Robust Principal Component Analysis

Manini Patel (1644004)
MTech(CSE - 2nd sem), SEAS, AU

***Abstract*—In this 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.Principal Component Analysis(PCA) plays a central role in statistics, engineering and science where it is widely used for data analysis and dimensionality reduction and also for the representation of shape, appearance, and motion.Its one drawback is that it uses least squares estimation technique and hence it is very sensitive to "outliers" which occurs due to noise,missing values,corrupted data etc.. Because of the ubiquity of corrupted data in real-world applications, much research has focused on developing robust algorithms.In this paper approach has been explored and proposed to robustifying PCA in which aim is to recover a low-rank matrix from highly corrupted measurements.Application Robust PCA is Video Surveillance, Face Recognition, Latent Semantic Indexing and Ranking and Collaborative Filtering.**

Keywords: Principal Component Analysis (PCA), Principal Component Pursuit (PCP),Low Rank Matrix, Sparse Matrix, Convex Optimization.

I. INTRODUCTION

Suppose we are given a data matrix M such that M can be decomposed as

$$M = L_0 + S_0$$

where L_0 has low rank and S is sparse and both matrix are of arbitrary magnitude. We do not know the low-dimensional column and row space of L_0 , similarly the non zero entries of S_0 are not known. There are many prior attempts to solve or atleast alleviate the above mentioned problem.

A. *Classical Principal Component Analysis* [1]

To solve the dimensionality and scale issue we must leverage on the fact that such data matrix are intrinsically lower in dimension, thus are indirectly sparse in some sense. Perhaps the simplest assumption is that the data in matrix all lie near some lower dimensional subspace, hence we can stack all the data points as column vector of a matrix M, and this column vector can be represented mathematically,

$$M = L_0 + S_0$$

where L_0 is essentially low rank and N_0 is a small perturbation matrix.Classical Principal Component seeks the best rank-k estimate of L_0 by solving

$$\textit{minimize } \|M - L\|$$

$$\textit{Subject to rank(L) } \leq k$$

$\| \; M \; \|$ denotes the l^2 norm; that is, the largest singular value of M.This problem can be efficiently solved via the singular value decomposition (SVD) and enjoys a number of optimality properties when the noise N_0 is small and independent and identically distributed Gaussian.

Solution given by truncated SVD :

$$\mathrm{L}_0 = U\Sigma V^T = \sum_i \mathrm{u}_i v_i^T \implies L = \sum_i^r \sigma_i u_i v_i^T$$

$$\mathrm{r} = \mathrm{rank}(L_0)$$

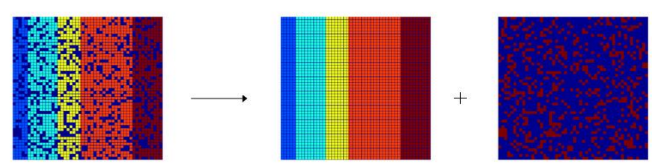
B. *Robust Principal Component Analysis* [1]

Robust Principal Component Analysis is the problem of recovering the low-rank and sparse components. Under suitable assumptions on the rank and incoherence of L_0, and the distribution of the support of S_0, the components can be recovered exactly with high probability, by solving the Principal Component Pursuit (PCP) problem given by

$$\text{minimize } \| \mathrm{L} \|_* + \lambda \| \mathrm{S} \|_1$$
$$\text{subject to } \mathrm{L} + \mathrm{S} = \mathrm{M}$$

nuclear norm: $\|\mathrm{L}\|_* = \sum_i \sigma_i(L)$(sum of singular values of L)

l_1 norm: $\| \mathrm{S} \|_1 = \sum_{ij} |S_{ij}|$ (sum of absolute values)



Matrix of corrupted observations   *Underlying low-rank matrix*   *Sparse error matrix*

II. PRIOR ASSUMPTIONS

While separating the data matrix M, what if M has both sparse and low-rank component has given an example of a matrix M which is equal to $e_1 e_1^*$ which has 1 on top left left corner and remaining values as 0. Another issue arises certain direction in original data is poorly represented. In such cases, M is both low-rank and sparse. Hence we impose an incoherence condition which asserts that for small values of

μ, S_0 is not sparse. The other condition is that the sparse matrix should not have low-rank i.e. we assume that the sparsity pattern of the sparse component is selected uniformly at random.[1]

$$max_i \parallel U * e_i \parallel^2 \leq \frac{\mu r}{n_1} and max_i \parallel V * e_i \parallel^2 \leq \frac{\mu r}{n_2}$$

The incoherence parameter μ, which measures how column spaces and row spaces of L are aligned with previous basis and between themselves. In above discussed situations, value of μ is higher. But for smaller values of μ, the singular vectors are randomly spread out.

III. THEOREMS

Under the above mentioned essential assumptions, the simple PCP approach perfectly recovers low rank and sparse component exactly with large probability.

A. *Theorem-1*

Suppose $L_0 is n * n$, obeys the above mentioned prior assumptions. Fix any$n * n matrix|\Sigma$ of signs. Suppose that the support set$-\Omega$ of S_0 is uniformly distributed among all sets of cardinality m, and that $sgn([S_0]_{ij}) = \Sigma_{ij} forall(i;j) \in \Omega$. 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

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

B. *Matrix completion from grossly corrupted data*

In many situations, it is possible that some values can be corrupted. In some applications, some entries may be missing as well. Denoting P_Ω as an orthogonal projection on linear space of matrices having support on $\Omega \subset [n1] \times [n2]$, As we have only few entries of $L_0 + S_0$, we can write Y as,

$$Y = P_\Omega(L_0 + S_0) = P_{\Omega_{obs}} L_0 + S\prime_0$$

C. *Theorem-2*

Suppose $L_0 in n \times n$,obeys the incoherence conditions and that obeys the prior assumptions ans 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 $1 - cn^{-10}$, Principle Component Pursuit with$\lambda = 1/\sqrt{0.1n}$

is exact, that is $L_0' = L_0$ provided that,

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

Here,ρ_τ and τ_s are positive numerical constants. For $n1 \times n2$ matrices, we take $\lambda = \frac{1}{\sqrt{0.1 n_{(1)}}}$ succeeds from m = 0.1 n1n2 corrupted entries with probability at least $1 - cn_{(1)}^{-10}$ (1) provided that $rank(L_0) \leq \rho_r \eta \mu^{-1}(log n_{-1})^{-2}$

IV. ALGORITHM

We use the Augmented Langrange Multiplier method to solve this convex optimization problem. The algorithm is generalized to wide range of problems. The rank remains bounded by rank(L0) throughout the optimization. The augmented langrangian here is:

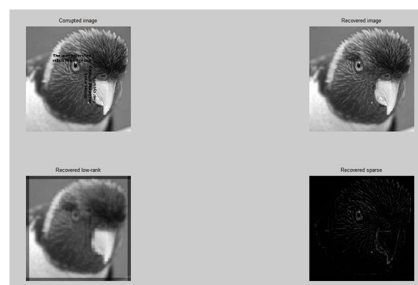
$$l[L, S, Y] = \parallel L \parallel_* + \lambda \parallel S \parallel_1 (Y, M - L - S) + \frac{\mu}{2} \parallel M - L - S \parallel_F^2$$

Below is the proposed Alternating Directions methods,which is a special case of augmented Lagrange multiplier(ALM).

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

V. APPLICATION AND IMPLEMENTATION

Robust PCA has numerous application such as Video Surveillance where it is often required to identify the activities that stand out from the background, face recognition where we effectively model low-dimensional for imagery data, Latent Semantic Indexing, Matrix Completion and Recommendation System.



CONCLUSION

We can conclude from the above explanation and the results that one can disentangle the low-rank and the sparse components exactly by convex programming, this provably works under quite broad conditions. Also the above method can be used for matrix completion and matrix recovery from sparse errors and this also works in the case when there are both incomplete and corrupted entries.

REFERENCES

[1] Candès, E. J., Li, X., Ma, Y., and Wright, J. 2011. Robust principal component analysis? J. ACM 58, 3, Article 11 (May 2011), 37 pages.
[2] Yi Ma, Compressive Principal Component Pursuit, 2012
[3] C HANDRASEKARAN , V., S ANGHAVI , S., P ARRILO , P., AND W ILLSKY , A. 2009. Rank-sparsity incoherence for matrix decomposition. Siam J. Optim.
[4] https://github.com/dlaptev/RobustPCA