Robust Principal Component Analysis

Report by Madhav Chavda
Roll No: 1401006

***Abstract*—Singular Value Decomposition and Principal Component analysis (PCA) are widely used techniques for dimension reduction and having the most part of the data. But this method is sensitive when a corrupted value comes into picture and especially in the outliers. In this period of time where the world has a huge amount of data, and also the corrupted ones it is not worth using this algorithm. This paper discusses about the mechanism of Robust Principal Component Analysis which is advanced version of PCA. It talks about exact recovery of corrupted low rank matrices via convex optimization. This convex optimization approach is capable enough to recover exact low rank matrices and identifies the corrupted points.**

I. Introduction

In today's world where everyday millions of amount of data are required to be analysed, we need a mechanism to deal with this large amount of data. There is a very chance of data being corrupt too. So, can we hope to recover the corrupted large amount of data? For this we define a mechanism in which we disentangle the data matrix into a low rank matrix (L_0) and a sparse matrix (S_0) i.e. $D = L_0 + S_0$. However, both he matrices are of arbitrary magnitude and dimension. We do not know low dimensional row and column space of L_0. Also the non-zero entries in S_0 and number of those entries are not known. With all this can we hope to recover the low rank and sparse matrix both accurately and efficiently? In many application domain data lies in billions of dimension, thus to handle the file we reduce the dimensionality of data which make them lie in some low dimensional subspace i.e. $M = L_0 + N_0$. We have classical PCA which says

$$\begin{aligned} &minimize \quad && ||M - L|| \\ &subject\ to \quad && rank(L) \leq k \end{aligned}$$

The above approach, classical approach, is efficiently solved SVD when noise N_0 is small independent and identically distributed Gaussian. PCA is most widely used statistical tool for data analysis and dimensionality reduction. But due to its sensitivity to grossly corrupted observations, this is not very accurate. A single corrupted entry in M in the outliers can take L arbitrarily far from true L_0. And in today's modern world the gross errors are much occurring phenomenon.

The paper proposed algorithm can be considered as an idealized version of robust PCA. We aim to recover a low rank matrix L_0 from highly corrupted measurements. This can be modeled as $M = L_0 + S_0$, unlike the small noise term in N_0.

II. Principal Component Pursuit (PCP)

The paper uses method of principal component pursuit instead of classical PCA for some of the above mentioned drawbacks. The PCP exactly recovers L_0 and S_0. The PCP is given as follows,

$$\begin{aligned} &minimize \quad && ||L||_* + \lambda ||S||_1 \\ &subject\ to \quad && L + S = M \end{aligned}$$

Where $M \in R^{n_1 x n_2}$,

$||L||_* := \sum_i \sigma_i(L)$ denote the nuclear norm of low rank matrix L that is the sum of the singular value of L, and let $||S||_1 := \sum_{ij} |S_{ij}|$ denote the l_1 norm of matrix S seen as sum of absolute values of element S.

The PCP theoretically works even when rank of L_0 grows almost linearly in dimension of M. Also when errors in S_0 are up to a constant fraction of entries. Algorithmically this can be solved by efficient and scalable algorithms at a cost not much higher than classical PCA. Empirically, the simulations and experiments suggests that this works under surprisingly broad conditions for many types of real data.

A. *When does separation makes sense*

Suppose the matrix M has 1 in top left corner and zero elsewhere. Then, M can be a low rank matrix and as well a sparse matrix too. In that case we have identifiability issue. Thus, to make problem more meaningful, we need to impose that a low rank component cannot be sparse. In that case the incoherence comes into picture. This is an assumption concerning the singular vectors of low rank component.

$$L_0 = U \sum V^* = \sum_{i=1}^{r} \sigma_i u_i v_i$$

Where r is rank of matrix. U and V are left and right singular vectors respectively. The incoherence condition with μ parameter states,

$$max_i ||U^* e_i||^2 \leq \frac{\mu r}{n_1}, \qquad max_i ||V^* e_i||^2 \leq \frac{\mu r}{n_2} \tag{1}$$

$$||UV^*||_\infty \leq \sqrt{\frac{\mu r}{n_1 n_2}} \tag{2}$$

Another identifiability issue arises if the sparse matrix has low rank. This will occur if, say, all the nonzero entries of S occur in a column or in a few columns. Suppose for instance, that the first column of S_0 is the opposite of that of L_0, and that all the other columns of S_0 vanish. Then in that case we wouldn't be able to recover L_0 and S_0 by any of the method since $M = L_0 + S_0$ will have column space equal to or included in L_0. Thus, sparsity pattern of sparse component is selected uniformly at random.

B. Main Result

Theorem 1.1: Suppose L_0 is nxn, obeys (2)-(3). Fix any $n \times n$ matrix $\sum$ of $signs$. Suppose that the support set Ω of S_0 is uniformly distributed among all sets of cardinality m, and that sgn($[S_0]_{ij}$) = $\sum_{ij}$ 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

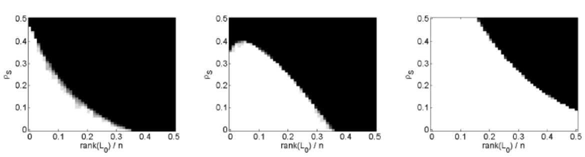
$$rank(L_0) \leq \rho_r n \mu^{-1} (logn)^{-2} \; and \; m \leq \rho_s n^2 \qquad (3)$$

In this equation, ρ_r and ρ_s are positive 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 $rank(L_0) \leq \rho_r n_2 \mu^{-1} (logn_1)^{-2} \; and \; m \leq \rho_s n_1 n_2$. It is the mathematical analysis which reveals the correctness of this value.

Matrices L_0 whose singular vectors—or principal components—are reasonably spread can be recovered with probability nearly one from arbitrary and completely unknown corruption patterns (as long as these are randomly distributed). The only “piece of randomness” in the assumptions concerns the locations of the nonzero entries of S_0; everything else is deterministic. This also has no tuning parameters. The S_0 can be modelled as; take an arbitrary matrix S and set to zero its entries on the random set c this gives S_0.

C. Exact Recovery from Varying Fractions of Error

We study the effect of different fraction of errors in our algorithm. The below given graphs are for understanding the exact recovery for different errors.



(a) Robust PCA, Random Signs (b) Robust PCA, Coherent Signs (c) Matrix Completion

The x-axis is rank(L_0)/n and y-axis shows the fraction of errors. The black pixels are the unsuccessful and the white are the successful ones. This means that for a value of rank(L_0)/n say 0.2, we have corresponding values of fraction of errors say 0.3, which means that at that point of rank(L_0)/n i.e. 0.2, the algorithm can afford 30 percent of errors. If we compare (a), (b) and (c) graphs we come to know that we can afford maximum errors if it's a matrix completion problem than the matrix recovery problem, which has corruption in the entries. The matrix recovery problem with coherent signs are more affordable than the matrix recovery problem with random signs.

D. Algorithms

There are many algorithm through which the concept of PCP and disentanglement of L_0 and S_0 can be done. We discuss here some algorithms which can be used. Iterative Thresholding algorithm is one such algorithm which performs nuclear norm minimization by repeatedly shrinking singular values. This essentially reduces complexity of each iteration. However, for disentanglement of L_o and S_0 requires 10^4 iterations.Thus, the idea of improving convergence with continuous technique comes into picture. Among the continuous techniques, Accelerated Proximal Gradient(APG) algorithm was developed for matrix completion. A similar APG algorithm was proposed by Lia Et Al [2009b] which had convergence rate of O(1/k^2), which makes it 50 times faster than the iterative thresholding algorithm.
However, despite its good convergence, the practical performance of APG highly depend on design of good continuous schemes. For getting good accuracy and convergence across wide range of problem, the paper uses Augmented Lagrange Multiplier (ALM) to solve convex PCP problem. ALM has more accuracy than APG and that to also in fewer iterations. The ALM algorithm also has a property that the rank of the iterates often remains bounded by rank(L_0) throughout the optimization, allowing them to be computed especially efficiently. The algorithm is given as follows:

Algorithm 1 PCP by Alternating Directions

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

III. Conclusion

This article discusses about a technique known as robust principal component analysis which can disentangle the low rank and sparse component exactly and accurately by convex optimization algorithm. The algorithm can do it for matrix completion problem as well as matrix recovery problem with large number of corruption. The article uses Augmented Lagrange Multiplier (ALM) with Principal Component Pursuit (PCP). It can be used in video Surveillance,Face Recognition,Latent Semantic Indexing etc problems.

IV. References

[1]//http://ieeexplore.ieee.org/abstract/document/937541/

[2]http://papers.nips.cc/paper/3704-robust-principal-component-analysis-exact-recovery-of-corrupted-low-rank-matrices-via-convex-optimization

[3]Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices by Convex Optimization. John Wright , Yigang Peng, Yi Ma Visual Computing Group Microsoft Research Asia

[4]Analysis of Robust PCA via Local Incoherence. Huishuai Zhang, Yi Zhou, Yingbin Liang.