FRPCA: Fast Robust Principal Component Analysis

Alaa E. Abdel-Hakim
Electrical Engineering Department
Assiut University
Assiut, Egypt
alaa.hakim@ieee.org

Motaz El-Saban
Microsoft Research
Cairo Innovation Lab
Cairo, Egypt
motazel@microsoft.com

Abstract

While the performance of Robust Principal Component Analysis (RPCA), in terms of the recovered low-rank matrices, is quite satisfactory to many applications, the time efficiency is not, especially for scalable data. We propose to solve this problem using a novel fast incremental RPCA (FRPCA) approach. The low rank matrices of the incrementally-observed data are estimated using a convex optimization model that exploits information obtained from the preestimated low-rank matrices of the original observations. The evaluation results supports the potential of FRPCA for fast, yet accurate, recovery of the low-rank matrices. The proposed FRPCA boosts the efficiency of the traditional RPCA by multiple hundreds of times, while scarifying less than 1% *of accuracy.*

1 Introduction

The rapid increase of applications, where data usually lie in dimensions up to six-digit figures, mandates exploiting subspace recovery techniques. PCA is considered as one of the well-known successful tools in this aspect [5]. While small-noise distortions have small impact on the estimated principal components, gross distortions, even sparse, may ruin the entire operation of PCA [5, 8]. RPCA [8] is one of the approaches that have been developed to "robustify" the classical PCA. While RPCA demonstrates very good performance for such applications, the existing algorithms for solving RPCA suffer from efficiency problems with scalable input data [7]. This represents a severe limitation for the use of RPCA in applications which exhibit scalable data and real-time requirements, e.g. video background subtraction, tracking, ... etc.

Several algorithms have been developed to solve the RPCA convex optimization problems. For instance, interior point solvers, e.g. [1], have quite good convergence rates. Nonetheless, the complexity of these kinds of algorithms ($O(m^6)$, where m is the dimension of the input data) causes a very difficult scalability problem. First-order methods, e.g. iterative thresholding algorithms for L1-normalization [4, 9], represented a better alternative in terms of scalability. Nevertheless, iterative thresholding is reported to exhibit very slow convergence [8, 7].

Lin, et.al [7] have proposed two algorithms for solving this problem. The first one is a first-order accelerated proximal gradient algorithm. The second is a dual algorithm, which solves the convex optimization problem via a dual model. The dual algorithm does not perform a full SVD computation, and hence can accommodate larger input matrices. As stated in [7], although these two algorithms are good for some applications to be performed in a "reasonable amount of time," they are unsuitable for direct use in applications involving larger matrices.

To overcome this limitation, Lin et.al [6] have presented a faster Augmented Lagrangian Multipliers method, (ALM), for solving the RPCA problem. Two versions of ALM were presented: Exact Augmented Lagrangian Multipliers, (EALM) and Inexact Augmented Lagrangian Multipliers, (IALM). Both algorithms were proven to run about five times faster than the state-of-the-art algorithms, which were presented in [7]. Nonetheless, since the core computation of these algorithms depends on SVD computations, the time complexity of that solution is $O[\min(nm^2, mn^2)]$, where n and m are the dimensions of the input observations. So, the time efficiency of the algorithm is linearly degraded with the increase of n, if $n > m$ and quadratically if $m > n$; which is the most common case, where the number of instances, n, is less than the dimensions of a single instance, m. Therefore, for some specific applications that have real-time characteristics, e.g. background subtraction in video streams, tracking, surveillance applications, ... etc, the time efficiency of these algorithms is unacceptable.

We suggest a mathematical model that takes into consideration the accuracy constraints of the classical RPCA. At the same time, the proposed model integrates abstract information of the offline-recovered low-rank of the in-

Preprint submitted to 21st International Conference on Pattern Recognition.
Received April 7, 2012.

put data rather than including the entire dataset in the online low-rank recovery phase. By using the proposed approach, a considerable improvement in efficiency is achieved with a small, or negligible, decrease in accuracy.

2 Problem Formulation

Assume that high-dimensional observed data, $D \in \mathbb{R}^{m\times n}$, is composed of a low rank term, $A \in \mathbb{R}^{m\times n}$, and an additive error corruption term, $E \in \mathbb{R}^{m\times n}$. The classical low rank recovery problem is formulated as follows [8]: *Given $D = A + E \in \mathbb{R}^{m\times n}$, where A is unknown low rank and E is unknown sparse, recover A.*

To overcome the complexity problem, we propose to reduce n by considering the added "online" observations only for the optimization problem, in addition to some partial information induced from the original input observations. So, the problem of the proposed FRPCA approach is formulated as follows:

FRPCA Problem: Assume that an original "offline" set of input observations $D_{org} \in \mathbb{R}^{m\times n}$ is modeled using the traditional RPCA as $D_{org} = A_{org} + E_{org}$, where A_{org} and E_{org} are low rank and sparse matrices, respectively. D_{org} is augmented by another "online" observation set $D' \in \mathbb{R}^{m\times k}$ to construct an augmented set $D_{aug} \in \mathbb{R}^{m\times(n+k)}$. Given the low rank and sparse terms of the original input data, A_{org} and E_{org}, recover the low rank and sparse terms of the augmented observations, A_{aug} and E_{aug}, which satisfy the condition: $D_{aug} = A_{aug} + E_{aug}$.

3 The FRPCA Model

The solution of the FRPCA problem, i.e. the recovery of A_{aug} and E_{aug}, optimizes for minimum rank of A_{aug} and the maximum sparsity of E_{aug} under the condition: $D_{aug} = A_{aug} + E_{aug}$. At the same time, for complexity purposes, this optimization process should not involve the entire augmented dataset. For this purpose, A_{aug} can be divided into two main portions: $A'_{org} \in \mathbb{R}^{m\times n}$ and $A' \in \mathbb{R}^{m\times k}$, which are the corresponding low rank terms of the original and the incremented observations, respectively. In other words, $A_{aug} = [A'_{org}|A'] \in \mathbb{R}^{m\times(n+k)}$. Under the condition that $k \ll n$, we can assume that $A'_{org} \simeq A_{org}$, and hence $E'_{org} \simeq E_{org}$. In other words, the impact of adding few more observations on the preestimated low rank terms can be neglected. We call this assumption the "*Robustness Assumption*" Eq. 1.

$$A'_{org} \cong A_{org} \quad E'_{org} \cong E_{org} \quad s.t. \quad k \ll n \tag{1}$$

When estimating the low-dimension representation of a number of high-dimensional observations, every instance "contributes" to the solution. So, adding few more instances to the input observations will have minimum impact on the obtained solution of the original data. In practice, as shown in the next validation section, k can grow up to as much as n, or even more, with acceptable deviations from the original values of A_{org} and E_{org}. Thus, the solution of FRPCA comes down to finding A' and E' rather than A_{aug} and E_{aug}, since A'_{org} and E'_{org} are substituted by A_{org} and E_{org}, according to the robustness assumption of Eq. 1.

So, the solution of the FRPCA problem can be formulated as follows:

$$\begin{aligned}(A', E') \quad = \quad & \arg_{A',E'} \min rank(A') + \gamma_1 \parallel E' \parallel_0 \\ & +\gamma_2 f(A_{org}, A') \\ & s.t. D_{aug} = A_{aug} + E_{aug}\end{aligned} \tag{2}$$

where γ_1 and γ_2 are arbitrary coefficients. The constraint of Eq. 2 can be reformulated as follows:

$$D_{aug} = [D_{org}|D'] = [A'_{org}|A'] + [E'_{org}|E'] \tag{3}$$

By applying Eq. 3:

$$D_{aug} = [D_{org}|D'] = [A_{org}|A'] + [E_{org}|E'] \tag{4}$$

Since the first portion of the constraint of Eq. 1, $D_{org} = A_{org} + E_{org}$, has been already achieved when the low-rank terms of the original observations were recovered, the constraint of Eq. 2 can be rewritten as:

$$D' = A' + E' \tag{5}$$

On the other hand, the model of Eq. 2 is highly non-convex [8]. An acceptable approximation to convert the model of Eq. 2 to the convex optimization form is to use the nuclear norm to represent the rank and the L1-norm to replace the zero-norm as a representation of sparsity.

The last term of Eq. 2 represents the constraints of the relationship between the incremented observations and the original ones. This relation is represented by an objective function $f(A_{org}, A')$, which enforces the condition of having the recovered low-rank terms of the incremented observations as similar as possible to the original input observations. In this work, we selected $f(A_{org}, A')$ as shown in Eq. 6. Squared Frobenius norm is a strictly convex function, which complies with the convex optimization problem [3].

$$f(A_{org}, A') = \parallel (\frac{1}{n}\sum_{j=1}^{n} A^j_{org}) * M - A' \parallel^2_F \in \mathbb{R}^{m\times k} \tag{6}$$

where A^j_{org} is the j^{th} column of the original low-rank matrix, A_{org}; and M is a unit row-vector $\in \mathbb{R}^{1\times m}$. Thus,

Preprint submitted to 21st International Conference on Pattern Recognition.
Received April 7, 2012.

the final form of the proposed FRPCA model is shown in Eq. 7

$$
\begin{aligned}
(A', E') \quad = \quad & \arg_{A',E'} \min \parallel A' \parallel_* + \gamma_1 \parallel E' \parallel_1 \qquad (7) \\
& + \gamma_2 \parallel [(\frac{1}{n} \sum_{j=1}^{n} A_{org}^j) * M] - A' \parallel_F^2 \in \mathbb{R}^{m \times k} \\
& s.t. D' = A' + E'
\end{aligned}
$$

By using this solution, the time complexity of the algorithm is reduced to $O(mk^2)$. Given that the number of the new observations $k \ll n$, a large gain, in terms of execution time, is achieved.

4 Robustness Assumption Validation

We validate the robustness assumption using two sets of data: synthetic and real data. The validation procedure that we follow depends on estimating the low-rank matrices of the two datasets using the traditional RPCA. The two state-of-the-art algorithms, EALM and IALM [6] are used for solving the RPCA problem. The problem is resolved after adding new input observations. The effect of the added instances on the recovered low-rank matrices is investigated.

One hundred 144×192 images were generated from an original gray-bar image. This set of distorted images, D_{org}, is generated by distorting the original gray-bar image such that the pixels inside three randomly-positioned 30×30 boxes are negated. The low rank matrices, A_{org}, are estimated for the entire training set (an offline step). Different distorted observations, in a similar manner to D_{org} distortion, are added gradually to D_{org} to construct an augmented set D_{aug}. This leads to increasing the size of the augmented data set by one per each step. We compare A_{org} with the corresponding columns of A_{aug}, A'_{org}.

Two kinds of error are evaluated for the recovered low-rank matrices, A'_{org}. The first, e_1, is the error measured against the ground truth, i.e. the original images. e_1 is calculated: $e_1 = \frac{\|A'_{org} - I * M\|_F}{\|I * M\|_F}$, where I is the input synthetic gray-bar image stacked in a column vector and M is a unit row-vector $\in \mathbb{R}^{1 \times m}$. The second error value, e_2, is used as a measure of the deviation between A'_{org} and A_{org}, and is given by: $e_2 = \frac{\|A'_{org} - A_{org}\|_F}{\|A_{org}\|_F}$.

For real datasets, we use a 240×320 video sequence [2]. Figure 1 shows the e_2 values, for both EALM and IALM. Although the error values are larger than those of the synthetic case, we can see that for added observations up to 60% of the size of the original observations, the error does not exceed 1%. For synthetic data, adding new elements, of size up to 1.5X of the original observation set, leads to error values below 8×10^{-8}. The

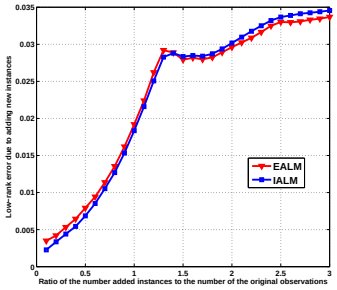


Figure 1. Error curves using the real data.

synthetic validation curves are omitted for space limitations.

5 Evaluation Results

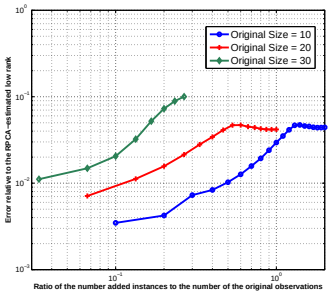
The proposed approach is expected to have positive impact on efficiency, in terms of the execution time. The price paid for this improvement is a degradation in the accuracy. Therefore, the evaluation experiment is designed to highlight these two effects.

Two implementations of the proposed FRPCA are evaluated in this section. The first implementation is the Exact-FRPCA (EFRPCA). In this implementation, the convex optimization problem is solved using Lagrangian Multipliers in a similar algorithm to EALM, but using the FRPCA mathematical model of Eq. 7. Similarly, the second implementation is the Inexact-FRPCA (IFRPCA), which calculates partial singular values of the FRPCA model for the sake of accelerating the recovery process.

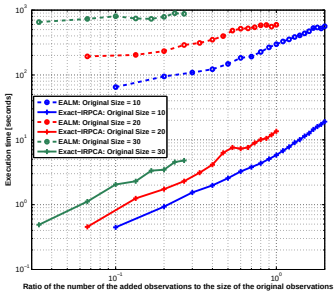
The proposed FRPCA was evaluated using the same two sets of data, which were used in the robustness validation section. EFRPCA and IFRPCA were run on both sets using a Matlab R2010a code that has been executed on an Intel Core 2 Duo CPU E8500 with 1.16GHz and 3.00GB RAM PC. The accuracy and execution time were compared to the fastest state-of-the-art algorithms of RPCA: EALM and IALM.

Figure 2 shows the evaluation results for the real data using the EFRPCA versus EALM. The error values are calculated according to the formula of Eq. **??**. As shown in the figure, the accuracy is decreased with the increase of the original data size. This is expected, since the approximation of the low-rank terms of the original data will be looser. Nonetheless, if the number of the incrementally-added instances is kept small, the ratio of the incremented data to the size of the original data is decreased. This causes a horizontal movement to left along the error curve, i.e. a decrease in the calculated error. At the same time, a larger improvement in the execution time is achieved, as shown in Fig. 2(b). The error values are larger than those of the synthetic case, and this is expected as discussed earlier in the robustness validation

Preprint submitted to 21st International Conference on Pattern Recognition.
Received April 7, 2012.

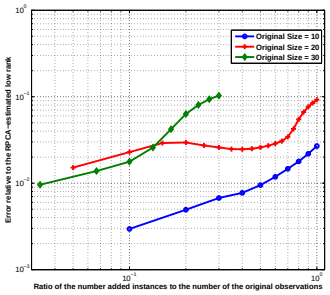


(a) The error results

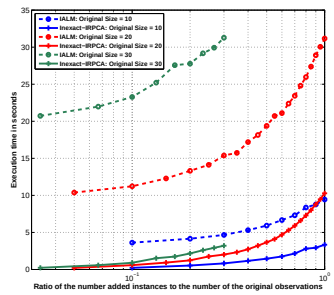


(b) The execution time results

Figure 2. Evaluation results using real data for the EFRPCA vs. EALM.



(a) The error results



(b) The execution time results

Figure 3. Evaluation results using real data of the IFRPCA vs. IALM.

section. However, for the targeted operation range for the size of the added instances to be around 10%, or less, of the size of the original observations, the error values are in the acceptable range of 1-2%. Figure 3 shows similar results for the IFRPCA and IALM.

For synthetic data, up to 30% of incrementally-added observations, the error value did not exceed 0.1%. This speed gain varies depending on the size of the added instances from 10X to 200X. The synthetic detailed evaluation curves are omitted for space limitations.

6 Conclusions and Future Work

In this paper, we proposed a novel approach to fast recovery of low-rank matrices of higher-dimensional data. The proposed approach, FRPCA, uses a convex optimization model that exploits information about the low-rank terms of the observed data, which were calculated so far. FRPCA decreases the time complexity of the traditional RPCA from $O(mn^2)$ to $O(mk^2), k \ll n$. For two experimental datasets, FRPCA improved the efficiency by multiple-hundred times. The accuracy of the recovered low-rank was slightly decreased by less than $1 - 2\%$. For future work, we plan to investigate more objective functions, which relates the optimized low-rank to the low-rank terms of the original observations.

References

[1] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, March 2004.

[2] A. Criminisi. *Database of monocular sequences labelled into foreground and background layers*. `http://research.microsoft.com/en-us/projects/i2i/data.aspx`.

[3] J. Dattorro. *Convex Optimization & Euclidean Distance Geometry*. Lulu.com, July 2006.

[4] E. T. Hale, W. Yin, and Y. Zhang. Fixed-Point Continuation for l1-Minimization: Methodology and Convergence. *SIAM Journal on Optimization*, 19(3):1107–1130, 2008.

[5] I. Jolliffe. *Principal Component Analysis*. Springer Verlag, New York, New York, 1986.

[6] Z. Lin, M. Chen, and L. Wu. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *Analysis*, math.OC:2209–2215, 2009.

[7] Z. Lin, A. Ganesh, J. Wright, L. Wu, M. Chen, and Y. Ma. Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–18, 2009.

[8] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices by convex optimization. In *proceedings of Neural Information Processing Systems (NIPS)*, December 2009.

[9] W. Yin, S. Osher, D. Goldfarb, and J. Darbon. Bregman Iterative Algorithms for L1-Minimization with Applications to Compressed Sensing. *SIAM J. Imaging Sciences*, 1(1):143–168, 2008.