Robust PCA with Partial Subspace Knowledge

Jinchun Zhan
Dep. of Electrical & Computer Eng.
Iowa State University, Ames, Iowa, USA
Email: jzhan@iastate.edu

Namrata Vaswani
Dep. of Electrical & Computer Eng.
Iowa State University, Ames, Iowa, USA
Email: namrata@iastate.edu

***Abstract*—In recent work, robust PCA has been posed as a problem of recovering a low-rank matrix L and a sparse matrix S from their sum, $M := L + S$ and a provably exact convex optimization solution called PCP has been proposed. Suppose that we have a partial estimate of the column subspace of the low rank matrix L. Can we use this information to improve the PCP solution, i.e. allow recovery under weaker assumptions? We propose here a simple modification of the PCP idea, called modified-PCP, that allows us to use this knowledge. We derive its correctness result which shows that modified-PCP indeed requires significantly weaker incoherence assumptions than PCP, when the available subspace knowledge is accurate. Extensive simulations are also used to illustrate this. Finally, we explain how this problem naturally occurs in many applications involving time series data, e.g. in separating a video sequence into foreground and background layers, in which the subspace spanned by the background images is not fixed but changes over time and the changes are gradual. A corollary for this case is also given.**

I. Introduction

Principal Components Analysis (PCA) is a widely used dimension reduction technique that finds a small number of orthogonal basis vectors, called principal components (PCs), along which most of the variability of the dataset lies. Accurately computing the PCs in the presence of outliers is called robust PCA. Outlier is a loosely defined term that refers to any corruption that is not small compared to the true data vector and that occurs occasionally. As suggested in [1], an outlier can be nicely modeled as a sparse vector. The robust PCA problem occurs in various applications ranging from video analysis to recommender system design in the presence of outliers, e.g. for Netflix movies, to anomaly detection in dynamic networks [2]. In video analysis, background image sequences are well modeled as forming a low-rank but dense matrix because they change slowly over time and the changes are typically global. Foreground is a sparse image consisting of one or more moving objects. In recent work, Candes et al and Chandrasekharan et al [2], [3] posed the robust PCA problem as one of separating a low-rank matrix L (true data matrix) and a sparse matrix S (outliers' matrix) from their sum, $M := L + S$. They showed that by solving the following convex optimization called principal components' pursuit (PCP)

$$\begin{array}{ll} \text{minimize}_{\tilde{L},\tilde{S}} & \|\tilde{L}\|_* + \lambda\|\tilde{S}\|_1 \\ \text{subject to} & \tilde{L} + \tilde{S} = M \end{array} \tag{1}$$

it is possible to recover L and S exactly with high probability under mild assumptions. This was among the first recovery guarantees for a practical (polynomial complexity) robust PCA algorithm. Since then, the batch robust PCA problem, or what is now also often called the sparse+low-rank recovery problem, has been studied extensively, e.g. see [4], [5], [6], [7], [8], [9], [10], [11], [12], [13].

Contribution: In this work we study the following problem. Suppose that we have a partial estimate of the column subspace of the low rank matrix L. How can we use this information to improve the PCP solution, i.e. allow recovery under weaker assumptions? We propose here a simple modification of the PCP idea, called *modified-PCP*, that allows us to use this knowledge. We derive its correctness result (Theorem 2.1) by adapting the proof given in [2]. The result shows that modified-PCP indeed requires significantly weaker incoherence assumptions than PCP, as long as the available subspace knowledge is accurate.

In many applications involving time series data, e.g. in video, the subspace spanned by a set of consecutive columns of L does not remain fixed, but instead changes over time and the changes are gradual. Also, often an initial short sequence of low-rank only data (without outliers) is available, e.g. in video analysis, it is easy to get an initial background-only sequence. Thus, for this application modified-PCP can be used to design a piecewise batch solution to the robust PCA problem that will require weaker assumptions for exact recovery than PCP. This is made precise in Corollary 2.2.

Modified-PCP is motivated by the modified-CS [14] idea. Modified-CS solves the problem of sparse recovery with partial support knowledge. Its idea is to try to find the vector that is sparsest on the complement set of the known support subject to the data constraint. Modified-PCP uses a similar idea for the vector of singular values of the low rank matrix.

A. Notation

For a matrix X, we denote by X^* the transpose of X; denote by $\|X\|_\infty$ the ℓ_∞ norm of X reshaped as a long vector, i.e., $\max_{i,j} |X_{ij}|$; denote by $\|X\|$ the operator norm or 2-norm; denote by $\|X\|_F$ the Frobenius norm; denote by $\|X\|_*$ the nuclear norm; denote by $\|X\|_1$ the ℓ_1 norm of X reshaped as a long vector. Let $\|\mathcal{A}\|$ denote the operator norm of operator $\mathcal{A}$, i.e., $\|\mathcal{A}\| = \sup_{\{\|X\|_F=1\}} \|\mathcal{A}X\|_F$; let $\langle X, Y\rangle$ denote the Euclidean inner product between two matrices, i.e., trace(X^*Y); let sgn(X) denote the entrywise sign of X.

We let $\mathcal{P}_\Theta$ denote the orthogonal projection onto linear subspace Θ. We use Ω to denote the support set of S, i.e.,

U.S. Government work not protected by U.S. copyright

$\Omega = \{(i, j) : S(i, j) \neq 0\}$. We also use Ω to denote the subspace spanned by all matrices supported on Ω.

Given two matrices B and B_2, $[B\ B_2]$ constructs a new matrix by concatenating matrices B and B_2 in the horizontal direction. Let B_{rem} be a matrix containing some columns of B. Then $B \setminus B_{\text{rem}}$ is the matrix B with columns in B_{rem} removed.

We say that a matrix U is a *basis matrix* if $U^*U = I$ where I is the identity matrix.

B. Problem Definition

We are given a data matrix $M \in \mathbb{R}^{n_1 \times n_2}$ that satisfies

$$M = L + S \tag{2}$$

where S is a sparse matrix with support set Ω and L is a low rank matrix with rank r and with reduced SVD

$$L = U \Sigma V^*, \tag{3}$$

We assume that we are given an $n_1 \times r_G$ basis matrix G so that $L_{\text{new}} := (I - GG^*)L$ has rank smaller than r. The goal is to recover L and S from M using G.

We explain the above a little more. With G as above, U can be rewritten as

$$U = [(GR \setminus U_{\text{extra}})\ U_{\text{new}}], \tag{4}$$

where $U_{\text{new}} \in \mathbb{R}^{n_1 \times r_{\text{new}}}$ with $r_{\text{new}} < r$ and $U_{\text{new}}^* G = 0$; R is a rotation matrix and U_{extra} contains r_{extra} columns of GR. Let V_{new} be the right singular vectors of the reduced SVD of $L_{\text{new}} := (I - GG^*)L = U_{\text{new}} U_{\text{new}}^* L$, i.e.

$$L_{\text{new}} = U_{\text{new}} \Sigma_{\text{new}} V_{\text{new}}^* \tag{5}$$

Another way to explain the above is that $G = [(U \setminus U_{\text{new}})\ U_{\text{extra}}]R^*$. Also, let $U_0 := [U \setminus U_{\text{new}}]$ and let r_0 be its rank.

C. Proposed Solution: Modified-PCP

Denote by Γ the linear space of matrices with column span equal to that of the columns of G, i.e.

$$\Gamma := \{GX^*,\ X \in \mathbb{R}^{n_2 \times r_G}\}, \tag{6}$$

and by $\Gamma^\perp$ its orthogonal complement. Clearly,

$$L_{\text{new}} + GX^* + S = M \tag{7}$$

Inspired by PCP [2], we try to recover L_{new} and S by solving the following **Modified PCP** (mod-PCP) program

$$\begin{array}{ll} \text{minimize}_{\tilde{L},\tilde{S}} & \|\tilde{L}_{\text{new}}\|_* + \lambda \|\tilde{S}\|_1 \\ \text{subject to} & \tilde{L}_{\text{new}} + G\tilde{X}^* + \tilde{S} = M \end{array} \tag{8}$$

Once S is recovered, we can also get $L = M - S$. Its column subspace can be estimated as the left singular vectors with nonzero singular values.

II. Correctness Result

We first state the assumptions required for the result and then give the main result and discuss it.

A. Assumptions

As explained in [2], we need a denseness assumption on the singular vectors of L in order to ensure that they can be separated from the sparse matrix S. We also need that S is not low rank. One way to ensure that S is full rank w.h.p. is by selecting the support of S uniformly at random [2]. We assume this here too. For denseness, as we will show next, for modified-PCP, the following assumptions suffice. Assume that, for a small enough parameter μ,

$$\max_i \|[G\ U_{\text{new}}]^* e_i\|^2 \leq \frac{\mu r}{n_1}, \tag{9}$$

$$\max_i \|V_{\text{new}}^* e_i\|^2 \leq \frac{\mu r}{n_2}, \tag{10}$$

and

$$\|U_{\text{new}} V_{\text{new}}^*\|_\infty \leq \sqrt{\frac{\mu r}{n_1 n_2}}. \tag{11}$$

Recall that the columns of the "subspace knowledge" matrix G span a subspace of range(L) and $U_{\text{new}}, V_{\text{new}}$ are the left and right singular vectors of $L_{\text{new}} := (I - GG')L$. As we explain later, the last two assumptions, and particularly the last one, are weaker than those required by PCP as long as r_{extra} is not too large compared to r. The requirements are much weaker when $r_{\text{new}} \ll r$ also.

B. The Result

We can claim the following.

Theorem 2.1: Let $n_{(1)} = \max(n_1, n_2)$ and $n_{(2)} = \min(n_1, n_2)$. Fix any $n_1 \times n_2$ matrix Υ of signs. Assume the model given in Sec I-B and assume that L satisfy (9), (10) and (11); the support set of S is uniformly distributed with size at most m; and

$$r \leq \rho_r n_{(2)} \mu^{-1} (\log n_{(1)})^{-2},\ m \leq \rho_s n_{(1)} n_{(2)} \tag{12}$$

for some constants ρ_r, ρ_s; the signs of S are fixed, i.e., $sgn([S]_{ij}) = \Upsilon_{ij}$, for all $(i, j) \in \Omega$. Then Modified-PCP (8) with $\lambda = 1/\sqrt{\max\{n_1, n_2\}}$ recovers S exactly (and hence also $L = M - S$) with probability at least $1 - cn_{(1)}^{-10}$ for some numerical constant c.

Proof: The proof is obtained by adapting the proof techniques of [2] to our problem. We provide a brief outline in Sec III and a complete proof in [15].

C. Discussion

For simplifying our discussion, first just assume that range(G) = range(U_0) so that $r_G = r_0$ and $r_{\text{extra}} = 0$. Also suppose that $n_{(1)} = n_1$ and $n_{(2)} = n_2$, i.e. the matrix has fewer columns than rows. The PCP program of [2] is (8) with no subspace knowledge available, i.e. $G_{PCP} = []$ (empty matrix). With this, Theorem 2.1 simplifies to the corresponding result for PCP. Thus, $U_{\text{new},PCP} = U$ and $V_{\text{new},PCP} = V$ and so (9) is the same but (10) and (11) get replaced by stronger requirements: $\max_i \|V^* e_i\|^2 \leq \frac{\mu r}{n_2}$ and $\|UV^*\|_\infty \leq \sqrt{\frac{\mu r}{n_1 n_2}}$. The last condition (11) for PCP is the most difficult to satisfy. We demonstrate this in Fig 3 by plotting the μ required for all the three conditions for the matrices used in our simulations. As can be seen, for mod-PCP we are able to relax the requirement significantly: we need a much smaller value of μ. From Theorem 2.1, a smaller μ means that, for a given rank r, a smaller value of n_2 will suffice. We demonstrate this in Fig 1 for $r_{\text{new}} = 9$ and $r = 29$.

A similar argument can be made even when range(G) contains only a few extra directions, i.e. r_{extra} is not too large. In that case (9) will be slightly stronger but (10) and (11) will be weaker. Thus, the overall requirement will still be weaker since (11) is the dominant condition. If $r_{\text{new}} \ll r$ the requirements will be significantly weaker.

Consider an online / recursive robust PCA problem where data vectors $y_t := s_t + \ell_t$ come in sequentially and where the subspace changes over time. Starting with an initial knowledge of the subspace, the goal is to estimate the subspace spanned by $\ell_1, \ell_2, \dots \ell_t$ and to recover the s_t's. Assume the following subspace change model introduced in [16]: $\ell_t = P_{(t)} a_t$ where $P_{(t)} = P_j$ for all $t_j \le t < t_{j+1}$, $j = 0, 1, \dots J$. At the change times, P_j changes as $P_j = [(P_{j-1} R_j \setminus P_{j,\text{old}}) \ P_{j,\text{new}}]$ where $P_{j,\text{new}}$ is a $n \times c_{j,\text{new}}$ basis matrix that satisfies $P_{j,\text{new}}^* P_{j-1} = 0$; R_j is a rotation matrix; and $P_{j,\text{old}}$ is a $n \times c_{j,\text{old}}$ basis matrix. Let $r_j := \text{rank}(P_j)$. Clearly, $r_j = r_{j-1} + c_{j,\text{new}} - c_{j,\text{old}}$. If $\sum_j (c_{j,\text{new}} - c_{j,\text{old}}) \le c_0$, then $r_j \le r_{\max} = r_0 + c_0$.

Suppose that the initial subspace, range(P_0), is perfectly known. Then modified-PCP provides a piecewise batch exact recovery solution to the above problem that would need a weaker set of denseness assumptions than those required by PCP. The following corollary is immediate.

Corollary 2.2: Assume the model described above. For a given $\alpha < n_1$, suppose that

1) *we solve modified-PCP at every $t = t_0 + k\alpha - 1$, $k = 1, 2, \dots,$ using the last α measurements, i.e. using $M = [y_{t_0+(k-1)\alpha}, y_{t_0+(k-1)\alpha+1}, \cdots y_{t_0+k\alpha-1}]$*
2) *$t_j - t_{j-1}$ is an integer multiple of α*
3) *the initial subspace range(P_0) is exactly known.*
4) *for all $j = 1, 2, \dots J$, (9), (10), and (11) hold with $n_2 = \alpha$, $r = r_j$, $G = P_{j-1}$, $U_{new} = P_{j,new}$, and V_{new} being the right singular vectors of $U_{new}^* [\ell_{t_j}, \ell_{t_j+1}, \dots \ell_{t_j+\alpha-1}]$ and with $\mu = \mu_j$*
5) *for all $j = 1, 2, \dots J$, the bound of Theorem 2.1 holds with $n_{(1)} = n_1$, $n_{(2)} = \alpha$, $r = r_j$ and $\mu = \mu_j$*

then, we can recover L and S exactly and in a piecewise batch fashion with probability at least $(1 - cn^{-10})^J$.

Proof: Denote by Θ_j the event that L_j, S_j are exactly recovered from matrix $M = [y_{t_j}, y_{t_j+1}, \dots y_{t_j+\alpha-1}]$. This also implies that P_j is exactly recovered. Using conditions 4 and 5 and Theorem 2.1, we get that probability $\mathbb{P}(\Theta_j | \Theta_{j-1}) \ge 1 - cn^{-10}$. By condition 3, $\mathbb{P}(\Theta_0) = 1$. Also, clearly, conditioned on Θ_{j-1}, Θ_j is independent of $\Theta_{j-2}, \Theta_{j-3}, \cdots, \Theta_0$. Thus $\mathbb{P}(\Theta_1, \Theta_2, \cdots, \Theta_J | \Theta_0) = \mathbb{P}(\Theta_1|\Theta_0)\mathbb{P}(\Theta_2|\Theta_1) \cdots \mathbb{P}(\Theta_J|\Theta_{J-1}) \ge (1 - cn^{-10})^J$. ■

In [17], [18], [16], Qiu et al studied the online / recursive robust PCA problem and proposed a novel recursive algorithm called ReProCS. With the subspace change model described above, they also needed the following "slow subspace change" assumption: $\|P_{j,\text{new}}^* \ell_t\|_2$ is small for sometime after t_j and increases gradually (a model was assumed for this). However, modified-PCP does not need this. Moreover, the performance guarantee they provide is not a correctness result. Lastly, even with perfect initial subspace knowledge, ReProCS cannot achieve exact subspace recovery while, as shown above, modified-PCP can. On the other hand, ReProCS is a fully recursive and fast algorithm while modified-PCP is neither.

In simulation experiments that we will show in Sec IV, for uniformly randomly selected support sets, and when the "slow subspace change" assumption does not hold, modified-PCP significantly outperforms both ReProCS and PCP. However, in situations that involve correlated support change, e.g., a moving object in a video sequence, the difference is not clear and depends on whether the slow subspace change assumption holds or not. This case requires further study (not shown here). Correlated support change usually results in S also being rank deficient. Hence, in this case, mod-PCP will require a different approach to selecting λ and will require more assumptions.

Other recent work on algorithms for recursive / online robust PCA includes [19], [20], [21], [22], [23]. In [22], [23], two online algorithms for robust PCA (that do not model the outlier as a sparse vector but as a vector that is "far" from the data subspace) have been partly analyzed. Other somewhat related work includes online algorithms for low-rank matrix completion and dictionary learning [24], [25].

III. Proof Outline

The overall proof approach is similar to that in [2]. The first step involves starting with the KKT conditions and relaxing them to find a set of conditions under which L_{new}, S is the unique minimizer of (8) (Lemma 3.1). These conditions are further relaxed to get a set of conditions on the dual certificate that are easy to satisfy, as is also done in [2] (Lemma 3.2). Finally the golfing scheme [26], [2] is used to construct this dual certificate and to show that it indeed satisfies the required conditions.

The proof needs the following linear space of matrices.

$$\Pi := \{[G\, U_{\text{new}}]X^* + Y V_{\text{new}}^*, \ X \in \mathbb{R}^{n_2 \times (r_G + r_{\text{new}})}, Y \in \mathbb{R}^{n_1 \times r_{\text{new}}}\},$$

Lemma 3.1: (similar to [2, Lemma 2.4], [3, Proposition 2]) If $\|\mathcal{P}_\Omega \mathcal{P}_\Pi\| < 1$, i.e., $\Omega \cap \Pi = \{0\}$, (L_{new}, S) is the unique solution to Modified-PCP (8) if there is a pair (W, F) obeying

$$U_{\text{new}} V_{\text{new}}^* + W = \lambda(\text{sgn}(S) + F),$$

with $\mathcal{P}_\Pi W = 0$, $\|W\| < 1$, $\mathcal{P}_\Omega F = 0$ and $\|F\|_\infty < 1$.

Lemma 3.2: (similar to [2, Lemma 2.5]) If $\|\mathcal{P}_\Omega \mathcal{P}_\Pi\| \le 1/2$ and $\lambda < 1$, (L_{new}, S) is the unique solution to Modified-PCP (8) if there is a pair (W, F) obeying

$$U_{\text{new}} V_{\text{new}}^* + W = \lambda(\text{sgn}(S) + F + \mathcal{P}_\Omega D)$$

with $\mathcal{P}_\Pi W = 0$ and $\|W\| \le \frac{1}{2}$, $\mathcal{P}_\Omega F = 0$ and $\|F\|_\infty \le \frac{1}{2}$, and $\|\mathcal{P}_\Omega D\|_F \le \frac{1}{4}$.

The conditions needed by the corresponding lemma in [2] are stronger because (a) $\|U_{\text{new}} V_{\text{new}}^*\|_\infty$ is smaller than $\|UV^*\|_\infty$; and (b) Π is a smaller subspace of matrices than $T := \{UX^* + YV^*, \ X \in \mathbb{R}^{n_2 \times r}, Y \in \mathbb{R}^{n_1 \times r}\}$ used in [2].

Construction of a dual certificate W that satisfies the conditions of Lemma 3.2 requires using Π instead of T, U_{new} instead of U and V_{new} instead of V and following the solution approach of [2].

IV. Simulation Experiments

We first give below the algorithm used to solve modified-PCP. Next, we give recovery error comparisons for simulated data. Finally, we show a comparison of the assumptions needed by PCP and mod-PCP for the simulated data.

A. Algorithm for solving Modified-PCP

We give below an algorithm based on the Inexact Augmented Lagrange Multiplier (ALM) method [27] to solve the modified-PCP program, i.e. solve (8). This algorithm is a direct modification of the algorithm designed to solve PCP in [27]. Using the same ideas, along with a accurate recovery result for the basis pursuit denoising (BPDN) [28] problem, it should be possible to prove that the output of the algorithm converges to the solution of modified-PCP.

For the modified-PCP program (8), the Augmented Lagrangian function is:

$$\begin{aligned}\mathbb{L}(\tilde{L}_{\text{new}}, \tilde{S}, Y, \tau) = \; & \|\tilde{L}_{\text{new}}\|_* + \lambda\|\tilde{S}\|_1 + \langle Y, M - \tilde{L}_{\text{new}} - \tilde{S} \\ & - G\tilde{X}^* \rangle + \frac{\tau}{2}\|M - \tilde{L}_{\text{new}} - \tilde{S} - G\tilde{X}^*\|_F^2,\end{aligned}$$

Thus, with similar steps in [27], we have following algorithm. In Algorithm 1, Lines 3 solves $\tilde{S}_{\text{new},k+1} = \arg\min_{\tilde{S}} \|\tilde{L}_{\text{new},k}\|_* + \lambda\|\tilde{S}\|_1 + \langle Y, M - \tilde{L}_{\text{new},k} - \tilde{S} - G\tilde{X}_k^* \rangle + \frac{\tau}{2}\|M - \tilde{L}_{\text{new},k} - \tilde{S} - G\tilde{X}_k^*\|_F^2$; Line 4-6 solve $\tilde{L}_{\text{new},k+1} = \arg\min_{\tilde{L}_{\text{new}}} \|\tilde{L}_{\text{new}}\|_* + \lambda\|\tilde{S}_{k+1}\|_1 + \langle Y, M - \tilde{L}_{\text{new}} - \tilde{S}_{k+1} - G\tilde{X}_k^* \rangle + \frac{\tau}{2}\|M - \tilde{L}_{\text{new},k} - \tilde{S}_{k+1} - G\tilde{X}_k^*\|_F^2$. The soft-thresholding operator is defined as

---

Algorithm 1 (Modified RPCA via the Inexact ALM Method)[27, Algorithm 5]

Input: Measurement matrix $M \in \mathbb{R}^{n_1 \times n_2}$, λ, G.

1: $Y_0 = M/\max\{\|M\|, \|M\|_\infty/\lambda\}$; $\tilde{S}_0 = 0$; $\tau_0 > 0$; $v > 1$; $k = 0$.
2: **while** not converged **do**
3: $\tilde{S}_{k+1} = \mathfrak{S}_{\lambda\tau_k^{-1}}[M - G\tilde{X}_k - \tilde{L}_{\text{new},k} + \tau_k^{-1}Y_k]$.
4: $(\tilde{U}, \tilde{\Sigma}, \tilde{V}) = \text{svd}((I - GG^*)(M - \tilde{S}_{k+1} + \tau_k^{-1}Y_k))$;
5: $\tilde{L}_{\text{new},k+1} = \tilde{U}\mathfrak{S}_{\tau_k^{-1}}[\tilde{\Sigma}]\tilde{V}^T$.
6: $\tilde{X}_{k+1} = G^*(M - \tilde{S}_{k+1} + \tau_k^{-1}Y_k)$
7: $Y_{k+1} = Y_k + \tau_k(M - \tilde{S}_{k+1} - \tilde{L}_{\text{new},k+1} - G\tilde{X}_{k+1})$.
8: $\tau_{k+1} = \min(v\tau_k, \bar{\tau})$.
9: $k \leftarrow k + 1$.
10: **end while**

Output: $\hat{L}_{\text{new}} = \tilde{L}_{\text{new},k}, \hat{S} = \tilde{S}_k, \hat{L} = M - \tilde{S}_k$.

$$\mathfrak{S}_\epsilon[x] = \begin{cases} x - \epsilon, & \text{if } x > \epsilon; \\ x + \epsilon, & \text{if } x < -\epsilon; \\ 0, & \text{otherwise,} \end{cases} \tag{13}$$

We use yall1 [29] to solve Line 5. Parameters are set as suggested in [27], i.e., $\tau_0 = 1.25/\|M\|, v = 1.5, \bar{\tau} = 10^7\tau_0$ and iteration is stopped when $\|M - \tilde{S}_{k+1} - \tilde{L}_{\text{new},k+1} - G\tilde{X}_{k+1}\|_F/\|M\|_F < 10^{-7}$.

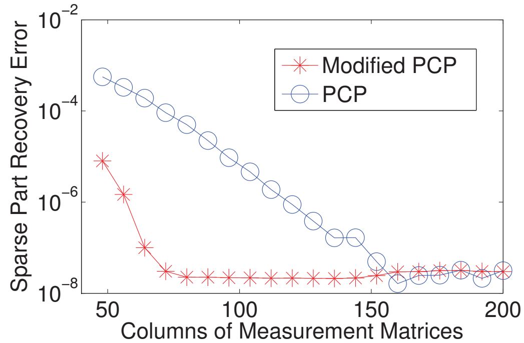

Fig. 1: Recovery result comparison with different columns

The above algorithm is slow primarily because of the projected sparse recovery step. In ongoing work, we are working on using ideas from existing work so that this is not needed.

B. Recovery Error Comparisons

In the simulations shown here we used $G = U_0$. The data was generated as follows. We generated $[U_0 \; U_{\text{new}}]$ by orthonormalizing an $n_1 \times (r_0 + r_{\text{new}})$ matrix with entries i.i.d. $\mathcal{N}(0, 1/n_1)$. Then we generated a matrix X_1 of size $r_0 \times d$ and a matrix X_2 of size $(r_0 + r_{\text{new}}) \times n_2$ with entries i.i.d. $\mathcal{N}(0, 1/n_1)$. We generate the support set of size m uniformly at random for sparse matrix S and assign value ± 1 with equal probability to entries in the support set.We set $M_{\text{new}} = [U_0 \; U_{\text{new}}]X_2 + S$ and $M_0 = U_0X_1$. We computed $G = U_0$ as the left singular vectors with nonzero singular values of M_0 and this was used as the partial subspace knowledge for modified-PCP. Modified-PCP solved (8) with $M \equiv M_{\text{new}}$ using Algorithm 1. PCP solved (8) with $M \equiv [M_0 \; M_{\text{new}}]$ and $G_{PCP} = []$ using the algorithm from [27]. Sparse recovery error is calculated as $\|S - \hat{S}\|_F^2/\|S\|_F^2$ and averaged over 100 Monte Carlo simulations. We plot it against $r := r_0 + r_{\text{new}}$ in Fig 1. For this figure, we used $n_1 = 200, r_0 = 20, d = 30$, $r_{\text{new}} = 9, m = 15n_2$, and n_2 ranging from 40 to 200. Notice that this is the situation where $n_2 \leq n_1$ so that $n_{(2)} = n_2$ and $n_{(1)} = n_1$. This situation typically occurs for time series applications, where one would like to use fewer columns to still get exact/accurate recovery. We compare mod-PCP and PCP. As we can see, PCP needs many more columns than mod-PCP for exact recovery. Here we say exact recovery when $\|S - \hat{S}\|_F^2/\|S\|_F^2$ is less than 10^{-7}.

In Figure 2, we show comparisons with increasing rank. We used $n_1 = 200, r_0 = 20, d = 40, n_2 = 80, m = 1200$ and r_{new} ranging from 1 to 20. We compare results for modified-PCP, PCP and ReProCS [16]. Clearly modified-PCP significantly outperforms both the others because the rank r is quite large for PCP and the "slow subspace change" assumption required by ReProCS does not hold.

C. Assumptions' Comparisons

We compute and plot the μ required for (9), (10), and (11) for PCP and mod-PCP. As shown in Figure 3, we can see

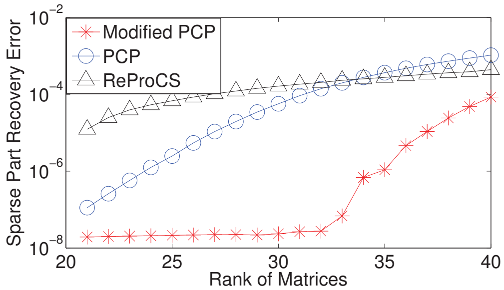

Fig. 2: Recovery result comparison with different rank

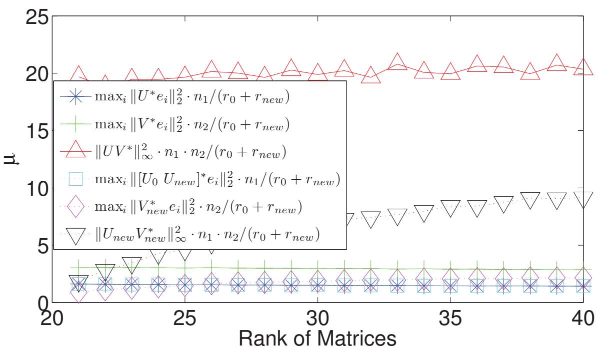

Fig. 3: Comparing the μ required by PCP and modified-PCP. We do this for the data used to generate Fig 2.

that constraint (11) is the most difficult one to satisfy. For mod-PCP, clearly this needs a much smaller value of μ.

We also computed smallest μ required for (9), (10) and (11) for one real video sequence, the lake sequence taken from http://www.ece.iastate.edu/~chenlu/ReProCS/ReProCS.htm. The initial low-rankification and computation of P_0 was done as explained in [16]. For this case, we got $r_0 = 26$, $r_{\text{new}} = 4$, and $\mu = 466$ for PCP, $\mu = 20$ for mod-PCP.

V. Conclusions

In this work we studied the following problem. Suppose that we have a partial estimate of the column subspace of the low rank matrix L. How can we use this information to improve the PCP solution? We proposed a simple modification of PCP, called *modified-PCP*, that allows us to use this knowledge. We derive its correctness result (Theorem 2.1) by adapting the proof given in [2]. We can argue that it indeed requires weaker incoherence assumptions on the low-rank matrix than PCP, as long as the number of extra directions in the available subspace knowledge is not too large. The requirements are significantly weaker when the number of unknown (new) directions is also small. Simulation experiments further illustrate these claims. Ongoing work includes designing a fast and/or recursive algorithm for modified-PCP.

References

[1] J. Wright and Y. Ma, "Dense error correction via l1-minimization," vol. 56, no. 7, pp. 3540–3560, July 2010.

[2] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of ACM*, vol. 58, no. 3, 2011.

[3] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Rank-sparsity incoherence for matrix decomposition," *SIAM Journal on Optimization*, vol. 21, 2011.

[4] Arvind Ganesh, John Wright, Xiaodong Li, Emmanuel J Candes, and Yi Ma, "Dense error correction for low-rank matrices via principal component pursuit," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*. IEEE, 2010, pp. 1513–1517.

[5] T. Zhang and G. Lerman, "A novel m-estimator for robust pca," *arXiv:1112.4863v1*, 2011.

[6] M. McCoy and J. Tropp, "Two proposals for robust pca using semidefinite programming," *arXiv:1012.1086v3*, 2010.

[7] H. Xu, C. Caramanis, and S. Sanghavi, "Robust pca via outlier pursuit," *IEEE Tran. on Information Theorey*, vol. 58, no. 5, May 2012.

[8] Arvind Ganesh, Kerui Min, John Wright, and Yi Ma, "Principal component pursuit with reduced linear measurements," *arXiv:1202.6445*.

[9] Daniel Hsu, Sham M. Kakade, and Tong Zhang, "Robust matrix decomposition with outliers," *arXiv:1011.1518*.

[10] John Wright, Arvind Ganesh, Kerui Min, and Yi Ma, "Compressive principal component pursuit," *arXiv:1202.4596*.

[11] Min Tao and Xiaoming Yuan, "Recovering low-rank and sparse components of matrices from incomplete and noisy observations," *SIAM Journal on Optimization*, vol. 21, no. 1, pp. 57–81, 2011.

[12] Yudong Chen, Ali Jalali, Sujay Sanghavi, and Constantine Caramanis, "Low-rank matrix recovery from errors and erasures," *IEEE Trans. Inform. Theory*, vol. 59(7), pp. 4324–4337, 2013.

[13] Alekh Agarwal, Sahand Negahban, and Martin J Wainwright, "Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions," *The Annals of Statistics*, vol. 40, no. 2, pp. 1171–1197, 2012.

[14] N. Vaswani and W. Lu, "Modified-cs: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Sig. Proc.*, vol. 58(9), pp. 4595–4607, September 2010.

[15] Jinchun Zhan and Namrata Vaswani, "Robust pca with partial subspace knowledge," *http://home.engineering.iastate.edu/~jzhan/modPCPfull.pdf*.

[16] C. Qiu, N. Vaswani, and L. Hogben, "Recursive robust pca or recursive sparse recovery in large but structured noise," *ICASSP, 2013, under revision for IEEE Tran. Info. Th., also at arXiv: 1211.3754[cs.IT]*.

[17] C. Qiu and N. Vaswani, "Real-time robust principal components' pursuit," in *Allerton*, 2010.

[18] C. Qiu and N. Vaswani, "Recursive sparse recovery in large but correlated noise," in *Allerton*, 2011.

[19] Jun He, Laura Balzano, and Arthur Szlam, "Incremental gradient on the grassmannian for online foreground and background separation in subsampled video," in *IEEE Conf. on Comp. Vis. Pat. Rec. (CVPR)*, 2012.

[20] Clemens Hage and Martin Kleinsteuber, "Robust pca and subspace tracking from incomplete observations using l0-surrogates," *arXiv:1210.0805 [stat.ML]*, 2013.

[21] Alaa E Abdel-Hakim and Motaz El-Saban, "Frpca: Fast robust principal component analysis for online observations," in *Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE, 2012, pp. 413–416.

[22] J. Feng, H. Xu, and S. Yan, "Online robust pca via stochastic optimization," in *Adv. Neural Info. Proc. Sys. (NIPS)*, 2013.

[23] J. Feng, H. Xu, S. Mannor, and S. Yan, "Online pca for contaminated data," in *Adv. Neural Info. Proc. Sys. (NIPS)*, 2013.

[24] Yuejie Chi, Yonina C. Eldar, and Robert Calderbank, "Petrels: Parallel subspace estimation and tracking by recursive least squares from partial observations," *IEEE Trans. Sig. Proc.*, December 2013.

[25] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro, "Online learning for matrix factorization and sparse coding," *The Journal of Machine Learning Research*, vol. 11, pp. 19–60, 2010.

[26] David Gross, Yi-Kai Liu, Steven T Flammia, Stephen Becker, and Jens Eisert, "Quantum state tomography via compressed sensing," *Physical review letters*, vol. 105, no. 15, pp. 150401, 2010.

[27] Zhouchen Lin, Minming Chen, and Yi Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

[28] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals," *IEEE Trans. Info. Th.*, pp. 1030–1051, March 2006.

[29] Junfeng Yang and Yin Zhang, "Alternating direction algorithms for l1 problems in compressive sensing," Tech. Rep., Rice University, June 2010.