BOUNDED SUBSET SELECTION WITH NONINTEGER COEFFICIENTS

Masoud Alghoniemy

Electrical Engineering Department
University of Alexandria,
Alexandria 21544, Egypt

Ahmed H. Tewfik

Electrical Engineering Department
University of Minnesota,
Minneapolis, MN 55455

ABSTRACT

The subset selection problem is known to be NP hard. It was recently shown that by relaxing the requirement that the reconstructed signal be equal to the original, one ends with a bounded error subset selection that admits a solution in polynomial time. In the bounded error subset selection problem, the reconstructed signal is allowed to differ from the original signal by a bounded error. This bounded error formulation is natural in many applications, such as coding. In this paper, we improve the accuracy and reduce the complexity of the previously proposed approach for solving the bounded error subset selection problem. In particular, unlike the previously proposed approach for solving the bounded error subset selection problem, our new algorithm accommodates cases where the coefficients of the closest sparse approximation to the underlying signal in the dictionary are not necessarily one. Our new algorithm is based on weighting the dictionary vectors by the minimum l_2 norm solution and relaxing the integer constraint on the coefficients of the dictionary vectors. It is shown to guarantee high signal accuracy and sparsity. Compared with the Basis Pursuit and the Method of Frames (MoF) algorithms, the proposed algorithm has a better rate-distortion behavior.

1. INTRODUCTION

Classical subset selection arises in many signal processing applications. For example, Fourier bases are suitable for signals which are rich in harmonics while wavelet dictionaries can be used for signals that include transients. Hence, overcomplete dictionaries are required to deal with inherently complex signals. In the subset selection problem (SSP), 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 basis 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

$$\mathbf{A}\mathbf{x} = \mathbf{b}. \tag{1}$$

It is known that the SSP is NP-hard [1]. Several strategies have been developed for solving the SSP. For example, the Method of Frames (MoF) finds the solution that minimizes the l_2 norm [2]. On the other hand, the solution of the Basis Pursuit (BP) algorithm minimizes the l_1 norm [3]. Mallat *et. al* developed the Matching Pursuit (MP) technique in which the signal is iteratively decorrelated from the basis vector which has maximum correlation with the residual [4]. The Best Orthogonal Basis (BOB) is designed for wavelet and cosine packet dictionaries which finds the solution based on a minimum entropy criterion [5]. The previous techniques do not necessarily find the sparsest solution since the optimization criterion does not address the sparsity issue. The authors in [6] found the sparse solution for certain structured matrices.

The authors in [7] reformulated the subset selection problem as a bounded error problem by relaxing the equality constraint in (1). Thus, the Bounded Subset Selection Problem (BSSP) finds $\mathbf{x}$ such that

$$\mathbf{b}_{min} \le \mathbf{A}\mathbf{x} \le \mathbf{b}_{max} \tag{2}$$

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 which may represent the error introduced by the masking model as in audio coding [8], or simply a constant perturbation. It is shown in (2) that the BSSP is not NP hard and in fact admits a solution with $O(N^{3.5})$ complexity for primal-dual log barrier interior point method. The relaxation introduced in the BSSP allowed more freedom which is used to find the sparse solution to (2). Sparseness is imposed in (2) by minimizing the number of non-zero elements in the solution vector $\mathbf{x}$. This is achieved by solving the following binary integer program

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

$$subject \ \ to \begin{cases} \mathbf{b}_{min} \le \mathbf{A}\mathbf{x} \le \mathbf{b}_{max} \\ x_k \in \{0,1\}. \end{cases}$$

By limiting $x_k \in \{0,1\}$, the solution vector $\mathbf{x}$ plays the role of selecting the minimum number of basis vectors from the dictionary $\mathbf{A}$ according to the constraint (2). Note that the integer constraint $x_k \in \{0,1\}$ adds to the complexity of the solution. At the same time, it limits the achievable reconstruction accuracy since the reconstructed vector cannot be a linear combination of *weighted* vectors, i.e., the model (3) assumes that the signal under consideration, $\mathbf{b}$, can be represented as an exact sum of *non-weighted* vectors which is not the case in most situations.

In this paper we overcome the shortcomings of the approach in (2) by allowing signal representations using scaled dictionary vectors while preserving the sparseness of the solution. This is achieved by introducing a weighting matrix, $\mathbf{W}$, derived from the minimum l_2 norm solution and relaxing the integer constraint. The relaxation has the effect of reducing the complexity of the algorithm. The proposed approach is discussed in the next section. Simulation results are given in section 3. The simulation results show that the proposed approach has a better rate-distortion behavior than traditional solutions to the SSP.

2. THE PROPOSED APPROACH

In order to overcome the shortcomings of the model in (3), We consider solving the problem in two steps:

- **Step 1:** Find a sparse solution starting from a feasible one.
- **Step 2:** Increase the accuracy of the solution obtained in step 1 while preserving its sparseness.

Observe that in the classical SSB or the BSSP in (2), the solution vector $\mathbf{x}$ is responsible for *both* the sparsity and the accuracy of the solution. In particular, when $x_k \neq 0$, this means that the k^{th} basis is selected with weight equal to x_k. Hence, $\mathbf{x}$ is responsible for achieving *both* the sparsity and accuracy of the solution. The proposed two-step algorithm separates the sparsity issue from the accuracy requirement. In particular, $\mathbf{x}$ is only responsible for achieving a sparse solution while accuracy comes from weighting the dictionary vectors according to the l_2 norm solution and finally computing an optimal l_2 norm solution over the set of retained vectors.

2.1. Step 1: Finding a sparse solution from the minimum l_2 norm

To partially address the accuracy issue, we introduce the weighted dictionary $\mathbf{A_w} = \mathbf{AW}$, where $\mathbf{W}$ is a $M \times M$ diagonal matrix whose elements are the MoF solution of (1). Note that the minimum l_2 norm solution is obtained by selecting all vectors in the dictionary $\mathbf{A}$. However, its sparseness is not guaranteed. The sparsity issue can be addressed by solving the following BSSP

$$\min_{\mathbf{x}} \ \mathbf{1}^T\mathbf{x} \tag{4}$$

$$subject \ \ to \begin{cases} \mathbf{b}_{min} \le \mathbf{AWx} \le \mathbf{b}_{max} \\ x_k \in \{0,1\}. \end{cases}$$

Since x_k can only be either 0 or 1, $\mathbf{x}$ takes care of the sparseness issue. This is clear by noticing that $\mathbf{x}$ selects the appropriate *weighted* vectors from $\mathbf{A_w}$.

It should be noted that the norm of the error in the reconstructed signal introduced by solving (4) is guaranteed to be lower than the corresponding error introduced by solving (3), i.e., without weighting the dictionary, for the same sparsity, due to the effect of $\mathbf{W}$. However, the complexity is still high due to the integer constraint imposed $\forall x_k$. In order to reduce the complexity of the problem, we relax the integer constraint by letting $0 \le x_k \le 1$. Hence, we solve instead the following problem:

$$\min_{\mathbf{x}} \ \mathbf{1}^T\mathbf{x} \tag{5}$$

$$subject \ \ to \begin{cases} \mathbf{b}_{min} \le \mathbf{AWx} \le \mathbf{b}_{max} \\ x_k \in [0,1]. \end{cases}$$

Problem (5) can be solved using any of the standard linear programming packages, such as lp_solve package [9]. We may now ask the question, "what is the effect of relaxing the integer constraint on both the sparsity and the accuracy of the solution? It is natural to expect that the solution of (5) would be less sparse than that of (4). But since the weights of the vectors have been taken care of by $\mathbf{W}$, one would expect that the values of x_k would be gathered either around zero or one. This is exactly what we observe in practice. For example, Fig. 3 and Fig. 7 describe the histogram of x_k for the Werner Sorrows and the FM signals analyzed using the proposed algorithm with D4 wavelet and cosine packet dictionaries, as explained in section 3. It is clear that the majority of x_k are packed around 0 and 1, justifying the use of the relaxation in (5). Since the amplitudes and the number of $x_k \notin \{0,1\}$ are small, sparsity can be improved by simple rounding. In particular, by rounding the $x_k's$, the resultant coefficient vector will be a binary vector. Another choice is to disregard $x_k \neq 1$. We have found experimentally that rounding or simply retaining the coefficients $x_k = 1$ yields essentially equivalent performance from a reconstruction error standpoint. For example, for the FM signal shown in Fig. 6 analyzed using the CP dictionary of size 1792, the number of ones after rounding the solution vector $\mathbf{x}$ is 494 coefficients out of 1792 and the 2-norm of the reconstruction error is 5.9692e-014. On the other hand, the number of the coefficients $x_k = 1$ without rounding is equal to 423 and the 2-norm of the reconstruction error is 7.9480e-014 which is almost the same as if we did the rounding. This is also clear from the histogram in Fig. 7.

2.2. Step 2: Minimizing the l_2 norm while preserving sparseness

Step 2 further improves accuracy. In particular, relaxing the integer solution increases the error since we started from the minimum l_2 norm solution. Hence a correcting step is required using only the selected vectors to preserve the sparseness that was achieved in the previous step. Assume that the solution to (5) yields q basis vectors out of M, then define an $M \times q$ selection matrix $\mathbf{S}$ with elements equal to 0 or 1 and exactly one 1 in each row and column. For each column of $\mathbf{S}$ the k^{th} location equals to 1 if $x_k \neq 0$. Hence the columns of $\mathbf{AS}$ are the q selected vectors by the solution of (5). Step 2 finds the solution vector $\mathbf{x}^*$, of dimension q, which minimizes the l_2 norm of the error considering only the q selected vectors. i.e.,

$$\mathbf{x}^* = \min_{\mathbf{x}} \ \| \mathbf{ASx} - \mathbf{b} \|_\mathbf{2} \tag{6}$$

It should be noted that step 2 does not reduce the sparsity of the solution as it only deals with q vectors. In conclusion, the reconstructed signal $\tilde{\mathbf{b}} = \mathbf{ASx}^*$ is guaranteed to satisfy

$$\mathbf{b}_{min} \le \tilde{\mathbf{b}} \le \mathbf{b}_{max} \tag{7}$$

and the solution is maximally sparse.

3. SIMULATION

Simulation was performed on different signals with different dictionaries. The Atomizer package was used in the simulation which provided us with signals and dictionaries [10]. The lp_solve package was also used for solving the mixed integer linear programs [9]. A 512 samples of the Werner Sorrows signal shown in Fig. 1 was analyzed using the D4 wavelet packet dictionary of size 3072. In the model we used, $\vec{\epsilon_1} = \vec{\epsilon_2} = \vec{0.2}$. Fig 2 shows a zoomed plot for the reconstructed signal along with the upper and lower bounds. It is clear that the reconstructed signal satisfies (7). The 2-norm of the error was 2.3327e-13 and the CPU running time was 23 sec. The Sparseness of the solution can be noticed from the histogram of the integer-relaxed x_k coefficients in Fig. 3 which indicates the packing of the coefficients around 0 and 1 as explained in the previous section. Fig. 4 compares the sorted $x_k's$ resulted from the proposed algorithm with both the BP and the MoF solutions. It is clear that the proposed approach has better sparseness characteristics compared with the BP and the MoF algorithms. This is due to the fact that step 1 addresses the sparsity issue directly by minimizing the number of non-zero elements in the solution vector while step 2 does not reduce the sparseness of the solution. On the other hand, Fig. 5 shows the 2-norm of the reconstruction error for the "Waerner Sorrows" signal vs. the number of retained coefficients for the proposed algorithm, BP, and the MoF, respectively. It is clear that the proposed algorithm has a better rate-distortion behavior over the BP and the MoF algorithms. The same experiments were performed on 256 samples from the FM signal shown in Fig. 6 with the Cosine Packet (CP) dictionary of size 1792 with $\vec{\epsilon_1} = \vec{\epsilon_2} = \vec{0.1}$. The 2-norm of the reconstruction error was found to be 7.948e-14 and the CPU running time was 55.5 sec. As is clear from Fig. 8, the proposed approach has a better sparseness property compared with the BP and the MoF solution. Also the proposed algorithm has a better rate-distortion characteristics compared with the BP and the MoF as illustrated by Fig. 9.

4. CONCLUSION

The bounded subset selection problem was reviewed and an algorithm for computing a sparse solution to the problem was presented. We found that by relaxing the integer constraint in the BSSP model, one can achieve better signal approximation in terms of sparseness and reconstruction accuracy. The solution is obtained in two steps. The first step finds a sparse solution by weighting the dictionary using weights derived from the minimum l_2 norm solution to the problem . The second step serves as a correction step and minimizes the l_2 norm of the sparse representation. Compared with the Basis Pursuit and the MoF algorithms, the proposed approach has a better rate-distortion behavior.

5. 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., "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] 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.

[6] Nafie M., Tewfik A.H., and Ali M., "Deterministic and iterative solutions to subset selection problems", *IEEE Trans. on Signal Processing,* pp. 1591-1601, July 2002.

[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.* May 2004.

[8] Painter T., and Spanias A., "Perceptual Coding of Digital Audio." *Proceedings of the IEEE,* Vol. 88, No. 4, pp. 451-513, April 2000.

[9] ftp://ftp.es.ele.tue.nl/pub/lp_solve/

[10] http://www-stat.stanford.edu/ Atomizer

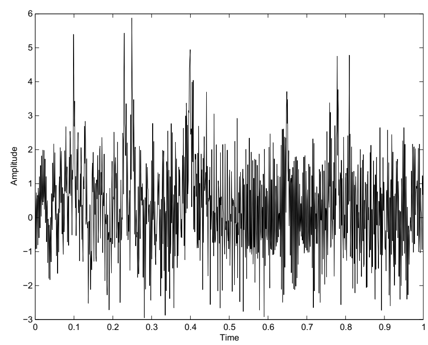

Figure 1: Werner Sorrows Signal

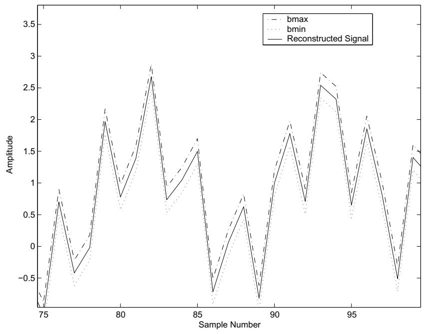

Figure 2: Samples of the Reconstructed Werner, $\mathbf{b}_{min}$, and $\mathbf{b}_{max}$ for $\vec{\epsilon} = 0.\vec{2}$.

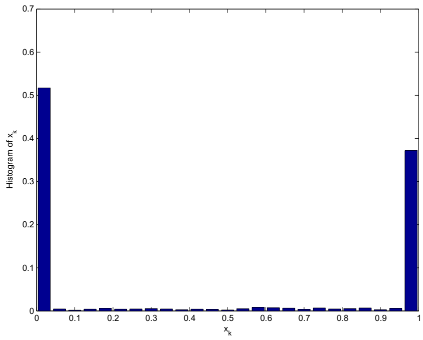

Figure 3: Histogram of integer-relaxed x_k "Werner Sorrows"

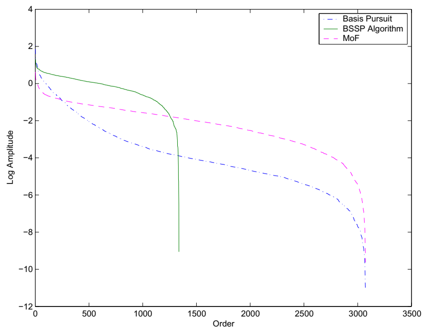

Figure 4: Sorted Coefficients "Werner Sorrows"

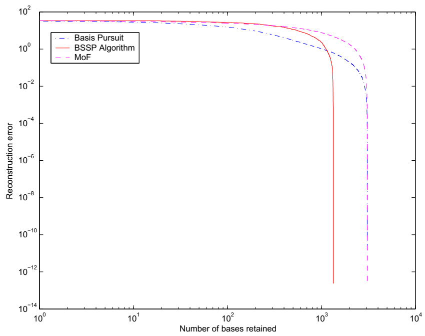

Figure 5: Reconstruction error vs. order "Werner Sorrows"

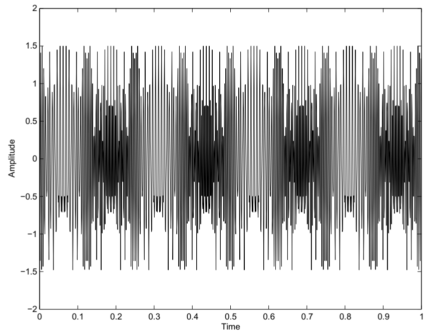

Figure 6: FM Signal

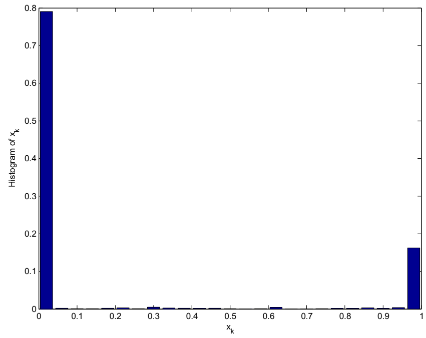

Figure 7: Histogram of integer-relaxed x_k "FM"

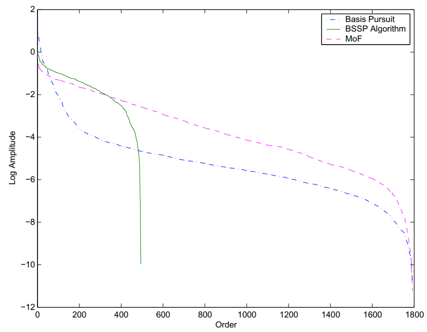

Figure 8: Sorted Coefficients "FM"

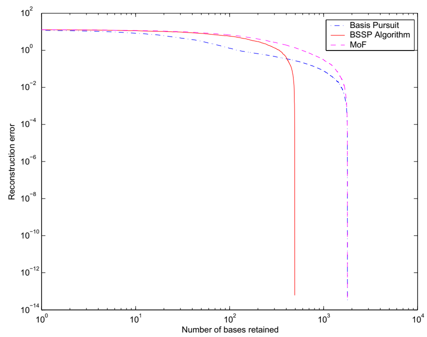

Figure 9: Reconstruction error vs. order "FM"