# Robust Principal Component Analysis

SEAS (Ahmadabad University )
Nishi Shah
Enroll : 1401099

***Abstract*—Principle Component Analysis(PCA) is widely used for statistical techniques. But PCA is very sensitive to outliers as it breaks down with one corrupted data point. Robust Principle Component Analysis(RPCA) is modified version of PCA. It works well with grossly corrupted observations by decomposing large data matrix into low-rank and sparse matrix. RPCA has many real life important applications particularly when the data under study can naturally be modeled as a low-rank plus a sparse contribution. This paper suggests that under some suitable assumptions, it is possible to recover both the low-rank and the sparse components exactly by solving a very convenient convex program called Principal Component Pursuit(PCP).**

**Keywords: Principal components, outliers, nuclear-norm minimization, $l_1$-norm minimization, duality, low-rank matrices, sparsity, video surveillance.**

## I. Introduction

Suppose a large data matrix M is given. Which maybe decomposed as:

$$M = L_0 + S_0$$

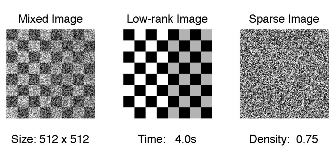


Fig. 1: Decomposition of a large data matrix.

As shown in the figure above $L_0$ has low rank and $S_0$ is sparse.We do not know the low-dimensional column and row space of $L_0$, not even their dimension. Similarly, we do not know the locations of the nonzero entries of $S_0$, not even how many there are. In this paper we aim to recover the low-rank matrix $L_0$ and $S_0$ from M.

## II. Application

Applications in which data can be decomposed in low-rank and sparse matrix, some of these are given below:

- Video surveillance :- Suppose we want to identify some activity happening in the video. The video is sequence of thousands or even ten thousands frames. So, the low-rank matrix gives the stationary background and the sparse matrix captures the moving objects, that's how we can identify the activity in a video surveillance.
- Face recolonization :- Knowing the fact that images of a convex, lambertian surface under varying illuminations span a low dimensional subspace, thus low dimensional models are mostly effective for imagery data especially images of human face. But it becomes crucial in the cases like self-shadowing image, face with speculators or saturation in brightness. This problem can also be solved by PCP, by taking $\lambda = \frac{1}{n_1}$.
- Latent semantic indexing. :- In web search engines, it is often needed to analyze a large amount of documents. That is where Latent Semantic Indexing(LSI) is used. In which common data of the documents are decomposed as low-rank matrix($L_0$)and a few key words are decomposed as sparse matrix($S_0$).
- Ranking and collaborative filtering. :- Now a days importance of e-commerce websites, advertising etc are increasing so as the competition. Thus, keeping record of user's interest is necessary. To complete the incomplete ranking of users is the problem of completing low-rank matrix and also simultaneously correct the error which can be almost solved by Robust PCA.

## III. About Robust PCA

Theoretically, we can recover exact $L_0$ and $S_0$ by Principle Component Pursuit (PCP)which estimates

$$minimize \ \ ||L||_* + \lambda||S||$$

$$Subject \ \ to \ \ L + S = M$$

To avoid situations like, sparse matrix($S_0$) can be interpreted

$$\boldsymbol{M} = \boldsymbol{e}_1\boldsymbol{e}_n^* = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

Fig. 2: low-rank or sparse matrix?.

as low-rank matrix($L_0$) or the other way round, we will assume that sparsity pattern of the sparse component is selected uniformly at random.

### A. *Theorem*(1.1)

Theorem 1.1 shows that incoherent low-rank matrices can be recovered from non-vanishing fractions of gross errors in polynomial time. In above equation to obtain the correct answer we will have to choose the correct value of $\lambda$. Which is stated in Theorem(1.1) in the paper. In which $\lambda$ is taken as $1/\sqrt{(n_{(1)})}$ which is universal and doesn't dependent on $L_0$ and $S_0$. Thus, the equation will be as given below:

$$minimize||L||_* + \frac{1}{\sqrt{(n_{(1)})}}||S||_1$$

where,
$(n_{(1)}) = max(n_{(1)}, n_{(2)})$
This works for the larger value of rank, order of $n/(logn^2)$. When $\mu$ is not too large.The only assumption in this could have been location of nonzero values in sparse matrix. Thus, for deriving an equation at a random set zero is set which will give $S_0$.

These conditions are phrased in terms of two quantities. The first is the maximum ratio between the $l_{\inf}$ norm and the operator norm, restricted to the subspace generated by matrices whose row or column spaces agree with those of $L_0$. The second is the maximum ratio between the operator norm and the $l_{\inf}$ norm, restricted to the subspace of matrices that vanish off the support of $S_0$. when the product of these quantities are small, then recovery is exact for a certain interval of the regularization parameter. Suppose $n_1 = n_2 = n$ for simplicity and $\mu_0$ is the smallest quantity satisfies the above equation then, recovery occurs whenever

$$\max_i(i : [S_0]_i j) \neq 0X\sqrt{\mu_0 r/n} < 1/12$$

### *B. Theorem(1.2)*

We assume that $\Omega_o bs$ is uniformly distributed among all sets of cardinality m obeying $m = 0.1n_2$. Suppose for simplicity, that each observed entry is corrupted with probability $\tau$ independently of the others. Then, there is a numerical constant c such that with probability at least $1 - cn^-10$,PCP with $\lambda = 1/\sqrt{0.1n}$ is exact, that is, $L = L_0$, provided that,

$$rank(L_0) \leq \rho rn\mu - 1(logn) - 2and\tau \leq \tau_s$$

. In this equation, $\rho r$ and $\tau_s$ are positive numerical constants. For general, $n_1 X n_2$ rectangular matrices, PCP with $\lambda = 1/sqrt(0.1n_1)$ succeeds from $m = 0.1n_1n_2$ corrupted entries with probability at least $1 - cn_1{}^-10$, provided that $rank(L_0) \leq \rho rn(2)\mu - 1(logn(1)) - 2$.

If all the entries are available, that is, $m = n_1n_2$, then it turns out to be Theorem 1.1. If $\tau = 0$, we have a pure matrix completion problem from about a fraction of the total number of entries and the theorem guarantees perfect recovery as long as r obeys, which, for large values of r, matches the strongest results available.

## IV. Dual Certificates

We introduce a simple condition for the pair $(L_0, S_0)$ to be the unique solution to PCP. These conditions, given in the lemma in the paper, are stated in terms of a dual vector, the existence of which certifies optimality.

Lemma : Assume that $||P_\Omega P_T|| < 1$.With the standard notations, $(L_0, S_0)$ is the unique solution if there is a pair (W,F) obeying

$$UV^* + W = \lambda(sgn(S_0) + F)$$

, with $p_T W = 0, ||W|| < 1, p_\Omega F = 0and||F||_{\inf} < 1$.

## V. Algorithms

The theorem above shows that incoherent low-rank matrices can be recovered from non-vanishing fractions of gross errors in polynomial time. For small problem sizes, PCP can be performed sing off-the-shelf tools such as interior pint methods. Despite the superior converge rates, interior point methods are typically limited to small problems, say $n < 100$, due to $O^{n6}$ complexity of computing a step direction.

### *A. Principal component pursuit by alternating directions*

initialize : $S_0 = Y_0 = 0, \mu > 0$
while not converged do
compute $L_{h+1} = D_{1/\mu}(M - S_h + \mu^{-1}Y_h)$;
compute $S_{k+1} = S_{\lambda/\mu}(M - L_{h+1} + \mu^{-1}Y_h)$;
compute $Y_{h+1} = Y_h + \mu(M - L_{h+1} - S_{h+1})$;
end while
output : L,S.

## VI. Conclusion

The paper gives us the conclusion that universal value of $\lambda$ which is $\lambda = 1/\sqrt{n}$, works with high probability for recovering any low-rank, incoherent matrix. In a paper by Chandrasekaran , the parameter $\lambda$ is data-dependent and may have to be selected by solving a number of convex programs. The difference in the results is because of the different different assumptions about the origin of large data matrix M. The main advantage in the paper is that $\lambda$ has a fixed universal value which doesn't change according to $L_0$ or $S_0$. Which has been proven very useful in practical world where it is difficult to manage a large data matrix.

## VII. References


- Robust Principal Component Analysis (Emmanuel J. Candes and Xiaodong LI, Stanford University)
- CAND'ES, E. J., AND RECHT, $B$.2009. Exact matrix completion via convex optimzation. Found. Comput. $Math$.9,717–772.CAND'ES, E. J., ROMBERG, J.,
- https://www.youtube.com/watch?v=$DK8RTamIoB8$
- CHANDRASEKARAN, V., SANGHAVI, S., PARRILO, P., AND WILLSKY, A. 2009. Rank-sparsity incoherence for matrix decomposition. Siam J. Optim., to appear http://arxiv.org/abs/0906.2220.