Robust Principle Component Analysis

Hardil Mehta
Information and Communication technology(ICT)
School of Engineering and Applied Science
Ahmedabad,
Email: Hardil.m.btechi14@ahduni.edu.in

***Abstract*—Over the past decade there has been an explosion in terms of the massive amounts of high dimensional data in almost all fields of science and engineering.In order to be able to even make sense of data in such high dimensionality and scale, one has to leverage on the fact that such data often have low intrinsic dimensionality.An approach is presented here, which assumes that the data all lie in near some low-dimensional subspace.In terms of Mathematics,it can be said as all the data points are stacked as column vectors of matrix M,the matrix should have low rank. This approach, known as Principal Component Analysis.We would therefore like a method that is able to extract the principal components (the low rank structure) of measurement data even in the presence of such gross but sparse corruptions.The Proposed Robust PCA framework is used for this task. Even though a positive fraction of entries in the data matrix is corrupted, this approach recovers the principle components of data matrix, thus this convincing property of principle component pursuit leads us to a principled approach to robust principle component analysis.This approach also applies to a situation when fraction of entries are missing as well.**

***Index Terms*—nuclear norm minimization, $\ell 1$-norm minimization,, duality, low-rank matrices, sparsity.**

I. Introduction

Given an observed matrix M that is formed as a superposition of a low-rank matrix L_0 and a sparse matrix S_0,we have

$$M = L_0 + S_0$$

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 Principle Component Analysis

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,

$$M = L_0 + N_0$$

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

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

Throughout this article $||M||$ denotes $\ell 2$ norm. In classical PCA it is assumed that N_0 is small.

B. Robust Principle Component Analysis

Robust Principal Component Analysis is the problem of recovering the low-rank and sparse components. Under suitable assumptions on the rank and incoherence of L0, and the distribution of the support of S0, the components can be recovered exactly with high probability, by solving the Principal Component Pursuit (PCP) problem given by

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

Principal component pursuit minimizes a linear combination of the nuclear norm of a matrix L and the norm of M L. Minimizing the $\ell 1$ norm is known to favor sparsity, while minimizing the nuclear norm $||L||_* = \sum_{\sigma \in \sigma(L)} \sigma$ is known to favor low-rank matrices (intuitively, favors sparsity of the vector of singular values).

The applications of the above mentioned Robust PCA along with convex optimization and other multiplier algorithms are ***video surveillance, Face recognition, Latent Semantic Indexing*** and many more field.

II. Problem and approach to solution

The separation problem seems impossible to solve since the number of unknowns to infer for L_0 and S_0 is twice as many as the given measurements in $M \in R^{n_1 * n_2}$.In this article, we are going to see that it can be solved by tractable convex optimization. Let $\parallel M \parallel := \sigma_i j |M_i j|$ denote the nuclear norm of matrix M,sum of singular values of M.
Under weak assumptions ,the Principal Component Pursuit(PCP) estimate solving

$$\begin{aligned} &\text{minimize} \parallel \text{L} \parallel_* + \lambda \parallel \text{S} \parallel_1 \\ &\text{subject to rank(L)} \leq \text{K} \end{aligned}$$

exactly recovers the low-rank L_0 and the sparse S_0. It is guaranteed to work even if the rank of L_0 grows almost linearly in the dimension of the matrix, and the errors in S_0 are up to a constant fraction of all entries.Algorithmically, we will see that this problem can be solved by efficient and scalable algorithms, at a cost not so much higher than the classical PCA. Empirically, our simulations and experiments suggest this works under surprisingly broad conditions for many types. of real data.

A. *Incoherence of the low rank component L0*

For instance, suppose the matrix M is equal to $e_1 e_1*$ (this matrix has a one in the top left corner and zeros everywhere else).Then since M is both sparse and low-rank, how can we decide whether it is low-rank or sparse? To make the problem meaningful, we need to **impose that the low-rank component L_0 is not sparse**.Using incoherence introduced in Candes and Recht[2009] for the matrix completion problem;it is an assumption concerning the singular vectors of the low rank component.SVD of $L_0 \in R^{n_1 * n_2}$ as

$$L_0 = \mathrm{U}\Sigma V^*$$

where r is the rank of the matrix, σ1 ,..., σr are the positive singular values, and U = [u1 ,...,ur], V = [v1 ,...,vr] are the matrices of left-singular and right-singular vectors.Then the incoherence conditions are given by

$$max_i ||U * e_i||_2^2 \leq \frac{\mu r}{n_1}, max_i ||V * e_i||_2^2 \leq \frac{\mu r}{n_2} - (1)$$

and

$$||UV*||_\infty \leq \sqrt{\frac{\mu r}{n_1 n_2}} - (2)$$

These conditions require the singular vectors to be spread enough with respect to the canonical basis. Intuitively, if the singular vectors of the low-rank matrix L0 are aligned with a few canonical basis vectors, then L0 will be sparse and hard to distinguish from the sparse corruption matrix S0.

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 it is clear that we would not be able to recover L_0 and S_0 by any method whatsoever since M = L_0 + S_0 would have a column space equal to, or included in that of L_0 . To avoid such meaningless situations, we will assume that the **sparsity pattern of the sparse component is selected uniformly at random.**

III. MAIN RESULT

Under these minimal assumptions, the simple PCP solution perfectly recovers the low-rank and the sparse components, provided the rank of the **low-rank component is not too large**, and that the **sparse component is reasonably sparse**.

A. *Theorem 1.1*

Suppose $M \in R^{n*n}$ satisfies incoherence conditions (1) and (2) that the support of S_0 is uniformly distributed among all sets of cardinality m,Then $\exists$c such that with high probability over the choice of support of S_0 Principal Component Pursuit (1.1) with $\lambda = 1/[\sqrt{n}]$is exact, that is, L = L_0 and S = S_0 , provided that

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

Above, ρr and ρs are positive numerical constants.In the general rectangular case, where L_0 is $n_1 x n_2$, PCP with λ = $\frac{1}{\sqrt{n_{(1)}}}$ 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 $m \leq \rho_s n_1 n_2$. Matrix L_0 whose principal components are spreaded can be recovered with probability almost onefrom arbitary and completely unknown corruption patterns. It also works for higher ranks like n/$log(n)^2$ when μ is not large. Minimizing

$$||L||_* + \frac{1}{\sqrt{n_{(1)}}} ||S||_1$$

where,

$$n_{(1)} = max(n_1, n_2)$$

under the assumption of theorem, this always gives correct answer. Here we chose λ = $\frac{1}{\sqrt{n_{(1)}}}$ but it is not clear why that has happened. It has been due to mathematical analysis why we are taking that value.

IV. CONCLUSION

From the above results, we can say that if conditions in equation (1) and (2) are satisfied then we can get a low rank and sparse matrix from matrix M. The main phenomena is to select a uniformly random sparse matrix. So this approach can be used to decompose a given matrix in two components i.e. low rank and sparse matrix.

REFERENCES

[1] Cands, 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] https://github.com/dlaptev/RobustPCA