Robust Principle Component Analysis.

Report by Rushita Thakkar Roll number 1401004
Email: rushita.t.btechi14@ahduni.edu.in

***Abstract*—The paper basically shows an algorithm for solving the Robust Principle Component Analysis problem i.e. Recovering a low rank matrix with an unknown fraction of entries being arbitrary corrupted or missing .The data matrix is basically a super-position of low rank component and sparse component.It is shown that under some specific conditions and assumptions,it is possible to recover both low rank component as well as sparse components(errors).Recovering the Low-rank component and the Sparse component is possible by solving convex optimization program called Principle Component pursuit by minimizing a weighted combination of nuclear norm and of l_1 norm.This suggests a possibility of principled approach to robustify the Principle component Analysis.This can be useful in foreground background separation video surveillance,face recognition etc.**

Keywords: PCA,Matrix norms,low rank matrices,sparsity,outliers,video surveillance

I. Introduction

In case of Big Data,high dimensionality is a curse.Classical Principle Component Analysis is widely used statistical tool for **dimension reduction** where the concept is that all data points lie near a low dimensional subspace.If we stack all the data points as column vectors of a Matrix M,Matrix should have a low rank.

i.e. $M = L_0 + S_0$,

L_0 is the low rank matrix and S_0 is the sparse matrix. Classical PCA is very **sensitive to outliers**,even 1 single corrupted entry in M can render the estimated L' arbitrarily far from true L_0.Corrupted and Missing entries are very common in modern day applications such as image processing,data analysis,bio-informatics etc..The $Robust$ PCA approach aims at recovering Low rank Matrix L_0 from highly **corrupted** as well as **missing entries** from data matrix M.We do not know the low dimensional row and column space of L_0 and the locations of non-zero entries of Sparse component S_0

II. Notations and Definitions

$\| M \|$: **2-norm** i.e. Largest singular value of M
$\| M \|_*$:**Nuclear norm** i.e. $\sum_i \sigma_i(M_i)$
$\| M \|_1$: l_1-norm i.e $\sum_{ij} |M_i|$
$\| M \|_F$:**Frobenius norm** i.e.$\sqrt{trace(M^*M)}$
$\langle M, N \rangle$: The **Euclidean inner product** of two matrices i.e. trace(M*N)
$S_\tau : \mathbb{R} \rightarrow \mathbb{R}$ is the **shrinkage operator** $S_\tau[x] = sgn(x)max(|x| - \tau, 0)$
$D_\tau(X)$ is the **singular value threshold operator** given by $D_\tau(X) = US_\tau(\Sigma)V^*$ is any singular value decomposition

III. Main Results

As mentioned above,the high dimensional data can be projected onto low sub-space using PCA i.e. they lie on some **low-dimensional subspace** and are sparse in some basis. Writing the data matrix M as superposition of a matrix N_0(noise) and **Low rank matrix** L_0.

M = $L_0 + N_0$

Classical PCA seeks best rank-*k* estimation of L_0 by solving:

$$\begin{array}{c} \text{minimize} \;\; \| M - L \| \\ \text{subject to} \;\; rank(L) \leq k \end{array}$$

Here the objective function is to minimize the noise.The problem can be solved efficiently via **Singular Value Decomposition** and enjoys optimality properties when the noise N_0 is small.

In case of Robust PCA,the aim is to separate a low rank matrix and a sparse matrix of arbitrary large magnitude.Under weak assumptions,the *Principle Component Pursuit* estimate solving

$$\begin{array}{c} \text{minimize} \; \| L \|_* + \lambda \| S \|_1 \\ \text{subject to} \; L + S = M \end{array}$$

exactly recovers the low-rank L_0 and sparse S_0.The paper shows that,at cost not much higher than Classical PCA.This result works under surprisingly broad conditions for many types of real data.

Now,in order to separate efficiently and correctly,it is necessary that the **low rank** component is **not sparse** and the **sparse matrix** is **not low rank** .This is the **Incoherence condition**.If this condition is not satisfied, it would be extremely difficult to recover the components. μ correlation between the basis vectors of column space and row space.$L_0 = U\Sigma V^*$.The condition is:

$$\begin{array}{c} max_i \| U^* e_i \|^2 \leq \frac{\mu r}{n_1} \;, max_i \| V^* e_i \|^2 \leq \frac{\mu r}{n_2} \\ \text{and} \; \| UV^* \| \leq \sqrt{\frac{\mu r}{n_1 n_2}} \end{array}$$

The above condition asserts that L_0 and S_0 are **not orthogonal**.Thus,for small values of μ,the singular vectors are not sparse.

A. *Theorem 1*

Suppose L_0 is $n \times n$ obeying incoherence condition.Fix any 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 $sgn([S_0]) = \Sigma_{ij}$ for all $(i, j) \in$ .Then there is a numerical constant c such that with probability at least $1 - cn^{-10}$(over *the choice of support S_0),PCP with $\lambda = 1/\sqrt{n}$ is exact,that is $L' = L_0$ and $S' = S_0$ provided that*

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

ρ_s and ρ_s are numerical constants So,just placing any random values with random signs in the Sparse matrix ,just keeping in mind the support set recovering low rank component and sparse component is possible with **probability nearing 1**.This also works for large values of rank i.e. order of $n/(logn)^2$.The piece of **randomness** in assumptions are locations of non zero entries of S_0,everything else is deterministic.Also the choice of $\lambda = 1/\sqrt{n_{(1)}}$ is universal where $n_{(1)}$ = max(n1,n2)

B. *Matrix Completion from Grossly Corrupted data*

P_Ω is the orthogonal projection onto the linear space of matrices supported on $\Omega \subset [n_1] \times [n_2]$,

$$P_\Omega = \begin{cases} X_{ij}, (i,j) \in \Omega \\ 0, (i,j) \notin \Omega, \end{cases}$$

As we have only few entries of $L_0 + S_0$ which can be written as: $Y = P_{\Omega_{obs}}(L_0 + S_0) = P_{\Omega_{obs}} + S_0^{'}$ So,we only see the entries in Ω_{obs}So here we have very few entries and some entries are corrupted.But recovering L and S is possible if we have **under sampled but perfect** $P_{\Omega_{obs}}(L_0)$ by solving the convex program.

$$\begin{gathered} \text{minimize } \parallel L \parallel_* + \lambda \parallel S \parallel_1 \\ \text{subject to } P_{\Omega_{obs}}(L+S) = P_{obs}L_0 \end{gathered}$$

C. *Theorem 2*

Suppose L_0 in $n \times n$,obeys the incoherence conditions and that Ω_{obs} 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,

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

So,perfect **recovery from incomplete and corrupted entries** is possible by convex programming. Here,ρ_r and τ_s are positive numerical constants.For $n_1 \times n_2$ matrices we take $\lambda = 1/\sqrt{0.1n_{(1)}}$ succeeds from $m = 0.1n_1n_2$ corrupted entries with probability at least $1 - cn_{n(1)}^{-10}$ provided that $rank(L_0) \leq \rho_r n_{(2)}\mu^{-1}(logn_{n(1)})^{-2}$

If τ = 0,we have pure matrix completion problem.

IV. ALGORITHM

Here the **Principle Component Pursuit** is solved using the **Augmented Lagrange Multiplier** Algorithm.This works across a wide range of problem settings with no tuning parameters.The rank iterates often remains bounded by $rank(L_0)$ throughout the optimization.The augmented Lagrangian for this problem:

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

The solution is obtained by repeatedly setting $(L_k, S_k) = \arg min_{L,S} l(L,S,Y_k)$ and the updating the Lagrange multiplier by $Y_{k+1} = Y_k + \mu(M - L_k - S_k)$ Now,

$$\begin{gathered} \arg min_S l(L,S,Y) = S_{\lambda/\mu}(M - L - \tfrac{1}{\mu}Y) \\ \arg min_L l(L,S,Y) = D_{1/\mu}(M - S - \tfrac{1}{\mu}Y) \end{gathered}$$

Here the strategy is to first **minimize l w.r.t L fixing S,then minimize l w.r.t S fixing L**,and the finally **update the Lagrange Multiplier** Y based on residual $M - L - S$

Algorithm 1 ALM Algorithm

Result:

initialize :$S_0 = Y_0, \mu > 0$
while *not converged* **do**
  compute $L_{k+1} = D_{1/\mu}(M - S - \frac{1}{\mu}Y)$
  compute $S_{k+1} = S_{\lambda/\mu}(M - L - \frac{1}{\mu}Y)$
  compute $Y_{k+1} = Y_k + \mu(M - L_k - S_k)$
  end while
  return *L,S*
end

Here $\mu = n_1n_2/4 \parallel M \parallel_1$.Algorithm is **terminated** when $\parallel M - L - S \parallel_F \leq \delta \parallel M \parallel_F$ with $\delta = 10^{-7}$

V. VIDEO FOREGROUND BACKGROUND SEPARATION APPLICATION

Separation of Low rank component and the sparse component can be really very useful for foreground and background separation in a video.So basically the part that remains same or **stationary** throughout can be modelled as low rank whereas the **moving objects** in the video frame can be separated as Sparse matrix as it occupies few pixels in the frame.So we can just focus on the moving part of the video if we want.This can be extremely useful for surveillance and military purposes where we are interested in recording unusual activities of the opposite party.

VI. CONCLUSION

In applications such as live video streaming ,where streaming of huge amount of data.Our aim is to process this data as quickly and efficiently as possible.Robust PCA algorithm in this paper **ensures dimension reduction** separating the low rank component as well as the sparse component.It **corrects the errors** in the data and also **completes the matrix** where the entries are missing.The cost of separation and recovering in not much higher than Classical Principle Component Analysis.

VII. REFERENCES

REFERENCES

[1] Candès, E. J., Li, X., Ma, Y., Wright, J. (2011). Robust principal component analysis?. Journal of the ACM (JACM), 58(3), 11. Chicago

[2] E. Candes, X. Li, Y. Ma, and J. Wright. Robust principal component analysis. 2009. http://www-stat.stanford.edu/ candes/papers/RobustPCA.pdf

[3] E. Candes, X. Li, Y. Ma, and J. Wright. Robust principal component analysis. 2011. https://github.com/dlaptev/RobustPCA