Robust Principal Component Analysis?

Kishan Raval, 1401117
School of Engineering and
Applied Science
Ahmedabad University
Email: kishan.r.btechi14@ahduni.edu.in

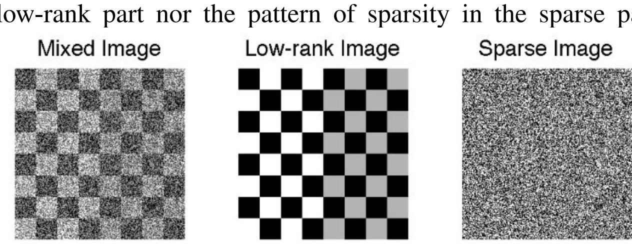
***Abstract*—We talk about a ubiquitous problem in every domain working with large data, where we have a data matrix, which is the superposition of a low-rank component and a sparse component. our aim is to separate both the low-rank(i.e. original data) matrix and sparse (corrupted) Matrix from the data Matrix. Here we extract them exactly by solving a complex program called Principal Component Analysis. Under certain conditions, we prove that the separation is possible exactly by minimizing the weighted combination of nuclear norm and l_1 norm. Then we claim that we can extract these components exactly even is some fraction of the entries are corrupted and some of them are missing too. We show algorithm to solve the convex optimization problem and then show some of the real life application problem and then show some of the real life applications of the same such as foreground detection from video, Matrix completion and Netflix Recommendation system.**

Key words: Principal Component Analysis, Singular Value Decomposition, l_1 norm minimization, nuclear norm, sparsity, duality.

I. Introduction

Principal Component Analysis is widely used in many areas of applied statistics. It is natural since interpretation and visualization in a fewer dimensional space is easier than in many dimensional space. But in real life situations, the data observed has some fractions of outliers. Principal Component Analysis is not prone to corrupted values. If we organize all the points and then apply SVD, hen SVD fails in such cases where we also have corrupted features.

Robust PCA deals with the problem of making PCA robust to outliers and grossly corrupted observations. As, shown in the figure 1, the Robust PCA problem can be formulated mathematically as the problem of decomposing a matrix consisting of sum of low-rank matrix and a sparse matrix into these two components, without any prior knowledge of the low-rank part nor the pattern of sparsity in the sparse part.



[Figure 1]

II. The Matrix Separation Problem

Let the data matrix M has above discussed vectors L_0 and S_0. Then, ideally

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

and we aim to extract L_0 and S_0 exactly. Initially, as the number if unknowns are higher and we do not know the amount of errors in L_0, it seems daunting to extract them. Hence, this seems to be a hard problem, as we do not know how to solve this non-convex, we try to relax the problem to a convex optimization problem as following:

$$minimize \quad ||L||_* + \lambda||S||_1$$

$$Subject\ to \quad M = L_0 + S_0 \tag{2}$$

Where, $||L||_* = \Sigma_i \sigma_i(L)$ denote the nuclear norm of matrix L and $||S||_1 = \Sigma_{ij}|M_{ij}|$ of the matrix S denotes the l_1 norm

Under these relaxed assumptions, Principal Component Pursuit(PCP) estimates low-rank L_0 and sparse S_0 exactly. The solution holds true even if rank of L grows linearly with the dimensions of M and S_0 has constant fractions of entries.

III. Prior Assumptions

While separating the data matrix M, what if M has both sparse and low-rank component? [1] 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.

$$max_i||U * e_i||^2 \leq \frac{\mu r}{n_1}$$

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

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.

IV. MAIN RESULT

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

A. Theorem 1.1

Suppose L_0 is $n * n$, obeys the above mentioned prior assumptions. 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 $sgn([S_0]_{ij}) = \Sigma_{ij}$ for all $(i,j)\epsilon\Omega$. 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 $\lambda = 1/\sqrt{n}$ is exact, that is, $\hat{L} = L_0$ and $\hat{S} = S_0$, provided that

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

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 $\Omega \subset [n_1] \times [n_2]$,

As we have only few entries of $L_0 + S_0$, we can write Y as,

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

Here, we only see the entries in Ω_{obs}. So, even though data matrix is undersampled, we can recover L_0 from $P_{\Omega_{obs}} L_0$.

Following approach recovers the low-rank component exactly.

$$\begin{aligned} & minimize \quad \| L \|_* + \lambda \| S \|_1 \\ & subject\ to \quad P_{\Omega_obs}(L+S) = Y \end{aligned} \tag{3}$$

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,

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

Here,ρ_r and τ_s are positive numerical constants. For $n_1 \times n_2$ matrices, we take $\lambda = \frac{1}{\sqrt{0.1 n_{(1)}}}$ succeeds from $m = 0.1 \; n_1 n_2$ corrupted entries with probability at least $1 - cn_{(1)}^{-10}$ provided that rank$(L_0) \leq \rho_r n \mu^{-1}(logn_{-1})^{-2}$

V. 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(L_0) throughout the optimization. The augmented langrangian here is:

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

which solves PCP by repeatedly setting L_k and S_k as $argmin_{L,S} l(L,S,Y_k)$

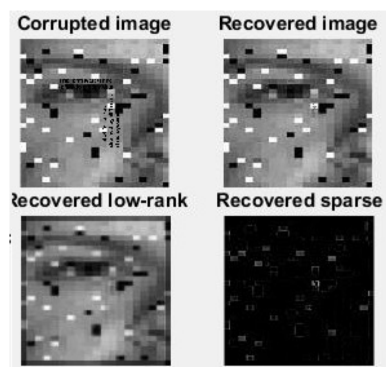
Below is the proposed Alternating Directions methods, which is a special case of augmented Lagrange multiplier(ALM).

Algorithm 1 Principal Component Pursuit 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 $Y_{k+1} = Y_k + \mu(M - L_{k+1} + S_{k+1})$
output L, S.

APPLICATION AND IMPLEMENTATION

Robust PCA can be 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.



[Figure 2]

Above image is showing the result of separation of a corrupted inage into low rank and sparce components. Rank of the low rank component comes out to 47 and cardinality of sparse matrix is 217931.

REFERENCES

[1] Cands, E. J., Li, X., Ma, Y., Wright, J. (2011). Robust principal component analysis?. Journal of the ACM (JACM), 58(3), 11. Chicago
[2] Jolliffe, I. (2002). Principal component analysis. John Wiley and Sons, Ltd. Chicago
[3] E. Candes, X. Li, Y. Ma, and J. Wright. Robust principal component analysis. 2009. http://www-stat.stanford.edu/ candes/papers/RobustPCA.pdf