Matrix rigidity and the ill-posedness of Robust PCA and matrix completion

Jared Tanner†, Andrew Thompson‡, Simon Vary∗

∗Mathematical Institute, University of Oxford
Oxford OX1 3DP, UK
†The Alan Turing Institute, British Library
London NW1 2DB, UK
‡National Physical Laboratory
Teddington TW11 0LW, UK

Abstract—Robust Principal Component Analysis (RPCA) [2], [4] and low-rank matrix completion (MC) [13] are extensions of PCA to allow for outliers and missing entries respectively. It is well-known that solving these problems requires a low coherence between the low-rank matrix and the canonical basis. However, the well-posedness issue in both problems is an even more fundamental one: in some cases, both Robust PCA and matrix completion can fail to have any solutions due to the set of low-rank plus sparse matrices not being closed, which in turn is equivalent to the notion of the matrix rigidity function not being lower semicontinuous [10]. An analogy can be drawn to the case of sets of higher order tensors not being closed under canonical polyadic (CP) tensor rank, rendering the best low-rank tensor approximation unsolvable [6] and hence encourage the use of multilinear tensor rank [5].

Robust PCA (RPCA) solves a low-rank plus sparse matrix approximation with the sparse component allowing for few but arbitrarily large corruptions in the low-rank structure; that is, a matrix \( M \in \mathbb{R}^{m \times n} \) is decomposed into a low-rank matrix \( L \) plus a sparse matrix \( S \)

\[
\min_{X \in \mathbb{R}^{m \times n}} \| X - M \|_F, \quad \text{s.t.} \quad X \in \text{LS}_{m,n}(r,s),
\]

where \( \text{LS}_{m,n}(r,s) \) is the set of \( m \times n \) matrices that can be expressed as a rank \( r \) matrix \( L \) plus a sparsity \( s \) matrix \( S \)

\[
\text{LS}_{m,n}(r,s) = \{ L + S \in \mathbb{R}^{m \times n} : \text{rank}(L) \leq r, \| S \|_0 \leq s \}.
\]

It is well known that the minimization in (1) need not have a unique solution without further constraints, such as the singular vectors of the low-rank component being uncorrelated with the canonical basis as quantified by the incoherence condition [3], [13].

Herein we highlight the presence of a more fundamental difficulty: There are matrices for which RPCA and MC can have no solution in that their constituents diverge even while the objective is minimized to zero. This is not because of the ambiguity between possible solutions or lack of information about the matrix, but instead because \( \text{LS}_{m,n}(r,s) \) is not a closed set. Moreover, this is not an isolated phenomenon, as sequences of \( \text{LS}_{m,n}(r,s) \) matrices converging outside of the set can be constructed for a wide range of ranks, sparsities and matrix sizes.

Theorem 1 (LS_{m,n}(r,s) is not closed). The set of low-rank plus sparse matrices \( \text{LS}_{m,n}(r,s) \) is not closed for \( r \geq 1, s \geq 1 \) provided \( r+1 \) \((s+2) \leq n \), or provided \( (r+2)^{\frac{1}{2}} s^{\frac{1}{2}} \leq n \) where \( s \) is a squared positive integer multiple of \( r \).

Theorem 1 implies that there are matrices \( M \) such that problem (1) is ill-posed in that the objective can be decreased to zero with the constituents \( L \) and \( S \) diverging with unbounded energy. The problem size bounds in Theorem 1 allow for matrices with \( r = O(n^\alpha) \) to have number of corruptions of order \( s = O(n^{2-2\alpha}) \) for \( l \in [0, 1/2] \), which for constant rank allows \( s \) to be quadratic in \( n \), and for \( l \in (1/2, 1] \) to have the number of corruptions of order \( s = O(n(1-l)) \).

The phenomenon described in Theorem 1 is best illustrated by considering the minimization of \( \min_{X \in \text{LS}_{3,3}(1,1)} \| X - M(\epsilon) \|_F \) for the following two matrices

\[
M(\epsilon) = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad M(2) = \begin{bmatrix} 1 & -2 & -2 \\ -2 & 0 & 0 \\ -2 & 0 & 0 \end{bmatrix}.
\]

There exists the following sequence of matrices \( X_\epsilon \in \text{LS}(1,1) \)

\[
X_\epsilon = \begin{bmatrix} 2 & -1 & -1 \\ -1 & \epsilon & \epsilon \\ -1 & \epsilon & \epsilon \end{bmatrix} = \begin{bmatrix} 1/\epsilon & -1 & -1 \\ -1 & \epsilon & \epsilon \\ -1 & \epsilon & \epsilon \end{bmatrix} + \begin{bmatrix} 2 - 1/\epsilon & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},
\]

\[
L_\epsilon, S_\epsilon
\]

which can decrease the objective function \( \| X_\epsilon - M(\epsilon) \|_F = 2\epsilon \to 0 \) as \( \epsilon \to 0 \), but at the cost of the constituents \( L_\epsilon \) and \( S_\epsilon \) diverging with unbounded energy. Moreover, the sequence which minimizes the error converges to a matrix \( M(1) \) lying outside of the feasible set \( \text{LS}(1,1) \) and is in the set \( \text{LS}(1,2) \) instead. As a consequence, RPCA for matrices \( M(1) \) and \( M(2) \) in (3) does not have a global minimum. Likewise, we could consider an instance of the MC problem in which the top left entry of \( M(1) \) is missing and a rank 1 approximation is sought. We see that a rank 1 solution cannot be obtained, but the sequence \( L_\epsilon \) decreases the objective arbitrarily close to zero while the energy of the iterates grows without bounds, \( \| L_\epsilon \|_F \to \infty \).

We show that a number of widely considered nonconvex Robust PCA and low-rank matrix completion algorithms can follow diverging sequences similar to \( X_\epsilon \) in (4). Figure 1 and Figure 2 depict divergence for two popular nonconvex RPCA algorithms and two non-convex MC algorithms when applied to \( M(1) \) and \( M(2) \) in (3).

Convex relaxations of RPCA solving \( \min_{L+S=\text{LS}(1,1)} \| L \|_F + \lambda \| S \|_1 \) do not suffer from the divergence of constituents due to their explicit minimization of their norms. However, they suffer from sub-optimal performance. Figure 3 shows that for two widely considered convex relaxation algorithms, as the regularization parameter \( \lambda \) is increased from near zero it first produced a solution with \( r = 0 \) and \( s = 5 \), then at approximately \( \lambda = 1/2 \) transitions to solutions with overspecified degrees of freedom \( r = 2 \) and \( s = 5 \), and then for large values of \( \lambda \) gives solutions with \( r = 2 \) and \( s = 0 \).

This work brings to attention an overlooked issue: that both RPCA and MC can be ill-posed because the set of low-rank plus sparse matrices is not closed. We give lower bound of \( n(r,s) \geq (r+1)(s+2) \) and \( n(r,s) \geq (r+2)^{\frac{3}{2}} s^{\frac{3}{2}} \) in Theorem 1 and also conjecture the best attainable bound is achieved at \( n(r,s) \geq r + (s+1)^{1/2} \).
I. APPENDIX

Fig. 1: Solving for an LS(1,1) approximation to $M^{(1)}$ and $M^{(2)}$ using two non-convex Robust PCA algorithms. Despite the norm of the residual $\|M^{(1)} - (L^t + S^t)\|_F$ converging to zero, norms of the constituents $L^t$, $S^t$ diverge. We set AltMin parameters $r = 1$, $s = 1$ and for FastGD we set $\lambda = 3.23$ and stepsize $\eta = 1/6$ which corresponds to choosing $s = 1$.

(a) FastGD [17] applied to $M^{(1)}$. (b) AltMin [8] applied to $M^{(2)}$.

Fig. 2: Recovery of $M^{(1)}$ given a rank 1 constraint by two non-convex matrix completion algorithms. Despite the norm of the residual $\|y - P_{H_1}(X^t)\|_F$ converging to zero, the norm of the recovered matrix $X^t$ diverges.

(a) LMaFit [16] applied to $M^{(1)}$. (b) CGIHT with restarts [1] applied to $M^{(1)}$.

Fig. 3: Recovered ranks, sparsities and objective values for two convex Robust PCA algorithms that solve $\min_{L, S = M^{(2)}} \|L\|_1 + \lambda \|S\|_1$ with varying choice of $\lambda$. Both PCP and IALM do not recover the $r = 1$, $s = 1$ solution for any $\lambda$. IALM recovers solutions with overspecified degrees of freedom $r = 2$, $s = 5$ for $\lambda$ roughly 1/2.


REFERENCES