Robust Principal Component Analysis

Report by Akhil Vavadia
SEAS-AU
Roll No: 1401095

***Abstract*—Principal Component Analysis (PCA) has been most widely a useful for statistical technique. One drawback of the Principal Component Analysis (PCA) methods is that they uses least squares estimation techniques and hence it is very sensitive to "outliers" because one single corrupted point will completely breaks down our data matrix. In computer vision applications, outliers typically occur within a sample due to some noise, errors, or occlusion. Therefore, this paper suggests that a substantial amount of low rank matrices, which cannot be recovered by standard robust PCA, become recoverable by solving a convenient convex program called Principal Component Pursuit(PCP). The result theoretically justifies the effectiveness of features in robust PCA.**

Keywords: Principal component, principal component pursuit, low rank matrix, sparse matrix, video surveillance

I. Introduction

Suppose we are given a large data matrix , and we know it can be decomposed as $M = L_0 + S_0$,where L_0 has low rank and S_0 is sparse;here both components are of arbitrary magnitude.We do not know the low-dimensional column and row space of L_0, not even their dimension. Similarly, we do not know the locations of the nonzero entries of S_0, not even how many there are. Can we hope to recover the low-rank and sparse components both accurately (perhaps even exactly) and efficiently?

Classical PCA: Principal component analysis the most widely used technique for dimensionality reduction in statistical data analysis. The classical PCA problem seeks to find the best (in the least-squares-error sense) low-dimensional subspace approximation to high-dimensional points.

minimize $||M - L||$
subject to rank(L)<= k.

Using the singular-value decomposition (SVD), the PCA finds the lower dimensional approximating subspace by forming a low-rank approximation to the data matrix, formed by considering each point as a column; the output of PCA is the (low-dimensional) column space of this low-rank approximation.

Classical PCA is extremely sensitive to the presence of ouliers(grossly corrupted data):a single grossly corrupted entry in data matrix could render our output(aproximation) arbitrarily far from the true output.The problem we study here can be considered an idealized version of Robust PCA, in which we aim to recover a low-rank matrix L_0 from highly corrupted measurements in data matrix $M = L_0 + S_0$.

There are many important applications in which the data under study can naturally be modeled as a low-rank plus a sparse contribution.Some of these are face recognition, video surveillance, latent semantic indexing, ranking and collaboration filtering, system identification, graphical modeling. A key application where this occurs is in video analysis where we have given video frames as data matrix and we have to separate a background from moving foreground objects.

II. Proposed Solution

Our problem of recovering low dimensional and sparse components from grossly corrupted data matrix can be solved by *convex optimization* ,to achieve that we will use method called *Principal Component Pursuit (PCP)*.Let $||L||_* := \sum_i \sigma_i(L)$ denote the nuclear norm of the matrix M, that is, the sum of the singular values of M, and let $||M||_1 = \sum_{ij} |M_{ij}|$ denote the l_1-norm of M seen as a long vector in R^{n1Xn2}.Then

minimize $||L||_* - \lambda ||S||_1$
subject to $L + S = M$.

exactly recovers the low-rank L_0 and the sparse S_0.This solves our problem at cost no so much higher than classical PCA.

When Does Separation Make Sense?

To avoid identifiability issue and make problem meaningful we make two below assumptions,

A. Low-rank component can not be sparse

This is an assumption concerning the singular vectors of the low-rank component.That is,

$$L_0 = U\sigma V^* = \sum_{i=1tor} \sigma_i u_i v^*$$

where r = rank(L_0).Then,the incoherence condition with parameter μ states that,

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

and

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

Here,since the orthogonal projection P_U onto the column space of U is given by $P_U = UU^T$, (1) is equivalent to

$max_i||P_U e_i||^2 <= \mu r/n_1$, and similarly for PV.the incoherence condition asserts that for small values of μ, the singular vectors are reasonably spread out-not sparse.

B. Sparse matrix can not be 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 it is clear that we would not be able to recover L_0 and S_0 since $M = L_0 + S_0$ would have a column space included in that of L_0.

III. Main Result

Under these minimal assumptions, the simple PCP solution perfectly recovers the low-rank and the sparse components.We found two theorems considering two situations as mentioned below

A. Data matrix with corrupted entries:

if L_0 is nXn of

$$rank(L_0) \leq \rho_r n \mu^{-1} (logn)^{-2} \tag{3}$$

and S_0 is also nXn with random sparsity pattern of cardinality $m <= \rho_s n^2$ then with probability $1 - O(n^{-10})$ and PCP with $\lambda = 1/\sqrt{n}$ we can recover exact L_0 and S_0.
Here,for rectangular matrices $n = max(n1Xn2)$.A rather remarkable fact is that there is no tuning parameter in our algorithm.Further, it is not a priori very clear why $\lambda = 1/\sqrt{n}$ is a correct choice no matter what L_0 and S_0 are. It is the mathematical analysis which reveals the correctness of this value. one can obtain results with larger probabilities of success, that is, of the form $1 - O(n^{-\beta})$ for $\beta > 0$ at the expense of reducing the value of ρ_r .

B. Data matrix with both corrupt and missing entries:

if L_0 is nXn of

$$rank(L_0) \leq \rho_r n \mu^{-1} (logn)^{-2} \tag{4}$$

and with Ω random set of size $m = 0.1n^2$ and each observed entry is corrupted with probability $\tau < \tau_s$ then with probability $1 - O(n^{-10})$ and PCP with $\lambda = 1/\sqrt{0.1n}$ we can recover exact L_0 and S_0.
That is,

$$\begin{aligned}&\text{minimize } ||L||_* - \lambda||S||_1\\ &\text{subject to } L_{ij} + S_{ij} = M_{ij} \text{ (i,j) } \epsilon\ \Omega_{obs}.\end{aligned}$$

which is same as matrix completion problem. In short, perfect recovery from incomplete and corrupted entries is possible by convex. optimization.

IV. Algorithm

Algorithm 1 given below is a special case of a more general class of augmented Lagrange multiplier algorithms known as alternating directions methods

ALGORITHM 1: (Principal Component Pursuitby Alternating Directions [Lin et al. 2009a; Yuan and Yang 2009])

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

References

[1] Robust Principal Component Analysis? by Emmanuel j. Cande's and Xiaodong Li, Stanford University,Yi Ma, University of Illinois at Urbana-Champaign, John Wright, Microsoft Research Asia.

[2] Chandrasekaran, V., Sanghavi, S., Parrilo, P.,AND Willsky, A. 2009. Rank-sparsity incoherence for matrix decomposition. Siam J. Optim., to appear http://arxiv.org/abs/0906.2220.

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

[4] 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