Robust Principal Component Analysis

Akash Soni, *Roll Number: 1401047*
School of Engineering and Applied Science, Ahmedabad University

***Abstract*—This paper is about a phenomenon in which we can recover each component of data of data matrix, which is the superposition of a low-rank component and a sparse component. We prove that under some suitable assumptions, it is possible to recover both the low-rank and the sparse components by solving Principal Component Pursuit; among all feasible decompositions, simply minimize a weighted combination of the nuclear norm and of the ℓ_1 norm. This suggests the possibility of a principled approach to robust principal component analysis. An algorithm for solving this optimization problem is discussed, and present applications in the area of video surveillance, where the methodology allows for the detection of objects in a cluttered background, and in the area of face recognition, where it offers a principled way of removing shadows and specularities in images of faces.**

***Keywords*—Algorithms, Theory,Principal components, robustness vis-a-vis outliers, nuclear-norm minimization, $\ell_1 - norm$ minimization, duality, low-rank matrices, sparsity, video surveillance**

I. Introduction

The standard PCA algorithm constructs the optimal (in a least square sense) subspace approximation to observations by computing the eigen vectors or Principal Components (PCs) of the sample covariance or correlation matrix. Its broad application can be attributed to primarily two features: its success in the classical regime for recovering a low-dimensional subspace even in the presence of noise, and also the existence of efficient algorithms for computation. Indeed, PCA is nominally a non-convex problem, which we can, nevertheless, solve, thanks to the magic of the SVD which allows us to maximize a convex function. It is well-known, however, that precisely because of the quadratic error criterion, standard PCA is exceptionally fragile, and the quality of its output can suffer dramatically in the face of only a few (even a vanishingly small fraction) grossly corrupted points. Such non-probabilistic errors may be present due to data corruption stemming from sensor failures, malicious tampering, or other reasons. Attempts to use other error functions growing more slowly than the quadratic that might be more robust to outliers, result in non-convex (and intractable) optimization problems.

A. Motivation

Suppose we are given a large data matrix M, and know that it may be decomposed as

$$M = L_0 + S_0,$$

where L_0 has low rank and S_0 is sparse; here, both components are of arbitrary magnitude. We do not know the low-dimensional column and row space of L_0, not even their dimension. Similarly, we do not know the locations of the nonzero entries of S_0, not even how many there are. To alleviate the curse of dimensionality and scale, we must leverage on the fact that such data have low intrinsic dimensionality. More precisely, this says that if we stack all the data points as column vectors of a matrix M, the matrix should (approximately) have low rank: mathematically,

$$M = L_0 + N_0,$$

where L_0 has low-rank and N_0 is a small perturbation matrix. Classical Principal Component Analysis (PCA) seeks the best (in an ℓ_2 sense) rank-k estimate of L_0 by solving

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

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

B. Message

The method that is going to be used is tractable convex optimization. Let $\|M\| := \sum_i \sigma_i(M)$ denote the nuclear norm of the matrix M, that is, the sum of the singular values of M, and let $|M|_1 = \sum_{ij} |(M_{ij})|$ denote the 1-norm of M seen as a long vector in $R^{n_1 \times n_2}$. Then we will show that under rather weak assumptions, the Principal Component Pursuit (PCP) estimate solving

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

exactly recovers the low-rank L_0 and the sparse S_0.Theoretically, this 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.

II. Understanding

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. Write the singular value decomposition of $L_0 \in R^{n_1 x n_2} as L_0 = U \sum V^* = \sum_{i=1}^{r} \sigma_i u_i v_i^*$,where r is the rank of the matrix, $\sigma_1, ..., \sigma_r$are the positive singular values, and $U = [u_1, ..., u_r], V = [v_1, ..., v_r]$ are the matrices of left- and right-singular vectors. Then, the incoherence condition with parameter μ states that

$$max_i \|U^* e_i\|^2 \leq \frac{\mu r}{n_1},$$
$$max_i \|V^* e_i\|^2 \leq \frac{\mu r}{n_2}$$

and

$$\|UV^*\|_\infty \leq \sqrt{\frac{\mu r}{n_1 n_2}}.$$

Here, $\|M\|_\infty = max_{i,j}|M_{ij}|$, that is, is the ℓ_∞ norm of M seen as a long vector. Note that since the orthogonal projection P_U onto the column space of U is given by $P_U = UU^*$, is equivalent to $max_i \|P_U e_i\|^2 \leq \mu r/n_1$, and similarly for PV.

III. Theorem

Suppose L_0 is n x n. Fix any n x n matrix $\sum$ of signs. Suppose that the support set ω of S_0 is uniformly distributed among all sets of cardinality m, and that $sgn([S_0]_{ij}) = \sum$ for all (i, j) $\in \omega$. Then, there is a numerical constant c such that with probability at least $1 - cn^{-10}$ (over the choice of support of S_0), Principal Component Pursuit with $\lambda = 1/\sqrt{n}$, returns exact low-rank and sparse matrix provided that

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

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

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

.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 $\lambda = 1/\sqrt{n_{(1)}}$.

IV. Algorithm

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

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
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 $\delta = 10^{-7}$

V. 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. In the application, there is 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.

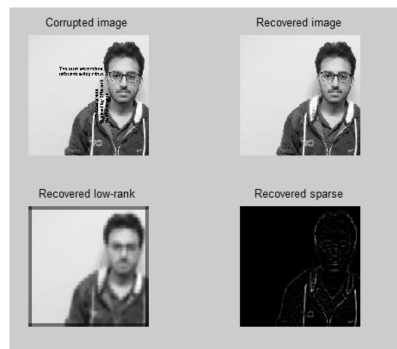

Fig. 1. L_0 and S_0 recovered from corrupted data

As shown above from the corrupted image M having dimons 256×256, thus $M \epsilon R^{256x256}$ we have successfully recovered L_0 and have recorded speculations and image shadowing in S_0. The program took 25.02 $seconds$ to converge with 318 iterations. The rank of low-rank matrix ; **rank(L)= 30** which suggests it is indeed a low rank. The cardinality of set of the sparse matrix S_0 is **card(S)= 222796**. The error rate observed was 4.035123.

VI. Acknowledgement

The discussion here is solely based on the work done on Robust PCA by Emmanuel J. Candes, Xiaodong Li, Yi Ma and John Wright, the simulation also follows the idea of the same literature.

VII. REFERENCES

[1] BASRI, R.,AND JACOBS, D. 2003. Lambertian reflectance and linear subspaces. IEEE Trans. Patt. Anal. Mach. Intel. 25, 2, 218-233.
[2] BECK, A.,AND TEBOULLE, M. 2009. A fast iterative shrinkage-thresholding algorithm. SIAM J. Imag. Sci. 2, 1, 183-202.