REDUCED COMPLEXITY BOUNDED ERROR SUBSET SELECTION

Masoud Alghoniemy

Dept. of Electrical Engineering
University of Alexandria
Alexandria 21544, Egypt

Ahmed H. Tewfik

Dept. of Electrical and Computer Engineering
University of Minnesota
Minneapolis, MN 55455

ABSTRACT

A reduced complexity version of the Bounded Error Subset Selection (BESS) algorithm is proposed. By relaxing the integer constraint in the original BESS algorithm, we show that the BESS problem can be reformulated as an ordinary linear program instead of an integer program with exponential worst-case complexity. We retain the sparseness of the representation in the modified BESS by weighting the dictionary with the minimum 2-norm solution of the subset selection problem corresponding to the BESS problem at hand. The proposed algorithm is compared to the Basis Pursuit, Orthogonal Matching Pursuit, and the Best Orthogonal Basis algorithms. It is shown that the proposed algorithm has a better packing property and an improved rate-distortion behavior.

1. INTRODUCTION

Sparse signal representations find applications in many signal processing areas such as coding, signal restoration, direction finding, source localization, and linear inverse problems, to name a few. In the subset selection problem (SS), it is required to find the best signal representation for a signal vector $\mathbf{b}$ using an overcomplete dictionary represented by the N-dimensional vectors spanning the column space of the matrix $\mathbf{A}$. By construction, the number of vectors M in the dictionary is such that $M \gg N$. Thus, it is required to find the sparsest vector $\mathbf{x}$ (the vector $\mathbf{x}$ with the minimum number of non-zero solution) such that $\mathbf{Ax} = \mathbf{b}$. It is known that the SS is NP-hard [1]. Several strategies have been developed for solving the SS problem. In particular, the Method of Frame (MoF) finds the solution which minimizes the 2-norm of the solution vector [2]. However, the MoF does not address the sparseness issue. The Basis Pursuit (BP) algorithm, which can be solved using linear programming, finds the solution that minimizes the l_1-norm of the solution vector [3]. In practice, BP algorithm yields reasonably sparse signal representations. Matching Pursuit (MP) is an iterative greedy algorithm in which the signal is iteratively decorrelated from the dictionary vector which has maximum correlation with the residual [4]. A variant of MP called the Orthogonal Matching Pursuit (OMP) performs an extra step of orthogonalization before each iteration [5]. However, both MP and OMP are greedy algorithms that lack a global optimality criterion. The Best Orthogonal Basis (BOB) uses an entropy measure over orthogonal bases to provide a near-optimal solution [6]. However, as will be seen in the simulation section, BOB fails to find a good sparse representation for some signals when the orthogonal dictionary does not include the non-orthogonal signal components.

The proposed algorithm analyzes a perturbed version of the signal under investigation. This is a reasonable approach due to the normal presence of noise, masking effect, or channel distortion. Sparseness is imposed explicitly by minimizing the number of non-zero coefficients in the solution vector. We demonstrate via simulation the ability of the proposed algorithm to find sparser signal representations with smaller approximation errors than other algorithms.

2. THE BOUNDED ERROR SUBSET SELECTION

The Bounded Error Subset Selection (BESS) has been introduced by the authors in [7, 8] as a reformulation of the classical subset selection problem. It has been shown that by introducing a perturbation vector $\vec{\epsilon}$ to the signal under investigation, $\mathbf{b}$, one can obtain a maximally sparse representation of the signal from the overcomplete dictionary $\mathbf{A}$. Sparseness is imposed by minimizing the number of non-zero coefficients in the solution vector $\mathbf{x}$. This can be achieved by constraining the solution coefficients to take only 0-1 values. In particular, the BESS solves the following integer program:

$$\min_{\mathbf{x}} \ \mathbf{1}^T\mathbf{x} \quad (1)$$
$$s.\ t. \begin{cases} \mathbf{b}_{min} \leq \mathbf{Ax} \leq \mathbf{b}_{max} \\ x_k \in \{0, 1\}. \end{cases}$$

where, $\mathbf{b}_{min} = \mathbf{b} - \vec{\epsilon_1}$ and $\mathbf{b}_{max} = \mathbf{b} + \vec{\epsilon_2}$. Here, $\vec{\epsilon_1}$ and $\vec{\epsilon_2}$ are error vectors or simply constant perturbations. Since $x_k \in \{0, 1\}$, the value of the objective function represents the number of non-zero coefficients in the solution vector, $\mathbf{x}$. This measure of sparseness is minimized directly while keeping the reconstruction error, $\vec{\epsilon_r}$, bounded by

$$\| \vec{\epsilon_r} \|_\infty \leq max\{\| \vec{\epsilon_1} \|_\infty, \| \vec{\epsilon_2} \|_\infty\}. \quad (2)$$

In practice, we chose the error vectors to be equal constant perturbation vectors of value ϵ, hence $\| \vec{\epsilon_r} \|_\infty \leq \epsilon$. Thus, unlike the classical subset selection problem where the l_2-norm of the error is minimized, the BESS tries to increase the sparseness of the representation while keeping the l_∞-norm of the error bounded.

Fig. 1 provides a geometrical interpretation of the BESS in the 2-D case. In particular, by introducing two error vectors $\vec{\epsilon_1}$ and $\vec{\epsilon_2}$ that are not equal in general, the solution vector $\mathbf{x}$ is allowed to take values such that the reconstructed signal $\mathbf{Ax}$ can only be located inside the bounding rectangle as shown in Fig. 1. Hence, the $\| \vec{\epsilon_r} \|_\infty$ is controlled by the size of the bounding rectangle. If $\vec{\epsilon_1} = \vec{\epsilon_2}$, then the rectangle reduces to a square in 2-D, or a hyper-cube in the N-dimensional case, centered at the signal point $\mathbf{b}$.

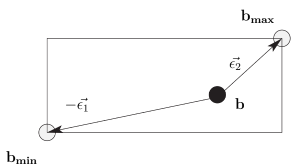


Figure 1: 2-D Geometric interpretation for the BESS.

3. RELAXING THE INTEGER CONSTRAINT

Although the formulation (1) is optimal in the sense that it minimizes the number of non-zero coefficients of the solution, it suffers from the following shortcomings:

- The complexity of the binary integer program (1) is high. In fact, it has an exponential worst-case complexity.
- The underlying model with $x_k \in \{0, 1\}$, implicitly assumes that the signal under investigation, $\mathbf{b}$, is composed of *non-weighted* dictionary vectors. In particular, it does not accommodate the fact that, in general, an acceptable approximation to the signal as a direct sum of entries from the given dictionary $\mathbf{A}$ may not exist.

In order to overcome the previous shortcomings one needs to introduce a weighted dictionary and relax the integer constraint. One solution is to directly relax the integer constraint in (1). This would introduce the required weighting effect by letting the elements of $\mathbf{x}$ take values other than zero and one. However, this leads to losing the explicit expression for the required sparseness that characterized the original BESS formulation with the integer constraint. In particular, the value of the objective function in (1) no longer represents the number of non-zero elements in $\mathbf{x}$.

Note also that in the integer BESS formulation of (1), the solution vector $\mathbf{x}$ is responsible for *both* the sparseness and accuracy of the solution. In particular, when $x_k \neq 0$, this means that the k^{th} dictionary vector is selected with weight equal to x_k. In contrast, let us now examine a different formulation of the problem that separates the sparseness issue from the accuracy requirement. This is achieved by introducing the weighted dictionary $\mathbf{A_w} = \mathbf{AW}$ where $\mathbf{W}$ is a diagonal weighting matrix, $\mathbf{W} = diag(w_1, w_2, \cdots, w_M)$, where w_i represents the weight corresponding to the k^{th} dictionary vector. The diagonal entries w_i should be selected in such a way that the reconstructed signal $\hat{\mathbf{b}}$ is in the neighborhood of $\mathbf{b}$, i.e., inside the corresponding hyper-cube. A possible choice is to use the minimum l_2-norm solution of the corresponding subset selection problem, i.e., take $\mathbf{w}$ to be the minimum l_2-norm solution to $\mathbf{Ax} = \mathbf{b}$. This guarantees that $\hat{\mathbf{b}}$ would be inside a sphere co-centered with the cube at $\mathbf{b}$ and at the same time $\|\mathbf{b} - \hat{\mathbf{b}}\|_\mathbf{2}$ is minimized.

Now, by introducing the weighted dictionary $\mathbf{A_w}$, we can reformulate the BESS problem as:

$$\min_{\mathbf{x}} \ \mathbf{1}^T\mathbf{x} \tag{3}$$

$$s.\ t. \begin{cases} \mathbf{b}_{min} \leq \mathbf{AWx} \leq \mathbf{b}_{max} \\ x_k \geq 0. \end{cases}$$

It should be noted that trying to solve (3) by only relaxing the integer constraint without introducing $\mathbf{W}$ may lead to a non feasible solution in the linear program. This is due to the fact that, in this case, there is no guarantee that $\hat{\mathbf{b}}$ is inside the bounding hyper-cube.

As we discuss below and demonstrate experimentally in the next section, solving (3) yields a sparse solution to the subset selection problem. In order to increase the accuracy of that solution, a correction step is performed after (3) is solved. The correction consists in updating the non-zero coefficients in the solution to (3). Specifically, we replace the vector of these non-zero coefficients by the vector that minimizes the l_2-norm of the error in approximating $\mathbf{b}$ using the dictionary vectors corresponding to these coefficients.

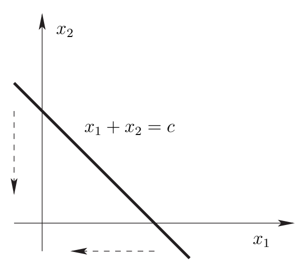


Figure 2: Minimizing the integer-relaxed objective function in 2-D.

Although the objective function in (3) does not explicitly represent the number of non-zero coefficients in $\mathbf{x}$, it does implicitly induce sparseness in the solution. This can be understood from Fig. 2 which illustrates the objective function to be minimized in the 2-D case. As mentioned earlier, introducing the weighting matrix $\mathbf{W}$ has the effect of reducing the magnitude of the entries x_k of $\mathbf{x}$. On the other hand, minimizing the objective function in (3) subject to the positivity constraint on the entries of $\mathbf{x}$ has the effect of pushing the coefficients x_k towards zero as shown in Fig. 2. This increases the sparseness of the solution. Indeed, the experimental results that we provide in the next section show that the vast majority of coefficients x_k in the solution $\mathbf{x}$ to (3) are zero.

Finally, note that (3) is an ordinary linear program that can be solved using fast LP algorithms such as the interior-point algorithms. In contrast, the complexity of the integer program (1) is high.

4. SIMULATION

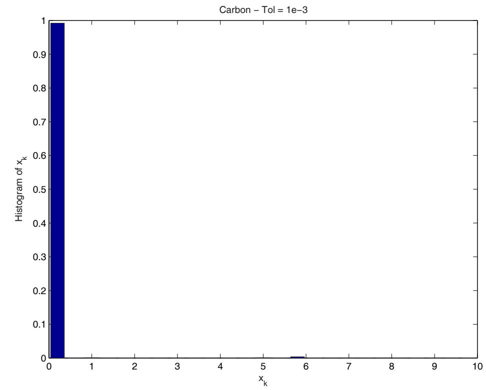
The proposed algorithm (BESS) has been compared to the well-known methods for sparse signal representation, namely, Basis Pursuit, Orthogonal Matching Pursuit, and Best Orthogonal Basis with l_1 entropy. Simulations were performed on different signals and different dictionaries derived from the ATOMIZER package [9]. The free-ware lp_solve was used to solve the linear program on Pentium-III 866 MHz. [10].

For illustration purpose, several signals were analyzed to demonstrate the advantage of using the proposed algorithm. Table 1 summarizes information about the analysis results, including dictionary size, the tolerance ϵ used, the consumed CPU time in seconds, the sparseness of the solution (the ratio of the number of non-zero coefficients in the solution to the length of $\mathbf{x}$), and finally, the 2-norm of the reconstruction error.

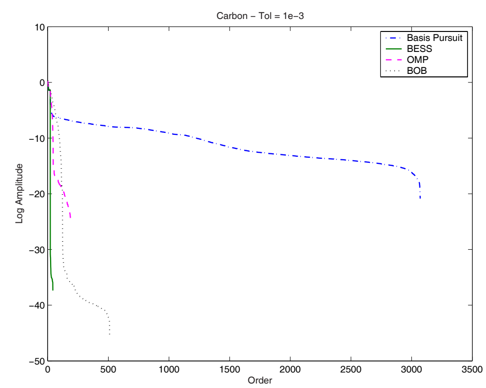
Table 1: SIMULATION RESULTS

Signal	Dictionary	ϵ	Time	Sparseness	$\Vert \vec{\epsilon_r} \Vert_2$
FM	$CP : 256 \times 1792$	1e-3	51.74	0.1445	4.16e-13
Werner	$WP : 512 \times 3072$	1e-3	36.13	0.167	2.72e-13
Doppler	$CP : 256 \times 2048$	1e-6	29.46	0.125	6.414e-14
Carbon	$WP : 512 \times 3072$	1e-3	19	0.014	5.35e-15

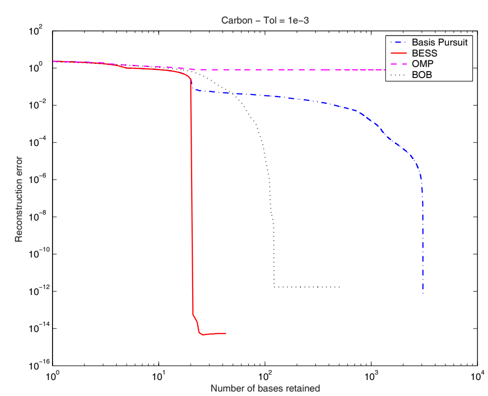
To compare the proposed algorithm to BP, OMP, and BOB algorithms, we provide histograms of the magnitude of the sorted solution coefficients and the norm of the reconstruction error vs. the number of bases in Figures 3 and 4. Specifically, Figs. 3 and 4 show the results that we obtained for the Carbon and Doppler signals respectively. Sparseness of the solution can be verified from the histogram of the solution vectors for both signals in Fig. 3(a) and Fig. 4(a). Note that both figures indicate the high probability of occurrence of zero. Fig. 3(b) and Fig. 4(b) reveal the packing property of the coefficients for the BESS as compared to the other algorithms. As can be seen from Fig. 3(c) OMP fails to represent the Carbon signal properly. In contrast, the proposed algorithm was able to represent it using fewer coefficients than either BOB or BP. Similarly, the proposed algorithm succeeded in sparsely representing the Doppler signal compared to the other techniques as shown in Fig. 4(c).



(a) Coefficients Histogram



(b) Sorted Coefficients

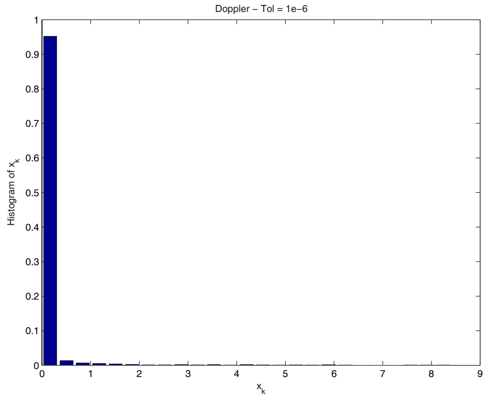


(c) Rate-Distortion

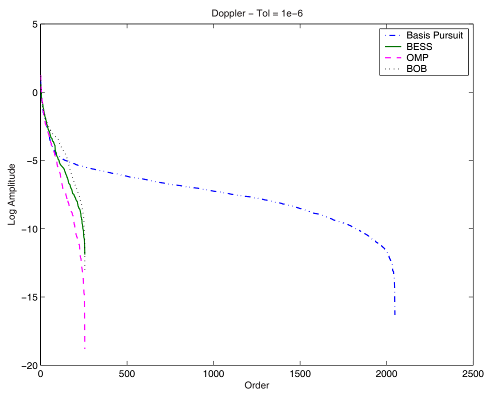
Figure 3: Carbon Signal

5. CONCLUSION

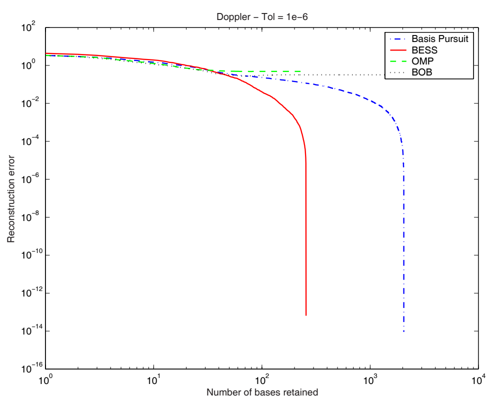
An integer-relaxed version of the Bounded Error Subset Selection (BESS) algorithm was proposed. This has the effect of reducing the complexity of the algorithm from an integer program to an ordinary linear program which can solved efficiently. It has been shown that the reduction in complexity does not come on the expense of reducing the sparseness of the signal representation. Several signals were analyzed using the proposed algorithm and compared to the Basis Pursuit, Orthogonal Matching Pursuit, and Best Orthogo-



(a) Coefficients Histogram



(b) Sorted Coefficients



(c) Rate-Distortion

Figure 4: Doppler Signal

nal Basis algorithms. Simulation shows the potential of the BESS algorithm in sparse signal representation.

6. REFERENCES

[1] B. Natarajan, "Sparse Approximate Solutions to Linear Systems," SIAM J. Comp., Vol. 24, pp. 227-234, Apr. 1995.

[2] Daubechies I., "Time-Frequency Localization operators: A Geometric Phase Space Approach," *IEEE Trans. on Info. Theory,* Vol. 34, No. 4, pp. 605-612, July 1988.

[3] Chen S., and Donoho D., Saunders M. A., "Atomic Decomposition By Basis Pursuit," SIAM J. on Scientific Computing, Vol. 20, No. 1, pp. 33-61, 1998.

[4] Mallat S., and Zhang Z., "Matching Pursuit with Time-Frequency Dectionaries," *IEEE Trans. on Signal Processing,* Vol. 41, No. 12, pp. 3397-3415, Dec. 1993.

[5] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," *in Proc. 27th Asilomar Conference on Signals, Systems and Computers, A. Singh, ed.*, IEEE Comput. Soc. Press, Los Alamitos, CA, 1993.

[6] Coifman R., and Wickerhauser M., "Entropy-based Algorithms for Best Basis Selection," *IEEE Trans. on Info. Theory,* Vol. 38, No. 2, pp.713-718, March 1992.

[7] Masoud Alghoniemy and Ahmed H. Tewfik, "A Sparse Solution to the Bounded Subset Slection Problem: A Network Flow Model Approach," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing.*, Vol. 5, pp. 89-92, May 2004.

[8] Masoud Alghoniemy and Ahmed H. Tewfik, "Bounded Subset Selection with Noninteger Coefficients," *Proceedings of the Eusipco-2004 Conference.*, pp. 317-320, Sept. 2004.

[9] http://www-stat.stanford.edu/atomizer

[10] lp_solve@yahoogroups.com