Phaseless PCA: Low-Rank Matrix Recovery from Column-wise Phaseless Measurements

Seyedehsara Nayer [1] **Praneeth Narayanamurthy** [1] **Namrata Vaswani** [1]

Abstract

This work proposes the first set of simple, practically useful, and provable algorithms for two inter-related problems. (i) The first is low-rank matrix recovery from magnitude-only (phaseless) linear projections of each of its columns. This finds important applications in phaseless dynamic imaging, e.g., Fourier ptychographic imaging of live biological specimens. Our guarantee shows that, in the regime of small ranks, the sample complexity required is only a little larger than the order-optimal one, and much smaller than what standard (unstructured) phase retrieval methods need. (ii) We also study a dynamic extension of the above which allows the low-dimensional subspace from which each image/signal (each column of the low-rank matrix) is generated to change with time. We introduce a simple tracking algorithm that also has a provable guarantee when the subspace changes are piecewise constant.

1. Introduction

In recent years, there has been a resurgence of interest in the classical "phase retrieval (PR)" problem (Fienup, 1982; Gerchberg & Saxton, 1972). This involves recovering an n-length signal $\mathbf{x}^*$ from the *magnitudes* of its Discrete Fourier Transform (DFT) coefficients. While practical PR methods have existed for a long time, the focus of the recent work has been on obtaining convergence guarantees for these and newer algorithms. For this, the generalized PR problem has been introduced which replaces DFT by inner products with arbitrary design vectors, $\mathbf{a}_i$. Thus, the goal is to recover $\mathbf{x}^*$ from $\mathbf{y}_i := |\langle \mathbf{a}_i, \mathbf{x}^* \rangle|, i = 1, 2, \dots, m$. This line of work includes convex relaxation methods (Candes et al., 2013a;b) as well as non-convex methods (Netrapalli et al., 2013b; Candes et al., 2015; Chen & Candes, 2015; Zhang et al., 2016; Wang et al., 2016). It is easy to see that, without extra assumptions, PR requires $m \geq n$. The best known guarantees – see (Chen & Candes, 2015) and follow-up works – prove exact recovery high probability (whp) with order-optimal number of measurements (samples): $m = Cn$. Here and below, C is reused often to refer to a constant more than one. Much of the above work assumes that the $\mathbf{a}_i$'s are independent and identically distributed (iid) standard Gaussian vectors. There is also very recent work on guarantees for PR methods that only need a random initialization (instead of a carefully designed spectral initialization). Due to lack of space we do not discuss these here. These need more than Cn measurements.

A natural approach to reduce the sample complexity is to impose structure on the unknown signal(s). There is little work on structured PR with the exception of sparse PR which has been extensively studied e.g., (Jaganathan et al., 2012; Shechtman et al., 2014; Szameit et al., 2012). Low-rankness is the other common structure. There seem to be two natural approaches to exploit it. The first is to consider recovery from phaseless dense linear projections of the *entire* low-rank matrix. However, this model implicitly assumes that the signal/image, whose measurements are available, can be reshaped into a low-rank matrix. This is an assumption that is often not valid. A more reasonable (and commonly used) model is to assume that a set, or a time sequence, of signals/images together form a low-rank matrix and we have *phaseless linear projections of each signal or image*. This latter model finds applications in many dynamic imaging problems where the phase cannot be recovered. For example, in Fourier ptychography, it allows for imaging of slowly changing dynamic scenes such as live biological specimens in vitro. Similarly, it can also be used for various other dynamic phaseless imaging problems that occur in X-ray and sub-diffraction imaging or in astronomy.

We study the above problem: *recover a low-rank matrix from phaseless linear projections of each of its columns.* We also study its dynamic extension, which is a more useful problem setting when dealing with a long time sequence of signals/images. Versions of these problems were briefly studied in (Vaswani et al., 2017; Nayer & Vaswani, 2018; 2019) where a set of heuristics were proposed and experimentally evaluated, along with a partial attempt to analyze

[1]Department of Electrical and Computer Engineering, Iowa State University, USA. Correspondence to: Seyedehsara Nayer <sarana@iastate.edu>.

Proceedings of the 36^{th} International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019.

the first step of one of them. An application to dynamic ptychography was demonstrated in (Jagatap et al., 2018).

Phaseless Columnwise Low Rank Matrix Recovery (Ph-Co-LRMR). This involves recovering a rank r matrix $\mathbf{X}^* \in \mathbb{R}^{n\times q}$ from measurements of the form

$$\mathbf{y}_{ik} := |\mathbf{a}_{ik}{}'\mathbf{x}_k^*|,\ i = 1,\dots,m,\ k = 1,\dots,q. \quad (1)$$

Here and below, we use $'$ to denote vector or matrix transpose. Also, $\mathbf{x}_k^*$ is the k-th column of $\mathbf{X}^*$. Let $\mathbf{X}^* \overset{SVD}{=} \mathbf{U}^*\mathbf{\Sigma}^*\mathbf{B}^*$ denote its singular value decomposition so that $\mathbf{U}^* \in \mathbb{R}^{n\times r}$, $\mathbf{B}^* \in \mathbb{R}^{r\times q}$, and $\mathbf{\Sigma}^* \in \mathbb{R}^{r\times r}$ is a diagonal matrix. *Observe that this notation is a little non-standard, if the SVD was $\mathbf{U}^*\mathbf{\Sigma}^*\mathbf{V}^{*\prime}$, we are letting $\mathbf{B}^* := \mathbf{V}^{*\prime}$. Thus, columns of $\mathbf{U}^*$ and rows of $\mathbf{B}^*$ are orthonormal.* We use $\sigma_{\max}^*, \sigma_{\min}^*$ to denote the maximum and minimum singular values of $\mathbf{X}^*$ and $\kappa = \sigma_{\max}^*/\sigma_{\min}^*$ to denote its condition number. Also, let

$$\tilde{\mathbf{B}}^* := \mathbf{\Sigma}^*\mathbf{B}^*.$$

We write the SVD as above to make it easier to specify the dynamic problem setting. Also, the QR decomposition of an estimate of $\tilde{\mathbf{B}}^*$, denoted $\hat{\mathbf{B}}$, will be written as $\hat{\mathbf{B}} \overset{QR}{=} \mathbf{R}_B\mathbf{B}$ with $\mathbf{B}$ being an $r \times q$ matrix with orthonormal rows (or equivalently $\hat{\mathbf{B}}' \overset{QR}{=} \mathbf{B}'(\mathbf{R}_B)'$).

The only assumption we need on $\mathbf{X}^*$ is right incoherence (incoherence of right singular vectors). In our notation, this means

$$\max_k \|\mathbf{b}_k^*\|^2 \le \mu^2 r/q$$

with $\mu \ge 1$ being a constant. Notice that this implies that $\|\tilde{\mathbf{b}}_k^*\|^2 \le {\sigma_{\max}^*}^2\mu^2 r/q$ and $\|\mathbf{x}_k^*\|^2 \le \mu^2\kappa^2\|\mathbf{X}\|_F^2/q$.

For guarantees, we assume $\mathbf{a}_{ik} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I})$ (iid standard Gaussian).

Dynamic Ph-Co-LRMR or Phaseless Subspace Tracking (ST). The low-rank assumption is equivalent to assuming that $\mathbf{x}_k^* = \mathbf{U}^*\tilde{\mathbf{b}}_k^*$ where $\mathbf{U}^*$ specifies a fixed r-dimensional subspace. For long signal/image sequences, a better model (allows r to be smaller) is to let the subspace change with time. As is common in time-series analysis, the simplest model for time-varying quantities is to assume that they are piecewise constant. We adopt this approach here. Let $k_0 = 1$, and let k_j denote the j-th subspace change time, for $j = 1, 2, \dots, J$. Assume that

$$\mathbf{x}_k^* = \mathbf{U}_j^*\tilde{\mathbf{b}}_k^*, \text{ for all } k_j \le k \le k_{j+1}, j = 0, 1, \dots, J, \quad (2)$$

where $\mathbf{U}_j^*$ is an $n \times r$ matrix with orthonormal columns that represent the subspace it spans (we call it a "basis matrix"). Let $k_{J+1} = q_{\text{full}}$ denote the total number of frames. Thus $\mathbf{X}^*$ is now $n \times q_{\text{full}}$.

The goal is to track the subspaces $\mathrm{Span}(\mathbf{U}_j^*)$ on-the-fly; of course "on-the-fly" for subspace tracking means with a delay of at least r. Once this can be done, it is easy to also recover the matrix columns $\mathbf{x}_k^*$ (by solving a simple r-dimensional PR problem to recover the $\tilde{\mathbf{b}}_k^*$'s).

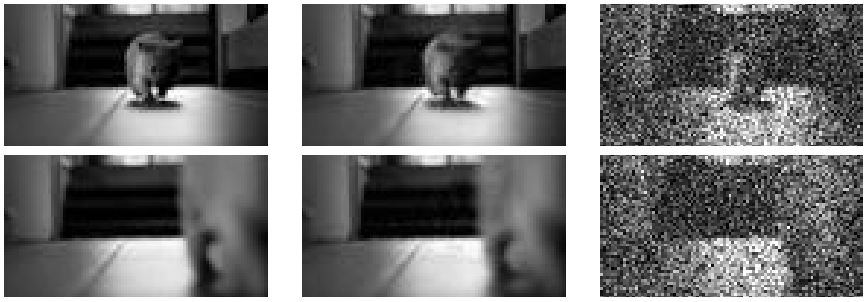
Notation. We use $\|.\|$ to denote the l2-norm of a vector or the induced 2-norm matrix. A tall matrix with orthonormal columns is referred to as a *"basis matrix"*. For two basis matrices $\mathbf{W}, \mathbf{D}$, we define the subspace error (distance) as $\mathrm{SE}(\mathbf{W}, \mathbf{D}) = \|\mathbf{D}_\perp{}'\mathbf{W}\| = \|\mathbf{W}_\perp{}'\mathbf{D}\| = \|(\mathbf{I} - \mathbf{W}\mathbf{W}')\mathbf{D}\|$. This measures the largest principal angle between the two subspaces. We often use terms like "estimate $\mathbf{W}$" when the goal is to really estimate its column span, $\mathrm{Span}(\mathbf{W})$. Since *we are working with real valued vectors and matrices*, the phase-invariant distance is just the sign invariant distance and is defined as $\mathrm{dist}(\mathbf{x}^*, \hat{\mathbf{x}}) = \min(\|\mathbf{x}^* - \hat{\mathbf{x}}\|, \|\mathbf{x}^* + \hat{\mathbf{x}}\|)$. Define the corresponding matrix distance as $\text{mat-dist}(\mathbf{X}^*, \hat{\mathbf{X}})^2 := \sum_{k=1}^q \mathrm{dist}(\mathbf{x}^*, \hat{\mathbf{x}})^2$.

We *reuse* the letters c, C to denote different numerical constants in each use with the convention that $C \ge 1$ and $c < 1$.

Contributions. This work has two key contributions.

1. It provides the first complete solution approach and guarantee for Ph-Co-LRMR (LRMR from column-wise phaseless linear projections). In the regime of small ranks r, we show that the required sample complexity is close to optimal. We also demonstrate the practical advantage of our proposed approach, AltMinLowRaP (Alt-Min for Phaseless Low Rank Recovery) over existing work, via extensive simulation experiments *and* a few experiments for recovering real videos (that are only approximately low-rank) from simulated coded diffraction pattern (CDP) measurements; see Fig. 1.
2. This work provides the first simple algorithm *and* a provable guarantee for Phaseless Subspace Tracking (ST) which is a dynamic extension of Ph-Co-LRMR.

Our proposed algorithm for Ph-Co-LRMR, AltMinLowRaP, relies on three key ideas. The first is a clever spectral initialization for obtaining the first estimate of $\mathrm{Span}(\mathbf{U}^*)$. The second is the observation that, if $\mathbf{U}^*$ were known, we only need to solve q small (r-dimensional) standard PR problems to recover the $\tilde{\mathbf{b}}_k^*$'s. Third, given an estimate of $\tilde{\mathbf{b}}_k^*$'s, we can obtain a new improved estimate of $\mathrm{Span}(\mathbf{U}^*)$ by solving a least squares (LS) problem. The key insight that helps obtain a significant sample complexity reduction over standard PR is the observation that, for both the initialization and the update steps for $\mathbf{U}^*$, conditioned on $\mathbf{X}^*$, we have access to mq mutually independent measurements. These are not identically distributed, however, the right incoherence assumption on $\mathbf{X}^*$ ensures that the distributions are similar



(a) Original (b) AltMinLowRaP (c) RWF

Figure 1. Comparison of recovery performance for the mouse video. The images are shown at $t = 60, 78$. For more video results, see (Nayer et al., 2019).

enough so that concentration holds with mq samples.

Assuming constant condition number, our guarantee shows that the sample complexity, mq, for recovering a rank r matrix of size $n \times q$ to ϵ accuracy is just $Cnr^5 \log(1/\epsilon)$. In the regime of small ranks r, this is close to the optimal complexity of $2nr$. Ignoring log factors, when $q \approx n$, this implies that only about r^5 samples per signal are required. For small r, this is a significant improvement over standard PR approaches which necessarily require $m \geq Cn$. Moreover, our guarantee for the dynamic setting (phaseless ST) shows that, if the subspace remains constant for a certain period of time before it changes, and if the largest principal angle of the change is not too small, we can both detect and track the changed subspace to within ϵ error with finite delay. This delay can be made small by increasing the number of measurements m per column.

2. Ph-Co-LRMR: Phaseless Column-wise Low-Rank Matrix Recovery

The goal is to recover $\mathbf{X}^* = \mathbf{U}^*\tilde{\mathbf{B}}^*$ from measurements of the form (1). We adopt an alternating minimization (AltMin) approach (Netrapalli et al., 2013a). We can rewrite $\mathbf{y}_{ik} = |\mathbf{a}_{ik}{}'\mathbf{U}^*\tilde{\mathbf{b}}_k^*| = |(\mathbf{U}^{*\prime}\mathbf{a}_{ik})', \tilde{\mathbf{b}}_k^*|$. If $\mathbf{U}^*$ were known, the problem of recovering each $\tilde{\mathbf{b}}_k^*$'s is an easy r-dimensional standard PR problem. Any PR solution can be used, here we use reshaped Wirtinger flow (RWF) (Zhang et al., 2016). Since $\mathbf{U}^*$ has orthonormal columns, the design vectors $(\mathbf{U}^{*\prime}\mathbf{a}_{ik})$'s are also standard Gaussian random vectors. Given an estimate of the $\tilde{\mathbf{b}}_k^*$'s, one can update the estimate of $\mathrm{Span}(\mathbf{U}^*)$ by standard LS followed by QR decomposition on its output to obtain a matrix with orthonormal columns. Thus, given an initial estimate of $\mathrm{Span}(\mathbf{U}^*)$, we have a simple alt-min algorithm that alternates between r-dimensional PR for updating the $\tilde{\mathbf{b}}_k^*$'s and LS for updating $\mathbf{U}^*$.

Since we only have an estimate of $\mathbf{U}^*$ at each iteration, the measurements for the PR step are not noise-free. The t-th iteration, it sees noise proportional to the $\mathrm{SE}(\mathbf{U}^t, \mathbf{U}^*)$. As a result the error in its estimate will be of the same level. Thus, there is no advantage in running the full RWF. Instead one can obtain a speed-up by letting the number of RWF iterations, $T_{RWF,t}$, at the t-th step grow with t.

To obtain the initialization, we develop a clever modification of the truncated spectral initialization idea from (Chen & Candes, 2015; Vaswani et al., 2017). First assume that r is known. We initialize $\hat{\mathbf{U}}$ as the first r left singular vectors of the following matrix.

$$\mathbf{Y}_U = \frac{1}{mq}\sum_{ik} \mathbf{y}_{ik}^2 \mathbf{a}_{ik}\mathbf{a}_{ik}' \mathbb{1}_{\left\{|\mathbf{a}_{ik}'\mathbf{x}_k^*|^2 \leq \frac{9}{mq}\sum_{ik}|\mathbf{a}_i'\mathbf{x}_k^*|^2\right\}}. \tag{3}$$

To simply understand why this works, consider the above matrix with the indicator function removed. Then it is not hard to see that its expected value equals $(1/q)[\mathbf{U}^*(\mathbf{\Sigma}^{*2})\mathbf{U}^{*\prime} + 2\mathrm{trace}(\mathbf{\Sigma}^{*2})\mathbf{I}]$, and so its span of top r singular vectors equals $\mathrm{Span}(\mathbf{U}^*)$. Hence, with large enough mq, the same should approximately hold for the original matrix. However, when using $\mathbf{Y}_U$ with the indicator function removed, a few "bad" measurements (those with very large magnitude $\mathbf{y}_{ik}^2$ compared to their empirical mean over i, k) can heavily bias its value. Hence a larger lower bound on mq will be needed to get a good initialization with high probability (whp). On the other hand, including the indicator function truncates the summation to only sum over the "good" measurements. Mathematically, this helps ensure that $\mathbf{Y}_U$ is close to a matrix that can be written as $\sum_{ik}\mathbf{w}_{ik}\mathbf{w}_{ik}{}'$ with $\mathbf{w}_{ik}$'s being sub-Gaussian vectors (instead of sub-exponential in the case without truncation).

We can also use $\mathbf{Y}_U$ to correctly estimate r whp by relying on the fact that when m and q are large the gap between its r-th and $(r+1)$-th singular value is close to ${\sigma_{\min}^*}^2/q$. With this idea, we estimate r as given in the first step of Algorithm 1.

By defining the $n \times m$ matrix $\mathbf{A}_k := [\mathbf{a}_{1,k}, \mathbf{a}_{2,k}, \ldots, \mathbf{a}_{m,k}]$ and $\mathbf{y}_k := [\mathbf{y}_{1,k}, \mathbf{y}_{2,k}, \ldots, \mathbf{y}_{m,k}]'$, and letting $|\mathbf{z}|$ denote element-wise magnitude of a vector, we can rewrite (1) as

Algorithm 1 AltMin-LowRaP: Alt-Min for Phaseless Low Rank Recovery

Parameters: $T, T_{RWF,t}, \omega$.
Partition m_{tot} measurements and design vectors into one for initialization and $2T$ disjoint sets for rest.
Set $\hat{r}$ as the largest index j for which $\lambda_j(\mathbf{Y}_U) - \lambda_n(\mathbf{Y}_U) \geq \omega$ where $\mathbf{Y}_U$ is in (3).
$\mathbf{U}^0 \leftarrow \hat{\mathbf{U}}^0 \leftarrow$ top $\hat{r}$ singular vectors of $\mathbf{Y}_U$ defined in (3).
for $t = 0 : T$ **do**
  $\hat{\mathbf{b}}_k^t \leftarrow RWF(\{\mathbf{y}_k^{(t)}, \mathbf{U}^{t\prime}\mathbf{A}_k^{(t)}\}, T_{RWF,t})$ for each $k = 1, 2, \cdots, q$
  Set $\hat{\mathbf{X}}^t = \mathbf{U}^t\hat{\mathbf{B}}^t$
  Compute QR decomp: $\hat{\mathbf{B}}^t \stackrel{\text{QR}}{=} \mathbf{R}_B^t\mathbf{B}^t$
  $\hat{\mathbf{C}}_k \leftarrow \text{Phase}(\mathbf{A}_k^{(T+t)\prime}\hat{\mathbf{x}}_k^t)$ for each $k = 1, 2, \cdots, q$
  Compute $\hat{\mathbf{U}}^{t+1}$ as $\arg\min_{\tilde{\mathbf{U}}} \sum_{k=1}^q \|\hat{\mathbf{C}}_k\mathbf{y}_k^{(T+t)} - \mathbf{A}_k^{(T+t)\prime}\tilde{\mathbf{U}}\mathbf{b}_k^t\|^2$
  Compute QR decomp: $\hat{\mathbf{U}}^{t+1} \stackrel{\text{QR}}{=} \mathbf{U}^{t+1}\mathbf{R}_U^{t+1}$
end for

$\mathbf{y}_k = |\mathbf{A}_k{}'\mathbf{x}_k^*|, k = 1, 2, \ldots, q$. This simplifies the writing of Algorithm 1. Also, as is commonly done in existing literature, e.g., (Netrapalli et al., 2013a), in order to obtain a provable guarantee in a simple fashion, a new (independent) set of m measurements in each new update of $\mathbf{U}^*$ and of the $\tilde{\mathbf{b}}_k^*$'s is used. Since we prove geometric convergence of the iterates, this increases the required sample complexity by a factor of only $\log(1/\epsilon)$). In simulations, this is not done.

Theorem 2.1 (Guarantee for AltMinLowRaP)**.** *Consider Algorithm 1. Assume that the $\mathbf{y}_{ik}$'s satisfy (1) with $\mathbf{a}_{ik}$ being iid standard Gaussian; $\mathbf{X}^*$ satisfies right-incoherence with parameter μ; and that the product $\mu\kappa$ is a constant. Set $T := C\log(1/\epsilon)$, $T_{RWF,t} = C(\log r + \log\kappa^2 + t(\log(0.7)/\log(1-c)))$, and $\omega = 0.25{\sigma_{\min}^*}^2/q$. Assume that, for the initialization step and for each new update, we use a new set of m measurements with m satisfying $mq \geq C\kappa^6\mu^2nr^5$ and $m \geq C\max(r, \log q, \log n)$. Pick $T = C\log(1/\epsilon)$. Then, w.p. at least $1 - Cn^{-10}$,*

$$\text{SE}(\mathbf{U}^*, \mathbf{U}^T) \leq \epsilon, \text{ mat-dist}(\hat{\mathbf{X}}^T, \mathbf{X}^*) \leq \epsilon\|\mathbf{X}^*\|_F$$

and $\text{dist}(\hat{\mathbf{x}}_k^T, \mathbf{x}_k^*) \leq \epsilon\|\mathbf{x}_k^*\|$ for each k. Moreover, after the t-th iteration,

$$\text{SE}(\mathbf{U}^*, \mathbf{U}^t) \leq 0.7^t\delta_{\text{init}}, t = 0, 1, 2, \ldots, T$$

where $\delta_{\text{init}} = \frac{c}{\kappa^2 r}$ and similar bounds also hold on the error in estimating $\mathbf{x}_k^*$s.

The time complexity is $mqnr\log^2(1/\epsilon)$ and memory complexity is $mqn\log(1/\epsilon)$.

We prove this in Sec. 5, proofs of the key lemmas can be found in (Nayer et al., 2019).

Theorem 2.1 implies that the sample complexity $m_{\text{tot}} = (2T+1)m$ needs to satisfy $m_{\text{tot}}q \geq C\kappa^6\mu^2nr^5\log(1/\epsilon)$ along with $m_{\text{tot}} \geq C\max(r, \log q, \log n)\log(1/\epsilon))$. The lower bound on just m is essentially redundant (except for very large q). As discussed earlier, in the regime of small r, this is close to the optimal value of $(n+q)r$. Also, clearly, it is significantly better than that of standard PR that does not use any structural assumptions. These need $m = Cn$.

The sample complexity gain is to be expected because we are exploiting extra structure. But what is also expected is that time complexity increases when doing that. For a given value of m and q, this is indeed true. AltMinLowRaP is about r times slower than the best PR methods such as TWF or RWF. These need time of order $mqn\log(1/\epsilon)$ to recover a set of q n-length signals. However, if we instead consider the time needed if, for each method, we use the least number of measurements needed for the method to provably give an ϵ-accurate estimate of the signals, then, in fact, we can argue that AltMinLowRaP is faster. More precisely, if we let $mq = Cnr^4\log^2(1/\epsilon)$ for AltMinLowRaP and $mq = Cnq$ for TWF/RWF, then AltMinLowRaP is faster as long as $r^5 < q$.

3. Dynamic Ph-Co-LRMR or Phaseless ST

Consider the subspace tracking (ST) problem. This assumes that the matrix $\mathbf{X}^*$ is of size $n \times q_{\text{full}}$ with the columns corresponding to a time sequence of signals or images. Thus measurements of the columns are obtained sequentially and hence there is benefit in developing an online (mini-batch) algorithm that works with measurements of short batches of α consecutive columns. The resulting algorithm is a simple modification of the static case idea along with a carefully designed subspace change detection step. To understand the change detection strategy, let $\hat{k}_j$ denote the estimated change times. Consider an α frame interval contained in $[k_j, k_{j+1})$. Assume that an ϵ-accurate estimate of the previous subspace $\mathbf{U}_{j-1}^*$ has been obtained by $\hat{k}_{j-1} + T\alpha$ and that this time is before k_j. Define the matrix

$$\mathbf{Y}_{U,det} := (I - \hat{\mathbf{U}}_{j-1}\hat{\mathbf{U}}_{j-1}{}')\mathbf{Y}_U(I - \hat{\mathbf{U}}_{j-1}\hat{\mathbf{U}}_{j-1}{}')$$

With a little bit of work (see proof of Lemma C.1 of (Nayer et al., 2019)), one can show that $\mathbf{Y}_{U,det}$ is close to a matrix $\mathbf{E}_{det}$ whose eigenvalues satisfy

$$\lambda_{\max}(\mathbf{E}_{det}) - \lambda_{\min}(\mathbf{E}_{det}) \geq 0.5(\text{SE}(\mathbf{U}_{j-1}^*, \mathbf{U}_j^*) - 2\epsilon)^2{\sigma_{\min}^*}^2/\alpha.$$

On the other hand, in an α frame interval contained in $[\hat{k}_{j-1} + T\alpha, k_j)$, this quantity can be upper bounded by $\text{SE}(\hat{\mathbf{U}}_{j-1}, \mathbf{U}_{j-1}^*)^2{\sigma_{\max}^*}^2/\alpha \leq \epsilon^2{\sigma_{\max}^*}^2/\alpha$. Thus, this quantity is small when the change has not occurred, and is large when the subspace has changed sufficiently. By using a large enough lower bound on the product $m\alpha$, and

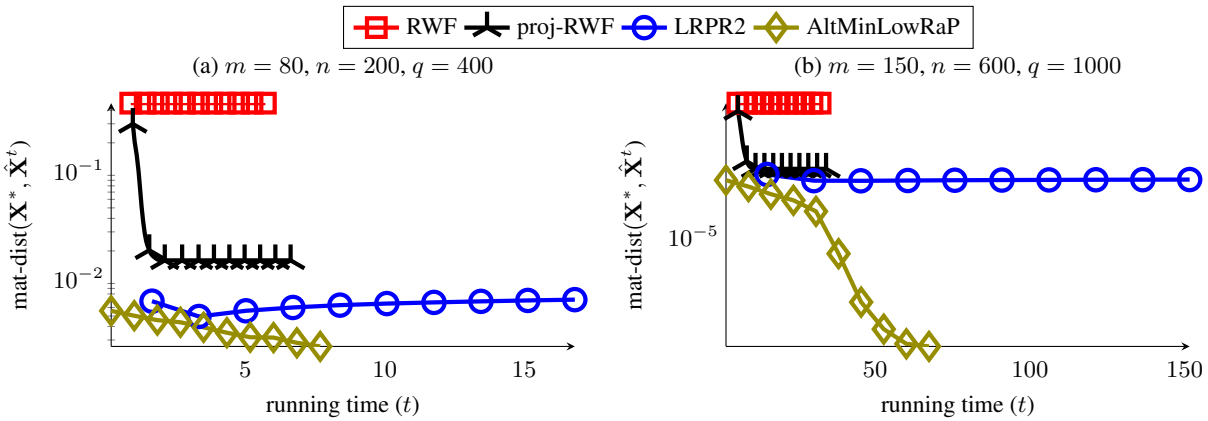


Figure 2. Error versus time plot with time in seconds. In the left plot $m = 80$, $n = 200$, $q = 400$ and in the right plot $m = 150$, $n = 600$, $q = 1000$. We compare with LRPR2 which is the only other existing Low-Rank Phase Retrieval algorithm (Vaswani et al., 2017), RWF (Zhang et al., 2016) and projected RWF. Notice that the first step (initialization) of LRPR2 is slower than our proposed method. This is likely due to the fact that the estimates of the rank are different for the two algorithms and thus the errors are also slightly different in the two cases. For the purpose of better illuatration, we only plot the error and time at the end of every 10 iterations for RWF and proj-RWF.

concentration bounds, the same can be shown for the difference between the max and min eigenvalues of $\mathbf{Y}_{U,det}$. We summarize the complete algorithm in Algorithm 2. It toggles between the "update" mode and the "detect" mode. It starts in the update mode, and enters the detect mode after T iterations of the update are done.

Once we have an ϵ-accurate estimate of the current subspace, it is straightforward to also recover the corresponding signals $\mathbf{x}_k^*$. This can simply be done via projected PR to recover $\tilde{\mathbf{b}}_k^*$'s. See last line of Algorithm 2. This borrows a similar idea from (Narayanamurthy & Vaswani, 2018).

Theorem 3.1 (Phaseless ST)**.** *Consider Algorithm 2. Pick any value of* $m \geq C \max(r, \log n, \log q_{\text{full}})$. *For this* m, *set* $\alpha = \frac{C\kappa^6 \mu^2 n r^5}{m}$. *Set* $T := C \log(1/\epsilon)$, *and the change detection threshold* $\omega = c/(\kappa\sqrt{r})$. *Assume that* $k_{j+1} - k_j \geq \alpha(T+2)$ *and that at, each change time,* k_j, $\text{SE}(\mathbf{U}_{j-1}, \mathbf{U}_j)^2 > c/(\kappa\sqrt{r})$. *Then, w.p. at least* $1 - Cn^{-10}$,

1. *we can detect the change with a delay of at most* 2α, *while ensuring no false detections:* $k_j \leq \hat{k}_j \leq k_j + 2\alpha$;
2. *for any* $\epsilon > 0$, *we can get an* ϵ*-accurate estimate with a delay of at most* $(T+3)\alpha$; *and we have the following subspace error bounds:*

$$\text{SE}(\mathbf{U}^{(k)}, \mathbf{U}_j^*) \leq \begin{cases} \text{SE}(\mathbf{U}_{j-1}^*, \mathbf{U}_j^*) + \epsilon & \text{if } k \in \mathcal{J}_{-1} \\ (0.7)^\ell \frac{c}{\kappa\sqrt{r}} & \text{if } k \in \mathcal{J}_\ell, \\ \epsilon & \text{if } k \in \mathcal{J}_{T+1} \end{cases}$$

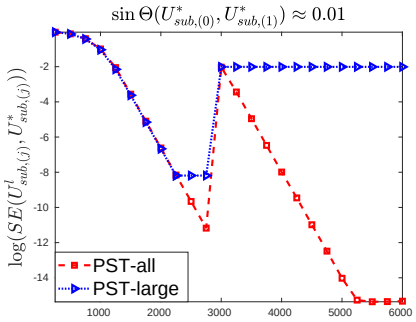
Here, $\mathcal{J}_{-1} := [k_j, k_j + 3\alpha]$, $\mathcal{J}_\ell := [k_j + (3+\ell)\alpha, k_j + (3+\ell+1)\alpha]$ *for* $\ell = 0, 1, 2, \ldots, T$ *and* $\mathcal{J}_{T+1} := [k_j + (T+4)\alpha, k_{j+1}]$.

Offline PST-large returns $\hat{\mathbf{X}}$ *that satisfies mat-dist*$(\hat{\mathbf{X}}, \mathbf{X}^*) \leq \epsilon$.

For its proof, see (Nayer et al., 2019).The above result shows that, if the subspace remains constant for at least $\alpha \log(1/\epsilon)$ frames, and if the amount of subspace change (largest principal angle of subspace change) is of order $1/\sqrt{r}$ or larger, then we can both detect the change and track the changed subspace to ϵ error within a delay of order $\alpha \log 1/\epsilon$. Moreover, for only at most 3α frames after a change, the subspace error does not reduce and is essentially bounded by the amount of change. After this, it decays exponentially every α frames.

Notice from the expression for α that, if we pick the smallest allowed value of m, then the required α (and hence the required delays) will be large. However, we are allowed to tradeoff m and α. If we let m grow linearly with n, then we will only need $\alpha \approx r^4$, which is, in fact, close to the minimum required delay of r. This also matches what is seen in existing works on provable ST in other settings (e.g., robust ST, ST with missing data, or streaming PCA with missing data) (Narayanamurthy & Vaswani, 2018; Mitliagkas et al., 2014). These are able to allow close to optimal detection and tracking delays but all these assume that m increases linearly with n. Also, the only other work that can also provably handle time-varying (piecewise constant) subspaces is (Narayanamurthy & Vaswani, 2018).

Improved Phaseless ST: PST-all. Notice from Theorem 3.1 that Algorithm 2 can only provably detect and track subspace changes that are larger than a small threshold. While this makes sense for detection, it should be possible to track all types of changes. By including a simple modification in Algorithm 2 (include the "update" step during the detection



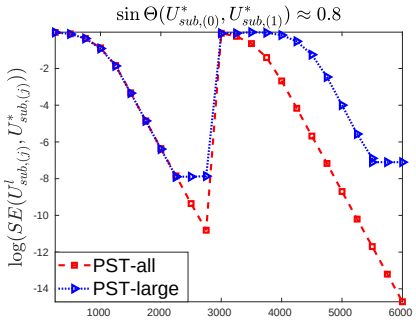


Figure 3. Plot of subspace error versus time at each α frames. Notice that for the cases where $\mathrm{SE}(\mathbf{U}_0^*, \mathbf{U}_1^*) = 0.8$ both algorithms are able to detect and track changes whereas when $\mathrm{SE}(\mathbf{U}_0^*, \mathbf{U}_1^*) = 0.01, 0.2$ only the PST-all algorithm works.

mode as well), we can empirically demonstrate that this is indeed true. See Fig 3. The proof that this is the case should also be possible, but needs careful changes to the analysis of the subspace update step. It requires the ability to deal with two different subspaces in one α-frame epoch. This is part of ongoing work. The modified algorithm is presented in (Nayer et al., 2019)

4. Related Work

Other work on phaseless low-rank recovery. The only other work that also studies Ph-Co-LRMR is (Vaswani et al., 2017). This introduced a series of heuristics to solve the Ph-Co-LRMR problem and evaluated them experimentally. It also attempted to provide a guarantee for obtaining an initial estimate of $\mathbf{X}^*$ also via a spectral initialization method. However, their guarantee is far from optimal even in the setting of small r. If we compare their main result (their Theorem 3.2) with ours, (i) it required the following lower bound on just m: $m \geq C \max(\sqrt{n}, r^4)/\epsilon^2$ in addition to a lower bound on mq that also depends on $1/\epsilon^2$. The dependence on $1/\epsilon^2$ is what makes this very large. (ii) However they only analyzed initialization, so another way is to compare their guarantee with just our Lemmas 5.1 and 5.3. They still need a lower bound of $m \geq C\sqrt{n}$ which is a significantly stronger requirement than ours. Also the lower bound of r^3 (ii) Finally, their result also needed a bound on each entry of $\tilde{\mathbf{b}}_k^*$ rather than just right incoherence.

There is no other work on phaseless ST except a recent short conference paper (Nayer & Vaswani, 2018) that proposed a complicated algorithm which relied on impractical assumptions of subspace change. It does not contain any guarantees.

Another seemingly related work is (Sanghavi et al., 2017). This attempts to recover an $n \times r$ matrix $\mathbf{U}^*$ from measurements $\mathbf{y}_i = \|\mathbf{a}_i' \mathbf{U}^*\|^2$. If $r = 1$, this is the standard PR problem. In the general case, this is related to covariance sketching, but not to our problem.

Dynamic structured data recovery. Our work can be interpreted as another key addition to a decade long body of work on dynamic structured high-dimensional data re-

Algorithm 2 Phaseless ST

$\hat{k}_0 \leftarrow 0, j \leftarrow 0, \ell \leftarrow 0$
Mode $\leftarrow$ update
for $k \geq 0$ **do**
  if Mode $=$ update **then**
    if $k = \hat{k}_j + (\ell+1)\alpha$ **then**
      if $\ell = 0$ **then**
        $\hat{\mathbf{U}}_{j,\ell} \leftarrow$ top r singular vectors of $\mathbf{Y}_u$
      end if
      $\hat{\mathbf{b}}_\tau \leftarrow RWF((\mathbf{y}_\tau, \hat{\mathbf{U}}_{j,\ell}{}'\mathbf{A}_\tau), T_{RWF,t})$, for $\tau \in [k-\alpha+1, k]$
      Compute QR decomposition $\hat{\mathbf{B}} \overset{QR}{=} \mathbf{R}_B \mathbf{B}$
      $\hat{\mathbf{C}}_\tau \leftarrow \text{Phase}\left(\mathbf{A}_\tau' \hat{\mathbf{U}}_{j,\ell} \hat{\mathbf{b}}_\tau\right)$, for $\tau \in [k-\alpha+1, k]$
      $\hat{\mathbf{U}}_{j,\ell+1} \leftarrow \arg\min_{\mathbf{U}} \sum_{\tau \in [k-\alpha+1,k]} \|\hat{\mathbf{C}}_\tau \mathbf{y}_\tau - \mathbf{A}_\tau{}'\mathbf{U}\hat{\mathbf{b}}_\tau\|^2$
      Compute QR decomposition $\hat{\mathbf{U}}_{j,\ell+1} \overset{QR}{=} \mathbf{U}_{j,\ell+1}\mathbf{R}_U$
      $\ell \leftarrow \ell + 1$
    end if
    if $\ell = L$ **then**
      $\mathbf{U}_j \leftarrow \mathbf{U}_{j,\ell}$, Mode $\leftarrow$ detect
    end if
  end if
  if Mode $=$ detect **then**
    if $\left(\lambda_{\max}\left(\mathbf{Y}_{U,\text{det}}\right) - \lambda_{\min}\left(\mathbf{Y}_{U,\text{det}}\right) \geq w\right)$ **then**
      $j \leftarrow j+1, \hat{k}_j \leftarrow k, \ell \leftarrow 0$, Mode $\leftarrow$ update
    end if
  end if
  Output $\mathbf{U}^{(k)} \leftarrow \hat{\mathbf{U}}_{j,\ell}$
end for Offline PST: For each $k \in [\hat{k}_j, \hat{k}_{j+1})$, output $\hat{\mathbf{x}}_k = \hat{\mathbf{U}}\hat{\mathbf{b}}_k$ with $\hat{\mathbf{U}} = basis([\hat{\mathbf{U}}_j, \hat{\mathbf{U}}_{j+1}])$ and $\hat{\mathbf{b}}_k$ is a (at most) $2r$-length vector obtained by RWF.

covery. Other problems in this category that have been extensively studied include dynamic compressive sensing (Vaswani & Zhan, 2016), dynamic robust PCA (or robust ST), see (Narayanamurthy & Vaswani, 2018) and references therein, and ST with missing data (Balzano et al., 2018). In terms of works with complete provable guarantees, there is the nearly optimal robust ST (NORST) approach and its precursors (Narayanamurthy & Vaswani, 2018), and recent papers on streaming PCA with missing data (Mitliagkas et al., 2014; Gonen et al., 2016). For robust ST, the problem setting itself implies $m = n/2$, while in the streaming PCA case, $m = cn$ is used. This is why both achieve close to optimal tracking delays (at least when the added unstructured noise is nearly zero). As noted earlier, our method can also achieve this if we let m grow with n. Unlike our work, most of these approaches (except NORST) cannot work with time-varying subspaces though.

Existing work on low-rank matrix recovery. Low-rank matrix sensing (LRMS) involves recovering a low-rank $\mathbf{X}^*$ from $\mathbf{y}_i = \langle \mathbf{A}_i, \mathbf{X}^* \rangle$ with $\mathbf{A}_i$ being dense matrices. This is the easier "global" measurements' setting, i.e., each $\mathbf{y}_i$ contains information about the entire $\mathbf{X}^*$. Initial solution approaches for LRMS borrow ideas from the compressive sensing literature, which is another instance of a problem with global measurements (of the sparse vector). In both these cases, it is possible to prove a simple (sparse or low-rank) restricted isometry property (RIP) which simplifies the rest of the analysis. Our problem setting is different from and more difficult than the above two since the measurements are not global. In this sense it can be compared to low-rank matrix completion (LRMC) but of course its measurement model is not like LRMC either. LRMC involves completely local measurements. Because of this, to allow for correct "interpolation", it requires assuming that $\mathbf{X}^*$ has dense rows and columns. This is imposed by assuming denseness (incoherence) of its left and right singular vectors. In our setting, since we have global measurements of each column, but not of the entire matrix, only right incoherence suffices.

Thus, even the linear version of our setting (suppose phase information was available) is clearly different from both LRMS and LRMC. However since no other complete guarantees exist for our problem or even for its linear version, we briefly compare our sample complexity with that of LRMC methods. The first iterative LRMC solution, AltMinComplete (Netrapalli et al., 2013a), needed a sample complexity of about $nr^5 \log(1/\epsilon)$. This is comparable to what we need. The best known guarantee is for a projected gradient descent solution from (Cherapanamjeri et al., 2016) and this needs $\Omega(nr^2 \log^2 n \log^2(\kappa/\epsilon)$ samples.

5. Proof of Theorem 2.1: overall lemmas

The proof is any easy consequence of the lemmas stated below and are proved in (Nayer et al., 2019).

Observe that $\mathbf{U}^0 = \hat{\mathbf{U}}^0$ since this initial estimate is obtained by SVD. We will also sometimes refer to it as $\mathbf{U}_{\text{init}}$.

Lemma 5.1 (Rank estimation and Initialization of $\mathbf{U}^*$)**.** *Pick a $\delta_{\text{init}} < 0.25$. Assume $mq \geq \kappa^4 nr^3/\delta_{\text{init}}^2$ and $m \geq \kappa^4 r^2 \log\max(n,q)/\delta_{\text{init}}^2$. Set the rank estimation threshold $\omega = 0.3{\sigma^*_{\min}}^2/q$. Assume also that $\mu\kappa \leq C$. Then, w.p. at least $1 - Cn^{-10}$, the rank is correctly estimated and*

$$\text{SE}(\mathbf{U}^*, \mathbf{U}_{\text{init}}) \leq \delta_{\text{init}}$$

Definition 5.2. *Define $\mathbf{g}_k^t := \mathbf{U}^{t\prime}\mathbf{x}_k^*$ and $\mathbf{e}_k := (\mathbf{I} - \mathbf{U}^t\mathbf{U}^{t\prime})\mathbf{x}_k^*$.*

It is easy to see that $\mathbf{x}_k^* = \mathbf{U}\mathbf{g}_k + \mathbf{e}_k$ and so $\mathbf{y}_{ik} = |(\mathbf{U}'\mathbf{a}_{ik})'\mathbf{g}_k + \mathbf{a}_{ik}'\mathbf{e}_k|$. Thus, we have a noisy PR problem to solve with the noise magnitude proportional to $\|\mathbf{e}_k\| \leq \text{SE}(\mathbf{U}^t, \mathbf{U}^*)\|\mathbf{x}_k^*\|$. Also, the solution estimates $\mathbf{g}_k^t = (\mathbf{U}^{t\prime}\mathbf{U}^*)\tilde{\mathbf{b}}_k^*$ which is just a rotated version of $\tilde{\mathbf{b}}_k^*$. We show in the next lemma that the error $\text{dist}(\mathbf{g}_k^t, \hat{\mathbf{b}}_k^t)$, is proportional to $\text{SE}(\mathbf{U}^t, \mathbf{U}^*)$. By triangle inequality the same is true for the error in $\hat{\mathbf{x}}_k^t := \mathbf{U}^t\hat{\mathbf{b}}_k^t$.

Lemma 5.3 (Recovery of $\tilde{\mathbf{b}}_k^*$'s)**.** *At iteration t, assume that $\text{SE}(\mathbf{U}^*, \mathbf{U}^t) \leq \delta_t$. Pick a $\delta_b < 1$. If $m \geq Cr$, and if we set $T_{RWF,t} = C\log\delta_t/\log(1-c)$, then w.p. at least $1 - 2q\exp\left(-c\delta_b^2 m\right)$, the following is true for each $k = 1, 2, \cdots, q$*

$$\begin{aligned}
\text{dist}\left(\mathbf{g}_k^t, \hat{\mathbf{b}}_k^t\right) &\leq C\delta_t\|\tilde{\mathbf{b}}_k^*\| = C\delta_t\|\mathbf{x}_k^*\|, \\
\text{dist}(\hat{\mathbf{x}}_k^t, \mathbf{x}_k^*) &\leq C\delta_t\|\mathbf{x}_k^*\|, \\
\textit{mat-dist}(\mathbf{G}^t, \hat{\mathbf{B}}^t) &\leq C\delta_t\|\tilde{\mathbf{B}}^*\|_F = C\delta_t\|\mathbf{X}^*\|_F \qquad (4)
\end{aligned}$$

Thus, if $m \geq C\max(r, \log n, \log q)/\delta_b^2$, then the above bounds hold w.p. at least $1 - Cn^{-10}$.

If each $\hat{\mathbf{b}}_k^t$ is close enough to $\mathbf{g}_k^t$ (which is a rotated version of $\tilde{\mathbf{b}}_k^*$), then, we would expect $\mathbf{B}^t$ to also satisfy the incoherence assumption. We show next that this is indeed true when $\delta_t \leq \frac{0.25}{C\sqrt{r}\kappa}$.

Lemma 5.4 (Incoherence of $\mathbf{B}^*$ implies incoherence of $\mathbf{B}^t$)**.** *At iteration t, assume that $\text{SE}(\mathbf{U}^*, \mathbf{U}^t) \leq \delta_t$ with $\delta_t \leq \frac{0.25}{C\sqrt{r}\kappa}$. If $\mathbf{B}^*$ is μ-incoherent, then, w.p. $1 - 2q\exp(-\delta_b^2 m)$, $\mathbf{B}^t$ is $\hat{\mu}$-incoherent with $\hat{\mu} = C\kappa\mu$.*

Finally, the next claims shows that the LS step to update $\hat{\mathbf{U}}$ reduces its error by a factor of 0.7 at each iteration. Its proof relies on the previous two lemmas.

Claim 5.5 (Descent Lemma)**.** *At iteration t, assume that $\text{SE}(\mathbf{U}^*, \mathbf{U}^t) \leq \delta_t$. If $\delta_t \leq \frac{c}{r\kappa^2}$, $mq \geq C\kappa^2\hat{\mu}^2 nr^2/\delta_t^2$ and*

$m \geq C \max(r, \log n, \log q)$ *then w.p. at least* $1 - cn^{-10}$,

$$\mathrm{SE}(\mathbf{U}^*, \mathbf{U}^{t+1}) \leq 0.7\delta_t := \delta_{t+1}.$$

Proof of Theorem 2.1. The $\mathrm{SE}(\mathbf{U}^*, \mathbf{U}^t)$ bounds are an immediate consequence of the initialization lemma, Lemma 5.1, and the above claim, Claim 5.5, along with setting $\delta_{\text{init}} = c/\kappa^2 r$ and $\delta_t = 0.7^{t-1}\delta_{\text{init}}$. The other bounds then follow by using Lemma 5.3. □

5.1. Proofs of the above lemmas: key ideas

The two lemmas that need the most effort to prove are Lemma 5.1 and Claim 5.5. The overall idea for proving Lemma 5.1 is borrowed from (Vaswani et al., 2017; Chen & Candes, 2015). But there are many important differences because we define $\mathbf{Y}_U$ differently in this work (the threshold in the indicator function now sums over all mq measurements). This simple change enables us to get a significantly improved result. It lets us use concentration over all the mq measurements (and design vectors) in each of the three steps of the proof. This is what helps eliminate the lower bound $m \geq Cr^4$ on just m that was needed in (Vaswani et al., 2017). However, this also means that the proofs are much more involved (more quantities now vary with k). It also means that we need the product $\mu\kappa$ to be a constant.

Overall idea of proof of Claim 5.5 is as follows. Using the top level approach of (Netrapalli et al., 2013a),

$$\mathrm{SE}(\mathbf{U}^*, \mathbf{U}^{t+1}) \leq \frac{\mathrm{MainTerm}}{0.9\sigma^*_{\min} - \mathrm{MainTerm}}.$$

with MainTerm is defined as

$$\frac{\max_{\mathbf{W}\in\mathcal{S}_W} |\mathrm{Term1}(\mathbf{W})| + \max_{\mathbf{W}\in\mathcal{S}_W} |\mathrm{Term2}(\mathbf{W})|}{\min_{\mathbf{W}\in\mathcal{S}_W} \mathrm{Term3}(\mathbf{W})}.$$

Here, $\mathcal{S}_W = \{\mathbf{W} \in \mathbb{R}^{n\times r} : \|\mathbf{W}\|_F = 1\}$ is the space of all $n \times r$ matrices with unit Frobenius norm,

$$\mathrm{Term1}(\mathbf{W}) := \sum_{ik} \mathbf{b}_k{}'\mathbf{W}'\mathbf{a}_{ik}\mathbf{a}_{ik}{}'\mathbf{U}^*(\tilde{\mathbf{B}}^*\mathbf{B}'\mathbf{b}_k - \tilde{\mathbf{b}}^*),$$

$$\mathrm{Term2}(\mathbf{W}) := \sum_{ik} (\mathbf{c}_{ik}\hat{\mathbf{c}}_{ik} - 1)(\mathbf{a}_{ik}{}'\mathbf{W}'\mathbf{b}_k)(\mathbf{a}_{ik}{}'\mathbf{x}^*_k),$$

$$\mathrm{Term3}(\mathbf{W}) := \sum_{ik} (\mathbf{a}_{ik}{}'\mathbf{W}'\mathbf{b}_k)^2,$$

and $\mathbf{c}_{ik}, \hat{\mathbf{c}}_{ik}$ are the phases (signs) of $\mathbf{a}_{ik}{}'\mathbf{x}^*_k$ and $\mathbf{a}_{ik}{}'\hat{\mathbf{x}}_k$.

Consider a fixed $\mathbf{W}$ first. Pick a $\delta < 1$. We can show that, if mq is large enough, whp,

$$(1-\delta)m \leq \mathrm{Term3} \leq (1+\delta)m$$

To bound Term1, we begin by showing that $\mathbb{E}[\mathrm{Term1}] = 0$. We then use a simple modification of the Bernstein-type inequality of (Vershynin, 2012), followed by using careful linear algebra tricks and Lemma 5.3 to show that, if mq is large enough, whp,

$$|\mathrm{Term1}| \leq m\delta\|\tilde{\mathbf{B}}^*(\mathbf{B}'\mathbf{B} - \mathbf{I})\| \leq Cm\delta\delta_t\|\mathbf{X}^*\|_F$$

for any $\delta < 1$. For Term2, we use Cauchy-Schwarz to get

$$|\mathrm{Term2}| \leq \sqrt{\mathrm{Term3}}\sqrt{\mathrm{Term31}}, \text{ where}$$
$$\mathrm{Term31} := \sum_{ik} (\mathbf{c}_{ik}\hat{\mathbf{c}}_{ik} - 1)^2(\mathbf{a}_{ik}{}'\mathbf{x}^*_k)^2$$

Consider Term31. Notice that $(\mathbf{c}_{ik}\hat{\mathbf{c}}_{ik} - 1)^2$ takes only two values - zero or one. It is zero when the signs are equal, else it is one. Thus, to start bounding $\mathbb{E}[\mathrm{Term31}]$, we can use Lemma 1 of (Zhang et al., 2016). The idea is this: the probability that the signs are unequal is large only when $(\mathbf{a}_{ik}{}'\mathbf{x}^*_k)^2$ is comparable in magnitude to $\mathrm{dist}^2(\mathbf{x}^*_k, \hat{\mathbf{x}}_k)$; and is very small otherwise. Since we have already bounded this distance by a small value (Lemma 5.3), whp, we are able to show that, with the same probability, $\mathbb{E}[\mathrm{Term31}] \leq m\delta_t^3\|\mathbf{X}^*\|_F^2$. Careful use of concentration bounds then implies that, if mq is large enough, whp, the same order bound holds for $|\mathrm{Term31}|$. Thus,

$$|\mathrm{Term2}| \leq Cm\sqrt{1+\delta}\sqrt{\delta_t}\delta_t\|\mathbf{X}^*\|_F$$

The above provides main ideas, the actual proofs need carefully developed epsilon-net arguments to bound the max or min of each of the terms over $\mathcal{S}_W$ (the argument for the min needs particular care). Combining everything and using $\|\mathbf{X}^*\|_F \leq \sqrt{r}{\sigma^*_{\max}}^2$, we can conclude that $\mathrm{SE}(\mathbf{U}^*, \mathbf{U}^{t+1}) \leq C(\delta + \sqrt{\delta_t})\sqrt{r}\kappa^2\,\delta_t$. We then set δ, δ_t to show that this bound is below $0.7\delta_t$.

6. Numerical Evaluation

Ph-Co-LRMR. Here we demonstrate the superiority of our algorithm with respect to few existing PR algorithms. We consider r unknown setting and show that our algorithm works well while performed for 100 independent trials.The results are summarized in Fig. 2. Notice that our algorithm outperforms existing techniques, The complete details of experiments are provided in (Nayer et al., 2019).

Dynamic Ph-Co-LRMR. Here we validate the PST and PST-all algorithms. For this experiment we generate the data similarly as in the previous experiment and we use $n = 300$, $r = 2$, $k_1 = 2992$, $q_{\text{full}} = 6000$, $m = 100$, $\alpha = 250$ and consider two values of subspace change: $\mathrm{SE}(\mathbf{U}^*_0, \mathbf{U}^*_1) = 0.01, 0.8$. The results for the two algorithms are shown in Fig. 3. Notice that in the first case PST does not improve the estimation error while PST-all does and in the second case both algorithms succeed.

Video Reconstruction. For the video results we consider the Coded Diffraction Pattern (CDP) measurements of the video. We provide the visual results in Fig 1. The details regarding the setup can be found in (Nayer et al., 2019).

References

Balzano, L., Chi, Y., and Lu, Y. M. Streaming pca and subspace tracking: The missing data case. *Proceedings of IEEE*, 2018.

Candes, E. J., Eldar, Y. C., Strohmer, T., and Voroninski, V. Phase retrieval via matrix completion. *SIAM J. Imaging Sci.*, 6(1):199–225, 2013a.

Candes, E. J., Strohmer, T., and Voroninski, V. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Comm. Pure Appl. Math.*, 2013b.

Candes, E. J., Li, X., and Soltanolkotabi, M. Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Trans. Info. Th.*, 61(4):1985–2007, 2015.

Chen, Y. and Candes, E. Solving random quadratic systems of equations is nearly as easy as solving linear systems. In *Adv. Neural Info. Proc. Sys. (NIPS)*, pp. 739–747, 2015.

Cherapanamjeri, Y., Gupta, K., and Jain, P. Nearly-optimal robust matrix completion. *ICML*, 2016.

Fienup, J. R. Phase retrieval algorithms: a comparison. *Applied optics*, 21(15):2758–2769, 1982.

Gerchberg, R. W. and Saxton, W. O. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 1972.

Gonen, A., Rosenbaum, D., Eldar, Y. C., and Shalev-Shwartz, S. Subspace learning with partial information. *The Journal of Machine Learning Research*, 17(1):1821–1841, 2016.

Jaganathan, K., Oymak, S., and Hassibi, B. Recovery of sparse 1-d signals from the magnitudes of their fourier transform. In *IEEE Intl. Symp. on Information Theory (ISIT)*, pp. 1473–1477. IEEE, 2012.

Jagatap, G., Chen, Z., Hegde, C., and Vaswani, N. Model corrected low rank ptychography. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 3988–3992. IEEE, 2018.

Mitliagkas, I., Caramanis, C., and Jain, P. Streaming pca with many missing entries. *Preprint*, 2014.

Narayanamurthy, P. and Vaswani, N. Nearly optimal robust subspace tracking. In *International Conference on Machine Learning*, pp. 3701–3709, 2018.

Nayer, S. and Vaswani, N. Phaseless subspace tracking. *IEEE conference GlobalSIP (arXiv preprint arXiv:1809.04176)*, 2018.

Nayer, S. and Vaswani, N. Phast: Model-free phaseless subspace tracking. *IEEE conference on ICASSP*, 2019.

Nayer, S., Narayanamurthy, P., and Vaswani, N. Phaseless pca: Low-rank matrix recovery from column-wise phaseless measurements. *arXiv preprint arXiv:1902.04972*, 2019.

Netrapalli, P., Jain, P., and Sanghavi, S. Low-rank matrix completion using alternating minimization. In *STOC*, 2013a.

Netrapalli, P., Jain, P., and Sanghavi, S. Phase retrieval using alternating minimization. In *Adv. Neural Info. Proc. Sys. (NIPS)*, pp. 2796–2804, 2013b.

Sanghavi, S., Ward, R., and White, C. D. The local convexity of solving systems of quadratic equations. *Results in Mathematics*, 71(3-4):569–608, 2017.

Shechtman, Y., Beck, A., and Eldar, Y. C. Gespar: Efficient phase retrieval of sparse signals. *IEEE Trans. Sig. Proc.*, 62(4):928–938, 2014.

Szameit, A., Shechtman, Y., Osherovich, E., Bullkich, E., Sidorenko, P., Dana, H., Steiner, S., Kley, E. B., Gazit, S., Cohen-Hyams, T., Shoham, S., Zibulevsky, M., Yavneh, I., Eldar, Y. C., Cohen, O., and Segev, M. Sparsity-based single-shot subwavelength coherent diffractive imaging. *Nature Materials*, 11:455–9, April 2012. doi: 10.1038/nmat3289.

Vaswani, N. and Zhan, J. Recursive Recovery of Sparse Signal Sequences from Compressive Measurements: A Review. *IEEE Trans. Sig. Proc.*, 64 (13):3523–3549, 2016.

Vaswani, N., Nayer, S., and Eldar, Y. C. Low rank phase retrieval. *TSP*, August 2017.

Vershynin, R. Introduction to the non-asymptotic analysis of random matrices. *Compressed sensing*, pp. 210–268, 2012.

Wang, G., Giannakis, G. B., and Eldar, Y. C. Solving systems of random quadratic equations via truncated amplitude flow. *arXiv preprint arXiv:1605.08285*, 2016.

Zhang, H., Zhou, Y., Liang, Y., and Chi, Y. Reshaped wirtinger flow and incremental algorithm for solving quadratic system of equations. In *Adv. Neural Info. Proc. Sys. (NIPS)*, 2016.