Accelerated Proximal Algorithm for Finding the Dantzig Selector and Source Separation Using Dictionary Learning

Hayat ULLAH*, Muhammad AMIR, Muhammad IQBAL, Ahmad KHAN, Wasim KHAN

Abstract: In most of the applications, signals acquired from different sensors are composite and are corrupted by some noise. In the presence of noise, separation of composite signals into its components without losing information is quite challenging. Separation of signals becomes more difficult when only a few samples of the noisy undersampled composite signals are given. In this paper, we aim to find Dantzig selector with overcomplete dictionaries using Accelerated Proximal Gradient Algorithm (APGA) for recovery and separation of undersampled composite signals. We have successfully diagnosed leukemia disease using our model and compared it with Alternating Direction Method of Multipliers (ADMM). As a test case, we have also recovered Electrocardiogram (ECG) signal with great accuracy from its noisy version using this model along with Proximity Operator based Algorithm (POA) for comparison. With less computational complexity compared with ADMM and POA, APGA has a good clustering capability depicted from the leukemia diagnosis.

Keywords: Accelerated Proximal Gradient Algorithm (APGA); Alternating Direction Method of Multipliers (ADMM); Dantzig Selector; Electrocardiogram (ECG) signal; Leukemia data

1 INTRODUCTION

Compressed sensing (CS) considers the recovery and separation of signals that have sparse representation in some known transform domain i.e. Fourier transform, wavelet transform, sine and cosine transforms etc. In CS, original signal is recovered from far fewer samples as compared to the conventional Shannon theory [1-5]. Most of the practical signals are sparse in one or other domain, for example the sinusoid shaped signals are mostly sparse in the Fourier domain unlike images which have sparse representation in cosine domain. Applying l_1 norm recovery techniques to these types of signals has produced better results.

The stationary and non-stationary signals using CS theory is discussed in [6] with the main goal to achieve high resolution Time-Frequency (TF) distributions. Composite signals are considered sparse by taking a few ambiguity functions around the origin and thus making it an undersampled problem. Instead of using regular optimization techniques for CS recovery, [7] uses two-dimensional L-statistics modified techniques for the reconstruction of signals highly contaminated by impulsive noise in the ambiguity domain. A few samples corrupted by noise are removed by sorting operation making a CS function. Some of the applications of sparse signal processing theory and CS, including decomposition of signal, speech processing, medical images denoising and reconstruction are given in [8-11].

On the other side in comparison with the propositions in [6, 7], the algorithms in [12, 13] use the overcomplete dictionaries for the recovery of sparse signals by using the low pass characteristics of the signals. If the restricted isometric property is satisfied by the measurement matrix, then recovery through l_1-norm of nearly sparse signals is guaranteed.

In [14] an attempt is made to separate sinusoids from other signals by considering the ambiguity function points along the zero time-lag. This approach is inspired by the theory used in direction of arrival and array signal processing. For the recovery of narrow band signals when corrupted by noise in the shape of non-stationary signals

using CS approach, [15] makes observations in the TF domain instead of time domain as in the case of standard CS theory. By taking short-time Fourier transform (STFT) of the signal and using different size of windows for mapping the local information of sparsity in STFT to the global information in Fourier transform. The non-stationary signals which are not sparse over the entire data may be sparse in a local window of STFT which can represent and reflect the local behavior of the composite signal. For a successful separation of non-stationary signals, the time-frequency regions corresponding to these signals are identified and removed by using the L-filter as mentioned in [16, 17]. L-filter statistics is applied first for the separation of the composite signal and is then followed by CS recovery techniques. The separation of the stationary and non-stationary signals is inspired by its application in radar signal processing, micro-Doppler effects and rigid body points. The signals reflected from the main body of the plane are considered as stationary signals and are sparse in the Fourier domain, while signals from the moving fans are treated as non-stationary signals [18]. To get the signal of interest, the overlapping signals in TF domain are removed keeping the points where narrowband signals exist i.e. stationary.

In some cases composite signals may be degraded by noise. The separation of the signal into its components without loss of a good portion of the signal is not easy and the separation becomes more difficult specifically in the case of undersampled composite signals. Proximity Operator based Algorithm is introduced in [19] which is used for separation of undersampled composite signals and hand writing (images) in the presence of noise [20]. Here Dantzig selector with overcomplete dictionaries is used for the separation of composite signals and images. In [19, 20], fixed point proximity operator based algorithm is used which is a special case of proximal mapping and which is also known as indicator operator or projection operator. In this work we have used the accelerated version of proximal mapping with overcomplete dictionaries for better convergence rates. The algorithm has better performance in terms of data clustering and source separation. The comparison of Dantzig selector with that of the Least

Absolut Shrinkage and Selection Operator (LASSO) is given in [21-24]. LASSO does the matching of the candidate solution to that of the observations while Dantzig selector model tries to limit the residuals. The solution produced by the Dantzig selector is always sparser in terms of l_1-norm [25].

The problem of finding the Dantzig selector is solved in four steps by the authors in [26]. At first the problem is formulated as conic, secondly, its formulation is as dual expression, third step is to apply smoothing, and fourth one is to solve it by optimal first order method. The algorithm is flexible enough and its significance is depicted from its usage in the solution of CS problems. In signal processing, machine learning mathematics etc. different types of convex cone problems arise, which are solved by this approach. Due to the stability and computational efficiency, the algorithm is compatible with that of the LASSO. The ADMM is used for finding the Dantzig selectors in [27]. They use the non-monotone gradient algorithm to solve this problem in multiple steps and compare this method with that of the first-order method used in [26] for finding the Dantzig selector. The use of the ADMM for finding the Dantzig selectors outperforms the method used in [26] in terms of the calculation time while producing the results of comparable quality. In this paper, we are using APGA for finding the Dantzig selector with overcomplete dictionaries and apply the model for source separation. We have used this model for the diagnosis of leukemia in patients along with ADMM. We have also successfully recovered and separated a noisy ECG signal using this model and POA.

2 MATERIAL AND METHODS
2.1 The Dantzig Selector with Overcomplete Dictionaries

Consider the estimation of a parametric vector $\beta \in R^n$ from a linear regression model:

$$y = X\beta + \xi \tag{1}$$

Where $y \in R^m$ is a vector of observations, $X \in R^{n \times m}$ sensing matrix with $m \ll n$, and $\xi_{m \times 1}$ is independent and identically distributed (i.i.d) noise vector. This makes it an underdetermined problem. To find the most suitable vector $\beta^* \in R^n$ out of possible candidate solutions along with splitting into its components is the goal of the research which makes it an underdetermined problem. If β is a vector of parameters and $s_{p \times 1}$ is a composite sparse signal in the overcomplete dictionary $\Psi_{n \times p}$, then $\beta_{n \times 1}$ is given by the relation:

$$\beta = \Psi_s \tag{2}$$

The vector β and the composite signal s are unknown, but in most of the problems one can imagine the overcomplete dictionary Ψ. For example, in the sinusoidal signals with periodic impulses, concatenate discrete Fourier transform and identity matrices for proper dictionary representation. Let $p = 2n$ and assume the $n \times p$ dictionary Ψ by horizontally concatenating orthonormal basis of appropriate dimensions, $\Psi = [\Phi \; \theta]$, and the components of β take the sparse representations $\beta_\Phi = \Phi s_\Phi$ and $\beta_\theta = \theta s_\theta$, with $\beta = \beta_\theta + \beta_\Phi$.

In a more commodious form:

$$\beta = [\Phi \; \theta] \begin{bmatrix} s_\Phi \\ s_\theta \end{bmatrix} \tag{3}$$

If the observations y are given, then for the recovery of $s = \begin{bmatrix} s_\Phi \\ s_\theta \end{bmatrix}$, and therefore, for the components β_Φ and β_θ, one can solve the Danzig selector problem using the overcomplete dictionaries.

$$s \in \min_{s \in C^{2n}} \left\{ \|s\|_1 : \left\| D^{-1} \Psi^T X^T \left(X\Psi_s \right) - y \right\|_\infty \le \delta \right\} \tag{4}$$

D is a diagonal matrix having elements $d_{ij} = diag \left\{ \left\| (X\Psi)_j \right\|_2 \right\}$ and is used for normalization of the dictionary couple. For D to be invertible, the entries of X must be i.i.d and Ψ having entries as nonrandom basis. Random sensing matrix is non coherent but also not orthogonal to fixed bases. It implies that for D to be invertible, $d_{ij} \ne 0$ for each j. The greater the randomness of sensing matrices, the greater the incoherence with the overcomplete dictionaries, the smaller the isometry constant and successful recovery through the l_1-norm, $\delta > 0$.

2.2 Proximal Algorithms

Dantzig selector is computed by characterizing a solution to model shown in Eq. (4) by employing the APGA and the use of l_1-norm. The model is then used for separation of composite signals. Suppose a composite function $f(x) = g(x) + h(x)$ and

$$\min f(x) = \min (g(x) + h(x)) \tag{5}$$

Assume $g(x)$ is convex, differentiable and $dom(g) = R^n$, while h is convex but not necessarily differentiable. APGA produces better convergence. For sparsity and separation, we use the overcomplete dictionaries so that the signal of interest has sparse representation in that dictionary and can be recovered from it.

According to [19], the proximal mapping (or the proximal operator) of the convex function h is defined as:

$$prox_h (x) = \begin{matrix} arg \min \\ u \end{matrix} \left(h(u) + \frac{1}{2} \|u - x\|_2^2 \right) \tag{6}$$

for $h(x)$, we have the following three different cases.

2.2.1 The Three Cases of Proximal Algorithms

CASE-1

When $h(x) = 0$: $prox_h(x) = u$ and is the same as the gradient descent algorithm and the update rule. i.e $f(x)$ is not differentiable but $g(x)$ is differentiable and t is the step size parameter, then the update rule:

$$x^+ = x - t\nabla g(x) \qquad (7)$$

The proximal algorithm is then reduced to:

$$prox_h(x) = {}^{arg\,min}_u \left(\frac{1}{2}\|u - x\|_2^2 \right) \qquad (8)$$

For this case the convergence rate is slow and is $O(1/\in)$.

CASE-2

When $h(x) = I_C(x)$: i.e. projected gradient descent or proximal gradient descent. If $C \in R^n$ is a closed and convex set then considering [20],

$$ {}^{arg\,min}_{x \in C} f(x) = {}^{arg\,min}_{x \in R^n} g(x) + I_c(x) \qquad (9)$$

Where:

$$I_c(x) = \begin{cases} 0 & x \in C \\ \infty & x \notin C \end{cases} \qquad (10)$$

Which is an indicator function of C.

Hence $prox_h(x) = {}^{arg\,min}_u \left(\frac{1}{2}\|u - x\|_2^2 + I_c(x) \right)$, which is again reduced to Eq. (8), i.e. Gradient projection descent will be reduced to gradient descent if $x \in C$, means if $I_C(x) = 0$. i.e. $prox_h(x) = P_C(x)$, projection operator onto C.

Hence, $prox_h(x) = {}^{arg\,min}_u \left(\frac{1}{2}\|u - x\|_2^2 \right)$, this method is used in [20], with overcomplete dictionaries they have separated the composite signals. The update step is:

$$x^k = prox\left(x^{(k-1)} - t_k \nabla g\left(x^{(k-1)}\right)\right) \text{ or } x^+ = P_C\left(x^{(k-1)} - t_k \Delta g\left(x^{(k-1)}\right)\right) \quad (11)$$

CASE-3

When $h(x) = \|x\|_1$: i.e. $prox_h$ is the "soft-threshold" (shrinkage) operation and is given by:

$$prox_h(x_i) = \begin{cases} x_i - 1 & x_i \geq 1 \\ 0 & |x_i| \leq 1 \\ x_i + 1 & x_i \leq -1 \end{cases} \qquad (12)$$

Or more compactly,

$$prox_h(x) = {}^{arg\,min}_u \left(\frac{1}{2}\|u - x\|_2^2 + \|x\|_1 \right) \text{ or } prox_h(x) = S_t(x) \quad (13)$$

It is also converging slowly and has a convergence rate of $O(1/\in)$.

2.2.2 Accelerated Proximal Algorithm

For all the three previously discussed cases, the convergence rate is $O(1/\in)$ in which the projection or the indicator operator $h(x) = I_c(x)$ as used in [20]. It performs usual gradient update and then projection back onto C. The method is faster than subgradient but it is not easy to define the proximal operator for most of the functions in closed form. For a specific problem someone must know the proximal operator in closed form. Also each iteration evaluates $prox_t(\ )$ twice as done in [20], iterations can be time consuming or fast, depending on h.

We use the accelerated proximal algorithm to find the Dantzig selector and then apply the overcomplete dictionaries to separate the signals into respective components. To do this we consider $h = 0$ and then apply the acceleration.

The acceleration idea to the proximal algorithm is first introduced by Nesterov [28], in which each step uses entire history of previous steps calling the proximal operator twice. But we use the extension introduced by Beck and Teboulle [29], which accelerates the proximal gradient algorithm by achieving the optimal $O(1/\sqrt{\in})$ convergence rate for composite signals as each step uses information from two last steps and makes one proximal call.

For the accelerated version, [29] introduces another term in the update equation, which makes Eq. (7) as follows:

$$v = x^{(k-1)} + \frac{k-2}{k+1}\left(x^{(k-1)} - x^{(k-2)}\right) \qquad (14)$$

$$x^k = prox_{t_k}\left(v - t_k \nabla g(v)\right) \qquad (15)$$

First step is to start with $k = 1$, which makes it just a usual proximal gradient update. After that, $v = x^{(k-1)} + \frac{k-2}{k+1}\left(x^{(k-1)} - x^{(k-2)}\right)$ which carries some "momentum" from previous iterations. The momentum term with the updates is given by:

$$y = (1 - \theta_k)x^{(k-1)} + \theta_k v^{(k-1)} \qquad (16)$$

$$x^k = prox_{t_k} y\left(v - t_k \nabla g(y)\right) \qquad (17)$$

$$v^{(k)} = x^{(k-1)} + \frac{1}{\theta_k}\left(x^{(k)} - x^{(k-1)}\right) \qquad (18)$$

where $\theta_k = \frac{2}{k+1}$, substitute the expression for $v^{(k)}$ in expression for y, will give the complete updated accelerated proximal gradient update.

$$y = x^{(k-1)} + \theta_k \left(\frac{1}{\theta_k - 1} - 1 \right)\left(x^{(k-1)} - x^{(k-2)} \right) \tag{19}$$

$$x^k = prox_{t_k} y\left(y - t_k \nabla g(y) \right) \tag{20}$$

2.3 The Dantzig Selector Formulation in Terms of Accelerated Proximal Algorithm

Consider Eq. (4), suppose $A = \Psi^T X^T X \Psi$, and $\gamma = \Psi^T X^T y$ and the set $S = \left\{ s : \|s - \gamma\|_\infty \le \delta \right\}$, then Eq. (4) will be given as:

$$\min_{s \in C^{2n}} \left\{ \|s\|_1 : \|As - \gamma\|_\infty \le \delta \right\} \tag{21}$$

If "s" is a solution to Eq. (21), the iterative update for Eq. (21) through accelerated proximal gradient algorithm is given:

$$v^{(k)} = s^{(k-1)} + \frac{1}{\theta_k}\left(s^{(k)} - s^{(k-1)} \right) \tag{22}$$

$$s^k = prox_{t_k} y\left(y - t_k \nabla g(y) \right) \tag{23}$$

where $y = s^{(k-1)} + \theta_k \left(\frac{1}{\theta_k - 1} - 1 \right)\left(s^{(k-1)} - s^{(k-2)} \right)$.

Pseudo code for summarizing the algorithm of APGA is given below.

Algorithm 1: Accelerated Proximal Gradient Algorithm (APGA) with overcomplete dictionaries

---

Parameters setting: y is given as the set of observations.

The sensing matrix X and the overcomplete dictionary B according to the problem are chosen.

Initialization: $v^0 = v^1 = 0$, $s^0 = 0$, $\beta^0 = 0$

Iterations: generate the sequence $\left\{ (s^k, v^k) : k \in N \right\}$ using Eq. (22) and Eq. (23).

while (stopping criterion not met) **do**

$v^{(k)} \leftarrow s^{(k-1)} \frac{1}{\theta_k}\left(s^{(k)} - s^{(k-1)} \right)$

$s^k \leftarrow prox_{t_k} y\left(y - t_k \nabla g(y) \right)$

$\theta_k = \sqrt{m t_k}$, t_k is step size

$k = k + 1$

end while

Post-processing: Use the appropriate processing scheme to construct the Dantzig estimator $\hat{s}$ from final output of the while loop.

Output: Get the approximate separate signals from $\hat{s}$ and Ψ using Eq. (3)

The stopping criteria is given in the following from [19, 20, 29].

1. $\dfrac{\left\| D^{-1} \Psi^T X^T \left(X \Psi s^k - y \right) \right\|_\infty}{\max\left\{ \|s^k\|_2, 1 \right\}} \le \sigma$, where $0 < \sigma < 1$, or

2. $supp(s^k)$ is stationary for some iterations, $supp(s^k)$ is support of $s^k : \left\{ i : x(i) \ne 0 \right\}$.

3. $stopingcriteria = norm\left(s - s^0 \right) \big/ norm(s) < \varepsilon$.

3 RESULTS AND DISCUSSION

In this section, we consider the classsification of two different types of Leukemia patients using the standard data set easily available online and also the recovery and separation of ECG signal from its noisy version using APGA with overcomplete dictionaries and ADMM.ECG (bio-signals), are corrupted by the noise due to muscular movements or by sensors, etc. We sparsely recover the components using both the algorithms with great accuracy. However, ADMM is not performing well in case of complex dictionaries [20].
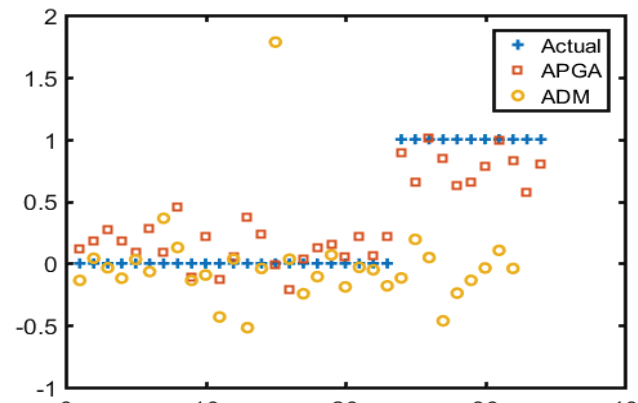

Figure 1 Actual values, recovered through APGA and ADMM for classification of Leukemia patients

3.1 Leukemia Diagnosis in Patients

We have taken the leukemia data set [19] and compared our algorithm with ADMM in [27] for diagnoses of 35 cancer affected patients, whether a patient is suffering from specific type of cancer. We assign "1" to the patients suffering from Acute Lymphocytic Leukemia and "0" to the patients having Acute Myelogenous Leukemia. From Fig. 1 it is clear that APGA values are more outspread (means dividing the data easily into two chunks to reflect the two types of Leukemia patients) than ADMM showing the tendency of APGA over ADMM for adaptation to linear regression model and separation of values into two different clusters. Tab. 1 shows the number of iterations and the run time taken by ADMM and APGA for the recovery of Dantzig selector β^* for different values of "δ".
δ is dependent on noise level and is given by $\delta = \sigma\sqrt{2\log n}$.

Table 1 The number of iterations and recovery time taken by APGA and ADMM for different values of δ

		Number of iterations		Run time in seconds	
		APGA	ADMM	APGA	ADMM
δ	0.07	80	1000	3	1167
	0.15	120	1040	4	100
	0.20	90	1015	3	93
	0.25	7	739	0.5	78
	0.30	7	817	0.5	78
	0.40	7	100	0.5	1345

3.2 Separation and Recovery of ECG Signal

A composite signal is generated by combining ECG signal and the sparse spike signal. The ECG signal is $\beta_\Phi = ecg(x)$, with $x = 0, 1, 2, 3, ..., 1023$. The sparse spike component β_θ is formed by choosing a set A of size 'a' uniformly at random.

The overcomplete dictionary Ψ is a concatenation of the discrete Fourier transform and the Euclidean Bases (or identity matrices). The composite signal is observed by $y = X\beta + \xi$. X is $m \times n$ sensing matrix with the elements chosen from normal distribution and the columns are normalized, ξ is a noise vector with zero mean and variance σ^2. The parameters used are $n = 1024$, $m = 512$, with stopping criteria $\varepsilon = 10^{-6}$ and $\eta = 6$ for $\sigma = 0.01$. APGA is applied to the noisy incomplete observations $y = X\Psi s + v$ for the separation of ECG signal from the sparse spike component. The step wise pseudo code is as under.


Pseudo Code 1: APGA for Separation of Composite Signal (ECG denoising)

Input: The set of observations y

Overcomplete Dictionary: The overcomplete dictionary used in the recovery and separation of ECG signal is a composite dictionary of discrete Fourier transform and Euclidean Bases (or Identity matrices).

1. Applying APGA to y observation for efficient establishing of coefficient vector $\hat{s}$ which is sparse and satisfies Eq. (4) and also recover the estimated Dantzig selector $\hat{\beta} = \Psi\hat{s}$.

2. The component signals β_Φ and β_θ are recovered by means of $\hat{s}$.
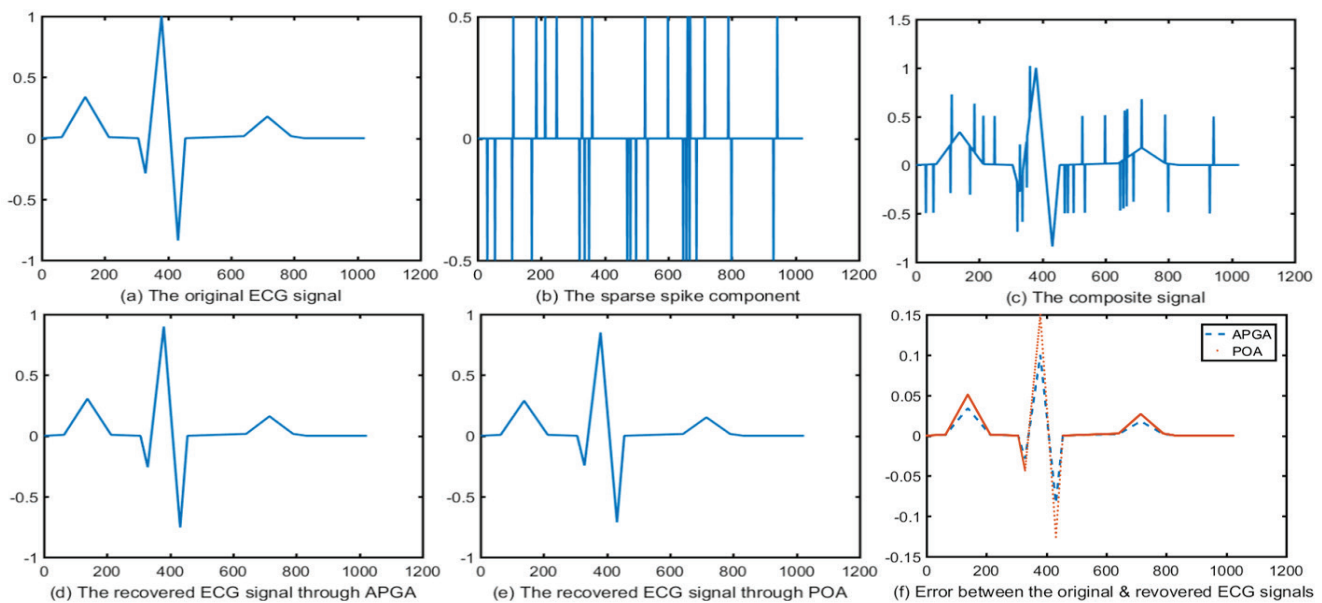
Figure 2 (a) The original ECG signal, (b) The sparse spike components, (c) The composite signal, (d) The recovered ECG signal through APGA, (e) The recovered ECG signal through POA, (e) The recovery error of the ECG signal $\beta_\Phi - \hat{\beta}_\Phi$, through APGA and POA

Simulation results for the recovery and separation of ECG signal through Accelerated Proximal Gradient Algorithm (APGA) and Proximity Operator based Algorithm used in [19, 20] and the error between the original and recovered signal by the algorithms are given in Fig. 2. The simulations are run in MATLAB R2016a using the system with Intel i5-5200U CPU (2.20GHz) and 8GB RAM.

Tab. 2 shows pixel wise error for a few pixels between the APGA and POA. Also Signal to Noise Ratio (SNR), Peak Signal to Noise Ratio (PSNR) and Mean Square Error (MSE) values are given in Tab. 2.

Table 2 Pixel wise recovery error by APGA and POA for a few pixels, SNR, PSNR and MSE are given in the table

Pixels number	150	375	425	725	1000
APGA	0.035	0.100	0.075	0.010	0.000
POA	0.050	0.150	0.130	0.025	0.005
	SNR	PSNR		MSE	
APGA	20.00	33.8523		$4.1188e^{-04}$	
POA	16.4782	30.3305		$9.2673e^{-04}$	

4 CONCLUSIONS

APGA with overcomplete dictionaries is used for the separation of composite signals and denoising. From Fig. 1 and Pseudo Code 1, it is clear that APGA based Dantzig selector model is a more intelligent learning algorithm than ADMM in terms of clustering and separation. This model is at least order one faster than ADMM in terms of the time taken by simulation. Also the results produced by ADMM using real valued dictionaries are good enough but in case of the complex valued dictionaries like Fourier transform dictionary, it is difficult to efficiently recover the composite signals. The use of APGA for finding the Dantzig selector and source separation in combination with the overcomplete dictionaries, is robust to noise levels and size of the problem and works well in both real and complex valued dictionaries. This model is equally applicable in medical signal processing such as Electroencephalogram (EEG) and Electrocardiogram (ECG) etc. Having applications in CS if the sparsity of the

noisy signals is improved then the method can be used in further compression of the signal.

5 REFERENCES

[1] Senturk, A. & Kara, R. (2016). An analysis of image compression techniques in wireless multimedia sensor networks/Analiza nacina sazimanja slike u bezicnoj visemedijskoj senzorskoj mrezi. *Tehnicki Vjesnik-Technical Gazette*, 23(6), 1863-1870. https://doi.org/10.17559/TV-20150313101530

[2] Li, H., Wu, F., & Zhang, X. (2018). Fusion of multispectral and panchromatic images via local geometrical similarity. *Tehnički vjesnik*, 25(2), 546-552. https://doi.org/10.17559/TV-20171023105844

[3] Sajjadieh, M. H. & Asif, A. (2015). Compressive sensing time reversal mimo radar: Joint direction and doppler frequency estimation. *IEEE Signal Processing Letters*, 22(9), 1283-1287. https://doi.org/10.1109/LSP.2015.2396650

[4] Sun, B., Feng, H., Chen, K., & Zhu, X. (2016). A deep learning framework of quantized compressed sensing for wireless neural recording. *IEEE Access*, 4, 5169-5178. https://doi.org/10.1109/ACCESS.2016.2604397

[5] Ullah, H., Amir, M., Haq, I. U., Khan, S. U., Rahim, M. K. A., & Khan, K. B. (2018). Wavelet Based De-noising Using Logarithmic Shrinkage Function. *Wireless Personal Communications*, 98(1), 1473-1488. https://doi.org/10.1007/s11277-017-4927-3

[6] Flandrin, P. & Borgnat, P. (2010). Time-frequency energy distributions meet compressed sensing. *IEEE Transactions on Signal Processing*, 58(6), 2974-2982. https://doi.org/10.1109/TSP.2010.2044839

[7] Stanković, L., Stanković, S., Orović, I., & Amin, M. G. (2013). Robust time-frequency analysis based on the L-estimation and compressive sensing. *IEEE Signal Processing Letters*, 20(5), 499-502. https://doi.org/10.1109/LSP.2013.2252899

[8] Khan, K. B., Khaliq, A. A., Shahid, M., & Shah, J. A. (2016). A new approach of weighted gradient filter for denoising of medical images in the presence of Poisson noise. *Tehnički vjesnik*, 23(6), 1755-1762. https://doi.org/10.17559/TV-20150620130712

[9] Giacobello, D. & Jensen, T. L. (2018). Speech dereverberation based on convex optimization algorithms for group sparse linear prediction. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 446-450. https://doi.org/10.1109/ICASSP.2018.8462560

[10] Peško, I., Trivunić, M., Cirović, G., & Mučenski, V. (2013). A preliminary estimate of time and cost in urban road construction using neural networks. *Tehnicki vjesnik/Technical Gazette*, 20(3)

[11] Shah, J. A., Qureshi, I. M., Omer, H., Khaliq, A. A., & Deng, Y. (2014). Compressively sampled magnetic resonance image reconstruction using separable surrogate functional method. *Concepts in Magnetic Resonance Part A*, 43(5), 157-165. https://doi.org/10.1002/cmr.a.21314

[12] Candes, E. J., Eldar, Y. C., Needell, D., & Randall, P. (2011). Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 31(1), 59-73. https://doi.org/10.1016/j.acha.2010.10.002

[13] Donoho, D. L. & Elad, M. (2003). Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ1 minimization. *Proceedings of the National Academy of Sciences*, 100(5), 2197-2202. https://doi.org/10.1073/pnas.0437847100

[14] Amin, M. G., Belouchrani, A., & Zhang, Y. (2000). The spatial ambiguity function and its applications. *IEEE Signal Processing Letters*, 7(6), 138-140. https://doi.org/10.1109/97.844632

[15] Stanković, L., Orović, I., Stanković, S., & Amin, M. (2013). Compressive sensing based separation of nonstationary and stationary signals overlapping in time-frequency. *IEEE Transactions on Signal Processing*, 61(18), 4562-4572. https://doi.org/10.1109/TSP.2013.2271752

[16] Djurovic, I., Stankovic, L., & Bohme, J. F. (2003). Robust L-estimation based forms of signal transforms and time-frequency representations. *IEEE Transactions on Signal Processing*, 51(7), 1753-1761. https://doi.org/10.1109/TSP.2003.812739

[17] Stankovic, L., Dakovic, M., & Thayaparan, T. (2014). Time-frequency signal analysis with applications. Artech house

[18] Stankovic, L., Thayaparan, T., Dakovic, M., & Popovic-Bugarin, V. (2013). Micro-Doppler removal in the radar imaging analysis. *IEEE Transactions on Aerospace and Electronic Systems*, 49(2), 1234-1250. https://doi.org/10.1109/TAES.2013.6494410

[19] Prater, A., Shen, L., & Suter, B. W. (2015). Finding Dantzig selectors with a proximity operator based fixed-point algorithm. *Computational Statistics & Data Analysis*, 90, 36-46. https://doi.org/10.1016/j.csda.2015.04.005

[20] Prater, A. & Shen, L. (2015). Separation of undersampled composite signals using the Dantzig selector with overcomplete dictionaries. *IET Signal Processing*, 9(3), 226-234. https://doi.org/10.1049/iet-spr.2014.0048

[21] Jing-yi, L., Hong, L., Dong, Y., & Yan-sheng, Z. (2016). A new wavelet threshold function and denoising application. *Mathematical Problems in Engineering*. https://doi.org/10.1155/2016/3195492

[22] James, G. M., Radchenko, P., & Lv, J. (2009). DASSO: connections between the Dantzig selector and lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(1), 127-142. https://doi.org/10.1111/j.1467-9868.2008.00668.x

[23] King, E. J., Kutyniok, G., & Zhuang, X. (2011). Analysis of data separation and recovery problems using clustered sparsity. *In Wavelets and Sparsity XIV*, 8138, 813818. International Society for Optics and Photonics https://doi.org/10.1117/12.892723

[24] Li, Q., Micchelli, C. A., Shen, L., & Xu, Y. (2012). A proximity algorithm accelerated by Gauss-Seidel iterations for L1/TV denoising models. *Inverse Problems*, 28(9), 095003. https://doi.org/10.1088/0266-5611/28/9/095003

[25] Zheng, S. & Liu, W. (2011). An experimental comparison of gene selection by Lasso and Dantzig selector for cancer classification. *Computers in biology and medicine*, 41(11), 1033-1040. https://doi.org/10.1016/j.compbiomed.2011.08.011

[26] Becker, S. R., Candès, E. J., & Grant, M. C. (2011). Templates for convex cone problems with applications to sparse signal recovery. *Mathematical programming computation*, 3(3), 165. https://doi.org/10.1007/s12532-011-0029-5

[27] Lu, Z., Pong, T. K., & Zhang, Y. (2012). An alternating direction method for finding Dantzig selectors. *Computational Statistics & Data Analysis*, 56(12), 4037-4046. https://doi.org/10.1016/j.csda.2012.04.019

[28] Nesterov, Y. (2013). Gradient methods for minimizing composite functions. *Mathematical Programming*, 140(1), 125-161. https://doi.org/10.1007/s10107-012-0629-5

[29] Beck, A. & Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1), 183-202. https://doi.org/10.1137/080716542

Contact Information:

Hayat ULLAH, PhD scholar,
(Corresponding author)
International Islamic University Islamabad (IIUI),
DEE, FET, Sector H-10, 44000 Islamabad, Pakistan
E-mail: hayat.phdee104@iiu.edu.pk

Muhammad AMIR, Dean, Professor,
International Islamic University Islamabad (IIUI),
DEE, FET, Sector H-10, 44000 Islamabad, Pakistan
E-mail: m.amir@iiu.edu.pk

Muhammad IQBAL, Assistant Professor,
International Islamic University Islamabad (IIUI),
DEE, FET, Sector H-10, 44000 Islamabad, Pakistan
E-mail: muhammad.iqbal@iiu.edu.pk

Ahmad KHAN, Assistant Professor,
COMSATS University Islamabad **(CUI)**,
Abbottabad Campus,
Department of Computer Science,
University Road, Tobe Camp, 22060 Abbottabad, Pakistan
E-mail: ahmadkhan@cuiatd.edu.pk

Wasim KHAN, Assistant Professor,
International Islamic University Islamabad (IIUI),
DEE, FET, Sector H-10, 44000 Islamabad, Pakistan
E-mail: wasim.khan@iiu.edu.pk