International Journal of Scientific Engineering and Applied Science (IJSEAS) – Volume-6, Issue-8, August 2020
ISSN: 2395-3470
www.ijseas.com

Recovery of low rank and sparse structure pursuit by alternative minimization

Amina Shiroz[1], Aparna P.R[2]
1 Sree Chitra Thirunal College of Engineering Pappanamcode, Thiruvananthapuram, India
2 Sree Chitra Thirunal College of Engineering Pappanamcode, Thiruvananthapuram, India

Abstract
Recovery of low rank and sparse structure refers to be decomposition of data matrix having high dimensional data into low dimensional space. In this paper, we address a non-convex alternative minimization algorithm for extract low rank and sparse component. Compare to conventional algorithms having convex relaxation methods, our proposed algorithm is more efficient for large scale problems and big data. By using notion of bounded difference of gradients and suitable intialization, proposed non convex algoirthm gets linearly convergence to the global optima under standard conditions. This alternative minimization scheme is employed to solve resulting constrained minimization algorithm. The numerical experiments/synthetic data show that the newly method ensures less computational complexity as compare with the popular convex relaxation based approach.This method exactly recovers the underlying low rank and sparse matrices from noisy measurements under a standard incoherence assumption. Algorithm provides a principal approach for Robust Principal Component Analysis and used in applications such as data analysis, with myriad applications ranging from web search to bio informatics to computer vision and image analysis for foreground and background seperation.
Keywords: **Big data, *L*ow rank matrix, Sparse matrix, Non–convex optimization, Alternative minimization, Matrix decomposition.**

1. Introduction

In data mining and machine learning, the major problem is big data or complex data. Typical approach for solve this big data is refers to be data decomposition, which split the given data into low rank component and sparse component from a noisy data matrix. In most of the machine learning setups, an underlying tenet is that the data often contains some type of structure that enables intelligent representation and processing. These structure is obtained by estimable mathematical tools for dealing with high dimensional data. This technique is termed as Dimensional Redution(DR). The objective of this technique is to extract useful infrormation from data by learning the approximatable low dimensional mapping of high dimensional data[2]. The dimensional reduction methods includes Principal Component Analysis[PCA][15], Linear Discriminant Analysis[LDA] [2,1] and independent component analysis[IDA]. All of this mostly used method for dimensional reduction is PCA. It acquire low dimensional information of high dimensional data. For this occasion, observed that PCA is delicate to outliers and to relative scaling of the original variables and noise.

To overcome this problem we proposed a newly method, Robust Principal Component Analysis[RPCA]. RPCA[3] refers to be decomposition of low rank component and sparse component of data matrix. It acquire low rank calculation of input matrix and it also eliminate sparse corruptions from the input matrix. The low rank component is minimum number of independent rows and columns of matrix and it often refers to low dimensional pattern of revelant data . In the case of sparse, corresponds to either grossly corrupted measurements or pattern irevelant data. Sparse matrix or sparse array have less number of non zero elements.

Conventional algorithms such as Augmented Lagrange Method[ALM][8], Alternative Direction Method of Multipliers[ADMM] and other non convex algorithms are also used for dimensional reduction. The performance of these algorithms is good, but better is our proposed algorithm. In the case of ALM it exactly recovers the low rank and sparse model. It holds true for high probability random low rank and sparse components. But the only disadvantage is, singular value decomposition in full dimension is perform at each iteration. So we can replace singular value decomposition with Lncozeration partial SVD technique, but it is

less efficeint. In ADMM, does not processed exact minimization. It opproximately minimize the variable. Inorder to recover this problem we go through our proposed algorithm

In this paper as apart of motivation consider a general model as:

$$Y = X\,L^{\star} + S^{\star} \qquad (1)$$

Here the parameter matrix being the superposition of low rank matrix and sparse matrix. Parameter matrix is define as set of real parameters that use constraints to estimate the real parameters. In this above equation X is the general design matrix $X \in R^{n*d_1}$ Design matrix is simply regression matrix which contains the explanatory variables of set of objects. Each row represents an individual objects and columns represents variables of their specified values of object. $L^{\star}$ represents the low rank matrix where $L^{\star} \in R^{d_1*d_2}$. Low rank matrix guarantees number of linearly independent columns (rows) in matrices is less than the total number of columns (rows) is matrices[10,19]. $S^{\star}$ represents the sparse matrix where $S^{\star} \in R^{d_1*d_2}$ [9]. It comprises of fewer number of non zero entries. And finally $Y \in R^{n*d_2}$ be the response or observe matrix. We observe the similar models with or without the noise. The mathematical result of robust principal component analysis is attained by the prevalent approach called convex relaxation. It retrieve the low rank component $L^{\star}$ and sparse component $S^{\star}$ together from X and Y. solve the following convex problem:

$$argmin_{L,S} \|Y - X(L+S)\|_F^2 + \lambda \|L\|_{\star} + \mu \|S\|_{1,1} \qquad (2)$$

rank function $\|L\|$ is the relaxed by the nuclear form means sum of the singular values and cardinality function $\|S\|_{1,1}$ is relaxed by the l_1 norm of S, signifies the sum of all entries in absolute values. λ, μ > 0 are the regularization parameters[5,6]. Equation 1 can be resolved by convex optimization algorithms such as proximal gradient and alternative methods of multipliers. These algorithms enjoys good convergence properties, due to the nuclear norm regularization. But they require singular value decomposition at each iteration, it is very expensive for high dimensional problems with large d_1 and d_2. To get better of this obstruction of existing optimization algorithms for low rank and sparse matrix learning we go through the reparametrizing process. So here we modify the unidentified low rank matrix L as the invention of two smaller matrices U and V so $L = UV^T$, where $U \in R^{d_1*r}$ and $V \in R^{d_2*r}$. U and V automatically satisfied the low rank constraint. Thus equation (2) change into following constrained problem:

$$argmin_{U,V,S} \frac{1}{2} \|Y - X(L+S)\|_F^2 \; subject \; \|S\|_{0,0} \leqslant s \qquad (3)$$

Where s > 0 is the tuning parameter. The proposed optimization problem can be compute by alternative minimization. An alternative minimization[12] scheme can be hired to search for an static point which satisfies the necessary optimization condition. It is done by solve one variable by fixing other two variables until convergence. In this case for example solve U by fixing V and S. In spite of the great achievement of reparameterization, the theoretical properties of non convex optimization[6] algorithms is very difficult because of equally limitation of factorization. Non convex optimization refers to be minimizing the function over non convex sets. Non convex functions which have more than one global optimal points. In this paper suggest a low rank plus sparse structure by means of alternative minimization algorithm[12]. It works under the linear convergence to the global optimal point under certain standard conditions. The crucial part of this algorithm is the notion of bounded difference of gradients which leads to sufficient condition for the convergence of non convex optimization[9]. Based on the notion the algorithm converges exponentially fast to global optimal point and recovers low rank and sparse components[18]. The main contributions of paper are;

- In this paper, a novel dimensional reduction is proposed by applying low rank and sparse constraints the proposed algorithm become more robust for noise and outliers.
- The algorithm can also be applied to extract the useful information from the data models. And is potentially more powerful and informative to extract features.
- The iterative way of alternative minimization[12,25] algorithm resolve the result non-convex optimization problem through constraints and linearly convergence to a global optimal point. It accurately mends low rank and sparse assemblies.
- By using numerical behavior and comparison it with existing algorithms for check the efficiency and validity of proposed algorithm. Synthetic database (MNIST database) is used to evaluate the performance of algorithm.

2. related work

2.1 Low rank representation via augmented Lagrange method

Low rank representation[6,10] is dimensional reduction technique and it represents the data models only through low rank structure by augmented lagrange method. It recovers the data matrix from corrupted observations only through row space. Compare to our proposed algorithm LLR does not have sparse component. This method reverse only for convex optimization[11] problems since we all know that proposed algorithm based on non convex optimization.

Consider the equation;

$$X=AZ \qquad (4)$$

X represents the sets of data vectors, each represented by linear combinations of the bases in a dictionary A =[$a_1, a_2, \ldots\ldots a_m$] and Z = $\left[Z_1, Z_2, \ldots\ldots Z_n\right]$ is the coefficient matrix

$$min_z rank(z) s.t X=AZ \qquad (5)$$

In the above optimization problem is difficult to solve the discrete nature of rank function. So in the above equation replace rank function by nuclear norm function.

$$min_Z \|Z\|, s.t X=AZ+E$$

$$min_{Z,E,J} \|J\|+\lambda\|E\|_{2,1} s.t X=XZ+E, Z=J \quad (6)$$

lagrage method mainly focused on replace constrainted optimization problems by series of unconstrainted problems. Apply lagrange method to equation (6) becomes:

$$L=\|J\|+\lambda\|E\|_{2,1}+\left[Y_1^t(X-XZ-E)\right]+\left[Y_2^t(Z-J)\right]+\frac{\mu}{2}\left(\|X-XZ-E\|_F^2\right)+(Z-J)_F^2 \quad (7)$$

$\|.\|$ indicates th nuclear norm of matrix. And E be the error matrix. Then the perform segmentation of data into respective subspaces, after that compute an affinity matrix and encodes pairwise affinities between vectors. Proposed algorithm recovers grossly measurements or noisy elements from data matrix. In the case of LLR data matrix exhibits as low rank srtucture and take part in dimensionality reduction[30,16]. In the case of alternative

minimization algorithm recovers low rank plus sparse components of data matrix.2.2 Generalized Prinicipal Component Analysis(GPCA)

GPCA drawn the data model from a union of multiple subspaces in the form of algebraic way. Subspaces that contains data point is obtained by taking the gradient of a polyniminal at that point(3). It provide sucess segmentation at certain conditions but it does not have restriction on the subspaces. This is the main disadvantage of this method. It is very difficult to estimate the polynominals of real data which leads to sensitive to noise. This method is has computational complexity and very expensive so, we go through the proposed method low rank plus sparse model.

In conventional algorithms for low rank and sparse model or RPCA is less accurate and sensitive to errors. Alternative minimization[26] has been observed in matrix completion, dictionary learning, expectation minimization algorithm, sparse coding etc but cannot suitable to our algorithm due to the presence of sparse component. Waters[16] proposed a greedy algorithm for extraction of low rank and sparse matrix from compressive measurements. Projection gradient method proposed by Kyrillidis and Cevher [11] over non convex sets recover low rank and sparse components from linear noisy environments. Feng [15] proposed a online algorithm based on robust principal component analysis [RPCA] which mainly focused on reparameterization $L=UV^T$ and l_1 norm and converges to optimal solution but they do not specify convergence guarantee. As part of modification of online algorithm Netrapalli[18] put forward a non convex optimization algorithm with provable guarantee. The resultant enjoys good convergence properties but it is confused whether it satisfy general low rank and sparse structure that show in equation (3) Their exist a similar model as ours, which derive low rank and sparse structure by it do not factorize the low rank as U and V and leads to singular value decomposition at each iteration. So this method involves time consuming and computationally expensive. This algorithm was suggest by Yan et al. finally one of the similar technique is proximal gradient descent algorithm which designed by Chen and Wainwright [22] , is fast low rank matrix estimation with provable guarantee however, they centred on symmetric positive semidefinite matrix. So compare with these all methods our algorithm enjoy good convergence to global optimal point under standard conditions and have lower per iteration. The proposed non- convex alternative minimizaton algorithm have less time consuming and computational complexity.

3. proposed method

Consider the linear model discuss in equation (1) To overcome time consuming and computational complexity parameterize the parameter $L=UV^T$ equation (1) and obtained as :

$$Y=X\left(UV^T\right)+S \quad (8)$$

Here we defined non-convex optimization algorithm for solve equation (3). Consider the loss function:

$$L\left(U,V,S\right)=\frac{1}{2}\|Y-X\left(UV^T+S\right)\|_F^2 \quad (9)$$

Loss function[20] is a method of evaluating how much alternative minimization algorithm models the given data models. It also shows the similarity of the prediction model to the expected outcome. Loss function converts learning problem into optimization problem. An optimization problems seeks to minimize a loss function.

3.1 Non convex optimization

optimization[6] means it is the action of making better through maximizing or minimizing process. Convex optimization has some desirable properties of linear programming problems. Convex function refers to minimizing a convex function over convex sets. Convex function have only one feasible region since it have one global optimal point. In the case of non convex optimization minimizing function with respect to non convex sets it holds multiple feasible region for that reason it carry many optimal points. Non convex function is so far but more general to non linear problems. So by implementing proposed algorithm (alternating minimization algorithm) into non convex optimization we can find global optimal point. Fig 1 shows graphical representation of convex and non convex function

The main point noticed while designing a non-convex algorithm is ensure massive modeling power for express a learning problems. This leads to increasing the perform of relaxation techniques. In most of the machine learning problems, models are trained and predict the performance on the base of this optimization techniques. For increasing the accuracy of learning and prediction problems structural constraints such as low rank and sparsity are used or objective itself designed a non convex function. This type of representation is mainly imposed in high dimensional space or train non linear models such as tensor models and deep learning networks[23].

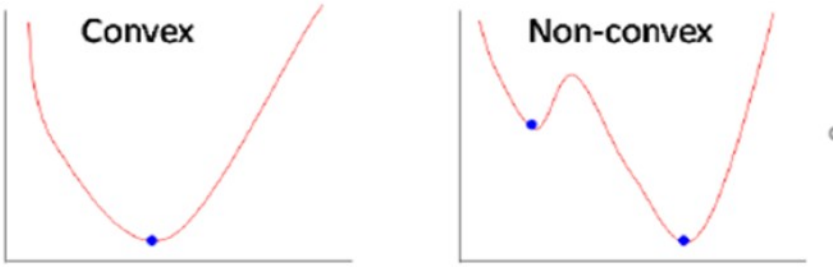


Fig 3. 1: convex function and non - convex function

3.2 non – convex optimization alternative minimization algorithm

inputs: number of iteration T, rank parameter *K*, sparsity parameter s

1. **initialize:** $V^{(0)} \epsilon R^{d_2 Xk}$ and $S^{(0)} \epsilon R^{d_1 Xd_2}$
2. **for** t = 0 to T – 1 **do**
3. $U^{(t+1)} = argmin_u \mathcal{L}(U, V^{(t)}, S^{(t)})$.
4. $\left\{U^{(t+1)}, R_1^{(t+1)}\right\}$ = QR($U^{(t+1)}$).
5. $V^{(t+1)} = argmin_v \mathcal{L}(U^{(t+1)}, V^{(t)}, S^{(t)})$.
6. $\left\{V^{(t+1)}, R_2^{(t+1)}\right\}$ = QR($V^{(t+1)}$).
7. $S^{t+1} = argmin_s \mathcal{L}(U^{(t+1)}, V^{(t+1)}, S)$.
8. $S^{(t+1)}$ = Truncate($S^{(t+1)}, S$)

End for

Our aim is to find U, V, S and we all know that rank of UV^T is r and the sparsity s of S. Therefore loss function $L(U,V,S)$ is to minimized. Before start algorithm we have to initialize U_0, V_0, S_0 and set number of iterations T, t = 0 to T – 1 displayed in step 1 and 2 . Alternative minimization [25,26]technique is described in given algorithm. In step 3 we have to minimize the loss function L with respect to U where the other two variables V, S remains fixed. Thus obtained the minimized function U^{t+1} . In next step 4 we apply QR decomposition to factorize U^{t+1} into U^{t+1} and $R_1^{(t+1)}$, where U^{t+1} is the orthogonal basis and $R_1^{(t+1)}$ is the corresponding upper triangular matrix. Note that QR decomposition is not necessary in practice, it is important in theoretical analysis. This process is repeated in step 5 and 6 with respect to V and obtained V^{t+1} and $R_2^{(t+1)}$. Although the reparameterization process of L[19,25] makes the optimization problem easier to analyze, it boosts the computational efficiency. In step 7 perform alternative minimization with respect to S to gain S^{t+1} . In final step (8) execute a truncation step to implement the sparsity of the iterate S. The truncate (. , s) function is defined as:

$$\left[truncate(S,s)\right]_{JK} = \underbrace{S_{JK}}_{if S_{JK}\ is\ among\ the\ tops\ large\ magnitude} = \underbrace{0}_{otherwise} \qquad (10)$$

To obtain S^{t+1} ,[14,20] Truncated function conserves the entries of S^{t+1} with top s large magnitudes[14] and set the rest to zero. s is the tuning parameter that carry out the sparsity level. This truncated step formulate the sparse structure through out the solution path. By alliteratively executing the alternative minimization and truncation step, the proposed algorithm converges to the global optimal solutions.

3.3 application

Main application of low rank plus sparse model[RPCA] is compressed sensing[16] and background modeling for the foreground separation in videos. Compression sensing is used for efficient acquiring and reconstructing the signal. Many research have been in the field of compressed sensing. Compressed sensing exploits the fact that an image is sparse in some appropriate basis to reconstruct under sampled data without loss of image information. Successful application of compressed sensing requires image sparsity. For this required nyquist Shannon theorem A sparse signal is one with less number of non zero elements. The number of zero valued elements divided by total number of elements is called sparsity of the matrix. The idea of compressed sensing can be extended to matrices enabling recovery of missing or corrupted entries of a matrices under low rank and sparse conditions. So sparsity signal play a vital role in compressed sensing.

The important and exact application of low rank plus sparse model[RPCA] is background modeling for foreground separation in video scenario fig(2). Image frames in each video can be exhibits as data matrix. This data matrix is taken as input for implementing the algorithm[24]. First of all we have to find the low rank(L) and sparse structure(S) of data matrix. In order to reduce computational complexity factorize $L=UV^T$,U and V are the small product of matrix L . Express these matrices as column vectors while set the input for execution of algorithm. After implementation of proposed algorithm obtained background separation and foreground estimation image[22]. Background of image frames remains static, and they are closely related so it exhibits as low rank structure and foreground varies frames to frames it reveals sparse structure.

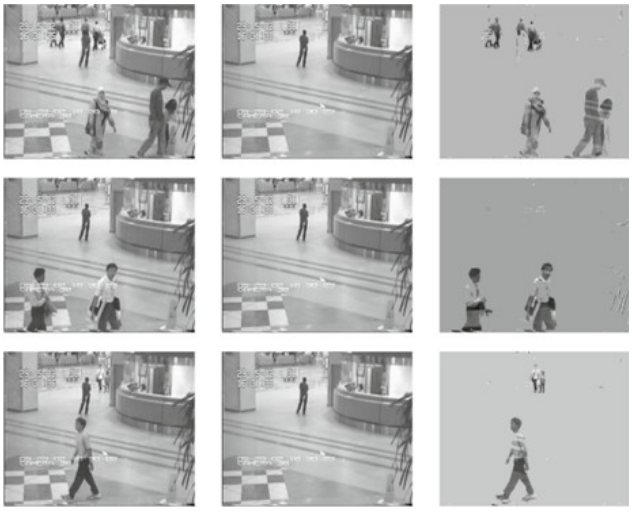


Fig 3.2: (1)original frame, (2) background estimation (3) foreground detection

Next section we discuss how much the algorithm attains linear convergence rate. Compare to convex relaxation methods such as proximal gradient descent and alternating direction methods of multipliers proposed algorithm is more efficient because of reparameterization and factorization. One among the crucial factor is it converges at standard conditions to make effective algorithm.

3.4 standard conditions and assumptions:

Consider the equation of singular value decomposition of $L^{\star} \epsilon R^{d_1 * d_2}$ be $L^{\star} = U^{\star} \Sigma^{\star} \bar{V}^{T\star}$ where $U^{\star} \epsilon R^{d_1 * r}$ and $V^{\star} \epsilon R^{d_2 * r}$ be the left and right matrices and Σ be the diagonal matrices such as diag($\sigma_1, \sigma_2, \ldots\ldots \sigma_r$) ϵ R^{r*r} .rank parameter K is obtained as $\frac{\sigma_1}{\sigma_r}$

These are the assumptions made used for implementing and theoretical analysis of algorithm

Assumption 1: Assume if rank ($L^{\star}$) ≤ r. $L^{\star}$ has incoherent μ, then singular value decomposition of $L^{\star}$ holds

$$max_{1 \le i \le d_1} \| e_i^T U \|_2 \le \mu \sqrt{\frac{r}{d_1}}$$

$$max_{1 \le i \le d_1} \| e_i^T U \|_2 \le \mu \sqrt{\frac{r}{d_1}} \qquad (11)$$

This incoherence assumption of μ implies that L is not too spiky. Since elements of those matrices are sparse and very few large entries. This is widely used in matrix completion and matrix decomposition.

Assumption 2: Assumption 2: This assumption is known as Restricted Isometric Property(RIP). This a usual assumption which satisfies design matrix X. This is mainly used in compression sensing. This is sufficient condition for sparse reconstruction. In this assumption, δ_r be the restricted isometric constant of rank where $0 < \delta_r < 1$

For a singular value decomposition of $L \epsilon R^{d_1 * d_2}$ (low rank) such that rank(L) ≤ r, [19] then the design matrix satisfies the following conditions

$$\left(1-\delta_r\right)\|L\|_F^2 \leq \|X\|_F^2 \leq \ 1+\ \delta_r\|L\|_F^2 \qquad (12)$$

For (sparse matrix) any S $\epsilon R^{d_1 * d_2}$ and sparse matrix should satisfy, $\|S\|_{0,0}$ ≤ s, then it holds,

$$\left(1-\zeta_s\right)\|S\|_F^2 \leq \|X\|_F^2 \leq \ 1+\ \zeta_s\|S\|_F^2 \qquad (13)$$

Several random matrix such as Gaussian, Bernoulli and partial Fourier matrices satisfy RIP property with high probability. ζ_s be the restricted isometric constant of sparse component. $0<\zeta_s<1$. In the case of identity design matrix $\delta_r=0, \delta_s$ =0.

With help of these assumptions, we can describe the linear convergence of the non convex optimization alternative minimization algorithm. Before go through the theorem we lay out some technical conditions with simplifies theorem through notation. Let

$$v=max\sqrt{\left(1+\delta_{2r}\right)\left(1+\zeta_{2s}\right)}\ , \frac{1+\delta_1}{1-\delta_1} \qquad (14)$$

And the condition be, dimension $d=min\left(d_1, d_2\right)$ is sufficient large, hence That is d is greater than or equal two, such that:

$$d \geqslant max\, 120\sqrt{2r}\,v\mu\sigma_1, 16\frac{\sqrt{r}}{7\sqrt{2}k\sigma_1 16-1}, 8\mu\sqrt{r}\left(3\sqrt{2}k\sigma_1-3\delta_1\sqrt{r}-1\right)^{2\backslash 3}s \qquad (15)$$

$\delta_1 \leq 8\mu\left(s\backslash d^{3\backslash 2}\right), \delta_1$ is sufficiently small.

For rotational simplification, we define

$$\gamma=\frac{3\sqrt{2}k}{1-24\sqrt{2}\,v\mu\sqrt{r}\,s\backslash d^{3\backslash 2}k\sigma_1} \qquad (16)$$

$$\rho=\frac{96\sqrt{2}\,v\mu\sqrt{r}\,s\backslash d^{3\backslash 2}k\sigma_1}{1-24\sqrt{2}\,v\mu\sqrt{r}\,s\backslash d^{3\backslash 2}k\sigma_1} \qquad (17)$$

then the convergence theorem be:

Assume the sparsity parameter is chosen as *S=4S*, S *in 4S* indicate the sparse matrices at initialization of algorithm. Suppose the initialization $\bar{U}^{(0)}, \bar{V}^{(0)}, \bar{S}^{(0)}$ *satisfy* $\|\bar{S}^{(0)}-S\|_2 \leq \sigma_1$,

$$\|\bar{V}^{(0)}-\bar{V}^{\star}\|_F \leq 2\mu\sqrt{\frac{\left(1+\zeta_{2s}\right)r}{\left(1+\delta_{2r}\right)}}\left(\frac{s}{d}\right)^{3/2} \qquad (18)$$

$$\|\bar{U}^{(0)}-\bar{U}^{\star}\|_F \leq 3\delta_1\sqrt{r}+\mu\sqrt{\frac{r}{d_1}\frac{s^{1.5}}{d_2}} \qquad (19)$$

under the condition in equation (15), we have to find the value of t, it hold that:

$$\|U^{(t+1)}-U^{\star}\|_F+\gamma\|V^{(t+1)}-V^{\star}\|_F\leq\rho\left(\|U^{(t)}-U^{\star}\|_F+\|V^{(t)}-V^{\star}\|_F\right) \quad (20)$$

where the value of γ and ρ are illustrated in equation (15) and (16). in addition $U^{\star},V^{\star},S^{\star}$ be the initial iteration of the convergence theorem. From this value it will converge and easy to attain. If we scale both design matrix X and response matrix Y, hence value of σ_1 is sufficient large and satisfy the condition $\|\bar{S}^{(0)}-S^{\star}\|_2$ is satisfied[20]. $\bar{U}^{(0)}$ and $\bar{V}^{(0)}$ are the true solution existing as standard in this literature[20] of alternative minimization[17]. By compare the algorithm of robust principal component analysis with (9) is not optimal. In circumstance of non convex optimization it became globally optimal with respect to other convex relaxation methods. The complete derivation can be found through reference[23,25,26].

4. experiments and results

In this section, discuss the performance of our proposed algorithm in the case of real data sets and synthetic data sets. Compare proposed method with proximal gradient descent[18], alternative direction method of multipliers[8] and other convex relaxation methods[6,29]. And also equate with augmented Lagrange method[8,6]. In the case of alternating projection method, we investigate under the base of example RPCA. It is check on the base of numerical results. For evaluation process, we have two use two measures, first measure is estimation error based on frobenius form. That is $\|\hat{L}-L^{\star}\|_F$ and the second measure is computational time per second. For implementing the proposed algorithm initialize $\bar{U}^{(1)}$ is the semi orthogonal matrix and $S^{(1)}$ is the zero matrix[20]. We set r as the true rank s as 4 s and K as rank parameter which obtained from diagonal matrix while computing singular value decomposition. Proposed algorithm is performed under python 3.7 on spyder(anaconda navigator). It works on Ubuntu 20.04 LTS (Linux).

4.1 numerical experiments

Synthetic data are used for numerical experiments, it is artificially manufactured data. Our first experiment is conduct using [RPCA] Robust Principal Component Analysis.

Robust Principal Component Analysis[RPCA]: Consider the general model in equation(1). S is taken as zero matrix and it is generated by uniform random distribution. In the stimulation[3] we set $d_1=d_2=d$. Here we accompany different settings, by changing the value of *d* and *r* and by changing the value of iteration. And we can use the value of d in between 100 and 5000, like that r varies from 5 to 50. The iterative value depends on convergence of proposed algorithm its about 15 times[24,27]. For this settings we can concluded as our algorithm works at lower per iteration which imposes high value of d. compare of settings with alternative direction of multipliers and proximal gradient descent. Here the estimation error of both *L* and *S* is identical and approximately zero.

Augmented Lagrange Method[ALM]: Numerical solvers on the nucleus plus l^1 norm model for robust principal component pursuit is used for comparison with our proposed algorithm. Among these ALM is better for practice. So we choose ALM for comparison with non- convex optimization alternative minimization algorithm the major cost of ALM or alternative minimization algorithm[8] depends on Singular Value Decomposition [SVD]. So in the case of ALM their perform both partial and full SVD in each iteration. This leads to high computational cost. When shift from full SVD to partial SVD, the rank of low rank component in previous iteration reach below a certain threshold. ADMM(alternative direction mothod of multipliers) [26]is used for partial updates of dual variables. It is also a variant augmented Lagrange scheme. Compare to our algorithm the problem is solved approximately by solving first variable and the other variable fixed and algorithm directly update the dual variable. This process repeat alliteratively until convergence. This method not equal to the exact minimization(alternative minimization algorithm)In the case of proximal gradient descent method, is also use for non convex differentiable algorithms[28]. It may contains two settings one is full observation setting and second one is partial observation setting. Each algorithm has two phases sparse

estimation and iterative phase. Here we consider gradient based iterations of the variables U, V, S. the main difference in partial observation and full observation settings is introducing of sampling rate. Code in run in MATLAB. Codes are available in online[6]. And the code of existing algorithms obtained from the corresponding authors.

Table 1:comparison of our algorithm with other convex relaxation based methods in term of $\|\hat{L} - L^{*}\|_F$ 10^{-2} and time in second for RPCA

Methods	D = 100, r = 5		D = 1000, r =10		D =500, r = 15		D=5000, r = 50	
Proximal gradient	error	time	error	time	error	time	error	time
	1.57	0.58	4.48	625	4.03	600	More than 1hr	More than 1 hr
ALM	1.21	2.03	1.36	2.48	1.50	2.00	3.06	5.58
ADMM	0.05	6.45	0.08	898.45	0.078	785	More than 1hr	More than 1hr
Non-convex	5.45	0.59	7.00	2.16	6.03	1.42	2.45	266
Proposed algorithm	0.00	0.0018	0.0015	3.00	0.002	2.16	0.003	98.58

4.2 Experiments on MNIST database.

For the implementation of algorithm we used a synthetic database produced using the MNIST(Modified National Institute of Standards and Technology database) obtained from keras(TensorFlow). It is a large database of handwritten digits from 0 - 9 in the form of 28×28 gray scale images. MNIST data set contains 60,000 training images and 10,000 testing images. Overall shape of this MNIST images 60,000x784 and its corresponding to 60,000 labels. database is intended to serve as machine learning algorithms. It is zip file or .gz extended file and is prepossessed in spyder platform. T. For normalizing the pixel value between 0 and 1, diving each value with maximum value 255.

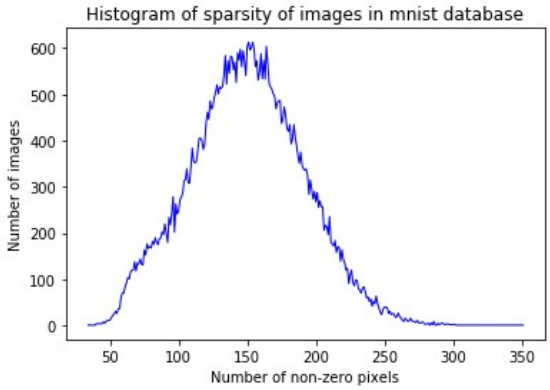


figure 4.1: histogram of sparsity of images in MNIST database

know we can discuss histogram of MNIST data set to know about sparsity. Figure 4.1 plot the histogram representation of sparsity. The mean sparsity of MNIST data set is obtained as 150. that is data set found sparsity in nature. So we can say that MNIST data set is sparsity data set. But we need to test our algorithm with both low rank as well as sparse data. So we have to create synthetic data both sparse and low

rank. Hence, manually we convert sparse MNIST data set into low rank plus sparse data set. It is done by adding a low rank matrices to MNIST data. For different digits different low rank matrix created randomly and add with a rank 2. For example consider a digit 0, one matrix is added for some number of digit 0 from the set of all digits. Like this another matrix is added for the rest of the digit 0 in the set of 60,000 digits such that rank of low rank matrix is 2 and overall matrix is 5 which is also a rank. The resultant obtained is sum of low rank plus sparse MNIST data. Thus the newly obtained data (L+S) have low rank (L)is 5 and sparse data both of size 60,000x784. The proposed algorithm is implemented on newly created MNIST data set.

Implementation: implement newly created MNIST data set to proposed algorithm at two different settings

1. Fix a digit(label fixed)study the average error performance of the algorithm for many images with same label. Keeping the number of observations fixed
2. fix an image(both label and data point fixed). Study the error performance of algorithm by increasing the number of observations starting from 20

First of all, consider first experiment is done by fixing the number of iterations and taking different images of the same digit and finding its average error. In this way find the average error of different images of different digits. The number of iterations and images are fixed for every data matrix. For our purpose we select the digit 0 as the single digit and the algorithm is implemented for different images of same digit 0 from the synthetic data set.

The algorithm recovers the image and error of each image is founded . We used l_2 norm as a error parameter. After finding the error for each images of the digit 0, average error is plotted versus the number of iterations. The average error plot obtained is shown in figure 4.2. error plot graph shows that average mean square error of different images of 0 out of 60,000 images tends to zero at about 30 iterations. Such that we can plot error versus iteration graph we obtain similar graph. About 20 times we repeat this procedure to obtain minimum value in graph(4.2) and (4.4). Graph 4.3 reconstructing a single image digit 0 and its error plot graph.

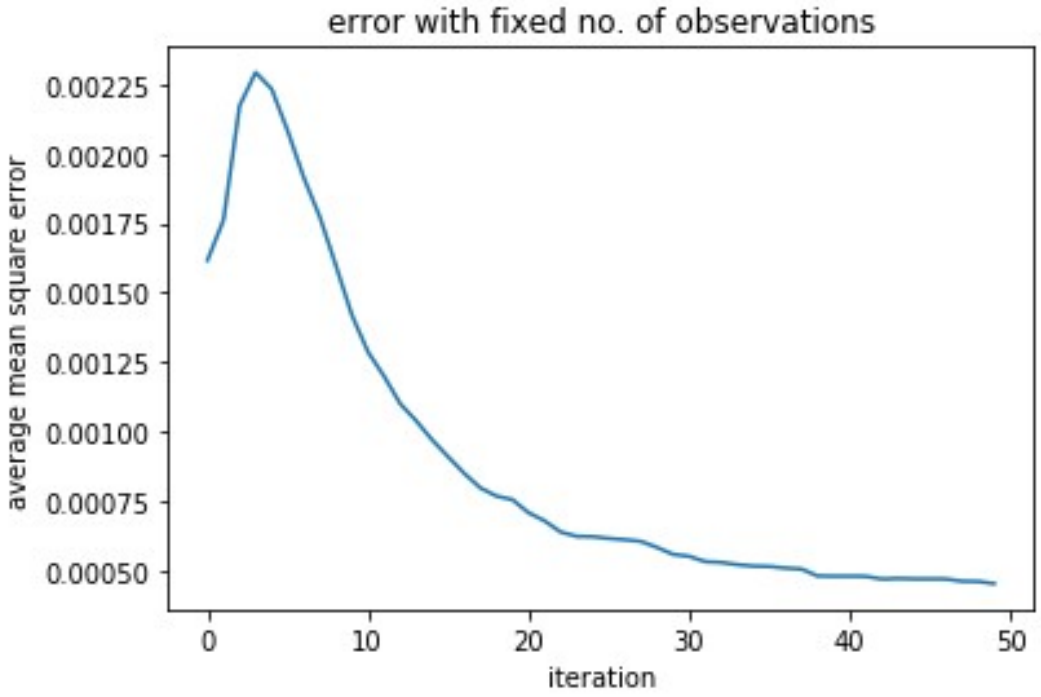
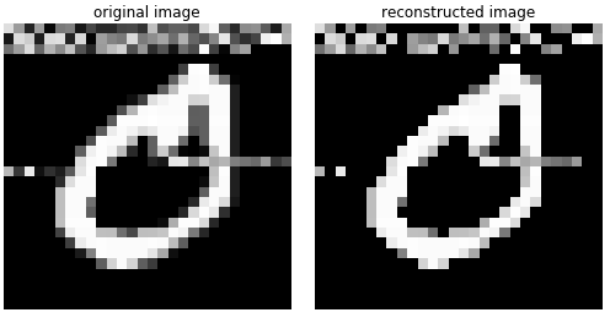
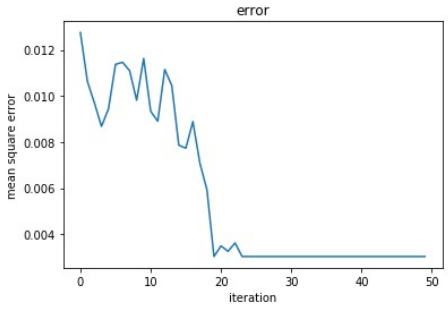


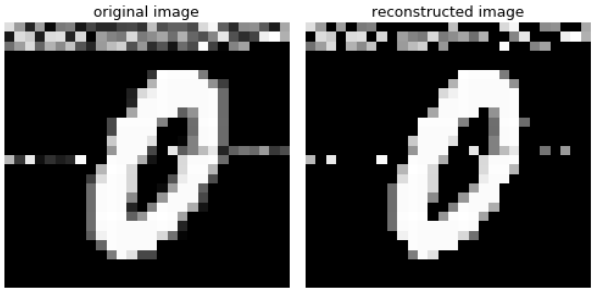
fig 4.2: average error from different images of digit 0



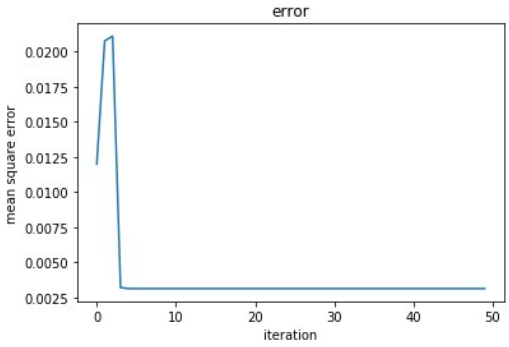
(a)



(b)



(c)



(d)

fig 4.3: (a),(c) original images and reconstructed sample images of 0, (b),(d) error plot v/s mean square error of different images 0

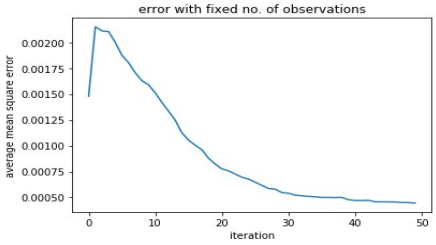
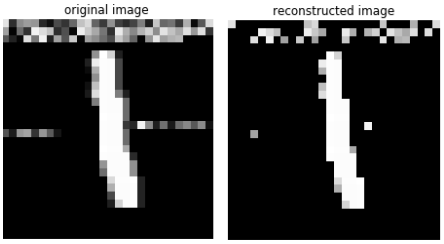


fig 4.4: average images of different images of 1



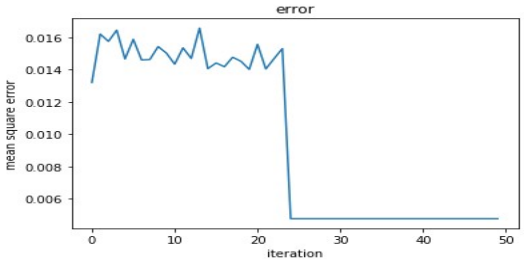


fig 4.5: error plots and reconstructed image of 1

The second experiment is done by taking only a single image of a single digit, we took the digits 0 – 9 into consideration for example we prefer the digit 1 and the algorithm is implemented with different number of observations. Here also the number of iteration is fixed. The data matrix taken is same for all the different observations and has a size of 28x28. The error for different number of observations for a single image of the digit is founded and plotted versus iterations.

Figure 4.6 shows the result of experiment done on digit 1. we start experiment from 20 observations of digit 1. figure 4.6 (b) shows the reconstructed images at 22 compare to original image it reduces the clarity and reduce the mean square error at 10 iterations shown in figure 4.7(a). Like that image of digit 1 at 24 and 26 observations figure 4.6(c,d) is less clear compare to original image and as we go through the error plot graphs, the mean square error attains approximate zero at iterations up to 38 and 12. In the case of 28 observations figure 4.6(e) the original image is approximately same as the reconstructed image and its error plot graph attains approximate zero at about 25 iterations. we continue this experiment from 0 – 9 and obtain the reconstructed image as original image at 28 observations. And the error plot graph attain zero below 30 iterations. From the resultant of the second experiment show that as the number of observations increase still 28 and the input/original image close to the reconstructed image.

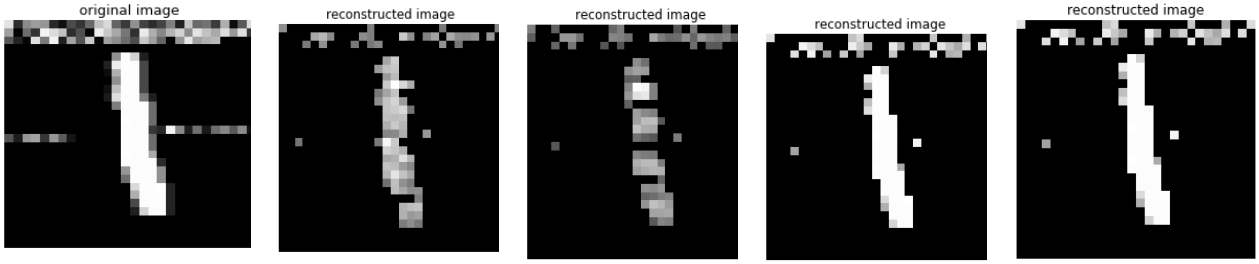


Fig 4.6 (a) Original image of digit 1 (b) reconstructed image with observations 22 (c) reconstructed image with observations 24 (d) reconstructed image with observations 26 (e)reconstructed image with observations 28

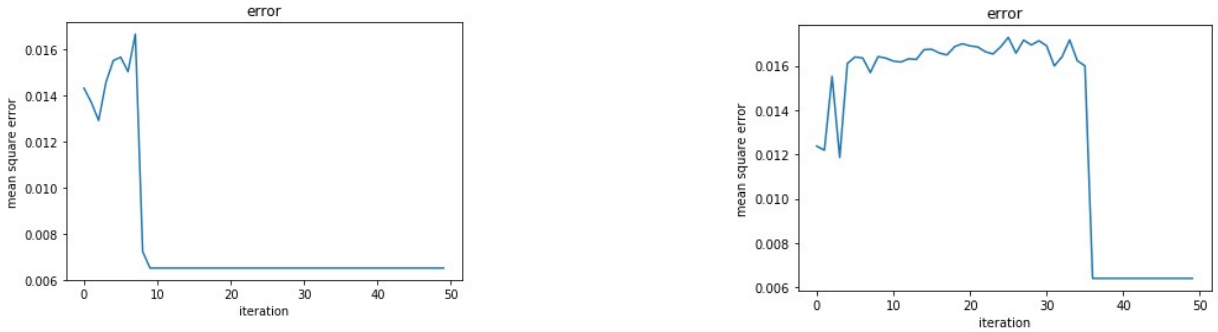


fig 4.7: (a)error plot graph of digit 1 with 22 observations (b) error plot graph of digit 1 with 24 observations

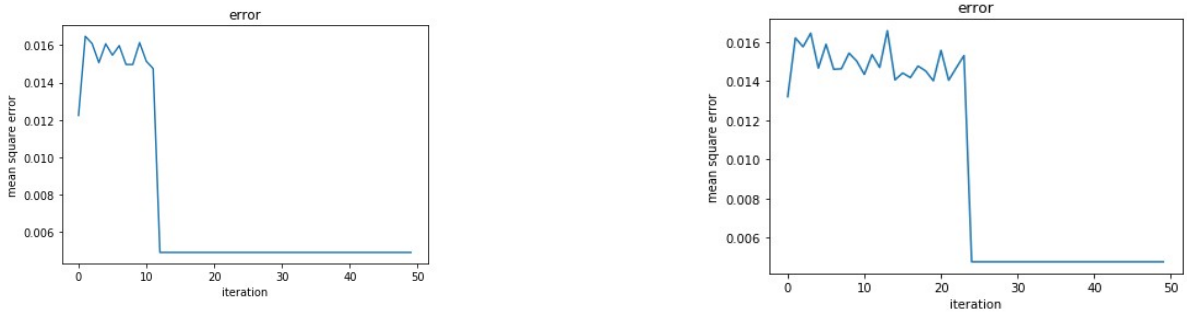


fig 4.7: (a) error plot graph of digit 1 with 26 observations (b) error plot graph of digit 1 with 28 observations

5. conclusion

In this paper proposed a novel statistical method called non convex optimization alternative minimization algorithm for recovers exact low rank and sparse structure pursuit under certain standard conditions. The proposed algorithm enjoys linear convergence to an global optimal point. Using this algorithm we can reduce the dimensional of data models. We applied our algorithm on the synthetic data produced from the MNIST data set. The algorithm was implemented for two cases, one with fixed number of observations and different number of images. Second with different number of observations and a single image and the error were plotted in each case. From this show that with certain initialization, the proposed non-convex optimization alternative minimization algorithm enjoys linear convergence to global optima and recovers low rank plus sparse matrices under standard conditions.

Acknowledgments

The authors would like to thank sincerely the anonymous reviewers for sharing creative ideas helping and instructive comments.

References

[1] Y.Lu, Z. Lai, Y. Xu, X. Li, D. Zhang and C. Yuan, “non negative discriminant matrix factorization”, IEEE trans. Circuits sysyt. Video tecnol., vol 27, no. 7 pp, 1392-1405, jul 2017
[2] J. P. Cunningham and Z. Ghahramani, “linear dimensional reduction: survey, insights, and generalizations” J.mach.learn.res.,vol 16,pp.2559-2900,2015
[3] Candès, E.J.,Li, X.,Ma,Y.,Wright, J.: “Robust principal component analysis?” J. ACM 58, 1–37 (2011)
[4] Jianhui Chen, Jiayu Zhou, and Jieping Ye. “Integrating low-rank and group-sparse structures for robust multi-task learning.” In KDD, pages 42–50. ACM, 2011
[5] Karim Lounici, Massimiliano Pontil, Alexandre B Tsybakov, and Sara Van De Geer. ”Taking advantage of sparsity in multi-task learning”. arXiv preprint arXiv:0903.1468, 2009.
[6]“Low-rank matrix recovery and completion via convex optimization”.http://perception.csl.illinois.edu/matrix rank/sample_code.html3. PROPACK–Software for large and sparse SVD calculations. http://sun.stanford.edu/rmunk/PROPACK/.
[7] Candès, E.J., Recht, B.:” Exact matrix completion via convex optimization.” Found. Comput. Math. 9, 717–772 (2009)
[8] Zhouchen Lin, Minming Chen, and Yi Ma. “The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices.” arXiv preprint arXiv:1009.5055, 2010
[9] Qi Yan, Jieping Ye, and Xiaotong Shen. “Simultaneous pursuit of sparseness and rank structures for matrix decomposition”. Journal of Machine Learning Research, 16:47–75, 2015
[10] Anastasios Kyrillidis and Volkan Cevher. “Matrix decomposition and low rank representation” IEEE 2012
[11] alps: “Accelerated low rank and sparse matrix reconstruction”. In Statistical Signal Processing Workshop(SSP), 2012 IEEE, pages 185–188. IEEE,2012.
[12] Lewis, A.S., Malick, J.: “Alternating projections on manifolds”. Math. Oper. Res. 33, 216–234 (2008
[13] Simonsson, L., Eldén, L.: Grassmann ”algorithms for low rank approximation of matrices with missing values”. BIT Numer. Math. 50, 173–191 (2010
[14] Xiao-Tong Yuan and Tong Zhang.”Truncated power method for sparse eigenvalue problems. Journal of Machine Learning” Research, 14(1):899–925, April 2013. ISSN 1532-4435
[15] SJiashi Feng, Huan Xu, and Shuicheng Yan. Online “robust pca via stochastic optimization.” In NIPS, pages 404–412, 2013
[16] Andrew E Waters, Aswin C Sankaranarayanan, and Richard Baraniuk. Sparcs: “Recovering low rank and sparse matrices from compressive measurements” .In NIPS, pages 1089–1097, 2011.

[17] Keshavan, R.H., Montanari, A., Oh, S.: "Matrix completion from a few entries". IEEE Trans. Inf. Theory 56, 2980–2998 (2010).
[18] Min, K., Zhang, Z., Wright, J., Ma, Y.:" Decomposing background topics from keywords by principal component pursuit." In: Proceedings of the 19th ACM international conference on Information and knowledge management, pp. 269–278 (2010).
[19] J. Zhou, J. Chen, and J. Ye. MALSAR: "Multi-task Learning via Structure" Al Regularization. Arizona State University, 2011. URL http://www.public. asu.edu/~jye02/Software/MALSAR.
[20] J. D. Gross, `"Recovering low-rank matrices from few coefficients in any basis," IEEE Trans. Inf. Theory, vol. 57, no. 3,pp.15481566,mar 2011
[21] Venkat Chandrasekaran, Sujay Sanghavi, Pablo A Parrilo, and Alan S Willsky. "Rank-sparsity incoherence for matrix decomposition". SIAM Journal on Optimization, 21(2):572–596, 2011
[22] statiscial modeling foreground object detection."http://perception.i2r.astar.edu.sg/bk_model/bk_index.html
[23] Qi Yan, Jieping Ye, and Xiaotong Shen. "Simultaneous pursuit of sparseness and rank structures for matrix decomposition. Journal of Machine Learning Research", 16:47–75, 2015.
[24] J. Xue, Y. Zhao, W. Liao, and J. C.-W. Chan, "Total variation and rank constraint RPCA for background subtraction," IEEE Access, vol. 6,pp. 4995549966, 2018.
[25] Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi."Low-rank matrix completion using alternating minimization." In STOC, pages 665–674, 2013
[26] Moritz Hardt. "Understanding alternating minimization for matrix completion." In FOCS, pages 651–660. IEEE, 2014
[27] Huang, P.-S., Chen, S. D., Smaragdis, P., Hasegawa-Johnson, M.: "Singing-voice separation from monaural recordings using robust principal component analysis." In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 57–60 (2012)
[28] Wen, Z.,Yin,W., Zhang,Y.: "Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm." Math. Prog. Comput. 4, 333–361 (2012)
[29] Chandrasekaran, V., Sanghavi, S., Parrilo, P.A., Willsky, A.S.: "Rank-sparsity incoherence for matrix decomposition" SIAM J. Optim. 21, 572–596 (2011)
[30] 13. Eckart, C., Young, G.: "The approximation of one matrix by another of lower rank." Psychometrika 1, 211–218 (2000)

First Author Amina Shiroz received B Tech degree in Electronics and Communication from Cochin University of science and technology in the year 2018. Now she is perusing M Tech under A P J Abdul Kalam Technological University with specialization in signal processing. Her research interest includes machine learning, computer vision, signal processing etc.

Second Author Aparna P R received B Tech degree in Electronics and Communication Engineering from the University of Kerala in the year of 2009. She secured First Rank in M Tech degree from Cochin University of Science and Technology with specialization in VLSI and Embedded system in the year of 2011. She started her carrier as Assistant professor in 2011. She is currently working as Assistant professor in the Department of Electronics and Communication engineering, Sree Chitra Thirunal College of Engineering, Trivandrum. Her research interest includes image processing, signal processing, VLSI, embedded systems etc.