

# Group Polytope Faces Pursuit for recovery of block-sparse signals

Aris Gretsistas and Mark D. Plumbley\*

Queen Mary University of London  
Centre for Digital Music  
Mile End Road, E1 4NS, London, UK  
aris.gretsistas@eecs.qmul.ac.uk

**Abstract.** Polytope Faces Pursuit is an algorithm that solves the standard sparse recovery problem. In this paper, we consider the case of block structured sparsity, and propose a novel algorithm based on the Polytope Faces Pursuit which incorporates this prior knowledge. The so-called Group Polytope Faces Pursuit is a greedy algorithm that adds one group of dictionary atoms at a time and adopts a path following approach based on the geometry of the polar polytope associated with the dual linear program. The complexity of the algorithm is of similar order to Group Orthogonal Matching Pursuit. Numerical experiments demonstrate the validity of the algorithm and illustrate that in certain cases the proposed algorithm outperforms the Group Orthogonal Matching Pursuit algorithm.

**Keywords:** block-sparsity, polytopes, sparse representations.

## 1 Introduction

Over recent years, the study of sparse representations [1] has seen an increasing interest among researchers and its significance has been highlighted in numerous signal processing applications ranging from signal acquisition to de-noising and from coding to source separation. Sparse representations are signal expansions that can accurately represent the signal of interest using a linear combination of a relatively small number of significant coefficients drawn from a basis or a redundant dictionary.

Let  $\mathbf{y} \in \mathbb{R}^M$  be the observed vector that we need to decompose and represent in the dictionary  $\mathbf{A}$  of size  $M \times N$  with  $M < N$  using a small number  $K$  of significant coefficients corresponding to the columns of the full rank matrix  $\mathbf{A}$ . The sparse representation problem can then be formulated:

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{1}$$

where  $\mathbf{x} = [x_1, \dots, x_N]^T$  is a  $K$ -sparse vector, namely it has only  $K = \|\mathbf{x}\|_0$  non-zero entries, with  $K \ll N$ . The above system of linear equations is said

---

\* This research is supported by ESPRC Leadership Fellowship EP/G007144/1 and EU FET-Open Project FP7-ICT-225913 "SMALL".

to be an underdetermined system, as the number of unknowns is larger than the number of equations. Such a system yields an infinite number of solutions. In *sparse coding* we are interested in obtaining the sparsest solution which has the smallest number of non-zero elements. Two well studied algorithms that can recover under certain conditions the sparse vector  $\mathbf{x}$  in equation (1) are *Basis Pursuit* (BP) [2] and *Orthogonal Matching Pursuit* (OMP) [3].

The conventional sparsity model assumes that the non-zero coefficients can be located anywhere in the sparse vector. However, block structures, which imply that the non-zero elements are grouped in blocks (or clusters) instead of being arbitrarily located throughout the vector  $\mathbf{x}$ , can appear in practical scenarios. More specifically, the sparse coefficients in multi-band signals [4] or harmonic signals [5] can be clustered in groups of dictionary atoms. In that special case of structured sparsity the block-sparse vector  $\mathbf{x}$  is treated as a concatenation of blocks of length  $d$ :

$$\mathbf{x} = \underbrace{[x_1 \dots x_d]}_{\mathbf{x}^T[1]} \underbrace{[x_{d+1} \dots x_{2d}]}_{\mathbf{x}^T[2]} \dots \underbrace{[x_{N-d+1} \dots x_N]}_{\mathbf{x}^T[P]} \quad (2)$$

where  $\mathbf{x}[p]$  denotes the  $p$ -th block and  $N = Pd$ . In [6] the block  $k$ -sparse vector is defined as the vector  $\mathbf{x} \in \mathbb{R}^N$  that has non-zero  $\ell_2$  norm for at most  $k$  indices out of  $P$ , namely:

$$\|\mathbf{x}\|_{2,0} = \sum_{p=1}^P I(\|\mathbf{x}[p]\|_2 > 0) \leq k \quad (3)$$

where  $I(\cdot)$  is the indicator function.

It follows that the redundant dictionary  $\mathbf{A}$  can also be represented as a concatenation of  $P$  block matrices:

$$\mathbf{A} = \underbrace{[\mathbf{a}_1 \dots \mathbf{a}_d]}_{\mathbf{A}^T[1]} \underbrace{[\mathbf{a}_{d+1} \dots \mathbf{a}_{2d}]}_{\mathbf{A}^T[2]} \dots \underbrace{[\mathbf{a}_{N-d+1} \dots \mathbf{a}_N]}_{\mathbf{A}^T[P]} \quad (4)$$

where  $\mathbf{A}[p]$  denotes the  $p$ -th column block matrix of size  $M \times d$ .

In order to solve the problem in equation (1) one can attempt the minimization of the mixed  $\ell_2/\ell_1$  norm [6]:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_{2,1} \quad \text{such that} \quad \mathbf{y} = \mathbf{A}\mathbf{x} \quad (5)$$

where  $\|\mathbf{x}\|_{2,1} = \sum_{p=1}^P \|\mathbf{x}[p]\|_2$ . Moreover, greedy algorithms can serve as alternatives to the optimization in equation (5) e.g. *Group Orthogonal Matching Pursuit* (G-OMP) [6].

## 2 Review of the Polytope Faces Pursuit algorithm

In this section we will review the original *Polytope Faces Pursuit* (PFP) algorithm, which we will generalize to group form in section 3. The traditional  $\ell_1$ -minimization problem can be converted to its *standard form* using nonnegative

coefficients:

$$\min_{\tilde{\mathbf{x}}} \mathbf{1}^T \tilde{\mathbf{x}} \quad \text{such that} \quad \mathbf{y} = \tilde{\mathbf{A}} \tilde{\mathbf{x}}, \tilde{\mathbf{x}} \geq 0 \quad (6)$$

where  $\mathbf{1}$  is a column vector of ones,  $\tilde{\mathbf{A}} = [\mathbf{A}, -\mathbf{A}]$  and  $\tilde{\mathbf{x}}$  is the  $2N$  nonnegative vector:

$$\tilde{\mathbf{x}}_i = \begin{cases} \max(x_i, 0) & 1 \leq i \leq N \\ \max(-x_{i-N}, 0) & N+1 \leq i \leq 2N. \end{cases} \quad (7)$$

The new linear program has a corresponding *dual linear program*:

$$\max_{\mathbf{c}} \mathbf{y}^T \mathbf{c} \quad \text{such that} \quad \tilde{\mathbf{A}}^T \mathbf{c} \leq \mathbf{1} \quad (8)$$

such that a bounded solution to (8) exists if and only if a bounded solution to (6) exists. Thus, we can initially look for a solution  $\mathbf{c}^*$  to (8) and use the Karush-Kuhn-Tucker (KKT) [7] conditions to solve the resulting system for  $\mathbf{x}^*$ .

The algorithm in an iterative fashion adds one vector at a time, the one with the maximum scaled correlation:

$$\mathbf{a}^k = \arg \max_{\mathbf{a}_i \notin \tilde{\mathbf{A}}^k} \frac{\mathbf{a}_i^T \mathbf{r}^{k-1}}{1 - \mathbf{a}_i^T \mathbf{c}^{k-1}}. \quad (9)$$

After updating the solution vector  $\tilde{\mathbf{x}}$  and the corresponding  $\mathbf{c}$  the algorithm iterates until the stopping criteria is met. The full PFP algorithm is given in [8].

### 3 Recovery of block-sparse signals via Group Polytope Faces Pursuit

#### 3.1 Group selection criterion

As has been described in [8], the Polytope Faces Pursuit algorithm, based on the conventional sparsity model, starts at  $\mathbf{c} = 0$  and adopts a path following approach towards the residual until it hits a face of the *polar polytope*  $P^* = \{\mathbf{c} \mid \pm \mathbf{a}_i^T \mathbf{c} \leq 1, \mathbf{a}_i \in \mathbf{A}\}$ , which is dual to the *primal polytope*  $P = \text{conv}\{\pm \mathbf{a}_i, \mathbf{a}_i \in \mathbf{A}\}$ . The next face encountered is the one along the current face towards the projected residual. More specifically, the path of the PFP algorithm at the  $k$ -th iteration can be defined as:

$$h^k = \mathbf{a}_i^T (\mathbf{c}^k + \alpha \mathbf{r}^k). \quad (10)$$

The next face will be encountered for the minimum  $\alpha$  such that  $h^k = 1$ . A little manipulation of this condition leads to the maximum scaled correlation of equation (9) as the atom selection criterion of the PFP algorithm.

In order to extend this to the block sparsity case, inspired from the work in [9] which proposes an implementation of the group LARS algorithm, at each step of the algorithm we are looking for a minimum  $\alpha$  such that:

$$\|\tilde{\mathbf{A}}[i]^T (\mathbf{c}^k + \alpha \mathbf{r}^k) - \mathbf{1}\|_2^2 = 0 \quad \text{for} \quad i = 1, \dots, P \quad (11)$$

where  $\tilde{\mathbf{A}}[i] = [\mathbf{A}[i], -\mathbf{A}[i]]$  is the  $M \times 2d$  doubled block matrix and in the above expression we consider only  $d$  atoms for which the inner product with the residual is nonnegative. After computations we end up with the following second order polynomial:

$$\lambda^2 \|\mathbf{1} - \tilde{\mathbf{A}}[i]^T \mathbf{c}\|_2^2 - 2\lambda(\mathbf{1}^T - \tilde{\mathbf{A}}[i] \mathbf{c}^T) \tilde{\mathbf{A}}[i]^T \mathbf{r} + \|\tilde{\mathbf{A}}[i]^T \mathbf{r}\|_2^2 = 0 \quad (12)$$

where  $\lambda = 1/\alpha$ . The discriminant of the above quadratic polynomial is given:

$$\Delta = 4((\mathbf{1}^T - \tilde{\mathbf{A}}[i] \mathbf{c}^T) \tilde{\mathbf{A}}[i]^T \mathbf{r})^2 - \|\tilde{\mathbf{A}}[i]^T \mathbf{r}\|_2^2 \|\mathbf{1} - \tilde{\mathbf{A}}[i]^T \mathbf{c}\|_2^2. \quad (13)$$

The discriminant of the polynomial of equation (12) will always be less or equal to zero, and therefore the polynomial will have two complex conjugate solutions. Considering that due to the nonnegative constraint of the solution vector we require that  $\tilde{\mathbf{A}}[i]^T \mathbf{r} > \mathbf{0}$  and also that it always holds  $\mathbf{1} - \tilde{\mathbf{A}}[i]^T \mathbf{c} \geq \mathbf{0}$ , it is straightforward to show that  $(\mathbf{1}^T - \tilde{\mathbf{A}}[i] \mathbf{c}^T) \tilde{\mathbf{A}}[i]^T \mathbf{r} = \|(\mathbf{1}^T - \tilde{\mathbf{A}}[i] \mathbf{c}^T) \tilde{\mathbf{A}}[i]^T \mathbf{r}\|_2 \leq \|\tilde{\mathbf{A}}[i]^T \mathbf{r}\|_2 \|\mathbf{1} - \tilde{\mathbf{A}}[i]^T \mathbf{c}\|_2$ , where the Cauchy-Schwarz inequality has been used. Therefore, it follows that  $\Delta \leq 0$ . Consequently, the Group Polytope Faces Pursuit (G-PFP) algorithm at each iteration will have to choose the group of dictionary atoms with the maximum  $\lambda$ , where:

$$\lambda = \frac{(\mathbf{1}^T - \tilde{\mathbf{A}}[i] \mathbf{c}^T) \tilde{\mathbf{A}}[i]^T \mathbf{r} \pm j \sqrt{\|\tilde{\mathbf{A}}[i]^T \mathbf{r}\|_2^2 \|\mathbf{1} - \tilde{\mathbf{A}}[i]^T \mathbf{c}\|_2^2 - ((\mathbf{1}^T - \tilde{\mathbf{A}}[i] \mathbf{c}^T) \tilde{\mathbf{A}}[i]^T \mathbf{r})^2}}{\|\mathbf{1} - \tilde{\mathbf{A}}[i]^T \mathbf{c}\|_2^2}. \quad (14)$$

In order to simplify equation (14) we take the squared absolute value of the complex conjugate solution and the group selection criterion of the G-PFP algorithm reduces to:

$$\tilde{\mathbf{A}}[i]^k = \arg \max_{\mathbf{A}[i] \notin \tilde{\mathbf{A}}^*} \frac{\|\tilde{\mathbf{A}}[i]^T \mathbf{r}^{k-1}\|_2}{\|\mathbf{1} - \tilde{\mathbf{A}}[i]^T \mathbf{c}^{k-1}\|_2}. \quad (15)$$

Note that when the block size is  $d = 1$  equation (15) reduces to the maximum scaled correlation of equation (9). In the next section we derive the dual linear program for group sparse signals and show that there exists an optimum primal-dual  $(\mathbf{x}^*, \mathbf{c}^*)$  pair.

### 3.2 Dual linear program of the group sparse recovery problem

The Lagrangian to the problem of equation (5) is:

$$\mathcal{L}(\mathbf{x}, \mathbf{c}) = \|\mathbf{x}\|_{2,1} - \mathbf{c}^T (\mathbf{A}\mathbf{x} - \mathbf{y}) \quad (16)$$

and subsequently, the differential of  $\mathcal{L}$  with respect to  $\mathbf{x}$  is:

$$\partial_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \mathbf{c}) = \partial_{\mathbf{x}} \|\mathbf{x}\|_{2,1} - \mathbf{A}^T \mathbf{c}. \quad (17)$$

It can easily be shown that the subdifferential  $\partial_{\mathbf{x}}\|\mathbf{x}\|_{2,1}$  is given by the expression  $\partial_{\mathbf{x}}\|\mathbf{x}\|_{2,1} = \mathbf{x}[p]/\|\mathbf{x}[p]\|_2$  when  $\|\mathbf{x}[p]\|_2 > 0$ . However, for the zero block-elements of  $\mathbf{x}$  the gradient is not defined, but  $\partial_{\mathbf{x}}\|\mathbf{x}\|_2$  coincides with the set of unit  $\ell_2$  norm vectors  $\mathcal{B}_{\ell_2}^r = \{\mathbf{u} \in \mathbb{R}^r \mid \|\mathbf{u}\|_2 \leq 1\}$  [10]. Therefore, for each  $p = 1, \dots, P$ , we have:

$$\partial_{\mathbf{x}}\|\mathbf{x}\|_{2,1} = \begin{cases} \mathbf{x}[p]/\|\mathbf{x}[p]\|_2 & \|\mathbf{x}[p]\|_2 > 0 \\ \mathcal{B}_{\ell_2}^r & \text{otherwise.} \end{cases} \quad (18)$$

It follows that  $\partial_{\mathbf{x}}\|\mathbf{x}\|_{2,1} \leq \mathbf{1}$ . The KKT conditions require that  $\mathbf{A}\mathbf{x} = \mathbf{y}$  and  $\partial_{\mathbf{x}}\mathcal{L}(\mathbf{x}, \mathbf{c}) = 0$ . Substituting equations (17) and (18) to the last expression we get the dual to the problem of equation (5):

$$\max_{\mathbf{c}} \mathbf{y}^T \mathbf{c} \quad \text{such that} \quad \|\mathbf{A}^T \mathbf{c}\|_{\infty} \leq 1. \quad (19)$$

Therefore, for the optimal  $\mathbf{x}^*$  exists a corresponding optimal  $\mathbf{c}^*$ . According to the KKT conditions for the primal-dual optimal  $(\mathbf{x}^*, \mathbf{c}^*)$  the necessary and sufficient conditions are  $\mathbf{A}\mathbf{x}^* = \mathbf{y}$  and  $\|\mathbf{A}^T \mathbf{c}^*\|_{\infty} \leq 1$ .

As already discussed, the G-PFP algorithm is based on the geometry of the polar polytope associated with dual linear program and searches the optimum vertex  $\mathbf{c}^*$  using a path following approach. In the following section the proposed algorithm is derived.

### 3.3 The proposed algorithm

Let us now derive the proposed algorithm for recovery of block sparse signals. The G-PFP algorithm is an iterative greedy algorithm that builds the solution vector in a similar way to the G-OMP algorithm. The algorithm at the  $k$ -th iteration uses equation (15) to identify the next group of atoms, where we consider only vectors  $\tilde{\mathbf{a}}_i$  for which  $\tilde{\mathbf{a}}_i^T \mathbf{r}^{k-1} > 0$  within each group of atoms due to the nonnegativity constraint of the solution vector and we exclude the groups that have already been selected in previous iterations. Note that the first iteration will be identical to the G-OMP algorithm as  $\mathbf{c}$  is initialized at zero.

Next the algorithm adds the selected group of atoms to the active set and updates the solution vector  $\tilde{\mathbf{x}}^k$ , the residual  $\mathbf{r}^k$  and the corresponding  $\mathbf{c}^k$ . The algorithm iterates till the stopping criteria are met. The resulting algorithm of G-PFP is given in Algorithm 1.

One of the most expensive computations of the algorithm is the calculation of the Moore-Penrose pseudo-inverse  $(\tilde{\mathbf{A}}^k)^\dagger$  required for the update of the solution vector  $\tilde{\mathbf{x}}^k$  and the corresponding  $\mathbf{c}^k$  at each iteration. As has already been in discussed in [11] for the conventional sparsity PFP algorithm, directional updates could be used (e.g. the method of conjugate gradient) instead of the Cholesky factorization method when dealing with large scale systems.

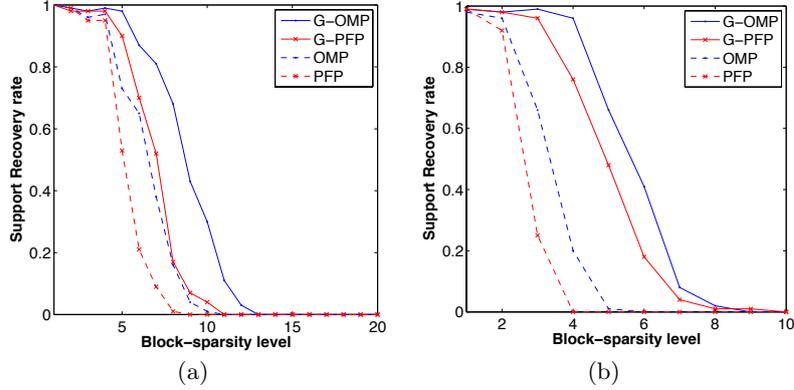
Note also that following LARS, pretty much as we did in [11] we omit the releasing step, which reduces the computational cost but is expected not to lead to a large change to the result.

**Algorithm 1** Group-Polytope Faces Pursuit (G-PFP)

- 
- 1: Input:  $\tilde{\mathbf{A}} = [\tilde{\mathbf{a}}_i], \mathbf{y}$
  - 2: Set stopping conditions  $l_{\max}$  and  $\theta_{\min}$
  - 3: Initialize:  $k \leftarrow 0, \mathcal{I}^k \leftarrow \emptyset, \tilde{\mathbf{A}}^k \leftarrow \emptyset, \mathbf{c}^k \leftarrow \mathbf{0}, \tilde{\mathbf{x}}^k \leftarrow \emptyset, \hat{\mathbf{y}}^k \leftarrow \mathbf{0}, \mathbf{r}^k \leftarrow \mathbf{y}$
  - 4: **while**  $|\mathcal{I}^k| < l_{\max}$  and  $\max_i \tilde{\mathbf{a}}_i^T \mathbf{r}^{k-1} > \theta_{\min}$  **do** {Find next face}
  - 5:    $k \leftarrow k + 1$
  - 6:   Find face:  
 $i^k \leftarrow \arg \max_{i \notin \mathcal{I}^{k-1}} \{ \|\tilde{\mathbf{A}}[i]^T \mathbf{r}^{k-1}\|_2 / \|1 - \tilde{\mathbf{A}}[i]^T \mathbf{c}^{k-1}\|_2 \mid \tilde{\mathbf{A}}[i]^T \mathbf{r}^{k-1} > 0 \}$
  - 7:   Add constraints:  
 $\tilde{\mathbf{A}}^k \leftarrow [\tilde{\mathbf{A}}^{k-1}, \tilde{\mathbf{A}}[i^k]], \mathcal{I}^k \leftarrow \mathcal{I}^{k-1} \cup \{i^k\}$
  - 8:    $\tilde{\mathbf{x}}^k \leftarrow (\tilde{\mathbf{A}}^k)^\dagger \mathbf{y}, \mathbf{c}^k \leftarrow (\tilde{\mathbf{A}}^k)^\dagger \mathbf{1}, \hat{\mathbf{y}}^k \leftarrow \tilde{\mathbf{A}}^k \tilde{\mathbf{x}}^k, \mathbf{r}^k \leftarrow \mathbf{y} - \hat{\mathbf{y}}^k$
  - 9: **end while**
  - 10: Output:  $\mathbf{c}^* = \mathbf{c}^k, \tilde{\mathbf{x}}^* \leftarrow \mathbf{0} +$  corresponding entries from  $\tilde{\mathbf{x}}^k$
- 

## 4 Simulation results

In the first experiment we attempted to quantify the performance of the proposed algorithm and compare against the group sparsity algorithm G-OMP and the standard sparsity algorithms OMP and PFP, using synthetic data. To do so, we randomly generated dictionaries of size  $40 \times 200$  by drawing from i.i.d. Gaussian matrices and normalizing them. The block  $k$ -sparse vector  $\mathbf{x}$  with block size  $d$  was generated by selecting uniformly at random the non-zero groups of atoms.

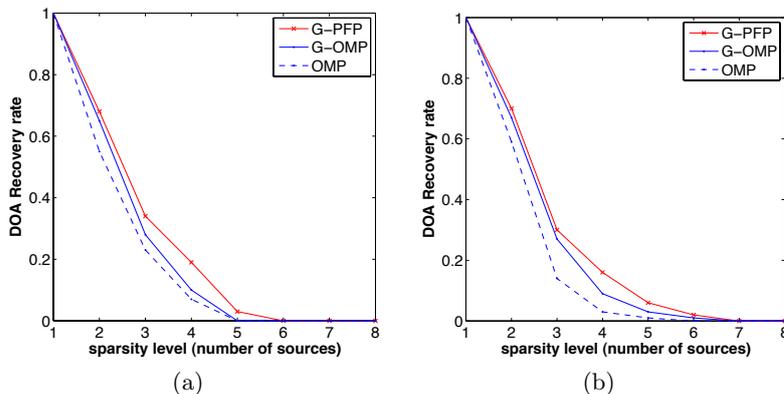


**Fig. 1.** Support recovery rates (over 100 trials) of G-OMP, G-PFP, OMP, PFP vs block-sparsity level  $k$  for a dictionary  $\mathbf{A} \in \mathbb{R}^{M \times N}$  with  $M = 40, N = 200$  and block size (a)  $d = 2$  and (b)  $d = 4$ .

Fig. 1(a)-(b) illustrates the support recovery rate of all tested algorithms for a variable sparsity level  $k$ , where the block size  $d$  has chosen equal to 2 and 4, respectively. The results has been averaged over 100 iterations. As can be

seen, the greedy group sparsity algorithms perform better in both cases and the performance gain increases with the block size. However, G-OMP shows the best success recovery rates apart from the case when  $d = 4$  for high sparsity levels, where G-PFP shows a slightly better performance.

For the second experiment, we chose to apply the algorithms to the problem of direction-of-arrival (DOA) estimation and compare their performance. In this case, we compared G-PFP against G-OMP and OMP. After discretization of the angular space, we formed the redundant dictionary  $\mathbf{A}$  of size  $M \times N$  containing the impulse responses of  $M = 8$  sensors uniformly spaced at half wavelength for all  $N = 181$  potential angles of arrival (resolution grid of  $1^\circ$ ). Assuming that the  $k \ll N$  plane waves impinge on the array from different angles (which has been chosen randomly) and taking  $d$  time-snapshots we formulated the resulting MMV problem as a block sparsity problem by appropriately interleaving the multiple vectors. Therefore, the  $d$  snapshots define the number of the size of each block.



**Fig. 2.** DOA recovery rates (over 100 trials) of G-OMP, G-PFP OMP vs block-sparsity level  $k$  (or number of sources). The numbers of sensors is  $M = 8$ , the angular grid resolution is set at  $1^\circ$  and the number of snapshots (or block size) is (a)  $d = 3$  and (b)  $d = 4$ .

Fig. 2(a)-(b) shows the recovery success rate of the true angles of arrivals averaged over 100 iterations when the number of snapshots and subsequently the block size is 3 and 4, respectively. For the specific setting in both cases G-PFP outperforms the other two algorithms achieving the highest recovery success rates. Considering the fact that the dictionary due to the small number of sensors chosen is quite block-coherent, the results suggest that the G-PFP algorithm can achieve better performance in distinguishing between correlated group of atoms.

## 5 Conclusions

We have introduced an algorithm for the block sparse recovery problem based on the PFP algorithm. The so-called G-PFP algorithm, which is a greedy algorithm of similar complexity to the G-OMP algorithm, adds one group of atoms at a time and iteratively builds the solution. Experiments on the support recovery of exact sparse block synthetic signals show that the proposed algorithm outperforms the standard PFP algorithm, but performs a little worse than G-OMP. However, on the DOA estimation problem the proposed algorithm showed better performance than G-OMP at all sparsity levels investigated.

Our future work will investigate and attempt to explain this behaviour of G-PFP in the coherent dictionary setting.

## References

1. M.D. Plumbley, T. Blumensath, L. Daudet, R. Gribonval, and M.E. Davies, "Sparse representations in audio and music: From coding to source separation," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 995-1005, 2010.
2. S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33-61, 1998.
3. Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA*, 1-3 Nov. 1993, pp. 40-44.
4. M. Mishali and Y. C. Eldar, "Blind multi-band signal reconstruction: Compressed sensing for analog signals," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 993-1009, Mar. 2009.
5. R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Transactions on Signal Processing*, vol. 51, no. 1, pp. 101-110, Jan. 2003.
6. Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042-3054, Jun. 2010.
7. S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
8. M.D. Plumbley, "On polar polytopes and the recovery of sparse representations," *IEEE Transactions on Information Theory*, vol. 53, no. 9, pp. 3188-3195, 2007.
9. M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 68, No. 1. (February 2006), pp. 49-67.
10. E. v. den Berg and M. P. Friedlander, "Joint-sparse recovery from multiple measurements," Technical Report, 2009.
11. A. Gretsistas, I. Damnjanovic and M.D. Plumbley, "Gradient Polytope Faces Pursuit for large scale sparse recovery problems," *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 2030-2033, 2010.