Expectation Maximization Based Matching Pursuit

Ali Cafer Gurbuz, *Member, IEEE,* Mert Pilanci, *Student Member, IEEE,* and Orhan Arikan, *Member, IEEE*

Abstract—A novel expectation maximization based matching pursuit (EMMP) algorithm is presented. The method uses the measurements as the incomplete data and obtain the complete data which corresponds to the sparse solution using an iterative EM based framework. In standard greedy methods such as matching pursuit or orthogonal matching pursuit a selected atom can not be changed during the course of the algorithm even if the signal doesn't have a support on that atom. The proposed EMMP algorithm is also flexible in that sense. The results show that the proposed method has lower reconstruction errors compared to other greedy algorithms using the same conditions.

Index Terms—sparse reconstruction, compressive sensing, greedy methods, expectation maximization

I. INTRODUCTION

SPARSE signal representations and the theory of compressive sensing (CS) [1], [2] has received considerable attention in many research communities and has a wide range of applications such as computational photography [3], medical imaging [4], radar [5], [6], sensor networks [7] and many more. The main application revolves around the classical underdetermined linear regression problem:

$$y = \Phi x + n \qquad (1)$$

where $y \in \Re^M$ and $n \in \Re^M$ are the measurement and noise vectors of dimension M. $\Phi \in \Re^{M \times N}$ is the compressive measurement matrix where $M < N$. x is the signal to be reconstructed which can be represented with a basis Ψ as $x = \Psi s$. Here it is assumed that the signal is K sparse meaning $||s||_{\ell_0} = K$ and the goal is to obtain the most sparse s that satisfies (1) among many solutions of (1).

CS shows that the sparse signal s, hence , x can be recovered with very high probability from $O(K \log(N/K))$ measurements by solving a convex ℓ_1 minimization problem of the following form.

$$min||s||_1 \quad s.t. \quad ||y - \Phi\Psi s||_2 < \epsilon \qquad (2)$$

The problem in (2) is convex and the global optimal solution is guaranteed. However, the computational complexity of (2) is high and instead suboptimal greedy algorithms are also used in many applications. Matching pursuit (MP) [8], orthogonal matching pursuit (OMP) [9], compressive sampling matching

pursuit (CoSamp) [10] iterative hard/soft thresholding (IHT) [11] belong to these greedy algorithms. Most of these greedy techniques work with selecting the mostly correlated columns of the dictionary and when a wrong or unreliable dictionary element (atom) is selected it can not be removed from the list of selected atoms and these selections affect the selection of following atoms resulting in a decrease of performance. There are new variants of OMP that tries to find and remove unreliable atoms [12] by backtacking and assigning reliability to the previously chosen atoms.

This paper focuses on the same sparse signal recovery problem using an expectation maximization [13] (EM) framework. The observed measurements y are actually the incomplete data that is observed about the system. The complete observations would be the measurements $y_i \quad i = 1, 2, ..., K$ corresponding to each sparsity element or atom. Note that the vector y_i is not the ith index of y, it represents the measurements if only the ith support of the signal was present. The resultant algorithm iteratively updates the measurements and the selected atom indexes allowing removal of unreliable atoms. Variants of the proposed EM-MP method that both requires and does not require an estimate sparsity level to be known priori are developed in the paper. The numerical simulation results show that the proposed EMMP method has lower average reconstruction errors in varying number of measurements or signal to noise ratio (SNR) levels compared to standard greedy techniques such as OMP or ℓ_1 minimization based basis pursuit.

The paper is organized as follows. The EMMP algorithm is explained in Section II. Simulation and test results are shown in Section III. Conclusions are discussed in Section IV.

II. EXPECTATION MAXIMIZATION MATCHING PURSUIT

The sparse reconstruction problem for the linear system of $y = \Phi\alpha$ can be written in the form of

$$y = [\phi_1\phi_2...\phi_N][\alpha_1\alpha_2...\alpha_N]^T \qquad (3)$$

and $y = \sum_{i=1}^{N} \alpha_i \phi_i$. A K-sparse solution means that only K of the α_i's are nonzero. The matching pursuit algorithm (MP) projects y to each column of Φ and select the atom that has the largest correlation and removes that projection from the measurements and iteratively continues the procedure until the termination criteria is met. OMP adds a least squares step where the measurements y is projected to the span of the selected atoms at each step and residual is calculated using least squares solution which prevents recurrent selection of atoms.

Here we would like to make the observation that the measurements y can be written as the summation of K measurement vectors as $y = \sum_{j=1}^{K} y_j$ where each measurement

vector is $\boldsymbol{y}_j = \alpha_j \boldsymbol{\phi}_j$. Here $\boldsymbol{y}_j$'s are not known or measured but if we had known or could estimate them, then it would be easy to find a one sparse solution in the linear system of $\boldsymbol{y}_j = \boldsymbol{\Phi} \boldsymbol{x}$. Since $\boldsymbol{y}_j$'s are not directly measured but their sum $\boldsymbol{y}$ is available about the system, $\boldsymbol{y}$ can be interpreted as the incomplete data about the system and $\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_K$ can be seen as the complete data. This interpretation guides us to the well known expectation maximization (EM) framework to obtain iteratively both the complete data $\boldsymbol{y}_j \quad j = 1, 2, ..K$ and their corresponding supports. The proposed EM based matching pursuit algorithm (EMMP) is summarized in Table I.

TABLE I
EMMP ALGORITHM

Input:
$\boldsymbol{\Phi}$ measurement matrix
$\boldsymbol{y}$ measurements
S sparsity level
ϵ termination criteria threshold

Initialization::
$CD = 0$ complete data matrix of size $M \times S$
$r = y$ residual vector

While loop, repeat until $||\boldsymbol{r}||_2 < \epsilon$
  for $i = 1 : S$
   Expectation:
   $\hat{\boldsymbol{y}}_i = \boldsymbol{y} - \sum_{j \neq i}(CD(:, j))$

   Maximization:
   $proj = \boldsymbol{\Phi}^T \hat{\boldsymbol{y}}_i$
   $\lambda = arg \max |proj|$
   $\boldsymbol{p} = \boldsymbol{\Phi}(:, \lambda)$
   $\boldsymbol{y}_i = \boldsymbol{p}\boldsymbol{p}^T \hat{\boldsymbol{y}}_i$

   Keep and Update:
   $\lambda_L(i) = \lambda \quad \lambda$ list
   $CD(:, i) = \boldsymbol{y}_i$
  end for loop
Calculate residual $\boldsymbol{r} = \boldsymbol{y} - \sum_{j=1}^{S} CD(:, i)$
end while loop

$\hat{\boldsymbol{\Phi}} = \boldsymbol{\Phi}(:, \lambda_L)$
$\boldsymbol{x} = zeros(N, 1)$
$\boldsymbol{x}(\lambda_L) = \min ||\boldsymbol{y} - \sum_{j=1}^{S} \boldsymbol{x}(\lambda_L(j))\hat{\boldsymbol{\Phi}}(:, j)||_2$
Output: $\boldsymbol{x}$ solution vector

The proposed EMMP algorithm for solving a linear system of $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{n}$ takes the measurement matrix $\boldsymbol{\Phi}$, the measurements $\boldsymbol{y}$, a given sparsity level S and a termination parameter ϵ as the inputs of the algorithm. Here S denotes the targeted sparsity level of the output solution. The effect of S on algorithm performance and a new algorithm that doesn't require S a priori is explained in Section III. Here Table I focuses on the main core of the EMMP algorithm.

EMMP algorithm starts with initialization of a complete data matrix CD of size $M \times S$ where M is the number of measurements. This matrix is initialized to zeros. Also

a residual vector $\boldsymbol{r}$ is initialized to $\boldsymbol{r} = \boldsymbol{y}$. The algorithm continues until the norm of the residual vector $\boldsymbol{r}$ is less then the input parameter ϵ. Until this termination criteria is met expectation and maximization steps are calculated for each complete data index. First the ith complete data component is estimated in expectation step as $\hat{\boldsymbol{y}}_i = \boldsymbol{y} - \sum_{j \neq i}(CD(:, j))$. Here $\hat{\boldsymbol{y}}_i$ is the estimate of the ith complete data component at the current iteration. In the maximization step $\hat{\boldsymbol{y}}_i$ is projected on to the columns of $\boldsymbol{\Phi}$ and the column $\boldsymbol{p}$ that gives the highest correlation with $\hat{\boldsymbol{y}}_i$ is selected. The ith complete data vector $\boldsymbol{y}_i$ is obtained by projecting $\hat{\boldsymbol{y}}_i$ to $\boldsymbol{p}$ vector as $\boldsymbol{y}_i = \boldsymbol{p}\boldsymbol{p}^T \hat{\boldsymbol{y}}_i$. The selected index list and the complete data matrix is updated by the current index of the selected $\boldsymbol{p}$ vector and the $\boldsymbol{y}_i$. This procedure is repeated S times for each sparsity level until the termination criteria of the while loop is met. There are various ways of terminating these types of greedy methods, here only one possibility where ℓ_2 norm of the residual is used. After each for loop the residual vector is updated by $\boldsymbol{r} = \boldsymbol{y} - \sum_{j=1}^{S} CD(:, i)$ and depending on the norm of this residual vector the while loop is terminated. The S sparse solution $\boldsymbol{x}$ can be reported as the least squares solution to the selected columns of the $\boldsymbol{\Phi}$ matrix. Next section details the performance of the outlined algorithm under various conditions compared to other sparse recovery methods.

III. SIMULATION RESULTS

In this section the average performance of the proposed EMMP algorithm is analyzed and compared to CS and OMP results. First the algorithms are tested for varying number of measurements. A sparse signal of dimension $N = 512$ and $K = 10$ and $K = 20$ sparsity levels are measured with a measurement matrix $\boldsymbol{\Phi}$ that is constructed randomly from the normal distribution. The number of measurements M is varied from 5 to 500. A white gaussian noise corresponding to signal to noise ratio (SNR) of 5dB is added to the measurements. The system is solved with CS (Eq. (2)), OMP and EMMP methods and the error norm between the constructed signals and the true signal is calculated. An oracle result that a priori knows the indexes of the nonzero components in the result is also calculated. This procedure is repeated for each measurement number 50 times with random sparse signals, random measurement matrices and random noise realizations each time. Figure 1(a,b) shows the average ℓ_2 norm error for the tested algorithms for $K = 10$ and $K = 20$ sparsity levels respectively.

Both results in Fig. 1 show that the average reconstruction error for the proposed EMMP algorithm is lower than the compared CS and OMP algorithms for high enough measurements. It is also important to note that the EMMP method archives the performance of the oracle result after some measurement number level where both CS and OMP have an error offset compared to oracle result. Increasing the sparsity level of the signal increases the required number of measurements for achieving lower reconstruction errors. For all methods including EMMP, doubling the sparsity level nearly doubled the required number of measurements which is an expected result consistent with the information that the required number
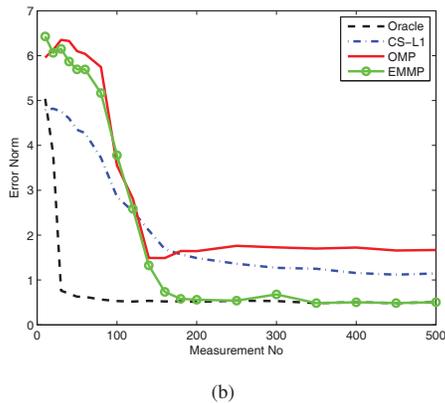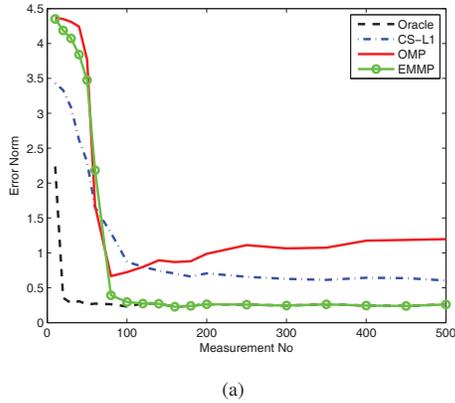
(a)

(b)

Fig. 1. (a) Average ℓ_2 norm error vs. compressive measurement number for the tested CS, OMP, EMMP and oracle results for sparsity levels (a) $K = 10$ and (b) $K = 20$

(a)

Fig. 2. Average ℓ_2 norm error vs. SNR for the tested CS, OMP, EMMP and oracle results

In the SNR regime where algorithms start to produce true reconstructions proposed EMMP algorithm outperforms the compared methods.

One of the important parameters of the EMMP algorithm is the S parameter controlling the sparsity level of the solution. To analyze the effect of this parameter a sparse signal with sparsity level $K = 20$ is randomly chosen and $M = 100$ measurements with a random gaussian measurement matrix are generated. An SNR level of 10dB is used. The sparse signal is reconstructed with the EMMP algorithm using $S = K - 3$, $S = K$ and $S = K + 3$. Hence lower, upper and correct number of sparsity levels are tested. The reconstructed and the true signals for each case are shown in Fig. 3(a-c) respectively.

of measurements is linearly dependent to the sparsity level of the signal. EMMP also follows this rule.

The results in Fig. 1 compare algorithm performances under a single SNR for varying number of measurements. It is also important to see the performance under various noise levels. To obtain the noise performance of the algorithm a sparse signal of dimension $N = 512$ and sparsity level $K = 20$ is measured with $M = 200$ compressive measurements. Again a random gaussian measurement matrix is used. WGN with SNR levels changing from -10 to 20 dB are tested. For each SNR level a new random sparse signal, with random measurement matrix and noise realizations are generated and this procedure is repeated 50 times. Figure 2 shows the average ℓ_2 norm error between the reconstructed signal and the true signal for CS, OMP, EMMP and oracle cases.

It can be seen from Fig. 2 that the EMMP has lower error values compared to OMP for nearly all SNR values. EMMP algorithm achieves the performance of the oracle at an SNR of 3 dB, while OMP achieves it nearly at an SNR level of 9 dB which shows a 6 dB SNR gain in performance. CS result shows slightly better performance in very low SNR regime compared to both OMP and EMMP but the error rates for all the algorithms in low SNR regime is high and reconstructed signals are in average not very close to the true signals.
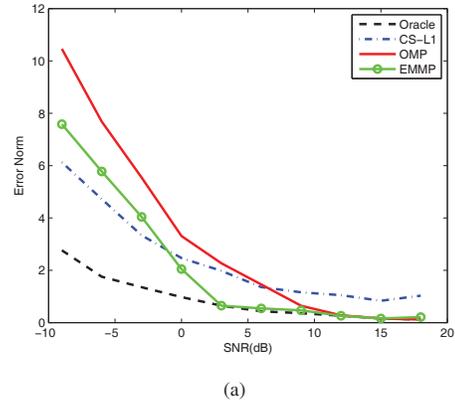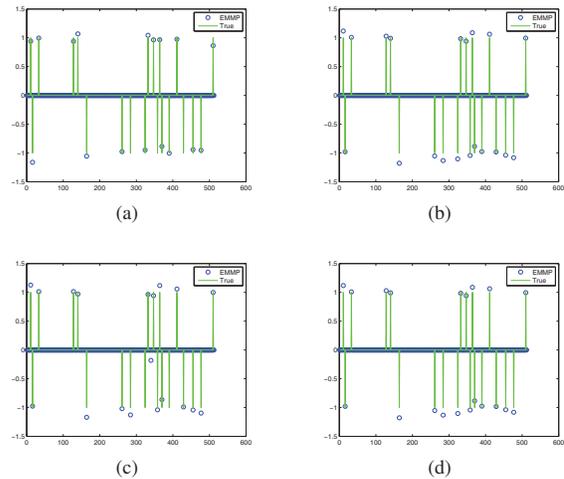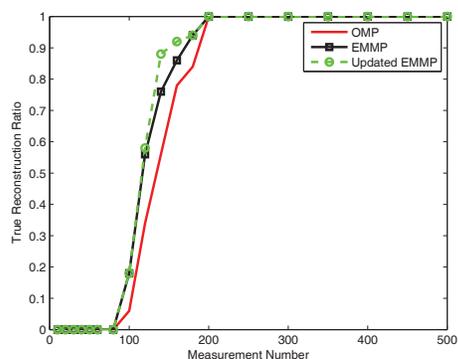

(a)

(b)

(c)

(d)

Fig. 3. True and reconstructed signals with EMMP algorithm when true sparsity is $K = 20$ and the sparsity level in EMMP is used as (a) $S = 17$, (b) $S = 20$ and (c) $S = 23$. (d) True and reconstructed signal with the updated EMMP algorithm that doesn't require an S parameter as an input.

Figure 3(a) shows that using a sparsity level S that is less than the true sparsity of the signal in EMMP creates a result where S of the K true nonzero signal components are reconstructed. Hence a best S sparse representation of the

signal is found. Using less sparsity level doesn't create nonzero components outside the support of the signal which is very important. In Fig. 3(b) sparsity level S is taken as same as K and the signal is correctly reconstructed. If S overshoots K, as in Fig. 3(c) still all support of the signal is correctly found but only one wrong nonzero component is reconstructed although $S = K + 3$ is used. Figure 3(d) is a result of an updated EMMP algorithm where no sparsity level S is required to be given as an input parameter. It is seen that the EMMP method creates the sparse signal correctly as in Fig. 3(b) without using an S parameter in the algorithm.

The updated EMMP algorithm can be basically summarized as follows. The EMMP algorithm is run with an S parameter starting from $S = 1$ and if the condition can not be satisfied for that S level within a fair amount of iterations then S parameter is increased and the EMMP algorithm is run again for the current S level. The iteration is stopped when the termination criteria is satisfied. By this way, EMMP searches for solutions with increasing sparsity and stops at the lowest sparsity level that it could satisfy the termination criteria. To analyze the performance of the updated EMMP algorithm a $K = 20$ sparse signal of dimension $N = 512$ is measured with varying number of measurements from 10 to 500. An SNR of 10dB is used. The sparse signal is reconstructed with EMMP, updated EMMP, and OMP methods. This procedure is repeated 50 times and the true reconstruction numbers are counted. A sample reconstruction is counted as true if the reconstructed signal satisfies the error condition i.e., $||\Phi * x - y|| < \epsilon$, it is a K or less sparse signal and the support set of the signal is not outside of the true signal support set. Figure 4 shows the true reconstruction performance of all the algorithms. It van be observed that both EMMP and updated EMMP algorithms perform similar and perform with higher true reconstruction rates compared to OMP method.

method treats the measurements as the incomplete data about the system and estimates the complete data corresponding to each support index with an iterative EM type algorithm. It is observed that the proposed EMMP method perform reconstruction with lower average reconstruction errors compared to CS and OMP algorithms.

REFERENCES

[1] D. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
[2] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Information Theory*, vol. 52, pp. 489–509, 2006.
[3] M. D. ad M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.
[4] M. Lustig, D. Donoho, and J. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007.
[5] V. Patel, G. Easley, D. H. Jr., and R. Chellappa, "Compressed synthetic aperture radar," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 244–254, 2010.
[6] A. C. Gurbuz, J. H. McClellan, and W. R. Scott Jr., "A compressive sensing data acquisition and imaging method for stepped frequency gprs," *IEEE Trans. Signal Processing*, vol. 57, no. 7, pp. 2640–2650, 2009.
[7] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," in *Int. Conf. on Information Processing in Sensor Networks (IPSN)*, April 2006, pp. 134 – 142.
[8] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, 1993.
[9] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Information Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[10] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comp. Harmonic Anal., arXiv math.NA 0803.2392*, 2008.
[11] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *preprint*, 2008.
[12] H. Huang and A. Makur, "Backtracking-based matching pursuit method for sparse signal reconstruction," *IEEE Signal Processing Letters*, vol. 18, p. 7, 2011.
[13] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Societys Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

(a)

Fig. 4. True reconstruction ratio vs the used number of compressive measurements for the tested OMP, EMMP and updated EMMP algorithms.

IV. CONCLUSIONS

This paper presented a novel expectation maximization based matching pursuit (EMMP) algorithm for sparse recovery in underdetermined linear system of equations. The proposed