Robust Principal Component Analysis

Palash Hariyani
SEAS
1401101

***Abstract*—The paper deals with a condition in which it tries to extract the individual components of a data matrix which is formed with a low-rank component and a sparse component. The paper achieves its goal under certain conditions where it is possible to exactly recover both the low-rank and sparse components by solving aver convenient convex program called Principal Component Pursuit. It creates a possibility of a principled approach to robust principal component analysis since our methodology and results assert that one can recover the principal components of a data matrix even though a positive fraction of its entries are arbitrarily corrupted. The paper then discusses an algorithm for solving the optimization problem and present applications in the area of video surveillance, where the method 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 the images of faces.**

I. Introduction

Today we face many problems of managing large data in various fields such as engineering and science. This creates a challenge in various sectors such as image and video processing. The data matrix is basically of very huge dimensions. In order to curb this problem, we can exploit the fact that such data have low intrinsic dimensionality. This takes one of the 3 following cases :

- They lie on some low-dimensional subspace.
- They are sparse in some basis.
- They lie on some low-dimensional manifold.

We assume that the data lie on some low-dimensional subspace. Thus, we can split the matrix into a low rank matrix(L_0) and a perturbation matrix(N_0).

$$M = L_0 + N_0$$

For this we need to minimize $||M-L||$ subject to rank$(L) \leq k$. This is efficient when noise is small, independent and identically distributed Gaussian.

II. Applications

Robust PCA has found many application in several fields -

A. Face Recognition

Images of convex, Lambertian surface under varying illuminations span a low-dimensional subspace. Images of a human face can be well approximated by a low-dimensional subspace. However, realistic face images often suffer from self-shadowing, specularities or starvations in bright performance.

B. Latent Semantic Indexing

Web search engines often need to analyze and index the content of an enormous corpus of documents. In this method, the basic idea is to gather a document-versus-term matrix M whose entries typically encode the relevance of a term to a document such as the frequency with which it appears in the document.

C. Video Surveillance

Given a sequence of surveillance video frames, we often need to identify activities that stand out from the background. If we stack the video frames as columns of a matrix M, then low rank component L_0 naturally corresponds to the stationary background and the sparse component S_0 captures the moving objects in the foreground.

D. Ranking and Collaborative Filtering

: In today's world, predicting the interests of users is gaining popularity in the fields of commerce and advertisement. Companies such as Amazon and Netflix now routinely gather user choices and rankings for various products and update their database accordingly. The problem faced here is to predict the rankings of a user based on incomplete rankings for a particular product. A small portion of the data can be even corrupted. We need to continuously complete the data matrix and get rid of errors. Here, we need to infer a low-rank matrix from a set of incomplete and corrupted entries. The results of this paper can be extended to solve this problem.

III. About Robust PCA

We are aware that PCA is the most widely used tool today for data analysis and dimensionality reduction. But it is quite risky and vulnerable as a singly grossly corrupted entry in M could render the estimated L arbitrarily far from the true L_0. Since these errors are ubiquitous in the modern machines, a need to make the PCA method robust was required. Unfortunately, none of the approaches yields a polynomial-time algorithm with strong performance guarantees under broad conditions. The problem which is stated in the paper is considered an idealized version of robust PCA, in which we aim to recover a low-rank matrix L_0 from highly corrupted measurements $M = L_0 + S_0$. Here, the entries in S_0 can have arbitrarily large magnitude and their support is assumed to be sparse but unknown. The number of unknowns to infer for L_0 and S_0 is twice as many as the given measurements in $M \in R^{n_1 X n_2}$. This problem can be solved by 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. We can thus, show that under weak assumptions, the Principal Component Pursuit (PCP) estimate solving : Minimize $||L||^* + \lambda||S||_1$ subject to $L + S = M$.

This recovers the low rank matrix L_0 and the sparse S_0.

IV. Theorems

A. THEOREM 1.1:

Fix any nXn matrix $\sum$ of signs. Suppose that the support set Ω of S_0 is uniformly distributed among all sets of cardinality m, and that $sgn([S_0]_ij) = \sum_{ij} \Omega$ 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), PCP with $\lambda = 1/\sqrt{n}$ is exact, i.e., $L = L_0$ and $S = S_0$ provided that

$$rank(L_0) \leq \rho rn\mu - 1(logn) - 2andm \leq \rho sn_2$$

In this equation, ρr and ρs are positive numerical constants. In the general rectangular case, where $L_0 isn_1 * n_2$, PCP with $\lambda = 1/sqrtn_1$ succeeds with probability at least $1-cn(1)-10$, provided that $rank(L_0) \leq \rho rn(2)\mu - 1(logn(1)) - 2andm \leq \rho sn_1n_2$.

The above theorem states that matrices L_0 whose singular vectors or principal components are reasonably spread can be recovered with probability nearly one from arbitrary and completely unknown corruption patterns. This also works for large values of the rank, on the order of $n/(logn)^2$ when μ is not too large.

B. THEOREM 1.2:

We assume that Ω_obs is uniformly distributed among all sets of cardinality m obeying $m = 0.1n_2$. Suppose for simplicity, that each observed entry is corrupted with probability τ independently of the others. Then, there is a numerical constant c such that with probability at least $1 - cn^-10$,PCP with $\lambda = 1/\sqrt{0.1n}$ is exact, that is, $L = L_0$, provided that,

$$rank(L_0) \leq \rho rn\mu - 1(logn) - 2and\tau \leq \tau_s$$

. In this equation, ρr and τ_s are positive numerical constants. For general, n_1Xn_2 rectangular matrices, PCP with $\lambda = 1/sqrt(0.1n_1)$ succeeds from $m = 0.1n_1n_2$ corrupted entries with probability at least $1 - cn_(1)^-10$, provided that $rank(L_0) \leq \rho rn(2)\mu - 1(logn(1)) - 2$.

If all the entries are available, that is, $m = n_1n_2$, then it turns out to be Theorem 1.1. If $\tau = 0$, we have a pure matrix completion problem from about a fraction of the total number of entries and the theorem guarantees perfect recovery as long as r obeys, which, for large values of r, matches the strongest results available.

V. Dual Certificates

We introduce a simple condition for the pair (L_0, S_0) to be the unique solution to PCP. These conditions, given in the lemma(as in the paper), are stated in terms of a dual vector, the existence of which certifies optimality.

Lemma : Assume that $||P_\Omega P_T|| < 1$.With the standard notations, (L_0, S_0) is the unique solution if there is a pair (W,F) obeying

$$UV^* + W = \lambda(sgn(S_0) + F)$$

, with $p_T W = 0, ||W|| < 1, p_\Omega F = 0and||F||_{\inf} < 1$.

VI. Algorithms

The theorem above shows that incoherent low-rank matrices can be recovered from non-vanishing fractions of gross errors in polynomial time. For small problem sizes, PCP can be performed sing off-the-shelf tools such as interior pint methods. Despite the superior converge rates, interior point methods are typically limited to small problems, say $n < 100$, due to O^{n6} complexity of computing a step direction.

A. Principal component pursuit by alternating directions

initialize : $S_0 = Y_0 = 0, \mu > 0$
while not converged do
compute $L_{h+1} = D_{1/\mu}(M - S_h + \mu^{-1}Y_h)$;
compute $S_{k+1} = S_{\lambda/\mu}(M - L_{h+1} + \mu^{-1}Y_h)$;
compute $Y_{h+1} = Y_h + \mu(M - L_{h+1} – S_{h+1})$;
end while
output : L,S.

VII. Conclusion

The analysis leads to the conclusion that a single universal value of λ, namely $\lambda = 1/\sqrt{n}$, works with high probability for recovering any low-rank, incoherent matrix. In Chandrasekaran et al.[2009], the parameter λ is data-dependent and may have to be selected by solving a number of convex programs. The distinction in the results is a consequence of differing assumptions about the origin of the data matrix M. We regard the universality of λ in the analysis as an advantage, since it may provide useful guidance in practical circumstances where the generative model for M is not completely known.

VIII. References

- Robust Principal Component Analysis (Emmanuel J. Candes and Xiaodong LI, Stanford University)
- CAND‘ES, E. J., AND RECHT, B.2009. Exact matrix completion via convex optimzation. Found. Comput. $Math$.9,717–772.CAND‘ES, E. J., ROMBERG, J.,
- $https : //www.youtube.com/watch?v = DK8RTamIoB8$