Robust Principal Component Analysis

Report by Mihir Gajjar
School of Engineering and Applied Science

***Abstract*—In this paper, it is proved that, given a data matrix which is the superposition of a low-rank component and a sparse component, under some suitable assumptions it is possible to recover both the low-rank and the sparse components exactly by solving a very convenient convex program called Principal Component Pursuit even though there are incomplete and corrupted entries in the given matrix.An algorithm for solving this optimization problem is discussed, and its applications are presented.**

***Index Terms*—Principal Components, Convex optimization, Nuclear-norm minimization, l_1-norm minimization, Low-rank matrices, Sparsity,robustness vis-a-vis outliers.**

I. Introduction

A. Motivation

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

$$M = L_0 + S_0, \tag{1}$$

where, L_0 has low rank and S_0 is sparse. Here, Both components are of arbitrary magnitude. We do not know the characteristics of L_0 and S_0. To reduce the effect of dimensionality and scale, the assumption is that all the data lie near some low-dimensional subspace. Mathematically,

$$M = L_0 + N_0, \tag{2}$$

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.

This problem can be efficiently solved via the singular value decomposition (SVD) and gives optimal results when the noise N_0 is small and independent and identically distributed Gaussian.

B. Robust PCA

PCA is widely used for data analysis and dimensionality reduction. But, in PCA, a single grossly corrupted entry in the data matrix M could change the estimated $\hat{L}$ arbitrarily far from the true L_0. And, gross errors are now present in every modern application. The problem we study here can be considered as an idealized version of Robust PCA, where we want to recover the low-rank matrix L_0 from highly corrupted measurements M = L_0 + S_0. Unlike the small noise term N_0 in classical PCA, The entries in S_0 can have arbitrarily large magnitude, and their support is assumed to be sparse but unknown.

II. Assumptions

(a) The low rank component L_0 is not sparse: If 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, we cannot decide whether it is low rank or sparse.
(b) The sparsity pattern of the sparse component is selected uniformly at random: If all the nonzero entries of S occur in a few columns, then the sparse matrix has low-rank. If the first column of S_0 is the opposite of that of L_0, and all the other columns of S_0 vanish, then we would not be able to recover L_0 and S_0 by any method since M = L_0 + S_0 would have a column space included in that of L_0.

III. Main Results

Theorem 1.1 says that under these minimal assumptions, PCP perfectly recovers the low rank and the sparse components of matrices L_0 whose singular vectors or principal components are reasonably spread with probability nearly one from arbitrary and completely unknown corruption patterns (as long as these are randomly distributed).Under the assumptions of the theorem, the Principal Component Pursuit (PCP) estimate solving

Minimize: $||L||_* + \frac{1}{\sqrt{n_{(1)}}}||S||_1$
where, $n_{(1)}$ = max(n1, n2),

always returns the correct answer. The choice $\lambda = \frac{1}{\sqrt{n_{(1)}}}$ is universal.(According to the proof, there are a whole range of correct values of lambda, from that a simple value is selected.) Now, We wish to recover L_0 from the available incomplete and corrupt entries of L_0. Theorem 1.2 proves that matrix completion is stable vis-a-vis gross errors.Let P_Ω be the orthogonal projection onto the linear space of matrices supported on $\Omega \subset$ [n1] $\times$ [n2] and imagine we only have available a few entries of L_0 + S_0, which we conveniently write as,

$$Y = P_{\Omega_{obs}}(L_0 + S_0) = P_{\Omega_{obs}} L_0 + S_0'; \tag{3}$$

that is, we see only those entries (i, j) $\in \Omega_{obs} \subset [n1] \times [n2]$. We propose recovering L_0 by solving the following problem:

Minimize: $||L||_*$ + $\lambda||S||_1$
Subject to: $P_{\Omega_{obs}}$(L + S) = Y.

Among all decompositions matching the available data, Principal Component Pursuit finds the one that minimizes the weighted combination of the nuclear norm, and of the l_1 norm. In short, perfect recovery from incomplete and corrupted entries is possible by convex optimization. A more careful study is likely to lead to a stronger version of Theorem 1.2 .

IV. Review of Literature

The matrix completion problem is that of recovering a low-rank matrix from only a small fraction of its entries.The ideas depart from the literature on matrix completion on the following fronts:

1. The results are of a different nature. Here, a given data matrix is separated in to low rank and sparse components.
2. Here, we have a fraction of entries available, but do not know which one,while the other is not missing but corrupted. The algorithm simultaneously detects the corrupted entries, and perfectly fits the low-rank component to the remaining entries that are considered reliable. In this sense, this methodology goes beyond matrix completion.
3. A new technique is introduced that allows to fix the signs of the nonzero entries of the sparse component. This technique is important because assuming independent signal signs may not make much sense for many practical applications when the involved signals can all be nonnegative.

V. Algorithm

The convex PCP problem is solved using an Augmented Lagrange Multiplier (ALM) algorithm introduced in Lin et al.[2009a] and Yuan and Yang [2009]. The ALM method operates on the augmented Lagrangian:

$$l(L,S,Y) = ||L||_* + \lambda||S||_1 + \langle Y, M-L-S\rangle + \frac{\mu}{2}||M-L-S||_F^2 \tag{4}$$

For matrices X, let D_τ(X) denote the singular value thresholding operator given by D_τ(X) = U$S_\tau(\sum)V^*$, where X = U$\sum V^*$ is any singular value decomposition.

$$\underset{L}{argmin} l(L,S,Y) = D_{\frac{1}{\mu}}(M - S + \mu^{-1}Y). \tag{5}$$

Similarly, Let S_τ : R $\longrightarrow$ R denote the shrinkage operator $S_\tau[x]$ = sgn(x)max($|x|$ - τ ,0) , and extend it to matrices by applying it to each element.

$$\underset{S}{argmin} l(L,S,Y) = S_{\frac{\lambda}{\mu}}(M - L + \mu^{-1}Y). \tag{6}$$

And then the Lagrange multiplier matrix is updated via
Y_{k+1} = Y_k + $\mu(M - L_k - S_k)$.

ALGORITHM: (Principal Component Pursuit by Alternating Directions [Lin et al. 2009a;Yuan and Yang 2009])

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

Hence, first we minimize *l* with respect to *L* (fixing *S*), then minimize *l* with respect to *S* (fixing *L*), and then finally update the Lagrange multiplier matrix *Y* based on the residual *M* - *L* - *S*. In this algorithm the dominant cost of each iteration is computing L_{k+1} via singular value thresholding. The rank of the iterates often remains bounded by rank(L_0) throughout the optimization. We choose $\mu = \frac{n_1 n_2}{4||M||_1}$ as suggested in Yuan and Yang [2009]. The algorithm is terminated when $||M - L - S||_F \leq \delta||M||_F$, with $\delta = 10^{-7}$.

VI. Applications

In practice, we need to solve the low-rank and sparse decomposition problem for matrices of extremely high dimension and under broad conditions. Depending on the application either the low-rank component or the sparse component could be the object of interest.

1. Video Surveillance: Used for separating the foreground and the background.L_0 corresponds to the stationary background and the sparse component S_0 captures the moving objects in the foreground.
2. Face Recognition: The images of a humans face can be well-approximated by a low-dimensional subspace. This method is able to effectively remove such defects in face images.

Other applications include Latent Semantic Indexing and Ranking and Collaborative Filtering.

VII. Conclusion

Under quite broad conditions, using convex programming, one can disentangle the low-rank and sparse components exactly even when there are both incomplete and corrupted entries. The convex optimization method Principal Component Pursuit gives efficient and accurate results. Future experiments should investigate whether either or both of these assumptions can be relaxed and also on developing algorithms that can be easily implemented on the parallel and distributed computing infrastructures.

A. Notation

$||M||$ denotes the operator norm or 2-norm of the matrix M. $||M||_F$ denotes the Frobenius norm of the matrix M. $||M||_*$ denotes the nuclear norm of the matrix M. $||M||_1$ denotes the l_1-norm of the matrix M. $||M||_\infty$ denotes the l_∞-norm of the matrix M. e_i is the ith canonical basis vector in Euclidean space (the vector with all entries equal to 0 but the ith equal to 1). $n_{(1)}$ = max(n_1, n_2) and $n_{(2)}$ = min(n_1, n_2), where $n_1 \times n_2$ is the dimension of the matrix. Calligraphic letters are used to manipulate linear transformations that act on the space of matrices. (eg. $P_\Omega X$)

References

[1] Candès, E. J., Li, X., Ma, Y., and Wright, J. 2011. Robust principal component analysis? J. ACM 58, 3, Article 11 (May 2011), 37 pages. DOI = 10.1145/1970392.1970395 http://doi.acm.org/10.1145/1970392.1970395

[2] Candès, E.J. & Recht, B. Found Comput Math (2009) 9: 717. doi:10.1007/s10208-009-9045-5

[3] [CIMAT2013].(2013, Dec 10).*Conferencia "Robust principal component analysis", por Emmanuel Candes*.[Video File]. Retrieved from https://www.youtube.com/watch?v=DK8RTamIoB8

[4] Meng F, Yang X, Zhou C (2014) The Augmented Lagrange Multipliers Method for Matrix Completion from Corrupted Samplings with Application to Mixed Gaussian-Impulse Noise Removal. PLoS ONE 9(9): e108125. doi:10.1371/journal.pone.0108125