

# Accelerated Wirtinger Flow for Multiplexed Fourier Ptychographic Microscopy

Emrah Bostan\*, Mahdi Soltanolkotabi<sup>†</sup>, David Ren\*, and Laura Waller\*

## Abstract

Fourier ptychographic microscopy enables gigapixel-scale imaging, with both large field-of-view and high resolution. Using a set of low-resolution images that are recorded under varying illumination angles, the goal is to computationally reconstruct high-resolution phase and amplitude images. To increase temporal resolution, one may use multiplexed measurements where the sample is illuminated simultaneously from a subset of the angles. In this paper, we develop an algorithm for Fourier ptychographic microscopy with such multiplexed illumination. Specifically, we consider gradient descent type updates and propose an analytical step size that ensures the convergence of the iterates to a stationary point. Furthermore, we propose an accelerated version of our algorithm (with the same step size) which significantly improves the convergence speed. We demonstrate that the practical performance of our algorithm is identical to the case where the step size is manually tuned. Finally, we apply our parameter-free approach to real data and validate its applicability.

## 1 Introduction

Fourier Ptychographic Microscopy (FPM) is a computational imaging technique that—guided by synthetic aperture principles—generates gigapixel images with both wide field-of-view and high resolution [?]. The system is realized by replacing a microscope’s illumination unit with a light-emitting diode (LED) array. As LEDs illuminate the sample from different angles, the camera captures multiple *intensity* images of different spatial frequency bands of the sample, without moving parts. Based on a nonlinear inverse problem, which is a type of phase retrieval [?], the low-resolution measurements are used to computationally generate a high-resolution image of the sample in both amplitude and phase.

---

\*Department of Electrical Engineering and Computer Sciences, University of California, Berkeley

<sup>†</sup> Ming Hsieh Department of Electrical Engineering, University of Southern California  
E. Bostan’s research is supported by the Swiss National Science Foundation (SNSF) under grant P2ELP2 172278. M. Soltanolkotabi’s research is supported by the Air Force Office of Scientific Research under award number FA9550-18-1-0078.

FPM has been established as a viable tool for bioimaging applications [?, ?, ?, ?], including *live cell imaging* [?] for studies such as stem cell development and drug discovery [?]. Sequential data collection (*i.e.* one image collected per LED) has limited temporal resolution, preventing dynamic imaging. This has been addressed by *multiplexed coded illumination*, where a random subset of LEDs are turned on at the same time to reduce the total number of images that need to be taken [?]. However, the multiplexed information must then be decoupled, which can make the reconstruction less robust to noise and model-mismatch [?]. This puts emphasis on the stability (such as convergence and sensitivity to hyper-parameters) of reconstruction algorithms for FPM, especially when multiplexed illumination is used.

Several algorithms have been proposed for solving the phase retrieval problem within the context of FPM [?]. Existing reconstructions have noticeably capitalized on gradient descent type methods (and their projected variants) for the multiplexed illumination case [?]. They provide a favorable trade-off between the reconstruction quality and compute time [?]. However, the very fundamental question of how one chooses the step size has not been rigorously investigated. Since the acquisition parameters (for instance, the illumination coding and number of measurements) can vary, the importance of a systematic approach to determine the step size is pronounced in practice. While there are known strategies for quadratic cost functions of the phase retrieval problem [?], these approaches do not apply to other cost functions including those that are effective for the FPM model at hand [?]. More importantly, these heuristics do not provide us with convergence guarantees except for idealized random models such as those in [?, ?, ?].

In this paper, we develop an auto-tuned algorithm for FPM with multiplexed coded illumination that is based on theoretical principles. To that aim, we propose a gradient descent algorithm called the accelerated Wirtinger Flow (AWF). Our main contributions are:

- The proposal of an analytical expression to select the step size, which makes the final algorithm free of tuning parameters. The framework is applicable to any type of LED selection for multiplexing.
- The stationary point convergence of the Wirtinger flow iterates with the chosen step size.
- The illustration of AWF’s convergence speed reaching its analogue with manually optimized step size. We also demonstrate the applicability of AWF to real data.

## 2 Forward Model

We start with a mathematical description of the measurement formation process in FPM with multiplexed coded illumination. Consider the setup in Figure 1, where an array of LEDs is used as the illumination source of a standard microscope. The coordinate vector is given by  $(\mathbf{r}, z)$  where  $\mathbf{r} = (r_1, r_2) \in \mathbb{R}^2$  denotes

the spatial location on a transverse plane perpendicular to the optical axis  $z$ . A *thin* sample (that is the 2D object to be imaged) is located at  $z = 0$ , which is also the rear focal plane of the objective lens.

The sample is characterized by a complex-valued *transmission function*

$$s(\mathbf{r}) := \sqrt{I_s(\mathbf{r})} \exp(j\phi_s(\mathbf{r})), \quad (1)$$

where  $j^2 = -1$  and the continuous mappings  $I_s : \Omega \rightarrow \mathbb{R}$  and  $\phi_s : \Omega \rightarrow \mathbb{R}$  represent the spatial *intensity* and *phase* maps (of the sample), respectively. The domain  $\Omega$  is assumed to be a compact subset of  $\mathbb{R}^2$ .

Placing the LED array sufficiently far away, each LED's illumination is modeled as a monochromatic plane wave at the sample plane  $z = 0$ . When the  $i$ th LED is switched on, the field exiting the sample (and entering the microscope objective) is given by  $s_i(\mathbf{r}) = s(\mathbf{r}) \exp(2\pi j \langle \boldsymbol{\xi}_i, \mathbf{r} \rangle)$ , where  $\boldsymbol{\xi}_i$  is the spatial frequency vector of the corresponding angle of illumination<sup>1</sup>. In effect, each  $s_i$  is uniquely described by shifting the Fourier transform<sup>2</sup> of the transmission function  $\hat{s}$  since it holds that  $\hat{s}_i(\cdot) = \hat{s}(\cdot - \boldsymbol{\xi}_i)$ .

Due to the finite-aperture objective lens, the exit field is low-pass filtered as it goes through the microscope. The process is specified by the *pupil* function  $\hat{p}$ , which is the Fourier transform of the coherent point spread function [?], and suppresses spatial frequencies beyond the diffraction limit (the pupil function is 1 inside the numerical aperture (NA) of the objective and 0 otherwise). Consequently, the camera captures the intensity of the lower-resolution field, expressed as:

$$I_i(\mathbf{r}) = \left| \mathcal{F}^{-1} \{ (\hat{p}(\cdot) \hat{s}(\cdot - \boldsymbol{\xi}_i)) \} \right|^2(\mathbf{r}). \quad (2)$$

We see in (2) that FPM can sample spatial frequency bands beyond the diffraction limit via angle-varying illumination, however phase information is lost. By sampling the spatial frequency bands with overlapping regions, FPM enables phase retrieval, but this typically requires excessive redundancy [?]. *Multiplexing* fixes this (*i.e.* less redundancy) without making the phase retrieval fail. In this approach, rather than turning on LEDs one at a time, a subset of LEDs are simultaneously lit [?]. Since LEDs are mutually incoherent with each other, the total intensity of the multiplexed measurement is expressed as the sum of the intensity if each LED was switched on individually:

$$I_{\mathcal{M}}(\mathbf{r}) = \sum_{i \in \mathcal{M}} I_i(\mathbf{r}), \quad (3)$$

where  $\mathcal{M}$  denotes the index set of the selected LEDs.

<sup>1</sup> $\boldsymbol{\xi}_i := (1/\lambda)(\cos(\alpha_i), \cos(\beta_i))$  where  $\lambda$  is the wavelength;  $\alpha_i$  and  $\beta_i$  are the angles of incidence on the axes  $r_1$  and  $r_2$ , respectively.

<sup>2</sup>The Fourier transform  $\hat{f}$  of a function  $f$  is defined by

$$\hat{f}(\boldsymbol{\xi}) = \mathcal{F}\{f\}(\boldsymbol{\xi}) := \iint f(\mathbf{r}) \exp(-j2\pi \langle \boldsymbol{\xi}, \mathbf{r} \rangle) \, d\mathbf{r}.$$

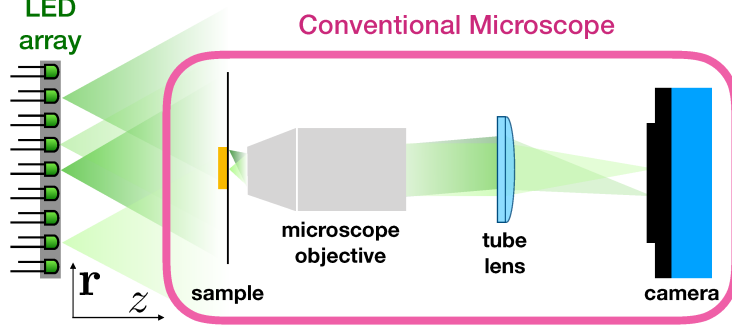


Figure 1: Optical setup and the geometry multiplexed FPM. Multiple LEDs simultaneously illuminate the sample from different angles. After passing through the microscope, sample’s image is generated at the camera plane where intensity-only measurements are recorded. The goal is to reconstruct both phase and amplitude with higher resolution than the objective’s diffraction limit.

Finally, we discretize (3) by considering its *amplitude-based* counterpart as such formulations are more robust to both noise and model mismatch [?]. To that end, let  $\mathbf{s} \in \mathbb{C}^n$  be a discretization of the Fourier transform of the transmission function and  $\mathbf{y}_{\mathcal{M}} \in \mathbb{R}_{\geq 0}^m$  the discrete samples of  $\sqrt{I_{\mathcal{M}}}$  at the camera plane. We note that  $m < n$  since the reconstructed transmission function has a higher space-bandwidth product than the measured image [?]. For each multiplexed measurement, the discretized forward model is hence given by

$$\mathbf{y}_{\mathcal{M}} = \sqrt{\sum_{i \in \mathcal{M}} |\mathbf{F}^H \mathbf{P} \mathbf{C}_i \mathbf{s}|^2}. \quad (4)$$

where  $\mathbf{C}_i \in \mathbb{C}^{m \times n}$  is the matrix representation of the  $m$ -pixel cropping centered at  $\xi_i$ ,  $\mathbf{P} \in \mathbb{C}^{m \times m}$  is a diagonal matrix generated from the discretized pupil function, and  $\mathbf{F}^H \in \mathbb{C}^{m \times m}$  represents the inverse of the discrete (normalized) Fourier transform [?]. Note that  $|\cdot|^2$  and  $\sqrt{\cdot}$  are element-wise operations.

### 3 Reconstruction Algorithm: Accelerated Wirtinger Flow

In this section we propose a reconstruction algorithm for recovering the sample’s amplitude and phase maps from a set of multiplexed intensity measurements. To this aim, we solve the following optimization problem:

$$\min_{\mathbf{s} \in \mathbb{C}^n} \left( \mathcal{J}(\mathbf{s}) := \sum_{k=1}^K \left\| \mathbf{y}_{\mathcal{M}_k} - \sqrt{\sum_{i \in \mathcal{M}_k} |\mathbf{F}^H \mathbf{P} \mathbf{C}_i \mathbf{s}|^2} \right\|_2^2 \right). \quad (5)$$

Here,  $K$  represents the total number of captured multiplexed images. An iterative approach to solve this problem is to consider gradient descent type updates of the form

$$\mathbf{s}_{t+1} \leftarrow \mathbf{s}_t - \mu_t \nabla \mathcal{J}(\mathbf{s}_t), \quad (6)$$

where  $\mu_t$  is the step-size at iteration  $t$ . Since the cost function  $\mathcal{J}$  is not complex-differentiable<sup>3</sup>, we shall rely on the notion of Wirtinger derivatives to define the gradient  $\nabla \mathcal{J}$  and, hence, refer to (6) as Wirtinger Flow (WF)<sup>4</sup> [?]. Still, the cost function is differentiable (even in the sense of Wirtinger derivatives) except for isolated points so that we use the notion of generalized gradients. This allows us to define the gradient at a non-differentiable point as one of the limit points of the gradient in a local neighborhood of the non-differentiable point [?]. For our cost function in (5), the generalized gradient takes the form

$$\nabla \mathcal{J}(\mathbf{s}) := \sum_{k=1}^K \sum_{i \in \mathcal{M}_k} \mathbf{C}_i^H \mathbf{P}^H \mathbf{F} \mathbf{e}, \quad (7)$$

where

$$\mathbf{e} = \left( \sqrt{\sum_{i \in \mathcal{M}_k} |\mathbf{A}_i \mathbf{s}|^2} - \mathbf{y}_{\mathcal{M}_k} \right) \odot \left( \frac{\mathbf{A}_i \mathbf{s}}{\sqrt{\sum_{i \in \mathcal{M}_k} |\mathbf{A}_i \mathbf{s}|^2}} \right)$$

with  $\mathbf{A}_i = \mathbf{F}^H \mathbf{P} \mathbf{C}_i$ . Here,  $\odot$  denotes the Hadamard (*i.e.* element-wise) product while  $\mathbf{P}^H$  and  $\mathbf{C}_i^H$  are the adjoint operators of  $\mathbf{P}$  and  $\mathbf{C}_i$ , respectively. Note that for a complex-valued vector  $\mathbf{a}$ , the element-wise division  $\mathbf{a}/|\mathbf{a}|$  results in a vector whose entries contain the phase of the entries of  $\mathbf{a}$ .

Based on bounded Hessian arguments<sup>5</sup>, we propose the following step size for the WF iterations in (6):

$$\mu_t = \mu := \frac{1}{\left\| \sum_{k=1}^K \sum_{i \in \mathcal{M}_k} \mathbf{C}_i^H \mathbf{P}^H \mathbf{P} \mathbf{C}_i \right\|_2}. \quad (8)$$

This proposed step size is general and can accommodate any illumination coding design represented by  $\mathcal{M}_k$ ,  $k = 1, \dots, K$ . Moreover, we establish that

$$\left\| \sum_{k=1}^K \sum_{i \in \mathcal{M}_k} \mathbf{C}_i^H \mathbf{P}^H \mathbf{P} \mathbf{C}_i \right\|_2 = \left\| \sum_{i \in \mathcal{A}} |\mathbf{P}_i|^2 \right\|_{\ell_\infty}, \quad (9)$$

where  $\mathcal{A} := \cup_{k=1}^K \mathcal{M}_k$  and  $\|\cdot\|_{\ell_\infty}$  is the maximum norm. Here,  $\mathbf{P}_i$  represents a shifted pupil function that is centered at  $\boldsymbol{\xi}_i$ . Therefore, under the assumption of an ideal pupil function, the step size is inversely proportional to the maximum redundancy factor of the sampling in the Fourier domain.

<sup>3</sup> $\mathcal{J}$  is a mapping from  $\mathbb{C}^n$  to  $\mathbb{R}_{\geq 0}$  so that it is not holomorphic.

<sup>4</sup>It is noteworthy that (6) is also closely related to the well-known Gerchberg-Saxton method [?].

<sup>5</sup>We omit the full derivation.

### 3.1 Theory for convergence to stationary points

We now provide a theoretical justification for our choice of step size in (8). To do so, we state the following theorem:

**Theorem 1.** *For  $k = 1, 2, \dots, K$ , let  $\mathbf{y}_{\mathcal{M}_k} \in \mathbb{R}_{\geq 0}^m$  denote the  $k$ th multiplexed FPM measurement according to (4). We run the updates*

$$\mathbf{s}_{t+1} \leftarrow \mathbf{s}_t - \mu_t \nabla \mathcal{J}(\mathbf{s}_t),$$

*with  $\nabla \mathcal{J}(\mathbf{s})$  as defined in (7) and the step size  $\mu_t$  obeying (8). Also, let  $\mathbf{s}_* \in \arg \min_{\mathbf{s} \in \mathbb{C}^n} \mathcal{J}(\mathbf{s})$  be a global optima. Then, it holds that*

$$\lim_{t \rightarrow \infty} \|\nabla \mathcal{J}(\mathbf{s}_t)\|_2 \rightarrow 0,$$

and

$$\min_{t \in \{1, 2, \dots, T\}} \|\nabla \mathcal{J}(\mathbf{s}_t)\|_2^2 \leq \frac{(\mathcal{J}(\mathbf{s}_1) - \mathcal{J}(\mathbf{s}_*))}{\mu T}.$$

Our theorem demonstrates that the WF iterates with our chosen step size converge to a point where the generalized gradient is zero. This is a non-trivial statement, as the cost function is non-smooth and there are many stationary points where the generalized gradient does not vanish. We note that the theorem does not imply convergence to a local optima.

### 3.2 Acceleration

We have shown that the WF iterates converge to a stationary point if the step size is set as in (8). However, the convergence rate is still rather slow. To overcome this challenge, inspired by the seminal work of Nesterov [?], we apply an acceleration method to our WF scheme. The accelerated WF (AWF) updates are

$$\mathbf{v}_{t+1} \leftarrow \mathbf{s}_t - \mu_t \nabla \mathcal{J}(\mathbf{s}_t); \tag{10a}$$

$$q_{t+1} \leftarrow 1/2 + (1/2)\sqrt{1 + 4q_t^2}; \tag{10b}$$

$$\mathbf{s}_{t+1} \leftarrow \mathbf{v}_{t+1} + (q_t - 1/q_{t+1})(\mathbf{v}_{t+1} - \mathbf{v}_t), \tag{10c}$$

where  $q_1 = 1$  and the step size  $\mu_t$  is kept the same. We note that Nesterov's acceleration scheme is derived for convex and smooth functions, which is not the case for the cost function in (5). Within our framework, it is used as a remedy for improving the convergence in practice, which we shall demonstrate in Section 4.

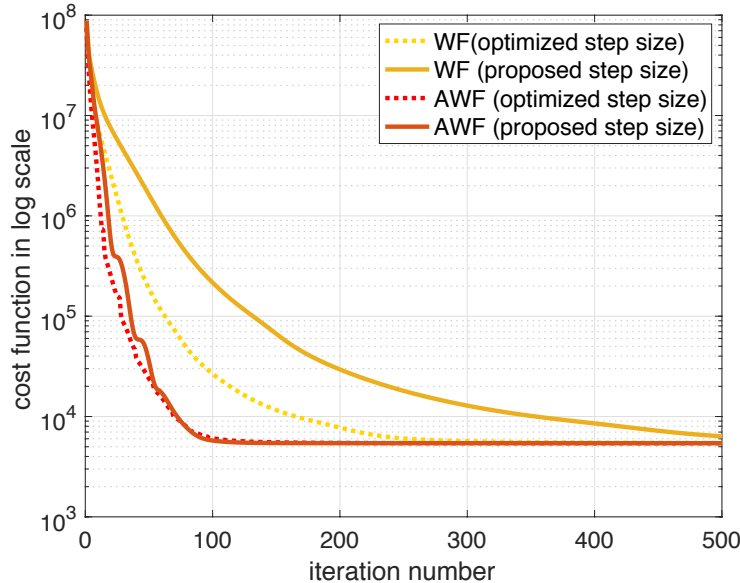


Figure 2: Evolution of the cost function throughout with iterations: Wirtinger Flow (WF) and its accelerated version (AWF) implement the updates in (6) and (10), respectively.

## 4 Numerical Results

We illustrate the practical benefits of the AWF algorithm with both simulated and experimental data. We start by investigating the efficiency of our analytical step size. We simulate the optical system in Figure 1 using the following physically-accurate parameters: separation of the LEDs is 4 mm; distance of the LED array to the sample is 77 mm; illumination wavelength is 514 nm; microscope objective has 0.1 NA with  $8\times$  magnification; pixel size is  $6.5\mu\text{m}$ . We consider a total number of 293 LEDs. Randomly chosen 4 LEDs are lit at the same time for multiplexing.

For comparison, we manually optimize the step size for WF where we aim at achieving the fastest-possible convergence speed while ensuring that the cost function decreases as the iterations proceed. Once the step size is tuned, we also incorporate the Nesterov’s acceleration. We run the algorithms for 500 iterations. All methods use the same initialization that sets a constant image for both amplitude and phase.

By looking at the convergence plots illustrated in Figure 2, we see that the AWF with our proposed step size is as efficient as its variant that uses a manually-tuned step size. This shows that our parameter-free approach does not compromise performance, providing us with a practical framework that does not require any tuning. We see that Nesterov’s acceleration notably improves the convergence speed for both cases. We also note that there is a significant

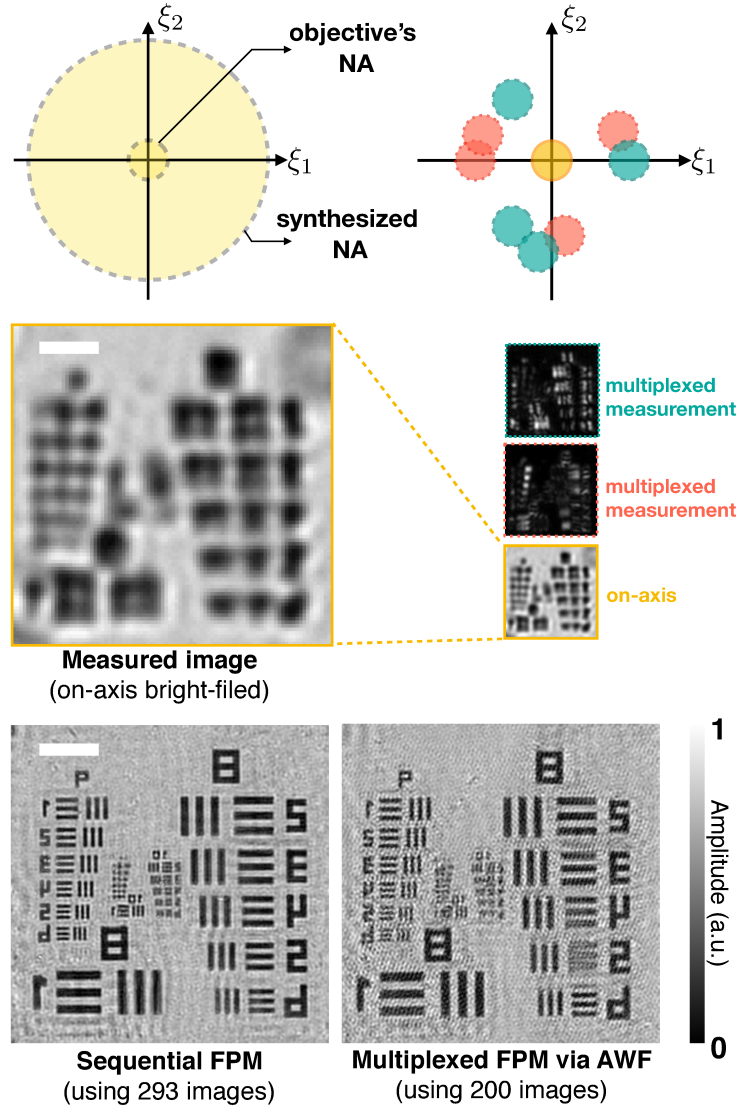


Figure 3: Application of the proposed algorithm to USAF resolution target. Scale bar is  $50\text{ }\mu\text{m}$ . See text for further information.

difference between the non-accelerated methods. This highlights the importance of acceleration strategies for nonlinear inverse problems.

For the experimental data, we image a USAF resolution target with a commercial Nikon TE300 inverted microscope that uses programmable  $32 \times 32$  LED array. Experimental parameters are the same as those used in the simulations.

We consider  $4 \times$  multiplexed measurements (total of 200 images) as well as



the sequential ones (total of 293 images) for validation. We use AWF with the proposed step size for both reconstructions (see Figure 3). Our experiments show that FPM drastically improves the spatial resolution (approximately  $5\times$ ) of the imaging system. We also see that the reconstruction obtained from multiplexed measurements (with 30% less data) is similar to the one from the sequential data, but exhibits more artifacts.

## 5 Conclusion

We introduced a theoretically-sound reconstruction algorithm for multiplexed FPM. Our main contribution has been the proposal of an analytical step size for which we established a stationary point convergence. Considering a Nesterov-type acceleration, we have shown that the practical convergence is as fast as the case where the step size is manually optimized.

## Acknowledgements

M. Soltanolkotabi is supported by the Packard Fellowship in Science and Engineering, a Sloan Research Fellowship in Mathematics, an NSF-CAREER under award #1846369, the Air Force Office of Scientific Research Young Investigator Program (AFOSR-YIP) under award #FA9550-18-1-0078, an NSF-CIF award #1813877, and a Google faculty research award.