Robust Principal Component Analysis

Hemang Jethava (1644002)
MTech(CSE - 2nd sem)
School of Engineering and Applied Science
Ahmedabad University
Email: hemangkumar.j.mtechcs16@ahduni.edu.in

***Abstract*—In the era of Big Data there are many important applications in which the data under study can naturally be modeled as a low-rank plus a sparse contribution.The paper talks about a situation where we have a data matrix, which is a superimposition of a low-rank component and a sparse component. The job of recovering the low-rank and the sparce components seems daunting but it is proven that under the appropriate assumptions, the low-rank and the sparse components can be recovered. This can be achieved by solving a program called Pricipal Component Pursuit. This suggests the chances of well structured approach to robust principal component analysis since our ndings give an insight that the principal components of the data matrix can be recovered even though a certain amount of its entries are corrupted. This can be extended to a situation where the certain entries are missing as well. This approach can be utilised in many applications.**

Keywords: Principal Component Analysis (PCA), Principal Component Pursuit (PCP),Low Rank Matrix, Sparse Matrix, Convex Optimization, low-rank matrices, outliers, sparsity, robustness, nuclear-norm minimization,

I. Introduction

The recent explosion in the enormous amounts of high dimensional data in science, engineering and society presents a challenge and an opportunity in many areas such as image,video and multimedia processing, data analysis, bio informatics and many more. This data routinely lies in thousands or even billions of dimensions.

Robust PCA is the modified version of the widely used method of PCA which is used as statistical tool for data analysis and dimensionality reduction. It aims at obtaining back the low rank and sparse components from grossly corrupted data. The errors in the data always exists in modern applications such as image processing, web data analysis, and bio informatics, where some measurements may be arbitrarily corrupted

A. Motivation

Suppose we are given a large data matrix which can be decomposed as

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

where L_0 has low rank and S_0 is sparse and both the components are of arbitrary magnitude.We do not know the low dimensional column and row space of L_0. Similarly, we have no information about the non-zero entries of S_0.

If we stack all data points as column vectors of a matrix M, then the matrix should have low rank.

$$M = L_0 + N_0, \tag{2}$$

where L_0 has low rank and N_0 is a small perturbation matrix.

$$\begin{aligned} &minimize \quad \parallel M - L \parallel \\ &subjectto \quad rank(L) \le k \end{aligned} \tag{3}$$

This problem can be efficiently solved via singular value decomposition and enjoys a number of optimality properties when the noise N0 is small and independent and identically distributed Gaussian.

There are a wide number of applications which can t the data under the study into low rank and sparse components. Some of those examples are video surveillance, face recognition and latent semantic indexing, ranking and collaborative filtering.

II. Understanding

In today's big data driven world, the data points can always be corrupted and thus the use of PCA would result into the breakdown of the system. So the use of Robust PCA is encouraged to solve this problem.

The method of dimensionality reduction as in principal component analysis requires a collection of points and columns of data matrix. Then we have to find the best low rank approximation of this matrix.

The problem of separation of the data matrix is a challenging task as the number of unknowns to infer L_0 and S_0 are very large in M $\in R^{n_1 \times n_2}$. Even though this problem can be solved by tractable convex optimization, under certain assumptions, the Principal Component Pursuit solving

$$\begin{aligned} &minimize \quad \parallel L \parallel_* + \lambda \parallel S \parallel_1 \\ &subjectto \quad L + S = M \end{aligned} \tag{4}$$

can recover the two components i.e. low-rank L_0 and sparse S0. This would work even if the rank of L_0 would increase in the dimension of the matrix and the errors in the sparse component S0 are around a certain percentage of all the entries. The above problem can be solved by certain algorithms with the cost not much higher than the classical PCA.

In certain cases the matrix is both low-rank and sparse. Suppose the matrix M is equal to $e_1 e_2^*$. This matrix has one as the entry in the top-left corner and zeros everywhere else. Then the matrix M turns out to be both low-rank and sparse. In this case the decomposition of the data matrix into the low-rank (L_0) and sparse (S_0) components cannot be done accurately and efficiently. The singular value decomposition of the low-rank component L_0 $\epsilon \mathrm{R}_{n1n2}$ can be written as

$$L_0 = U \in V^* = \Sigma_{i=1}^r \sigma_i u_i v_i^*,$$

To get a proper outcome we will have the following assumptions.

- Low-rank matrix will not be sparse.
- Sparse matrix will not be low-ranked.

Suppose all the nonzero entries of S occur in a column. For instance the first column of S_0 is the opposite of that of L_0 and other columns of S_0 vanish then we would not be able to recover L_0 and S_0 since M = L_0 + S_0 will have a column space equal to or included in that of L_0.

A. Theorem 1.1

Suppose L_0 is n n. Fix any n 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($[S_0]_{ij}$) $=\Sigma_{ij} forall(i,j) \in$ Ω Then, there is a numerical constant c such that with probability at least $1-$ cn_{10}, Principal Component Pursuit with $\lambda = \frac{1}{\sqrt{n}}$ is exact, i.e. ˆ L = L_0 and ˜S = S_0, provided that

$$rank(L_0) \leq p_r r n \mu_{-1}(logn)^{-2} and m \leq p_s n^2 (1.1)$$

The above theorem states that if L_0 obeys our assumption that it is a low-rank matrix and not sparse such that the value of λ = $1\frac{1}{\sqrt{n}}$ under the equation 1.1, then the probability of recovery can be maximized, under the assumtion of the theorem minimizing

$$L_0 + \frac{1}{\sqrt{n_1}} S_1, n_1 = max(n_1, n_2),$$

always returns the correct answer.

We can recover the low-rank component in the case of some of its entries being corrupted but here is a possibility that some of its entries may be completely missing.

B. theorem 1.2

Let us imagine that we have only a few entries of L_0 +S_0. Now, we wish to recover L_0 but we can see only some of those entries out of which some of them are corrupted and we don't know which one. This is a signicant extension to matrix completion problem.

$$\begin{aligned} minimize \quad & \| L \|_* \lambda + \| S \|_1 \\ subjectto \quad & P_{\Omega_{obs}}(L+S) = Y \end{aligned} \tag{5}$$

Suppose the matrix L_0 obeys the assumptions and Ω is uniformly distributed among all the sets of cardinality m = $0.1n^2$. Suppose the probability of entry being corrupted is T then the probability can be maximized 1-cn^{-10} with the value of $\lambda = \frac{1}{\sqrt{0.1 n_1}}$, provided that

$$rank(L_0) \leq p_r n \mu^{-1} (logn)^{-2} and T \leq T_s (1.2) \tag{6}$$

Thus the recovery from corrupted entries or incomplete entries is possible by convex optimization.

III. Algorithm

The convex PCP problem has been solved using the augmented Lagrange multiplier(ALM) algorithm which belongs to a class of algorithms for solving constrained optimization problems.

The ALM method functions on the augmented Lagrangian l(L,S,Y)=
$\| L \|_* + \lambda \| S \|_1 + (Y, M-L-S) + \mu/2 \| M-L-S \|_F^2$

The Lagrange multiplier algorithm would solve the PCP by iteratively setting (L_k,S_k) = $argmin_{L,S}$ l(L,S,Y_k). It would then update the Lagrange multiplier matrix using the equation Y_{k+1}= Y_k + μ(M - L_k - S_k).

We can avoid to solve the sequence of convex programs in our problem of decomposing the data matrix into low-rank and sparse components.

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

The singular value thresholding operator D(X) can be minimized as

$argmin_L$l(L,S,Y) = $D_{\frac{1}{\mu}}$(M - S + μ^{-1}Y)

Thus we minimize l w.r.t L by xing S and then we minimize s w.r.t S by xing L and then the Lagrange multiplier matrix Y is updated due to the residual M- L-S.

Algorithm 1 Principal Component Pursuit by Alternating Directions [Lin et al. 2009a;Yuan and Yang 2009]

- 1: Initialize:S_0=Y_0=0,μ ¿ 0
- 2: while not converged do
- 3: compute L_{k+1} = $D_{\frac{1}{\mu}}$ (M - S_k + $\mu^{-1} Y_k$);
- 4: compute S_{k+1} = $S_{\frac{\lambda}{\mu}}$ (M_{Lk+1} + $^1 Y_k$);
- 5: compute Y_{k+1} = Y_k + μ(ML_{k+1} - S_{k+1});
- 6: end while
- 7: end while
- 8: Output: L,S.

The choice of μ and the stopping criterion are very important for the implementation of the algorithm. For the above algorithm, the values are chosen as μ = n1n2/4$\| M \|_1$. We will terminate the algorithm in the case of $\| M-L-S \|_F \leq$ $\delta \| M \|_f$, with δ=10^7.

IV. Conclusion

The low-rank components can be obtained from the existing fractions of gross errors in polynomial time. The low computation cost is guaranteed and the efficiency is becoming practical for the real imaging problems. This is mainly due to the rapid development in scalable algorithms for non smooth convex optimization

References

[1] E. Cand'es, E. J., Li, X., Ma, Y., Wright, J. (2011). Robust principal component analysis?. Journal of the ACM (JACM), 58(3), 11. Chicago