Robust Principal Component Analysis

Jainil Vachhani

***Abstract*— This paper talks about the separation of a low rank component matrix and sparse matrix from a data matrix created by superposition of a low rank component and a sparse component. The paper uses a program known as Principal Component Pursuit which maximizes a weighted combination of the nulcear form. The paper provides an optimization algorithm, and provides a solution where a fraction of the matrix's entries are missing or are corrupted. The paper then concludes by providing present application in the filed of video surveillance, where the algorithm 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 face.**

I. INTRODUCTION

A. Motivation

In case of Big Data, high dimensionality is desired to be avoided, for which Principal Component Analysis(PCA) is used. The basic concept behind Principal Component Analysis is that all data points lie near a low dimensional subspace.
A large matrix can be decomposed as follows:

$$M = L_0 + S_0,$$

where, L_0 has a low rank component and S_0 is sparse component.The assumption here is that the data all lie near some low-dimensional subspace. Mathematically, it can be expressed as:

$$M = L_0 + N_0,$$

where, N_0 is a small perturbation matrix. Classical PCA looks for the best rank-k estimate of L_0 by solving

$$minimize||M - L|| \quad and \quad subject \quad to \quad rank(L) \leq k$$

The Robust PCA approach aims at recovering Low rank Matrix L from highly corrupted as well as missing entries from data matrix M. Also, it is assumed that the sparsity pattern of the sparse component is selected uniformly at random.

B. Applications

In the following applications the data can be modeled as a low rank matrix plus a sparse contribution.

- Video Surveillance - Given a sequence of surveillance video frames, we often need to identify activities that stand out from the background.
- Face Recognition - Images of a convex, Lambertian surface under varying illuminations span a low-dimensional subspace.
- Latent Semantic Indexing - Web search engines often analyze and index the content of an enormous corpus of document.
- Ranking and Collaborative Filtering - A document-versus-term matrix M whose entries typically encode the relevance of a term to a document such as the frequency it appears in the document. This idea is known as TFIDF(Term Frequency Inverse Document Frequency).

C. Main Result

High dimensional data have a low intrinsic dimensionality, which means that they lie on the same dimensional subspace. Suppose L_0 is $n * n$ matrix. Fix any $n * n$ matrix Σ of signs. Suppose that the support set Ω of S_0 0 is uniformly distributed among all sets of cardinality m, and that sgn[S_0] = Σ_{ij} for all (i,j) ϵ Ω

D. Implications for Matrix Completion from Grossly Corrupted Data

Suppose L_0 is $n*n$ and Ω is uniformly distributed among all sets of cardinality m. Each observed entry is corrupted with probability τ independently of others.Then, there exists a constant c such that with probability at least $1 - cn^{-10}$, *Principal Component Pursuit* with $\lambda = 1/\sqrt{0.1n}$ is exact provided that,

rank(L_0) $\leq \rho\ \nu^{-1}$ (log n^{-2}).

So,perfect recovery from incomplete and corrupted entries is possible by convex programming. And the above result proves that matrix completion is stable versus gross errors. However, if = 0, then we have pure matrix completion problem.

II. ARCHITECTURE OF THE PROOF

In this section, we provide the key steps for the proof of the above mentioned theorem.

A. An elimination theorem

Suppose the solution with input data $M_0 = L_0 + S_0$ is unique and exact, and consider $M_0^1 = L_0 + S_0^1$ where S_0^1 is a trimmed version S_0. Then, the solution with input M_0^1 is exact as well.

B. De-randomization

Suppose L_0 obeys the conditions of Theorem 1 and that the locations of the nonzero entries of S_0 follow the Bernoulli model, and the signs of S_0 0 are independent and identically distributed. Then, , if the PCP solution is exact with high probability, then it is also exact with at least the same probability for the model in which the signs are fixed and the locations are sampled from the Bernoulli model with parameter ρ_s
This theorem is convenient because to prove our main result, we only need to show that it is true in the case where the signs of the sparse component are random.

C. Dual Certificates

Assume that $|P_\Omega P| < 1$. With the standard notations, (L_0, S_0) is the unique solution if there is a pair (W, F) obeying
$UV + W = \lambda(sgn(S_0) + F)$

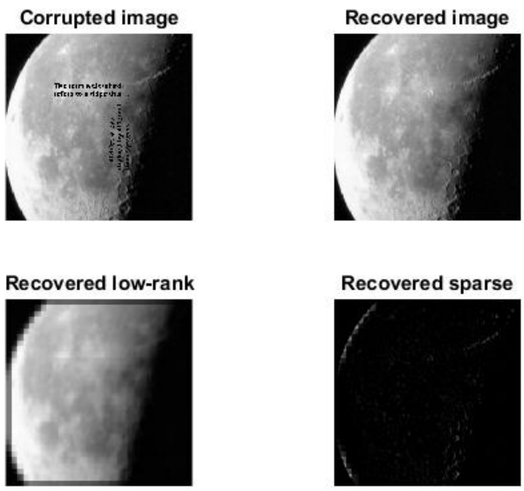
Assume $|P_\Omega P| < 1/2$ and $\lambda < 1$. Then with the same notation, (L_0, S_0) is the unique solution if there is a pair (W, F) obeying
$UV + W = \lambda(sgn(S_0) + F + P_\Omega D)$

D. Key Lemmas

Assume that S_0 is supported on a set Ω and that the signs of S_0 are independent and identically distributed symmetric. Then, the matrix W^s obeys :
$|W^s| < 1/4$ and
$|P_\Omega W^s|_\infty < \lambda/4$

III. NUMERICAL ANALYSIS

In this section, we perform numerical experiments corroborating our main results and suggesting their many applications in image and video analysis.



CONCLUSION

From the given constraints, we conclude the possibility to recover a low rank and sparse component individually, from a data matrix. We obtain solution, through a convex program called Principal Component Analysis and this method has various real world applications like Video Surveillance, Face Recognition etc.

REFERENCES

[1] John Wright, Yigang Peng, Yi Ma, Arvind Ganesh, Shankar Rao. Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices by Convex Optimization. Springer-Verlag, New York, New York, 1986.
[2] http://perception.csl.illinois.edu/matrix-rank/samplecode.html
[3] https://github.com/dlaptev/RobustPCA
[4] E. J. Cand‘es, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inform. Theory, 52(2):489509, 2006.