Robust Principal Component Analysis Project Report

Nisarg Tike
Roll No: 1401070

***Abstract*—Principal Component Analysis (PCA) is the most widely used statistical technique for dimensionality reduction. One major drawback of this method is that it is very sensitive to outliers i.e. one single corrupted data point in the data matrix will result into a large deviation from the expected output. In computer vision applications, outliers typically occur within a sample due to noise or sensor failures. This report is a summary of the paper "Robust PCA". The paper proposes a method which under suitable assumptions, makes possible the decomposition of a corrupted matrix into a low rank component and a sparse component.**

Keywords: Principal component, low rank matrix, sparse matrix

I. Introduction

Robust principal component analysis has received much attention in recent studies for its ability to recover the low rank model from sparse noise. Such sparse structure of noise is ubiquitous in many real time applications such as face recognition, video surveillance and system identification. A key application where this occurs is in video analysis where video frames are considered as data matrix and the objective is to separate background from moving foreground objects. The paper suggests that under suitable assumptions, it is possible to decompose a grossly corrupted data matrix M into a low rank component L_0 and sparse component S_0.

II. Assumptions

The objective is to decompose matrix M into a low rank component L_0 and a sparse component S_0 such that
M = L_0 + S_0. Certain assumptions are made for the proposed solution. The low rank matrix L_0 is not sparse and the sparse matrix S_0 is not low rank. Writing the singular value decomposition of L_0,

$$L_0 = U\Sigma V^T$$

Let r be the rank of matrix L_0. Then, the incoherence condition with parameter μ states that,

$$\max_i ||U^T e_i||^2 \leq \frac{\mu r}{n_1}$$
$$\max_i ||V^T e_i||^2 \leq \frac{\mu r}{n_2}$$
$$||UV^T||_\infty \leq \sqrt{\frac{\mu r}{n_1 n_2}}$$

Here, $||M||_\infty = \max_{i,j} |M_{ij}|$ is the l_∞ norm of M seen as a long vector. To avoid the sparse matrix S_0 from being low rank, the sparsity pattern of the sparse component is selected uniformly at random.

III. Important Theorems

Theorem 1.1. *Suppose L_0 is $n\times n$, obeys above assumptions. Fix any $n\times n$ matrix Σ of signs. Suppose that the support set Ω of S_0 is uniformly distributed among all sets of cardinality m, and that sgn($[S_0]_{ij}$) = Σ_{ij} for all (i, j) $\in \Omega$. Then, there is a numerical constant c such that with probability at least $1 - cn^{-10}$ (over the choice of support of S_0), Principal Component Pursuit with $\lambda = \frac{1}{\sqrt{n}}$ is exact, that is, $\hat{L} = L_0$ and $\hat{S} = S_0$, provided that*

$$\text{rank}(L_0) \leq \rho_r n \mu^{-1} (\log n)^{-2}$$
$$m \leq \rho_s n^2$$

In this equation, ρ_r and ρ_s are positive numerical constants. In the general rectangular case, where L_0 is $n_1 \times n_2$, PCP with $\lambda = \frac{1}{\sqrt{n_{(1)}}}$ succeeds with probability at least $1 - cn_{(1)}^{-10}$, provided that $rank(L_0) \leq \rho_r n_2 \mu^{-1} (\log n_{(1)})^{-2}$ and $m \leq \rho_s n_1 n_2$.

In the paper, following is proposed for recovering L_0.

$$\begin{aligned} &\text{minimize} && ||L||_* + \lambda ||S||_1 \\ &\text{subject to} && P_{\Omega_{obs}}(L + S) = Y \end{aligned}$$

Theorem 1.2. *Suppose L_0 is $n\times n$, obeys the conditions in the assumptions section, and that Ω_{obs} 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 τ independently of the others. Then, there is a numerical constant c such that with probability at least $1 - cn^{-10}$, Principal Component Pursuit with λ = $\frac{1}{\sqrt{0.1n}}$ is exact, that is, $\hat{L}$ = L_0, provided that*

$$\text{rank}(L_0) \leq \rho_r n \mu^{-1} (\log n)^{-2}$$
$$\tau \leq \tau_s$$

In this equation, ρ_r and τ_s are positive numerical constants. For general $n_1 \times n_2$ rectangular matrices, PCP with $\lambda = \frac{1}{\sqrt{0.1 n_{(1)}}}$ succeeds from m = $0.1n_1n_2$ corrupted entries with probability at least $1 - cn_{(1)}^{-10}$, provided that $rank(L_0) \leq \rho_r n_2 \mu^{-1} (\log n_{(1)})^{-2}$.

IV. Algorithm

A. Lagrange function

The convex PCP algorithm can be solved using an augmented Lagrange multiplier (ALM). The ALM method operates on the *augmented Lagrangian*

$$l(L,S,Y) = ||L||_{*} + \lambda||S||_1 + \langle Y, M-L-S\rangle + \frac{\mu}{2}||M-L-S||_F^2$$

Let S_τ denote the shrinkage operator.

$$S_\tau[x] = sgn(x)\max(|x|-\tau, 0)$$

$$\arg\min_S l(L,S,Y) = S_{\lambda/\mu}(M - L + \mu^{-1}Y)$$

Let $D_\tau(X)$ denote the singular value thresholding operator.

$$D_\tau(X) = US_\tau(\Sigma)V^T$$

$$X = U\Sigma V^T$$

$$\arg\min_L l(L,S,Y) = D_{1/\mu}(M - S + \mu^{-1}Y)$$

B. Alternating Directions

Following PCP by Alternating directions algorithm is implemented.

Algorithm 1 PCP by Alternating Directions

```
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
output: L,S
```

The dominant cost of Algorithm 1 for each iteration is computing L_{k+1} via singular value thresholding. The most important implementation details for this algorithm are the choice of μ and the stopping criterion. In the paper, the authors chose μ as $n_1n_2/4|M|||_1$ and the termination criterion was set as $||M-L-S||_F \leq \delta||M||_F$, with $\delta = 10^{-7}$.

V. Observations

The above algorithm was implemented in matlab and was tested on a corrupted image of moon. The stopping criterion was $||M-L-S||_F \leq 10^{-5}$. λ was set to be 0.02. The default max interation for the algorithm was set as 1000. The program ran for 234 iterations and successfully converged. Following results were obtained.

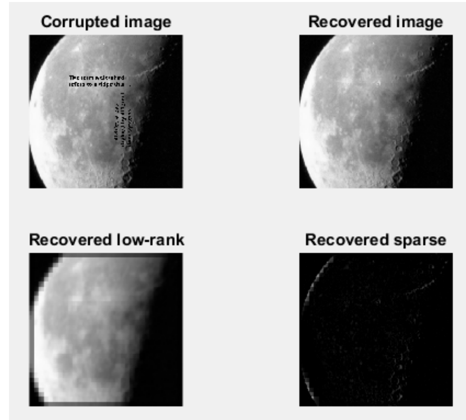

Figure 1.1 Recovering Low-rank component and sparse component from corrupted image.

Following are the details of the result.

Size of M	961 $\times$ 256
λ	0.02
μ	1.0
Rank(L)	28
nnz(S)	230467
Error	0.000010
Iterations	234
Elapsed time	10.756821 s

VI. Conclusion

From the observations above, it is indeed possible to decompose a corrupted matrix into a low-rank component and a sparse component with a very high accuracy given that the assumptions are satisfied.

References

Robust Principal Component Analysis, by Emmanuel J. Candes, Xiaodong Li, Yi Ma and John Wright http://statweb.stanford.edu/~candes/papers/RobustPCA.pdf

https://github.com/dlaptev/RobustPCA

https://en.wikipedia.org/wiki/Principal_component_analysis

https://en.wikipedia.org/wiki/Matrix_norm

http://file.scirp.org/pdf/JSIP_2013101711003963.pdf