Epitome of the paper on the Robust Principal Component Analysis

Deep C. Patel (1401010),
School of Engineering and Applied Science, Ahmedabad University

***Abstract*—This paper is the epitome of the Paper titled "Robust PCA?" written by E. Candes et al. [2011]. This paper will introduce the problem dealt by Robust PCA. Then the approach taken to solve the problem using Robust PCA will be shown. Then the assumptions made by the authors for the Robust PCA to work perfectly would be put forward. Then the link between the least rank matrix recovery problem and the matrix completion problem would be shown. And finally algorithm put foward by authors to solve this problem will be shown. The method finds its applications in computer vision, video surveillance, repairing and digitizing vintage films etc.**

***Keywords*—Robust PCA, PCA, Convex Optimization, Matrix Completion, Low rank matrix recovery**

I. Introduction

In this era of computing, availability of high speed internet and rapid development in Artificial Intelligence have triggered growth of daily data generation by the people and sensors. A huge amount of data is generated daily consisting of high definition imagery, user behavior, surrounding parameters etc. And these data requires analysis to be performed on them to draw out some interesting inference and use it in future work. But looking at this massive amount, the analysis faces an obstacle of "Curse of Dimensionality" by the high dimensional structure of the data. So, one of the solution to this problem is to project the data on the lower dimension subspace. In linear algebra terms it means that we want the low rank data matrix. The most common technique to do so is the Classical PCA. But looking at the huge amount of data we can safely assume that it would be grossly corrupted and filled with outlying observations. But even a fraction of error may corrupt PCA's estimate of low rank matrix . So we want PCA to be robust. The original paper discusses the method to make the PCA robust. It aims to decompose large data matrix M into low rank matrix L_0 and sparse error matrix S_0, with errors in M being of arbitrary magnitude. The model looks like

$$M = L_0 + S_0 \tag{1}$$

II. Approach to Decompose M

The problem discussed in above section is clearly as follows,

Problem. Given the observed data matrix M = $L_0 + S_0$, where L_0 and S_0 are unknown, but we know that L_0 is low rank matrix and S_0 is sparse matrix, recover L_0.

The straightforward solution to this problem is stated below,

$$\min_{L_0,S_0} \quad rank(L_0) + \gamma\|S\|_0 \quad Subj\ M = L_0 + S_0 \tag{2}$$

But, the (2) is highly non convex optimization and we don't know any efficient solution to it. So, now to get relatively easy solution to this problem, we have to relax (1) by replacing $rank()$ with $\| \, \|_n$ or nuclear norm and $\| \, \|_0$ or l^0-norm with $\| \, \|_1$ or l^1-norm as suggested by the authors of ([1], [2]).

So the new solution is as follows,

$$\min_{L_0,S_0} \quad \|L_0\|_n + \lambda\|S\|_1 \quad Subj\ M = L_0 + S_0 \tag{3}$$

As, stated by the authors of ([1], [2]), solving this convex optimization problem perfectly recovers the low rank matrix.

III. Conditions for the Sensible Separation

Not every time we will get a sensible separation of L_0 and S_0 from the data matrix M. To do so, our L_0 and S_0 should follow some conditions. These conditions are clearly stated in the [1] and are as follows,

1) *Incoherence condition for the low rank matrix* L_0

For proper separation our low rank matrix L_0 should not be sparse and for that it should meet the incoherence conditions. Let's say that SVD of $L_0 \,\epsilon\, R^{n_1 \times n_2}$ is,

$$L_0 = U\Sigma V^T \tag{4}$$

Let r be the $rank(L_0)$ then following are the incoherence conditions,

$$\max_i \|U^* e_i\|^2 \leq \frac{\mu r}{n_1}, \ \ \max_i \|V^* e_i\|^2 \leq \frac{\mu r}{n_2}, \tag{5}$$

$$\|UV^*\|_\infty \leq \sqrt{\frac{\mu r}{n_1 n_2}} \tag{6}$$

As stated by authors in [1], for small μ, the incoherence condition implies on low rank matrix L_0 that it would be sparse.

2) *Sparsity pattern for* S_0 *following Bernoulli distribution*

An issue arises if the sparse matrix S_0 has low rank. So, to resolve that issue we assume that our sparse matrix S_0 is drawn from Bernoulli sign and support model with parameter ρ_s. It means that entries of $sign(S_0)$ is independently distributed and each takes value 0 with probability $1-\rho_s$ and ± 1 with probability $\rho_s/2$ each. In this model the non zero entries of S_0 can have arbitrary magnitude.

IV. Supporting Theorem

The authors in [1] have stated a theorem which establishes the robustness of our method for recovering the low rank matrix L_0 stated as (3). We can unhesitatingly say that it is the base theorem behind this paper [1]. The theorem is stated as follows,

Theorem 1

Suppose that, $n_{(1)} = max(n_1, n_2)$ and $n_{(2)} = min(n_1, n_2)$, and

- L_0 is $n_1 \times n_2$ of $rank(L_0) \leq \rho_r n_{(2)} \mu^{-1} (\log n_{(1)})^{-2}$
- S_0 is $n_1 \times n_2$, random sparsity pattern of cardinality $m \leq \rho_s n_1 n_2$

Then with probability atleast $1 - cn_{(1)}^{-10}$ for some constant c, Principal Component Pursuit (PCP) with $\lambda = \frac{1}{\sqrt{n_{(1)}}}$ is exact, $i.e$,

$$\hat{L} = L_0,\ \hat{S} = S_0 \tag{7}$$

Here ρ_r and ρ_s are positive numerical constants.

V. Matrix Completion from Grossly Corrupted Data

There are some situations where the matrix entries are missing as well. To recover the low rank matrix we have to solve the following convex optimization problem

$$\begin{aligned} &\min_{L_0, S_0} \quad \|L_0\|_n + \lambda \|S\|_1, \\ &Subj\ M_{ij} = L_{ij} + S_{ij}, (i,j)\,\epsilon\,\Omega_{obs} \end{aligned} \tag{8}$$

Here Ω_{obs} is location of observed entries.

The authors in [1] have also stated a theorem which implies that perfect recovery of the low rank matrix from matrix with missing and grossly corrupted entries is possible using convex optimization. The theorem is as follows,

Theorem 2

Suppose that, $n_{(1)} = max(n_1, n_2)$ and $n_{(2)} = min(n_1, n_2)$, and

- L_0 is $n_1 \times n_2$ of $rank(L_0) \leq \rho_r n_{(2)} \mu^{-1} (\log n_{(1)})^{-2}$
- Ω_{obs} is random set of cardinality $m = 0.1 n_1 n_2$
- Each observed entry is independently corrupted with probability $\tau \leq \tau_s$

Then with probability atleast $1 - cn_{(1)}^{-10}$ for some constant c, Principal Component Pursuit (PCP) with $\lambda = \frac{1}{\sqrt{0.1 n_{(1)}}}$ is exact, $i.e$,

$$\hat{L} = L_0 \tag{9}$$

Here ρ_r and τ_s are positive numerical constants.

VI. Algorithm for Robust PCA

The authors of [1] have stated that the guarantee provided by the Theorem 1 is independent of particular algorithm used to solve Principal Component Pursuit. The authors have used the Augmented Langrage Multipliers algorithm for separation of the low rank and sparse error matrices from the data matrix M, which they claim as most suitable for practical use. The algorithm is stated below,

Algorithm 1 Augmented Lagrange Multiplier

```
procedure ALM
    initialize : (S_0 = Y_0 = 0, μ > 0);
    while not converged do
        compute L_{k+1} = D_{1/μ}(M − S_k + μ^{-1} Y_k);
        compute S_{k+1} = S_{λ/μ}(M − L_{k+1} + μ^{-1} Y_k);
        compute Y_{k+1} = Y_k + μ(M − L_{k+1} − S_{k+1});
    end while
return L,S
```

The algorithm repeatedly finds $(L_k,\ S_k)$ pair from threshold operator $D_{1/\mu}$ and shrinkage operator $S_{\lambda/\mu}$ respectively. Now it updates Langrage multiplier matrix Y_{k+1} by using,

$$Y_{k+1} = Y_k + \mu(M - L_{k+1} - S_{k+1}); \tag{10}$$

We terminate the algorithm when $\|M - L - S\|_F \leq \delta \|M\|_F$, with $\delta = 10^{-7}$.

VII. Conclusion

In this epitome of the "Robust PCA?" paper [1], we saw how author posed the problem of low rank matrix recovery and solved it using convex optimization algorithms. We also saw the theorem which guarantees the recovery of low rank matrix from the error matrix. And finally we saw the algorithm put forward by authors of [1] to separate the L_0 and S_0 from M.

References

[1] E. J. Cands, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?", Journal of the ACM, vol. 58, no. 3, pp. 1-37, Jan. 2011.

[2] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices by convex optimization", Advances in neural information processing systems, pp. 2080-2088, Apr. 2009.