Robust Principal Component Analysis.

Report by Ratnesh Jigish Shah
School of Engineering and Applied Sciences,Ahmedabad University
Email: ratneshshah1997@gmail.com

***Abstract*—In this era of Big Data, there are datasets of high dimension that can have a fraction of missing entries and also lots of entries are corrupted and the problem is to fill those missing entries and also to correct those erroneous data. Suppose we have a data matrix,which is the superposition of a low-rank component and a sparse component and we want to recover each component individually this can be done using convex optimization techniques which are discussed in the paper.**

Keywords: Robust PCA, Low rank, Sparse Matrix,,Convex Optimization, Principal Component pursuit, Matrix Norms, Augmented Lagrange Multiplier.

I. Introduction

PCA is one of the most widely used statistical tool for data analysis and dimensionality reduction but the problem is that it is very sensitive to outliers.A single error point can drastically change the PCA component since classical PCA gives the direction of maximum variance. Thus, a single grossly corrupted entry in data could render the estimated matrix arbitrarily far from the true matrix.Gross errors occur in many applications such as Image Processing, Web data Analysis,Bioinformatics,sensor failures etc. hence it is important to make PCA robust.

If we are given a large data matrix and we know that it can be decomposed as $M = L_0 + S_0$ where L_0 has low rank and S_0 is sparse.We have no information about the low-dimensional column and row space of L_0.We also have no information of the location of nonzero entries in the sparse matrix and our goal is to recover the low-rank and the sparse components accurately. One of the applications for this could be video surveillance where the low rank component would be the background which would be static while the sparse component would contain some activity on the foreground.

II. Theoretical Aspects

This problem of separating the low rank and the sparse component from the data matrix can be solved using convex optimization technique Principal Component Pursuit.

A. Recovering using PCP :

$$M = L_0 + S_0$$

L_0 is unknown(rank is unknown)
S_0 is unknown(No. of nonzero entries,magnitudes,locations)

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

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

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

Under weak assumptions this method exactly recovers the low rank L_0 and the sparse S_0.

B. Conditions:

For the above results to be true there are few conditions:

1. Low Rank Component cannot be Sparse
Suppose if we are given a matrix like this:

$$\begin{bmatrix} x_1 & x_2 & x_3 & \dots & x_n \\ x_1 & x_2 & x_3 & \dots & x_n \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

Now if this data matrix is corrupted such that the matrix is

$$\begin{bmatrix} x_1 & \mathbf{x} & x_3 & \dots & x_n \\ x_1 & \mathbf{x} & x_3 & \dots & x_n \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

We can not know what that entry would be as there is no information about x_2 in the corrupted matrix.Thus this condition says that we can not have rows and columns in the data matrix that are completely orthogonal to the rest.
To quantify this, the notion of Incoherence condition is introduced:

We have got a low rank component L_0 which has a column space and the row space which can be looked at by its SVD

i.e. $L_0 = U \sum V^* = \sum_{i=1}^{r} \sigma_i u_i v_i^*$ $\qquad r = \text{rank}(L_0)$.

Now, we are going to measure the correlation of the column space and the row space with the basis vector.

$$\max_i ||U * e_i||^2 \leq \frac{\mu r}{n_1} \qquad \max_i ||V * e_i||^2 \leq \frac{\mu r}{n_2}$$
$$||UV^*||_\infty \leq \sqrt{\frac{\mu r}{n_1 n_2}}$$

l_∞ norm: $||M||_\infty = max_{i,j}|M_{ij}|$
We take a basis vector and project it on to the column space and if the value if small than this means that the singular vectors are reasonably spread out hence not sparse.

2. Sparse Matrix cannot have Low Rank
This condition occurs when all the nonzero entries of S are present in a column or few columns.So, if the data in that whole column gets corrupted than there is no way to recover

the data.
Thus to avoid such cases, we will assume that sparse component is uniformly distributed.

III. Results

A. Theorem 1.1

It says that if we are given a low rank component of n x n and using the support set of S which is uniformly distributed if we corrupt the entries in the low rank component than by solving this,

$$\min ||L||_* + \lambda ||S||_1$$

with $\lambda = 1/\sqrt{n}$ we recover the low rank component with probability atleast 1 − cn^{-10} such that the recovered components are exactly the same i.e. $\hat{L}$ = L and $\hat{S}$ = S provided that

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

ρ_r and ρ_s are positive numerical constants and m is the cardinality.

B. Matrix completion from grossly corrupted data

There may be situation where the entries are both missing and corrupted.Using convex optimization technique such as PCP,

$$\text{minimize } ||L||_* + \lambda ||S||_1$$
$$\text{subject to } L_{ij} + S_{ij} = M_{ij}, (i,j) \in \Omega_{obs}$$

Ω_{obs} locations of observed entries.
Theorem:

- L_0 is n x n as before, rank(L_0) $\leq \rho_r n \mu^{-1} (logn)^{-2}$
- Ω_{obs} random set of size $m = 0.1n^2$
- Each observed entry is corrupted with probability $\tau \leq \tau_s$.

Then with probability $1 - cn^{-10}$, PCP with $\lambda = \frac{1}{\sqrt{0.1n}}$ is exact:

$$\hat{L} = L_0$$

IV. Algorithm

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

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

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

Scalar Shrinkage: $S_\tau[x]$ = sgn(x)max($|x| - \tau$,0)
Componentwise Thresholding $S_\tau(X)$
Singular Value Thresholding $D_\tau(X)$

$$D_\tau(X) = US_\tau(\sum)V^* \qquad X = U\sum V^*$$

Lagrangian:
L(L,S;Y) = $||L||_* + \lambda ||S||_1 + \frac{1}{\tau} < Y, M - L - S > + \frac{1}{2\tau} ||M - L - S||_F^2$
When iterates L_{k+1} have low rank:

- Only need to compute few singular values at each step, this requires us to compute those singular vectors of $M - S_k + \mu^{-1} Y_k$ whose corresponding singular vectors exceed the threshold μ.
- We shrink scalar values at each iteration

We simply choose $\mu = \dfrac{n_1 n_2}{4||M||_1}$ as suggested in Yuan and Yang[2009].

V. Applications

There are many important applications where the data matrix can be modelled as a combination of low rank and the sparse components.

- Image Filtering: Shadows and specularities can be removed from the face images.
- Video Surveillance: Background Modeling from surveillance video.
- User rating prediction:The problem is to use incomplete ranking provided by the users on some of the products to predict the preference of any user on any of the products.

VI. 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] 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