IMPROVEMENT OF ORTHOGONAL MATCHING PURSUIT STRATEGIES BY BACKWARD AND FORWARD MOVEMENTS

Miroslav Andrle and Laura Rebollo-Neira

Aston University,
Birmingham B4 7ET, UK

ABSTRACT

An approach is presented generalising a recently introduced refinement of Orthogonal Matching Pursuit methodologies. The proposed strategy evolves by backward and forward movements which, if applied at the end of an Orthogonal Matching Pursuit method, are guaranteed to improve upon such an approach.

1. INTRODUCTION

We are concerned with the following nonlinear approximation problem: A given signal is approximated by a linear combination of elementary signals, called *atoms*, which are drawn from a large, and in general redundant, set called a *dictionary*. The linear combination is said to be 'sparse' if the number of atoms intervening in the approximation is 'small' in comparison to the number of sample points that would be needed to store the target signal. Over the last fifteen years the signal processing community has been interested in this problem and developed algorithms for finding sparse solutions [1–11].

When the dictionary is orthonormal, nonlinear approximation problems can be solved in practice without difficulties. Moreover, if the dictionary is sufficiently close to orthogonal (almost *incoherent*) the problem may still be tractable [12–14]. However, research indicates that approximation of signals using *coherent* redundant dictionaries offers a real gain in approximation quality. Unfortunately the problem of constructing the best approximation of a signal as a linear superposition of k atoms drawn from a coherent dictionary is an intractable NP-hard problem even if the dictionary is finite. Hence, it is normally addressed by heuristic stepwise greedy algorithms which do not seek for the optimal solution. We refer to those algorithms as greedy Pursuit Strategies. Since the seminal paper of Mallat and Zhang [2] introducing the Matching Pursuit (MP) method in signal processing a number of modifications to this approach have been proposed. Two basic variations improving MP convergence rate are Orthogonal Matching Pursuit (OMP) [3,4] and Optimised Orthogonal Matching Pursuit (OOMP) [8]. We refer to both methods as Orthogonal Matching Pursuit Strategies (OMPS).

In this communication we would like to extend the refinement of OMPS we have recently proposed [11]. If applicable, such a refinement, which is based on interchange of pair of atoms, is guaranteed to improve upon OMP or OOMP methods. The generalisation that we introduce here entails interchanges of more than two atoms. We show that, without increasing the complexity too much such generalisation may produce a significant gain in the sparseness of a representation. The paper is organised as follows: Section 2 introduces the proposed strategy improving OMPS results. In Section 3 the approach is illustrated by numerical simulations. The conclusions are drawn in Section 4.

2. SETTING UP THE PROBLEM

We deal mathematically with a signal, f, by considering it as an element of an inner product space, $\mathscr{H}$, equipped with an inner product $\langle\cdot,\cdot\rangle$ and induced norm $||\cdot|| = \langle\cdot,\cdot\rangle^{1/2}$. Thus, the distance between two signals in $\mathscr{H}$ is the norm of their difference.

A finite dictionary $\mathcal{D}$ is a finite collection of normalised to unity elementary signals (atoms). We denote each atom by α_i so that the corresponding dictionary is given as $\mathcal{D} = \{\alpha_i\}_{i\in\mathcal{I}}$, where $\mathcal{I}$ is a set of labels. The k-term approximation of f by k selected atoms $\{\alpha_{l_1},\ldots,\alpha_{l_k}\}$ is expressed in the form

$$f^{(k)} = \sum_{i=1}^{k} c_i^{(k)}\alpha_{l_i}. \tag{1}$$

The coefficients $\{c_i^{(k)}\}_{i=1}^k$ yielding the approximation $f^{(k)}$ which is the orthogonal projection of f onto the subspace $V_k = \mathrm{span}\{\alpha_{l_i}\}_{i=1}^k$ can be computed as: $c_i^{(k)} = \langle\beta_i^{(k)}, f\rangle$, $i = 1,\ldots,k$, where $\{\beta_i^{(k)}\}_{i=1}^k$ is the unique biorthogonal sequence, $\langle\beta_j^{(k)}, \alpha_{l_i}\rangle = \delta_{i,j}$, satisfying $\mathrm{span}\{\beta_i^{(k)}\}_{i=1}^k = V_k$. Thus, Eq. (1) can be recast in the fashion

$$f^{(k)} = \sum_{i=1}^{k} \alpha_{l_i}\langle\beta_i^{(k)}, f\rangle = \hat{P}_{V_k} f, \tag{2}$$

This work has been supported by EPSRC of UK (GR/R86355/01). Project: "Biorthogonal techniques for optimal signal representation," web page http://www.ncrg.aston.ac.uk/Projects/BiOrthog.

where $\hat{P}_{V_k}$ is the orthogonal projector operator onto V_k. This is the Least Square (LS) approximation of f in V_k because it minimises the square distance $||f - f^{(k)}||^2$. Nevertheless, finding the subspace V_k for which the k-term approximation of f is optimal in the LS sense is a NP-hard problem. In practice it is often addressed by heuristic selection criteria which are only stepwise optimal.

Consider for instance the problem of deciding how effectively interchange an element in (1) with an element from $\mathcal{D}$ so as to produce an improved k-term approximation of f. In order to use the LS selection criterion to eliminate an atom α_{l_j} from (1) we should find the index j for which the quantity

$$q_i = \frac{|c_i|^2}{||\beta_i^{(k)}||^2},\ i = 1, \ldots, k \tag{3}$$

is minimised [10]. Les us denote $V_{k\setminus j}$ to the subspace arising by removing the atom α_{l_j} from V_k, i.e.,

$$V_{k\setminus j} = \text{span}\{\alpha_{l_1}, \ldots, \alpha_{l_{j-1}}, \alpha_{l_{j+1}}, \ldots, \alpha_{l_k}\}.$$

The LS selection criterion of the atom to replace α_{l_j}, that we call OOMP criterion, entails to find (for $||\nu_n|| \neq 0$) the index in $\mathcal{I}$ for which the functional

$$e_n = \frac{|\langle \nu_n, f\rangle|^2}{||\nu_n||^2},\ \nu_n = \alpha_n - \hat{P}_{V_{k\setminus j}}\alpha_n, \tag{4}$$

is maximised. Here $\hat{P}_{V_{k\setminus j}}$ is the orthogonal projector onto the subspace $V_{k\setminus j}$. It is appropriate to point out that the OMP criterion would maximise $|\langle \nu_n, f\rangle|^2$ instead of e_n.

Since it readily follows [11] that

$$\hat{P}_{V_{k\setminus j}}\alpha_n = \hat{P}_{V_k}\alpha_n - \frac{\beta_j^{(k)}\langle \beta_j^{(k)}, \alpha_n\rangle}{||\beta_j^{(k)}||^2}, \tag{5}$$

the computation of the sequence ν_n in (4) is a simple operation provided that $\hat{P}_{V_k}$ and $\beta_j^{(k)}$ are available.

We call *backward step* to the process of eliminating one atom from (1) and *forward step* to the process of adding one atom in (1). The implementation details to realise both steps are given in [11] and the corresponding MATLAB codes are available at [15].

The refinement of OMPS we proposed in [11] consists of swapping of pairs of atoms realised by consecutive backward and forward steps which are executed at the end of an OMP method. The aim of the present effort is to show that the generalisation of the swapping procedures to involving, say r, backward steps followed by r forwards steps (or vise versa) with r ranging from 1 to a fixed value, may produce a significant gain in the sparseness of the representation without much increase of complexity.

In order to study the results produced by this approach we implement the r backward and forward movements for $r = 1, \ldots, r_{\max}$. For each r the backward and forward operations are repeated until the whole process, if performed once more, would not improve the approximation. Out of the results obtained for the different values of r we select the value r_{opt} yielding the minimum approximation error. We term the corresponding solution *the best approximation*. The operational steps are sketched in Alg. 1.

Algorithm 1 Sketch of the algorithm for backward and forward movements.

```
Apply OOMP method to choose k atoms approximating f
ε_0 = ||f − f^(k)||^2
for r = 1 : r_max
  Reset variables to the OOMP output
  ε'_r = ε_0
  repeat
    ε_r = ε'_r
    for s = 1 : r
      Calculate the functional q_i                {cf. Eq. (3)}
      j = arg min_i q_i                {Backward LS criterion}
      Apply backward step for α_{l_j}   {see [11, Alg. 1a]}
      ε'_r = ε'_r + q_j
    end
    for s = 1 : r
      Calculate the functional e_n                {cf. Eq. (4)}
      n_s = arg max_n e_n                {Forward LS criterion}
      Apply forward step for α_{n_s}    {see [11, Alg. 2a]}
      ε'_r = ε'_r − e_{n_s}
    end
  until ε'_r ≥ ε_r
end
r_opt = arg min_r ε_r
```

We also assess the gain in sparseness for each r-value by performing some extra backward steps [10] to eliminate atoms so as to produce an approximation of the same quality as the one available before the implementation of backward and forward movements. We term the approximation corresponding to the minimum number of atoms *the sparsest approximation*.

3. NUMERICAL SIMULATIONS

We illustrate the proposed approach by using a highly coherent cubic B-spline dictionary on the interval $[0, 8]$ which arises by merging two B-spline dictionaries with prototype atoms of support 2^{-2} and 2, respectively, translated by a distance 2^{-6}. The details on how to construct such dictionaries are given in [16] and the MATLAB codes available at [15].

The cumulative coherence function, $\mu_1(p)$, which is equal to the maximum absolute sum of the inner products between one atom and p other distinct atoms [14], is plotted in Fig. 1 for $p = 1, \ldots, 200$.

The signals to be approximated are constructed by taking randomly 200 dictionary atoms and combining them with random coefficients from $[-10, 10]$. In this way we generated

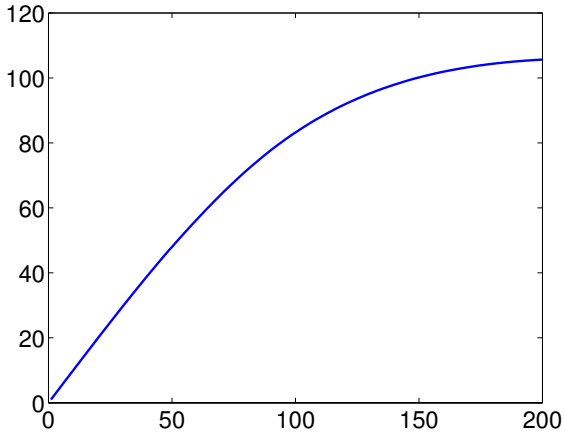

Fig. 1. The cumulative coherence function $\mu_1(p)$ for the B-spline dictionary in hand ($p = 1, \ldots, 200$).

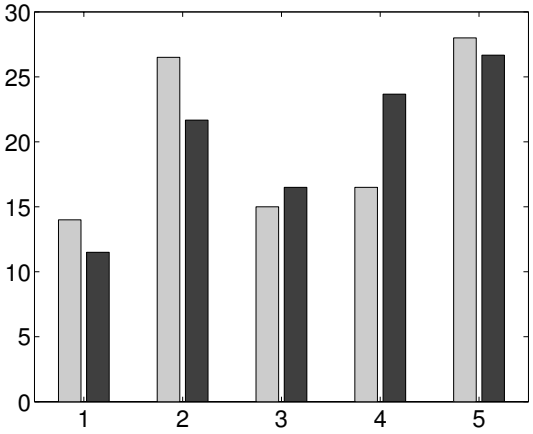

Fig. 2. Histogram showing the number of times that the best approximation occurs for a given value of r (light bars). Histogram showing the number of times that the sparsest approximation for the given error occurs vs r (dark bars).

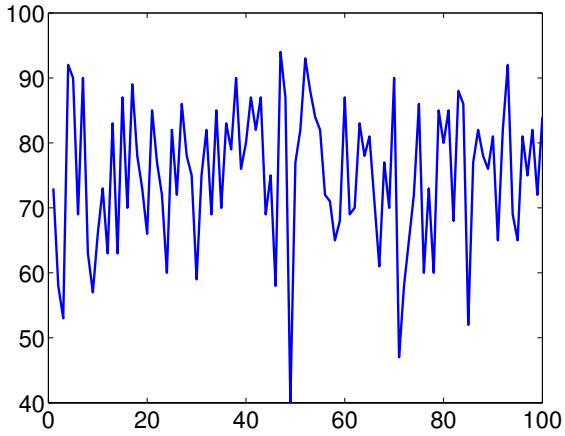

Fig. 3. The number of atoms that the sparsest approximation gains in relation to the OOMP approach.

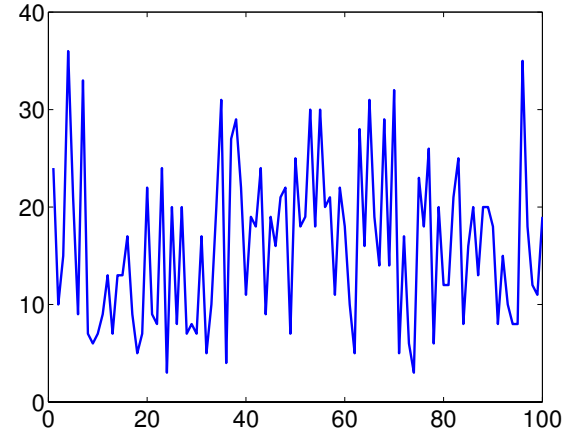

Fig. 4. The difference of the number of atoms involved in the sparsest and least sparse approximations.

100 different signals that we aim at approximating. Since the signals are restricted to be a linear combination of dictionary atoms the problem is *exact sparse* [14]. However, because our dictionary is highly coherent (see Fig. 1) neither OMP nor OOMP can find the exact solution. Hence we approximate each signal by using $k = 200$ atoms selected by the OOMP approach. Due to the fact that the approach does not solve the problem exactly we end with an approximation up to some error. For each simulation we applied the above described r backward and forward movements with $r = 1, \ldots, 5$. For each value of r the light bar in Fig. 2 represents the number of times that the best approximation occurs for that value of r. It should be pointed out that although the best approximation considerably improves the approximation error it is still not the exact solution.

The dark bar of Fig. 2 corresponds to the number of times that after applying some extra backward steps at the end of the proposed movements, in order to match the OOMP approximation error, the sparsest approximation for the given error occurs at the corresponding r-value.

The curve of Fig. 3 plots the number of atoms that the sparsest approximation gains in relation to the OOMP approach. Let us remark that in the 100 simulations we have run the mean value of the gain is 75 atoms. As a measure of the difference in sparseness that may be obtained by varying r, the curve of Fig. 4 plots the difference of the number of atoms involved in the sparsest and least sparse approximations. It is observed that such a difference is in some cases significant. In the 100 random simulations we present here the mean of that quantity is 16 atoms. It appears then that by letting r vary and choosing the best result a significant improvement in sparseness can be achieved. We should stress that the

proposed strategy does not have a big impact in complexity. As discussed in [11], the complexity of a backward/forward swap is roughly twice the complexity of a forward move in an OMPS. Thus, for each value of r the complexity of r-backward/forward movements is obtained by multiplying by r the complexity of a single backward/forward swap. Considering that these operations are implemented at the end of OMPS and are guaranteed to improve upon that output, we understand that the increment in complexity is not very significant in relation to the complexity of OMPS.

4. CONCLUSIONS

A generalisation of a recently introduced refinement of OMPS has been considered. The new strategy is based on interchanges of r atoms in an atomic decomposition with r dictionary atoms. The previous refinement is recovered by fixing $r = 1$. However, as was shown here, by letting r range from one to a fixed value $r_{\max}$ a significant gain in sparseness can be achieved. In order to stress this point 100 random simulations have been performed. It was found that by the sparsest approximation a mean value of 75 atoms can be saved to produce an approximation equivalent to that of OOMP. This result is impressive, considering that in all the signals the OOMP approach involves 200 atoms.

From our experiments is not possible to discriminate a universal optimum value for r. Moreover, it is appropriate to remark that there is not theoretical reason for the existence of such universal value. Our proposal is then to allow r range between one and some maximum value $r_{\max}$ in order to discriminate the optimal value for the particular situation. This implies an increment of complexity of approximately $r_{\max}$ times the complexity of the previous refinement corresponding to $r_{\max} = 1$.

Finally we would like to remark that, although the proposed r backward and forward movements could be implemented at any stage of the OOMP approach, we propose to realise the movements at the end of the OOMP process to be able to guarantee an improvement on OOMP if some movements can be realised.

5. REFERENCES

[1] S. Chen, C.F.N. Cowan, and P.M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Transaction on Neural Networks*, vol. 2, no. 2, pp. 302–309, 1991.

[2] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[3] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proceedings of the 27th Annual Asilomar Conference in Signals, System and Computers*, 1993, vol. 1, pp. 40–44.

[4] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Const. Approx.*, vol. 13, no. 1, pp. 57–98, 1997.

[5] B.D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Sig. Proc.*, vol. 47, no. 1, pp. 187–200, 1999.

[6] S.O. Aase, J.H. Husøy, K. Skretting, and K. Engan, "Optimized signal expansions for sparse representation," *IEEE Trans. on Signal Process.*, vol. 49, no. 5, pp. 1087–1096, 2001.

[7] S.J. Reeves, "An efficient implementation of the backward greedy algorithm for sparse signal reconstruction," *IEEE Signal Proc. Let.*, vol. 6, no. 10, pp. 266–268, 1999.

[8] L. Rebollo-Neira and D. Lowe, "Optimized orthogonal matching pursuit approach," *IEEE Signal Proc. Let.*, vol. 9, no. 4, pp. 137–140, 2002.

[9] R. Gribonval and P. Vandergheynst, "On the exponential convergence of matching pursuits in quasi-incoherent dictionaries," *IEEE Trans. Inform. Theory*, vol. 52, no. 1, pp. 255–261, 2006.

[10] M. Andrle, L. Rebollo-Neira, and E. Sagianos, "Backward-optimized orthogonal matching pursuit approach," *IEEE Signal Proc. Let.*, vol. 11, no. 9, pp. 705–708, 2004.

[11] M. Andrle and L. Rebollo-Neira, "A swapping-based refinement of orthogonal matching pursuit strategies," *Signal Processing*, in press.

[12] R. Gribonval and M. Nielsen, "Highly sparse representations from dictionaries are unique and independent of the sparseness measure," preprint, 2003.

[13] R. Gribonval, R. Figueras i Ventura, and P. Vandergheynst, "A simple test to check the optimality of a sparse signal approximation," *Signal Processing*, in press.

[14] J.A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[15] M. Andrle and L. Rebollo-Neira, "Biorthogonal techniques for optimal signal representation," 2003, Web page http://www.ncrg.aston.ac.uk/Projects/BiOrthog.

[16] M. Andrle and L. Rebollo-Neira, "Cardinal B-spline dictionaries on a compact interval," *Appl. Comput. Harmon. Anal.*, vol. 18, no. 3, pp. 336–346, 2005.