Phase retrieval using Gaussian basis functions

SEUNG-WHAN BAHK

Laboratory for Laser Energetics, University of Rochester
250 East River Road, Rochester, NY 14623, USA
sbah@lle.rochester.edu

Abstract: A wavefront map of arbitrary aperture can be represented by a sum of Gaussian basis functions. We compare its performance with other representation methods such as Zernike polynomials and Fourier modes. Gaussian basis functions can be applied to phase retrieval of high-frequency wavefront maps. Experiments show good agreement with direct measurements by a wavefront sensor. Optimum measurement conditions are discussed based on the statistics of wavefront properties.

1. Introduction

In imaging science, phase retrieval is a collection of techniques that enables the deduction of a 2-D wavefront map of the imaging system aberrations based on the observed image or point-spread functions. One effective approach in phase retrieval is the nonlinear optimization approach [1]. It is based on directly optimizing the representative parameters of the wavefront so that they minimize the error metric comparing the measured and calculated images. The optimization can be efficient because the search gradient is often available in the analytic form. The parameters that represent the wavefront can be point-by-point values of the map or coefficients of an orthonormal polynomial set. The pointwise representation is not limited to a specific aperture geometry but it suffers from phase-discontinuity artifacts in the retrieved wavefront in the presence of noise. Optimization using orthonormal polynomial representation, on the other hand, inherently guarantees the smoothness of the retrieved wavefront. Polynomial representation is tied to a specific aperture geometry on which the polynomials are constructed. The Zernike and Legendre polynomials, for example, are constructed over a circle and a rectangle, respectively. The orthonormal polynomials constructed over hexagonal and circular sector apertures are available [2, 3]. The Gram–Schmidt process allows one to construct orthonormal sets over an arbitrary aperture [4]. These polynomials tend to produce unrealistically high slopes around the edge of high-order modes. Another class of wavefront maps based on the Fourier modes, such as used in designing distributed phase plates for inertial fusion experiments [5], cannot be easily represented by polynomials or vice versa. To add flexibility for nonstandard apertures and generalize classes of wavefronts that can be represented, we explore the possibility of using Gaussian functions as basis functions for representing an arbitrary wavefront map, where an array of small, equally spaced Gaussian functions are superposed to form the wavefront.

Our approach is similar to the method of sieves with Gaussian kernels [6] but is based on a slightly different point of view. The method of sieves preserves the framework of point-by-point optimization, where the intermediate map is smoothed at each iteration indirectly by the spatially smoothed search direction, whereas the smoothing is automatically achieved here by use of the modal approach. Another similar approach uses the superposition of complex Gaussian beamlet functions, where it is necessary to use both amplitude and phase to take advantage of simple propagation properties of Gaussian beams [7]. Here, we use only the phase part represented by Gaussian functions and the propagation is handled using the global field. We first show the effectiveness of Gaussian modes for representing different types of wavefronts compared with polynomial and Fourier modes. We then formulate the nonlinear optimization phase-retrieval

approach for the Gaussian modes and present the analytic search gradient expression. In the subsequent sections, an optimum distance condition for the phase-retrieval measurement is discussed, followed by a description of experiments utilizing the Gaussian modes and the optimum measurement condition. Excellent agreement with direct measurements is demonstrated.

2. Gaussian modes

We show that an arbitrary wavefront $W(x, y)$ can be represented by equally spaced Gaussian functions with width w_0:

$$W(x,y) \approx \sum_{m,n} \exp\left\{-\left[\frac{x - x_{\text{grid}}(m)}{w_0}\right]^2 - \left[\frac{y - y_{\text{grid}}(n)}{w_0}\right]^2\right\} := \sum_{m,n} g(m,n) G(m,n;x,y), \quad (1)$$

where w_0 is the Gaussian function width and $x_{\text{grid}}(m)$, $y_{\text{grid}}(n)$ are the locations of the Cartesian grid on which the Gaussian functions are arranged with the spacing S. The grid format does not necessarily have to be Cartesian. It can be hexagonal for example. We will use Cartesian grid throughout this article. The $g(m, n)$ are real-valued coefficients. Although the Gaussian modes are not orthogonal, the least squares fit leads to the following approximate relation from which the coefficients $g(m, n)$ can be found:

$$\sum_{p,q}\left[\iint G(m,n;x,y)G(p,q;x,y)dxdy\right] g(p,q) \approx \iint G(m,n;x,y)W(x,y)dxdy. \quad (2)$$

Mapping the double indices that correspond to Gaussian modes into a single index, the above equation can be formulated into a simple matrix equation. The coefficient $g(p, q)$ can be found by a standard matrix inversion algorithm. The spacing S can be adjusted according to the *a priori* knowledge of the spatial frequency distribution. Four modes are required per period in the case of a purely sinusoidal wavefront. The inversion of the above linear equations is not essential for the phase retrieval but it is helpful for estimating the number of modes on existing maps or to make the transition from an intermediate estimate to a different representation mode.

An example of grid distribution is shown in Fig. 1. The grid points are indicated by red dots. The black circles around the red dots with the radius w_0 represent the width of the Gaussian functions. The optimum Gaussian width parameter w_0 with respect to the grid size S is numerically found to be about $2S$. Having a smaller grid per beam size ensures better fit. Also having more Gaussian modes around the perimeter of the beam ensures less error in the fit around the edge. In the example given, 12 grids per beam size and the outermost grid positioned up to 110% of the beam size exhibited 0.02 waves rms error.

Gaussian modes can represent a wavefront constructed either by polynomials or Fourier modes. Simple polynomial terms such as linear or quadratic functions are not easily represented by Fourier modes. Likewise, high-frequency Fourier modes are not easily represented by polynomials. In Fig. 2, the results of Gaussian mode fit against a quadratic function and another represented by random Fourier modes are shown. The quadratic wavefront shown in Fig. 2(a) is represented by 240 Gaussian modes in Fig. 2(b). The resulting error is 0.024 waves in rms, as shown in Fig. 2(c). A random Fourier mode wavefront shown in Fig. 2(d) is also well represented by Gaussian modes [Fig. 2(e)]. The rms error is 0.005 waves [Fig. 2(f)].

3. Application in nonlinear optimization phase retrieval

Nonlinear optimization in phase retrieval requires availability of the analytic gradient form of the error metric for efficient convergence. It will be shown that the gradient has the same form as

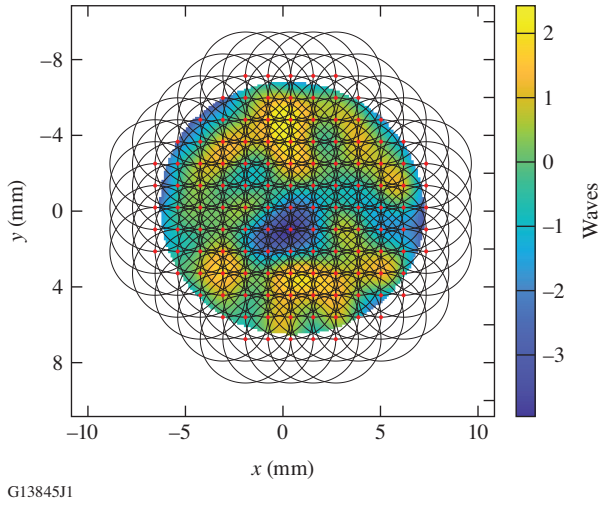

Fig. 1. Grid distribution example for a circular wavefront map

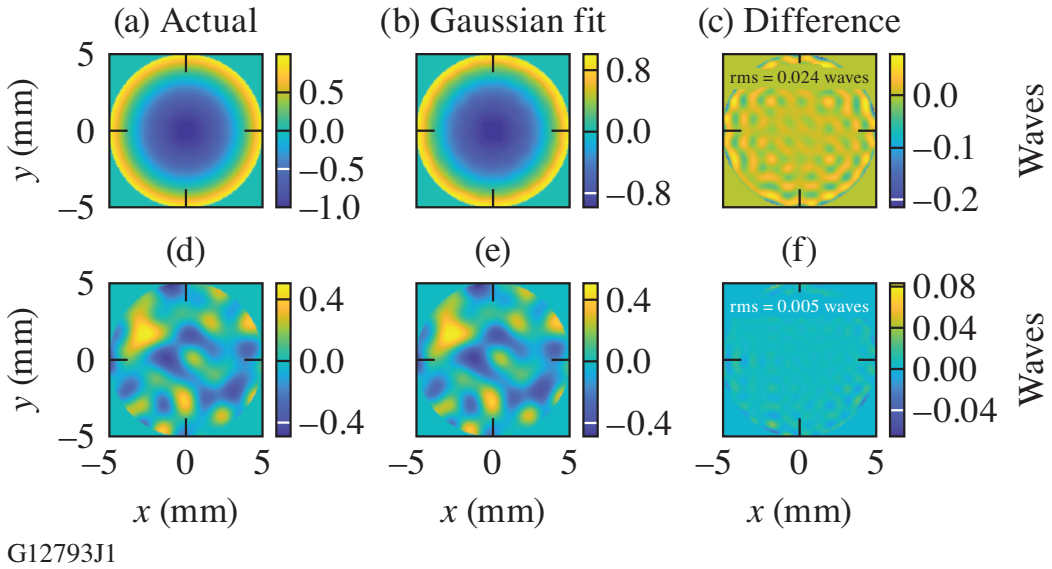

Fig. 2. Example of Gaussian mode representation for (a) a quadratic wavefront and (d) a random Fourier mode wavefront. [(b),(e)] Respective reconstruction using Gaussian modes. [(c),(f)] Residual errors.

the one shown by Fienup for the case of Zernike basis functions [1]. Although Gaussian basis functions are not orthogonal, the non-orthogonality does not affect the optimization formalism. Given the pupil plane intensity I_{NF}, its corresponding complex field $u^{(n)}$ at the nth iteration whose phase is represented by Gaussian functions $G_k(x, y)$ is

$$u^{(n)} = \sqrt{I_{\mathrm{NF}}} \exp\left[2\pi i \sum_k \alpha_k^{(n)} G_k(x, y)\right], \tag{3}$$

where $\alpha_k^{(n)}$ is the Gaussian mode coefficient for the kth mode at the nth iteration. We deal with two types of error metrics. The ϵ_1, which is the squared sum of amplitude difference and the ϵ_2, which is the squared sum of intensity difference as defined below:

$$\epsilon_1^{(n)} = \iint \left(\sqrt{I_{\mathrm{FF}}} - |U^{(n)}|\right)^2 dx'dy', \tag{4}$$

$$\epsilon_2^{(n)} = \iint \left(I_{\mathrm{FF}} - |U^{(n)}|^2\right)^2 dx'dy', \tag{5}$$

where I_{FF} is the measured far-field intensity and (x', y') is the far-field coordinates.

The differential pupil plane complex field $\Delta u^{(n)}$ back-propagated from $u^{(n)}$ and I_{FF} is given by

$$\Delta u_1^{(n)} = \mathcal{P}^{-1}\left\{\sqrt{I_{\mathrm{FF}}}\exp[i\arg(U^{(n)})] - U^{(n)}\right\} \tag{6}$$

for the amplitude error metric ϵ_1 and

$$\Delta u_2^{(n)} = \mathcal{P}^{-1}\left\{2U^{(n)}\left(|U^{(n)}|^2 - I_{\mathrm{FF}}\right)\right\} \tag{7}$$

for the intensity error metric ϵ_2. $U^{(n)} = \mathcal{P}(u^{(n)})$. $\mathcal{P}(\cdot)$ is the propagation operator that converts the field from the pupil plane to the far-field plane.

The gradient expression using the Parseval theorem is

$$\frac{\partial \epsilon^{(n)}}{\partial \alpha_k^{(n)}} = 4\pi \mathrm{Im} \iint \Delta u^{(n)*}(x, y) u^{(n)}(x, y) G_k(x, y) dxdy, \tag{8}$$

where it is assumed that $\Delta u^{(n)}$ corresponding to the specific error metric type is used.

There is a similarity between the Gaussian mode approach and the method of sieves. The above expression can be rewritten as

$$\frac{\partial \epsilon^{(n)}}{\partial \alpha_k^{(n)}} = 4\pi \mathrm{Im} \iint \Delta u^{(n)*}(x, y) u^{(n)}(x, y) G(x - x_k, y - y_k) dxdy. \tag{9}$$

The method of sieves preserves the framework of the point-by-point optimization except that its search gradient is spatially smoothed by convolving with a blurring function H. This smoothing in the search map reduces the possibility of phase discontinuity artifacts in the result in the presence of noise. The gradient map for the method of sieves is

$$\frac{\partial \epsilon^{(n)}}{\partial W^{(n)}(x', y')} = 4\pi \mathrm{Im} \iint \Delta u^{(n)*}(x, y) u^{(n)}(x, y) H(x - x', y - y') dxdy, \tag{10}$$

where $W^{(n)}(x', y')$ is the wavefront value at (x', y') at the nth iteration. The comparison of the above two expressions shows that the Gaussian mode approach is a spatially discretized version of the method of sieves given the blurring function is a Gaussian function. Phase-retrieval simulation using either approach produces comparable results given the same Gaussian width, although the method of sieves will be slower because of more convolution operations to perform.

4. Optimum propagation distance for phase retrieval

Before discussing experimental results, we discuss the optimum distance for phase retrieval. Some authors discussed the estimate of optimum distance for phase retrieval for a sinusoidal

wavefront with small amplitude (p–v $\ll 1$ wave) [8]. According to this, the optimum defocus distance Δz for a focusing beam with the focal length F is

$$\Delta z = \pm 2\lambda \frac{F^2}{\Lambda^2}, \tag{11}$$

where Λ is the period of the wavefront or can be considered as the minimum feature size of the wavefront. Because of the small amplitude constraint, this criterion does not always work.

We derive an optimum distance criterion distinguishing cases whether it is dominated by diffraction effect or geometrical focusing. The crossing between the two regimes is often indicated by the Fresnel number $N_F = \frac{\Lambda^2}{\lambda L}$, where L is the beam propagation distance and Λ is the feature size we are interested in. A large Fresnel number indicates that the beam propagates according to the geometrical rule, whereas a small number indicates that the diffraction effect is more dominant. An alternative way to look at this is to determine whether the Rayleigh distance $z_R = \Lambda^2/\lambda$ is larger or smaller than the beam propagation distance L. Beam propagation within the Rayleigh distance is considered geometrical.

We will first estimate the optimum distance based on the geometrical optics and compare this distance against the Rayleigh distance. If the geometrical optimum distance is shorter than the Rayleigh distance, we can use the geometrical estimation as the optimum. Otherwise the optimum distance is about the same as the Rayleigh distance. The geometrical optimum distance is estimated as follows.

Suppose a converging beam with an unknown radius of curvature R. When we can find a plane at a distance L where the beam is focusing, we can easily infer that the beam has a radius of curvature $R = L$. In this sense the optimum distance to quantify the converging nature of the beam is the focal length. We can apply a similar concept to find out the optimum distance or local focal length associated with the local radius of curvature as shown in Fig. 3(a). The local curvature can be calculated using the Laplacian.

$$\nabla^2 W = \frac{1}{\lambda R}, \tag{12}$$

where the wavefront W is in units of waves. The above expression is to be understood as the differential operation being done locally and R is the local radius of curvature. The optimum distance $L = \pm R = \pm 1/(\lambda \nabla^2 W)$. We extend this to a more-complicated wavefront and apply statistical argument. Given a wavefront as shown in Fig. 3(b), the local wavefront curvature is not constant throughout the aperture. Instead it follows the normal distribution in most cases. An example is shown in Fig. 3(c). We can use the standard deviation of the Laplacian as a representative number of the total radius of curvature. In this case the optimum distance $L_{\text{opt}} = \pm 1/(\lambda\sigma)$, where $\sigma := \sqrt{\langle |\nabla^2 W|^2 \rangle - \langle \nabla^2 W \rangle^2}$.

The above discussion is summarized as below:

$$L_{\text{opt}} = \begin{cases} \pm z_R = \pm \Lambda^2/\lambda & \text{for } \sigma < 1/\Lambda^2 \\ \pm 1/(\lambda\sigma) & \text{for } \sigma > 1/\Lambda^2. \end{cases} \tag{13}$$

It can be shown that the first condition is equivalent to Eq. (11) except for the factor of 1/2 when the propagation distance is mapped to a focusing geometry (Sec. 5).

We performed single-plane phase retrieval simulations by varying distances and the spatial frequency of the wavefront. Five cases were simulated where a high-order wavefront of 300- or 600-μm feature size was added to a low-order polynomial wavefront. The peak-to-valley was also varied from 0.5 to 2 waves. The 600-μm, two-wave p–v wavefront had same slopes as a 300-μm with one wave p–v (Fig. 4). Four cases perform decently at around $L = L_{\text{opt}}$. One case with a high-amplitude wavefront (the green line labeled as "Low order + 600 μm, 2 wv") performed best at $L \sim \frac{1}{4} L_{\text{opt}}$. This case corresponds to the diffractive regime and $\frac{1}{4} L_{\text{opt}} \sim z_R$.

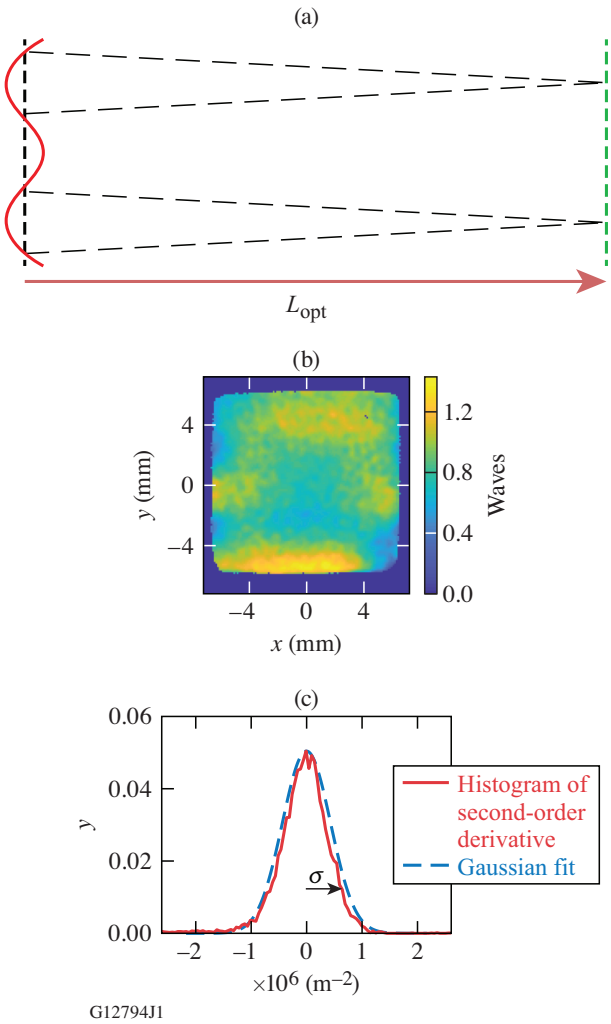

Fig. 3. (a) Geometrical focusing of local curvature, (b) a wavefront example, and (c) histogram of local wavefront curvature

5. Experimental demonstration

We experimentally test the performance of the Gaussian mode approach in phase retrieval and demonstrate the validity of the optimum distance discussed above. The experimental setup is shown in Fig. 5(a). A random wavefront is introduced on a spatial light modulator (SLM). The wavefront is imaged to a commercial high-resolution Shack–Hartmann sensor (Imagine Optic, HASO128) through a leaky mirror. The sensor was referenced against a Zygo GPI interferometer. The direct measurement is compared against the phase-retrieval results. The phase retrieval was performed using one near-field measurement at the pupil plane (P0) and two far-field measurements at the planes P1 and P2. The main beam is focused through a 1-m focal length lens that can simulate propagation distance from 0 to infinity. The mapping between the free-space propagation distance L and the corresponding distance L' in the converging beam space with focal length F is given by $L' = L/(1 + L/F)$ as shown in Fig. 6.

Three wavefronts with different spatial frequencies and amplitudes were introduced on the

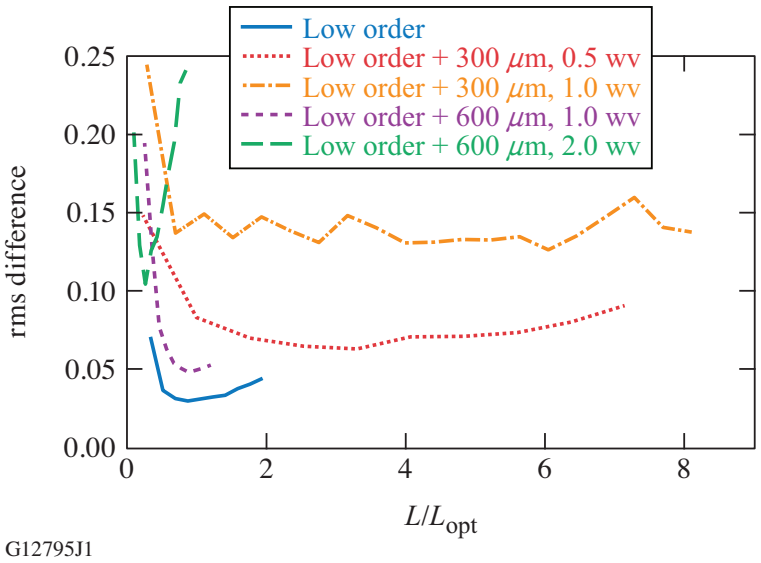

Fig. 4. Optimum distance simulation.

SLM. In the first experiment, the test wavefront consisting of only low-order polynomials was introduced. The optimum propagation distance is estimated to be 3.1 m and the two measurement planes were positioned at 2.5 m and 6.3 m. The secondary distance L_2 was set about 2× longer than L_{opt} to increase sensitivity to low-order terms. The comparison between the retrieved wavefront and the directly measured wavefront using a Shack–Hartmann sensor shows excellent agreement. The rms difference is 0.04 waves. The 2-D maps and lineouts are shown in Fig. 7.

In the second test, high-frequency Fourier modes were added to the low-order wavefront. The optimum distance in this case was estimated to be 2.7 m. The measurement planes are the same as the first experiment. The comparison against the direct measurement is shown in Fig. 8 with the rms difference of 0.05 waves. The lineout comparison on the right shows good agreement of the positions of the small peaks. The number of Gaussian modes was increased from 1467 in the low-order case to 4999.

In the third test, the amplitudes of high-frequency modes were increased while the maximum spatial frequency was reduced. This corresponds to the last case in the simulation where the diffraction effect overtakes the geometrical effect. The optimum distance is the Rayleigh distance, which in this case is 4× shorter than the optimum distance estimated from the geometrical effect. The measurement planes were repositioned at 0.25 and 2.46 m with the geometrical optimum distance being 1.2 m. The resulting retrieved wavefront again shows very good agreement in terms of peak locations and the wavefront amplitude with the direct measurement (Fig. 9). The rms difference is 0.07 waves and the number of modes used in the retrieval is 2777.

The experiments used *a priori* knowledge of the wavefront under test to determine the number of Gaussian modes and the optimum measurement plane locations. Without *a priori* knowledge, the optimum distance can be found by scanning the images along the propagation axis and choosing the plane where caustic-like diffraction occurs. The caustic lines form when the locally curved wavefront collapses, which physically corresponds to the geometrical optimum condition we discussed. The number of Gaussian modes needed also can be determined by analyzing the modulation frequency of the caustic-plane image. The spatial frequency at which these caustic foci occur provides information about the number of modes required for phase retrieval.

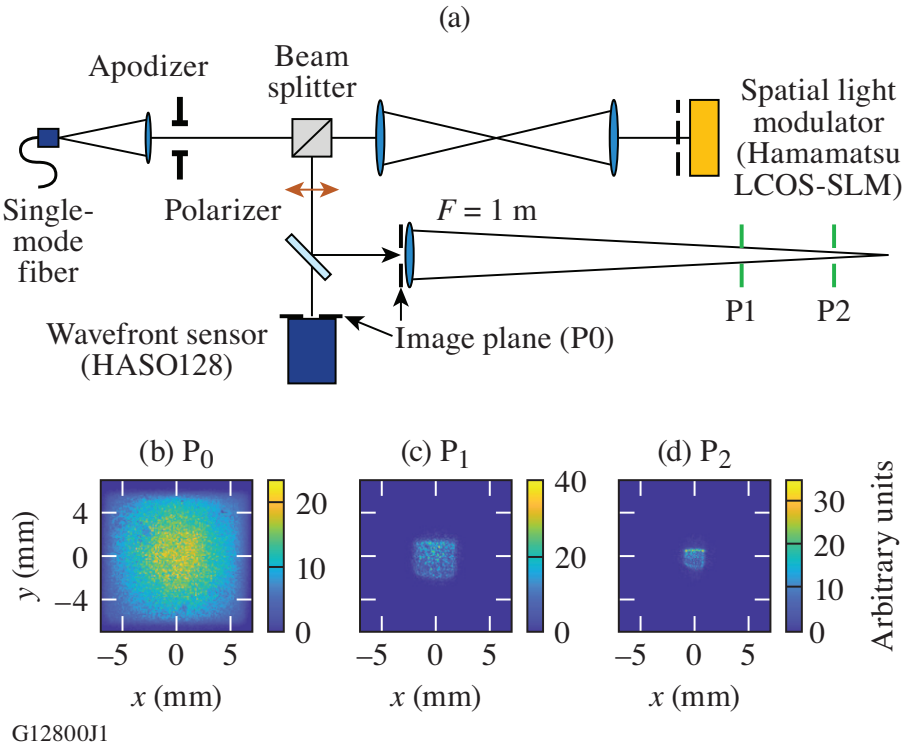

Fig. 5. (a) Experimental setup. The polarizer is used together with the spatial light modulator whose phase retardation works in only one polarization direction. Beam profiles at (b) the near-field plane (P0) and [(c),(d)] two far-field planes (P1 and P2), respectively.

6. Conclusion

We presented a wavefront representation method based on Gaussian basis functions. It is flexible for an arbitrary aperture and can be used to represent from polynomial type wavefronts to high-frequency Fourier modes. The way the Gaussian mode can be used in the nonlinear optimization approach for phase retrieval was discussed. The explicit form of the search gradient was given. The similarity of this approach with the method of sieves was discussed. The determination of optimum distance for measurement required for phase retrieval was discussed on the basis of the behavior of the local curvature. We extended this concept for a more-complicated wavefront using statistical parameters and simplified the expression.

The experiments utilizing the Gaussian basis functions and the optimum distance estimates resulted in the excellent agreement of phase retrieval with direct wavefront measurement. The test wavefronts from low- to high-frequency features were used in the demonstration.

The knowledge of the test wavefront regarding its curvature statistics and dominant spatial frequency terms helps to set the optimum measurement distance and the number of modes. Despite the lack of such information in advance, increasing the number of diverse measurement planes will in general cover most of the frequency bands of the test optic.

The performance of Zernike or Legendre basis functions on phase retrieval is similar to the Gaussian approach in terms of accuracy and speed. For 400 modes, the time it took to converge below 0.04 waves rms was 123 seconds for Gaussian modes and 117 seconds for Legendre modes on 2.4 GHz 8 Core Intel Core i9 processor without using GPU. For 5000 modes, it takes about 5

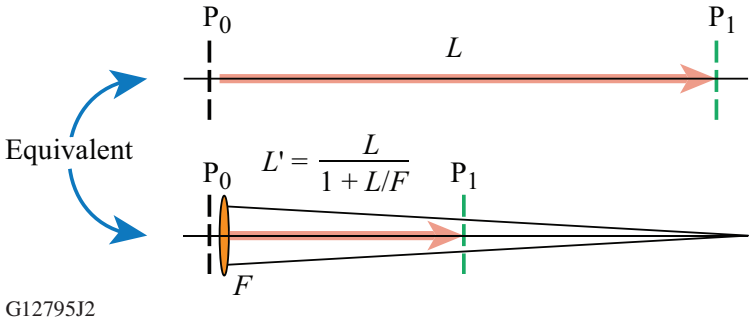

Fig. 6. Distance mapping between the free propagating beam and the converging beam with focal length F.

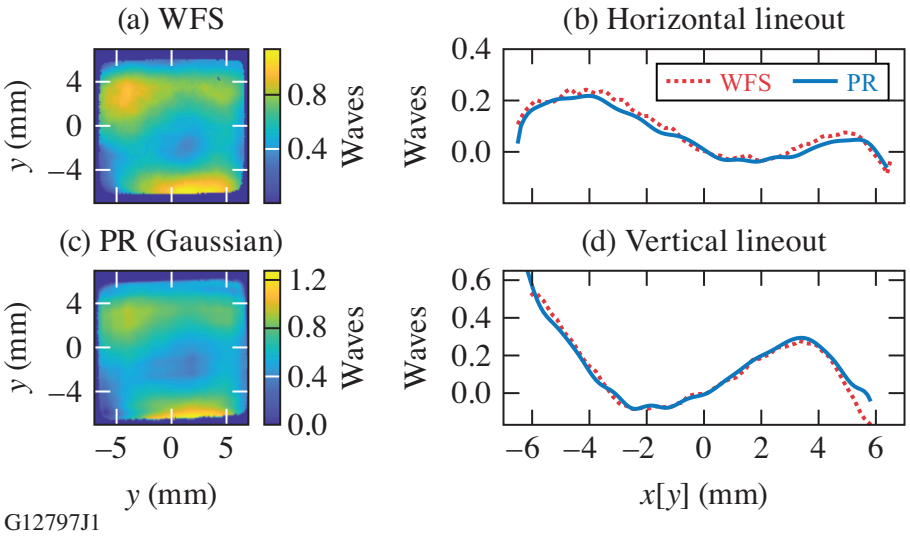

Fig. 7. Wavefront comparison in the low-order wavefront phase-retrieval experiment. WFS: wavefront sensor, PR: phase retrieval. L1=2.5 m, L2=6.3 m, L_{opt}=3.1 m. rms difference = 0.04 waves. # of modes = 1467.

minutes for both. The polynomial basis functions, however, tend to develop high slopes artifact around the beam profile edge when it deviates from the ideal circle or rectangle as the number of modes increases and higher order modes are being used. Therefore the Gaussian approach can be considered more suitable for laser beam characterization whereas the polynomials approach is suitable for metrology applications where the aperture of the optic is well defined.

Funding. National Nuclear Security Administration (DE-NA0003856); University of Rochester; New York State Energy Research and Development Authority.

Acknowledgments. The author thanks J. R. Fienup for helpful comments. This report was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the U.S.

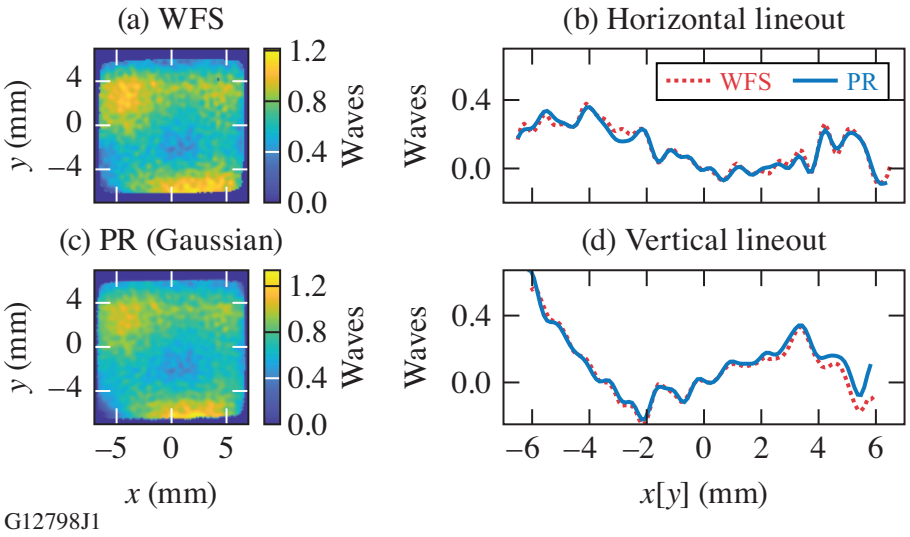

Fig. 8. Wavefront comparison in the low-amplitude high-order wavefront phase-retrieval experiment. The high-order Fourier modes were added to the low-order wavefront. L1=2.5 m, L2=6.3 m, L_{opt}=2.7 m. rms difference = 0.05 waves, # of modes = 4999.

Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are available upon request.

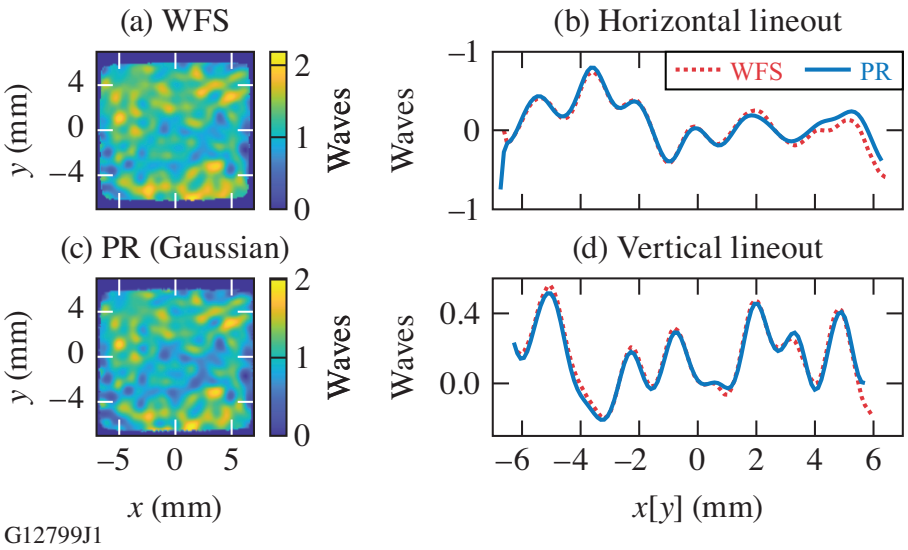

Fig. 9. Wavefront comparison in the large-amplitude, high-order wavefront phase-retrieval experiment. The high-order Fourier modes were added to the low-order wavefront. L1=0.25 m, L2=2.46 m, L_{opt}=1.2 m. rms difference = 0.07 waves, # of modes = 2777.

References

1. J. R. Fienup, "Phase-retrieval algorithms for a complicated optical system," Appl. Opt. **32**, 1737–1746 (1993).
2. C. F. Dunkl, "Orthogonal polynomials on the hexagon," SIAM J. on Appl. Math. **47**, 343–351 (1987).
3. J. A. Díaz and V. N. Mahajan, "Orthonormal aberration polynomials for optical systems with circular and annular sector pupils," Appl. Opt. **52**, 1136–1147 (2013).
4. V. N. Mahajan and G.-m. Dai, "Orthonormal polynomials in wavefront analysis: analytical solution," J. Opt. Soc. Am. A **24**, 2994–3016 (2007).
5. J. A. Marozas, T. J. B. Collins, J. D. Zuegel, P. W. McKenty, D. Cao, S. Fochs, and P. B. Radha, "Continuous distributed phase-plate advances for high-energy laser systems," J. Physics: Conf. Ser. **717**, 012107 (2016).
6. S. T. Thurman, R. T. DeRosa, and J. R. Fienup, "Amplitude metrics for field retrieval with hard-edged and uniformly illuminated apertures," J. Opt. Soc. Am. A **26**, 700–709 (2009).
7. P. J. Piscaer, A. Gupta, O. Soloviev, and M. Verhaegen, "Modal-based phase retrieval using gaussian radial basis functions," J. Opt. Soc. Am. A **35**, 1233–1242 (2018).
8. B. H. Dean and C. W. Bowers, "Diversity selection for phase-diverse phase retrieval," J. Opt. Soc. Am. A **20**, 1490–1504 (2003).