SPARSE APPROXIMATION WITH AN ORTHOGONAL COMPLEMENTARY MATCHING PURSUIT ALGORITHM

Gagan Rath and Christine Guillemot

IRISA-INRIA, France

ABSTRACT

This paper presents the orthogonal extension of the recently introduced complementary matching pursuit (CMP) algorithm for sparse approximation [1]. The CMP algorithm is analogous to the matching pursuit (MP) but done in the row-space of the dictionary matrix. It suffers from a similar sub-optimality as the MP. The orthogonal complementary matching pursuit algorithm (OCMP) presented here tries to remove this sub-optimality by updating the coefficients of all selected atoms at each iteration. Its development from the CMP follows the same procedure as of the orthogonal matching pursuit (OMP). In contrast with OMP, the residual errors resulting from the OCMP may not be orthogonal to all the atoms selected up to the respective iteration. Though the residual energy may increase over the OMP during the first iterations, it is shown that, compared with OMP, the convergence speed is increased in the subsequent iterations and the sparsity of the solution vector is improved.

1. INTRODUCTION

In many areas of signal processing, such as compression, denoising, time-frequency analysis, image analysis, compressed sensing, etc., one comes across a system of linear equations which is underdetermined. Such a system has infinite number of solutions, but, in the context of these applications, the interest primarily lies in finding the sparsest solution. That is, one wants to find the solution with the minimum number of non-zero elements that satisfies the system of equations exactly, or with certain approximation error less than a specified threshold. This problem is known as the sparse approximation in a redundant basis.

Consider the following system of equations:

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

where A is a matrix of dimension $K \times N$, $K < N$, and $\mathbf{b}$ is a known vector of dimension K. The columns of A are assumed to make a redundant basis for the K-dimensional space. The exact sparse approximation problem can thus be posed as

$$\min\{\|\mathbf{x}\|_0 \ : \ A\mathbf{x} = \mathbf{b}\}, \tag{2}$$

where the L_0 norm denotes the number of non-zero elements. The problem which allows some approximation error can be posed as

$$\min\{\|\mathbf{x}\|_0 \ : \ \|A\mathbf{x} - \mathbf{b}\|_p \leq \delta\}, \tag{3}$$

for some $\delta \geq 0$. The norm p is usually 2, but could be 1 or ∞ as well. The columns of A are called atoms and the collection of atoms is called a dictionary.

The above problems are NP-hard [2]. There are two heuristic approaches in the literature which find approximate solutions with tractable complexity. One is called the matching pursuit (MP), which is a sort of greedy algorithm [3, 4]. The other is known as the basis pursuit (BP), which relaxes the L_0 norm condition by L_1 norm and solves the problem through linear programming [5, 6]. MP algorithms have the advantages of ease of implementation and speed compared to the BP algorithms though the latter algorithms are claimed to produce more accurate solutions. In fact, it has been shown that the results for the OMP algorithm are comparable to those of BP [7] in certain applications.

In this paper, we present an orthogonal complementary matching pursuit (OCMP) algorithm for sparse approximation, which is an extension of the recently proposed CMP algorithm [1]. The algorithm is similar to OMP, but performed in the row space of the dictionary matrix. The development of the orthogonal complementary matching pursuit from the CMP follows the same line of development as of OMP from MP. Like OMP, the algorithm chooses one atom at each iteration that best approximates the residual, however, unlike OMP, the residual vector may not be orthogonal to the subspace spanned by the selected atoms. This has a beneficial impact on the convergence speed and the sparsity of the solution relative to OMP. We also show that, under the orthogonality of the rows of the dictionary matrix, the OCMP algorithm is equivalent to the OMP algorithm.

2. ORTHOGONAL MATCHING PURSUIT (OMP)

The MP algorithm [3] first finds the atom that has the highest correlation with the signal vector. It subtracts off the correlated part from the signal and then iterates the procedure on the resulting residual signal. Because of the sub-optimality [4], the convergence speed can be extremely slow. The OMP algorithm removes this drawback by updating the coefficients of the selected atoms at each iteration in order that the resulting residual vector is orthogonal to the subspace spanned by the selected atoms.

Let α_i, $1 \leq i \leq N$, denote the ith atom with $\|\alpha_i\|_2 = 1, \ \forall i$. At the jth iteration the algorithm computes

$$\alpha_j^{opt} = \underset{\alpha_i \in \mathcal{A}/\mathcal{A}_{j-1}^{opt}}{\arg\max} \ | < \mathbf{r}_{j-1}, \alpha_i > |, \tag{4}$$

where $\mathbf{r}_{j-1}$ denotes the residual at the $(j-1)$th iteration, $\mathcal{A}$ denotes the dictionary, $\mathcal{A}_{j-1}^{opt}$ denotes the set of atoms selected up to the $(j-1)$th iteration, and '/' denotes the set difference operator. Let A_j denote the matrix containing all the selected atoms in the previous iterations including α_j^{opt}. The approximation at the jth iteration is given as the projection of the original signal vector onto the subspace spanned by the columns of A_j, i.e., $\mathbf{b}_j = A_j(A_j^T A_j)^{-1} A_j^T \mathbf{b}$. Thus the coefficient vector at the jth iteration is given as $\mathbf{c}_j = (A_j^T A_j)^{-1} A_j^T \mathbf{b}$. The algorithm then updates the residual as

$$\mathbf{r}_j = \mathbf{b} - A_j(A_j^T A_j)^{-1} A_j^T \mathbf{b}. \tag{5}$$

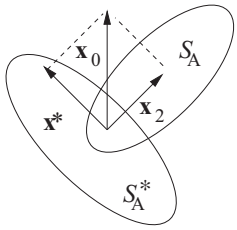

Fig. 1. Subspace representations of $\mathbf{x}_2$, $\mathbf{x}_0$, and $\mathbf{x}^*$

Because of the orthogonal projection, the new residual is orthogonal to all the atoms selected previously. Therefore, unlike the MP, once an atom is selected, it will not be selected in any of the subsequent iterations. This feature improves the convergence speed. The nonzero components of the sparse solution vector are equal to the components of the coefficient vector at the last iteration.

3. COMPLEMENTARY MATCHING PURSUIT

Consider the matrix A, whose columns are atoms. Since the columns of A are assumed to make a redundant basis, its rows are linearly independent. Thus the columns of A^T span a K dimensional subspace in the N-dimensional space. Let us denote this subspace by $\mathcal{S}_A$. The orthogonal complement of this subspace has dimension $N-K$. Let us denote this subspace by $\mathcal{S}_A^*$. Let G denote the $N \times (N-K)$ matrix whose columns are orthonormal and which span $\mathcal{S}_A^*$. G can be derived through well-known matrix operations such as QR factorization or Singular Value Decomposition (SVD) of A^T. Using the QR factorization, we can expand A^T as

$$A^T = [Q_1^{N\times K}\ G^{N\times(N-K)}] \begin{bmatrix} T_1^{K\times K} \\ \mathbf{0}^{(N-K)\times K} \end{bmatrix};$$

the columns of Q_1 are orthonormal and they make a spanning set for $\mathcal{S}_A$, and T_1 is a upper triangular matrix. The columns of G are orthonormal and they are orthogonal to the columns of Q_1. Clearly, $AG = \mathbf{0}_{K\times(N-K)}$.

Now consider the system of equations in Eqn.1. The minimum L_2 norm solution to this system of equation is the pseudo-inverse:

$$\mathbf{x}_2 = A^T(AA^T)^{-1}\mathbf{b}.$$

Clearly, $\mathbf{x}_2$ lies in $\mathcal{S}_A$. The sparsest solution, denoted as $\mathbf{x}_0$, can be expressed as $\mathbf{x}_0 = \mathbf{x}_2 + \mathbf{x}^*$, where $\mathbf{x}^*$ is some nonzero vector. Since $A\mathbf{x}_0 = A\mathbf{x}_2 = \mathbf{b}$, $A\mathbf{x}^* = \mathbf{0}_K$. Therefore $\mathbf{x}^*$ lies in $\mathcal{S}_A^*$. This is shown in Fig. 1. Now finding $\mathbf{x}_0$ is equivalent to finding $\mathbf{x}^*$.

Since $\mathbf{x}_0$ is sparse, $\|\mathbf{x}_0\|_0 \leq K$ implies that at least $N-K$ components of $\mathbf{x}^*$ have the same magnitudes but opposite signs as the corresponding components of $\mathbf{x}_2$. It has been shown in [1] that if these components of $\mathbf{x}^*$ are known, then the remaining components of $\mathbf{x}^*$ can be determined provided the rows of G having the same indices as these components make a spanning set for the $(N-K)$-dimensional space.

The exact solution can thus be obtained by exhaustively searching over all possible combinations of t, $N > t \geq N-K$, components from the L_2 norm solution $\mathbf{x}_2$, for which the components of $\mathbf{x}^*$ and $\mathbf{x}_2$ are equal in magnitude but opposite in sign. The remaining components of $\mathbf{x}^*$ can then be computed (see [1]) and subtracted from $\mathbf{x}_2$ to derive the sparse vector. An alternative approach is to follow a greedy pursuit algorithm, which we call complementary matching pursuit [1]. At first, the algorithm assumes that there are $(N-1)$ elements for which $\mathbf{x}^*$ and $\mathbf{x}_2$ have same magnitudes but opposite signs; in other words, the solution vector has only one nonzero element. Accordingly, for a given atom index, the algorithm computes $\mathbf{x}^*$ (by projecting $\mathbf{x}_2$ with 0 at that index onto the subspace $\mathcal{S}_A^*$) and subtracts it from $\mathbf{x}_2$ to compute the residual. The optimal atom is the one that gives the minimum residual error. This process is then iterated with the residual error until the desired error threshold is achieved. For details, the reader is referred to [1].

3.1. Orthogonal complementary matching pursuit (OCMP)

This algorithm is similar to the OMP algorithm. At every iteration, it identifies the atom that produces the smallest residual vector and then updates the coefficients of all the selected atoms. However, the residual vector is defined differently in this case. Instead of a K-dimensional vector as in the case of the OMP, here it is an N-dimensional vector.

Let $\mathbf{r}_j$ denote the residual at the jth iteration with initialization $\mathbf{r}_0 = \mathbf{x}_2$. Let R_i denote the set of indices except the ith index, i.e., $R_i \equiv \mathcal{I}/\{i\}$. Let $\mathbf{r}_j[R_i]$ denote the vector containing all the elements of $\mathbf{r}_j$ except the ith element. Similarly, let G_{R_i} denote the matrix containing all the rows of matrix G except the ith row. At the jth iteration the algorithm computes

$$\mathbf{e}_j^i = \mathbf{r}_{j-1} - G(G_{R_i}^T G_{R_i})^{-1} G_{R_i}^T \mathbf{r}_{j-1}[R_i], \quad \alpha_i \in \mathcal{A}/\mathcal{A}_{j-1}^{opt} \quad (6)$$

The second term on the right computes the orthogonal vector in $\mathcal{S}_A^*$ for the atom α_i. Since the columns of G are orthonormal with row norms less than 1, any $N-1$ rows of G make a redundant basis for the $(N-K)$-dimensional space. Hence, the above matrix inversion operation is valid. The optimal atom is thus found by minimizing the pth norm:

$$\alpha_j^{opt} = \underset{\alpha_i \in \mathcal{A}/\mathcal{A}_{j-1}^{opt}}{\arg\min} \|\mathbf{e}_j^i[R_i]\|_p. \quad (7)$$

where $\mathbf{e}_j^i[R_i]$ denotes the vector $\mathbf{e}_j^i$ without the ith component, and p can be 1, 2, or ∞. This step is the same as in CMP [1] except that the computations are performed only for the atoms which have not been selected yet. The second step updates the coefficients of all the selected atoms after each iteration, as in OMP. Let $\mathcal{I}_j$ denote the set of indices of atoms selected up to the jth iteration and let R_j denote its complementary set, i.e., $R_j \equiv \mathcal{I}/\mathcal{I}_j$. Let again G_{R_j} denote the matrix containing the rows of G having indices in R_j. The algorithm computes the orthogonal vector lying in $\mathcal{S}_A^*$ as $G(G_{R_j}^T G_{R_j})^{-1} G_{R_j}^T \mathbf{x}_2[R_j]$ and subtracts it from $\mathbf{x}_2$ to approximate the sparse vector $\mathbf{x}_0$:

$$\mathbf{u}_j = \mathbf{x}_2 - G(G_{R_j}^T G_{R_j})^{-1} G_{R_j}^T \mathbf{x}_2[R_j]. \quad (8)$$

It can be proved that the matrix inversion operation in Eqn. 8 is valid for all j till the last iteration (see [9] for the proof).

The updated coefficients are equal to the components of $\mathbf{u}_j$ with indices in $\mathcal{I}_j$. The new residual vector is obtained by setting those components of $\mathbf{u}_j$ to zero. That is, $\mathbf{c}_j = \mathbf{u}_j[\mathcal{I}_j]$ and $\mathbf{r}_j = \mathbf{u}_j \odot \mathbf{1}_{\mathcal{I}_j=0}$, where $\odot$ denotes element-wise multiplication, and $\mathbf{1}_{\mathcal{I}_j=0}$ denotes an N-dimensional vector with all elements equal to 1 except the elements at indices $\mathcal{I}_j$, which are equal to 0. The nonzero components of the sparse solution vector at the jth iteration are given by the vector $\mathbf{c}_j$. Thus, the approximation at the jth iteration is given as $\mathbf{b}_j = \sum_{k=1}^{j} c_k \alpha_k^{opt}$. And the resulting approximation error is

$$\epsilon_j = \mathbf{b} - \mathbf{b}_j = \mathbf{b} - \sum_{k=1}^{j} c_k \alpha_k^{opt}. \quad (9)$$

The algorithm terminates when the approximation error falls below the specified threshold, or when the number of atoms equals the specified limit. Notice that, unlike in OMP, here the residual error $\mathbf{r}_j$ is not the same as the approximation error ϵ_j. Further, the residual vector at any iteration is orthogonal to the columns of G. This can be proved as follows. Since the residual error contains zeros at the indices $\mathcal{I}_j$, $G^T\mathbf{r}_j = G_{R_j}^T\mathbf{u}_j[R_j]$. Using Eqn. 8, we can expand the right hand side as:

$$G_{R_j}^T(\mathbf{x}_2[R_j] - G_{R_j}(G_{R_j}^T G_{R_j})^{-1}G_{R_j}^T\mathbf{x}_2[R_j]) = \mathbf{0}.$$

This shows that the residual vectors lie in $\mathcal{S}_A$, as does $\mathbf{x}_2$. As a consequence of this orthogonality, in the $(j+1)$th iteration, the second term in Eqn. 6 will be equal to zero for any of the atoms selected up to the jth iteration, and thus the residual will remain unchanged for those atoms. Therefore, once an atom is selected, it will not be selected in the subsequent steps. The minimization in Eqn. 7 is thus performed only on the set of atoms not yet selected.

4. OCMP VS OMP

We derive the expressions for the coefficient of the optimal atom and the residual energy after the jth iteration. The non-zero components of the residual error are given as

$$\mathbf{r}_j[R_j] \quad = \quad \mathbf{x}_2[R_j] - G_{R_j}(G_{R_j}^T G_{R_j})^{-1}G_{R_j}^T\mathbf{x}_2[R_j] \quad (10)$$

Therefore, the residual error energy is

$$\|\mathbf{r}_j[R_j]\|^2 = \mathbf{x}_2^T[R_j](I_{N-K} - G_{R_j}(G_{R_j}^T G_{R_j})^{-1}G_{R_j}^T)\mathbf{x}_2[R_j].$$

Let $G_{\mathcal{I}_j}$ denote the matrix containing the rows of G with indices $\mathcal{I}_j$. Since $G^TG = I_{N-K}$, $G_{R_i}^T G_{R_i} = I_{N-K} - G_{\mathcal{I}_j}^T G_{\mathcal{I}_j}$. Using the matrix inversion lemma, and the orthogonality between the columns of G and $\mathbf{x}_2$, the above expression can be simplified as

$$\|\mathbf{r}_j[R_j]\|^2 = \|\mathbf{x}_2\|^2 - \mathbf{x}_2^T[\mathcal{I}_j](I_j - G_{\mathcal{I}_j}G_{\mathcal{I}_j}^T)^{-1}\mathbf{x}_2[\mathcal{I}_j], \quad (11)$$

where I_j denotes the identity matrix of order j. Now, using the orthogonality between G from A^T, it can be proved that $I_j - G_{\mathcal{I}_j}G_{\mathcal{I}_j}^T = A_j^T(AA^T)^{-1}A_j$ where A_j denotes the matrix containing only the atoms selected up to the jth iteration. Therefore

$$\|\mathbf{r}_j[R_j]\|^2 = \|\mathbf{x}_2\|^2 - \mathbf{x}_2^T[\mathcal{I}_j](A_j^T(AA^T)^{-1}A_j)^{-1}\mathbf{x}_2[\mathcal{I}_j]. \quad (12)$$

The coefficients associated with the selected atoms are given as

$$\mathbf{c}_j = \mathbf{x}_2[\mathcal{I}_j] - G_{\mathcal{I}_j}(G_{R_j}^T G_{R_j})^{-1}G_{R_j}^T\mathbf{x}_2[R_j]. \quad (13)$$

Since $\mathbf{x}_2$ is orthogonal to the columns of G, $G_{R_j}^T\mathbf{x}_2[R_j] + G_{\mathcal{I}_j}^T\mathbf{x}_2[\mathcal{I}_j] = \mathbf{0}$. Therefore,

$$\mathbf{c}_j = (I_j + G_{\mathcal{I}_j}(G_{R_j}^T G_{R_j})^{-1}G_{\mathcal{I}_j}^T)\mathbf{x}_2[\mathcal{I}_j]. \quad (14)$$

Using the matrix inversion lemma and simplifying, we get

$$\begin{aligned}\mathbf{c}_j &= (I_j - G_{\mathcal{I}_j}G_{\mathcal{I}_j}^T)^{-1}\mathbf{x}_2[\mathcal{I}_j] \\ &= (A_j^T(AA^T)^{-1}A_j)^{-1}A_j^T(AA^T)^{-1}\mathbf{b}. \quad (15)\end{aligned}$$

The approximation error is then

$$\|\epsilon_j\|^2 = \|\mathbf{b} - A_j\mathbf{c}_j\|^2 = \mathbf{b}^T\mathbf{b} - 2\mathbf{c}_j^T A_j^T\mathbf{b} + \mathbf{c}_j^T A_j^T A_j\mathbf{c}_j.$$

To compare these quantities with those in OMP, let us assume that OMP selects the same atoms up to the jth iteration. The residual error is

$$\|\mathbf{r}_{OMP}\|^2 = \mathbf{b}^T(I_K - A_j(A_j^T A_j)^{-1}A_j^T)\mathbf{b}, \quad (16)$$

Thus the approximation error with OCMP can be simplified as

$$\|\epsilon_j\|^2 = \|\mathbf{r}_{OMP}\|^2 + \|A_j\mathbf{c}_j - A_j(A_j^T A_j)^{-1}A_j^T\mathbf{b}\|^2.$$

Let $\mathbf{c}_{OMP}$ denote the coefficient vector which is given as $\mathbf{c}_{OMP} = (A_j^T A_j)^{-1}A_j^T\mathbf{b}$. Thus we can write $\mathbf{b} = A_j(A_j^T A_j)^{-1}A_j^T\mathbf{b} + \mathbf{r}_{OMP}$. Substituting this in Eqn. 15, we get

$$\mathbf{c}_j = \mathbf{c}_{OMP} + (A_j^T(AA^T)^{-1}A_j)^{-1}A_j^T(AA^T)^{-1}\mathbf{r}_{OMP}.$$

Now, if the rows of A are orthogonal with identical norms, then the second term in the above expression is equal to a null vector. This means that both OCMP and OMP result in the same coefficients. If they select the same atoms, their approximation errors are also identical. Therefore, when the rows of A are orthogonal, OCMP is equivalent to OMP.

Let us now consider the general case where the rows of A are not orthogonal. The above expressions show that, if $\mathbf{r}_{OMP}$ is zero, then $\mathbf{c}_j$ is equal to $\mathbf{c}_{OMP}$. This means that if OMP finds the exact solution, then the OCMP also gives the exact solution. This also shows that, in general, the approximation error with OCMP is not orthogonal to the subspace spanned by the selected atom. As a result, the approximation error is more than that of the OMP. However, this result of higher residual error does not extend to all subsequent iterations. On the contrary, due to the fact that the residual error is not orthogonal to the selected atoms, the convergence can be faster provided the error aligns more closely with any of the remaining atoms. The offset term thus may improve the accuracy of the selected atoms and the convergence speed by getting closer to the true coefficient magnitudes. Finally, similarly to the CMP [1], it can be shown that OCMP is equivalent to OMP on the system of equations $A^T(AA^T)^{-1}A\mathbf{x} = \mathbf{x}_2$. So the OCMP is basically an OMP applied in the row-space of the dictionary matrix.

5. SIMULATION RESULTS

In order to compare the proposed algorithm with OMP, we performed experiments with a random dictionary. The dictionary was created using the recently proposed K-SVD algorithm [8]. The dictionary consisted of 55 atoms with each atom having 32 elements. The atoms were normalized with respect to the L_2 norm. Then 1000 signal vectors of dimension 32 were generated, each created by a linear combination of certain number of atoms with randomly generated coefficients, but no additive noise. The atoms themselves were selected from random combinations.

We first compared the convergence of OCMP with those of the MP, CMP, and OMP algorithms. We generated signals with 8 random atoms and used them as input signals to the four algorithms. Fig. 2 shows the mean square residual errors at different number of iterations. The mean was computed over all the 1000 signal blocks and then normalized with respect to the mean signal energy. After a few iterations, the error dies out faster with OCMP than OMP. Higher convergence speed of OCMP with respect to CMP is expected.

We then varied the number of generating atoms for the input signal. We specified an error bound of 0.001 for each component of the signal vector and a maximum of 16 atoms in the approximation. Fig. 3 shows the number of atoms identified by different algorithms. We observe that, among the four algorithms, OCMP produces the best sparsity results. Note that the plots (with MP, CMP and OMP) tend to saturate because of the specified maximum of 16 atoms in the approximation. Fig. 4 shows the fraction of true atoms identified in the approximation. It is interesting to see that CMP identifies more correct atoms than not only MP, but also OMP. However, OCMP

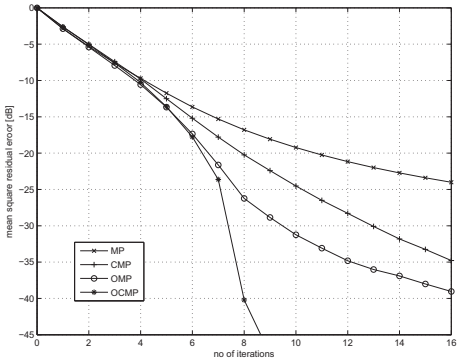

Fig. 2. Residual energy for different number of iterations. Number of generating atoms is 8.

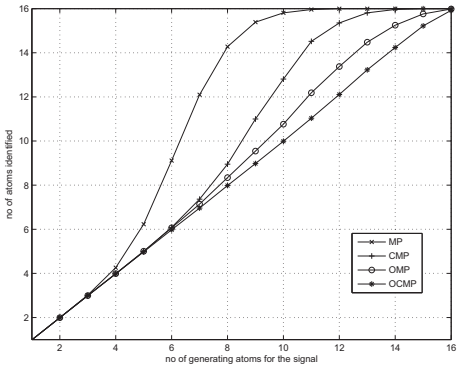

Fig. 3. No of atoms identified versus the number of actual atoms. Error bound per component 0.001, maximum 16 atoms.

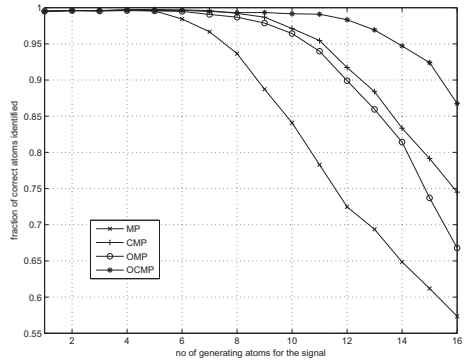

Fig. 4. Fraction of correct atoms detected versus the number of atoms. Error bound per component 0.001, maximum 16 atoms.

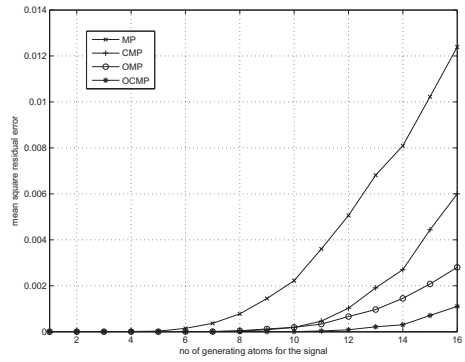

Fig. 5. Mean square residual error versus the number of atoms. Error bound per component 0.001, maximum 16 atoms.

outperforms all of them. Again the relative improvements are more pronounced for a higher number of generating atoms. Finally Fig. 5 shows the mean square error resulting from different algorithms. We see that OCMP results in the least error among the four algorithms. The relative gain over OMP becomes interesting as the number of generating atoms is increased.

6. CONCLUSIONS

In this paper, we have presented the orthogonal extension of the complementary matching pursuit which we introduced in [1]. We have derived the expressions for the coefficients of the solution vector and the residual errors, and shown that the algorithm is equivalent to an OMP algorithm performed in the row-space of the dictionary matrix. Simulations with a K-SVD optimized dictionary showed that the OCMP converges faster and produces a sparser solution than the OMP. Furthermore, in terms of identifying true atoms, it also outperforms the OMP algorithm.

7. REFERENCES

[1] G. Rath and C. Guillemot, "A complementary matching pursuit algorithm for sparse approximation," *European Signal Processing Conference (EUSIPCO) 2008*, Aug. 2008.

[2] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.,* 24, 2, pp. 227–234, April 1995.

[3] S. Mallat and Z. Zhang, "Matching pursuit in a time-frequency dictionary," *IEEE Trans. SP,* Vol. 41, No. 12, pp. 3397–3415, 1993.

[4] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad,"Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decompositions," *Proc. 27th Asilomar Conf. on Sig., Sys. and Comp.,* 1993.

[5] S. S. Chen, D. L. Donoho, and M. A. Saunders,"Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.,* 20, 1, pp. 33–61, 1988.

[6] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution," *Comm. Pure Appl. Math.,* 59, 6, pp. 797–829, 2006.

[7] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inform. Theory,* 53, 12, pp. 4655–4666, Dec. 2007.

[8] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. SP,* 54, 11, pp. 4311–4322, 2006.

[9] G. Rath and C. Guillemot, "Complementary Matching Pursuit Algorithms for Sparse Approximation," *submitted to IEEE Trans. SP,*, 2008.