

Phase retrieval with the transport-of-intensity equation. II. Orthogonal series solution for nonuniform illumination

T. E. Gureyev

School of Physics, The University of Melbourne, Parkville, Victoria 3052, Australia, and Division of Forest Products, Commonwealth Scientific and Industrial Research Organization, Clayton, Victoria 3169, Australia

K. A. Nugent

School of Physics, The University of Melbourne, Parkville, Victoria 3052, Australia

Received June 21, 1995; revised manuscript received November 28, 1995; accepted February 7, 1996

In a previous paper [J. Opt. Soc. Am. A **12**, 1932 (1995)] we presented a method for phase recovery with the transport-of-intensity equation by use of a series expansion. Here we develop a different method for the solution of this equation, which allows recovery of the phase in the case of nonuniform illumination. Though also based on the orthogonal series expansion, the new method does not require any separate boundary conditions and can be more easily adjusted for apertures of various shapes. The discussion is primarily for the case of a circular aperture and Zernike polynomials, but we also outline the solution for a rectangular aperture and Fourier harmonics. The latter example may have some substantial advantages, given the availability of the fast Fourier transform. © 1996 Optical Society of America.

1. INTRODUCTION

The problem of optical phase retrieval from intensity measurements plays an important role in many fields of physical research, e.g., optics, electron and x-ray microscopy,¹ crystallography,² diffraction tomography,³ and many others. In these disciplines phase recovery can be used as an essential component of the imaging technique and allows the acquisition of important additional information about the sample. Whereas intensity contrast reflects primarily the distribution of the imaginary part of the complex refraction index, the reconstructed phase provides information about its real part. More recently, phase retrieval has also become a major part of various adaptive optical systems that are being developed in astronomy,⁴ synchrotron x-ray optics,⁵ and ophthalmology.⁶ Here the recovered aberrations of the wave front are compensated with the help of a flexible mirror, resulting in a significant improvement in the imaging quality of the optical system.

Existing noninterferometric methods of phase retrieval, which attempt to recover the phase of an electromagnetic wave on the basis of direct measurements of its intensity, can be subdivided into two major categories according to the conditions of intensity measurements. In the first category, the intensity of a wave field is measured in the far (Fraunhofer) zone, so the complex amplitude of the scalar wave can be considered as a Fourier transform of the amplitude distribution in the object plane. If the latter distribution is bounded by a finite aperture (has a finite support), then its Fourier transform is an analytic function whose phase and intensity depend on each other.⁷ This dependence can be used for recovery of the phase from intensity data. In the second category,

intensity is measured in the Fresnel zone at two adjacent planes orthogonal to the optical axis (Fig. 1). Then the phase on the first of the planes is recovered by use of the information about the evolution of the intensity distribution.⁸ This approach is based on the transport-of-intensity equation (TIE) formalism. It was originally proposed by Teague^{8,9} and later developed by Roddier^{10,11} under the name of wave-front curvature sensing.¹² This approach is used in the present paper.

In a recent paper¹³ we proposed a new method for solution of the TIE in which both the data and the solution are expanded into a series of Zernike polynomials. Such an approach has several important advantages. First, the recovered phase is given in terms of classical Zernike aberrations, which are convenient for many applications.^{14–16} Second, the decomposition of the TIE into Zernike components reduces it to a relatively simple and well-conditioned system of algebraic equations the solution of which are the Zernike coefficients of the phase. We conducted a theoretical analysis of this algebraic system and found exact relations between individual Zernike aberrations and the evolution-of-intensity distribution in the wave front of a paraxial wave. Furthermore, the reduction of the TIE to an algebraic system allows effective methods for its analytical and numerical solution. In Ref. 13 we derived analytical expressions and presented numerical examples of the corresponding matrix and its inverse. It is important to note that the elements of these matrices do not depend on experimental data.

The results of Ref. 13 were obtained under two major assumptions. First, the beam for which the phase is to be reconstructed was required to have a circular cross section; second, illumination was assumed to be uniform within that cross section. These assumptions are often

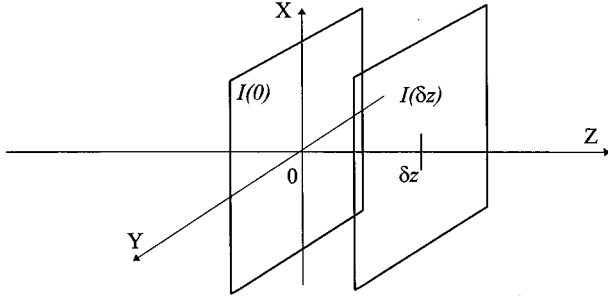


Fig. 1. Phase retrieval can be performed by using the two intensity measurements $I(x, y, 0)$ and $I(x, y, \delta z)$ on adjacent planes $z = 0$ and $z = \delta z$ orthogonal to the optical axis z .

used in adaptive optics.¹⁷ The first assumption, though apparently limiting, is of a technical character. It relates to the fact that the conventional Zernike polynomials are defined in a circular region. The assumption of uniform illumination is of more fundamental importance. It was an essential part of the original wave-front curvature sensing technique of Roddier.^{10,11} However, in many practically important problems in which phase recovery is required, the intensity distribution is not uniform. The new method allows us to remove both of the above assumptions.

The main result of the present paper is a new method for phase retrieval by the TIE, which can be used in the case of nonuniform illumination. We show that this approach not only allows one to deal with nonuniform illumination in a noniterative manner but also removes the major difficulty of some of the previous variants of the wave-front curvature sensing technique, namely, the necessity of distinguishing the boundary phase data from the intensity derivative inside the aperture.¹³ We also establish some important properties of the phase retrieval matrix corresponding to our method. In particular, we prove that this matrix is always invertible. These results provide the basis for the numerical implementation of the method. Its good performance is confirmed by computer simulations.

The outline of the paper is as follows. In Section 2 we describe the main properties of the TIE in the case of nonuniform intensity. We develop our method for phase retrieval by the TIE with use of Zernike polynomials in Subsection 3.A and Fourier harmonics in Subsection 3.B, illustrate the analytical results with several examples in Section 4 and summarize the main features of the method in Section 5. In Appendix A we show how the technique is simplified in the uniform intensity case.

2. TRANSPORT-OF-INTENSITY EQUATION IN THE CASE OF NONUNIFORM ILLUMINATION

In the paraxial (Fresnel) approximation with the optical axis parallel to the z coordinate, the slowly varying component $u(\mathbf{r}) = I^{1/2}(\mathbf{r})\exp[i\varphi(\mathbf{r})]$ of the scalar monochromatic wave $\exp(ikz)u(\mathbf{r})$ satisfies the paraxial equation

$$(2ik\partial_z + \Delta)u(x, y, z) = 0, \quad (1)$$

where $\mathbf{r} = (x, y, z)$, $\partial_z = \partial/\partial z$, $k = 2\pi/\lambda$ is the wave number, and $\Delta = \nabla^2 = \partial_x^2 + \partial_y^2$ is the two-dimensional Laplacian. If intensity is positive everywhere in some area Ω of the plane $z = z_0$, then, in Ω , Eq. (1) is equivalent to the following pair of equations for the intensity and phase^{8,9}:

$$2k\partial_z\varphi = -|\nabla\varphi|^2 + D(I), \quad (2)$$

$$k\partial_z I = -\nabla \cdot (I\nabla\varphi), \quad (3)$$

where $\nabla = (\partial_x, \partial_y)$ is the gradient operator in the plane and $D(I) = I^{-1/2}\Delta(I^{1/2})$. Equation (3) is the TIE. Teague^{8,9} was the first to suggest the use of the TIE for retrieval of the phase φ in Ω if the distributions of intensity and its z derivative are known there. Without loss of generality we can always assume that the domain Ω lies in the plane $z = 0$.

If we consider Ω to be the image of a finite illuminated aperture, then the intensity vanishes outside Ω . Note that according to a well-known property of Eq. (1), the image of a finite aperture cannot be finite; i.e., for an arbitrary large R points (x, y) exist such that $x^2 + y^2 > R^2$ and $I(x, y) > 0$ (here and below we omit the coordinate z from the list of arguments of functions if $z = 0$). However, it is reasonable to assume that if the illuminated aperture in the object plane has a finite size, then intensity values in the image plane will be negligible outside some finite area Ω . Certainly, in general, the intensity may also have zeros at some points inside Ω . At such points, however, the TIE [Eq. (3)] is not valid, and phase may become a multivalued function in Ω .¹⁸ To avoid this sort of complication we assume that Ω is a simply connected domain with smooth boundary $\Gamma = \partial\Omega$ and that the intensity $I(x, y) \equiv I(x, y, 0)$ is a smooth function such that

$$I(x, y) > 0 \quad \text{inside } \Omega, \quad (4)$$

$$I(x, y) \equiv 0 \quad \text{outside } \Omega \text{ and on } \Gamma. \quad (5)$$

Thus to retrieve the phase φ we must solve Eq. (3) with $I(x, y)$ satisfying relations (4) and (5) in area Ω . Let us consider the problem of the existence and uniqueness of the solutions to the problem posed by relations (3)–(5).

We state that the solution to problem (3)–(5) is always unique up to an arbitrary additive constant. Obviously, if φ is a solution to problem (3)–(5), then $\varphi + C$ is also a solution for an arbitrary constant C . Let us prove that if φ_1 and φ_2 are two solutions to problem (3)–(5), then $\varphi_2 = \varphi_1 + C$ with some constant C . If $\tilde{\varphi} = \varphi_2 - \varphi_1$, then

$$\begin{aligned} \nabla \cdot (I\nabla\tilde{\varphi}) &= \nabla \cdot (I\nabla\varphi_2) - \nabla \cdot (I\nabla\varphi_1) \\ &= k\partial_z I - k\partial_z I = 0, \\ 0 &= - \int_{\Omega} \int \tilde{\varphi} \nabla \cdot (I\nabla\tilde{\varphi}) dx dy \\ &= \int_{\Omega} \int I |\nabla\tilde{\varphi}|^2 dx dy - \int_{\Gamma} I \tilde{\varphi} \partial_n \tilde{\varphi} ds. \end{aligned} \quad (6)$$

Because φ_1 and φ_2 are the phases of the slowly varying complex amplitude $u(\mathbf{r})$, the product $I^{1/2}\partial_n \tilde{\varphi}$ is bounded:

$|I^{1/2}\partial_n\tilde{\varphi}| \ll k$. As intensity $I(x, y)$ tends to zero, when (x, y) tends to the boundary Γ , the product $I^{1/2}\tilde{\varphi}$ also tends to zero at the boundary. Therefore the integral over Γ in Eq. (6) is equal to zero. Hence it follows from Eq. (6) and relation (4) that $\nabla\tilde{\varphi} = 0$ in Ω , i.e., $\tilde{\varphi} = \text{constant}$.

We also state that for the phase solution to problem (3)–(5) to exist, the following condition must hold:

$$\int \int_{\Omega} \partial_z I dx dy = 0. \quad (7)$$

To prove Eq. (7) we multiply Eq. (3) by the constant function $E(x, y) \equiv 1$ and integrate by parts over Ω :

$$\begin{aligned} \int \int_{\Omega} k \partial_z I dx dy &= \int \int_{\Omega} E k \partial_z I dx dy \\ &= - \int \int_{\Omega} E \nabla \cdot (I \nabla \varphi) dx dy \\ &= \int \int_{\Omega} I \nabla \varphi \cdot \nabla E dx dy = 0. \end{aligned} \quad (8)$$

Equation (7) is a necessary condition for the solvability of problem (3)–(5). This condition reflects the energy conservation law and can be used to check the consistency of experimentally measured intensity data.

We must emphasize that despite its seeming simplicity, problem (3)–(5) is a difficult one for mathematical analysis.¹⁹ Serious complications arise from the degeneracy of the coefficient $I(x, y)$ at the boundary. Note, for example, that conventional partial differential equations in a bounded domain usually require some boundary conditions for the uniqueness of their solution. In contrast, as we just proved, problem (3)–(5) has no more than one solution (modulo constant phases) in the absence of any classical boundary conditions on the phase solution. This is a rather unusual situation from the point of view of the theory of partial differential equations.

3. SOLUTION OF THE TRANSPORT-OF-INTENSITY EQUATION WITH NONUNIFORM INTENSITY

In this section we construct a solution to the TIE [Eqs. (3)–(5)] by the method of orthogonal expansions.²⁰ Rather than applying the method in its abstract form within an area of an arbitrary shape, using some set of basis functions, we consider a more specific situation of a circular domain and Zernike polynomials. The general ideas of this method can be easily transposed to domains of a different shape with an appropriate orthonormal set of basis functions. An example involving the rectangular domain and Fourier harmonics is presented in Subsection 3.B.

A. Circular Domain and Zernike Polynomials

We do not reproduce here the definition and basic properties of the Zernike polynomials, which can be found in many sources.^{14–16} We will use the definition and notation adopted in Ref. 13, which differ only slightly from those introduced in Ref. 16.

Let Ω be a two-dimensional disk of radius R in the plane $z = 0$ and let (r, θ) be the polar coordinates in Ω . The Zernike polynomials $Z_j(r/R, \theta)$, $0 \leq r \leq R$, $j = 1, 2, 3, \dots$, make up a complete orthonormal set in Ω with respect to the scalar product:

$$\langle f, g \rangle = R^{-2} \int_0^{2\pi} \int_0^R f(r, \theta) g(r, \theta) r dr d\theta. \quad (9)$$

Let us multiply TIE (3) by the Zernike polynomial $Z_j(r/R, \theta)$ and integrate it over Ω :

$$\begin{aligned} -R^{-2} \int_0^{2\pi} \int_0^R \nabla \cdot (I \nabla \varphi) Z_j r dr d\theta \\ = R^{-2} \int_0^{2\pi} \int_0^R F Z_j r dr d\theta, \end{aligned} \quad (10)$$

where $F = k \partial_z I(r, \theta)$. The right-hand side of Eq. (10) is by definition the j th Zernike coefficient $F_j = \langle F, Z_j \rangle$ of the function F . On the left-hand side of Eq. (10) we decompose φ into Zernike terms,

$$\varphi(r, \theta) = \sum_{i=1}^{\infty} \varphi_i Z_i(r/R, \theta) \quad (11)$$

and integrate by parts, taking Eq. (5) into account. The integral over the boundary Γ disappears, and we obtain

$$\sum_{i=1}^{\infty} \varphi_i R^{-2} \int_0^{2\pi} \int_0^R I \nabla Z_i \cdot \nabla Z_j r dr d\theta = F_j. \quad (12)$$

Now it is convenient to introduce the matrix $\mathbf{M} = [M_{ij}]$ with elements

$$\begin{aligned} M_{ij} &= \int_0^{2\pi} \int_0^R I(r, \theta) \nabla Z_i(r/R, \theta) \cdot \nabla Z_j(r/R, \theta) r dr d\theta, \\ i, j &= 1, 2, 3, \dots \end{aligned} \quad (13)$$

Using this definition we can rewrite Eq. (12) as a system of algebraic equations for the unknown Zernike coefficients φ_i of the phase:

$$\sum_{i=1}^{\infty} M_{ij} \varphi_i = R^2 F_j, \quad j = 1, 2, 3, \dots, \quad \text{or} \quad \mathbf{M} \varphi = R^2 F. \quad (14)$$

To retrieve the Zernike coefficients of the phase, it is necessary to solve the algebraic system (14), i.e., to find the inverse matrix $\mathbf{M}^{-1}$. Note that by definition $Z_1 = \text{constant}$. Hence it follows from Eq. (13) that the first row and the first column of $[M_{ij}]$ consist of zeros. This fact has two consequences. First, a solution to Eq. (14) may exist only if $F_1 = 0$. This is a manifestation of the general solvability condition [Eq. (7)] of the TIE derived in Section 2. Second, the first Zernike coefficient φ_1 cannot be found from Eq. (14). This is a manifestation of another general property proved in the Section 2, which states that the phase can be found from the TIE only up to an arbitrary additive constant. Obviously, the solvability of the TIE in this situation is equivalent to the invertibility of matrix $[M_{ij}]$. We consider the finite (truncated) subsystems of Eq. (14) with i and j not exceeding some integer N and prove that the matrices $\mathbf{M}_{(N)} = [M_{ij}]_{(N)}$, $2 \leq i \leq N$, $2 \leq j \leq N$ are invertible.

For a given integer N and a function $g(r, \theta)$ defined in Ω , let us denote by $g_{(N)}$ a finite Zernike expansion

$$g_{(N)}(r, \theta) = \sum_{j=2}^N g_j Z_j(r/R, \theta). \quad (15)$$

Note that the sum in Eq. (15) starts from $j = 2$. The following equation is an $(N - 1)$ -dimensional approximation of Eq. (14):

$$\sum_{i=2}^N M_{ij} \varphi_i = R^2 F_j, \quad j = 2, \dots, N, \quad \text{or} \quad \mathbf{M}_{(N)} \varphi_{(N)} = R^2 \mathbf{F}_{(N)}. \quad (16)$$

Let us show that matrix $\mathbf{M}_{(N)}$ is invertible for an arbitrary integer N . By definition,

$$\begin{aligned} \sum_{j=2}^N \sum_{i=2}^N M_{ij} \varphi_i \varphi_j &= \sum_{j=2}^N \sum_{i=2}^N \varphi_i \varphi_j \int \int_{\Omega} I \nabla Z_i \cdot \nabla Z_j r dr d\theta \\ &= \int \int_{\Omega} I |\nabla \varphi_{(N)}|^2 r dr d\theta. \end{aligned} \quad (17)$$

If $\mathbf{M}_{(N)} \varphi_{(N)} = 0$ for some $\varphi_{(N)}$, then $\nabla \varphi_{(N)} = 0$ and hence $\varphi_{(N)} = 0$, because $\varphi_{(N)}$ is orthogonal to constants by definition (15). Thus we have proved that the null space of matrix $\mathbf{M}_{(N)}$ contains only the zero vector, which means that $\mathbf{M}_{(N)}$ is invertible.

Now we can solve Eq. (16):

$$\begin{aligned} \varphi_i &= R^2 \sum_{j=2}^N M_{ij}^{-1} F_j, \quad i = 2, \dots, N, \quad \text{or} \\ \varphi_{(N)} &= R^2 \mathbf{M}_{(N)}^{-1} \mathbf{F}_{(N)}. \end{aligned} \quad (18)$$

The convergence of the partial phase solutions $\varphi_{(N)}$ is a complex problem. Generally, the functions $\varphi_{(N)}$ may not converge in a pointwise metric; i.e., for a particular point (r, θ) in Ω it may happen that the difference $|\varphi_{(L)}(r, \theta) - \varphi_{(N)}(r, \theta)|$ becomes arbitrarily large when L and N tend to infinity. The sequence $\varphi_{(N)}$ may not converge in the least-squares metric, either. This means that problem (3)–(5) is ill-posed,²¹ which is a common situation for inverse problems.²² The mathematical theory of ill-posed problems is now well developed, and different methods for constructing so-called regularized solutions exist.²¹ In the case of orthogonal series expansions the conventional recipe is to truncate the series at some N that is “not too large.” More specifically, for a given precision δ_F of the right-hand-side function F and δ_I of the intensity $I(r, \theta)$, there exists an optimal number $N_{\text{opt}} = N_{\text{opt}}(\delta_F, \delta_I)$ such that the finite Zernike sum $\varphi_{(N_{\text{opt}})}$ is the best possible approximation for the exact solution φ among all $\varphi_{(N)}$. Furthermore, $|\varphi(r, \theta) - \varphi_{(N_{\text{opt}})}(r, \theta)| < \epsilon(\delta_F, \delta_I)$ in Ω , and ϵ tends to zero when δ_F and δ_I tend to zero. Particular methods for the determination of the optimal truncation number $N_{\text{opt}}(\delta_F, \delta_I)$ as well as other regularization methods can be found in Ref. 21.

B. Rectangular Domain and Fourier Harmonics

Here we give a brief outline of the application of the method developed in Subsection 3.A to the rectangular

domain $\Omega_{ab} = (0, a) \times (0, b)$ and the functions

$$W_{mn}(x, y) = \exp(i2\pi mx/a) \exp(i2\pi ny/b). \quad (19)$$

The Fourier harmonics W_{mn} with integer m and n constitute a complete orthonormal basis in Ω_{ab} with respect to the standard scalar product

$$\langle f, g \rangle_{ab} = \frac{1}{ab} \int_0^a \int_0^b f(x, y) g^*(x, y) dx dy, \quad (20)$$

where the asterisk denotes complex conjugation. Following the scheme of Subsection 3.A, one can reduce TIE (3) in Ω_{ab} to the system of algebraic equations for the unknown Fourier coefficients of the phase:

$$\sum_{i,j} \varphi_{ij} M_{mn}^{ij} = ab F_{mn}, \quad (21)$$

where $\varphi_{ij} = \langle \varphi, W_{ij} \rangle_{ab}$ and $F_{mn} = \langle F, W_{mn} \rangle_{ab}$ are the Fourier coefficients of the phase φ and intensity z derivative $F = k \partial_z I$, respectively,

$$M_{mn}^{ij} = (2\pi)^2 (imb/a + jna/b) \hat{I}_{m-i, n-j}, \quad (22)$$

$\hat{I}_{pq} = \langle I, W_{pq} \rangle_{ab}$ are the Fourier coefficients of the intensity distribution $I(x, y)$ in Ω_{ab} . Note that $[M_{mn}^{ij}]$ is a rectangular matrix with indices $m = 0, \pm 1, \dots, \pm M$, $n = 0, \pm 1, \dots, \pm N$, $i = 0, \pm 1, \dots, \pm I$, $j = 0, \pm 1, \dots, \pm J$. Invertibility of the square matrix $[M_{mn}^{ij}]$ with indices

$$\begin{aligned} i, m &= 0, \pm 1, \dots, \pm M, \quad j, n = 0, \pm 1, \dots, \pm N, \\ m^2 + n^2 &> 0, \quad i^2 + j^2 > 0 \end{aligned} \quad (23)$$

with arbitrary integers M and N can be verified exactly as in Eq. (17). The Fourier coefficients of the phase can be found by

$$\varphi_{ij} = ab \sum_{m,n} [M_{mn}^{ij}]^{-1} F_{mn}, \quad (24)$$

with the range of indices defined in Eq. (23). Algebraic system (24) can be further simplified by taking into account that both the phase φ and intensity z derivative $F = k \partial_z I$ are real functions.

This approach may have some advantages over the Zernike decomposition used in Subsection 3.A. First, the availability of the fast Fourier transform will considerably increase the speed of all calculations. Second, if an object does not display any circular symmetry, it may be better represented by a finite number of its Fourier components than by the same number of Zernike components. Third, some detectors (such as a CCD camera) naturally have a rectangular geometry, and the transformations of the measured intensity distributions into polar coordinates suitable for the Zernike representation may introduce additional errors in the reconstruction.

The above formulas take an especially simple form in the case of uniform intensity such that $I(x, y) \equiv I_0$ in Ω_{ab} and $I(x, y) \equiv 0$ outside Ω_{ab} . This simplification is due to the fact that the Fourier harmonics W_{mn} are the eigenfunctions of the Laplacian in the rectangle Ω_{ab} :

$$M_{mn}^{ij} = (2\pi)^2 I_0 (m^2 b/a + n^2 a/b) \delta_{mi} \delta_{nj}, \quad (25)$$

$$\varphi_{mn} = \frac{(ab)^2}{(2\pi)^2 (m^2 b^2 + n^2 a^2) I_0} F_{mn}, \quad (26)$$

where $m = 0, \pm 1, \dots, \pm M$, $n = 0, \pm 1, \dots, \pm N$, $m^2 + n^2 > 0$. Note that the last formula is clearly a stable one, as the higher-order Fourier harmonics of the intensity z derivative are divided by the increasing numbers $(m^2 b^2 + n^2 a^2)$. This technique of phase retrieval by the Fourier expansion of the TIE will be discussed in detail in the future.

4. NUMERICAL ASPECTS AND EXAMPLES OF PHASE RETRIEVAL

In this section we discuss some scaling properties of the new phase retrieval algorithm developed in the previous sections and present initial results of its numerical testing. These results demonstrate good accuracy and stability of the proposed algorithm on simulated data with and without noise. More comprehensive numerical tests as well as the application of the method to experimental data will be presented elsewhere.

The method for phase retrieval with the TIE that we have developed here is described in essence by formula (18) of Subsection 3.A. According to formula (18), to retrieve $N - 1$ Zernike coefficients φ_i of the phase with indices $i = 2, \dots, N$, one must calculate the $(N - 1) \times (N - 1)$ matrix elements M_{ij} , $i, j = 2, \dots, N$ using Eq. (13), invert matrix $\mathbf{M}_{(N)}$, and multiply it by the vector $R^2 F_{(N)} = R^2 [F_j]$, $j = 2, \dots, N$ of the Zernike coefficients of the z derivative $F = k \partial_z I$.

It is apparent that the calculation of the matrix $[M_{ij}]$ is the most computationally demanding among the operations needed for the retrieval of phase by our method. However, with careful programming we found that we were able to compute the 20×20 matrix $\mathbf{M}_{(21)}$ in approximately 40 s using an Intel-486DX2 66-MHz-based personal computer with the double integrals calculated over a 128×128 grid.

If the intensity distribution $I(r, \theta)$ in the plane $z = 0$ remains unchanged (though possibly nonuniform), then there is no need to recalculate matrix $\mathbf{M}_{(N)}^{-1}$. In such cases the only computations necessary for phase retrieval are those for the Zernike coefficients of the z derivative of intensity and the multiplication of $F_{(N)}$ by $\mathbf{M}_{(N)}^{-1}$. It is in some cases possible to perform these calculations at a speed suitable for adaptive optics applications. If the intensity distribution in the plane $z = 0$ is uniform, matrix $\mathbf{M}_{(N)}^{-1}$ can be exactly calculated analytically. In Tables 1 and 2 we present examples of matrices $\mathbf{M}_{(21)}$ and $\mathbf{M}_{(21)}^{-1}$ for the case of uniform illumination $I(r, \theta) \equiv 1$. The elements of these tables were explicitly calculated analytically before their numerical evaluation. Note that the elements M_{ij} defined in Eq. (13) are dimensionless; hence their values do not depend on the choice of measurement units. One can see that matrix $\mathbf{M}_{(21)}^{-1}$ is very sparse and that its elements generally become smaller as the indices i and j increase. This indicates good stability for phase recovery with this matrix. Examples of computer-simulated retrieval of the Zernike coefficients of phase

with use of $\mathbf{M}_{(21)}^{-1}$ and other matrices are given below. Further peculiarities of the method in the uniform-intensity case are discussed in Appendix A.

Let us consider the scaling of the matrix representation [Eqs. (16)] of the TIE. Zernike coefficients φ_i of the phase and the matrix elements M_{ij} appearing on the left-hand side of Eqs. (16) are dimensionless. On the right-hand side of Eqs. (16) we have the vector $R^2 F_{(N)}$ with components $R^2 F_j = N_F \delta_z I_j$, where $\delta_z I_j$, $j = 2, \dots, N$ are the dimensionless Zernike coefficients of the intensity difference $\delta_z I = I(r, \theta, \delta z) - I(r, \theta, 0)$; δz is the distance between the two planes of intensity measurements; and

$$N_F = \frac{2\pi R^2}{\lambda \delta z}. \quad (27)$$

Now all the scaling factors in the TIE are combined in one dimensionless parameter (27) which has the form of the Fresnel number. The magnitude of N_F determines the stability of the calculation of the right-hand side of Eqs. (16):

$$R^2 F_j = N_F R^{-2} \int_0^{2\pi} \int_0^R [I(r, \theta, \delta z) - I(r, \theta, 0)] Z_j r dr d\theta. \quad (28)$$

The expression inside the square brackets in Eq. (28) can be written as

$$I(\delta z) - I(0) \cong \partial_z I(0) \delta z + \partial_z^2 I(0) (\delta z)^2 / 2 + \sigma(I), \quad (29)$$

where we omitted the (r, θ) arguments in all functions and denoted the noise in the intensity measurements by $\sigma(I)$. Hence if N_F is too large, the noise term in Eq. (29) may be amplified. If, on the other hand, δz is large, then the right-hand side of Eqs. (16) may become contaminated by the higher-order terms in the Taylor expansion of the intensity difference. A method for choosing the optimal distance δz is given in Ref. 8.

The following variant of phase retrieval formula (18) was used in our numerical experiments:

$$\varphi_{(N)} = N_F \mathbf{M}_{(N)}^{-1} \delta_z I_{(N)}, \quad (30)$$

with $\delta_z I = I(r, \theta, \delta z/2) - I(r, \theta, -\delta z/2)$. We applied the half-step difference formula to increase the accuracy of our approximation for $\partial_z I$.

In the subsequent examples we used the following common computational parameters: wavelength $\lambda = 0.5 \mu\text{m}$; intensity in the planes $z = 0, \pm \delta z/2$, $\delta z = 2 \times 10^4 \mu\text{m}$ was calculated in the square regions $|x| \leq 1280 \mu\text{m}$, $|y| \leq 1280 \mu\text{m}$ around the origin of the (x, y) coordinates; grid step sizes were $\delta x = \delta y = 20 \mu\text{m}$; and the radius of the illuminated region was $R = 1010.5 \mu\text{m}$.

In Table 3 we present an example of matrix $\mathbf{M}_{(21)}^{-1}(I_{\text{var}})$ corresponding to the nonuniform intensity distribution $I_{\text{var}} = I_{\text{var}}(r, \theta)$ at $z = 0$, $I_{\text{var}}(r, \theta) = a + \exp[-r^2/(2\sigma^2)]$ in Ω , $a = -0.12974$, $\sigma = 500 \mu\text{m}$. To check that matrix $\mathbf{M}_{(21)}^{-1}(I_{\text{var}})$ is well-conditioned, we calculated its eigenvalues, given in Table 4. Obviously, the (dimensionless) eigenvalues in Table 4 are small, which ensures the stability of the retrieval of the corresponding Zernike coefficients of the phase.

We then simulated the phase distribution at the plane $z = 0$ with particular values of the Zernike aberrations

Table 1. Matrix $M_{(21)}$ in the Case of Uniform Intensity Distribution

[illegible]

Table 3. Matrix $M_{(21)}^{-1}(I_{\text{var}})$

i	j																			
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2	0.8874	0	0	0	0	0	0.0843	0	0	0	0	0	0	0	0.0054	0	0	0	0	0
3	0	0.8874	0	0	0	0.0843	0	0	0	0	0	0	0	0	0	0.0054	0	0	0	0
4	0	0	0.2783	0	0	0	0	0	0	0.0407	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0.5433	0	0	0	0	0	0	0	0.0525	0	0	0	0	0	0	0	0
6	0	0	0	0	0.5433	0	0	0	0	0	0.0525	0	0	0	0	0	0	0	0	0
7	0	0.0843	0	0	0	0.1657	0	0	0	0	0	0	0	0	0	0.0265	0	0	0	0
8	0.0843	0	0	0	0	0	0.1657	0	0	0	0	0	0	0	0	0.0265	0	0	0	0
9	0	0	0	0	0	0	0	0.4072	0	0	0	0	0	0	0	0	0	0.0364	0	0
10	0	0	0	0	0	0	0	0	0.4072	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0.0407	0	0	0	0	0	0	0.0884	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0.0525	0	0	0	0	0	0.1139	0	0	0	0	0	0	0	0	0
13	0	0	0	0.0525	0	0	0	0	0	0	0	0.1139	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0.3204	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0.3204	0	0	0	0	0	0
16	0.0054	0	0	0	0	0	0.0265	0	0	0	0	0	0	0	0.0633	0	0	0	0	0
17	0	0.0054	0	0	0	0.0265	0	0	0	0	0	0	0	0	0	0.0633	0	0	0	0
18	0	0	0	0	0	0	0	0	0.0364	0	0	0	0	0	0	0	0.0922	0	0	0
19	0	0	0	0	0	0	0	0.0364	0	0	0	0	0	0	0	0	0	0.0922	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2720	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.2720

Table 4. Eigenvalues of Matrix $M_{(21)}^{-1}(I_{\text{var}})$ from Table 3

μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9	μ_{10}	μ_{11}	μ_{12}	μ_{13}	μ_{14}	μ_{15}	μ_{16}	μ_{17}	μ_{18}	μ_{19}	μ_{20}
0.8972	0.8972	0.5496	0.5496	0.4113	0.4113	0.3204	0.3204	0.2867	0.272	0.272	0.1627	0.1627	0.1076	0.1076	0.088	0.088	0.08	0.0566	0.0566

Table 5. Original and Retrieved Zernike Coefficients of the Phase Obtained by the Method Developed in the Present Paper

Zernike Coefficient Number	Zernike Coefficients of the Original Phase	Retrieved Zernike Coefficients $I(0)=\text{constant}$	Retrieved Zernike Coefficients, $I(0)=\text{var}$	Retrieved Zernike Coefficients, $I(0)=\text{Var.}$ 3% noise in $I(0)$ and $I(\pm \delta z/2)$	Retrieved Zernike Coefficients, $I(0)=\text{Var.}$ Retrieval Algorithm with Constant $I(0)$
2	2.0	1.97184	1.999674	2.056881	2.458137
3	-3.0	-2.986262	-2.999917	-2.957571	-3.68747
4	1.0	-0.9992586	-0.9997382	1.092017	0.9089373
5	0	1.771131×10^{-5}	-1.517023×10^{-5}	-1.003192×10^{-2}	1.638103×10^{-5}
6	0	8.157735×10^{-5}	-1.476062×10^{-5}	7.149399×10^{-3}	1.61842×10^{-5}
7	0	2.762491×10^{-3}	-5.597042×10^{-5}	2.252985×10^{-2}	0.5827547
8	0	-3.641968×10^{-3}	-1.139274×10^{-4}	5.816795×10^{-3}	-0.3885038
9	0.5	0.5005603	0.5005359	0.4705202	0.5354175
10	0	-1.080498×10^{-2}	-2.513457×10^{-4}	-0.0222681	-7.608061×10^{-5}
11	0	-1.230595×10^{-2}	-1.298638×10^{-4}	-2.760892×10^{-2}	-0.2846218
12	0	-2.223438×10^{-6}	-3.37451×10^{-5}	-1.584348×10^{-2}	-1.25583×10^{-5}
13	0	-1.522851×10^{-5}	-3.371672×10^{-5}	0.0156432	-1.263799×10^{-5}
14	0	-2.776338×10^{-3}	-8.833348×10^{-5}	3.096427×10^{-2}	-1.486509×10^{-5}
15	0	1.073869×10^{-5}	2.805307×10^{-5}	-5.569787×10^{-3}	7.547897×10^{-6}
16	0	-1.746409×10^{-3}	1.496716×10^{-5}	-0.0175551	5.254851×10^{-2}
17	0	3.501717×10^{-3}	-1.647772×10^{-4}	8.886978×10^{-3}	-7.883243×10^{-2}
18	0	9.588387×10^{-4}	-9.141329×10^{-5}	8.625328×10^{-3}	-6.605975×10^{-7}
19	-0.5	-0.4804059	-0.4995432	-0.4844857	-0.2332728
20	0	9.337312×10^{-4}	5.236415×10^{-5}	-9.081652×10^{-3}	1.520897×10^{-5}
21	0.1	0.1004306	0.1000526	9.586543×10^{-2}	2.069442×10^{-2}

given in the second column of Table 5. The contour map of this phase is given in Fig. 2. The initial distribution of the complex amplitude at $z = 0$ was constructed with use of this phase and either uniform $I(r, \theta) \equiv 1$ or variable $I_{\text{var}} = I_{\text{var}}(r, \theta)$ intensity distribution defined above. We then calculated the intensity distributions at $z = \pm \delta z/2$, $\delta z = 2 \times 10^4 \mu\text{m}$, evaluating the Kirchhoff integrals, and found the difference $\delta_z I = I(r, \theta, \delta z/2) - I(r, \theta, -\delta z/2)$. At the next step the Zernike coefficients

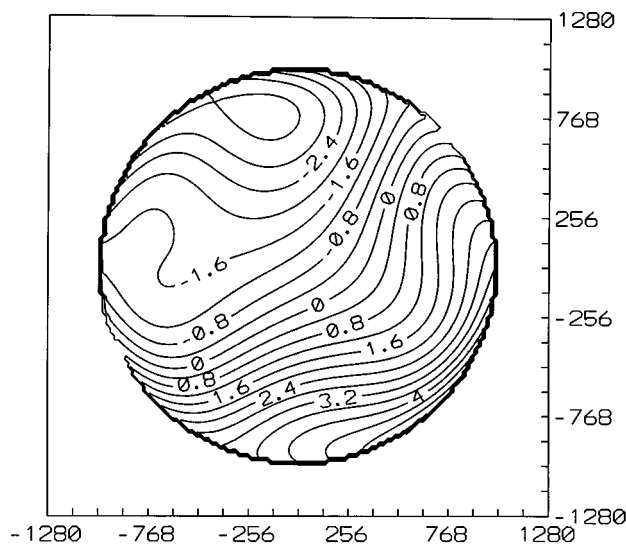


Fig. 2. Original distribution of phase in the plane $z = 0$ with the Zernike coefficients from the second column of Table 5.

of $\delta_z I_j, j = 2, \dots, 21$ were obtained. Then matrices $\mathbf{M}_{(21)}^{-1}(1)$ and $\mathbf{M}_{(21)}^{-1}(I_{\text{var}})$, shown in Tables 2 and 3, respectively, were applied to the product of N_F and the appropriate vector of the Zernike coefficients in accordance with formula (30). The vectors of the retrieved Zernike coefficients of the phase together with the Zernike coefficients of the original phase are presented in Table 5. The fifth column of Table 5 shows the retrieved coefficients of the phase when the pseudorandom Gaussian noise with zero mean and $I/\sqrt{N_c}$ variance, $N_c = 10^3$, was added to the intensity distributions in the planes $z = 0, \pm \delta z/2$. The contour map of the phase synthesized with these Zernike coefficients and its difference from the original one are presented in Figs. 3 and 4, respectively. These results demonstrate the high precision and stability of our method. The last column in Table 5 contains the retrieved Zernike coefficients obtained with the uniform intensity algorithm of Appendix A when the intensity was in fact nonuniform, $I(r, \theta, 0) = I_{\text{var}}$. The accuracy of the phase retrieval is low, which shows that intensity variation cannot be neglected.

In the second example, we simulated an initial phase distribution $\varphi(r, \theta, z = 0)$, which cannot be exactly represented as a linear combination of the first 21 Zernike polynomials. As a consequence, the phase $\varphi(r, \theta, z = 0)$ shown in Fig. 5 is different from its projection $\varphi_{21}(r, \theta, z = 0)$ (Fig. 6) on the space spanned by the first 21 Zernike polynomials. We considered this example to verify that the residual error $\varphi - \varphi_{21}$ is not amplified in the reconstruction performed in accordance with the proposed algorithm. In order to prove that, we calculated the intensity distributions at $z = \pm \delta z/2$, $\delta z = 2 \times 10^4 \mu\text{m}$, evaluating the Kirchhoff integrals with the initial

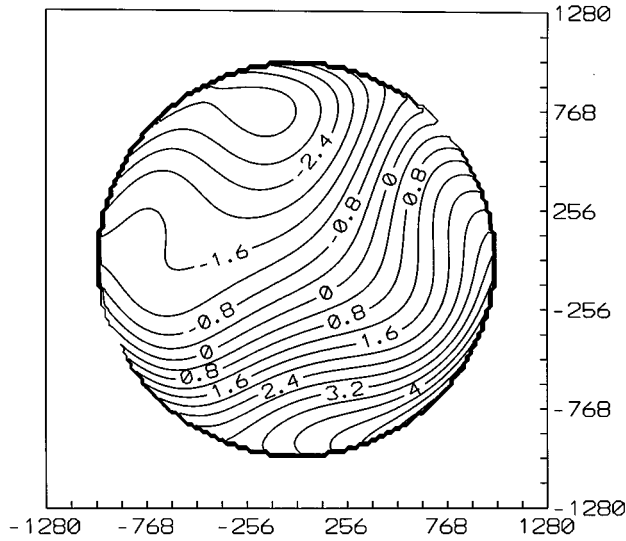


Fig. 3. Reconstructed phase in the plane $z = 0$ with the Zernike coefficients from the fifth column of Table 5.

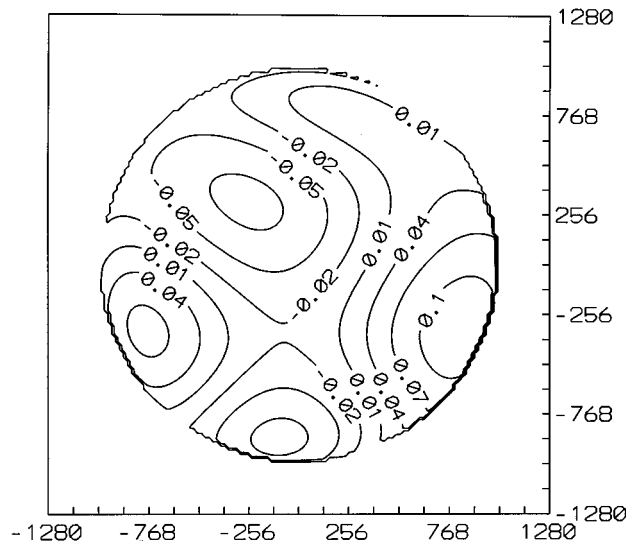


Fig. 4. Difference between the original phase from Fig. 2 and the reconstructed phase from Fig. 3.

complex amplitude $\exp[i\varphi(r, \theta, z = 0)]$, and found the difference $\delta_z I = I(r, \theta, \delta z/2) - I(r, \theta, -\delta z/2)$. We then applied phase reconstruction formula (30) with matrix $\mathbf{M}_{(21)}^{-1}(1)$ from Table 2 to retrieve the Zernike coefficients with $j = 2, \dots, 21$ of the original phase. These coefficients are given in the third column of Table 6, and the second column of this table contains the Zernike coefficients of the original phase $\varphi(r, \theta, z = 0)$. One can see that the values in these two columns are very close to each other, the mean square error

$$\Delta_2 = \left(\sum_{j=2}^{21} \left| \varphi_j^{\text{orig.}} - \varphi_j^{\text{reconst.}} \right|^2 \right)^{1/2} \left/ \left(\sum_{j=2}^{21} \left| \varphi_j^{\text{orig.}} \right|^2 \right)^{1/2} \right.$$

being equal to 8.29×10^{-2} . This result confirms the stability of our algorithm with respect to the error caused

by the presence of higher-order Zernike components in the original phase. We also added the pseudorandom Gaussian noise with zero mean and $I/\sqrt{N_c}$ variance, $N_c = 10^4, 10^3$, and 400 to the intensity distributions in the planes $z = \pm \delta z/2$. The Zernike coefficients with $j = 2, \dots, 21$ of the phase reconstructed from these noise-contaminated intensities are presented in the last three columns of Table 6. Again, our reconstruction algorithm shows good stability with respect to noise, with the error Δ_2 being equal to 0.11, 0.21, and 0.28, respectively.

5. SUMMARY

We presented a new approach for phase retrieval using the TIE and based on the method of orthogonal expan-

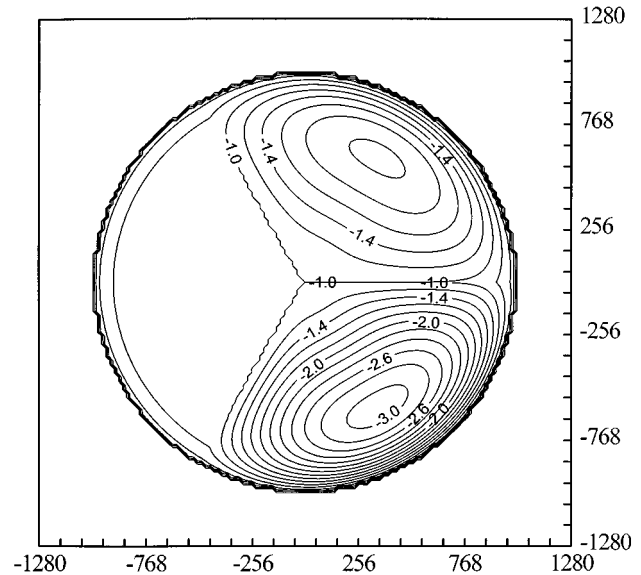


Fig. 5. Original phase distribution $\varphi(r, \theta, z = 0)$ used in the second example.

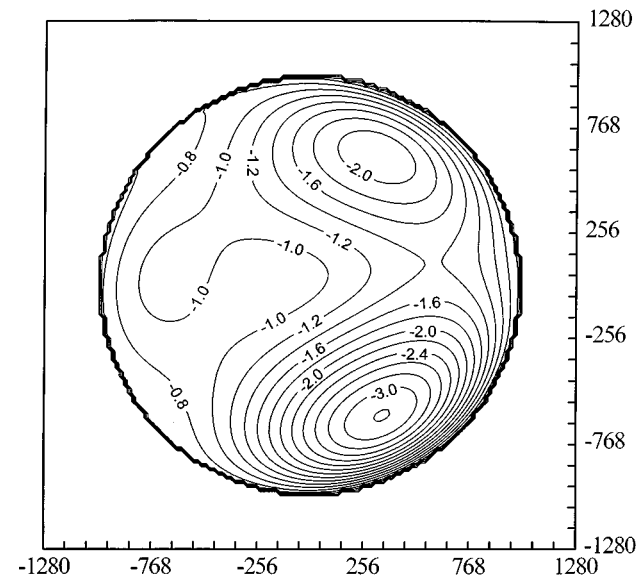


Fig. 6. Projection of the original phase distribution $\varphi(r, \theta, z = 0)$ used in the second example onto the linear space spanned by the first 21 Zernike polynomials.

Table 6. Original and Retrieved Zernike Coefficients of the Phase from the Second Example

Zernike Coefficient Number	Zernike Coefficients of the Original Phase	Retrieved Zernike Coefficients, No Noise	Retrieved Zernike Coefficients, 1% Noise in $I(\pm \delta z/2)$	Retrieved Zernike Coefficients, 3% Noise in $I(\pm \delta z/2)$	Retrieved Zernike Coefficients, 5% Noise in $I(\pm \delta z/2)$
2	-0.5151554	-0.513874	-0.4615785	-0.5483311	-0.3370395
3	0.2974385	0.2974176	0.3289021	0.2172351	0.2174085
4	0.1830336	0.1807383	0.1759277	0.0253246	4.693471×10^{-2}
5	0.1853739	0.1848741	0.1932727	4.417339×10^{-2}	0.1705668
6	0.3210833	0.3222703	0.3397285	0.385162	0.3138469
7	-0.1402221	-0.1391483	-0.1365951	-0.1602742	-9.774832×10^{-2}
8	0.2428355	0.240916	0.2356524	0.2611011	0.2014243
9	5.096811×10^{-6}	-7.777215×10^{-4}	-1.939028×10^{-3}	-1.506611×10^{-2}	5.610845×10^{-2}
10	0.273345	0.272554	0.2611552	0.2257029	0.1976958
11	0.4939752	0.5306388	0.5334223	0.50519	0.5823827
12	-0.2487529	-0.2407488	-0.243799	-0.2587623	-0.2435496
13	-0.1435919	-0.1390933	-0.1413112	-0.1294015	-0.1495548
14	1.769599×10^{-2}	-7.333863×10^{-3}	-1.529732×10^{-2}	-1.363889×10^{-2}	5.605185×10^{-3}
15	-1.052375×10^{-2}	-4.925628×10^{-4}	2.42282×10^{-3}	-1.915842×10^{-2}	1.076338×10^{-2}
16	0.1171886	8.999977×10^{-2}	9.020595×10^{-2}	8.307607×10^{-2}	9.308077×10^{-2}
17	-6.765875×10^{-2}	-5.186649×10^{-2}	-5.489764×10^{-2}	-3.411375×10^{-2}	-6.750885×10^{-2}
18	-0.2739193	-0.2170425	-0.212617	-0.2080416	-0.2000826
19	6.296654×10^{-6}	-2.308932×10^{-4}	-1.324174×10^{-4}	1.98962×10^{-3}	-0.0208768
20	2.813452×10^{-2}	-2.554167×10^{-3}	4.843787×10^{-4}	-3.370615×10^{-4}	-0.0120572
21	0.0162313	1.588301×10^{-3}	2.592367×10^{-4}	8.556461×10^{-3}	-8.649234×10^{-3}

sions. The new technique has two important features. First, it allows recovery of the phase in an area of a plane orthogonal to the optical axis when the intensity distribution is not uniform in that area. Second, it does not require boundary phase data or the values of the wave-front boundary slopes for phase retrieval. We proved that if the intensity is positive in a given area of a plane and vanishes outside it, then the phase solution to the TIE is unique up to an additive constant in the absence of any boundary conditions.

We then presented an algorithm for the construction of truncated orthogonal series approximations $\varphi_{(N)}$ for the phase. A truncated series solution $\varphi_{(N)}$ in the plane $z = z_0$ is obtained as a product of the phase retrieval matrix $N_F \mathbf{M}_{(N)}^{-1}$ and the orthogonal series expansion $\delta_z I_{(N)}$ of the intensity difference $\delta_z I(z_0) = I(z_0 + \delta z) - I(z_0)$ between two adjacent planes $z = z_0$ and $z = z_0 + \delta z$. The scaling factor $N_F = kR^2/\delta z$ appearing in the phase retrieval matrix has the form of the Fresnel number. It depends on the wavelength $\lambda = 2\pi/k$, the radius R of the illuminated area in the plane $z = z_0$, and the distance δz between the two planes of intensity measurements. The dimensionless matrix $\mathbf{M}_{(N)}$ depends only on the intensity distribution in a given area of the plane $z = z_0$ and the elements of the orthogonal basis over which the phase expansion is performed. We gave analytical expressions for the elements of matrix $\mathbf{M}_{(N)}$ corresponding to a circular area and the Zernike polynomials. In Subsection 3.B we also presented matrix $\mathbf{M}_{(N)}$ in the case of a rectangular aperture and Fourier harmonics. We proved that the inverse matrix $\mathbf{M}_{(N)}^{-1}$ always exists. Some numerical examples of matrices $\mathbf{M}_{(N)}$ and $\mathbf{M}_{(N)}^{-1}$ are given in Tables 1–3. We also considered the stability of the phase retrieval, which depends on the magnitude of N_F as well as on the eigenvalues of $\mathbf{M}_{(N)}^{-1}$. It turned out that the maxi-

mal eigenvalue of $\mathbf{M}_{(N)}^{-1}$ may increase with N , so the optimal truncation number N_{opt} for $\varphi_{(N)}$ must be chosen depending on the noise level in intensity measurements. The results of numerical simulations performed with the proposed method demonstrated good accuracy and stability for phase retrieval from intensity data.

APPENDIX A

In this appendix we elaborate the proposed method for phase recovery in the uniform intensity case and compare it with the approach developed in Ref. 13.

Let us consider a particular intensity distribution in the plane $z = 0$:

$$I(r, \theta) = I_0 H(R - r), \quad I_0 = \text{const.}, \quad H(t) = \begin{cases} 1, & t > 0 \\ 0, & t < 0 \end{cases} \quad (\text{A1})$$

Intensity $I(r, \theta)$ defined by Eqs. (A1) is uniform inside the circle Ω of radius R , vanishes outside Ω , and has a step-like discontinuity at the boundary $\Gamma = \partial\Omega$. This is the form of intensity distribution considered in Roddier's wave-front curvature sensing technique. Note that in our development in Sections 2 and 3 we assumed intensity $I(r, \theta)$ to be a smooth function. Therefore the discontinuous intensity [Eq. (A1)] should be understood as a limit of a sequence of smooth functions; e.g.,

$$I_n(r, \theta) = I_0 H_n(R - r), \quad I_0 = \text{const.} \quad (\text{A2})$$

$H_n(t)$ is a sequence of smooth functions converging to $H(t)$, $H_n(t) = 1$ when $t \geq 0$, $H_n(t) = 0$ when $t \leq (-1/n)$, and $0 < H_n(t) < 1$ when $(-1/n) < t < 0$. The functions $I_n(r, \theta)$ have finite but increasing gradients near the boundary Γ when $n \rightarrow \infty$. TIE (3) with intensity distribution (A2) acquires the form

$$kI_0^{-1}\partial_z I_n = -H_n(R-r)\Delta\varphi + \delta_n(R-r)\partial_r\varphi(R,\theta), \quad (\text{A3})$$

where $\delta_n(r) = \partial_r H_n(r)$ are smooth functions converging to the Dirac delta function $\delta(r)$. Let us multiply Eq. (A3) by $Z_j(r/R, \theta)$ and integrate it over the circle Ω_n of radius $R_n = R + (1/n)$:

$$\begin{aligned} & -R_n^{-2} \int_0^{2\pi} \int_0^{R_n} H_n(R-r)\Delta\varphi Z_j r dr d\theta \\ & + R_n^{-2} \int_0^{2\pi} \int_0^{R_n} \delta_n(R-r)\partial_r\varphi(R,\theta) Z_j r dr d\theta = F_j^n, \end{aligned} \quad (\text{A4})$$

$$F_j^n = kI_0^{-1}R_n^{-2} \int_0^{2\pi} \int_0^{R_n} \partial_z I_n Z_j r dr d\theta. \quad (\text{A5})$$

Now we integrate the first integral in Eq. (A4) by parts, taking into account that $\partial_r[H_n(R-r)Z_j] = -\delta_n(R-r)Z_j + H_n(R-r)\partial_r Z_j$. Then the second integral in Eq. (A4) cancels, and we obtain

$$\begin{aligned} & -R_n^{-1} \int_0^{2\pi} H_n(R-R_n)\partial_r\varphi Z_j d\theta + R_n^{-2} \\ & \times \int_0^{2\pi} \int_0^{R_n} H_n(R-r)\nabla\varphi \cdot \nabla Z_j r dr d\theta = F_j^n. \end{aligned} \quad (\text{A6})$$

The first integral in Eq. (A6) is equal to zero, because $H_n(R-R_n) = H_n(-1/n) = 0$. If we now substitute truncated Zernike expansion (15) of φ into Eq. (A6), we will obtain the algebraic system analogous to Eqs. (16) for the unknown Zernike coefficients φ_i of the phase:

$$\sum_{i=2}^N M_{ij}^n \varphi_i = R_n^2 F_j^n, \quad j = 2, \dots, N, \quad (\text{A7})$$

with

$$M_{ij}^n = \int_0^{2\pi} \int_0^{R_n} H_n(R-r)\nabla Z_i(r/R, \theta) \cdot \nabla Z_j(r/R, \theta) r dr d\theta, \quad i, j = 2, \dots, N. \quad (\text{A8})$$

Obviously, with $n \rightarrow \infty$ the matrix elements M_{ij}^n converge to

$$M_{ij}^\infty = \int_0^{2\pi} \int_0^R \nabla Z_i(r/R, \theta) \cdot \nabla Z_j(r/R, \theta) r dr d\theta, \quad (\text{A9})$$

and system (A7) transforms into

$$\sum_{i=2}^N M_{ij}^\infty \varphi_i = R^2 F_j^\infty, \quad j = 2, \dots, N, \quad (\text{A10})$$

$$F_j^\infty = kI_0^{-1}R^{-2} \int_0^{2\pi} \int_0^R \partial_z I(r, \theta) Z_j r dr d\theta. \quad (\text{A11})$$

The matrix $\mathbf{M}_{(N)} = [M_{ij}^\infty]_{(N)}$ does not depend on experimental data. It corresponds to matrix (13) in the case of uniform intensity distribution (A1). Matrix $\mathbf{M}_{(N)}$ is invertible, which can be proved exactly as in Eq. (17). The approximate phase solution $\varphi_{(N)}$ can be obtained by the formula

$$\varphi_i = R^2 \sum_{j=2}^N [M_{ij}^\infty]^{-1} F_j, \quad i = 2, \dots, N, \quad \text{or}$$

$$\varphi_{(N)} = R^2 \mathbf{M}_{(N)}^{-1} \mathbf{F}_{(N)}^\infty. \quad (\text{A12})$$

Examples of matrices $\mathbf{M}_{(N)}$ and $\mathbf{M}_{(N)}^{-1}$ for $N = 21$ are given in Tables 1 and 2.

It is important to note that phase retrieval matrix $\mathbf{M}_{(N)}^{-1}$ does not coincide with the phase retrieval matrix obtained in Ref. 13. In Ref. 13 we had to treat separately the intensity z derivative inside Ω and the delta-function-type part of that derivative localized near the boundary. These two terms can be clearly seen on the right-hand side of Eq. (A3). The first of these terms is proportional to the wave-front curvature inside Ω , and the second is proportional to the wave-front slopes at the boundary. Apparently, the left-hand side of Eq. (A3) must have the same structure. If we can represent it in the limit of $n \rightarrow \infty$ as a sum

$$kI_0^{-1}\partial_z I = f(r, \theta) + \delta(R-r)\psi(\theta), \quad (\text{A13})$$

with a smooth function $f(r, \theta)$ in Ω and a smooth function $\psi(\theta)$ on Γ , then Eqs. (A3) and (A13) imply that $-\Delta\varphi = f$ in Ω and $\partial_r\varphi = \psi$ on Γ . This is the classical Neumann boundary-value problem for the Poisson equation. It can be solved for the phase by the Zernike decomposition method as in Ref. 13, with the phase retrieval matrix consisting of separate blocks corresponding to f and ψ . The major difficulty of such an approach is the implementation of decomposition (A13), because the boundary delta functions are not easily measurable experimentally. Moreover, f is also nonzero in the vicinity of the boundary, which makes the problem of the separation of f and ψ difficult and unstable. In our present approach we have overcome this difficulty, as the intensity z derivative $k\partial_z I$ is treated in Eqs. (A12) as one entity. This is a major advantage of the proposed method in the case of a uniform intensity distribution.

ACKNOWLEDGMENTS

This research was supported by the Australian Research Council. The authors acknowledge helpful discussions with A. Roberts.

REFERENCES

1. M. Nieto-Vesperinas, *Scattering and Diffraction in Physical Optics* (Wiley, New York, 1991), Sec. 9.10.
2. R. P. Millane, "Phase retrieval in crystallography and optics," *J. Opt. Soc. Am. A* **7**, 394–411 (1990).
3. T. C. Wedberg and J. J. Stamnes, "Comparison of phase retrieval methods for optical diffraction tomography," *Pure Appl. Opt.* **4**, 39–54 (1995).
4. R. K. Tyson, *Principles of Adaptive Optics* (Academic, San Diego, Calif., 1991).
5. J. Susini, R. Baker, M. Krumrey, W. Schwegle, and Å. Kwick, "Adaptive x-ray mirror prototype: first results," *Rev. Sci. Instrum.* **66**, 2048–2052 (1995).
6. A. W. Dreher, J. F. Bille, and R. N. Weinreb, "Active optical depth resolution improvement of the laser tomographic scanner," *Appl. Opt.* **28**, 804–808 (1989).
7. M. V. Klibanov, P. E. Sacks, and A. V. Tikhonravov, "The phase retrieval problem," *Inverse Probl.* **11**, 1–28 (1995).

8. M. R. Teague, "Deterministic phase retrieval: a Green's function solution," J. Opt. Soc. Am. **73**, 1434–1441 (1983).
9. M. R. Teague, "Irradiance moments: their propagation and use for unique retrieval of phase," J. Opt. Soc. Am. **72**, 1199–1209 (1982).
10. F. Roddier, "Curvature sensing and compensation: a new concept in adaptive optics," Appl. Opt. **27**, 1223–1225 (1988).
11. F. Roddier, "Wavefront sensing and the irradiance transport equation," Appl. Opt. **29**, 1402–1403 (1990).
12. C. Roddier and F. Roddier, "Wave-front reconstruction from defocused images and the testing of ground-based optical telescopes," J. Opt. Soc. Am. A **10**, 2277–2287 (1993).
13. T. E. Gureyev, A. Roberts, and K. A. Nugent, "Phase retrieval with the transport-of-intensity equation: matrix solution with use of Zernike polynomials," J. Opt. Soc. Am. A **12**, 1932–1941 (1995).
14. M. Born and E. Wolf, *Principles of Optics*, 6th ed. (Pergamon, Oxford, 1980), Sec. 9.2.
15. S. N. Bezdid'ko, "The use of Zernike polynomials in optics," Sov. J. Opt. Technol. **41**, 425–429 (1974).
16. R. J. Noll, "Zernike polynomials and atmospheric turbulence," J. Opt. Soc. Am. **66**, 207–211 (1976).
17. P. Hickson, "Wave-front curvature sensing from a single defocused image," J. Opt. Soc. Am. A **11**, 1667–1673 (1994).
18. T. E. Gureyev, A. Roberts, and K. A. Nugent, "Partially coherent fields, the transport-of-intensity equation, and phase uniqueness," J. Opt. Soc. Am. A **12**, 1942–1946 (1995).
19. O. A. Oleinik and E. V. Radkevič, *Second Order Equations with Non-negative Characteristic Form* (Plenum, New York, 1973).
20. L. V. Kantorovich and V. I. Krylov, *Approximate Methods of Higher Analysis* (Wiley, New York, 1964).
21. A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems* (Wiley, New York, 1977).
22. P. C. Sabatier, ed., *Inverse Methods in Action* (Springer, Berlin, 1990).