

Aberration recovery by imaging a weak diffuser

GAUTAM GUNJALA,^{*} STUART SHERWIN, AAMOD SHANKER, AND LAURA WALLER

Department of Electrical Engineering & Computer Sciences, University of California, Berkeley, California 94720, USA

^{*}ggunjala@berkeley.edu

Abstract: We present a computational method for field-varying aberration recovery in optical systems by imaging a weak (index-matched) diffuser. Using multiple images acquired under plane wave illumination at distinct angles, the aberrations of the imaging system can be uniquely determined up to a sign. Our method is based on a statistical model for image formation that relates the spectrum of the speckled intensity image to the local aberrations at different locations in the field-of-view. The diffuser is treated as a wide-sense stationary scattering object, eliminating the need for precise knowledge of its surface shape. We validate our method both numerically and experimentally, showing that this relatively simple algorithmic calibration method can be reliably used to recover system aberrations quantitatively.

© 2018 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

1. Introduction

Aberrations are unavoidable in optical imaging systems; they distort incident wavefronts and prevent diffraction-limited resolution. Significant effort and expense goes into correcting aberrations, either physically or digitally. Particularly in computational imaging schemes (*e.g.* Fourier ptychography [1–3]) where both phase and amplitude are recovered, digital correction of aberrations is convenient [4]. We model aberrations using a real-valued function – the wavefront error function (WEF) – which can be either space-invariant or locally space-invariant. Image reconstruction methods that rely on accurate forward models (including aberrations) can particularly benefit from digital aberration correction for improved accuracy, reliability and effectiveness. Hence, there is a need for practical methods of aberration measurement.

Although there are existing methods that can accurately measure the WEF, many are complicated, requiring expensive hardware or precisely calibrated test objects. Interferometric techniques [5, 6] can be difficult to incorporate into existing optical systems. Pre-calibration by imaging known test objects (*e.g.* pinholes, fluorescent beads) [7,8] works well but requires precise knowledge of the test object. Adaptive optics methods both measure and correct aberrations [9–12]. Although there are advantages to physically (vs. digitally) correcting aberrations, the hardware required is expensive and not compatible with all commercial systems.

Computational techniques for WEF measurement include the Zemlin tableau method, used in transmission electron microscopy (TEM) to measure axial coma [13–16]. This involves imaging amorphous media with various tilts to produce a tableau of diffractograms which can be used to estimate the magnitudes of low-order aberrations. Other computational techniques seek to jointly estimate both the object being imaged and the system aberrations [17], which only work well if the object is suitable. Here, we aim to combine these two approaches, enabling use of the Zemlin tableau method for higher-dimensional aberration models through optimization and greatly simplifying the joint estimation problem by using a carefully chosen object.

We propose a computational method for quantitative WEF measurement, requiring only the imaging of a weak scattering object (a diffuser) under several different angles of plane-wave illumination. The procedure is outlined in Fig. 1. The exact surface shape of the diffuser need not be known, but should satisfy the weak object approximation (WOA) [18–20] such that it can be modeled by few statistical parameters [21, 22]. To create an inexpensive object that satisfies this,

we partially index-match a diffuser with oil. With appropriate illumination angles, the spatial spectra of the resulting intensity images uniquely specify the WEF. We use an optimization procedure to solve the inverse problem and determine the WEF parameters from the raw data.

Our method has several advantages over existing techniques for WEF estimation. Unlike interferometers or Shack-Hartmann wavefront sensors, we measure aberrations at the sensor plane, thus avoiding any potential mismatch between different optical paths and enabling recovery of field-varying aberrations. The only hardware requirements are steerable illumination and an uncalibrated diffuser, which need not be precisely known *a priori*. This simple and versatile method enables aberration calibration in systems where traditional techniques may be difficult to apply, and the lowered cost of calibration may allow more routine analysis.



Fig. 1. Overview of our wavefront error function (WEF) recovery procedure. (a) A weak diffuser is placed in the image plane of the system to be characterized. (b) A calibration speckle image is captured and its spectrum is used to estimate the statistical parameters of the diffuser. (c) Speckle images are measured for N distinct illumination angles and their spectra are shown. (d) Measurements are processed and input to (e) a nonlinear least-squares problem in order to recover the Zernike coefficients and (f) reconstruct the WEF.

2. Methods

In this section, we first derive the forward model that specifies how aberrations manifest in the images we collect. The model will assume a traditional imaging system that can be described either by a single complex optical transfer function in the case of space invariant aberrations, or local transfer functions in the case of spatially-varying aberrations. The phase of these functions is the WEF, which we will parametrize using Zernike polynomials, enumerated by the standard Noll indices [23]. Once the forward model is derived, we will formulate the objective function for the inverse problem and describe how we solve it to recover the coefficients of the WEF.

2.1. Forward model

We model the measured intensity image, I(x), in terms of the electric field, E, and the coherent pupil function associated with the imaging system, P:

$$I(\mathbf{x}) = |E_i(\mathbf{x})|^2 = |P(\mathbf{x}) * E_o(\mathbf{x})|^2,$$
(1)

where $x \in \mathbb{R}^2$ is the spatial coordinate, * denotes convolution and subscripts *i* and *o* indicate the image and object planes, respectively [24]. Taking the Fourier transform of Eq. (1), we have

$$\widehat{I}(\boldsymbol{u}) = \widehat{E}_i(\boldsymbol{u}) * \widehat{E}_i^*(-\boldsymbol{u}) = \widehat{P}(\boldsymbol{u})\widehat{E}_o(\boldsymbol{u}) * \widehat{P}^*(-\boldsymbol{u})\widehat{E}_o^*(-\boldsymbol{u}),$$
(2)

where $u \in \mathbb{R}^2$ is the NA-normalized spatial frequency coordinate and hatted quantities are related to their counterparts by Fourier transform [25].

We refer to the complex-valued function P(u) as the band-limited pupil function. In our analysis, we assume that it has no amplitude variation within the aperture since typical optical elements (*e.g.* lenses, mirrors) have trivial absorption. Thus, the pupil is written as

$$P(\boldsymbol{u}) = \exp\left[iW(\boldsymbol{u})\right] \cdot \operatorname{Circ}(\boldsymbol{u}),\tag{3}$$

where *i* is the imaginary unit, the $Circ(\cdot)$ function is the characteristic function of the unit disk, and the real-valued function *W* denotes the WEF.

If the weak scattering object (the diffuser) is illuminated by an angled plane wave, it causes a Fourier domain shift of u_0 . The electric field at the object plane is then

$$E_o(\boldsymbol{x}) = \exp\left[i2\pi \left(\boldsymbol{u}_{\boldsymbol{0}}^{\mathsf{T}} \boldsymbol{x}\right)\right] \cdot \exp\left[i\varphi(\boldsymbol{x})\right].$$
(4)

We then apply the weak object approximation to the object term of Eq. (4) [26]. This is a linearization of the complex exponential, which implies that our object has trivial absorption and small deviations in optical path length. Mathematically, we substitute $\exp[i\varphi(\mathbf{x})]$ with $1 + i\varphi(\mathbf{x})$ and the Fourier transform of the resulting expression is

$$\widehat{E_o}(\boldsymbol{u}) = \delta(\boldsymbol{u} - \boldsymbol{u_0}) + i\widehat{\varphi}(\boldsymbol{u} - \boldsymbol{u_0}).$$
(5)

Substitution of Eq. (5) into Eq. (2) leads to our forward model (see Appendix A for details):

$$\widehat{I}(\boldsymbol{u}) = \delta(\boldsymbol{u}) \big[|\widehat{P}(\boldsymbol{u}_0)|^2 + \gamma \big] + i \widehat{\varphi}(\boldsymbol{u}) \big[\widehat{P^*}(\boldsymbol{u}_0) \widehat{P}(\boldsymbol{u} + \boldsymbol{u}_0) - \widehat{P}(\boldsymbol{u}_0) \widehat{P^*}(-\boldsymbol{u} + \boldsymbol{u}_0) \big].$$
(6)

For pupil recovery, we shall primarily be concerned with the latter term in Eq. (6). We define

$$\widehat{I}_{\varnothing}(\boldsymbol{u}) = i\widehat{\varphi}(\boldsymbol{u}) \Big[\widehat{P^*}(\boldsymbol{u}_0)\widehat{P}(\boldsymbol{u}+\boldsymbol{u}_0) - \widehat{P}(\boldsymbol{u}_0)\widehat{P^*}(-\boldsymbol{u}+\boldsymbol{u}_0) \Big],$$
(7)

where \widehat{I}_{\emptyset} represents the DC-suppressed counterpart of \widehat{I} . We introduce the symmetric and anti-symmetric decomposition operators, denoted by $S\{\cdot\}$ and $\mathcal{R}\{\cdot\}$, respectively:

$$S\{f(u+u_0)\} = \frac{f(u+u_0) + f(-u+u_0)}{2} , \ \mathcal{A}\{f(u+u_0)\} = \frac{f(u+u_0) - f(-u+u_0)}{2}. \ (8)$$

Using this notation and substituting Eq. (3), Eq. (7) simplifies to

$$\widehat{I}_{\varnothing}(\boldsymbol{u}) = 2i \exp\left[i\mathcal{A}\left\{W(\boldsymbol{u}+\boldsymbol{u}_{0})\right\}\right]\widehat{\varphi}(\boldsymbol{u})\sin\left(\mathcal{S}\left\{W(\boldsymbol{u}+\boldsymbol{u}_{0})\right\}\right).$$
(9)

We remark that this simplification requires both difference operands in Eq. (7) to be nonzero, such that interference occurs and produces contrast in intensity. Since each instance of the pupil

 \widehat{P} is band-limited by the Circ(·) function, the region where the interference condition is met is given by the set $\mathscr{U} = \{u : Circ(u + u_0) = 1\} \cap \{u : Circ(u - u_0) = 1\}$. As a result, Eq. (9) is valid only within the region where the pupils overlap, as shown in Fig. 1(d). Finally, by taking the magnitude of both sides of Eq. (9), we arrive at our real-valued forward model,

$$\left|\widehat{I_{\mathscr{Q}}}(\boldsymbol{u})\right| = 2\left|\widehat{\varphi}(\boldsymbol{u})\right| \left|\sin\left(\mathcal{S}\left\{W(\boldsymbol{u}+\boldsymbol{u}_{0})\right\}\right)\right|, \ \boldsymbol{u} \in \mathscr{U}.$$
(10)

Our model indicates that for plane-wave illumination, the output intensity spectrum is determined by the *symmetric* component of the shifted WEF. We emphasize that this model only describes the region of the spectrum where the pupil interferes with its conjugate. For off-axis illumination, this is the region of overlap of the two shifted circles determined by the NA of the system.

2.2. Diffuser as a calibration object

Equation (10) suggests that the object must be known in order to recover the WEF; however, we can avoid this by using an object with a random surface whose statistics are known (or can be measured via calibration). Our partially index-matched holographic diffuser (Edmund Optics 10°, #54-493) has uncorrelated surface roughness with significant power in high spatial frequencies, extending at least to the band-limit of the optical system being characterized (see Fig. 2). This leads to an efficient model for the object as the product of a deterministic window function and a stationary random signal [21]. Hence, our calibration step is reduced to the estimation of a few parameters, which can be recovered directly. Stationarity also eliminates the need for precise alignment of the diffuser in the imaging system, an advantage over techniques that use known test objects. Mathematically, the diffuser can be modeled by:

$$|\widehat{\varphi}(\boldsymbol{u})| = |\widehat{\varphi_d}(\boldsymbol{u})|\eta(\boldsymbol{u}),\tag{11}$$

where $|\widehat{\varphi_d}(\boldsymbol{u})|$ is a radially-symmetric Gaussian window function, and $\eta(\boldsymbol{u}) \sim \text{Rayleigh}(\sigma)$ is a single realization of Rayleigh-distributed white noise, independent for each pixel. As a consequence of stationarity, arbitrary lateral motion of the diffuser will change the values of $\eta(\boldsymbol{u})$ but not the distribution parameter σ . We will assume in this section that the window function and distribution parameter are known; in practice, we estimate them from a simple single-image calibration procedure (outlined in Appendix B).



Fig. 2. (a) Measured phase, $\varphi(\mathbf{x})$, of a section of a 10° holographic diffuser (prior to index matching). (b) Spectrum of full intensity measurement (DC-suppressed and zeroed outside normalized cutoff frequency) displays rings due to defocus. (c) Radial average of spectrum.

We may now substitute Eq. (11) into Eq. (10) and collect known quantities on the left-hand side of the equation. This implies division of the spectrum of each measurement by the known window function, thus isolating the sinusoidal signal from the multiplicative noise (due to diffuser speckle). For the *j*-th illumination angle, we then have a processed measurement, M_i , given by

$$M_{j}(\boldsymbol{u}) \equiv \frac{|\widehat{I_{\emptyset,j}}(\boldsymbol{u})|}{2|\widehat{\varphi_{d}}(\boldsymbol{u})|} = \eta(\boldsymbol{u}) \left| \sin \left(\mathcal{S} \left\{ W(\boldsymbol{u} + \boldsymbol{u}_{j}) \right\} \right) \right|, \ \boldsymbol{u} \in \mathcal{U}_{j}.$$
(12)

At this point, we see that each processed measurement depends only on the WEF, and the diffuser only appears via the multiplicative noise term $\eta(u)$.

2.3. Inverse problem

Next, we detail our inverse problem approach to recovering the WEF. For efficient computation, we rewrite the argument of the sinusoid in Eq. (12) as a linear operator, $\mathbf{A}_j \in \mathbb{R}^{m \times n}$, which acts on the WEF's Zernike coefficients, $\mathbf{c} \in \mathbb{R}^n$. The following is a description of the construction of \mathbf{A}_j . The map from $W(\mathbf{u}) \to W(\mathbf{u} + \mathbf{u}_j)$, where \mathbf{u}_j is a known shift due to illumination angle, can be represented by a linear operator on Zernike coefficients, $\Sigma_j \in \mathbb{R}^{n \times n}$. If the illumination angle is unknown, it can be estimated from the acquired image using a procedure outlined in [27]. We then take the symmetric part of $W(\mathbf{u} + \mathbf{u}_i)$ by selecting only the coefficients of symmetric basis polynomials. If s_j is an indicator that the *j*-th basis function is symmetric, then this operation is given by diag($\mathbf{s} \in \mathbb{R}^{n \times n}$. Finally, we evaluate the resulting coefficients on a grid of *m* pixels through the evaluation map, $\Psi \in \mathbb{R}^{m \times n}$. We now define the design matrix for the *j*-th illumination angle by $\mathbf{A}_j \equiv \Psi \operatorname{diag}(\mathbf{s}) \Sigma_j$. Using these matrices and vectorizing the intensity spectrum, window function and white noise, we rewrite Eq. (12) as:

$$\mathbf{m}_{j} \equiv \frac{|\widehat{I_{\emptyset,j}}|}{2|\widehat{\varphi_{d}}|} = \eta \cdot \left| \sin \left(\Psi \operatorname{diag}(\mathbf{s}) \Sigma_{j} \mathbf{c} \right) \right| = \eta \cdot \left| \sin \left(\mathbf{A}_{j} \mathbf{c} \right) \right|.$$
(13)

Based on Eq. (7), aberrations of orders zero and one (piston, tip, and tilt) do not appear in the spectrum and thus cannot be estimated. As a result, corresponding rows and columns of the matrices defined above are omitted to avoid degeneracy in the solution space.

Next, we consider the diffuser surface uncertainty, η , which acts as a realization of multiplicative noise. Although regularized likelihood functions have be formulated for this case [28], they usually involve computing logarithms, which is unstable for signals containing zeros (*e.g.* sine). Instead, we treat η as additive noise and formulate a least-squares problem by approximating the Rayleigh noise distribution as a Gaussian distribution of equal mean and variance:

$$\eta \cdot \left| \sin\left(\mathbf{A}_{j}\mathbf{c}\right) \right| \approx \left(\mathbb{E}[\eta] + v \right) \left| \sin\left(\mathbf{A}_{j}\mathbf{c}\right) \right|, \quad v \sim \mathcal{N}\left(0, \frac{4-\pi}{2}\sigma^{2}\right).$$
 (14)

The expectation is computed using an unbiased estimator: $\mathbb{E}[\eta] = \hat{\sigma} \left(\frac{\pi}{2}\right)^{1/2}$ leading to:

$$\eta \cdot \left| \sin \left(\mathbf{A}_{j} \mathbf{c} \right) \right| \approx \mathbb{E}[\eta] \cdot \left| \sin \left(\mathbf{A}_{j} \mathbf{c} \right) \right| + \nu \left| \sin \left(\mathbf{A}_{j} \mathbf{c} \right) \right|,$$
 (15)

in which the latter term can be treated as heteroscedastic additive noise [29], with aberrationdependent variance proportional to $|\sin(\mathbf{A}_j \mathbf{c})|^2$. If the aberrations are approximately known *a priori*, a weighted nonlinear least-squares (NLS) fit will be more robust to the changing noise variance across the field-of-view. However, we do not generally have this prior knowledge, so we use an unweighted NLS cost function, which in practice is suitably accurate.

Finally, we use the known (or estimated) illumination angles to define the interference regions \mathcal{U}_i for each measurement. The characteristic functions, $\mathbb{1}[\mathcal{U}_i]$, for each of these sets is a binary

vector in \mathbb{R}^m which determines whether a particular pixel of a spatial spectrum should contribute to the cost function. Using the approximations above, we formulate a cost function that penalizes inconsistency between the measurements and our forward model applied to a candidate WEF. The resulting optimization problem can be written as

$$\mathbf{c}^{\star} = \arg\min_{\mathbf{c}} \sum_{j=1}^{K} \left\| \mathbb{1}[\mathscr{U}_{j}] \circ \left(\mathbf{m}_{j} - \sigma\left(\frac{\pi}{2}\right)^{1/2} \left| \sin\left(\mathbf{A}_{j}\mathbf{c}\right) \right| \right) \right\|^{2}, \tag{16}$$

where \circ indicates element-wise (Hadamard) multiplication and the minimizer, c^* , gives the Zernike coefficients of the recovered WEF.

To solve the optimization problem in Eq. (16), we use Newton's method with step size determined by backtracking line search (see Appendix C) [30]. Since our objective function is non-convex, this method is not guaranteed to converge to the global minimum for any particular initialization. Therefore, we initialize the algorithm with multiple random points within a neighborhood of bounded total root mean square (RMS) wavefront error ($||\mathbf{c}||_2$), obtain the local minimum from each initialization, and report the local minimum with the lowest cost.



Fig. 3. WEF reconstruction from simulated data. (a) Example intensity measurement; (b) WEF with 18 random Zernike coefficients (Noll indices 4-21); (c) spectra of 4 intensity images with illumination angles described by azimuthal angle θ and deflection angle ϕ ; (d) processed spectra, $M_j(u)$, as in Eq. (13); (e) spectra produced by the forward model using the true WEF coefficients; (f) spectra produced by the forward model using the recovered WEF coefficients; (g) recovered WEF coefficients demonstrate accurate recovery (3.11% relative error) of 5th order WEF using four input images and reasonable initialization values.

3. Numerical simulations

To test our method against ground truth data, we created a simulation of the entire procedure, including object calibration and aberration recovery. First, we generate the electric field of a weak scattering object as described in [21] and a random set of Zernike coefficients for the WEF. We simulate a single defocused on-axis measurement via the angular spectrum method and perform object calibration, as described in Appendix B. In this case, one measurement uses on-axis illumination, and the other three have an axial deflection angle of 0.3° and azimuthal angles of 0° , 45° and 90° . The simulated diffuser has Gaussian parameter $\sigma=5 \ \mu m$ (close to that of a 10° diffuser with magnification of 1) and phase scaled to the range $\left[-\frac{\pi}{12}, \frac{\pi}{12}\right]$. The spectra of the simulated measurements are visibly distinct (see Fig. 3(c)). Using these images, we performed the processing routines outlined in Sec. 2.2-2.3 to recover the WEF parameters.

To test the robustness of our method, we simulated datasets for WEFs with random coefficients for aberrations (up to order 3) and magnitudes from 0.2π - 6π radians RMS, covering the typical range of aberrations for diffraction-limited imaging systems. A dataset consists of 4 intensity images (500×500 pixels) with illumination angles as specified and basic system information (*e.g.* wavelength, NA). For each magnitude, we generated 100 datasets with distinct WEFs.

Computation was implemented in MATLAB R2018a on a laptop (Intel Core i7-7700HQ processor 2.8 GHz, 16 GB of RAM). The average time to process all measurements was 0.9 seconds. The average time to recover the WEF from processed measurements using 100 random initializations and 100 iterations of Newton's method for each was 14.5 seconds. The reported run time reflects serial processing of initializations, but parallel processing would further reduce the run time. As shown in Fig. 4, at least one of the 100 initializations converged to within 5% relative error for roughly 95% of all WEFs with a magnitude of at most π radians RMS.

As the magnitude of the wavefront error increases above π radians RMS, the fraction of successfully recovered WEFs rapidly decreases to just 1% for 6π radians RMS aberrations. Despite this limitation, even in the case of strong aberrations, convergence is highly probable if the initialization is sufficiently close to the true coefficient vector. To demonstrate this, we performed a second set of simulations using only the 6π radian RMS WEFs, generating initializations by



Fig. 4. (a) Fraction of distinct WEFs successfully recovered as aberrations increase. Success is defined as convergence to within 5% relative error of ground truth for at least one of 100 initializations generated by 0-mean normally-distributed noise. Random initializations for small aberration magnitudes are likely to converge, with the probability decreasing as the aberration magnitude increases. (b) Fraction of distinct WEFs successfully recovered for the case of large aberrations (all 6π radians RMS) with increasing initialization distance. 100 initializations are generated by adding 0-mean normally-distributed noise of increasing magnitude to the true solution. This demonstrates that initializations with as much as 3π radians RMS error converge with high probability, suggesting that modestly accurate initializations will enable fitting to large aberrations.

adding increasingly strong normally-distributed noise to the true coefficients; this can be thought of as initializing with partial prior information. The results are shown in Fig. 4(b), where we see that convergence is highly probable if the initialization is within 3π radians RMS of the true solution, which represents 50% accuracy. These results suggest that our procedure will work well with no prior knowledge for sufficiently weak aberrations, and will require either many initializations or a modestly accurate initial guess for stronger aberrations. Such a guess can be obtained, for example, via ray-tracing using an ideal model of the system.

4. Experimental results

To demonstrate the accuracy of our method experimentally, we constructed the setup shown in Fig. 5. The key features of the system are adjustable-angle laser illumination, a holographic diffuser with index-matching oil in the object plane, and a deformable mirror (DM) in the Fourier plane to generate known WEFs for validation testing. It is important for the deviation in refractive index between the diffuser and the oil to be small, so that the weak object approximation is valid. In our experiment, we have $n_{\text{diffuser}} = 1.5805$ and $n_{\text{oil}} = 1.5890$.



Fig. 5. Experimental setup. (a) Melles-Griot HeNe laser (632 nm), (b) 4*f* system with mirror tilt used to control illumination angle at the object (diffuser) plane; (c) 10° holographic diffuser (Edmund Optics, #54-493) index-matched with oil; (d) Deformable Mirror (Iris AO PTT111, 7 mm pupil diameter, gold-coated) in Fourier plane to introduce controlled aberrations; (e) ThorLabs DCC1240C CMOS camera (1280 × 1024 pixels, 5.3 μ m pixels).

We tested the accuracy of our method by applying different aberration functions to the DM, acquiring images from 3 different illumination angles, and then passing the images into our reconstruction algorithm. Each aberration function is a single Zernike polynomial whose magnitude is significantly larger than the imaging system's native aberrations. Although only a single Zernike polynomial is applied to the DM at a time, we fit all Zernike coefficients of orders 2 and 3 (Noll indices 4-10). For experimental data, we modify the reconstruction procedure to remove a small neighborhood of low frequencies, as well as the DC, from each spectrum, since correlated surface roughness produces a peak of finite width. We also do not know the exact illumination angles, and so these are estimated using [27].

Table 1 and Fig. 6 summarize the results of our experiments, using the two initialization methods discussed in Sec. 3. The left columns of the table and Fig. 6(a) show results initialized with coefficients centered around the expected aberration coefficients, with added noise. An



Fig. 6. Experimental fit to aberrations Z5, Z6, Z7 and Z9 using (a) 200 centered initializations (actual coefficients plus zero-mean white noise) and (b) 10,000 zero-mean random initializations. Bars and points of a single color represent recovered and actual coefficients, respectively, for a particular aberration magnitude. Recovered values of Z5 and Z6 demonstrate the most accurate recovery, while Z7 and Z9 (polynomials of higher degree) incur larger relative error, especially when initializations have zero mean. The relative error increases when fringe contrast in the measurements is low.

	Centered Init. Error					Zero-mean Init. Error			
Magnitude	Z5	Z6	Z7	Z9	-	Z5	Z6	Z7	Z9
1.0			1.37	0.57				1.37	1.06
2.0	0.40	0.45	2.45	0.77		0.40	0.45	2.36	2.69
3.0	0.43	0.45	1.70	0.79		0.42	0.46	2.99	0.79
4.0	0.47	0.47	1.51	0.97		0.48	0.47	4.60	0.95
5.0	0.44	0.50	1.50	1.11		0.44	0.51	5.15	1.12

Table 1. ℓ^2 -Reconstruction Error^{*} for Second- and Third-Order Aberrations

*Error calculated as $\|\mathbf{c}_{true} - \mathbf{c}_{recovered}\|_2^2$ and expressed in $\pi \times radians$ RMS

accuracy of 0.4π - 0.5π radians RMS compared to the expected aberrations is achieved for both oblique and vertical astigmatism (Noll indices Z5 and Z6, respectively), while a somewhat larger error of 0.5π - π radians RMS is achieved for vertical trefoil (Z9) and π - 2π radians RMS for vertical coma (Z7). The error is calculated as the RMS difference between the recovered coefficients and the coefficients written to the DM. Similar performance is achieved for Z5 and Z6 via random initialization, shown in the right columns of Table 1 and Fig. 6(b). However, results for Z9 are somewhat degraded and those for Z7 are significantly worse. Differences in performance among the aberrations seem to be related to the fringe contrast in experimental measurements, shown in Fig. 7, where we see that Z5 and Z6 have high contrast, Z9 has lower



Fig. 7. Experimental fit to (a) oblique astigmatism (Z5), (b) vertical astigmatism (Z6), (c) vertical coma (Z7), and (d) vertical trefoil (Z9). (*Left column*) Input polynomial coefficient written to DM plotted vs. recovered coefficient in π radians RMS. (Right) Examples of measured and reconstructed intensity spatial spectra for 3 illumination angles and aberrations of varying magnitude (π radians RMS). Rows (a) and (b) show excellent quantitative agreement between expected and recovered coefficients. Rows (c) and (d) are somewhat worse, likely due to lower contrast in the interference fringes at the edges of the pupil.

contrast, and Z7 has substantially less contrast, particularly at the edge of the pupil. In all cases, Fig. 7 shows agreement between the measured and best-fit intensity spectra in areas with high fringe contrast, demonstrating the accuracy of our model and the capabilities of our solver.

4.1. Experimental limitations

It is important to note several experimental limitations of our method. First, we can see that the experimental spectra in Fig. 7 appear significantly more noisy than the simulated spectra in Fig. 3, which tends to reduce the accuracy of the fit. Another limitation is that by changing the angle of illumination, it is only possible to recover aberration coefficients with sign ambiguity. It is evident from Eq. (16) that replacing \mathbf{c} with $-\mathbf{c}$ does not alter the cost. This problem can be solved by acquiring images at multiple defocus distances, but for simplicity, we consider only the case of modulating illumination angle for aberration recovery up to a sign. As system aberrations become more significant, the region of attraction of the true solution generally becomes smaller, as demonstrated in Fig. 4. Finally, use of unweighted least-squares, as discussed in section 2.3, is not statistically optimal in that it is an unbiased but not minimum-variance estimator. The variance can be reduced by formulating a weighted least-squares problem to address the heteroscedasticity of the noise, with weights determined either by an estimate of the denoised signal (*e.g.* using median-filtering) or by iterative reweighting.

5. Conclusion

We have presented a self-calibrated computational method for aberration recovery in imaging systems using only a weak scattering object (a diffuser). Adopting a statistical model of the object allows a simplified forward model for image formation, relating the system aberrations to the spatial spectra of acquired intensity images. Measurements at different illumination angles provide sufficient diversity for uniquely specifying the WEF. This simple and inexpensive method can be used to reliably characterize optical imaging systems, including ones where traditional techniques may be incompatible.

Funding

STROBE: A National Science Foundation Science and Technology Center (DMR 1548924); Moore Foundation Data-Driven Discovery Initiative; Center for Design Enabled Nanofabrication (C-DEN) (86368-23800-44–EHAXN).

Acknowledgments

The authors thank Michael Helmbrecht (at Iris AO, Inc.) for providing the deformable mirror used in experiments (http://www.irisao.com/product.ptt111.html) and for providing technical assistance. The authors also thank Emrah Bostan for insightful discussions.

Appendix A: Derivation of forward model

We substitute Eq. (5) into Eq. (2), resulting in an expression for the spectrum of measured intensity. Since convolution distributes over addition, we can write this expression as the sum of four convolutions, as follows:

$$\hat{I}(u) = \left[\delta(u - u_0) \hat{P}(u) * \delta(-u - u_0) \hat{P}^*(-u) \right] + j \left[\delta(-u - u_0) \hat{P}^*(-u) * \hat{\varphi}(u - u_0) \hat{P}(u) \right]
- j \left[\delta(u - u_0) \hat{P}(u) * \hat{\varphi}^*(-u - u_0) \hat{P}^*(-u) \right] + \left[\hat{\varphi}(u - u_0) \hat{P}(-u) * \hat{\varphi}^*(-u - u_0) \hat{P}^*(-u) \right].$$
(17)

Since the weak phase of the object is noise-like (see Fig. 2) and hence, uncorrelated, we can assume that the interactions between the phase of the object and itself, captured by the fourth term in Eq. (17), contribute only to the DC frequency through some scalar γ . As such, simplification of the above results in

$$\hat{I}(\boldsymbol{u}) = \delta(\boldsymbol{u}) \left[\hat{P}(\boldsymbol{u}_0) \hat{P^*}(\boldsymbol{u}_0) + \gamma \right] + i \hat{P^*}(\boldsymbol{u}_0) \left[\hat{\varphi}(\boldsymbol{u}) \hat{P}(\boldsymbol{u} + \boldsymbol{u}_0) \right] - i \hat{P}(\boldsymbol{u}_0) \left[\hat{\varphi^*}(-\boldsymbol{u}) \hat{P^*}(-\boldsymbol{u} + \boldsymbol{u}_0) \right].$$
(18)

Since $\varphi(\mathbf{x})$ is a real-valued function, we know that its Fourier transform, $\hat{\varphi}(\mathbf{u})$, satisfies conjugate symmetry. Using this fact, we can factor Eq. (18) to obtain

$$\hat{I}(\boldsymbol{u}) = \delta(\boldsymbol{u}) \Big[|\hat{P}(\boldsymbol{u}_0)|^2 + \gamma \Big] + i\hat{\varphi}(\boldsymbol{u}) \Big[\hat{P}^*(\boldsymbol{u}_0) \hat{P}(\boldsymbol{u} + \boldsymbol{u}_0) - \hat{P}(\boldsymbol{u}_0) \hat{P}^*(-\boldsymbol{u} + \boldsymbol{u}_0) \Big].$$
(19)

The remaining steps are given in section 2.1.

Appendix B: Object calibration

From a single calibration image taken on-axis with several Rayleigh lengths of defocus, we perform a parametrized joint object-pupil reconstruction to characterize the diffuser. We formulate a 3-parameter nonlinear least-squares problem to recover the amplitude and standard deviation of $|\widehat{\varphi_d}|$ and the defocus, which is the dominant term of the pupil $W(\boldsymbol{u})$ in the calibration measurement. This problem can be efficiently and accurately solved using derivative-free simplex search due to its low dimensionality and the radial symmetry of the defocus kernel. The average runtime of this step is 1.3 seconds using the computer with specifications given in section 3. Examples of simulated data and fitted functions can be seen in Fig. 8. We then divide the spectrum of the measurement by the deterministic terms where division is numerically stable to obtain independent samples of white noise drawn from the Rayleigh distribution:

$$\frac{|\widetilde{I}_{\phi}(\boldsymbol{u})|}{2|\widehat{\varphi_{d}}(\boldsymbol{u})|\left|\sin\left(\mathcal{S}\{W(\boldsymbol{u})\}\right)\right|} = \eta(\boldsymbol{u}) \sim \operatorname{Rayleigh}(\sigma)$$
(20)

Finally, we obtain an unbiased estimate of the Rayleigh parameter, $\hat{\sigma}$, from *K* samples of $\eta(u)$, using the following unbiased estimator [31]:

$$\hat{\sigma} = \left(\sum_{j=1}^{K} \frac{\eta(\boldsymbol{u}_{j})^{2}}{2K}\right)^{1/2} \frac{4^{K}K!(K-1)!\sqrt{K}}{(2K)!\sqrt{\pi}} \approx \left(\sum_{j=1}^{K} \frac{\eta(\boldsymbol{u}_{j})^{2}}{2K}\right)^{1/2} \frac{e^{1}\sqrt{K}}{\sqrt{K-1}} \left(\frac{K-1}{K}\right)^{K},$$

where the approximation is made using Stirling's formula [32]. A histogram of isolated white noise and the fitted distribution are shown in Fig. 8.

Appendix C: Derivatives of cost function

For the *j*-th measured image, $M_j(\boldsymbol{u})$, we define

$$\mathbf{e}_j = M_j(\boldsymbol{u}) - \sigma \left(\frac{\pi}{2}\right)^{1/2} \left| \sin \left(\mathbf{A}_j \mathbf{c}\right) \right|.$$



Fig. 8. (a) Simulated calibration procedure, in which the window function of the scatterer spectrum is estimated from a low-order (defocus only) model. (b) The Rayleigh noise parameter is estimated from the residuals upon dividing the measured spectrum by the damped defocus kernel in regions where division is stable.

Then the first derivative of cost function, Eq. (16), is given by

$$\nabla_{\mathbf{c}} f = -2\sigma \left(\frac{\pi}{2}\right)^{1/2} \sum_{j=1}^{K} \mathbf{e}_{j}^{\mathsf{T}} \operatorname{diag}\left(\cos\left(\mathbf{A}_{j}\mathbf{c}\right) \circ \operatorname{sgn}\left(\sin\left(\mathbf{A}_{j}\mathbf{c}\right)\right)\right) \mathbf{A}_{j}.$$

The Hessian matrix for the cost function is given by

$$\nabla_{\mathbf{c}}^{2} f = 2\sigma \left(\frac{\pi}{2}\right)^{1/2} \sum_{j=1}^{K} \mathbf{A}_{j}^{\mathsf{T}} \operatorname{diag}\left(\cos^{2}\left(\mathbf{A}_{j}\mathbf{c}\right) + \mathbf{e}_{j} \circ \sin\left(2\mathbf{A}_{j}\mathbf{c}\right) \circ \operatorname{sgn}\left(\sin\left(\mathbf{A}_{j}\mathbf{c}\right)\right)\right) \mathbf{A}_{j}.$$

It is important to note here that the expressions derived above are *generalized* derivatives [33]. The forward model is not differentiable at points where $\sin(\mathbf{A}_j \mathbf{c}) = 0$ due to the absolute value operation. However, since this operation is applied point-wise, we can treat each occurrence of zero as an isolated discontinuity and use the notion of generalized gradients to choose a *subgradient* for these points [34]. By using the signum function, we define this subgradient to be zero at non-differentiable points. As a result, the algorithm will only fail to make a valid update when $\mathbf{c} = 0$. As long as the algorithm is not initialized here, this point will be avoided with probability one.

Appendix D: Selection of illumination angles

Although it is difficult to rigorously analyze the selection of measurements for nonlinear least squares, in the limiting case of small aberrations, we make the approximation of $sin(x) \approx x$, and the problem reduces to linear least-squares, with the design matrix:

$$\mathbf{A} = \left(\begin{array}{c} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \dots \\ \mathbf{A}_K \end{array} \right).$$

In this case, the expected error on the coefficient vector, δc is given by:

$$\mathbb{E}\left[\|\delta \mathbf{c}\|_{2}^{2}\right] = \operatorname{Tr}\left[\mathbf{A}^{\dagger} \Sigma (\mathbf{A}^{\dagger})^{\mathsf{T}}\right],$$

where Σ is the measurement error covariance matrix and \mathbf{A}^{\dagger} is the pseudo-inverse of \mathbf{A} . Since measurements are independent, this is a diagonal matrix whose entries depend on the specific aberration. However, taking expectations over aberrations, Σ reduces to a scaled identity matrix, meaning that selecting measurements to minimize the metric Tr $[(\mathbf{A}^{\mathsf{T}}\mathbf{A})^{-1}]$ will provide a set of measurements that is robust to arbitrary aberrations.

We evaluated this metric for sample planes consisting of 3-8 equally-spaced angles at a range of radii. As shown in Fig. 9(a), the optimal radius is about 0.25 for any number of images. Taking the uniformly-spaced sample plans at the optimal radius for each number of images as initializations, no significant further improvement in the metric can be achieved via gradient-based optimization, meaning that these sample plans are at least locally optimal.

To evaluate these sample plans, we ran our reconstruction algorithm on the 100 random aberrations of magnitude π radian RMS. First, we initialize at the true coefficient value to quantify how the number of images impacts the accuracy. As shown in Fig. 10(a), the error scales approximately as $\frac{1}{\sqrt{N}}$, as would be expected for a linear least-squares fit. Next, we ran the optimization for 100 random initializations for each aberration. Fig. 10(b) shows that the frequency of convergence improves significantly with additional images.



Fig. 9. (a) Trace[$(\mathbf{A}^{\mathsf{T}}\mathbf{A})^{-1}$] evaluated for sample plans consisting of uniformly-spaced angles at radii varying from 0.05-0.5 of the pupil. (b) Uniformly-spaced sample plans with optimized radii (blue) and gradient-based optimization from these initializations.



Fig. 10. (a) Initializing at the true coefficients to evaluate how coefficient error scales with the number of images. (b) Initializing randomly to evaluate how the probability of convergence scales with the number of images.

References

- G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy," Nat. Photonics 7, 739 (2013).
- X. Ou, G. Zheng, and C. Yang, "Embedded pupil function recovery for Fourier ptychographic microscopy," Opt. Express 22, 4960–4972 (2014).
- J. Chung, J. Kim, X. Ou, R. Horstmeyer, and C. Yang, "Wide field-of-view fluorescence image deconvolution with aberration-estimation from Fourier ptychography," Biomed. Opt. Express 7, 352–368 (2016).
- A. Thust, M. Overwijk, W. Coene, and M. Lentzen, "Numerical correction of lens aberrations in phase-retrieval HRTEM," Ultramicroscopy 64, 249–264 (1996).
- P. Hariharan, B. Oreb, and T. Eiju, "Digital phase-shifting interferometry: a simple error-compensating phase calculation algorithm," Appl. Opt. 26, 2504–2506 (1987).
- M. P. Rimmer and J. C. Wyant, "Evaluation of large aberrations using a lateral-shear interferometer having variable shear," Appl. Opt. 14, 142–150 (1975).
- B. M. Hanser, M. G. Gustafsson, D. Agard, and J. W. Sedat, "Phase-retrieved pupil functions in wide-field fluorescence microscopy," J. Microsc. 216, 32–48 (2004).
- J. R. Fienup, J. C. Marron, T. J. Schulz, and J. H. Seldin, "Hubble space telescope characterized by using phase-retrieval algorithms," Appl. Opt. 32, 1747–1767 (1993).
- 9. R. K. Tyson, Principles of Adaptive Optics (CRC, 2015).
- 10. H. W. Babcock, "Adaptive optics revisited," Science 249, 253-257 (1990).
- N. Ji, D. E. Milkie, and E. Betzig, "Adaptive optics via pupil segmentation for high-resolution imaging in biological tissues," Nat. Methods 7, 141 (2010).
- A. Roorda, F. Romero-Borja, W. J. Donnelly III, H. Queener, T. J. Hebert, and M. C. Campbell, "Adaptive optics scanning laser ophthalmoscopy," Opt. Express 10, 405–412 (2002).
- F. Zemlin, K. Weiss, P. Schiske, W. Kunath, and K.-H. Herrmann, "Coma-free alignment of high resolution electron microscopes with the aid of optical diffractograms," Ultramicroscopy 3, 49–60 (1978).
- E. Voelkl, "Using diffractograms to evaluate optical systems with coherent illumination," Opt. Lett. 28, 2318–2320 (2003).
- 15. J. C. Spence, Experimental High-Resolution Electron Microscopy (Oxford University, 1988).
- R. Erni, Aberration-Corrected Imaging in Transmission Electron Microscopy: An Introduction (World Scientific, 2010).
- R. G. Paxman, T. J. Schulz, and J. R. Fienup, "Joint estimation of object and aberrations by using phase diversity," JOSA A 9, 1072–1085 (1992).
- D. Hamilton, C. Sheppard, and T. Wilson, "Improved imaging of phase gradients in scanning optical microscopy," J. Microsc. 135, 275–286 (1984).
- 19. N. Streibl, "Three-dimensional imaging by a microscope," JOSA A 2, 121–127 (1985).
- 20. H. Rose, "Nonstandard imaging methods in electron microscopy," Ultramicroscopy 2, 251-267 (1976).
- G. Gunjala, A. Shanker, V. Jaedicke, N. Antipa, and L. Waller, "Optical transfer function characterization using a weak diffuser," in *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XXIII*, (SPIE, 2016), p. 971315.
- 22. A. Shanker, A. Wojdyla, G. Gunjala, J. Dong, M. Benk, A. Neureuther, K. Goldberg, and L. Waller, "Off-axis aberration estimation in an EUV microscope using natural speckle," in *Imaging Systems and Applications*, (Optical Society of America, 2016), pp. ITh1F–2.
- 23. R. J. Noll, "Zernike polynomials and atmospheric turbulence," JOSA 66, 207-211 (1976).
- 24. J. Goodman, Introduction to Fourier Optics (McGraw-hill, 2008).
- 25. J. W. Goodman, Statistical Optics (John Wiley & Sons, 2015).
- 26. E. J. Kirkland, Advanced Computing in Electron Microscopy (Springer Science & Business Media, 2010).
- R. Eckert, Z. F. Phillips, and L. Waller, "Efficient illumination angle self-calibration in fourier ptychography," Appl. Opt. 57, 5434–5442 (2018).
- G. Aubert and J.-F. Aujol, "A variational approach to removing multiplicative noise," SIAM J. on Appl. Math. 68, 925–946 (2008).
- 29. R. H. Shumway and D. S. Stoffer, Time Series Analysis and its Applications (Springer, 2011).
- 30. S. Wright and J. Nocedal, Numerical Optimization (Springer Science, 1999).
- 31. M. Siddiqui, "Statistical inference for Rayleigh distributions," J. Res. Natl. Bureau Standards, Sec. D 68, 1007 (1964).
- 32. K. Lange, Applied Probability (Springer Science & Business Media, 2010).
- 33. F. H. Clarke, "Generalized gradients and applications," Transactions Am. Math. Soc. 205, 247–262 (1975).
- 34. J.-B. Hiriart-Urruty and C. Lemaréchal, Fundamentals of Convex Analysis (Springer-Verlag, 2001).