Robust Principal Component Analysis

Harsh Patel
Information and Communication technology(ICT)
School of Engineering and Applied Science
Ahmedabad,
Email: harshypatel13@gmail.com

***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 which exactly recovers the sparse and low rank matrix. Among all feasible decomposition techniques of a matrix this programs output is a simple addition of weighted nuclear norm of sparse matrix and l1 norm of low rank matrix. Even though a positive fraction of entries in the data matrix is corrupted, this approach recovers the Principal components of data matrix, thus this convincing property of Principal component pursuit leads us to a principled approach to robust Principal component analysis. This approach also applies to a situation when fraction of entries are missing as well. Application of the discussed algorithm here, extends to face recognition where it allows to remove speculations regarding shadowing effect and in video surveillance where our 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 are many prior attempts to solve or atleast alleviate the above mentioned problem.

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 l^2 norm. In classical PCA it is assumed that N_0 is small.

B. Robust Principal Component Analysis

PCA is unequivocally 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 buttress 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**:* 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. Thus M is both sparse and low rank, thus to make the problem meaningful we assume that low rank matrix L_0 is not sparse. Also there is a possibility that the sparse matrix S_0 has all non-zero entries in few columns. To avoid such meaningless situations, we assume that the sparsity pattern of S_0 is uniformly random.

*2) **Claim**:* Let the data matrix $M \in R^{n_1 x n_2}$. Also the low rank matrix is L_0 and the sparse matrix is S_0.Let $||M||_* = \sum_i \sigma_i(P)$ denote the nuclear norm of any matrix M.Also $||M||_1$ denote the l_1 norm of any matrix P, then Principal component pursuit gives 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 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 the 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 rank any arbitrary rank nxn, 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 $sgn([S_0]_{ij}) = \sum_{ij} for\ all(i, j)\ \epsilon\Omega$.Then there is a numerical constant c such that with probability at least $1 - cn^{-10}$, *Principal Component Pursuit* with $\lambda = 1/\sqrt{n}$, returns exact low-rank and sparse matrix provided that

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

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

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

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

.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 $||M - L - S||_F \le \delta ||M||_F$ with $\delta = 10^{-7}$

III. Results

The results for the above proposed *Principal Component Pursuit* and *Alternating Direction methods* are simulated using Matlab as a tool 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.

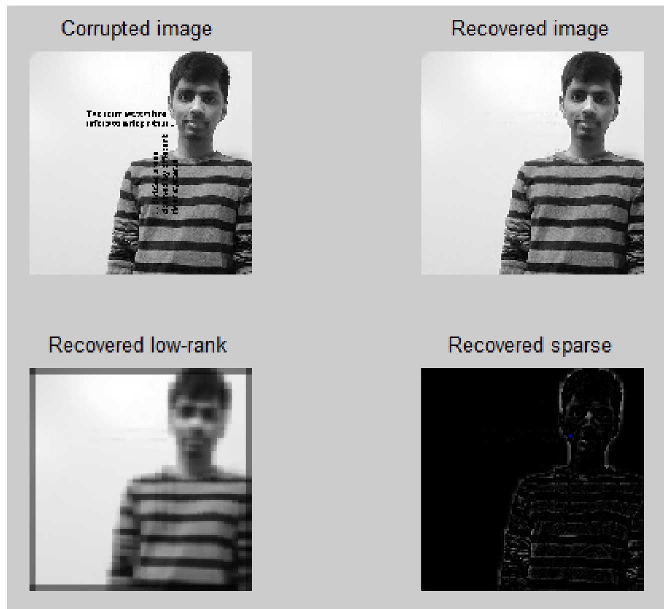

Fig. 1. L_0 and S_0 recovered from corrupted data

As shown above from the corrupted image M having dimensions 256x256, thus $M \epsilon R^{256x256}$ we have successfully recovered L_0 and have recorded speculations and image shadowing in S_0. The program took 25.08 *seconds* to converge with 317 iterations. The rank of low-rank matrix ;**rank(L)= 27** which suggests it is indeed a low rank. The cardinality of set of the sparse matrix S_0 is **card(S)= 223241**. The error rate observed was 2.59.

IV. Future Implementations

As stated initially in our assumption, the discussion here is limited to low-rank component being exactly low and the sparse component being exactly sparse. The algorithms given above can be implemented in the future if any one of the assumption is relaxed.

V. Reference

[1] Hotelling 933; Eckart and Young 1936; Jollife 1986
[2] Robust Principal Analysis, E.J. Candes, Li Ma
[3] Lin et al. 2009a; Yuan and Yang 2009