Compendium of Robust Principal Component Analysis

Ashutosh Kakadiya - 1401075
ashutosh.k.btechi14@ahduni.edu.in

School of Engineering and Applied Science, Ahmedabad University

***Abstract*—This report describe the implementation of the base paper "Robust Principal Component Analysis" by EMMANUEL J. CANDE'S and XIAODONG LI, Stanford University, YI MA, University of Illinois at Urbana-Champaign, Microsoft Research Asia, JOHN WRIGHT-Microsoft Research Asia. The main objective of this paper is to recover low-rank and sparse matrix from corrupted data matrix by Principal Component Pursuit.Algorithm that describe here is a special case of Augmented Lagrange Multiplier Algorithms. Video Surveillance, Face Recognition, Ranking and collaborative filtering are some applications that can be modeled as Low rank plus a sparse contribution.**

Keywords*—Convex optimization, duality, nuclear norm, principal component pursuit, robust regression.

I. Introduction

This document is a report of studying and implementation of the research paper on Robust Component Analysis. The main objective of that paper is to recover both L_0 a low rank matrix and S_0 a sparse matrix, which are obtain from decomposition of given large data matrix. Even when some entries in matrix S_0 arbitrarily corrupted and some entries are missing. This problem has been solved efficiently by using one of convex optimization method called principal component pursuit. The algorithm described in base paper is a special case of augmented Langrange multiplier algorithm, known as alternating directions methods. We can model many applications which the large data sets can be represent as a low rank plus sparse rank distribution like Video Surveillance (background and foreground separation), Face recognition (removing shadows or facial expressions).

II. Robust Principal Component Analysis (PCA)

- The given data matrix $M \in R^{n_1 X n_2}$ that would like to split into
$$M = L_0 + S_0$$
where L_0 is low rank and S_0 is sparse.
- Here Principal Component Pursuit problem is :
$$\begin{gathered}\text{minimize } ||L||_* + \lambda||S||_1 \\ \text{subject to } M = L_0 + S_0\end{gathered}$$
where $L, S \in R^{n_1 X n_2}$ and data matrix $M \in R^{n_1 X n_2}$

- The parameter $\lambda > 0$ enables a tradeoff between rank and sparsity.
- $||L||_* = \sum_{i=1}^{r} \sigma_i(L)$ is the nuclear norm[Appendix A].
- $||S||_0 = \sum_{i,j} S_{ij}$ is the ℓ_1-norm[Appendix A] of the matrix S thought as a vector.
- value of $\lambda = 1/\sqrt{n}$ (Maximum of matrix dimention).

Since, the main focus is to reduce the rank of given data matrix for less time complexity the best rank-k estimate of low rank matrix L_0 by solving
$$\begin{gathered}\text{minimize } ||M - L||_2 \\ \text{subject to } rank(L) \leq k\end{gathered}$$
This problem can be solved efficiently by Singular Value Decomposition by eliminating low magnitude singular values. Here $M - L$ is assumed as Gaussian Noise.

III. Assumptions

- Low rank matrix L_0 has to be not sparse, because of smaller value of μ.
- S_0 has uniformly distributed sparsity pattern.
- From matrix completion problem, we can write singular vectors of low rank component $L_0 \in R^{n_1 x n_2}$as
$$L_0 = U\Sigma U^* = \sum_{i=1}^{r} \sigma_i u_i v_i^*$$
where r is the rank of matrix of the positive singular values, and U and U are the matrices of left and right singular vectors.

IV. Supporting Theorem (Main Result)

$$\begin{gathered}\text{minimize } ||L||_* + \lambda||S||_1 \\ \text{subject to } M = L_0 + S_0\end{gathered}$$

- PCP recovers perfectly low-rank and sparse component under suitable assumptions.
  - rank(L_0) is small : $\leq O(n/(logn)^2)$
  - S_0 is reasonably small : $\leq O(n^2)$ nonzero entries.
- It works with fix universal value of $= 1/\sqrt{n}$. In rectangular case, $= 1/\sqrt{max(n_1, n_2)}$.
- **Theorem 1.1** This theorem only works for square matrix.
  - L_0 is nxn and S_0 is uniformely distributed among all sets of cardinality m. Then there is a numerical constant $c$such that probability at least $1 - cn^{-10}$ PCP with $= 1/\sqrt{n}$ is exact $\hat{L} = L_0$ and $\hat{S} = S_0$, provided that

$$rank(L_0) \le \binom{n}{k} \frac{n}{(logn)^2}, \qquad m \le \rho_s n^2$$

ρ_r and ρ_s are positive numerical constants.

V. Algorithm

Proposed algorithm is a special case of a general class of Lagrange multiplier (Alternating Directions method).We can solve the minimization constrain problem by the series of non constrain optimization problem.

Algorithm :

1. **initialize**: $S_0 = Y_0, \mu > 0$
2. **while** not converged **do**:
3. compute $L_{k+1} = D_{1/\mu}(M - S_k + \mu^{-1} Y_k)$;
4. compute $S_{k+1} = S_{/\mu}(M - L_{k+1} + \mu^{-1} Y_k)$;
5. compute $Y_{k+1} = Y_k + \mu(M - L_{k+1} - S_{k+1})$;
6. **end while**
7. **output:** L,S.

Here, augmented Lagrangian function $l(L, S, Y) = ||L||_* + \lambda ||S||_1 + (Y, M - L - S) + \mu/2 ||M - L - S||_F^2$ is minimizing iteratively by setting

$$(L^{(k)}, S^{(k)}) = \underset{(L,S)}{\text{arg min}}\ \ell(L, S, Y^{(k)})$$

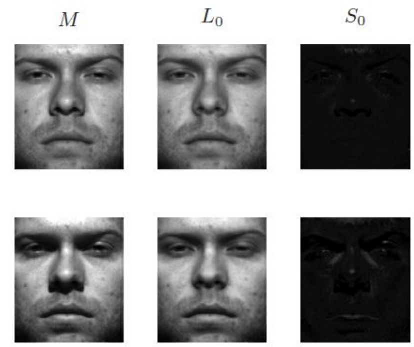
$Y$is Lagrange multiplier matrix which is updated by μ in every iteration until desired solution is not achieved. μ is a tunning parameter which determines the convergence rate of the algorithm.

VI. Applications

1) Surveillance video :
   - Given continuous data frames of given video, we want to extract out the non-stationary activity from background.
   - If we treated the given data frame as matrix D_0 then low-rank component would be the stationary objects as low-rank matrix $L_0$and S_0 sparse matrix would be the moving object (walking person at mall).

(a) Original frames (b) Low-rank $\hat{L}$ (c) Sparse $\hat{S}$

   - sequence of 200 frames each with resolution $176X144$.
2) Removing Face Illumination :
   - Well-aligned face images of a person under different varying illumination is closed to low-dimensional linear subspace.
   - From a set of images which are slightly corrupted by sparse errors, we can retrieve clear images by PCA for various face recognition applications.



   - Yale B database, $192X168$ images of a subject under 58 different illuminations.

VII. Conclusion

Under certain assumptions and λ value, we can efficiently recover Low rank and sparse matrix from given corrupted data by proposed ALM method.Time complexity(Convergence rate) is too high, but in many other journals[6] methods are described with suitable value of λ or speeding up the SVD, we can significantly reduce time complexity for Real Time Principal Component Pursuit.

Appendix A
Key Terms

1) Nuclear Norm.
   - The Nuclear norm is sum of all singular values of a matrix.
   - $||A||_1 = \sum_{i=1}^{n} \sigma_i(A)$.
2) ℓ_1 norm
   - The one norm of given vector $||v||$ is defined as the sum of absolute values of its components:
   - $||V||_1 = \sum_{i=1}^{n} |v_i|$

References

[1] E. J. Cand‘es, X. Li, Y. Ma, and J. Wright, ”Robust Principal Component Analysis?,” 2009. [Online]. Available: http://statweb.stanford.edu/ candes/papers/RobustPCA.pdf. Accessed: Jan. 31, 2017.

[2] E. J. Candes, X. Li, and J. Wright, ”Robust principal component analysis,” in Wikipedia, Wikimedia Foundation, 2017. [Online]. Available: https://en.wikipedia.org/wiki/Robust principal component analysis. Accessed: Jan. 27, 2017.

[3] S. P. Boyd and L. Vandenberghe, Convex optimization. Cambridge, UK: Cambridge University Press, 2004.

[4] University of Illinois, ”Low-rank matrix recovery and completion via convex optimization,” in http://perception.csl.illinois.edu/, 2010. [Online]. Available: http://perception.csl.illinois.edu/matrix-rank/references.html. Accessed: Jan. 28,31, 2017.

[5] Z. Lin, M. Chen, L. Wu, and Y. Ma, ”The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices,” p. 20.

[6] G. Pope, M. Baumann, C. Studer, and G. Durisi, ”Real-Time Principal Component Pursuit,” IEEE.