On the Difference Between Orthogonal Matching Pursuit and Orthogonal Least Squares

Thomas Blumensath, *Member, IEEE,* Mike E. Davies, *Member, IEEE*

***Abstract*— Greedy algorithms are often used to solve underdetermined inverse problems when the solution is constrained to be sparse, i.e. the solution is only expected to have a relatively small number of non-zero elements. Two different algorithms have been suggested to solve such problems in the signal processing and control community, orthogonal Matching Pursuit and orthogonal Least Squares respectively. In the current literature, there exist a great deal of confusion between the two strategies. For example, the later strategy has often be called orthogonal Matching Pursuit and has repeatedly been "re-discovered" in several papers. In this communication we try to pull together some of the literature and clarify the difference between the methods.**

***Index Terms*— Sparse signal approximation algorithms, orthogonal Matching Pursuit, orthogonal Least Squares**

I. Introduction

In this communication, we discuss two greedy strategies that were developed to solve the following problem. Given a vector $\mathbf{x} \in \mathbb{R}^{N_x}$ and a matrix $\mathbf{\Phi} \in \mathbb{R}^{N_x \times N_s}$, find a vector $\mathbf{s}$ such that the squared error is small, while $\mathbf{s}$ has only a small number of non-zero elements. For the discussion here, we use the term algorithm to mean any computational procedure that gives a particular result, i.e. we here discuss two different algorithms, which can be implemented using different computational steps.

A. Notation and theoretical overview

Let Γ^n be the index set of non-zero elements of $\mathbf{s}$, i.e. $\Gamma^n = \{i : s_1 \neq 0\}$, where s_i is the i^{th} element in the vector $\mathbf{s}$. We use this set to index sub-matrices of $\mathbf{\Phi}$ that only contain those columns with indices in Γ^n, i.e $\mathbf{\Phi}_{\Gamma^n}$ is a sub-matrix of $\mathbf{\Phi}$. We use the same notation for sub-vectors, i.e. $\mathbf{s}_{\Gamma^n}$ is a sub-vector of $\mathbf{s}$. The individual columns of $\mathbf{\Phi}$ will be the vectors ϕ_i. We assume throughout that $\|\phi_i\| = 1$.

With this notation, Orthogonal Matching Pursuit (OMP) and Orthogonal Least Squares (OLS), both 'greedily' build up the set Γ^n, adding a single element in each iteration. After updating Γ^n, both algorithms approximate $\mathbf{x}$ using:

$$\mathbf{x} \approx \hat{\mathbf{x}}^n \mathbf{\Phi}_{\Gamma^n} \mathbf{s}^n_{\Gamma^n} = \mathbf{\Phi}_{\Gamma^n} \mathbf{\Phi}^{\dagger}_{\Gamma^n} \mathbf{x}, \tag{1}$$

where $\mathbf{\Phi}^{\dagger}_{\Gamma^n}$ is the More-Penrose pseudo inverse of $\mathbf{\Phi}_{\Gamma^n}$. Therefore, $\mathbf{\Phi}_{\Gamma^n} \mathbf{\Phi}^{\dagger}_{\Gamma^n} \mathbf{x}$ is the closest approximation (in the Euclidean distance) to $\mathbf{x}$ using linear combinations of the columns in $\mathbf{\Phi}_{\Gamma^n}$.

The difference in the algorithms is in how they select the new index to be added to Γ in each iteration.

B. Confusion in the literature

Before discussing the algorithms in more details we here review the literature on the subject. The algorithm we call Orthogonal Matching Pursuit was first proposed in the signal processing literature by Davis, Mallat and Zhang in [1] and at the same time appeared in a paper by Pati, Rezaiifar and Rrishnaprasad [2]. The algorithm is discussed in some detail in the book [3], which greatly contributed to its current popularity.

A trail of confusion can, however, be traced back to [1], where the authors state "*This type of algorithm was first introduced for control applications [4]*". However, the algorithm proposed by Chen, Billinges and Luo in [4], which is the algorithm we call orthogonal Least Squares, differs from OMP as we will show below.

OLS has not played a prominent role in the signal processing literature, which led to the repeated discovery of this algorithm by a range of authors over the years. For example, the optimized orthogonal Matching Pursuit algorithm by Rebollo-Neira and Lowe [5] as well as the algorithm proposed by Natarajan [6] are incarnations of OLS. The paper "A fast orthogonal matching pursuit algorithm" by Gharavi-Alkhansari and Huang [7] has not helped the matter, as the discussed algorithm is not the OMP algorithm of [2] and [1] as the name of the paper suggests, but is in fact the OLS algorithm of [4].

OLS type algorithms have also appeared in the statistics literature, for example the forward selection algorithm [8] [9] [10] has been suggested for the subset selection problem in regression. This method is also sometimes known as forward selection, a name that is however normally used for a method that also allows elements to be deleted [8]. Again, OLS and OMP are not commonly distinguished, which has led many prominent authors to make statements of the form " *Orthogonal Matching Pursuit, (also called (...) stepwise regression in other fields)* ", [11].

II. Orthogonal Matching Pursuit

Orthogonal Matching Pursuit was developed as an improvement to Matching Pursuit [12] [1]. It therefore shares many of the properties of Matching Pursuit. In particular, the defining quality for us, which differentiates this method from OLS, is

The authors are with IDCOM & Joint Research Institute for Signal and Image Processing, Edinburgh University, King's Buildings, Mayfield Road, Edinburgh EH9 3JL, UK (Tel.: +44(0)131 6505659, Fax.: +44(0)131 6506554, e-mail: thomas.blumensath@ed.ac.uk, mike.davies@ed.ac.uk).

This research was supported by EPSRC grant D000246/1. MED acknowledges support of his position from the Scottish Funding Council and their support of the Joint Research Institute with the Heriot-Watt University as a component part of the Edinburgh Research Partnership.

the selection procedure, which it has inherited from Matching Pursuit.

In each iteration, orthogonal Matching Pursuit calculates a new signal approximation $\hat{\mathbf{x}}^n$. The approximation error $\mathbf{r}^n = \mathbf{x} - \hat{\mathbf{x}}^n$ is then used in the next iteration to determine which new element is to be selected. In particular, the selection is based on the inner products between the curent residual $\mathbf{r}^n$ and the colum vectors ϕ_i of $\mathbf{\Phi}$. Let these inner products be

$$\alpha_i^n = \phi_i^T r^n. \tag{2}$$

The new element is then selected for which the magnitude of α_i^n is largest, i.e.

$$i_{max}^n = \arg_i \max |\alpha_i^n| \tag{3}$$

so that

$$\Gamma^{n+1} = \Gamma^n \cup i_{max}^n. \tag{4}$$

OMP can therefore be summarised as:

- Initialise: $\mathbf{r}^0 = \mathbf{x}$, $\mathbf{s}^0 = \mathbf{0}$, $\Gamma^0 = \emptyset$
- for $n = 1; n := n + 1$ till stopping criterion is met
  - $\alpha_i = \phi_i^T r^{n-1}$ for all $i \notin \Gamma^{n-1}$
  - $i_{max} = \arg_i \max |\alpha_i|$
  - $\Gamma^n = \Gamma^{n-1} \cup i_{max}$
  - $\mathbf{s}_{\Gamma^n}^n = \mathbf{\Phi}_{\Gamma^n}^\dagger \mathbf{x}$
  - $\mathbf{r}^n = x - \mathbf{\Phi}\mathbf{s}^n$

The algorithm as described here does not represent a usable strategy in many applications, as we have here written the algorithm in a form that requires the solution to an inverse problem in each iteration. The most efficient implementation of the strategy, which, however, requires some additional storage, is based on QR factorisation [1]. Let $\mathbf{\Phi}_{\Gamma^n} = \mathbf{Q}_{\Gamma^n}\mathbf{R}_{\Gamma^n}$ be such a factorisation, where $\mathbf{Q}_{\Gamma^n}^T\mathbf{Q}_{\Gamma^n} = \mathbf{I}$ is the identity and where $\mathbf{R}_{\Gamma^n}$ is upper triangular, further let $\mathbf{z}_{\Gamma^n} = \mathbf{R}_{\Gamma^n}\mathbf{s}_{\Gamma^n}$, then the orthogonal Matching Pursuit solution can be calculated as:

1) Initialise $\mathbf{r}^0 = \mathbf{x}, \mathbf{z}^0 = 0, \Gamma^0 = \oslash$
2) for $n = 1; n := n + 1$ till stopping criterion is met
   a) $\alpha_i = \phi_i^T r^{n-1}$ for all $i \notin \Gamma^{n-1}$
   b) $i_{max} = \arg_i \max |\alpha_i|$
   c) $\Gamma^n = \Gamma^{n-1} \cup i_{max}$
   d) Update $\mathbf{Q}_{\Gamma^n}$ and $\mathbf{R}_{\Gamma^n}$ such that $\mathbf{Q}_{\Gamma^n}\mathbf{R}_{\Gamma^n} = \mathbf{\Phi}_{\Gamma^n}$, $\mathbf{Q}_{\Gamma^n}^T\mathbf{Q}_{\Gamma^n} = \mathbf{I}$ and $\mathbf{R}_{\Gamma^n}$ is upper triangular.
   e) $\mathbf{z}_{\Gamma^n} = [\mathbf{z}_{\Gamma^{n-1}}; z^n]$
   f) $\mathbf{r}^n = \mathbf{r}^{n-1} - z^n\mathbf{q}$
3) $\mathbf{s}_{\Gamma^n} = \mathbf{R}_{\Gamma^n}^{-1}\mathbf{z}_{\Gamma^n}$
4) Output: $\mathbf{r}^n, \mathbf{s}^n$

Here $z^n = \mathbf{q}^T\mathbf{x}$ and $\mathbf{q}$ is the new column vector added to $\mathbf{Q}_{\Gamma^n}$ in the current iteration. The required QR factorisation is most efficiently implemented using a form of modified Gram-Schmidt procedure [13], however, contrary to the modified Gram-Schmidt method, vectors are only orthogonalised, once they have been selected and are to be added to $\mathbf{Q}$. Using this approach, one needs to orthogonalise a single vector in each iteration in order to add a new column to $\mathbf{Q}$. Note also that we do not need to calculate $\mathbf{s}^n$ until the last iteration, in which we can invert $\mathbf{R}$ using back substitution[13].

III. Orthogonal Least Squares

The selection step used in OLS differs from the one used in Matching Pursuit and OMP in that it selects the vector ϕ_i, that will lead to the minimum residual error after orthogonalisation. It is important to realise that the OMP selection procedure does not select that element that, after orthogonal projection of the signal onto the selected elements, minimise the residual norm (see for example the discussion in [5]). Though, once the elements are selected, both OMP and OLS minimise the residual *given the selected elements*.

Using this selection procedure, we can write the OLS algorithm as:

- Initialise: $\mathbf{r}^0 = \mathbf{x}$, $\mathbf{s}^0 = \mathbf{0}$, $\Gamma^0 = \emptyset$
- for $n = 1; n := n + 1$ till stopping criterion is met
  - $i_{max} = \arg_i \min_{\Gamma^n:\Gamma^n=\Gamma^{n-1}\cup i} \|\mathbf{x} - \mathbf{\Phi}_{\Gamma_i^n}\mathbf{\Phi}_{\Gamma_i^n}^\dagger\mathbf{x}\|_2$
  - $\Gamma^n = \Gamma^{n-1} \cup i_{max}$
  - $\mathbf{s}_{\Gamma^n}^n = \mathbf{\Phi}_{\Gamma^n}^\dagger \mathbf{x}$
  - $\mathbf{r}^n = x - \mathbf{\Phi}\mathbf{s}^n$

Again, faster implementations are based on QR factorisation [4], however, one now uses the Modified Gram-Schmidt procedure to calculate the QR factorisation, i.e. in each iteration, the dictionary $\mathbf{\Phi}$ is split into two parts, $\mathbf{\Phi}_{\Gamma^n}$ and $\mathbf{\Phi}_{\Lambda^n}$, where the set Λ^n now contains all those indices not in Γ^n. The Modified Gram-Schmidt procedure [13] then iteratively updates a QR factorisation of $\mathbf{\Phi}_{|Gamma^n} = \mathbf{Q}_{\Gamma^n}\mathbf{R}_{\Gamma^n}$ as with OMP, however, at the same time, it keeps a modified copy of $\mathbf{\Phi}_{\Lambda^n}$, say $\hat{\mathbf{\Phi}}_{\Lambda^n}$, whose columns are the columns of $\mathbf{\Phi}_{\Lambda^n}$ that have been made orthogonal to all elements in $\mathbf{Q}_{\Gamma^n}$ and, crucially, that have been *re-normalised*. OLS can then be written as:

1) Initialise $\mathbf{r}^0 = \mathbf{x}, \mathbf{z}^0 = 0, \Gamma^0 = \oslash$
2) for $n = 1; n := n + 1$ till stopping criterion is met
   a) $\alpha_i = \hat{\phi}_i^T r^{n-1}$ for all $i \in \Lambda^{n-1}$
   b) $i_{max} = \arg_i \max |\alpha_i|$
   c) $\Gamma^n = \Gamma^{n-1} \cup i_{max}$
   d) Update $\mathbf{Q}_{\Gamma^n}$ and $\mathbf{R}_{\Gamma^n}$ such that $\mathbf{Q}_{\Gamma^n}\mathbf{R}_{\Gamma^n} = \mathbf{\Phi}_{\Gamma^n}$, $\mathbf{Q}_{\Gamma^n}^T\mathbf{Q}_{\Gamma^n} = \mathbf{I}$ and $\mathbf{R}_{\Gamma^n}$ is upper triangular.
   e) Update $\hat{\mathbf{\Phi}}_{\Lambda^n}$
   f) $\mathbf{z}_{\Gamma^n} = [\mathbf{z}_{\Gamma^{n-1}}; z^n]$
   g) $\mathbf{r}^n = \mathbf{r}^{n-1} - z^n\mathbf{q}$
3) $\mathbf{s}_{\Gamma^n} = \mathbf{R}_{\Gamma^n}^{-1}\mathbf{z}_{\Gamma^n}$
4) Output: $\mathbf{r}^n, \mathbf{s}^n$

The important difference here to OMP is that we calculate the inner products used for the selection of elements using the columns of the matrix $\hat{\mathbf{\Phi}}_{\Lambda^n}$.

IV. A geometric interpretation of the selection procedure

To further clarify the difference in the selection procedure between the two algorithms it is beneficial to look at the graphical representation given in figure 1.

Assume that the previous signal approximation is along the vertical axis. The current residual (Labelled by r in the figure) is then orthogonal to this direction, i.e. it has to lie within the orthogonal subspace indicated by the shaded area. Importantly, the columns of $\mathbf{\Phi}_{\Lambda^n}$, i.e. these elements of $\mathbf{\Phi}$ that have not

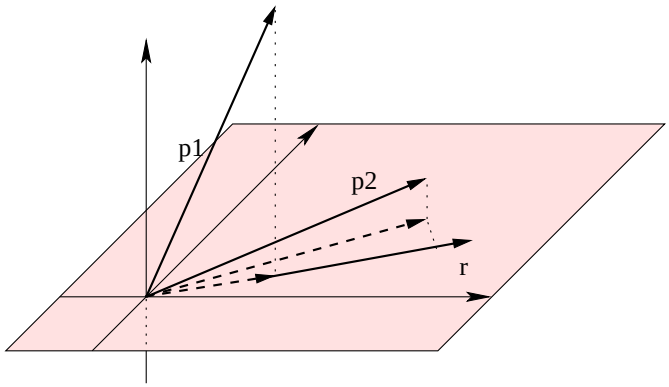

Fig. 1. Graphical representation of the problem.

so far been selected, do not have to lie within this subspace. We have here drawn two of these vectors as examples, these are labels p1 and p2 in the figure.

OMP selects the new element based on the inner product, i.e. based on the angle between the vectors and the current residual[1]. In the figure, this would be element p2.

OLS, on the contrary, selects the element that is able to best approximate the current residual, i.e. OLS will select the element with the smallest angle *after* this element is projected into the orthogonal subspace. The projection of p1 and p2 into the orthogonal subspace is here shown with dashed lines. It is clear that the closest angle after projection, is here with p1 and not p2. (Note that we have here not drawn the normalised projections. The normalisation is really only required to facilitate the calculation of the angle, which is only equivalent to the inner product formulation used here when the vectors all are of equal length.)

V. A word on computational complexity

In general, implementing OLS using the QR factorisation as discussed here, is more costly than using the QR factorisation for OMP, whenever the number of selected elements is smaller than the number of columns in $\mathbf{\Phi}$, because in the OLS implementation one has to orthogonalise all elements, while in OMP one only orthogonalises those elements which are being selected.

VI. Discussion and conclusion

Greedy algorithms are used increasingly in signal processing. In this paper we have reviewed two common methods, orthogonal Matching Pursuit and orthogonal Least Squares. We have discussed the difference in the greedy selection step. The similarity of the approaches has led to some confusion in recent literature and we hope that this communication contributes to the clarification of some points. In a nut shell, in order to distinguish between the two algorithms, which can appear in many different disguises, such as the QR factorisation based implementation discussed here, one has to take a closer look at the selection procedure. If the selection procedure is based on the largest inner product with the *original* elements in the dictionary, the algorithm is OMP. This selection step can be disguised if the implementation is based on a form of Gram-Schmidt orthogonalisation, in which not-selected elements are orthogonalised to the selected elements. In this case, if the elements are *not* normalised when taking the inner product, then the algorithm selects in the same manner as OMP, normalising the orthogonalised elements before (or during) the calculation of the inner products leads to the OLS selection procedure, which is guaranteed to select the element resulting in the smallest error after projection.

[1]Remember that we assume the ϕ_i to have unit norm, which implies that the angle is proportional to the inner product.

References

[1] S. Mallat, G. Davis, and Z. Zhang, "Adaptive time-frequency decompositions," *SPIE Journal of Optical Engineering*, vol. 33, pp. 2183–2191, July 1994.
[2] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in 27^th *Asilomar Conf. on Signals, Systems and Comput.*, Nov. 1993.
[3] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1999.
[4] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification.," *International Journal of Control*, vol. 50, no. 5, pp. 1873–1896, 1989.
[5] L. Rebollo-Neira and D. Lowe, "Optimized orthogonal matching pursuit approac," *IEEE Signal Processing Letters*, vol. 9, pp. 137–140, Apr 2002.
[6] B. K. Nataratjan, "Sparse approximat solutions to linear systems," *SIAM Journal on Computing*, vol. 24, pp. 227–234, Apr 1995.
[7] M. Gharavi-Alkhansari and T. S. Huang, "A fast orthogonal matching pursuit algorithm," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1389–1392, 1998.
[8] N. R. Draper and H. Smith, *Applied Regression Analysis*. Joh Wiley and Sons, 1966.
[9] R. R. Hocking, "The analysis and selection of variables in linear regression," *Biometrics*, vol. 32, no. 1, pp. 1–49, 1976.
[10] A. Miller, *Subset selection in regression*. Chapman and Hall, 2nd ed., 2002.
[11] D. Donoho, Y. Tsaig, I. Drori, and J. Starck, "Sparse solutions of underdetermined linear equations by stagewise orthogonal matching pursuit.," 2006.
[12] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.
[13] G. H. Golub and F. Van Loan, *Matrix Computations*. Johns Hopkins University Press, 3rd ed., 1996.