Backward Optimized Orthogonal Matching Pursuit Approach

Miroslav Andrle, Laura Rebollo-Neira*, and Evangelos Sagianos
NCRG, Aston University,
http://www.ncrg.aston.ac.uk
Birmingham B4 7ET,
United Kingdom

Abstract

A recursive approach for shrinking coefficients of an atomic decomposition is proposed. The corresponding algorithm evolves so as to provide at each iteration a) the orthogonal projection of a signal onto a reduced subspace and b) the index of the coefficient to be disregarded in order to construct a coarser approximation minimizing the norm of the residual error.

1 Introduction

Adopting the terminology early introduced in [1] we call atomic decomposition of a signal to the linear expansion:

$$f = \sum_{n=1}^{N} c_n \alpha_n, \tag{1}$$

where the atoms α_n are elements of a non-orthogonal sequence in the space of the signal f, which is assumed to be a Hilbert space $\mathcal{H}$. Within the general Matching Pursuit (MP) framework [1–5] the atoms are chosen, by different criteria, from a in general redundant set which is called a dictionary. The problem of selecting atoms in order to construct the signal representation with the minimum possible

1

number of coefficients is a very complex problem. In particular, the problem of constructing the optimal approximation with N-atoms selected so as to minimize the Euclidean distance between the signal and the approximation is a NP-hard problem [3]. Thus, in this line only suboptimal solutions are actually feasible. In a previous publication a suboptimal iterative pursuit strategy, which is only optimal at each iteration step, has been introduced with the name of Optimized Orthogonal Matching Pursuit (OOMP) approach [5]. Such a technique is based on an iterative method for computing dual atoms β_n^N ; $n = 1, \ldots, N$ giving rise to a representation of the orthogonal projector operator onto the subspace generated by the set of atoms α_n ; $n = 1, \ldots, N$. These atoms, which happen to be biorthogonal to atoms α_n ; $n = 1, \ldots, N$ [6], allow to compute the coefficients in (1) by computing inner products, i.e. $c_n^N = \langle \beta_n^N, f \rangle$ (the superscript N indicates that the dual atoms, and therefore the coefficients, depend of the number N of dictionary atoms being considered). The OOMP approach tackles the problem of selecting the new atom α_{N+1} to improve the approximation. Moreover, the coefficients of the atomic decomposition are recursively modified in order to yield an optimal approximation in the enlarged subspace. Here we consider the reverse situation: We assume that an atomic decomposition is given and we wish to eliminate some coefficients. For such an end we propose a technique, that we term Backward Optimized Orthogonal Matching Pursuit (BOOMP) which consists of the following elements: a) a recursive approach to modify the coefficients of the atomic decomposition when one of the coefficients is to be disregarded and b) the criterion to select such a coefficient. Although the technique can be applied to reduce coefficients of any atomic decomposition, regardless of how such a decomposition is obtained, in this letter we focus on its implementation as a posterior step of OOMP. The reason for taking this route

2

is that the implementation of BOOMP is really straightforward on the outputs of OOMP. In addition, we believe that this way of introducing the approach makes more clear its purpose and also its implementation steps.

We would like to stress that even the construction of suitable suboptimal signal approximations by using dictionaries is a complex theoretical and practical challenge [7, 8]. This communication aims at enhancing the fact that, since suboptimal forward and backward approximations are in general not reversible, application of the proposed backward pursuit approach, after a forward pursuit selection of atoms, may result in a gain with respect to sparseness of the representation. This is clearly illustrated here by a simple example.

2   Adaptive MP strategies

The MP approach is a technique to compute adaptive signal representations by iterative selection of atoms [1]. In its original form this technique does not yield at each iteration the linear expansion of the selected atoms that approximate the signal at best in a minimum distance sense. A later refinement, which does provide such an approximation, has been termed Orthogonal Matching Pursuit (OMP) [2, 3, 9]. However, since OMP selects the atoms according to the MP prescription, the selection criterion is not optimal in the sense of minimising the residual of the new approximation. The requirement of such minimization has led to the recently introduced OOMP approach [5]. This technique is implemented by means of an adaptive biorthogonalization method which, within the workings of the selection process, generates the set of biorthogonal atoms yielding orthogonal projections [5, 6]. Such biorthogonal atoms are used to compute the coefficients of the atomic

decomposition and are obtained through the following recursive equations:

$$\beta_n^{k+1} = \beta_n^k - \beta_{k+1}^{k+1}\langle\alpha_{k+1}, \beta_n^k\rangle \,; \quad n = 1, \ldots, k, \tag{2}$$

$$\beta_{k+1}^{k+1} = \frac{\psi_{k+1}}{||\psi_{k+1}||^2}, \tag{3}$$

where the set $\psi_j \,; \, j = 1, \ldots, k+1$ is obtained inductively, from $\psi_1 = \alpha_1$, by orthogonalization of atoms $\alpha_j \,; \, j = 1, \ldots, k+1$. These atoms are selected from the dictionary by minimizing, at each iteration step, the norm of the residual error in approximating the signal [5].

Let us stress once more that, since all the above mentioned pursuit strategies evolve by fixing the atoms selected in the previous steps, there is plenty of room for possible improvement with regard to compression of the representation.

Improving compression after the OOMP procedure implies having to eliminate some coefficients of the atomic resulting decomposition. For the coarser approximation to be optimal in a minimum distance sense, the remaining coefficients must be recalculated [10–12]. This feature of non-orthogonal expansions is a major difference with orthogonal ones and has motivated an adaptive approach to modify biorthogonal atoms in order for then to yield orthogonal projections when the corresponding subspace is reduced [12].

Let us suppose that OOMP has selected N atoms to represent a given signal up to some predetermined precision. Let us denote V_N to the subspace spanned by such atoms i.e, $V_N = \text{span}\{\alpha_1, \ldots, \alpha_N\}$ and let $\tilde{V}_{N-1}$ be the subspace which is left by removing one atom, say the j-th one, i.e. $\tilde{V}_{N-1} = \text{span}\{\alpha_1, \ldots, \alpha_{j-1}, \alpha_{j+1}, \ldots, \alpha_N\}$. Since the biorthogonal atoms $\beta_n^N \,; \, n = 1, \ldots, N$ are available (as output of the OOMP procedure), to construct the orthogonal projector of f onto $\tilde{V}_{N-1}$ we just

need to modify the atoms β_n^N ; $n = 1, \ldots, j-1, j+1, \ldots, N$ as follows [12]

$$\beta_n^{N-1} = \beta_n^N - \frac{\beta_j^N \langle \beta_j^N, \beta_n^N \rangle}{||\beta_j^N||^2} ; \quad n = 1, \ldots, j-1, j+1, \ldots, N. \tag{4}$$

In writing the above equation we have re-defined the superscript $N-1$. Now this upper index indicates that the biorthogonal atoms are modified in order to account for the deleting of any one atom (not necessarily the last element of the spanning set). In the next section we discussed how these adaptive backward equations generate the proposed BOOMP approach.

3 Backward Optimized Orthogonal Matching Pursuit

After the selection of N atoms the OOMP approach provides a representation of a signal f as given by [5]

$$f_{V_N} = \hat{P}_{V_N} f = \sum_{n=1}^N c_n^N \alpha_n, \tag{5}$$

$\hat{P}_{V_N} f$ indicates the orthogonal projection of the signal f onto V_N and coefficients c_n^N are obtained as $c_n^N = \langle \beta_n^N, f \rangle$. Theorem 1 below proves that, if we decide to eliminate the coefficient c_j^N from the above expansion, in order to obtain the optimal approximation of f in the reduced subspace $\tilde{V}_{N-1}$, the remaining coefficients c_n^N ; $n = 1, \ldots, j-1, j+1, \ldots, N$ should be modified as follows:

$$c_n^{N-1} = c_n^N - \frac{\langle \beta_n^N, \beta_j^N \rangle}{||\beta_j^N||^2} c_j^N ; \quad n = 1, \ldots, j-1, j+1, \ldots, N. \tag{6}$$

For the sake of organizing the corresponding proof let us first prove the following lemma:

Lemma 1. *Let signal $f_{\tilde{V}_{N-1}}$ be given by*

$$f_{\tilde{V}_{N-1}} = \sum_{\substack{n=1 \\ n \neq j}}^N c_n^{N-1} \alpha_n, \tag{7}$$

with coefficients c_n^{N-1} as in (6), and let f be a signal in $\mathcal{H}$. The difference $f - f_{\tilde{V}_{N-1}}$ is orthogonal to every function in $\tilde{V}_{N-1}$.

Proof. Using (6) and (7) we have:

$$f - f_{\tilde{V}_{N-1}} = f - \sum_{\substack{n=1 \\ n\neq j}}^{N} c_n^{N-1}\alpha_n = f - \sum_{\substack{n=1 \\ n\neq j}}^{N} c_n^N \alpha_n + \sum_{\substack{n=1 \\ n\neq j}}^{N} \frac{\langle \beta_n^N, \beta_j^N \rangle}{||\beta_j^N||^2} c_j^N \alpha_n$$

$$= f - \sum_{n=1}^{N} c_n^N \alpha_n + c_j^N \alpha_j + \sum_{n=1}^{N} \frac{\langle \beta_n^N, \beta_j^N \rangle}{||\beta_j^N||^2} c_j^N \alpha_n - c_j^N \alpha_j. \qquad (8)$$

Since $\beta_j^N \in V_N$, it follows from (5) that $\sum_{n=1}^{N} \frac{\langle \beta_n^N, \beta_j^N \rangle}{||\beta_j^N||^2}\alpha_n = \frac{\hat{P}_{V_N}\beta_j^N}{||\beta_j^N||^2} = \frac{\beta_j^N}{||\beta_j^N||^2}$. Hence (8) turns out to be

$$f - f_{\tilde{V}_{N-1}} = f - \sum_{n=1}^{N} c_n^N \alpha_n + c_j^N \frac{\beta_j^N}{||\beta_j^N||^2}. \qquad (9)$$

Since by hypothesis $\sum_{n=1}^{N} c_n^N \alpha_n = \hat{P}_{V_N} f$, the difference $f - \sum_{n=1}^{N} c_n^N \alpha_n$ is orthogonal to every function in V_N. Furthermore, since $\langle \alpha_n, \beta_j^N \rangle = \delta_{n,j}$, by taking the inner product both sides of (9) with every function α_n ; $n = 1, \ldots, j-1, j+1, \ldots, N$ we obtain $\langle \alpha_n, f - f_{\tilde{V}_{N-1}} \rangle = 0$, which proves that $f - f_{\tilde{V}_{N-1}}$ is orthogonal to every function in $\tilde{V}_{N-1}$. $\qquad \square$

Lemma 2. *The coefficients c_n^{N-1} of the linear expansion*

$$f_{\tilde{V}_{N-1}} = \sum_{\substack{n=1 \\ n\neq j}}^{N} c_n^{N-1}\alpha_n \qquad (10)$$

minimizing the distance in $\tilde{V}_{N-1}$ to a given signal $f \in \mathcal{H}$ are obtainable, from c_n^N and β_n^N ; $n = 1, \ldots, N$, as prescribed in (6).

Proof. Let g be an arbitrary signal in $\tilde{V}_{N-1}$ and let us write $||f - g||^2$ as follows:

$$||f - g||^2 = ||f - f_{\tilde{V}_{N-1}} + f_{\tilde{V}_{N-1}} - g||^2. \qquad (11)$$

From Lemma 1 we know that $f - f_{\tilde{V}_{N-1}}$ is orthogonal to every function in $\tilde{V}_{N-1}$ and since $f_{\tilde{V}_{N-1}} - g$ is in $\tilde{V}_{N-1}$ we have

$$||f - g||^2 = ||f - f_{\tilde{V}_{N-1}}||^2 + ||f_{\tilde{V}_{N-1}} - g||^2, \tag{12}$$

from where we conclude that $||f - g||^2$ is minimized if $g \equiv f_{\tilde{V}_{N-1}}$. $\qquad \square$

Lemma 2 tells us how to proceed to disregard coefficients of a non-orthogonal linear expansion. Assuming that the coefficient c_j^N to be disregarded has been selected, in order to optimize the approximation in a minimum distance sense, the remaining coefficients should be modified as indicated in (6). The next theorem gives an answer to the question as to how to select the coefficient c_j^N to be neglected.

Theorem 1. *Let R_j be the residual resulting by disregarding a coefficient c_j^N for passing from approximation f_{V_N} to $f_{\tilde{V}_{N-1}}$ i.e., $f_{V_N} = f_{\tilde{V}_{N-1}} + R_j$. In order to minimize the norm of the residual R_j such coefficient is to be chosen as the one yielding a minimum value of the quantity*

$$\frac{|c_j^N|^2}{||\beta_j^N||^2}. \tag{13}$$

Proof. Since $R_j = f_{V_N} - f_{\tilde{V}_{N-1}} = \sum_{n=1}^{N} c_n^N \alpha_n - \sum_{\substack{n=1 \\ n \neq j}}^{N} c_n^{N-1} \alpha_n$, by using (6) we have:

$$R_j = \sum_{n=1}^{N} c_n^N \alpha_n - \sum_{\substack{n=1 \\ n \neq j}}^{N} c_n^N \alpha_n + \sum_{\substack{n=1 \\ n \neq j}}^{N} c_j^N \alpha_n \frac{\langle \beta_n^N, \beta_j^N \rangle}{||\beta_j^N||^2}$$

$$= c_j^N \alpha_j + c_j^N \sum_{n=1}^{N} \alpha_n \frac{\langle \beta_n^N, \beta_j^N \rangle}{||\beta_j^N||^2} - c_j^N \alpha_j \tag{14}$$

As already discussed, $\sum_{n=1}^{N} \alpha_n \langle \beta_n^N, \beta_j^N \rangle = \beta_j^N$. Then, from the last equation it follows that $R_j = \frac{c_j^N \beta_j^N}{||\beta_j^N||^2}$. Consequently, in order to minimize $||R_j||^2$ the coefficient c_j^N to be neglected is the one minimizing (13). $\qquad \square$

Theorem 1 leads to a recursive algorithm for shrinking coefficients. We call such algorithm BOOMP, because, at each iteration, it selects the atom to be deleted

according to a selection criterion which is equivalent to the one proposed by OOMP [5] for forward approximations. Moreover, BOOMP is a natural complement of OOMP because its implementation on the output of OOMP is extremely simple. The few necessary steps are describe below.

BOOMP algorithm

Let us assume that atoms α_n ; $n = 1, \ldots, N$ have been selected by the OOMP approach in order to approximate a signal f. Hence, the biorthogonal set β_n^N and the corresponding coefficients c_n^N ; $n = 1, \ldots, N$ are also known. The BOOMP approach for reducing coefficients evolves as follows:

- Select the index j of the coefficient c_j^N to be disregarded as the one yielding a minimum value of the quantity $\frac{|c_j^N|^2}{||\beta_j^N||^2}$ as j ranges from 1 to N.

- Modify the corresponding biorthogonal atoms and coefficients as prescribed in (4) and (6) respectively.

- Set $N = N - 1$ and repeat the above steps until the coarsest acceptable approximation is reached.

Example

We illustrate now by a simple example the main remark of this communication: namely that BOOMP can improve the compression performance of the forward OOMP approach. We construct a dictionary of Mexican hat wavelets given by the functions

$$\alpha_{m,n}(t) = 2^{\frac{m}{2}} \alpha(t2^m - 0.2n) \qquad \text{with} \qquad \alpha(t) = \frac{2}{\sqrt{3}} \pi^{-\frac{1}{4}} (1 - t^2) e^{-\frac{t^2}{2}}. \qquad (15)$$

By considering scales $m = 0, 1, 2, 3$, and 4 to cover the $[0, 4]$ interval we have a dictionary of 665 atoms. The signal f to be represented is a chirp generated by the

8

MATLAB instructions:

```
t=0:0.01:4; f=chirp(t,0,1,2);
```

In order to produce a high quality representation of this chirp the OOMP approach selects $N = 60$ atoms (the norm of the residual error is 0.0544). By applying the proposed BOOMP to the OOMP approximation we reduce the number of coefficients up to 34 and the approximation is the one depicts in the top left graph of Figure 1. The norm of the residual error with respect to the true signal is in this case 1.18. However, if rather than applying the BOOMP approach we stop the OOMP approach at iteration 34, so as to have the same number of coefficients as in the previous case, the approximation is the one shown in the top right graph of Figure 1 and the norm of the residual error is 1.77. It is clear from the graphs that, in addition to yielding the smallest residual error with the same number of coefficients, the approximation obtained by the BOOMP approach is overall visually superior to the OOMP one with the same number of coefficients. The bottom left and right graph depict, respectively, the absolute value of the difference between the chirp signal and the corresponding approximations.

4   Conclusions

A recursive approach for shrinking coefficients of an atomic decomposition has been proposed. The approach is based on an adaptive technique which allows to modify biorthogonal functions in order to yield orthogonal projectors onto a reduced subspace. A criterion for disregarding coefficients has been discussed. Such criterion leads to an iterative procedure that we have termed BOOMP, because it evolves so as to fulfil identical requirements to those of the OOMP method. Accordingly, BOOMP provides at each iteration a) the coefficient of the atomic decomposition

to be deleted in order to construct a coarser approximation minimizing the norm of the residual error b) the coefficients of such decomposition rendering optimal approximation in the same sense. The approach is a good complement to OOMP, very simple to implement and definitely worth trying in any case. However it is appropriate to stress that situations for which the forward approach renders better results that the combination with the backward one should certainly exist. This is a consequence of lack of global optimality in both directions. Finally we would like to remark that if, rather than (13), one decided to apply another criterion for disregarding coefficients (see [13, 14] for some alternative ones) in order to leave an approximation minimizing the distance to the signal the remaining coefficients should be modified as prescribed in (6).

Acknowledgements

References

[1] S. Mallat and Z. Zhang, "Matching Pursuit in time-frequency dictionary", *IEEE Transactions on Signal Processing*, Vol 41, 3397–3415 (1993).

[2] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuits: recursive function approximation with applications to wavelet decompo-

sition", *Proceedings of the 27th Asilomar Conference in Signals, Systems, and Computers*, 1993.

[3] G. M. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations", *J. of Conts. Approx.*, Vol 13, 57–98 (1997).

[4] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic Decomposition by Basis Pursuit", SIAM, *Journal on Scientific Computing*, Vol 20, 1, 33–61 (1998).

[5] L. Rebollo-Neira and D. Lowe, "Optimised Orthogonal Matching Pursuit Approach", *IEEE Signal Processing Letters*, Vol 9, 4, 137–140 (2002).

[6] L. Rebollo-Neira, "Recursive biorthogonalization approach and orthogonal projectors", math-phi/0209026 (2002).

[7] A. DeVore, "Nonlinear approximation", *Acta Numer.*, 51–150 (1998).

[8] V. N. Temlyakov, "Greedy algorithms and M-term approximation with regard to redundant dictionaries", *Journal of Approximation Theory* Vol 98, 1, 117–145 (1999).

[9] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1998.

[10] M. Porat and Y. Y. Zeevi, "Gram-Gabor Approach to optimal image representation", *SPIE's Visual communications and image processing '90*, Vol 1360, 1474–1478 (1990).

[11] T. Genossar and M. Porat, "Optimal Bi-Orthonormal Approximation of Signals", *IEEE Trans. on Systems, Man. and Cybernetics*, Vol. SMC-22, No. 3, 449–460 (1992).

11

[12] L. Rebollo-Neira, "Backward Adaptive Biorthogonalization", *International Journal of Mathematics and Mathematical Sciences*, in press (2003), math-ph/0211066.

[13] A. Antoniadis and G. Oppenheim, Editors, *Lecture Notes in Statistics: Wavelets and Statistics*, Springer, 1995.

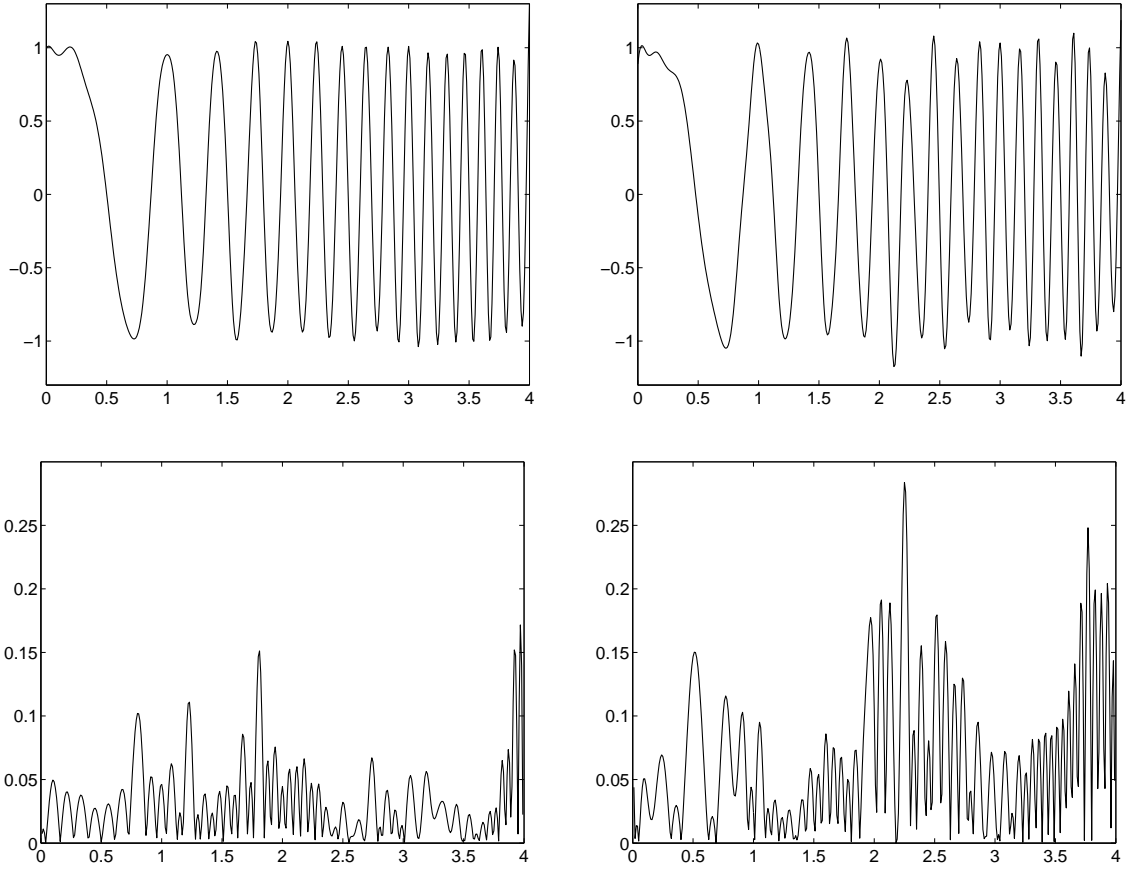[14] M. Jansen, *Noise reduction by wavelet thresholding,* New York, Springer, 2001.

Figure 1: The top left graph represents the approximation obtained by applying the BOOMP approach for reducing the 60 coefficients of the high quality OOMP approximation up to 34. The top right graph corresponds to the OOMP approximation resulting by stopping the approach after the selection of 34 atoms. The bottom left graph depicts the absolute value of the difference between the chirp signal and the BOOMP approximation of the left top graph. The right bottom graph has the same description as the left one, but with respect to the approximation of the right top graph.