Sub-diffraction Super-resolution Imaging for Structured Data

Gauri Jagatap, Zhengyu Chen, Seyedehsara Nayer, Chinmay Hegde, *Senior Member, IEEE* and Namrata Vaswani, *Fellow, IEEE*

***Abstract*—We study the problem of recovering structured data from ptychographic measurements. Ptychography is a image acquisition scheme that uses an array of images to produce high-resolution images in microscopy as well as long-distance imaging, to mitigate the effects of diffraction blurring. The number of measurements are typically much larger than the size of the signal (image or video) to be reconstructed, which translates to high storage and computational requirements.**

The issue of high sample complexity can be alleviated by utilizing structural properties of the image (or video). In this paper, we first discuss a range of sub-sampling schemes which can reduce the amount of measurements in ptychographic setups; however, this makes the problem ill-posed. Correspondingly, we impose structural constraints on the signals to be recovered, to regularize the problem. Through our novel framework of recovery algorithms, we show that one can reconstruct high-resolution images (or video) from fewer samples, via simple and natural assumptions on the structure of the images (or video). We demonstrate the validity of our claims through a series of experiments, both on simulated and real data.

***Index Terms*—Phase retrieval, ptychography, structure, sparse, low-rank, sub-diffraction imaging, super-resolution.**

I. Introduction

A. Motivation

A COMMON problem in microscopy and long-distance imaging is diffraction blurring. When the aperture of the imaging lens is much smaller in comparison to (i) the size of the object to be imaged [4], or (ii) the distance of the object to be imaged [5], a diffraction pattern is observed. When the spatial resolution of the object is smaller than the diameter of this pattern, the image formed at the sensing plane is typically blurred. Consequently, the limited angular extent of the input aperture leads to significant loss in spatial resolution, and designing methods for super-resolution in diffraction-blurred imaging systems is of considerable interest.

Fourier ptychography [4] is a technique which mitigates the effects of diffraction blurring by constructing a large *synthetic* aperture. Practically, this setup can be implemented by either spatially moving a single camera aperture [6], or by an array of fixed cameras [4], similar to those used in light-field cameras; each of the cameras measure different parts of the Fourier spectrum of the desired images. The image formation at the sensing plane is typically complex in nature, due to phase shifts induced by the optical lens setup. However, the sensing apparatus is incapable of estimating the phase of the complex values, and only the magnitudes can be measured.

This setup can be molded to that of the classical problem of *phase retrieval* [7], [8], [9], which is a non-linear, ill-posed inverse problem. In phase retrieval, the goal is to reconstruct a discretized image (or video) of size n (or nq) from noisy, magnitude-only observations of the image's discrete Fourier transform (DFT) coefficients. A generalized version of this problem replaces the DFT coefficients with a generic linear operator constructed by sampling certain families of probability distributions. Several algorithmic approaches for this generalized case have emerged in the recent literature, accompanied by strong theoretical guarantees on the accuracy of reconstruction [10], [11], [12], [13], [14].

A fundamental challenge in Fourier ptychography is the requirement of an over-complete set of observations. To reconstruct a length-n signal, one requires $m \gg n$ samples. This value of m can be typically very large, which can pose severe limitations in terms of data storage and computational load. To reduce this sample complexity, one can leverage low-dimensional modeling assumptions made on the signal. Exploitation of low-dimensional structures in signals has been well studied in the case of linear measurements. For instance, a natural structural assumption on image data is sparsity [15]. Further, more refined structured sparsity assumptions (such as block sparsity) can also be imposed to enable image reconstruction from an even smaller set of measurements [16], [17], [18].

Similarly, for video data, one can consider the scenario of estimating a dynamic slowly changing scene with a moving target. Then, without structural assumptions, for a video with q frames, one requires $m = \Omega(nq)$ measurements. To alleviate this, a *low-rank* assumption can be imposed on the video in order to reduce the sample complexity, a concept which has been well exploited in recent literature [19].

B. Our contributions

In this paper, we design and validate a series of sample-efficient algorithms for sub-diffraction imaging using the Fourier ptychography framework that exploits structure. Moreover, we introduce two practical "sub-sampling" strategies for Fourier ptychography. These strategies can be easily incorporated into pre-existing measurement setups. In particular, we make the following contributions:

1) We leverage underlying (structured) sparsity of natural image data in various transform domains, to present a

This work was supported in part by NSF grants CCF-1566281, CCF-1815101, CAREER CCF-1750920, and a gift from the Black and Veatch Foundation. The authors would like to thank Lei Tian and Ashok Veeraraghavan for useful discussions. Parts of this paper have appeared in the IEEE ICASSP 2018 conference [1], [2] and the IEEE ICIP 2018 conference [3].

family of reconstruction algorithms for recovering super-resolved sparse images from sub-sampled measurements.
2) We leverage underlying low-rank structure in video data and propose a novel reconstruction algorithm for recovering super-resolved slowly changing videos from sub-sampled measurements.
3) We propose a model-error correction strategy for our low-rank ptychography algorithm which accounts for inaccuracies in estimating the low-rank nature of data correctly.
4) We support our claims for reduced sample complexity requirements through a series of experiments, on both synthetically generated and real data.

Sparse data model: For sparse image data, we propose an approach based on a line of previous work [20], [21] wherein we had developed an algorithmic framework for improving sample-complexity of classical phase retrieval. This paper extends this line of work to the (more practically relevant) setting of Fourier ptychography.

Low-rank data model: For video data which satisfies the low-rank model, we adapt the algorithmic framework introduced in [22], [23] and extend to the setting of Fourier ptychography. For real-world videos that need not fit the low-rank model perfectly, we propose a novel modeling-error correction stage which allows for application of our approach to a broad class of video data.

II. Prior Work

A. Fourier ptychography

In the literature on Fourier ptychography, the majority of papers focus on the experimental merits of the procedure [6], [4], [24], [25], albeit without structural constraints. Recent work [26], [27], [28] provides analysis on the convergence guarantee of phase retrieval problem for Short Time Fourier Transform (STFT) measurements, which can be extended to the setting of ptychography; however, only simple test cases (that consider 1-D signals of specific length) have been analyzed until now.

In [29] the authors discuss the experimental robustness of various phase retrieval algorithms in the context of Fourier ptychography, and conclude that amplitude-based recovery methodologies are more effective in combating noise, aberrations and model mismatch.

In [5], authors proposed a way of adapting this super-resolution methodology for long-distance imaging, which they solve via alternation minimization. There exist several choices for the phase retrieval procedure in all of these setups. Most papers utilize first-order methods such as Wirtinger flow [30], [31] and Alternating Minimization [5]. Meanwhile in [32], [33], the authors use a Newton-step based alternating gradient descent, for the same setup.

Exploiting structure in the context of ptychography had not been explored in literature until very recently. Zhang et. al. study the problem of exploiting sparsity with threshold-based gradient descent [34], [35]. However they use sparsity as a regularization and do not study the problem in the context of under-sampled measurements. Our method explicitly addresses the sample-complexity issue, and is extensible to a large class of structured sparsity models.

Very recently, Shamshad et. al. [36] discuss a deep generative priors strategy for sub-sampled Fourier Ptychography under sparsity priors. Since their methodology is training-based, it requires large number of example images to learn the generative model accurately. This can be highly prohibitive in the context of microscopic or long-distance images, as the acquisition time and costs associated with generating such datasets will be very high.

To the best of our knowledge, there does not exist any prior work that considers low-rank structure in the context of ptychography.

B. Phase retrieval

Initially studied in the 1970s [37], phase retrieval is a classic problem and challenge in optical imaging and signal processing area. Traditionally, the alternating minimization framework is utilized; one can estimate the missing phase information of the measurements, and subsequently the signal coefficients, within the same iteration of this algorithm. Since this problem is inherently non-convex in nature, convergence of such algorithm to the desired ground truth signal value, is not always guaranteed, unless initialized properly [1]. For the case of multi-variable Gaussian measurements, Netrapalli et. al. provide the first set of guarantees [13].

Subsequently, a gradient descent based approach, which utilizes the Wirtinger gradient [12], [40] to minimize an ℓ_2-squared empirical loss function was developed, for Gaussian as well as Coded Diffraction Pattern (CDP) measurements. This line of work as well as subsequent papers[14], [41], [42] is now well established with near-optimal results.

Similarly, convex formulations of the same problem exist, with the majority of algorithms relying on *lifting* the problem from an n-dimensional space to an n^2-dimensional space, and attempting to solve a low-rank constrained problem in the larger space [10]. However, these methods are computationally expensive.

C. Sparse phase retrieval

Sparsity assumptions have recently been introduced in the context of phase retrieval. A series of approaches have emerged that use alternating minimization [13], [20], convex relaxation [41], [38], [43] and iterative thresholding [44], [45]. In all of the above, authors give a sample complexity of $\mathcal{O}\left(s^2 \log n\right)$ for stable recovery for s-sparse signals. In case of $s \ll n$, this result is an improvement compared to the standard requirement of $\mathcal{O}(n)$ measurements. Additionally, subsequent work [20], [21] suggests that modeling the sparsity into specific structures such as blocks or trees, leads to a lowered sample complexity (to $\mathcal{O}(s \log n)$). Related other works also show a similar complexity ($\mathcal{O}(s \log n)$), albeit for some more carefully designed measurements [46], [47].

[1]Exceptions to this are [38],[39], however this comes at the cost of higher computational or sample complexity.

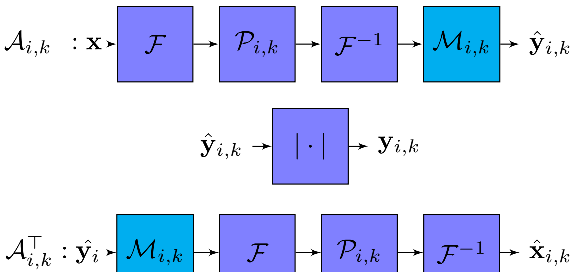


Fig. 1: *Sampling procedure for single image, using operator* $\mathcal{A}_{i,k}$. $\mathcal{M}_{i,k}$ *indicates the sub-sampling step. Indices* i *and* k *correspond to different cameras and video frames respectively.*

D. Low-rank matrix recovery

In classic signal processing, the low-rank matrix recovery problem has been studied in the context of matrix completion and robust PCA [48], [49], [50]. Our previous work [22] gave the first result on using low-rank model in the context of phase retrieval. However, all of the works mentioned above require generic linear matrix measurements, and the applicability of such methods for Fourier Ptychography has not been studied thus far.

III. Paper Outline

We describe the paper organization in detail. In Section IV, we describe lay the groundwork for the Fourier Ptychography measurement model used in the rest of the paper. In particular, in Section IV-A, we introduce the optical setup used to acquire conventional ptychography measurements. In Section IV-B we discuss sub-sampling strategies to reduce the number of measurements. In Section IV-C, we introduce the mathematical formulation for the measurement setup. In Section IV-D, we discuss the conventional reconstruction procedure used for inverting ptychographic measurements.

Further, we discuss signal reconstruction under our two main structural assumptions. In Section V, we establish the still image data model, with a sparsity prior and set up the main optimization problem. In Section VI, similarly, we establish the video data model, with a low-rank prior and the corresponding optimization problem. In both Sections V and VI, we introduce and describe our algorithms for reconstructing structured data from sub-sampled Fourier ptychography measurements.

We first report our experimental findings for sparse ptychography, in Section VII, for simulation (Section VII-A) and real data (Section VII-B) measurements. We then report our experimental findings for low-rank ptychography, in Section VIII, for simulation (Section VIII-A) and real data (Section VIII-B) measurements. Finally, in Section IX, we compare our sparsity and low-rank models in the context of the measurement setup described in Section IV.

IV. Fourier Ptychography Setup

A. Optical setup

The setup in Fourier ptychography, such as that described in [5], [33], involves imaging an object using a series of optical sensing operations. The object is illuminated by coherent light. The transformed beam of light from the illumination pattern then passes through a thin lens which is located in front of the object, leading to a thin lens effect that can be modeled via a Fourier transform operation. The Fourier domain image is captured by a camera array with limited-size aperture pupils. In the setting of [5], such camera array is realized by either a physical grid of N cameras, or by a single translating camera. In [33], the multi-camera setup is replaced by a single fixed lens but with grid of LEDs with programmable illumination angles or patterns. Effectively, both of these setups simulate a large synthetic aperture. The effect of the lens array on the image plane is equal to an inverse Fourier operation. Finally, the image (in the form of the light beam) is received by an optical sensor that records the absolute value of the complex image.

In this paper, in order to decrease sample complexity we also use an additional “sub-sampling” mask, in which we mute the measurements corresponding to a fraction of pixels (or cameras) constituting the measurement setup. This step is incorporated via an element-wise masking operation $\mathcal{M}$. This masking operation is discussed in further detail in Section IV-B. For capturing static images, the imaging procedure is summarized as in Figure 1. For capturing videos, the same setup is used except that different sub-sampling masks $\mathcal{M}$ are used for each of the q frames.

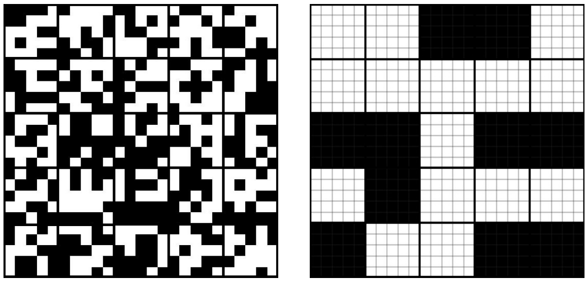
B. Sub-sampling strategies

Sub-sampling can be done in two ways: pixel-wise and camera-wise. Camera-wise sub-sampling corresponds to randomly switching off a different set of cameras at different times (refer Figure 2(b)), while pixel-wise corresponds to “switching off” different randomly selected pixels at different times (refer Figure 2(a)). Both strategies help save power (pixel-wise requires careful camera design in which individual pixel sensors can be turned off to save power). This strategy is similar to that used in compressed sensing literature [51]. Camera-wise sub-sampling can also result in a proportional reduction in data acquisition time in case “multiple cameras” are simulated by moving a single camera to different locations.

Random pixel patterns: We construct a sub-sampling mask in which the elements of the mask are picked up according to a Bernoulli distribution. If i is an index for a given camera in the camera array, then elements b_j^i corresponding to different pixels of a camera, are independent standard Bernoulli random variables. The mask resembles the operation of a diagonal matrix with 1s and 0s on the diagonal. Pixels corresponding to 1s are retained and those corresponding to 0s are discarded. A total of $m = f \times (nN)$ measurements are retained, *in expectation*, from all N cameras, where f denotes the fraction of samples (or pixels), and is also the probability associated with the Bernoulli random variable and n is the size of the original image frame. Figure 2 (a) represents an illustration.

In this case, for an input signal (vectorized image) $\mathbf{v} \in \mathbb{C}^n$, the sub-sampling mask operates as

$$\mathcal{M}_i(\mathbf{v})_j = b_j^i \cdot (\mathbf{v})_j, \tag{1}$$



(a) (b)

Fig. 2: *Construction of camera array masks via (a) random pixel and (b) random camera arrangements.*

where $\Pr(b^i_j = 1) = f$ and $\Pr(b^i_j = 0) = 1 - f$.

Randomly chosen cameras: Another sub-sampling strategy is to turn some cameras "on" or "off". We use sampling masks $\mathcal{M}_i$, which are picked up from a Bernoulli distribution $b \in \mathbb{R}^N$, with elements b_i being independent standard Bernoulli random variables. In terms of the sampling mask, for a vector input $\mathbf{v} \in \mathbb{C}^n$, the sub-sampling mask,

$$\mathcal{M}_i(\mathbf{v}) = b_i \cdot \mathbf{v}, \tag{2}$$

where $\Pr(b_i = 1) = f$ and $\Pr(b_i = 0) = 1 - f$. Figure 2 (b) represents an illustration of this setup.

C. Mathematical formulation of measurement setup

We discuss the mathematical model for recovering a multi-dimensional signal, from sub-sampled Fourier ptychographic measurements problem. We consider a matrix $\mathbf{X}$, with columns being vectorized images and q such images frames

$$\mathbf{X} := [\mathbf{x}_1, \ldots \mathbf{x}_k, \ldots, \mathbf{x}_q], \ \mathbf{X} \in \mathbb{C}^{n \times q}$$

where each frame is indexed by k. Henceforth, we denote the index set $\{1, \ldots q\}$ as $[q]$ for simplicity of notation. In the case of a single image frame, $q = 1$. For a video that is sufficiently slow changing, the rank of matrix $\mathbf{X}$ can be assumed to be no greater than r, where $r \ll \min(n, q)$. Each individual frame of the video $\mathbf{x}_k$ is fed to the measurement setup described in in Fig. 1. The measurements corresponding to a specific camera i, and image frame k, where i spans different cameras or LEDs ($i = 1, 2, \ldots, N$ or $i = [N]$ for simplicity of notation) is $\mathbf{y}_{i,k} \in \mathbb{R}^n$. The linear operators $\mathcal{A}_{i,k} : \mathbb{C}^n \to \mathbb{C}^n$ represent the series of operations represented in Fig. 1, prior to the camera sensor. Effectively, the measurements can be stacked into a long vector

$$\mathbf{y} = \begin{bmatrix} |\mathcal{A}_{1,1}(\mathbf{x}_1)| \\ \vdots \\ |\mathcal{A}_{i,k}(\mathbf{x}_k)| \\ \vdots \\ |\mathcal{A}_{N,q}(\mathbf{x}_q)| \end{bmatrix} = |\mathcal{A}(\mathbf{X})|$$

in which $\mathbf{y} \in \mathbb{C}^{nNq}$ and the measurement operators $\mathcal{A}_{i,k}$ can be stacked vertically into a long effective operator $\mathcal{A}$.

The forward operator $\mathcal{A}_{i,k}$ is effectively the sequence of operations:

$$\mathcal{A}_{i,k} = \mathcal{M}_{i,k} \mathcal{F}^{-1} \mathcal{P}_{i,k} \mathcal{F} \tag{3}$$

in which, $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fourier and inverse Fourier operations, and $\mathcal{P}_{i,k}$ is a pupil mask correspond to the i^{th} camera and k^{th} frame. The collection of operators $\{\mathcal{P}_{i,k}\}$, for all i, constitute a series of bandpass filters which cover different parts of the Fourier spectrum of a given frame k.

The sub-sampling mask $\mathcal{M}_{i,k}$ is different from camera to camera as well as from frame to frame.

D. Existing recovery methods

The problem of phase retrieval involves recovering a signal $\mathbf{x}$ (or single frame) from phase-less measurements of the form

$$\mathbf{y} = |\mathcal{A}(\mathbf{x})|.$$

A common recovery method uses alternating minimization [37], [13], which involves re-formulating the recovery as the solution to a non-convex problem:

$$\min_{\mathbf{C}, \mathbf{x}} \|\mathbf{y} - \mathbf{C} \cdot \mathcal{A}(\mathbf{x})\|_2 \,,$$

where the diagonal matrix $\mathbf{C} = \mathrm{diag}(\mathrm{phase}(\mathcal{A}(\mathbf{x})))$ captures the missing (complex) phase information from the measurements.

Algorithm 1 Alternating minimization for phase retrieval

Input: $\mathcal{A}, \mathbf{y}, t_0$
Initialize $\mathbf{x^0}$ s.t. $\min_\phi \left\|e^{i\phi}\mathbf{x}^0 - \mathbf{x}^*\right\|_2 \leq \delta \left\|\mathbf{x}^*\right\|_2$.
for $t = 0, \cdots, t_0 - 1$ **do**
  $\mathbf{C}^{t+1} \leftarrow \mathrm{diag}\left(\mathrm{phase}(\mathcal{A}(\mathbf{x^t}))\right)$,
  $\mathbf{x^{t+1}} \leftarrow \mathrm{argmin}_{\mathbf{x}} \left\|\mathcal{A}(\mathbf{x}^t) - \mathbf{C}^{t+1}\mathbf{y}\right\|_2$.
end for
Output $\mathbf{z} \leftarrow \mathbf{x}^{t_0}$.

Our recovery method is described in Algorithm 1. It involves an alternating procedure in which one estimates the missing phase information $\mathbf{C}$ and estimates the signal $\mathbf{x}$. A crucial requirement for the convergence of AltMinPhase is that a "good" initialization $\mathbf{x}^0$ is provided.

In the subsequent sections, we discuss the recovery of both sparse images and low-rank videos, in the context of the Fourier ptychography measurement setup. In Section IX, we compare these two models under the aforementioned sub-sampled measurement setup.

V. Still Image Data: Sparsity Model

In this section, we discuss an algorithm to estimate a single image from phaseless measurements using *fewer* samples than is required conventionally by alternating minimization. To do this, we utilize prior knowledge of the underlying sparsity of the image to formulate a new non-convex optimization problem:

$$\min_{\mathbf{x}} \sum_{i=1}^{N} \| |\mathcal{A}_i(\mathbf{x})| - \mathbf{y}_i \|_2^2, \ s.t. \ \mathbf{x} \in \mathfrak{M}_s^b, \tag{4}$$

Algorithm 2 Model-based CoPRAM for Ptychography

1: **Input:** $\mathcal{A}_1, \ldots \mathcal{A}_N, \mathbf{y}, s, b, t_0$

2: $x_j^0 \leftarrow \sqrt{\frac{1}{N}\sum_{i=1}^{N} y_{i,j}^2}$, j indexes signal entries $j = [n]$.

3: **for** $t = 0, \cdots, t_0 - 1$ **do**

4: $\quad \mathbf{C}^{t+1} \leftarrow \operatorname{diag}\left(\operatorname{sign}\left(\mathcal{A}(\mathbf{x^t})\right)\right)$,

5: $\quad \mathbf{x^{t+1}} \leftarrow \textsc{ModelCoSaMP}\Big(\frac{\mathcal{A}}{\sqrt{nN}}, \frac{\mathbf{C}^{t+1}\mathbf{y}}{\sqrt{nN}}, s, b, \mathbf{x}^t\Big)$.

6: **end for**

7: **Output** $\mathbf{z} \leftarrow \mathbf{x}^{t_0}$.

where $\mathbf{x}$ is a vectorized image. Here, $\mathfrak{M}_s^b$ is called a *sparsity model*, and denotes the set of all s-sparse signals whose non-zero coefficients can be grouped into blocks with uniform block length b. (The standard sparsity model can be represented by assigning $b = 1$.) To solve (4), we adapt the Compressive Phase Retrieval with Alternating Minimization (CoPRAM) framework, first introduced in [20], [21]. This procedure is shown in Algorithm 2.

The algorithm contains two stages: (i) initialization and (ii) sparse signal estimation, which we discuss in detail as follows.

A. Initialization

The initialization for solving the problem in (4) is a crucial step since the formulation is non-convex. It is therefore important to design an initialization that is as close to the ground truth of the signal to be recovered as possible. There exists a range of alternatives which can be chosen for this purpose, and we discuss this choice of initialization in detail in Section VII.

Typically in the literature, the choice of initialization is either (a) the observed intensity values from a small set of cameras placed at or near the center of the camera array [33] (b) an average of the intensity values from all cameras of the camera array [5]. In [5], the authors use the average $\frac{1}{N}\sum_i^N \mathbf{y}_i$, of the observed intensity values $\mathbf{y}_i$ from each camera, as the initial estimate $\mathbf{x}^0$. Another choice of initialization is to directly use the intensity values recorded by the central camera (indexed by $c \in [N]$), $\mathbf{y}_c$, which is essentially a low-resolution image that needs to be super-resolved.

In this paper, for the initialization stage, we improve upon the one given in [5] by using root-mean-squared measurements as the estimator $\sqrt{\frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_i^2}$, where $\mathbf{y}_i^2$ is an element-wise squaring operation (line 1 of Algorithm 2). We establish experimentally that this initialization is superior to that in [5]. A similar initialization strategy has been discussed in [28].

This is also a deviation from the conventional spectral initialization for phase retrieval as discussed in [13], [11], [20]. While a spectrally-obtained initial estimate succeeds for generic (Gaussian) linear measurements both in theory and practice, it unfortunately fails for the Fourier ptychographic setup. The intuition behind average or root mean squared initialization is as follows.

If the measurements were not phaseless, then $\mathbf{y}_{i,k}$ would contain random samples of a bandpass filtered version of the signal (with different i's corresponding to different random samples of different bands). Hence summing (or averaging) all the $\mathbf{y}_{i,k}$'s, would provide a good initial estimate of the $\mathbf{x}_k$. The same would also be true if the operation before the step of taking phaseless measurements returned a vector with all non-negative entries. In our setting, neither is exactly true, however the same idea still returns a good enough initial estimate. We believe the reason is that the image itself is all non-negative and hence its low-pass filtered measurements are definitely all non-negative as well. These likely dominate the summation, and because of this, the same approach works even though we are often removing the sign of negative entries as well (the higher frequency entries can be negative). Experimentally we have observed that instead of averaging, taking the root mean squared estimate gives a slightly better initial estimate. This is better because the large (low pass) entries dominate even more in this estimate than in a simple average.

B. Sparse signal estimation

Once we have a coarse estimate for the initialization of the CoPRAM algorithm, we then refine this estimate using a variant of alternating minimization. Specifically, at any given iteration, we first estimate the phase (line 3 of Algorithm 2) by applying the forward operator $\mathcal{A}$ to the signal estimate $\mathbf{x}^t$. Next, we assign this estimated phase into our observed intensity measurements, and subsequently obtain the next signal estimate $\mathbf{x}^{t+1}$ using a sparse recovery algorithm (line 4 of Algorithm 2) such as CoSaMP [52]. Moreover, in order to incorporate structural assumptions beyond sparsity, the only modification is to replace the sparse recovery method by any other stable structured sparse recovery method, such as model CoSaMP [17] (line 4 of Algorithm 2).

In [20] we have demonstrated (both theoretically and numerically) that the estimates $\mathbf{x}^{t+1}$ of the above alternating minimization technique for Gaussian measurements, converges to the solution $\mathbf{x}$ at a linear rate, using an appropriate termination condition.

The basic idea is that the "phase noise" induced due to the estimation error can be suitably bounded provided the initial estimate is good enough. Below, we empirically demonstrate that for the case of Fourier ptychographic measurements, similar gains can be achieved using our algorithm, as long as a good initialization is provided.

VI. Video Data: Low Rank Model

We develop a reconstruction method that exploits the assumption that a sequence of slowly changing images is often well approximated by a low rank matrix (with each column of the matrix being one image arranged as a 1D vector). For real videos, this means that the first few singular values of $\mathbf{X}$ contain most of the energy.

In the ideal scenario in which the video is *exactly* low-rank, the desired $\mathbf{X}$ will be the solution to the non-convex optimization problem:

$$\begin{aligned} \operatorname*{argmin}_{\mathbf{X}} \sum_{k=1}^{q}\sum_{i=1}^{N} \|\mathbf{y}_{i,k} - |\mathcal{A}_{i,k}(\mathbf{x}_k)|\|_2^2, \\ \text{s.t.} \quad \operatorname{rank}(\mathbf{X}) \le r, \end{aligned} \tag{5}$$

where r represents the rank-parameter. To solve (5), we adapt the *low-rank phase retrieval* (LRPR) algorithm in [22]. As above, our recovery algorithm consists of primarily two stages: (i) initialization, and (ii) low-rank matrix estimation. We call this adaptation the Low Rank Ptychography (LRPtych) algorithm.

In real-world applications, the exact low-rank assumption on the target video may not necessarily hold. Mathematically, the desired $\mathbf{X}$ can be written as $\mathbf{X} = \tilde{\mathbf{X}} + \mathbf{E}$ where $\mathbf{E}$ encodes the modeling error and $\tilde{\mathbf{X}}$ is *exactly* low rank.

To correct for this modeling error, we introduce an additional estimation stage. In this third stage, we invoke the *model correction* subroutine, to fix any errors that may have propagated due to inaccuracy in selecting the rank r, from the standard LRPtych algorithm. This stage, coupled with LRPtych, constitutes the Modified Low Rank Ptychography (or MLRPtych) framework. Mathematically, this represents the following optimization problem:

$$\hat{\mathbf{X}} := \tilde{\mathbf{X}} + \underset{\mathbf{E}}{\operatorname{argmin}} \sum_{k=1}^{q} \sum_{i=1}^{N} \| \mathbf{y}_{i,k} - |\mathcal{A}_{i,k}(\mathbf{x}_k + \mathbf{e}_k)| \|_2^2 \quad (6)$$

where $\mathbf{E} = [\mathbf{e_1}, \mathbf{e_2}, \ldots \mathbf{e_q}]$, $\mathbf{E} \in \mathbb{R}^{n \times q}$ is the modeling error.

In Algorithm 3, we summarize the three stages of our Modified Low Rank Ptychography algorithm. Our algorithm relies on the fact that a rank-r matrix $\mathbf{X}^*$ can be written as $\mathbf{X}^* = \mathbf{U}\mathbf{B}$, where $\mathbf{U}$ is a matrix of size $n \times r$ with mutually orthonormal columns, and $\mathbf{B}$ is a matrix of size $r \times q$.

In keeping with the requirements for phase retrieval algorithms, initialization is a key factor in obtaining an appropriate reconstruction of the video data matrix $\mathbf{X}$. For the low-rank matrix recovery stage, we introduce a subspace based alternating minimization method, which estimates the missing phase information and signal information in an alternating pattern. Further details of these three stages of Algorithm 3 are discussed below.

A. Initialization

The original LRPR algorithm used a spectral initialization approach that was a modification of the ideas in [12] to the low rank set up. However after experimental probing, we observe that borrowing the approach of LRPR does not work for the current application. We believe this is so because the measurement setup does not capture the properties of the Gaussian and CDP model discussed in [12].

Instead, we use the same initialization idea as described in Section V-A. We obtain the initial guess for each individual image frame as $\mathbf{x}_k^0 = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_{i,k}^2}$, where $\mathbf{y}_{i,k}^2$ is element-wise squared. Moreover, we follow this by computing a rank-r approximation of the resulting matrix and using its components to initialize $\mathbf{U}$ and $\mathbf{B}$. (Refer lines 1-5 of Algorithm 3 for this procedure).

A reduced singular value decomposition (reducedSVD) is applied on the video estimate $\mathbf{X}_0 = [\mathbf{x}_1^0, \ldots \mathbf{x}_q^0]$, with given rank r to obtain $\mathbf{U^0}, \mathbf{S^0}, \mathbf{V^0}$ respectively. This initialization ensures that the future estimates of $\mathbf{U}^t \in \mathbb{R}^{n \times r}$ estimate an r-dimensional subspace. Similarly, the corresponding coefficients in terms of $\mathbf{B}^0 = \mathbf{S}^0 \cdot \mathbf{V}^{0\top}$ are extracted.

Algorithm 3 Modified Low Rank Ptychography (MLRPtych)

(Initialization)
1: **Input:** $\mathbf{y}_k, \mathcal{A}_{i,k}, r$
2: $x_{k,j}^0 \leftarrow \sqrt{\frac{1}{N}\sum_{i=1}^{N} y_{i,k,j}^2}$, j indexes signal entries $j = [n]$.
3: $[\mathbf{U}^0, \mathbf{S}^0, \mathbf{V}^0] \leftarrow ReducedSVD((\mathbf{X}^0), r)$
4: $\mathbf{b}_k^0 \leftarrow (\mathbf{S}^0\mathbf{V}^{0\top})_k$, $k = [q]$.
(Low-rank matrix recovery stage)
5: **for** $t = 1, 2, \ldots, T$ **do**
6: a) $\mathbf{C}_k^t \leftarrow \operatorname{diag}(phase(\mathcal{A}_k(\mathbf{U}^{t-1}\mathbf{b}_k^{t-1})))$, $k = [q]$
7: b) $\mathbf{U}^{tmp} \leftarrow \operatorname{argmin}_{\tilde{\mathbf{U}}} \sum_k \left\| \mathbf{C}_k^t \mathbf{y}_k - \mathcal{A}_k(\tilde{\mathbf{U}}\mathbf{b}_k^{t-1}) \right\|^2$
8: c) $\mathbf{U}^t \leftarrow QR(\mathbf{U}^{tmp})$
9: d) $\mathbf{b}_k^t \leftarrow \operatorname{argmin}_{\tilde{\mathbf{b}}_k} \left\| \mathbf{C}_k^t \mathbf{y}_k - \mathcal{A}_k(\mathbf{U}^t \tilde{\mathbf{b}}_k) \right\|^2$, $k = [q]$
10: **end for**
11: **Intermediate output:** $\tilde{\mathbf{X}}^0 = \mathbf{U}^T\mathbf{B}^T$
(Modeling-error correction stage)
12: **for** $k = [q]$ **do**
13: $\tilde{\mathbf{x}_k}^0 = \mathbf{U}^T\mathbf{b}_k^T$
14: $\hat{\mathbf{x}_k^0} = \tilde{\mathbf{x}_k}^0 + \mathbf{e}_k^0$
15: **for** $t = 1, 2, \ldots, T'$ **do**
16: e) $\mathbf{C}_k^t \leftarrow \operatorname{diag}(phase(\mathcal{A}_k(\hat{\mathbf{x}_k}^t)))$
17: f) $\mathbf{e}_k^t \leftarrow \operatorname{argmin}_{\mathbf{e}}(\left\| \mathbf{C}_k^t \mathbf{y}_k - \mathcal{A}_k(\hat{\mathbf{x}_k}^t + \mathbf{e}) \right\|_2^2 + \tau \|\mathbf{e}\|_2^2)$
18: g) $\hat{\mathbf{x}_k}^{t+1} = \hat{\mathbf{x}_k}^t + \mathbf{e}_k^t$
19: **end for**
20: **end for**
21: **Output:** $\mathbf{X}^* = \hat{\mathbf{X}}^{T'+1}$

This initialization procedure critically ensures that a low rank structure is imposed in subsequent estimates of $\mathbf{X}$.

B. Low-rank matrix recovery

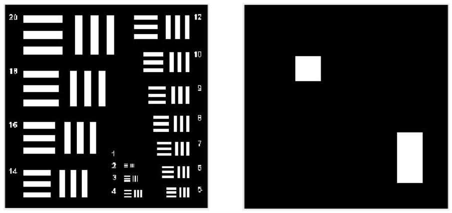
Once we obtain an initial estimate, we then refine it using a procedure similar to the LRPR2 algorithm of [22], which is an alternating-minimization algorithm that alternates between three steps: estimating the phase of the measurements $\mathbf{C}$, and the components $\mathbf{U}$ and $\mathbf{B}$ of the low rank matrix $\mathbf{X}$.

Specifically break down the Algorithm 3, in Line 10, we obtain an estimation of the missing phase information $\mathbf{C}_k^t$, for each frame k. In Line 11, we estimate an r-dimensional subspace $\mathbf{U}^t$, by utilizing the conjugate gradient (CG) method to obtain a fast, approximate solution, and thus avoid any need for explicit matrix inversions. In Line 12, we similarly estimate the coefficients $\mathbf{b}_k^t$ by using QR decomposition to obtain $\mathbf{b}_k^t$ in an efficient nammer.

C. Modeling-error correction

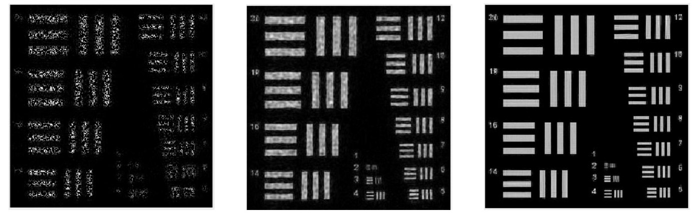
Finally, we proceed to the modeling error correction stage (lines 16-21 of Algorithm3), an idea similar to that used in iterative back projection (IBP)[53]. The output at the end of the low-rank matrix estimation stage, in Line 15, is exactly rank r. However, for most real videos, the low-rank model assumption, is often inconsistent, and cannot describe the video characteristics precisely.

We introduce new notation, to demarcate the real video as $\mathbf{X}^* = \tilde{\mathbf{X}} + \mathbf{E}$. In the modeling error correction stage,



(a) Spatially sparse (b) Block sparse

Fig. 3: *(a) Resolution chart, used as ground truth (b) simulated block sparse image, used as ground truth for experimental analysis.*



(a) Center SSIM=0.3517 (b) IERA SSIM=0.3369 (c)CoPRAM SSIM=0.8740

Fig. 4: *30% samples, (a) center image, reconstruction using (b) IERA (c) CoPRAM (spatial sparsity).*

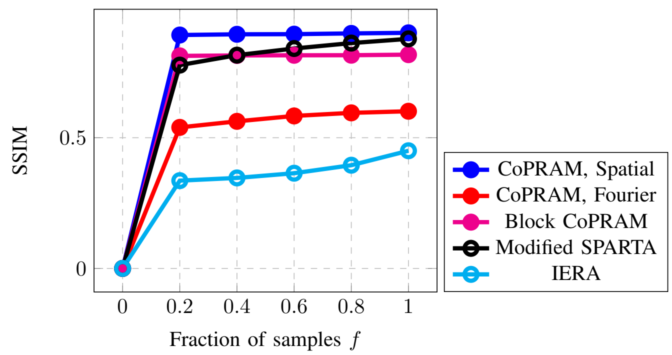


Fig. 5: *Variation of SSIM with sub-sampling ratio, with sparsity $s = 0.25n$, (block size $b = 4 \times 4$ for Block CoPRAM).*

we claim to produce $\hat{\mathbf{X}}^{t'} \to \mathbf{X}^*$. This stage, much like the previous stage involves alternatively estimating the modeling error $\mathbf{E} = [\mathbf{e}_1, \dots \mathbf{e}_q]$, and the missing phase information from the measurements.

We initialize this stage as $\hat{\mathbf{X}}^0 = \tilde{\mathbf{X}}^0 + \mathbf{E}^0$ where $\tilde{\mathbf{X}}^0$ is the output from the previous stage, and $\mathbf{E}^0 = \mathbf{0}$ initializes the modeling error on real videos. In lines 20 to 22, we use an alternative minimization method to estimate this model error, by alternatively updating $\mathbf{C}$ (step (e) of Algorithm 3) and $\mathbf{E}$ (step (f), and subsequently step (g) of Algorithm 3, $\hat{\mathbf{X}}$). We impose an ℓ_2 regularization on $\mathbf{e}_k$ to ensure that the error term is minimized.

In the next section we describe some experimental results based on our Model-based CoPRAM and MLRPtych algorithms.

VII. Experimental Results: Sparse Model

A. Simulation results

In this section, we demonstrate the performance of the sparse ptychography algorithms discussed in the previous sections on synthetically generated ptychographic measurements, with known ground truth values.

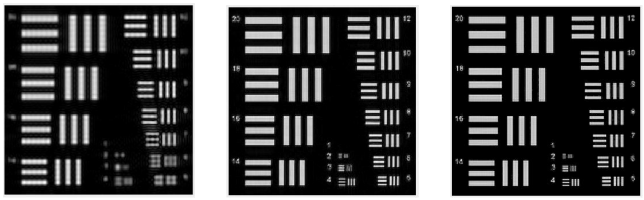
We describe the effect of enforcing the sparsity constraint in various domains as follows. We use two different datasets: (i) a simulated USAF resolution chart as shown in Figure 3 (a), and (ii) a simulated image which is specifically block sparse as shown in Figure 3 (b). The resolution chart provides a good way to inspect the recovery of finer details, at varying spatial resolutions. The parameters fed to the main algorithm are as follows: we used a $n = 256^2(256 \times 256)$ image of the Resolution Chart (resChart) as the ground truth. The camera array consists of $N = 81(9 \times 9)$ cameras, each with aperture diameter 72.75 pixels and overlap of 0.72 between consecutive cameras. A sub-sampling factor of $f = 0.3$ picks up 30% of the original number of measurements. To implement this, we generated masks $\mathcal{M}_i$ as in (1). For the sparse phase retrieval algorithm CoPRAM, we enforce a sparsity of $s = 0.25n$. The reconstruction procedure relies heavily on the extent of overlap, hence the norm of the reconstructed images is not preserved. We use Structural Similarity Index (SSIM) [54] as a metric to appropriately capture the quality of reconstruction, as it compares the two images in terms of luminance, contrast and structure, instead of utilizing a straightforward distance measure.

We employ CoPRAM by enforcing sparsity in spatial basis and compare the reconstruction from sub-sampled magnitude-only measurements, to those from Iterative Error Reduction Algorithm (IERA) [5], which is an application of AltMinPhase in the context of Fourier ptychography.

Sub-sampling via random pixel patterns: The results via the random pixel sub-sampling discussed in Section IV-B are displayed in Figure 4 for the input image in Figure 3. It can be noted that we can also impose sparsity in a wavelet basis (such as Haar) and we expect to achieve similar improvements in the SSIM.

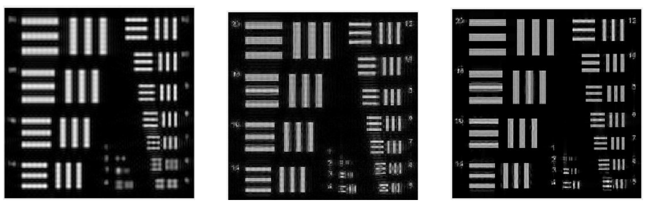
We have also analyzed the variation of the SSIM with different sub-sampling rates. For this, we used CoPRAM while assuming sparsity in the spatial basis for the input image in Fig. 3. We also invoked Block CoPRAM, (refer Sec. VII-A for details) which assumes block sparsity in the spatial domain. For comparison, we used IERA and also a modified version of another sparse phase retrieval algorithm called SPARTA [45], where we have used the same initialization as in line 1 of Algorithm 2. These results can be found in Figure 5.

Sub-sampling via randomly chosen cameras: The results via the randomly chosen cameras sub-sampling strategy discussed in Section IV-B are discussed here. We utilize this strategy to test the robustness of CoPRAM against IERA, under the sparsity assumption. We switch off $\approx 50\%$ of the cameras (for this experiment, 38 cameras are active, from 81 total), where the camera locations are picked according to (2) (the central camera is kept “on” by default). The results are displayed in Figure 6 for the input image in Figure 3. We



(a) Center SSIM=0.3927 (b) IERA SSIM=0.4225 (c) CoPRAM SSIM=0.9053

Fig. 6: *Using 50% cameras, (a) center image, reconstruction using (b) IERA (c) CoPRAM (spatial sparsity).*



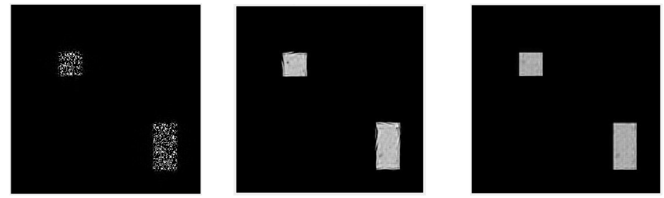
(a) Center SSIM=0.3674 (b) IERA SSIM=0.3088 (c) CoPRAM SSIM=0.6124

Fig. 7: *Using 0.12 overlap between consecutive cameras, (a) center image, reconstructed image using (b) IERA (c) reconstructed image using CoPRAM with sparsity constraint in spatial basis.*

observed that enforcing sparsity in the spatial domain gives a better reconstruction (Fig. 6 (d)).

Effect of decreased aperture overlap: One of the issues of the implementation in [5] is that they require consecutive camera arrays to have overlap with each other. This is physically impractical if one wants to implement a camera array in the same plane. However, with no camera overlap, their experiments perform poorly (oversampling is imperative for standard phase retrieval strategies). On the other hand CoPRAM uses a sparsity constraint to improve quality of reconstruction (Note: for this setup $f = 1$). For this experiment, we changed the amount of overlap between two cameras from 0.72 to 0.12. The results of this experiment suggest a superior reconstruction when CoPRAM is invoked, with sparsity in spatial basis (SSIM=0.6124) as compared to IERA (SSIM=0.3088) and the input center image (SSIM=0.3674) are displayed in Figure 7 for the input image in Figure 3. We observed that enforcing sparsity in the spatial domain gives a better reconstruction.

Extension to block sparsity: Since we were able to demonstrate the advantage of sparse modeling to reduce number of samples required for good reconstruction, we also applied CoPRAM to images with block sparsity (in the spatial domain). Instead of using CoSaMP (line 4 of Algorithm 2), we use a block variant of model-based CoSaMP [17] (we call this Block CoPRAM). For this experiment, we synthetically generated a block sparse image (Fig. 3 (b)), and measured it using the random sub-sampling pattern described in (1), with an low overlap of 0.12 between adjacent cameras. The reconstructions are displayed in Fig. 8, showing pronounced improvement when Block CoPRAM is used.



(a) Center SSIM=0.99687 (b) CoPRAM SSIM=0.99995 (c) Block CoPRAM SSIM=0.99998

Fig. 8: *Using 0.12 overlap and 30% samples (a) center image, reconstructed image using (b) CoPRAM (spatial sparsity) (c) Block CoPRAM (spatial block sparsity).*

TABLE I: Comparison of SSIM values for recovery from different initializations.

Initialization	Spectral	Center	Mean	RMS
SSIM	0.2328	0.8812	0.8908	0.8958

(a) Center (b) IERA (c) CoPRAM

Fig. 9: *Center image (a) and reconstruction using 30% pixels, via (b) IERA (c) CoPRAM, for a USAF imprint imaged via ptychography setup.*

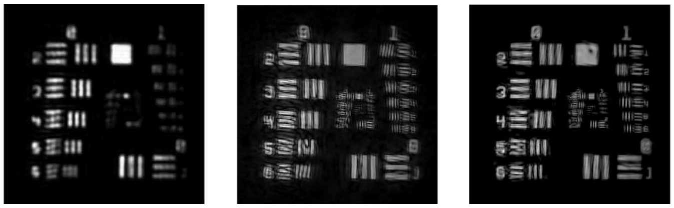
Effect of different initialization schemes: Several initialization schemes were compared. Specifically, we tried (i) spectral initialization, (ii) central camera image (iii) mean of absolute measurements, (iv) root-mean-squared (RMS) absolute measurements. The results from all of these initialization schemes in terms of SSIM, for the setting of 30% samples, using uniform random pixel sub-sampling, with CoPRAM, is tabulated in Table I. It is clear that the root-mean-squared measurements are a better initialization.

B. Real data experiments

For the sparse model, we used a USAF imprint imaged via the ptychographic setup, which is described in detail in Section VII. B. of [5]. The input image is 200×200 pixels, the camera array consists of $N = 529 (23 \times 23)$ cameras, each camera lens with aperture diameter spanning 56 pixels and spacing of 15.8 pixels (rounded to closest integer value) between consecutive pupils. The sparsity is assumed to be $s = 0.25n$. The reconstruction using uniform random pixel sub-sampling, by retaining 30% of the measurements and assuming sparsity in spatial basis is displayed in Figure 9.

Similarly, the results from uniform random camera sub-sampling by using 50% cameras is shown in Figure 10.

Perceptually, we results from CoPRAM are show better resolution and are in keeping with our findings from our simulation data experiments. In conclusion, the results of our



(a) Center (b) IERA (c) CoPRAM

Fig. 10: *Center image (a) and reconstruction using 50% cameras, via (b) IERA (c) CoPRAM for a USAF imprint imaged via ptychography setup.*

algorithm are well-applicable in real-world sparse imaging scenarios.

VIII. Experimental Results : Low-rank Model

A. Simulation results

In this section, we demonstrate the performance of the low-rank ptychography algorithms discussed in the previous sections on synthetically generated ptychographic measurements, with known ground truth values. We apply Algorithm 3 for two different patterns of under-sampling. The settings used for this experiment are as follows: the data is sized as $180 \times 180 \times q$, where q varies for different videos: $q = 112$ for "Bacteria" (B) video, $q = 148$ for "SleepingDog" (D) video, $q = 140$ for "Fish" (F) videos (all videos used for this implementation can be found at [55]). The aperture diameter of each camera considered is 40 pixels, overlap between consecutive cameras is of factor 0.48 and number of cameras in the camera array is 81 (9×9). We run lines 9-14 of MLR-Ptych algorithm for 5 iterations ($T = 5$) and lines 19-23 for 10 iterations ($T' = 10$). We compare the results of our algoirhtm to the basic AltMinPhase or IERA framework, for 250 outer iterations. In addition, we run original LR-Ptych algorithm, *without* modeling correction (lines 9-14 of Algorithm 3) for 5 iterations, as a comparison. The rank considered for all videos for is $r = 20$.

Sub-sampling via random pixel patterns: In the first set of experiments (refer Fig. 11,12), we consider random pixel under-sampling, as discussed in IV-B, with sub-sampling ratio f. In Fig. 12, we provide a visual comparison between the three algorithms (MLRPtych, LRPtych and IERA) that we tested in the experiment, for a fixed frame of the video of a fish (labeled as "F"). In Fig. 11 we compare the SSIM values from the reconstruction.

Sub-sampling via randomly chosen cameras: In the second set of experiments (refer Fig. 13,14), we consider a simpler and more feasible under-sampling strategy of turning a fraction of cameras from the camera array "on", as discussed in Section IV-B. We see similar trends of improved performance of MLRPtych w.r.t. IERA and LRPtych (see Fig. 14, in terms of SSIM, in both sets of experiments. It is also interesting to note that even under the scenario where we consider all measurements ($f = 1$), we see an improved recovery for the MLRPtych algorithm w.r.t. IERA. A visual comparison of the performance of both algorithms on "Bacteria" (B) video can be seen in Figure 14.

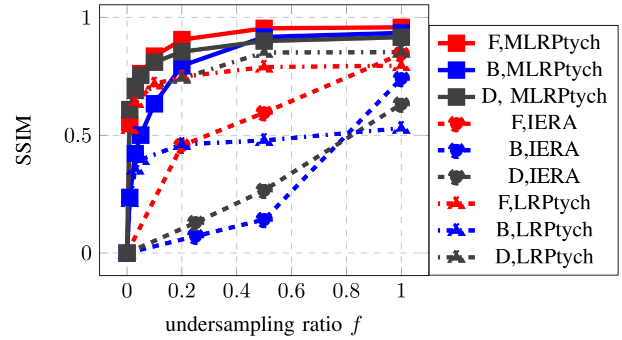
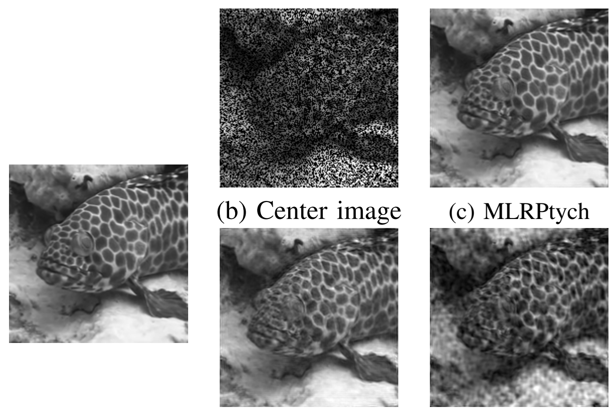


Fig. 11: Variation of SSIM of recovery of different algorithms, with random pixel sub-sampling, at different sub-sampling ratios f.



(a) Ground truth (b) Center image (c) MLRPtych (d) LRPtych (e) IERA

Fig. 12: *Visual comparison of super-resolved reconstructions via (c) MLRPtych, (d) LRPtych, (e) IERA for ptychography using* 50% *of measured pixels from low-resolution input (b), with known ground truth (a).*

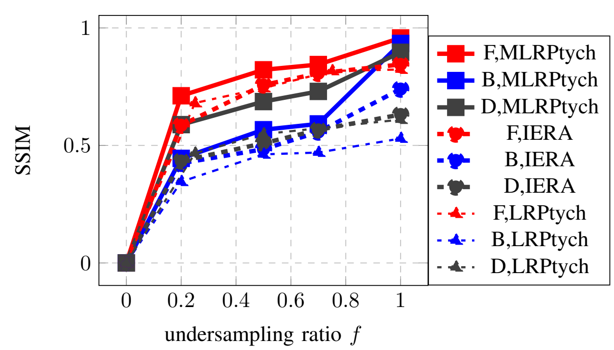
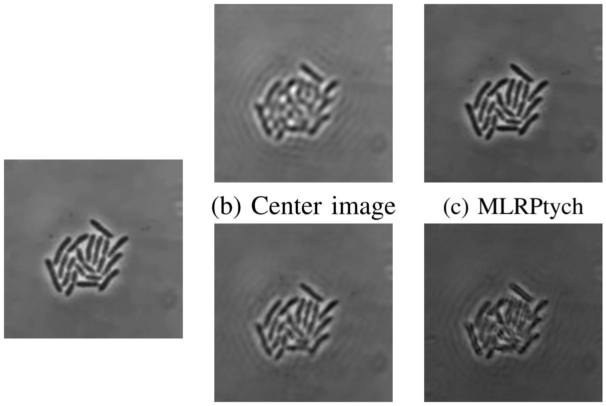


Fig. 13: *Variation of SSIM of recovery of different algorithms, with random camera sub-sampling, at different sub-sampling ratios f.*

The reconstruction metric, as well as perceptual quality suggests that MLRPtych (and LRPtych) give improved reconstruction with respect to conventional algorithms which do not consider a low-rank structure, using fewer measurements. We now demonstrate similar gains for experimentally obtained ptychographic measurements of biological cells.



(a) Ground truth (b) Center image (c) MLRPtych (d) LRPtych (e) IERA

Fig. 14: *Visual comparison of super-resolved reconstructions via (c) MLRPtych, (d) LRPtych, (e) IERA for ptychography using* 50% *of cameras from low-resolution input (b), with known ground truth (a).*

B. Real data experiments

For the low-rank model, we source the data captured by a multiplexed-LED illumination microscopic system implemented by Tian et. al. [33].

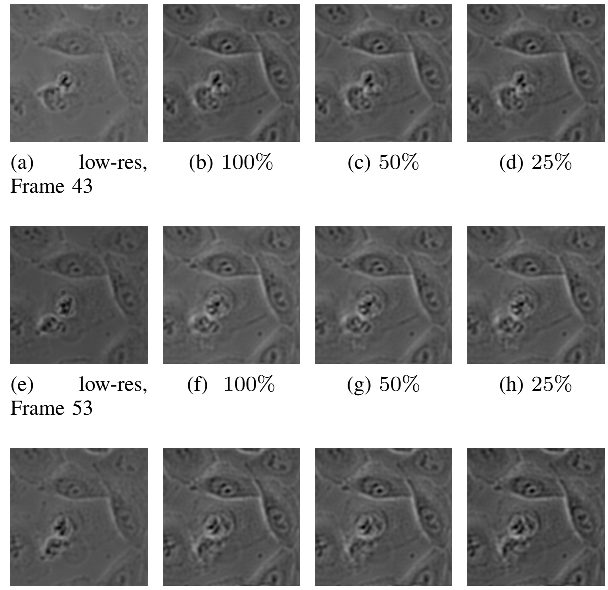
The setting used in such system is as follows. The total number of LEDs is 293 ($N = 293$) with overlap of 92.1%. Size of measurement from each LED is 100×100. Length of video $q = 98$. The size of recovered frames is 500×500. The rank considered for LRPtych is $r = 20$.

A low-rank regularization is useful in reducing the effect of noisy or erroneous, as well as sub-sampled measurements. With the simulation results, we have demonstrated the improved recovery of (approximately) low-rank videos, using much fewer samples. In this section we show similar gains on biological data acquired via a ptychography setup.

Sub-sampling via random pixel patterns: In the first set of experiments we utilize the random pixel sub-sampling strategy discussed in Section IV-B. The results of the reconstruction under various sub-sampling ratios f, for LRPtych, are shown in Figure 15.

Sub-sampling via randomly chosen cameras: In the second set of experiments, we utilize the random camera pattern discussed in Section IV-B to sub-sample measurements. In Figure 16, we show the results of reconstruction under the uniform random camera sub-sampling strategy.

In Table II, we compare the SSIM of reconstruction under different algorithms (implementation by Tian et. al. [33] which we call AltGrad, and LRPtych), and sub-sampling schemes, while using the $f = 1$, or "full" measurement case as the baseline. We note that LRPtych is capable of achieving superior performance as compared to AltGrad, under this metric. Further discussion on these experiments can be found in [55].

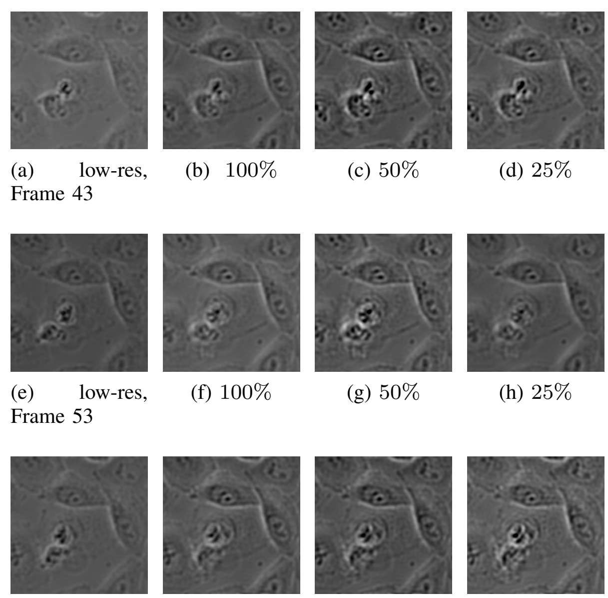


(a) low-res, Frame 43 (b) 100% (c) 50% (d) 25%

(e) low-res, Frame 53 (f) 100% (g) 50% (h) 25%

(i) low-res, Frame 63 (j) 100% (k) 50% (l) 25%

Fig. 15: *(a),(e),(i) show the low-resolution input images for Frames 43,53 and 63 respectively, and the results for pixel-wise sub-sampling are shown in (b)-(d) for frame 43, (f)-(h) for frame 53 and (j)-(l) for frame 63, using* $100\%, 50\%, 25\%$ *measurements.*



(a) low-res, Frame 43 (b) 100% (c) 50% (d) 25%

(e) low-res, Frame 53 (f) 100% (g) 50% (h) 25%

(i) low-res, Frame 63 (j) 100% (k) 50% (l) 25%

Fig. 16: *(a),(e),(i) show the low-resolution input images for Frames 43,53 and 63 respectively, and the results for camera-wise sub-sampling are shown in (b)-(d) for frame 43, (f)-(h) for frame 53 and (j)-(l) for frame 63, using* $100\%, 50\%, 25\%$ *measurements.*

IX. Low-rank v/s Block Sparse Phase Retrieval

For the sake of completeness, we compare the performance of Block Sparse variant of CoPRAM with the Low Rank

TABLE II: Comparison of reconstruction SSIM with that of full measurements.

sub-sample ratio	$f=1$	pixel $f=0.5$	pixel $f=0.25$	camera $f=0.5$	camera $f=0.25$
AltGrad	N/A	0.5711	0.4748	0.5951	0.5603
LRPtych	N/A	0.9979	0.9930	0.9218	0.8219

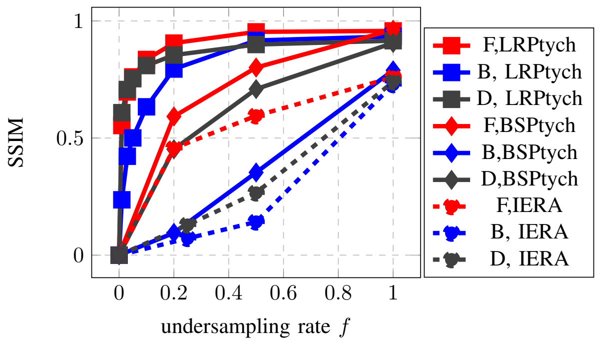


Fig. 17: Variation of SSIM of reconstructed image obtained using LRPtych, BSPtych (apply block sparsity on video signal), and IERA versus sampling rates for three videos "Fish" (F), "Dog" (D), "Bacteria" (B).

Ptychography algorithm. Note that a low-rank video can be considered to be *approximately* block sparse, though it may not be the best model for such kind of setups. To demonstrate this, we compare the performances of model-based CoPRAM with a block sparsity assumption, which assumes block sparsity in wavelet domain of a video signal (instead of low rank) and use same dynamic ptychography measurement set-up used for the LRPtych formulation by showing the SSIM verses pixel-wise under-sampling rate f in Fig. 17, for three videos of a fish (F), dog (D) and bacteria cell (B) respectively (Section VIII-A). We call this implementation BSPtych, and highlight that this implementation is different from that in Section V which considers a different measurements setup. As the videos used here are not typical for those under which the wavelet block sparsity model would hold , we can see that the performance of block sparsity based algorithm is not as good as low rank based one, but it is still better than IERA which uses no structure. Moreover, the measurement setup itself, is not identical to that used in Algorithm 2 for the reconstruction procedure. The block-sparse formulation considers the entire video volume to be a single image frame, where the block sparsity is modeled across the time (or frame) axis. The measurement setup in this scenario considers the video volume to be a single image, with each frame being a single column, which differs from the setup we use for the sparse formulation of the problem, in which the image frame is not vectorized. Because these two formulations are inconsistent, we argue that we require two different models for low-rank and block sparse formulations.

References

[1] Z. Chen, G. Jagatap, S. Nayer, C. Hegde, and N. Vaswani, "Low rank fourier ptychography," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 6538–6542.

[2] G. Jagatap, Z. Chen, C. Hegde, and N. Vaswani, "Sub-diffraction imaging using fourier ptychography and structured sparsity," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 6493–6497.

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

[4] G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution fourier ptychographic microscopy," *Nature photonics*, vol. 7, no. 9, pp. 739–745, 2013.

[5] J. Holloway, M. S. Asif, M. K. Sharma, N. Matsuda, R. Horstmeyer, O. Cossairt, and A. Veeraraghavan, "Toward long-distance subdiffraction imaging using coherent camera arrays," *IEEE Transactions on Computational Imaging*, vol. 2, no. 3, pp. 251–265, 2016.

[6] S. Dong, R. Horstmeyer, R. Shiradkar, K. Guo, X. Ou, Z. Bian, H. Xin, and G. Zheng, "Aperture-scanning fourier ptychography for 3d refocusing and super-resolution macroscopic imaging," *Optics express*, vol. 22, no. 11, pp. 13 586–13 599, 2014.

[7] J. Fienup, "Phase retrieval algorithms: a comparison," *Applied optics*, vol. 21, no. 15, pp. 2758–2769, 1982.

[8] S. Marchesini, "Phase retrieval and saddle-point optimization," *JOSA A*, vol. 24, no. 10, pp. 3289–3296, 2007.

[9] K. Nugent, A. Peele, H. Chapman, and A. Mancuso, "Unique phase recovery for nonperiodic objects," *Physical review letters*, vol. 91, no. 20, p. 203902, 2003.

[10] E. Candes, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, 2013.

[11] E. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM review*, vol. 57, no. 2, pp. 225–251, 2015.

[12] E. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval via wirtinger flow: Theory and algorithms," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1985–2007, 2015.

[13] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," in *Advances in Neural Information Processing Systems*, 2013, pp. 2796–2804.

[14] G. Wang, G. Giannakis, and Y. C. Eldar, "Solving systems of random quadratic equations via truncated amplitude flow," *IEEE Transactions on Information Theory*, 2017.

[15] D. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[16] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Royal Stat. Soc. Stat. Meth.*, vol. 68, no. 1, pp. 49–67, 2006.

[17] M. D. R. Baraniuk, V. Cevher and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.

[18] C. Hegde, P. Indyk, and L. Schmidt, "Fast algorithms for structured sparsity," *Bul. of the EATCS*, vol. 1, no. 117, pp. 197–228, Oct. 2015.

[19] M. A. Davenport and J. Romberg, "An overview of low-rank matrix recovery from incomplete observations," *arXiv preprint arXiv:1601.06422*, 2016.

[20] G. Jagatap and C. Hegde, "Fast, sample-efficient algorithms for structured phase retrieval," *Adv. Neural Inf. Proc. Sys. (NIPS)*, pp. 4924–4934, 2017.

[21] ——, "Towards sample-optimal methods for solving random quadratic equations with structure," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 2296–2300.

[22] N. Vaswani, S. Nayer, and Y. C. Eldar, "Low-rank phase retrieval," *IEEE Transactions on Signal Processing*, vol. 65, no. 15, pp. 4059–4074, 2016.

[23] E. Candes, X. Li, and M. Soltanolkotabi, "Phase retrieval from coded diffraction patterns," *Applied and Computational Harmonic Analysis*, vol. 39, no. 2, pp. 277–299, 2015.

[24] A. Maiden, M. Humphry, F. Zhang, and J. Rodenburg, "Superresolution imaging via ptychography," *J. Opt. Soc. Am. A*, vol. 28, no. 4, pp. 604–612, Apr 2011.

[25] R. Horstmeyer, R. Chen, X. Ou, B. Ames, J. Tropp, and C. Yang, "Solving ptychography with a convex relaxation," *New Journal of Physics*, vol. 17, no. 5, p. 053044, 2015.

[26] K. Jaganathan, Y. Eldar, and B. Hassibi, "Stft phase retrieval: Uniqueness guarantees and recovery algorithms," *IEEE J. Select. Top. Sig. Proc.*, vol. 10, no. 4, pp. 770–781, 2016.

[27] L. Li, C. Cheng, D. Han, Q. Sun, and G. Shi, "Phase retrieval from multiple-window short-time fourier measurements," *IEEE Sig. Proc. Lett.*, vol. 24, no. 4, pp. 372–376, 2017.

[28] T. Bendory, Y. Eldar, and N. Boumal, "Non-convex phase retrieval from stft measurements," *IEEE Trans. Info. Theory*, 2017.

[29] L.-H. Yeh, J. Dong, J. Zhong, L. Tian, M. Chen, G. Tang, M. Soltanolkotabi, and L. Waller, "Experimental robustness of fourier ptychography phase retrieval algorithms," *Optics express*, vol. 23, no. 26, pp. 33 214–33 240, 2015.

[30] L. Bian, J. Suo, G. Zheng, K. Guo, F. Chen, and Q. Dai, "Fourier ptychographic reconstruction using wirtinger flow optimization," *Optics express*, vol. 23, no. 4, pp. 4856–4866, 2015.

[31] L. Bian, J. Suo, J. Chung, X. Ou, C. Yang, F. Chen, and Q. Dai, "Fourier ptychographic reconstruction using poisson maximum likelihood and truncated wirtinger gradient," *Scientific reports*, vol. 6, p. 27384, 2016.

[32] X. Ou, G. Zheng, and C. Yang, "Embedded pupil function recovery for fourier ptychographic microscopy," *Optics express*, vol. 22, no. 5, pp. 4960–4972, 2014.

[33] L. Tian, X. Li, K. Ramchandran, and L. Waller, "Multiplexed coded illumination for fourier ptychography with an led array microscope," *Biomedical optics express*, vol. 5, no. 7, pp. 2376–2389, 2014.

[34] Y. Zhang, P. Song, J. Zhang, and Q. Dai, "Fourier ptychographic microscopy with sparse representation," *Scientific Reports*, vol. 7, no. 1, p. 8664, 2017.

[35] Y. Zhang, Z. Cui, J. Zhang, P. Song, and Q. Dai, "Group-based sparse representation for fourier ptychography microscopy," *Optics Communications*, vol. 404, pp. 55–61, 2017.

[36] F. Shamshad, F. Abbas, and A. Ahmed, "Deep ptych: Subsampled fourier ptychography using generative priors," *arXiv preprint arXiv:1812.11065*, 2018.

[37] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane picture," *Optik*, 1972.

[38] Y. Chen, Y. Chi, and A. Goldsmith, "Exact and stable covariance estimation from quadratic sampling via convex programming," *IEEE Trans. Inform. Theory*, vol. 61, no. 7, pp. 4034–4059, 2015.

[39] J. Sun, Q. Qu, and J. Wright, "A geometric analysis of phase retrieval," *Foundations of Computational Mathematics*, vol. 18, no. 5, pp. 1131–1198, 2018.

[40] Y. Chen and E. Candes, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," in *Advances in Neural Information Processing Systems*, 2015, pp. 739–747.

[41] H. Ohlsson, A. Yang, R. Dong, and S. Sastry, "Cprl–an extension of compressive sensing to the phase retrieval problem," in *Adv. Neural Inf. Proc. Sys. (NIPS)*, 2012, pp. 1367–1375.

[42] H. Zhang and Y. Liang, "Reshaped wirtinger flow for solving quadratic system of equations," in *Advances in Neural Information Processing Systems*, 2016, pp. 2622–2630.

[43] K. Jaganathan, S. Oymak, and B. Hassibi, "Sparse phase retrieval: Convex algorithms and limitations," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*. IEEE, 2013, pp. 1022–1026.

[44] T. Cai, X. Li, and Z. Ma, "Optimal rates of convergence for noisy sparse phase retrieval via thresholded wirtinger flow," *Ann. Stat.*, vol. 44, no. 5, pp. 2221–2251, 2016.

[45] G. Wang, G. Giannakis, J. Chen, and M. Akçakaya, "Sparta: Sparse phase retrieval via truncated amplitude flow," in *Proc. IEEE Int. Conf. Acoust., Speech, and Sig. Proc. (ICASSP)*. IEEE, 2017, pp. 3974–3978.

[46] M. Iwen, A. Viswanathan, and Y. Wang, "Robust sparse phase retrieval made easy," *Ap. Comp. Har. An.*, vol. 42, no. 1, pp. 135–142, 2017.

[47] S. Bahmani and J. Romberg, "Efficient compressive phase retrieval with constrained sensing vectors," in *Adv. Neural Inf. Proc. Sys. (NIPS)*, 2015, pp. 523–531.

[48] P. Jain, P. Netrapalli, and S. Sanghavi, "Low-rank matrix completion using alternating minimization," in *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*. ACM, 2013, pp. 665–674.

[49] E. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, p. 717, 2009.

[50] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Advances in neural information processing systems*, 2009, pp. 2080–2088.

[51] C. Quinsac, A. Basarab, and D. Kouamé, "Frequency domain compressive sampling for ultrasound imaging," *Advances in Acoustics and Vibration*, vol. 2012, 2012.

[52] D. Needell and J. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and computational harmonic analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[53] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical models and image processing*, vol. 53, no. 3, pp. 231–239, 1991.

[54] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[55] Z. Chen, "Low Rank Ptychography," http://www.public.iastate.edu/ zhengyuc/lowrankrst.html, 2018.