Robust Principal Component Analysis

Nand Bharat Parikh
Information and Communication Technology(ICT)
School of Engineering and Applied Science (SEAS)

***Abstract*—Given a data matrix which is a superposition of sparse and a low rank matrix; under some suitable assumptions this article proposes a convex optimization program called Principal Component Pursuit(PCP) which exactly recovers the sparse and low rank matrix. It suggests that even though a positive fraction of entries in the data matrix is corrupted, it is still possible to recover the Principal components of data matrix, thus this convincing property of Principal component pursuit leads us to a principled approach for robust Principal component analysis. This approach also applies to a situation when fraction of entries are missing as well, hence it can be extended to matrix completion problem. Application of the discussed algorithm here, extends to face recognition where it allows to remove speculations regarding shadowing effect and in video surveillance where this method helps us in detecting objects in cluttered background.**

I. Introduction

Suppose we are given a data matrix M such that M can be decomposed as $M = L_0 + S_0$, where L_0 has low rank and S_0 is sparse 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 has been many prior attempts to solve or at-least alleviate the above mentioned problem.One such attempt is the classical Principled Component Analysis.

A. Classical Principal 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

$$minimize \quad ||M - L||$$

$$subject\ to \quad rank(L) \leq k.$$

Throughout this article $||M||$ denotes 2-norm, which can be solved using Singular Value Decomposition. In classical PCA it is assumed that N_0 is small and independent and identically distributed Gaussian.

B. Robust Principal Component Analysis

PCA is unmistakably the best statistical tool for data analysis and dimensionality reduction. However its brittleness to small corrupted data in data matrix puts its validity in jeopardy, as this small corruption could render the estimated $\hat{L}$ arbitrarily far from true L_0. Such problems are ubiquitous in modern applications such as image processing, web data analysis and many more.The problems mentioned above are the idealized version of robust PCA, where we recover low rank matrix from highly corrupted data matrix M such that $M = L_0 + S_0$;where unlike classical PCA S_0,can have arbitrarily large magnitude and their support is assumed to be sparse.

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. Algorithms

In this section we discuss Principal Component Pursuit *(PCP)* algorithms to successfully retrieve low rank matrix and sparse matrix from a corrupted given data matrix, also to bolster its applicability to large scale problems we rely on convex optimization program. For the experiments performed in this section, we have used *Alternating Direction Method (ADM)* which is a special case of more general Augmented Lagrange multiplier (ALM).

A. Principal Component Pursuit

*1) **Assumptions**:* In practice, There is high possibility that the data matrix M has only the top left corner 1 and all other entries in the matrix are 0, making the matrix M both low-rank and sparse.To avoid this,we assume that low rank matrix L_0 is not sparse. Also,similarly there is a possibility that the sparse matrix S_0 has all non-zero entries in few columns.So, we assume that the sparsity pattern of S_0 is uniformly random.

*2) **Claim**:* Let the data matrix $M \in R^{n_1 \times n_2}$. Also the low rank matrix is L_0 and the sparse matrix is S_0.Let $||M||_* = \sum_i \sigma_i(M)$ denote the nuclear norm of any matrix M.Also $||M||_1$ denote the l_1 norm of any matrix M, then Principal component pursuit gives us estimate,

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

$$subject\ to \quad L + S = M$$

The above estimate exactly recovers the Low-rank matrix L_0 and the sparse matrix S_0. Theoretically the claim is true even if the rank of matrix L_0 almost grows linearly and the errors in S_0 are upto a constant factors of all entries. Empirically we can solve this problem by efficient and scalable algorithms, at a cost not much higher than classical PCA.

*3) **Main Result**:* Throughout this article, we define $n_{(1)} = max(n_1, n_2)$ and $n_{(2)} = min(n_1, n_2)$. Suppose L_0 is a square matrix of any arbitrary rank $n \times n$, such that it obeys the assumptions given above. Suppose that the support set Ω of S_0 is uniformly distributed among all sets of cardinality m, and that $sgn([S_0]_{ij})$ = $\sum_{ij} for\ all(i,j) \in \Omega$.Then there is a numerical constant c such that with probability at least $1 - cn^{-10}$, *Principal Component Pursuit* with λ = $1/\sqrt{n}$, returns exact low-rank and sparse matrix provided that

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

. In the above equation ρ_r and ρ_s are positive numerical constants. In general case this nxn;dimension of L_0 is $n_1 \times n_2$, PCP with $\lambda = 1/\sqrt{n_{(1)}}$, succeeds with the probability of at least $1 - cn_{(1)}^{-10}$, provided that $rank(L_0) \leq \rho_r n_{(2)} \mu^{-1} (logn_{(1)})^{-2}$ and $m \leq \rho_s n_1 n_2$. Thus the the claim we made can be restated

$$\begin{aligned} minimize \quad & ||L||_* + 1/\sqrt{n_{(1)}}||S||_1 \\ subject\ to \quad & L + S = M \end{aligned}$$

.Here it is to note that the parameter λ has not to be balanced between L_0 and S_0 and is independently found to be $\lambda = 1/\sqrt{n_{(1)}}$.

B. Alternating Directions Methods

The below proposed Alternating Directions methods is a special case of augmented Lagrange multiplier(ALM).

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.

In the above given algorithm we have updated the value S_0 and L_0. Let $S_\tau : R \longrightarrow R$ denote the shrinkage operator $S_\tau[x] = sgn(x)max(|x| - \tau, 0)$ and extend it to matrices by applying it to each element it shows that

$$arg\ min_S\ l(L, S, Y) = S_{\lambda/\mu}(M - L + \mu^{-1}Y)$$

. Similarly for matrices X, let $D_\tau(X)$ denote singular value threshold operator given by $D_\tau(X) = US_\tau(\sum)V^*$ and thus

$$arg\ min_L\ l(L, S, Y) = D_{1/\mu}(M - S + \mu^{-1}Y)$$

.Here we suggest $\mu = n_1 n_2 / 4||M||_1$ and we terminate the algorithm when $||M - L - S||_F \leq \delta||M||_F$ with $\delta = 10^{-7}$

III. Results

The results for the above proposed *Principal Component Pursuit* and *Alternating Direction methods* are simulated on MatLAB and a user input image.

A. Applications in Face Recognition

In the application discussed below we have taken a corrupted data matrix M which is a corrupted image and from this corrupted data matrix we have achieved the L_0 low-rank matrix and S_0 completely.Here the speculation and the shadowing effect of the image are stored in S_0.

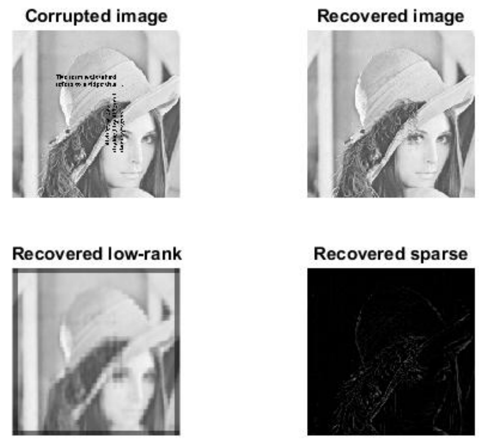

Fig. 1. L_0 and S_0 recovered from corrupted data

As shown above from the corrupted image M having dimensions 320×320, thus $M \epsilon R^{320 \times 320}$ we have successfully recovered L_0 and have recorded speculations and image shadowing in S_0. The program took 8.014 $seconds$ to converge with 154 iterations. The rank of low-rank matrix ;**rank(L)= 26** which suggests it is indeed a low rank. The cardinality of set of the sparse matrix S_0 is **card(S)= 231091**. The error rate observed was 3.37.

IV. Conclusion

Analysis of the above result, lead to the conclusion that with a single universal value $\lambda = 1/\sqrt{n_{(1)}}$, works with high probability for recovering the low-rank matrix L_0 and sparse matrix S_0 accurately under the stated assumptions.And gives one huge advantage over other algorithms as we don't need to compute λ for different kind of data matrix M.

References

[1] Robust Principal Analysis, E.J Candes, Li Ma
[2] Lin et al. 2009a;Yuan and Yang 2009