. Do

Department of Electrical and Computer Engineering,
University of Illinois at Urbana Champaign.

ABSTRACT

Recent studies in linear inverse problems have recognized the sparse representation of unknown signal in a certain basis as an useful and effective prior information to solve those problems. In many multiscale bases (e.g. wavelets), signals of interest (e.g. piecewise-smooth signals) not only have few significant coefficients, but also those significant coefficients are well-organized in trees. We propose to exploit this sparse tree representation as additional prior information for linear inverse problems with limited numbers of measurements. In particular, our proposed algorithm named Tree-based Orthogonal Matching Pursuit (TOMP) is shown to provide significant better reconstruction compared to methods that only use sparse representation assumption.

Index Terms— Inverse problems, sparse representations, tree structures, greedy algorithms, Tree-based Orthogonal Matching Pursuit

1. INTRODUCTION

Recently, there have been a series of remarkable work showing that one can effectively exploit the knowledge that the unknown signal has **sparse representation** in a certain basis as a prior information for signal reconstruction from limited number of measurements [1, 2, 3]. Suppose we know that an unknown signal of length M can be well approximated using only K significant coefficients ($K \ll M$) in a certain basis. However, the locations of these K significant coefficients are unknown and depend on the signal. It has been shown in [1, 2, 3] that under certain conditions, with high probability the unknown signal can be reconstructed accurately using only $O(K \log M)$ *non-adaptive linear measurements*.

In many multiscale bases (e.g. wavelets), signals of interest (e.g. piecewise-smooth signals) not only have few significant coefficients, but also *those significant coefficients are well-organized in trees*. In fact, these embedded tree structures for significant wavelet coefficients have been used successfully in compression, modeling, and approximation tasks.

We propose to exploit the **sparse tree representation** as additional prior information for signal reconstruction with limited numbers of measurements. Intuitively, a general sparse representation with K coefficients can be described with $2K$ numbers: K for the values of the significant coefficients and another K for the locations of these coefficients. If the K significant coefficients are known to be organized in trees then the indexing cost is significantly reduced and hence the total description of the unknown signal. Exploiting this embedded tree structure in addition to the sparse representation prior in inverse problems would potentially lead to: (1) better reconstructed signals; (2) reconstruction using fewer measurements; and (3) faster reconstruction algorithms.

In this paper, we describe the *Tree-based Orthogonal Matching Pursuit* (TOMP) algorithm that was first presented in [4] for signal reconstruction using sparse tree representation. The idea of using sparse tree representation for signal reconstruction was independently developed in [5] with the main focus on obtaining fast reconstruction algorithm. Whereas, our focus is to exploit the sparse tree representation to obtain better reconstruction compared to the method like Orthogonal Matching Pursuit [3] and Basis Pursuit [1, 2] that only uses sparse representation assumption.

2. SIGNAL RECONSTRUCTION USING SPARSE TREE REPRESENTATION

For an unknown signal \mathbf{y} of length M , suppose that we can only make a limited number N of non-adaptive linear measurements ($N \ll M$):

$$\mathbf{M}\mathbf{y} = \mathbf{b}, \quad (1)$$

where \mathbf{M} is a fixed $N \times M$ measurement matrix, and \mathbf{b} is a length- N vector that contains the measured data. Our inverse problem is to reconstruct signal \mathbf{y} from \mathbf{b} . Suppose \mathbf{y} has basis expansion by an orthonormal transform matrix \mathbf{T} as

$$\mathbf{y} = \mathbf{T}\mathbf{x}. \quad (2)$$

Let $\mathbf{A} = \mathbf{M}\mathbf{T}$ which is also a known $N \times M$ matrix, then our inverse problem (1) becomes

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \quad (3)$$

and solving \mathbf{x} from (3) will give \mathbf{y} .

Since $N \ll M$, both (1) and (3) are under-determined systems. Thus in order to solve this inverse problem we need to exploit prior knowledge about signal \mathbf{y} . We consider the case where \mathbf{y} is a piecewise-smooth signal, which models typical signals of interest. Figure 1 shows an example of the wavelet representation of such a signal. We notice that the wavelet representation \mathbf{x} of a piecewise-smooth signal has the following two special properties.

- P1** *Vector \mathbf{x} has sparse structure; i.e. only few entries in \mathbf{x} are nonzero or significant.*
- P2** *Those significant entries of \mathbf{x} are well organized in a tree structure (see below).*

As can be seen from Figure 1, property **P2** is due to the fact that nonzero coefficients correspond to discontinuity points in the original signal s . In particular, we notice that *if a coefficient is nonzero or significant then its ancestors are likely nonzero or significant*. Exploiting this property **P2** is the key point of our proposed algorithm, whereas existing algorithms such as Basis Pursuit [1, 2] and Orthogonal Matching Pursuit [3] only employ **P1** to solve the inverse problem (3).

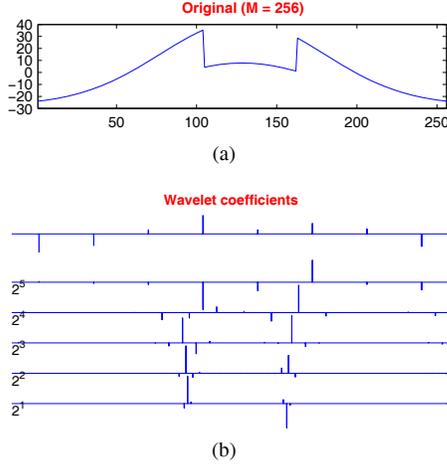


Fig. 1. Example of sparse tree representation: (a) original piecewise-smooth signal with $M = 256$ samples; (b) wavelet representation with about $K = 32$ significant coefficients.

3. ORTHOGONAL MATCHING PURSUIT (OMP)

As described in [3] OMP is an efficient greedy method to solve equation (3). In this method, the measurement matrix \mathbf{A} is considered as a dictionary with all columns \mathbf{a}_i 's of \mathbf{A} as atoms. Start with $\mathbf{r}_0 = \mathbf{b}$, at the k -iteration step the OMP algorithm selects the k -th atom as

$$\lambda_k = \arg \max_{1 \leq i \leq M} |\langle \mathbf{r}_{k-1}, \mathbf{a}_i \rangle|, \quad (4)$$

and it updates the residual as

$$\mathbf{r}_k = \mathbf{b} - P_{\text{span}\{\mathbf{a}_{\lambda_1}, \mathbf{a}_{\lambda_2}, \dots, \mathbf{a}_{\lambda_k}\}} \mathbf{b}. \quad (5)$$

Here, P denotes the orthogonal projection onto a subspace. The OMP algorithm in parallel applies the Gram-Schmidt orthogonalization upon chosen atoms for efficient computation of projections. Ideally, the number of iterations is equal to the number of nonzeros in \mathbf{x} and then the residual becomes zero.

4. TREE-BASED ORTHOGONAL MATCHING PURSUIT (TOMP)

The OMP method in Section 3 only exploits the property **P1** of the wavelet representation \mathbf{x} in the inverse problem $\mathbf{A}\mathbf{x} = \mathbf{b}$. We now motivate an improved version of OMP, named *Tree-based Orthogonal Matching Pursuit (TOMP)*, that in additionally exploits the property **P2** of \mathbf{x} .

The inputs of TOMP are an $N \times M$ measurement matrix \mathbf{A} where $N \ll M$, a data vector \mathbf{b} of length N and two tuning coefficients: the *relaxing coefficient* α and the *downward extending coefficient* l as will be explained later. TOMP returns a reconstruction vector $\hat{\mathbf{x}}$ of length M that solves $\mathbf{A}\mathbf{x} = \mathbf{b}$ and has sparse tree structure.

The key idea of TOMP is to *select from a limited search space in \mathbf{A} at each time a few columns associated with a section of the tree* that are corresponding to significant coefficients in \mathbf{x} . Let Λ_k be the *Selected Set* of columns \mathbf{a}_i 's in matrix \mathbf{A} from the beginning until before step k . Let \mathbf{U}_k be the Gram-Schmidt version of Λ_k .

Whenever a new column is added to Λ_k , it is Gram-Schmidt orthogonalized versus all existing columns in \mathbf{U}_k , then be added to \mathbf{U}_k .

This reconstruction algorithm consists of two phases: retrieving the scaling coefficients and retrieving the wavelet coefficients in \mathbf{x} . Firstly we set the Selected Sets Λ_0 and \mathbf{U}_0 to be empty and initialize the residual: $\mathbf{r}_0 = \mathbf{b}$. In Phase 1, we select all columns of \mathbf{A} corresponding to the scaling coefficients since these coefficients are nonzeros:

$$\Lambda_1 = \mathbf{A}\{\text{all scaling coefficients}\} \quad (6)$$

All vectors in Λ_1 are sequentially orthogonalized using Gram-Schmidt and stored in \mathbf{U}_1 . At the end of Phase 1, the residual is updated by:

$$\mathbf{r}_1 = \mathbf{r}_0 - P_{\mathbf{U}_1} \mathbf{r}_0 = \mathbf{b} - P_{\mathbf{U}_1} \mathbf{b} \quad (7)$$

In Phase 2, we focus on retrieving the wavelet coefficients which have sparse-tree structure. Phase 2 is the repetition of many iterations. Let \mathbf{C}_k be the *Candidate Set* of vectors \mathbf{a}_i 's which are drawn from \mathbf{A} at the beginning of step k as will be described later. Iteration k receives $\Lambda_k, \mathbf{U}_k, \mathbf{r}_k$ and the Candidate Set \mathbf{C}_k from the previous iteration.

For the first iteration, we initialize the counter $k = 1$. We denote $D_l(\mathbf{a}_i)$ as all the *Descendants* of \mathbf{a}_i on the tree down to l levels, where l is called the *downward extending coefficient* of the Candidate Set. The Candidate Set \mathbf{C}_1 is initialized as:

$$\mathbf{C}_1 = \mathbf{A}_{W(1)} \cup D_l(\mathbf{A}_{W(1)}) \quad (8)$$

where $\mathbf{A}_{W(1)}$ are the columns of \mathbf{A} corresponding to all the wavelet coefficients at level 1 (all the highest fathers). TOMP then investigates the inner products of the current residual with all vectors in \mathbf{C}_k :

$$c_i^{(k)} = |\langle \mathbf{r}_k, \mathbf{a}_i \rangle|, \quad \text{where } \mathbf{a}_i \in \mathbf{C}_k \quad (9)$$

$$c_i^{*(k)} = \max\{c_i^{(k)} : \mathbf{a}_i \in \mathbf{C}_k\} \quad (10)$$

Those vectors in \mathbf{C}_k which give close results to the value $c_i^{*(k)}$ are put into the *Finalist Set* \mathbf{F}_k :

$$\mathbf{F}_k = \{\mathbf{a}_i : \mathbf{a}_i \in \mathbf{C}_k, c_i^{(k)} \geq \alpha c_i^{*(k)}\} \quad (11)$$

We denote $\mathbf{H}(\mathbf{a}_i)$ as the *History Set* of \mathbf{a}_i which includes \mathbf{a}_i and all of its fathers. For each \mathbf{a}_i in the Finalist Set \mathbf{F}_k , we form $\{\mathbf{H}(\mathbf{a}_i) \cap \mathbf{C}_k\}$ called a *Subtree* corresponding to \mathbf{a}_i , which is totally inside \mathbf{C}_k . We orthogonalize each Subtree against the Selected Set \mathbf{U}_k to get the *orthogonalized Subtree* $\{\mathbf{H}(\mathbf{a}_i) \cap \mathbf{C}_k\}$ using Gram-Schmidt. We then compare the projections of the residual \mathbf{r}_k onto those orthogonalized Subtrees to select the Subtree that gives the largest projection:

$$\mathbf{a}_{i_k} = \arg \max_{\mathbf{a}_i \in \mathbf{C}_k} \|P_{\text{span}\{\mathbf{H}(\mathbf{a}_i) \cap \mathbf{C}_k\}} \mathbf{r}_k\|_2 \quad (12)$$

And we add the newly chosen Subtree into our Selected Sets:

$$\Lambda_{k+1} = \Lambda_k \cup \{\mathbf{H}(\mathbf{a}_{i_k}) \cap \mathbf{C}_k\} \quad (13)$$

$$\mathbf{U}_{k+1} = \mathbf{U}_k \cup \{\mathbf{H}(\mathbf{a}_{i_k}) \cap \mathbf{C}_k\} \quad (14)$$

Equations (12)-(14) mark the major differences between OMP and our proposed TOMP. Intuitively, the OMP algorithm sequentially selects the next "good" atom based only on its alignment (i.e. large inner product) with the current residual. In contrast, the TOMP

algorithm sequentially compares and selects the next "good" Subtree, group of atoms related in tree.

The residual is updated by deducting its projection onto the selected Subtree:

$$\mathbf{r}_{k+1} = \mathbf{r}_k - P_{\{\mathbf{H}(\mathbf{a}_{i_k}) \cap C_k\}} \mathbf{r}_k \quad (15)$$

To prepare for the next step, we use the idea from [5] that only expands the Candidate Set by adding its Descendants down to l levels as:

$$C_{k+1} = D_l(\{\mathbf{H}(\mathbf{a}_{i_k}) \cap C_k\}) \cup C_k \setminus \{\mathbf{H}(\mathbf{a}_{i_k}) \cap C_k\} \quad (16)$$

The reason for this update scheme is *if the father is significant then the children is likely to be significant* and should be added to the Candidate Set. In addition, a limited-size Candidate Set with items of high probability of being nonzero can save computational costs and deliver more accurate choice [5].

Equations (9) to (16) are repeated for each iteration until the number of selected atoms in Λ_k reaches a predefined portion (e.g. half) of the number of rows in \mathbf{A} . This is because we cannot expect to recover a signal with more degrees of freedom than the number of measurements. Another stopping rule in the noisy case is $\|\mathbf{r}_k\|_2^2 \leq \varepsilon$ where ε is an optionally selected limit according to the noise level in measurement data. For example, in the case when \mathbf{b} is affected by an additive white Gaussian noise of variance σ^2 , ε can be chosen as $N\sigma^2$, the noise energy, where N is the length of \mathbf{b} .

At the end of TOMP, nonzero coefficients of $\hat{\mathbf{x}}$ are determined by the columns in Λ_k :

$$P_{\text{span}\{\mathbf{a}_l: \mathbf{a}_l \in \Lambda_k\}} \mathbf{b} = \sum_{\mathbf{a}_\lambda \in \Lambda_k} \hat{x}_\lambda \mathbf{a}_\lambda \quad (17)$$

5. EXPERIMENTAL RESULTS AND DISCUSSIONS

Throughout this section, the test signals are piecewise smooth signals as in Figure 3 (a). Wavelet transform of type 8-level Daubechies 4 is applied upon the test signals to get sparse-tree signals in wavelet domain. The signals are then sampled by matrices of i.i.d. Gaussian entries (zero mean and variance 1) which all columns are normalized to unit norm.

In the first experiment, we count the number of multiplications from inner products performed by OMP and TOMP at different values of α and l . The piecewise smooth signal of length 16384 is reconstructed from 300 measurements. Figure 2 (a) shows that large α and small l lead to small number of multiplications. This is because the searching space's size is reduced. When $\alpha \geq 0.5$ and $l \leq 3$, TOMP requires at most about 10^8 multiplications while OMP requires about 1.9×10^8 multiplications. Hence the limited search space helps TOMP to reduce computational costs.

Through the experiments using the above test signal and wavelet transform, we observe that TOMP gives the highest reconstruction SNR at $\alpha = 0.975$ and $l = 2$. From now on, the presented TOMP results are achieved at these two values. In the first experiment, TOMP gives a reconstruction SNR of 23.47 dB over 16.51 dB of OMP while TOMP requires about one thirtieth the number of multiplications required by OMP. However, the effects of α and l on different test signals and wavelet transforms should be investigated in later studies.

A single reconstruction case is in Figure 3. From 300 measurements, TOMP can reconstruct the length-4096 piecewise smooth signal with SNR = 25.25 dB. OMP and BP fail to reconstruct the signal,

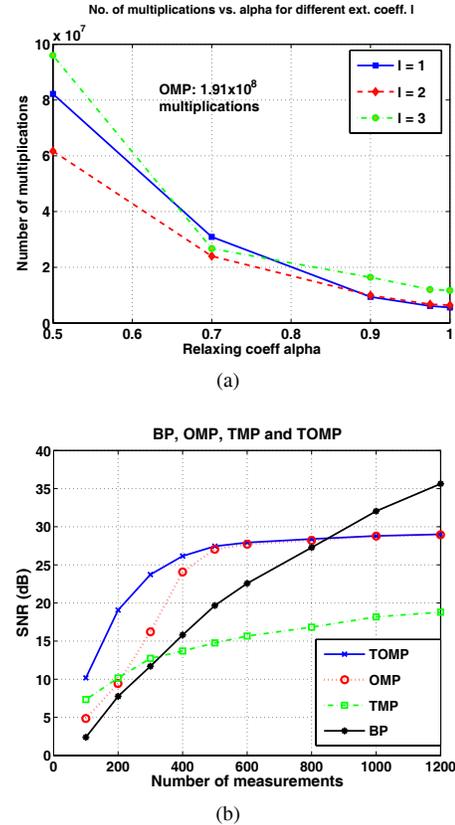


Fig. 2. TOMP compared to other methods: (a) TOMP compared to OMP in the number of multiplications from inner products when varying α and l ; (b) Reconstruction SNRs using BP, OMP, TMP and TOMP.

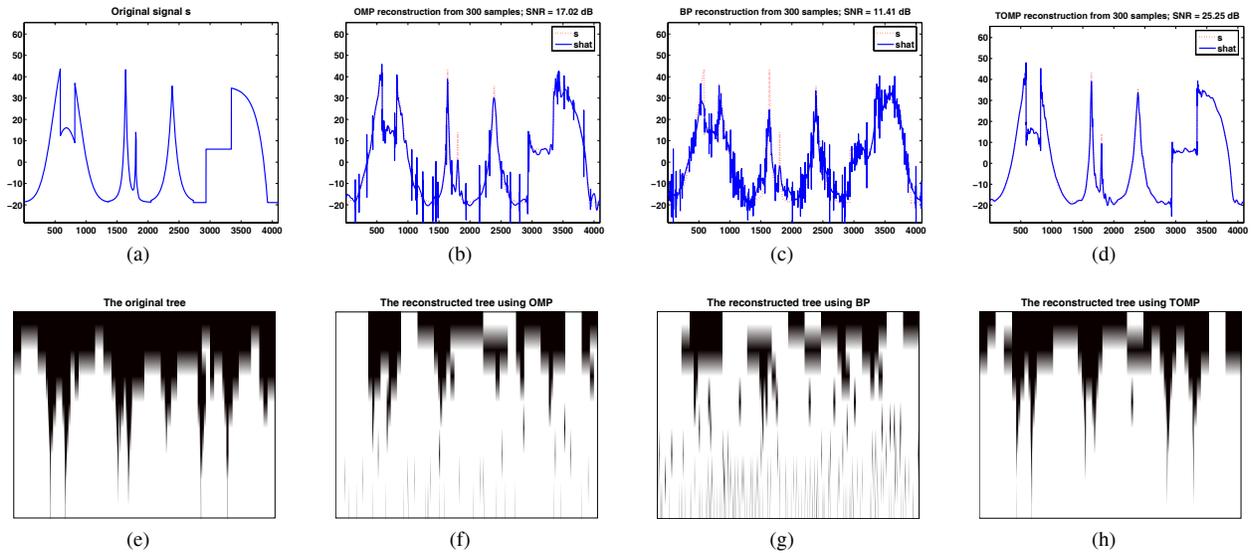


Fig. 3. An example piecewise-smooth signal of length 4096 and its reconstructions from 300 linear measurements using OMP, BP and TOMP, together with tree structures showing locations of significant coefficients and those recovered by OMP, BP and TOMP.

only give 17.02 dB and 11.41 dB. TOMP gains more than 8 dB over OMP and 13 dB over BP. As could be seen in Figure 3 (e-h), TOMP can preserve the tree structure of the original signal while OMP and BP cannot.

Finally, we reconstruct the piecewise smooth signal of length 4096 from different numbers of measurements using TOMP, OMP, TMP [5] and BP [1, 2]. For each number of measurements, we average the reconstruction SNR over 10 trials with different randomly generated measurement matrices M . We focus on the region of low number of measurements and compare the SNRs resulted from TOMP and other methods. In Figure 2 (b), TOMP can gain around 7 dB over OMP and 10 dB over BP.

6. CONCLUSION

Current algorithms such as BP and OMP employ the sparse structure of unknown signals as prior knowledge to solve the inverse problems. In this paper, we go further by introducing a new algorithm named TOMP (Tree-based Orthogonal Matching Pursuit) which in additionally exploits the tree structure of the signals. Our experiments show that TOMP can be better than both OMP and BP in reconstruction quality when the number of measurements is limited, and better than OMP in computational efficiency. The number of iterations in TOMP is less than in OMP since TOMP selects not only one element but a whole group on a tree (a Subtree) at once. In addition, with the limited search space introduced by the concept of Candidate Set, TOMP can also reduce the number of multiplications resulted from computing the inner product.

In summary, with the tree structure employed as a prior knowledge, we have achieved better reconstruction in both computation and SNR. However, the proposed TOMP algorithm in this paper still has not achieved the best performance since we have not fully understood the role of the relaxing coefficient α and the downward extending factor l . As for future studies, we will investigate the effects of α and l on different types of signals and wavelet transforms, on noisy signals and modify TOMP for very large size signals such as images.

We will also investigate other reconstruction algorithms that can exploit more the tree structures, and studying the performance analysis of these algorithms under sparse tree models.

7. REFERENCES

- [1] D. L. Donoho, "Compressed sensing," manuscript, Oct. 2004.
- [2] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," submitted for publication, Nov. 2004.
- [3] J. A. Tropp and A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," *IEEE Trans. Inform. Th.*, Apr. 2005, submitted.
- [4] C. La and M. N. Do, "Signal reconstruction using sparse tree representation," in *Proc. of SPIE Conf. on Wavelet Applications in Signal and Image Processing*, San Diego, Aug. 2005.
- [5] M. F. Duarte, M. B. Wakin, and R. G. Baraniuk, "Fast reconstruction of piecewise smooth signals from random projections," in *Proc. of Workshop on Signal Processing with Adaptive Sparse Structured Representations*, Rennes, France, Nov. 2005.