Iterative zonal wave-front estimation algorithm for optical testing with general-shaped pupils

Weiyao Zou and Jannick P. Rolland

College of Optics and Photonics–CREOL & FPCE, University of Central Florida, Orlando, Florida 32816

Received October 27, 2004; accepted November 16, 2004

An iterative zonal wave-front estimation algorithm for slope or gradient-type data in optical testing acquired with regular or irregular pupil shapes is presented. In the mathematical model proposed, the optical surface, or wave-front shape estimation, which may have any pupil shape or size, shares a predefined wave-front estimation matrix that we establish. Owing to the finite pupil of the instrument, the challenge of wave front shape estimation in optical testing lies in large part in how to properly handle boundary conditions. The solution we propose is an efficient iterative process based on Gerchberg-type iterations. The proposed method is validated with data collected from a 15×15-grid Shack–Hartmann sensor built at the Nanjing Astronomical Instruments Research Center in China. Results show that the rms deviation error of the estimated wave front from the original wave front is less than $\lambda/130$–$\lambda/150$ after ~12 iterations and less than $\lambda/100$ (both for $\lambda=632.8$ nm) after as few as four iterations. Also, a theoretical analysis of algorithm complexity and error propagation is presented.

OCIS codes: 220.4840, 220.4610, 120.6650.

1. INTRODUCTION

Wave-front estimation from wave-front slope data is a classic mathematical problem in optical testing. It converts the wave-front slope data into wave-front optical path differences (OPDs) or, equivalently, wave-front phases defined as the OPDs multiplied by $2\pi/\lambda$. We shall refer to the OPDs as the wave-front values. Let us suppose that the wave-front differences or other type of slope data have been obtained from a slope wave-front sensor. The task is how to estimate the wave-front values from such data, which is a numerical solution to the Neumann boundary problem of the Poisson equation.[1]

Mathematical methods and algorithms for wave-front estimation as it applies to optical testing have been contributed by many authors.[1–14] Of most interest to the work presented in this paper are those algorithms that can handle general pupil shapes. Such algorithms can be categorized into Fourier-transform (FT)-based algorithms and linear least-squares (LS)-based algorithms. For FT-based algorithms, Gerchberg and Saxton proposed in 1972 iterative fast-Fourier-transform (FFT)-based phase retrieval from amplitude measurements in the aperture and the image planes.[15] The Gerchberg–Saxton algorithm was later improved further by Fienup for phase retrieval.[16] Freischlad and Koliopoulos in 1985 proposed a discrete Fourier-transform-based algorithm for model estimation from wave-front slope measurements for square-shaped pupils.[12,13] Later, Freischlad extended this algorithm for general pupil shapes.[14] In Freischlad's method, additional LS matrix equations needed to be set up to generate the missing slope data for extending the general-shaped pupil to a square one. Roddier and Roddier proposed a technique to extrapolate the wave front outside of the pupil employing Gerchberg-type iterations and obtained an FFT-based algorithm for irregular-shaped pupils.[17,18] Finally, Zou and Zhang proposed in 2000 a linear LS-based algorithm to handle any pupil shapes. The algorithm used zero padding of the slope data outside of the sampling pupil, and the discontinuity of the test domain extension introduces remarkable deviation errors (i.e., up to $\lambda/4$ peak to valley). Such errors may not be acceptable for most optical tests.[19]

In this paper we first present in Section 2 the wave-front estimation approach that we have pursued; we briefly review in Section 3 the work done by Zou and Zhang in 2000.[19] We then present in Section 4 the proposed iterative wave-front estimation algorithm followed in Section 5 by two examples of wave-front estimations. Finally, we discuss in Section 6 the theoretical complexity and in Section 7 the noise-error propagation of the proposed algorithm in relation to previous algorithms.

2. MOTIVATION AND PROPOSED APPROACH

Usually, changes in the pupil shape or size will cause changes in the wave-front estimation matrix and require a new setup of the wave-front estimation process for each new size or shape. Thus a mathematical formulation that could automatize the wave-front estimation for any size and shape pupil would not only help in maximizing the efficiency of the testing process but also in minimizing the potential errors necessarily associated with setting up new matrices for wave-front estimation.

While the Zou and Zhang LS algorithm suffered remarkable deviation errors owing to discontinuous boundary conditions, its strength of formulizing a predefined matrix for wave-front estimation provides immediate plug-in of wave-front slope measurements for any pupil shape or size.[18] To illustrate the errors associated with discontinuous boundary conditions and the error propagation within the pupil of interest, we now give an example.

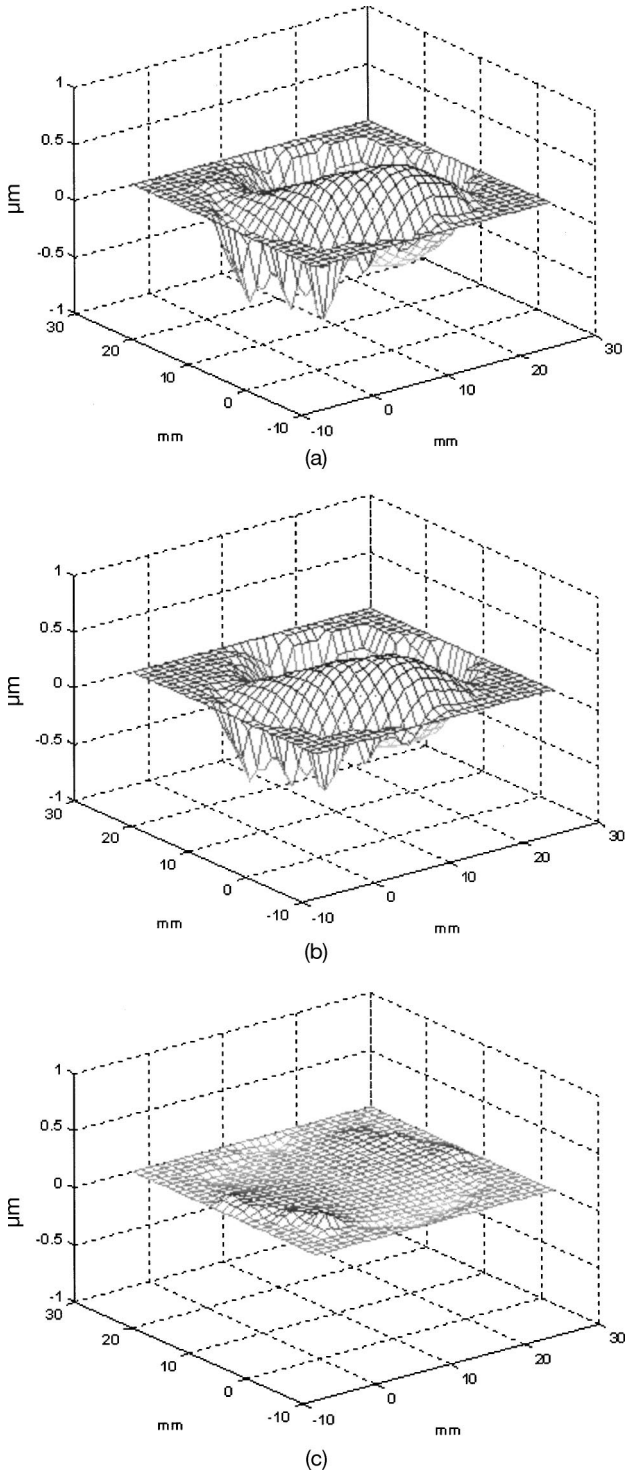

Fig. 1. (a) Ground-truth or original wave front, (b) wave front estimated from measured slope data with the algorithm without iteration, (c) wave-front deviation error computed as the difference between the ground-truth and the estimated wave front.

An original wave front (i.e., an estimated wave front that is considered to represent ground truth of wave-front estimation, as will be further explained in Section 5) is shown in Fig. 1(a), and the estimated wave front from slope data with Zou and Zhang's algorithm is shown in Fig. 1(b). The difference between the estimated and the original wave front is shown in Fig. 1(c), which represents the deviation errors associated with the domain extension made across the boundary. Therefore, while the domain extension technique is quite useful for developing a predefined and regular estimation matrix for any irregular pupil shapes, the challenge lies in how to establish continuity constraints at the boundaries across the original pupil to remove the deviation errors in the estimated wave front.

In this paper we propose to build on the Zou and Zhang[19] algorithm by adding Gerchberg-type iterations for wave-front estimation, which will permit extrapolation of the slope data outside of the pupil in order to satisfy continuity boundary conditions. In the iteration process proposed, the slope data in the pupil under test will be replaced with the sampled raw slope data iteratively until the estimated wave front converges to a solution.

3. REVIEW OF WAVE-FRONT ESTIMATION FROM SLOPE DATA BY ZOU AND ZHANG

Zou and Zhang proposed a noniterative linear LS method to perform wave-front estimation based on a domain extension technique.[19] The proposed algorithm was motivated by the need in practical settings to automatize the estimation process regardless of the pupil shape or size of the input slope data set. Regardless of the method adopted, a sampling geometry had to be first considered in performing wave-front estimation. There are basically three sampling geometries available for wave-front estimates: the Fried geometry,[5] the Hudgin geometry,[6] and the Southwell geometry,[10] illustrated in Fig. 2. Owing to its superiority over other geometries in error propagation,[10] Zou and Zhang adopted the Southwell geometry, which is characterized by taking the wave-front slope measurements and wave-front values estimation at the same nodes. In a problem with discrete slope measurements as a starting point, a two-dimensional array of discrete values w_i ($i=1,2,3,\ldots t\times t$) was used to map the estimated wave-front values. Furthermore, an interlaced array of j nodes was introduced to facilitate the estimations of wave-front slopes at the midpoints between wave-front nodes. Figure 3 shows the geometry in one direction (e.g., the y direction) with both the nodes i and the interlaced nodes j. Let us denote the wave-front slopes at the nodes i in the y and z directions as s_{y_i} and s_{z_i} ($i=1,2,3\ldots t\times t$), respectively. The slope data between two adjacent nodes was assumed to change linearly with distance,[20] which allowed us to perform linear interpola-

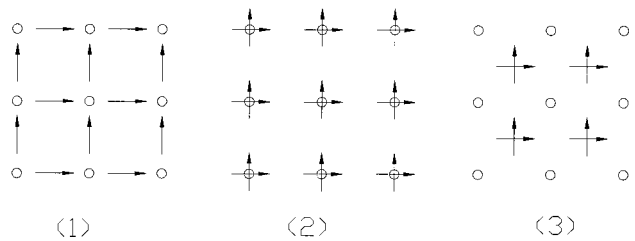

Fig. 2. Wave-front estimation schemes. From left to right: the (1) Hudgin, (2) Southwell, and (3) Fried models. In these figures, the small circles symbolize the wave-front values and the small arrows represent the partial derivatives of the wave front.

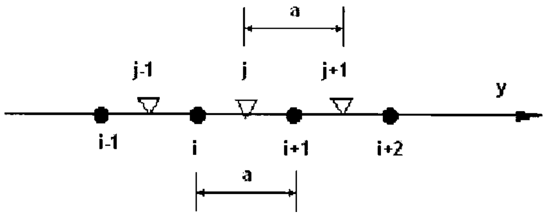

Fig. 3. Double sampling grid systems illustrated in the y direction.

tion to estimate the slopes between nodes. The slope at node j was then estimated as an average of the slopes at nodes i and $i+1$ by

$$s_{y_j}=\frac{1}{2}(s_{y_i}+s_{y_{i+1}}), \tag{1}$$

where the slope s_{yj} could also be expressed as the difference quotient of the wave-front values at nodes i and $i+1$ to their separation a, so that

$$s_{y_j}=\frac{w_{i+1}-w_i}{a}. \tag{2}$$

By combining Eqs. (1) and (2), a relationship between the wave-front slopes and the wave-front values at $i+1$ and i was established as

$$w_{i+1}-w_i=\frac{a}{2}(s_{y_i}+s_{y_{i+1}}). \tag{3}$$

Similarly, in the z direction and accounting for the sign convention shown in Fig. 3,

$$w_i-w_{i+1}=\frac{a}{2}(s_{z_i}+s_{z_{i+1}}). \tag{4}$$

The wave-front estimation model of Zou and Zhang[19] followed these steps:

1. Without loss of generality, the authors supposed that the regular square net had a $t\times t=m$ grid points.
2. Furthermore, the authors supposed that the original sampling domain Ω_0 (i.e. exit pupil, simply connected domain or multiple connected domains) was embedded into a regular square domain Ω_1 that contained the sampling domain Ω_0. Then the square domain Ω_1 was thought of as being composed of two parts: the real part Ω_0 and the imaginary part $\Omega_1\setminus\Omega_0$, both shown in Fig. 4.
3. The grid points in Ω_1 were indexed sequentially from 1 to m row by row (the grid points could also be indexed equivalently column by column as an alternative).
4. The slopes were set to zero in the imaginary part $\Omega_1\setminus\Omega_0$.

From the geometry adopted, Eqs. (3) and (4) were written in matrix form as

$$\begin{bmatrix}
-1 & 1 & & & & & & & & & & \\
 & -1 & 1 & & & & & & & & & \\
 & & \ddots & \ddots & & & & & & & & \\
 & & & -1 & 1 & & & & & & & \\
 & & & & -1 & 1 & & & & & & \\
 & & & & & -1 & 1 & & & & & \\
 & & & & & & \ddots & \ddots & & & & \\
 & & & & & & & -1 & 1 & & & \\
 & & & & & & & & \ddots & \ddots & & \\
 & & & & & & & & & -1 & 1 & \\
 & & & & & & & & & & -1 & 1 \\
 & & & & & & & & & & \ddots & \ddots \\
 & & & & & & & & & & -1 & 1 \\
1 & 0 & 0 & \cdots & 0 & -1 & & & & & & \\
 & 1 & 0 & 0 & \cdots & 0 & -1 & & & & & \\
 & & \ddots & \ddots & \ddots & \ddots & \ddots & & & & & \\
 & & & & \ddots & \ddots & \ddots & & \ddots & \ddots & & \\
 & & & & & & 1 & 0 & 0 & \cdots & 0 & -1 \\
 & & & & & & & 1 & 0 & 0 & \cdots & 0 \quad -1
\end{bmatrix}
\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_t \\ w_{t+1} \\ w_{t+2} \\ \vdots \\ w_{2t} \\ \vdots \\ w_{m-2t+1} \\ w_{m-2t+2} \\ \vdots \\ w_{m-t} \\ w_{m-t+1} \\ \vdots \\ w_{m-1} \\ w_m \end{bmatrix}
\begin{bmatrix} c_{2,1} \\ c_{3,2} \\ \vdots \\ c_{t,t-1} \\ c_{t+2,t+1} \\ \vdots \\ \vdots \\ c_{m-1,m} \\ d_{t+1,t} \\ d_{t+2,t} \\ \vdots \\ d_{2t,t} \\ \vdots \\ d_{i+t,1} \\ \vdots \\ d_{m-1,m-t-1} \\ d_{m,m-t} \end{bmatrix}, \tag{5a}$$

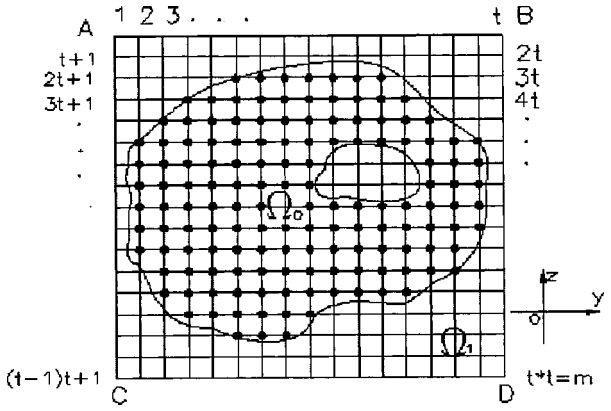

Fig. 4. Domain extension for an irregular-shaped pupil.

or

$$\mathbf{CW} = \mathbf{S}, \tag{5b}$$

where $c_{i+1,i}$ and $d_{i+t,i}$ were defined as

$$c_{i+1,i} = \frac{a}{2}(s_{y_{i+1}} + s_{y_i}), \tag{6a}$$

$$d_{i+t,i} = \frac{a}{2}(s_{z_{i+1}} + s_{z_i}). \tag{6b}$$

The corresponding normal equation set was written as

$$\mathbf{C}^{\mathrm{T}}\mathbf{CW} = \mathbf{C}^{\mathrm{T}}\mathbf{S}. \tag{7}$$

The desired solution region was embedded in a larger square region to obtain the predefined matrix for any pupil shape. After the authors determined the zero point of the estimated wave front, the matrix equation became positive. The authors employed the Cholesky decomposition method in solving the normal matrix.[19] Results were limited by up to $\lambda/4$ peak-to-valley deviation errors, as shown in Fig. 1.

4. WAVE-FRONT ESTIMATION FOR ANY PUPIL SHAPE: A GERCHBERG-TYPE ITERATIVE LINEAR LEAST-SQUARES APPROACH

The method proposed expands on the method by Zou and Zhang[19] by proposing to add an iterative process with the promise to improve the accuracy of the final wave-front estimation within the measured pupil. The merit of the iterative procedure is that it yields minor rms deviation errors while it still shares a predefined estimation matrix for any pupil shape and size. As such, all the matrix coefficients are predetermined and known once and for all. The proposed algorithm detailed hereafter will first require calculating slope data from the estimated wave front in order to enable the iterative process. Such computations will be presented first.

A. Wave-Front Slope Computations

Slope computation from a known wave front could be thought of simply as the inverse problem of wave-front estimation from slope data. In this case, one would inverse the equations established by Zou and Zhang[19] to obtain a matrix equation set for slope computation. However, such a resultant matrix is rank deficient; this property is intrinsically linked to the Southwell geometry chosen for the problem. However, this geometry as described above is optimal in terms of lowest noise propagation, and thus such geometry will be conserved in the algorithm. Therefore additional independent equations will be required for slope extractions from the estimated wave front. We choose to establish such equations on the basis of curvature estimates. We shall first describe the matrix formulation for the y-direction slope computations and then provide the matrix formulation for z-direction slope computation.

1. Wave-Front Slope Computation in the y Direction

For the slope at the midpoint between the nodes i and $i+1$, Eq. (3) may be written as

$$s_{y_{i+1}} + s_{y_i} = e_j, \tag{8}$$

where

$$e_j = \frac{2}{a}(w_{i+1} - w_i), \quad i = 1,2 \ldots m-1, \text{ but } i \neq t, 2t, 3t, \ldots m. \tag{9}$$

In matrix form, Eq. (8) may be written as

$$\mathbf{A}_1\mathbf{S}_y = \mathbf{E}, \tag{10}$$

which is not a full-rank matrix equation set. Curvature-based equations are then considered in order to determine a unique solution for slope computation. The curvature at a midpoint node $j+1$ is proportional to the slope difference between adjacent points $i+1$ and $i+2$. According to Fig. 3, we have

$$s_{y_{i+2}} - s_{y_{i+1}} = f_{j+1}, \tag{11}$$

where

$$f_{j+1} = \left(\frac{w_{i+3} - w_{i+1}}{2a} - \frac{w_{i+2} - w_i}{2a}\right) = \frac{1}{2a}(w_{i+3} - w_{i+2} - w_{i+1} + w_i). \tag{12}$$

If Eq. (12) is divided by the grid separation a, it will actually be a discrete approximation of the wave-front curvature at node $j+1$, which is of $O(a^3)$ precision as shown in Eq. (A12) in Appendix A. In matrix form, Eq. (11) may be expressed as

$$\mathbf{A}_2\mathbf{S}_y = \mathbf{F}. \tag{13}$$

When Eqs. (10) and (13) are combined, a matrix-form equation set may be written as

$$\mathbf{AS}_y = \mathbf{U}, \tag{14}$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix}, \tag{15}$$

$$\mathbf{S}_y = [S_{y1} \quad S_{y2} \quad , \ldots , \quad S_{ym}]^{\mathrm{T}}, \tag{16}$$

$$\mathbf{U} = \begin{bmatrix} \mathbf{E} \\ \mathbf{F} \end{bmatrix}, \tag{17}$$

with

$$\mathbf{A}_1 = \mathrm{diag}[\mathbf{D}_1, \mathbf{D}_1, \ldots, \mathbf{D}_1], \tag{18}$$

$$\mathbf{A}_2 = \mathrm{diag}[\mathbf{D}_2, \mathbf{D}_2, \ldots, \mathbf{D}_2], \tag{19}$$

where

$$\mathbf{D}_1 = \begin{bmatrix} 1 & 1 & & & \\ & 1 & 1 & & \\ & & \ddots & \ddots & \\ & & & 1 & 1 \end{bmatrix}, \tag{20}$$

$$\mathbf{D}_2 = \begin{bmatrix} -1 & 1 & & & & \\ & -1 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & -1 & 1 & \\ & & & & -1 & 1 \end{bmatrix}. \tag{21}$$

Then the normal equation set for the wave-front slope extraction in the y direction can be written as

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{S}_y = \mathbf{A}^{\mathrm{T}}\mathbf{U}. \tag{22}$$

2. Wave-Front Slope Computation in the z Direction

Similarly, the slope-based equations along the z direction are given by

$$s_{z_i} + s_{z_{i+t}} = g_j, \quad i = 1,2, \ldots, m-t, \tag{23}$$

where

$$g_j = \frac{2}{a}(w_i - w_{i+t}). \tag{24}$$

In matrix form, Eq. (23) may be written as

$$\mathbf{B}_1\mathbf{S}_z = \mathbf{G}, \tag{25}$$

which is not a full-rank matrix equation. To get a full-rank equation set, we add the curvature-based equations

$$s_{z_{i+t}} - s_{z_{1+2t}} = h_{j+t}, \tag{26}$$

where

$$h_{j+t} = \frac{1}{2a}(w_i - w_{i+t} - w_{i+2t} + w_{i+3t}), \tag{27}$$

and $i = 1,2, \ldots t; t+1, t+2, \ldots 2t-3, \ldots, m-3t$. The derivation of Eq. (27) is given in Appendix A [i.e., Eq. (A14)]. In matrix form, Eq. (26) becomes

$$\mathbf{B}_2\mathbf{S}_z = \mathbf{H}. \tag{28}$$

Combining Eqs. (25) and (28) in a matrix-form equation set, we obtain

$$\mathbf{B}\mathbf{S}_z = \mathbf{V}, \tag{29}$$

where

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix}, \tag{30}$$

$$\mathbf{S}_z = [S_{z1} \quad S_{z2} \quad \ldots \quad S_{zm}]^{\mathrm{T}}, \tag{31}$$

$$\mathbf{V} = \begin{bmatrix} \mathbf{G} \\ \mathbf{H} \end{bmatrix}, \tag{32}$$

and

$$\mathbf{B}_1 = \begin{bmatrix} \mathbf{I}_t & \mathbf{I}_t & & & \\ & \mathbf{I}_t & \mathbf{I}_t & & \\ & & \ddots & \ddots & \\ & & & \mathbf{I}_t & \mathbf{I}_t \\ & & & & \mathbf{I}_t \end{bmatrix}, \tag{33}$$

$$\mathbf{B}_2 = \begin{bmatrix} \mathbf{I}_t & -\mathbf{I}_t & & & \\ & \mathbf{I}_t & -\mathbf{I}_t & & \\ & & \ddots & \ddots & \\ & & & \mathbf{I}_t & -\mathbf{I}_t \\ & & & & \mathbf{I}_t \end{bmatrix}, \tag{34}$$

$$I_t = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}_{t \times t}. \tag{35}$$

Then the normal equation set for the z-direction slope extraction may be written as

$$\mathbf{B}^{\mathrm{T}}\mathbf{B}\mathbf{S}_z = \mathbf{B}^{\mathrm{T}}\mathbf{V}. \tag{36}$$

B. Gerchberg-Type Iterative Wave-Front Estimation Algorithm

On the basis of the equation sets just established for wave-front slope computations combined with the algorithm of wave-front estimation from slope data given by Zou and Zhang[19] in Section 2, we shall now detail the new iterative LS-based wave-front estimation algorithm illustrated in the flow chart in Fig. 5. With Eq. (7), the wave-front values are first estimated from the slope data in Ω_1. The matrix equation sets given by Eqs. (22) and (36) then serve to compute the y and z slopes in Ω_1 from the estimated wave front. The computed slopes are compared with the original slope data within Ω_0. If the differences are negligible (i.e., less than a termination criterion), the estimation wave front over Ω_1, among which only the wave front within Ω_0 is of interest, is output. Otherwise, the slope data in Ω_0 are replaced with the original measured slope data, while the slope data in the extended

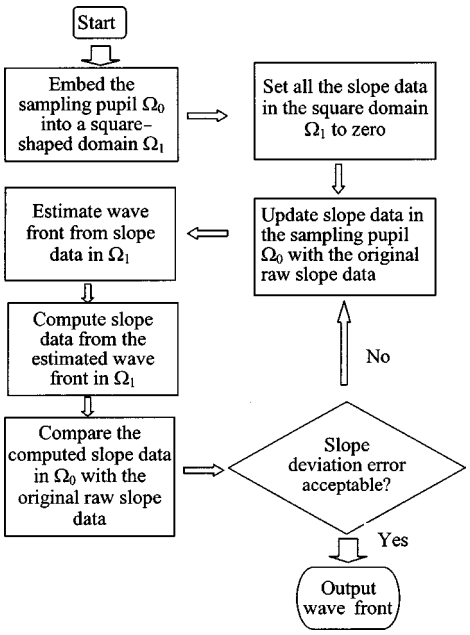

Fig. 5. Flow chart of the Gerchberg-type iterative least-squares wave-front estimation algorithm based on the domain extension technique.

area $\Omega_1 \setminus \Omega_0$ are kept unchanged. The iterative process continues until the established termination criterion is reached.

Such iterative process is referred to as the Gerchberg-type iterations, because the iterative process bears analogy to the Gerchberg–Saxton algorithm, which consists in replacing the computed amplitude of a discrete complex function in the pupil under test with the sampled amplitude across iterations until both amplitude and phase converge to a solution.[15] The iterative algorithm presented in this paper replaces the slope data in the pupil under test with the sampled raw slope data iteratively until the estimated wave front converges to a unique solution. In a fundamental sense, the Gerchberg–Saxton iterations were based on FTs, whereas the algorithm detailed in this paper is based on the linear LS method. The algorithm proposed here bears similarity to the algorithm proposed by Roddier and Roddier in 1991 in the sense that both algorithms use Gerchberg-type iterations to extrapolate the wave front outside the boundary.[17] A basic difference is that Roddier and Roddier's algorithm is based on FFTs instead of the linear LS method.

Without the iterative process, the slope data inside the pupil under test are from wave-front measurements, and the slopes outside of the pupil under test are zero. Therefore the slope data crossing the original pupil boundaries between Ω_0 and $\Omega_1 \setminus \Omega_0$ are not continuous, and such discontinuous boundary conditions yield severe errors in the estimated wave front not only at the edge of the pupil but also within the pupil of interest through propagation of errors, as shown in Fig. 1. This is because the slopes do not satisfy the derivative continuity condition of the Poisson equation, so deviation errors will be induced. The iterative process makes possible a continuous practical extrapolation of the slope data outside the optical pupil Ω_0, while it does not interfere with the internal region of Ω_0. The iterative algorithm converges quickly to an unbiased solution, while at the same time the smoother the wave-front surface under construction, the smaller the residue deviation error, as expected, and the fewer iterations needed. Theoretically, the deviation error of this unique solution will decrease to zero. However, measurement noise prohibits the deviation errors from reaching zero, so they stagger to its minimum.

5. RESULTS

To validate and assess the capability of the algorithm across irregular shaped pupils, we present two examples, one with a circular 30-mm-diameter pupil and the other with the same-size pupil but with a 10% central obstruction. Both data sets were acquired from a previous experiment reported in a previous paper.[19] The 30-mm-diameter pupils, with a sampling grid of $2 \times 2\text{-mm}^2$ element size in both cases, were conjugated to a 500-mm-diameter mirror under test. The obstructed wave front was obtained by considering the slope data within the obstructed pupil only. To establish the ground truth for each set of data, we estimated the wave front from the same set of slope data without domain extension but with the conventional iterative or direct methods such as the Jacobi iterative method, the Gauss elimination method, and the Cholesky decomposition method.[19,21] All these methods yield exactly the same estimated wave front, which we thus consider to represent ground truth (i.e., the original wave front), against which the proposed iterative wave-front estimation algorithm could be assessed.

A. Case 1: Circular Pupil without Central Obstruction

A circular pupil without obstruction is a simply connected domain. The considered 30-mm-diameter pupil with an array of 161 Shack–Hartmann grid points is shown in Fig. 6. The points outside the circular pupil in the square lattice are the imaginary grid points. The ground-truth wave front is also shown in Fig. 6. Deviation-error maps of the wave front estimated by the Gerchberg-type iterative algorithm with several numbers of iterations (i.e., $i = 0, 1, 2, 3, 4, 13$) are shown in Fig. 7. Results show that for $\lambda = 632.8$ nm, the rms deviation error was reduced from $\lambda/16$ to $\lambda/129$ after 13 iterations, where it reached its minimum. The residual deviation is 12% of its original, and 88% of the deviation error was removed.

B. Case 2: Circular Pupil with a 10% Central Obstruction

A 30-mm circular pupil with a 10% central obstruction is shown in Fig. 8. Such a percent of obscuration is common for astronomical telescope mirrors. In testing such mirrors during the fabrication process and telescope assembly, an algorithm that permits testing of any pupil shapes without any additional steps in preparing and setting up for such a test provides key advantages not only in time efficiency but also in minimizing the risk of test-induced errors. Deviation-error maps of the wave-front values estimated by the Gerchberg-type iterative algorithm with several numbers of iterations (i.e., $i = 0, 1, 3, 5, 7, 10$) are shown in Fig. 9. Results show that for $\lambda = 632.8$ nm, the rms deviation error was reduced from $\lambda/14$ to $\lambda/154$ after

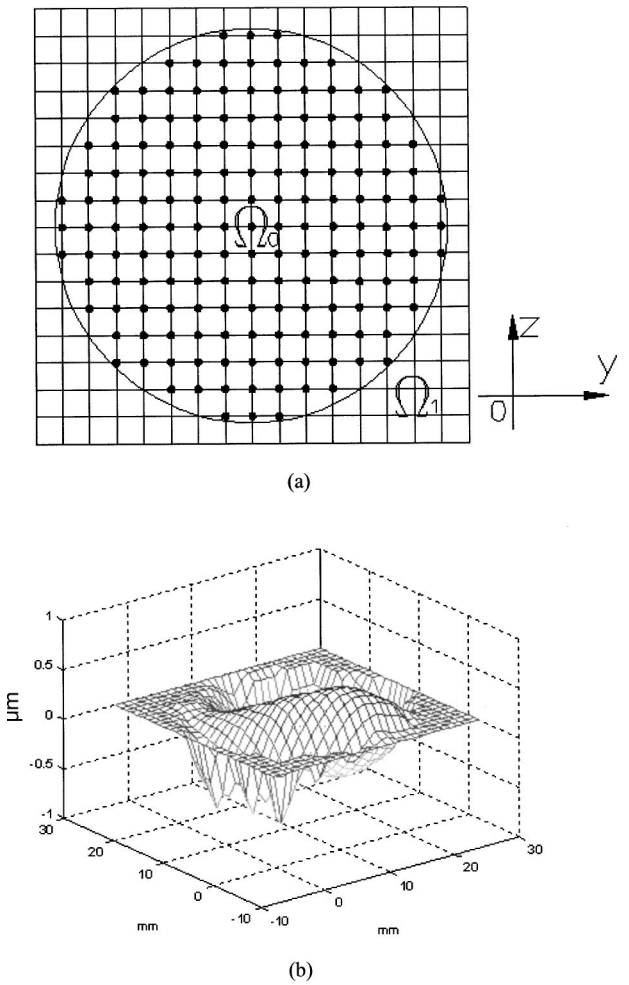

Fig. 6. (a) 30-mm-diameter circular pupil without central obstruction shown within the extended domain Ω_1. (b) The ground-truth wave front within the circular pupil Ω_0 on a vertical scale of ± 1 μm.

ten iterations, where it reached its minimum. The residual deviation error is 9% of its original, and 91% of the deviation error was removed.

C. Algorithm Convergence

The deviation error reduction through Gerchberg-type iterations was found to be efficient. Specifically, the final deviation error after a maximum of four iterations for the two examples considered was less than $\lambda/100$ for λ equal to 632.8 nm, as shown in Figs. 7 and 9. The convergence indicated by the rms wave-front error in units of wavelength as a function of the number of iterations is plotted in Fig. 10. Such a finding is high performance for optical testing, and the algorithm can be said to be very efficient.

6. THEORETICAL ANALYSIS OF COMPUTATIONAL COMPLEXITY

In this iterative wave-front estimation algorithm, three $m \times m$ (note that $m = t \times t$) linear equation sets need to be solved at each iteration. One is the equation set for wave-front estimation from the slope data, and the two others are the equation sets for y-slope and z-slope computations from a known wave front. The three equation sets are highly sparse. If we define a fill-in factor, an indicator of matrix sparsity, as the quotient of the number of nonzero elements to the total number of the matrix elements, then the fill-in factor of the wave-front estimation matrix $\mathbf{C}^T\mathbf{C}$ is $(5t-4)/t^3$, and the fill-in factors of the slope computation matrices $\mathbf{A}^T\mathbf{A}$ and $\mathbf{B}^T\mathbf{B}$ are both $(t+4)/t^3$. For example, the fill-in factors of the wave-front estimation matrix and the slope computation matrices are 4.6% and 1.4%, respectively, for $t=10$, and they decrease to 0.05% and 0.01% at $t=100$.

A. Spatial Complexity of the New Linear Least-Squares-Based Solution

Besides their high sparsity, all three matrices are symmetrical, positive, and banded once the wave-front zero point has been determined for the wave-front estimation. The extremely regular and symmetrical banded matrices allow efficient computations in solving linear equation sets because the nonzero elements in these matrices are regularly patterned with the numbers $4,3,2,1,-1$ only. Therefore the matrix storage problem is avoided in this algorithm except for approximately t^3 elements space reserved for the banded Cholesky decomposition of $\mathbf{C}^T\mathbf{C}$, which is a much smaller part than the $t^4/2$ element space set for a conventional Cholesky decomposition. Naturally, a space of $3t^2$ elements is necessary in each algorithm for storing the slope data and the wave-front values.

B. Computational Complexity of the New Linear Least-Squares-Based Solution

Before we discuss the computational complexity of our algorithm, we shall introduce "FLOPS," an acronym for "floating-point operations," to denote the arithmetic operations that a computer performs, such as multiplications or additions (or subtractions).[22] The positive-definite slope-extraction matrices $\mathbf{A}^T\mathbf{A}$ and $\mathbf{B}^T\mathbf{B}$ are banded and diagonal with semibandwidths of 2 and t, respectively. In computing the slope data from a wave front, it is an advantage to employ a direct solution method such as the Cholesky method to solve the normal equation sets, because we can decompose the matrices $\mathbf{A}^T\mathbf{A}$ and $\mathbf{B}^T\mathbf{B}$ into two unique triangular matrices by several simple derivations once and for all. Thereby no more Cholesky decompositions are needed in computation. The computations needed in solving the two systems of equations are substitutions only, which incur arithmetic costs of $\sim 4m$ times the bandwidth,[22] yielding $8t^2$ FLOPS for $\mathbf{A}^T\mathbf{A}$ and $4t^3$ FLOPS for $\mathbf{B}^T\mathbf{B}$. As a comparison, the computational cost needed for substitutions in solving an equation without exploiting the band structure is $2t^4$ FLOPS.[22]

To solve the linear system of equations for wave-front estimation, we first set the zero point for the wave front under construction to make the matrix $\mathbf{C}^T\mathbf{C}$ positive definite before the Cholesky decomposition can be performed. Since it is a banded sparse matrix with a semibandwidth of t, we can solve this equation set by a banded Cholesky decomposition method, which requires $\sim t^4$ FLOPS for decomposition and $4t^3$ FLOPS for substitutions.[22] As a comparison, employing the conventional Cholesky method to

solve this equation set without exploiting the band structure of the matrix yields $\sim\frac{1}{3}t^6+2t^4$ FLOPS.[22]

Other direct solution methods are also available in solving the above three equation sets, such as the Gaussian elimination and the singular-value decomposition (SVD) methods, but such methods are more computationally expensive. Generally, the conventional Gaussian elimination method requires $\sim\frac{2}{3}t^6$ FLOPS, and the SVD

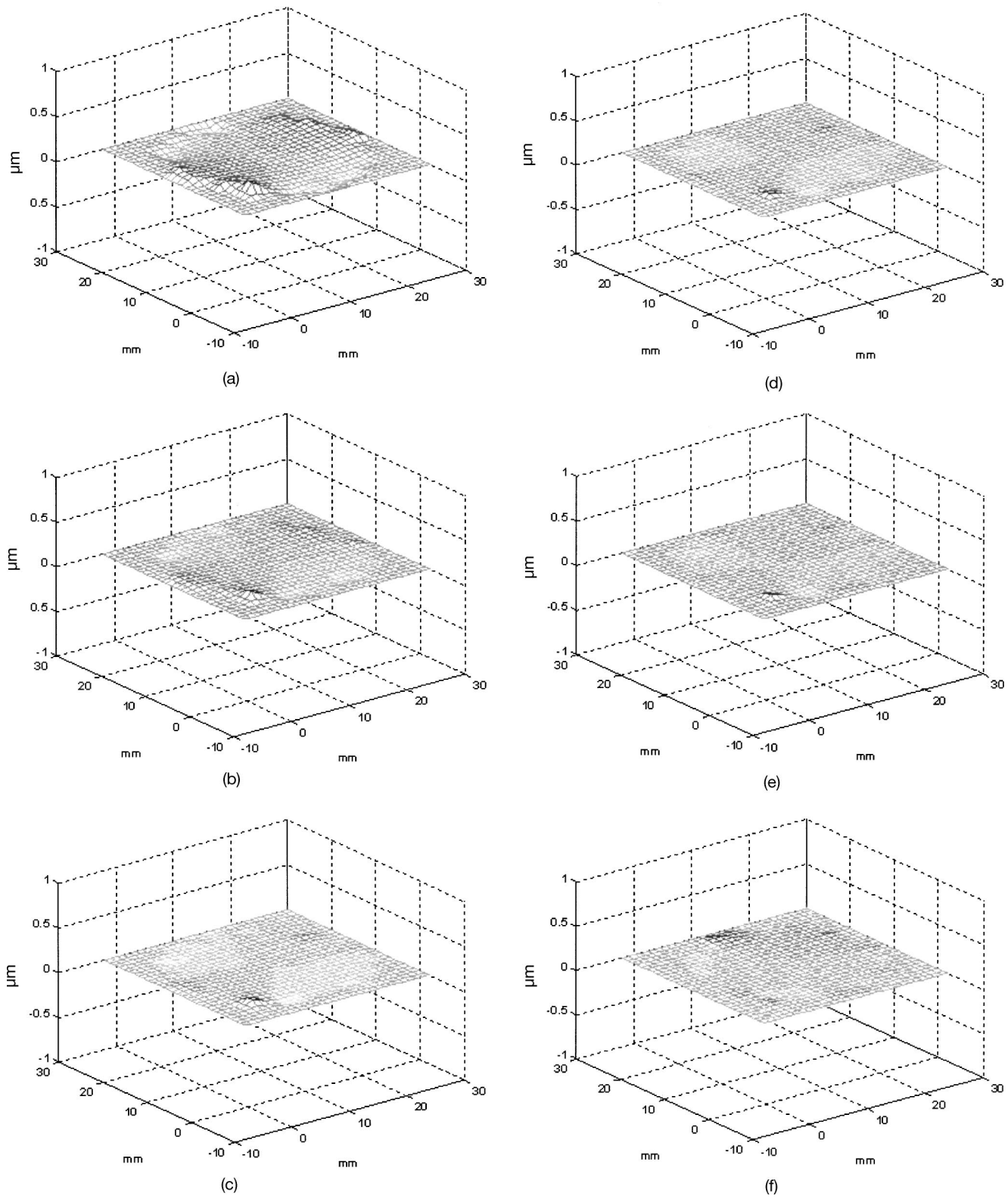

Fig. 7. Wave-front deviation error (on a vertical scale of ±1 μm, $\lambda=632.8$ nm) for a 30-mm-diameter circular pupil without obstruction across a sampled 15×15 point grid, for the number of iterations (a) $i=0$, rms=$\lambda/16$; (b) $i=1$, rms=$\lambda/37$; (c) $i=2$, rms=$\lambda/63$; (d) $i=3$, rms=$\lambda/86$; (e) $i=4$, rms=$\lambda/105$; (f) $i=13$, rms=$\lambda/129$.

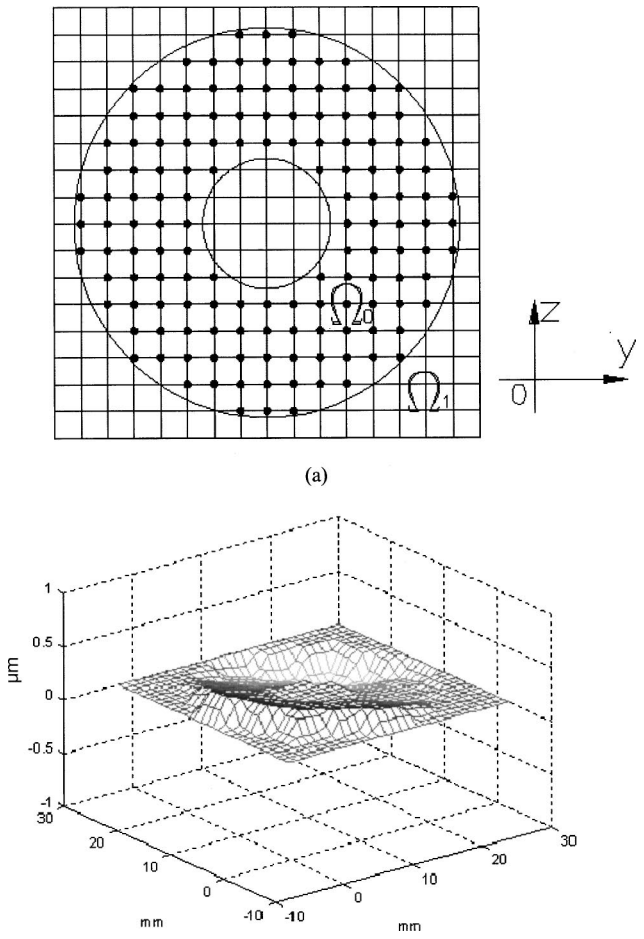

Fig. 8. (a). A 30-mm diameter circular pupil with a 10% central obstruction. (b). The ground-truth wave front at this pupil on a vertical scale of ±1 μm (right).

method requires $\sim 12t^6$ FLOPS.[23] Because the SVD method yields a unique solution with minimum norm for a rank-deficient LS problem, it is a good method in practice if the computational complexity is not a constraint.

An alternative for solving linear systems is to use iterative methods, such as the successive over-relaxation (SOR) algorithm, which is said to be one of most efficient among the classical iterative methods. The convergence rate of the SOR method is closely related to the problem model, the discretization mesh size, the relaxation factor, and the grid indexing orders. However, we can estimate the iteration counts needed for the optimal SOR method to converge to a solution with an error-reduction factor of 10^{-8} by[22,23]

$$R_{\omega b} = -8\left(\log_{10}\left(1 - \frac{2\pi}{t-1}\right)\right)^{-1} \approx 2.94(t+1) \approx 3t. \quad (37)$$

If it requires approximately $5t^2$ FLOPS operations in each iteration,[22] then the computational cost needed for solving Eq. (7) with the optimal SOR method is $\sim 15t^3$ FLOPS. The analysis indicates that the banded Cholesky method requires fewer FLOPS of computational cost for a small grid size ($t < 11$), but for a large grid size the SOR method is computationally less expensive. The complexity of the optimal SOR method increases with a cubic curve, whereas the complexity of banded Cholesky method increases with a quartic curve.

C. Comparison of Complexity with Fast-Fourier-Transform-Based Iterative Algorithms

1. Comparison of Computational Complexity

The FFT of a data set of length $m = 2^q$ (q is a positive integer) requires approximately $m \log_2(m)$ complex multiplications, which is equivalent to $5m \log_2(m)$ FLOPS of arithmetic operations according to a detailed analysis by Brigham.[24,25] The FFT-based iterative algorithm proposed by Roddier and Roddier[18] requires computation of two FFTs in addition to the computations of the y and the z slopes from the wave front at each iteration.[17] Therefore, if we assume that the computation of the slope extractions takes the same computational cost in both algorithms and we ignore this cost, the computational cost needed in one iteration of the FFT-based iterative algorithm is approximately $(20 \log_2(t) + 7)t^2$ FLOPS, which is usually much smaller than that of the optimal SOR-based (i.e., $15t^3$ FLOPS) iterative algorithm. Thus, on the basis of the evaluation of the computational complexity of one iteration, the FFT-based algorithm is superior in performance to the proposed algorithm, which could be significant for large values of grid-array size t ($t \times t = m$). Also, the required number of iterations for each algorithm will significantly affect the overall computational time. The number of iterations is a factor of the overall grid-array size, measurement-noise levels, etc. The FFT-based algorithms usually converge slowly; for example, the Gerchberg–Saxton algorithm requires at least tens to hundreds or even thousands of iterations to converge to a solution, while we found that the algorithm proposed in this paper converges to less than $\lambda/100$ deviation error after a maximum of approximately four iterations.

2. Comparison of Spatial Complexity

It has been established that the space complexity required for a FFT is $\sim O(t^4)$, which corresponds to the storage of complex matrix arrays. For the Gerchberg-type iterative algorithm that we have proposed with the banded Cholesky solution method, the spatial complexity is only $\sim O(t^3)$.

7. ERROR PROPAGATION

If we neglect the perturbations introduced by the rounding errors, wave-front errors may occur from two sources: the algorithm discretization errors that depend on the basic estimation scheme adopted and the wave-front sensor measurement error, such as the CCD centroiding error. The discretization errors of the wave-front estimation scheme adopted in this paper have been discussed in detail previously.[19] In this paper we discuss the error propagation of the wave-front estimation from the wave-front measurements. For convenience, we define the noise coefficient as Southwell did.[10]

Let us define a wave-front estimation matrix **C**. The wave-front slope measurement error vector may be given

as $S' = (s'_1 \quad s'_2 \quad \cdots \quad s'_m)^T$, and the wave-front error vector may be expressed as $W' = (w'_1 \quad w'_2 \quad \cdots \quad w'_m)^T$. If we ignore the algorithm discretization errors, we can write the error propagation matrix equation as

$$\mathbf{C}^T\mathbf{C}\mathbf{W}' = a\mathbf{C}^T\mathbf{S}', \tag{38}$$

where $\mathbf{C}$ was defined in Eq. (5) and a is the distance between grid points. If we introduce the Euclidian norm of

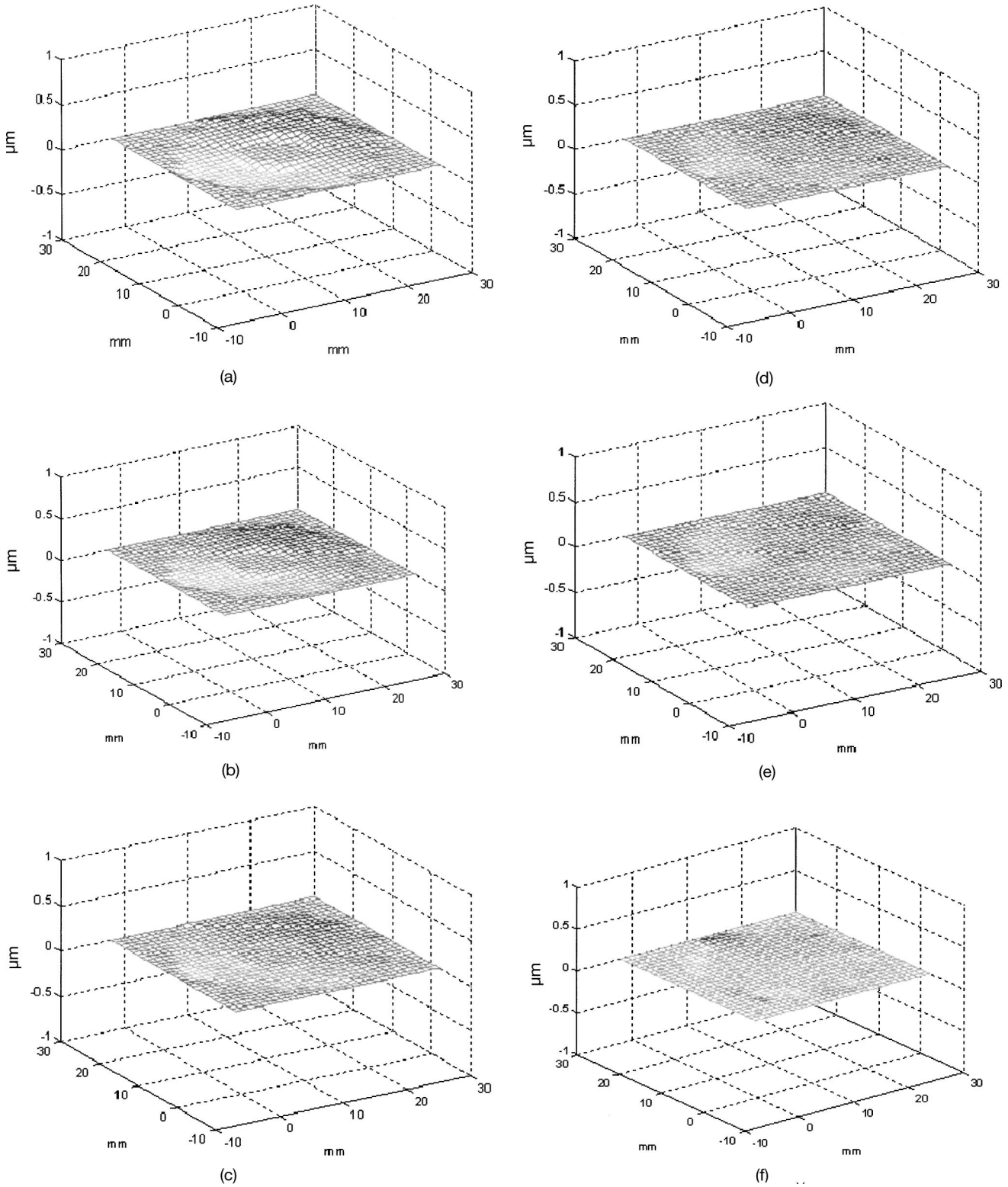

Fig. 9. Wave-front deviation error (on a vertical scale of ±1 μm, $\lambda = 632.8$ nm) for a 30-mm-diameter circular pupil sampled with a 15 × 15 grid and with a 10% central obstruction, for the number of iterations (a) $i = 0$, rms = $\lambda/14$; (b) $i = 1$, rms = $\lambda/26$; (c) $i = 3$, rms = $\lambda/63$; (d) $i = 5$, rms = $\lambda/107$; (e) $i = 7$, rms = $\lambda/135$; (f) $i = 10$, rms = $\lambda/154$.

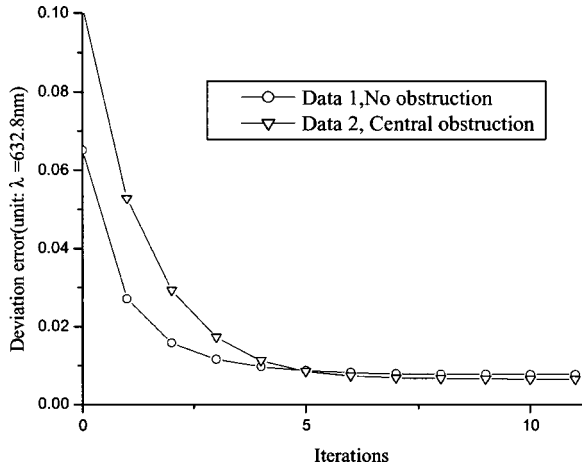

Fig. 10. Plot of RMS deviation errors in units of wavelength as a function of the number of iterations for the two data sets considered.

vector $\mathbf{X}$ as

$$\|\mathbf{X}\|_2 = (\mathbf{X}^T\mathbf{X})^{1/2} \tag{39}$$

and the corresponding matrix norm for a matrix $\mathbf{C}$ as[21]

$$\mathrm{lub}_2(\mathbf{C}) = \max_{X\neq 0}\left(\frac{\mathbf{X}^T\mathbf{C}^T\mathbf{C}\mathbf{X}}{\mathbf{X}^T\mathbf{X}}\right)^{1/2} = [\rho(\mathbf{C}^T\mathbf{C})]^{1/2}, \tag{40}$$

where X is an arbitrary vector and $\rho(\mathbf{C}^T\mathbf{C})$ is the spectral radius of $\mathbf{C}^T\mathbf{C}$, and if $\mathbf{C}^T\mathbf{C}$ is invertible, Zou *et al.* proved that[19]

$$\|\mathbf{W}'\|_2 \leqslant a\frac{[\mathrm{cond}(\mathbf{C}^T\mathbf{C})]^{1/2}}{\mathrm{lub}_2(\mathbf{C})}\|\mathbf{S}'\|_2, \tag{41}$$

where $\mathrm{cond}(\mathbf{C}^T\mathbf{C})$ is defined as the matrix condition number of $\mathbf{C}^T\mathbf{C}$, and

$$\mathrm{cond}(\mathbf{C}^T\mathbf{C}) \coloneqq \mathrm{lub}_2(\mathbf{C}^T\mathbf{C})\mathrm{lub}_2[(\mathbf{C}^T\mathbf{C})^{-1}] = \rho(\mathbf{C}^T\mathbf{C})\rho[(\mathbf{C}^T\mathbf{C})^{-1}]. \tag{42}$$

Since $\rho(\mathbf{C}^T\mathbf{C}) = |\lambda_{\max}|$, then $\rho[(\mathbf{C}^T\mathbf{C})^{-1}] = |\lambda_{\min}|^{-1}$, where $\lambda_{\max}$ and $\lambda_{\min}$ are the maximum and minimum of the eigenvalues of matrix $\mathbf{C}^T\mathbf{C}$, respectively. Therefore the condition number of $\mathbf{C}^T\mathbf{C}$ can be written as

$$\mathrm{cond}(\mathbf{C}^T\mathbf{C}) = \frac{|\lambda_{\max}|}{|\lambda_{\min}|}. \tag{43}$$

Then Eq. (41) becomes

$$\|\mathbf{W}'\|_2 \leqslant \frac{a\|\mathbf{S}'\|_2}{\sqrt{|\lambda_{\min}|}}. \tag{44}$$

If the wave-front slope errors are independent and have the same variance σ_s^2, then consider

$$\|\mathbf{W}'\|_2 = \sqrt{m}\left(\frac{1}{m}\sum_{i=1}^{m}|w_i'|^2\right)^{1/2} = t\sigma_w, \tag{45}$$

$$\|\mathbf{S}'\|_2 = \sqrt{m}\left(\frac{1}{m}\sum_{i=1}^{m}|s_i'|^2\right)^{1/2} = t\sigma_s, \tag{46}$$

where σ_w and σ_s are the rms errors of the wave front and of the wave-front slope measurements, respectively. From Eqs. (44)–(46), we have

$$\sigma_w \leqslant \frac{a\sigma_s}{|\lambda_{\min}|^{1/2}}. \tag{47}$$

Let σ_d be the rms error of the wave-front difference measurements with $\sigma_d = a\sigma_s$; then we have

$$\sigma_w \leqslant \gamma\sigma_d, \tag{48}$$

where

$$\gamma = |\lambda_{\min}|^{-1/2}. \tag{49}$$

It is a limit estimation of error-propagation coefficients, where γ^2 is the limit ratio of the mean square of the wave-front error to the mean square of the wave-front difference error, which is called the noise coefficient or error-propagation coefficient,[1,5,6,10,14] where

$$\sigma_w^2/\sigma_d^2 \leqslant \gamma^2 = |\lambda_{\min}|^{-1}. \tag{50}$$

Equation (50) points to the fact that the error-propagation coefficient is limited by the reciprocal of the minimum eigenvalue of the normal matrix.

The analysis is also applicable to the error propagation of slope computation provided by Eqs. (22) and (36). The problem is then reduced to evaluating the minimum eigenvalue of the normal matrix. Since the normal equation matrix is symmetric, the classical Jacobi method can be employed to compute eigenvalues.[26] The eigenvalues of the two slope-computation matrices of Eqs. (22) and (36) are found to be $2-\sqrt{2}$, $2+\sqrt{2}$, and 4 when $t>4$. For $t=4$, the eigenvalues are $2-\sqrt{2}$ and $2+\sqrt{2}$ only. Thereby the condition numbers of these two matrices are both $4/(2-\sqrt{2}) \approx 6.83$ at $t>4$ and $(2+\sqrt{2})/(2-\sqrt{2}) \approx 5.83$ at $t=4$. Therefore for y- (or, equivalently, z-) slope computations, we have

$$\sigma_s^2/\sigma_w^2 \leqslant \frac{1}{2-\sqrt{2}} = 1 + \frac{\sqrt{2}}{2} \approx 1.71, \qquad t \geqslant 4. \tag{51}$$

For the wave-front estimation matrix in Eq. (7), the estimation of the error propagation coefficient is a little more complex because the eigenvalues of this matrix are sensitive to the variation of the wave-front zero point, the matrix dimension size, and even the parity of the number of the matrix dimension. Employing the classical Jacobi method to compute the eigenvalues of Eq. (7), we obtain the curve of the noise-coefficient limit versus the grid size of the wave front, which is shown in Fig. 11. Results show that the wave-front estimation has a better performance in error propagation when the dimension number of the estimation matrix is odd. Therefore an odd number of the sampling grid array is preferable to its closest even number. Making a least-squares fitting of this curve, we can express the relationship quantitatively as

$$\sigma_w^2/\sigma_d^2 \leqslant \gamma^2 = \begin{cases} -44 + 28.58\, e^{t/8.925}, & t \text{ is even} \\ -31.875 + 20.61\, e^{t/7.766}, & t \text{ is odd} \end{cases}. \quad (52)$$

The error propagation of parity dependence is also reflected in the curve of the matrix condition numbers as shown in Fig. 12. By making a least-squares fitting of this curve, we obtained the condition number of the wave-front estimation matrix by

$$\mathrm{cond}_2(\mathbf{C}^{\mathrm{T}}\mathbf{C}) = \begin{cases} -243.442 + 150.87\, e^{t/7.518}, & t \text{ is odd} \\ -355.157 + 223.750\, e^{t/6.83}, & t \text{ is even} \end{cases}. \quad (53)$$

When the matrix dimension becomes larger, the Jacobi method converges slowly. It takes approximately 68,000 iterations to obtain the eigenvalues with 10^{-7} accuracy for the case of $t=15$. We obtained that at $t=15$ the maximum eigenvalue is 7.87 and the minimum eigenvalue is 0.009; therefore the condition number is 865. Such a condition number indicates that the error propagation in the wave-front estimation is stable and slow.

Some relatively subtle differences between the Roddier and Roddier[18] algorithm and the algorithm proposed here is the fact that in performing FFTs, the square-array matrix satisfies $m=2^q$, and therefore m must be even. In the case of the proposed algorithm, there is no such requirement, and we have shown that odd matrix sizes yield lower error propagation than even matrix sizes.

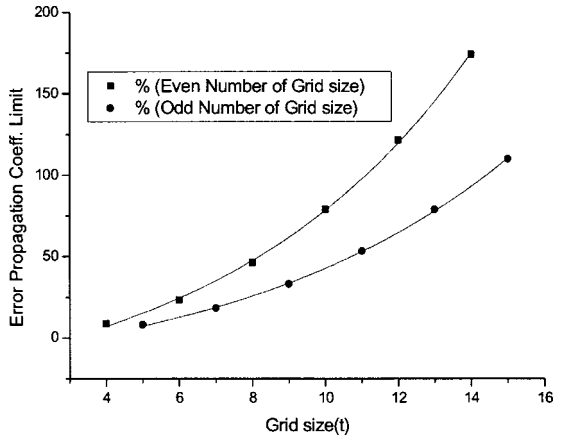

Fig. 11. Noise-coefficient limit versus the dimension size of the sampling grid.

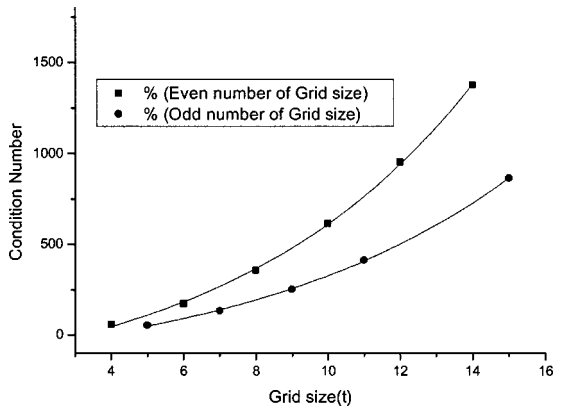

Fig. 12. Normal matrix condition number versus grid dimension size.

8. CONCLUSION

Klaus Freischlad pointed out in 1992 that a wave-front estimation algorithm suitable for practical optical testing must have the following properties[14]: (1) The wave-front estimates must be unbiased; (2) the error-propagation coefficient must be slow; (3) the computation must be efficient, especially for large data sets, and the necessary memory space should be small enough to be applicable in the laboratory; and (4) finally, the algorithm should be easily adaptable to various pupil shapes. In this paper we combined a Gerchberg–Saxton-type iterative process with a linear least-squares method to obtain a practical unbiased wave-front estimation algorithm for optical testing that combines the accuracy of the iterative wave-front extrapolation technique with the efficiency of the linear sparse matrix. With the domain extension technique, we obtained a predefined wave-front estimation matrix and the associated predefined slope-computation matrices for any pupil shape or size. The matrix coefficients are thus fully predetermined and known once and for all. An analysis of the error propagation showed that the wave-front estimation matrix is well-conditioned, yielding low propagation errors. A U.S. patent application has been filed for this iterative wave-front estimation algorithm; its provisional application number is 60/507, 657.

APPENDIX A: DERIVATION OF EQS. (11), (12), (26), and (27)

Let us denote

$$\left.\frac{\partial^n W}{\partial y^n}\right|_i$$

as the nth derivative of the wave front at point i and

$$\left.\frac{\partial^n W}{\partial y^n}\right|_{i+1/2}$$

as the nth derivative of the wave front at the midpoint between points i and $i+1$. According to Taylor's series, we can write

$$w_i = w_{i+1/2} - \frac{a}{2}\left.\frac{\partial W}{\partial y}\right|_{i+1/2} + \frac{a^2}{4\times 2!}\left.\frac{\partial^2 W}{\partial y^2}\right|_{i+1/2} - \frac{a^3}{8\times 3!}\left.\frac{\partial^3 W}{\partial y^3}\right|_{i+1/2} + \frac{a^4}{16\times 4!}\left.\frac{\partial^4 W}{\partial y^4}\right|_{i+1/2} + O(a^5), \quad (A1)$$

$$w_{i+1} = w_{i+1/2} + \frac{a}{2}\left.\frac{\partial W}{\partial y}\right|_{i+1/2} + \frac{a^2}{4\times 2!}\left.\frac{\partial^2 W}{\partial y^2}\right|_{i+1/2} + \frac{a^3}{8\times 3!}\left.\frac{\partial^3 W}{\partial y^3}\right|_{i+1/2} + \frac{a^4}{16\times 4!}\left.\frac{\partial^4 W}{\partial y^4}\right|_{i+1/2} + O(a^5). \quad (A2)$$

By subtracting Eq. (A1) from Eq. (A2), we can write

$$w_{i+1} - w_i = a \left.\frac{\partial W}{\partial y}\right|_{i+1/2} + \frac{a^3}{4 \times 3!} \left.\frac{\partial^3 W}{\partial y^3}\right|_{i+1/2} + O(a^5). \tag{A3}$$

By adding Eqs. (A1) and (A2), we can write

$$w_{i+1} + w_i = 2w_{i+1/2} + \frac{a^2}{4} \left.\frac{\partial^2 W}{\partial y^2}\right|_{i+1/2} + \frac{a^4}{8 \times 4!} \left.\frac{\partial^4 W}{\partial y^4}\right|_{i+1/2} + O(a^6). \tag{A4}$$

The replacement of w with $\partial W/\partial y$ yields

$$\left.\frac{\partial W}{\partial y}\right|_{i+1} + \left.\frac{\partial W}{\partial y}\right|_{i} = 2 \left.\frac{\partial W}{\partial y}\right|_{i+1/2} + \frac{a^2}{4} \left.\frac{\partial^3 W}{\partial y^3}\right|_{i+1/2} + O(a^6). \tag{A5}$$

Thus

$$\left.\frac{\partial W}{\partial y}\right|_{i+1/2} = \frac{1}{2}\left(\left.\frac{\partial W}{\partial y}\right|_{i+1} + \left.\frac{\partial W}{\partial y}\right|_{i} \right) - \frac{a^2}{8} \left.\frac{\partial^3 W}{\partial y^3}\right|_{i+1/2} + O(a^6). \tag{A6}$$

And using Eq. (A6), we may express Eq. (A3) as

$$w_{i+1} - w_i = \frac{a}{2}\left(\left.\frac{\partial W}{\partial y}\right|_{i+1} + \left.\frac{\partial W}{\partial y}\right|_{i} \right) - \frac{a^3}{12} \left.\frac{\partial^3 W}{\partial y^3}\right|_{i+1/2} + O(a^5). \tag{A7}$$

Let us now replace w with $\partial W/\partial y$ in Eq. (A7), to yield

$$\left.\frac{\partial W}{\partial y}\right|_{i+1} - \left.\frac{\partial W}{\partial y}\right|_{i} = \frac{a}{2}\left(\left.\frac{\partial^2 W}{\partial y^2}\right|_{i+1} + \left.\frac{\partial^2 W}{\partial y^2}\right|_{i} \right) - \frac{a^3}{12} \left.\frac{\partial^4 W}{\partial y^4}\right|_{i+1/2} + O(a^5). \tag{A8}$$

Remember that[19]

$$w_{i+1} - 2w_i + w_{i-1} = a^2 \left.\frac{\partial^2 W}{\partial y^2}\right|_{i} + \frac{a^4}{12} \left.\frac{\partial^4 W}{\partial y^4}\right|_{i} + O(a^6), \tag{A9}$$

so

$$\left.\frac{\partial^2 W}{\partial y^2}\right|_{i} = \frac{w_{i+1} - 2w_i + w_{i-1}}{a^2} - \frac{a^2}{12} \left.\frac{\partial^4 W}{\partial y^4}\right|_{i} + O(a^4). \tag{A10}$$

Applying Eqs. (A10) in (A8), we have

$$\left.\frac{\partial W}{\partial y}\right|_{i+1} - \left.\frac{\partial W}{\partial y}\right|_{i} = \frac{a}{2}\left(\frac{w_{i+2} - 2w_{i+1} + w_i}{a^2} + \frac{w_{i+1} - 2w_i + w_{i-1}}{a^2} \right) - \frac{a^3}{24}\left(\left.\frac{\partial^4 W}{\partial y^4}\right|_{i} + \left.\frac{\partial^4 W}{\partial y^4}\right|_{i+1} \right) - \frac{a^3}{12} \left.\frac{\partial^4 W}{\partial y^4}\right|_{i+1/2} + O(a^5) \tag{A11}$$

or

$$\left.\frac{\partial W}{\partial y}\right|_{i+1} - \left.\frac{\partial W}{\partial y}\right|_{i} = \frac{1}{2a}(w_{i+2} - w_{i+1} - w_i + w_{i-1}) + O(a^5). \tag{A12}$$

Let us neglect the higher-order small-value terms on the right-hand side of Eq. (A12) and denote the first derivative of the wave front in the y direction as s_y; then we have

$$s_{y_{i+2}} - s_{y_{i+1}} = \frac{1}{2a}(w_{i+3} - w_{i+2} - w_{i+1} + w_i), \tag{A13}$$

where $i = 1, 2, \ldots t-3; t+1, t+2, \ldots 2t-3, \ldots m-3$.

Similarly, in the z direction we can write

$$s_{z_{i+t}} - s_{z_{i+2t}} = \frac{1}{2a}(w_i - w_{i+t} - w_{i+2t} + w_{i+3t}),$$

$$i = 1,2, \ldots t, t+1, t+2, \ldots 2t, \ldots, m-3t. \tag{A14}$$

.

ACKNOWLEDGMENTS

This research was supported in part by National Science Foundation grant IIS/HCI-0307189. We thank Greg Forbes for stimulating discussion about this work and Felix G. Hamza-Lup and Anand Santhanam for stimulating discussions about analysis of algorithm complexity.

Weiyao Zou can be reached by e-mail at wzou@creol.ucf.edu or wyzou_2000@yahoo.com, and Jannick Rolland can be reached by e-mail at Jannick@odalab.ucf.edu.

REFERENCES

1. R. J. Noll, "Phase estimates from slope-type wave-front sensors," J. Opt. Soc. Am. **68**, 139–140 (1978).
2. M. P. Rimmer, "Method for evaluating lateral shearing interferograms," Appl. Opt. **13**, 623–629 (1974).
3. J. C. Wyant, "Use of an ac heterodyne lateral shear interferometer with real-time wavefront correction systems," Appl. Opt. **14**, 2622–2626 (1975).
4. J. W. Hardy, J. E. Lefebvre, and C. L. Koliopoulos, "Real-time atmospheric compensation," J. Opt. Soc. Am. **67**, 360–369 (1977).
5. D. L. Fried, "Least-square fitting a wave-front distortion estimate to an array of phase-difference measurements," J. Opt. Soc. Am. **67**, 370–375 (1977).
6. R. H. Hudgin, "Wave-front reconstruction for compensated imaging," J. Opt. Soc. Am. **67**, 375–378 (1977).

7. B. R. Hunt, "Matrix formulation of the reconstruction of phase values from phase differences," J. Opt. Soc. Am. **69**, 393–399 (1979).
8. R. Cubalchini, "Modal wave-front estimation from phase derivative measurements," J. Opt. Soc. Am. **69**, 972–977 (1979).
9. J. Herrmann, "Least-square wave-front errors of minimum norm," J. Opt. Soc. Am. **70**, 28–35 (1980).
10. W. H. Southwell, "Wave-front estimation from wave-front slope measurements," J. Opt. Soc. Am. **70**, 998–1006 (1980).
11. J. Herrmann, "Cross coupling and aliasing in modal wave-front estimation," J. Opt. Soc. Am. **71**, 989–992 (1981).
12. K. Freischlad and C. L. Koliopoulos, "Wavefront reconstruction from noisy slope or difference data using the discrete Fourier transform," in *Adaptive Optics*, J. E. Ludman, ed., Proc. SPIE **551**, 74–80 (1985).
13. K. Freischlad and C. L. Koliopoulos, "Modal estimation of a wave-front difference measurements using the discrete Fourier transform," J. Opt. Soc. Am. A **3**, 1852–1861 (1986).
14. K. Freischlad, "Wavefront integration from difference data," in *Interferometry: Techniques and Analysis*, G. M. Brown, O. Y. Kwon, M. Kujawinska, and G. T. Reid, eds., Proc. SPIE **1755**, 212–218 (1992).
15. R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," Optik (Stuttgart) **35**, 237–246 (1972).
16. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21**, 2758–2769 (1982).
17. C. Roddier and F. Roddier, "Interferogram analysis using Fourier transform techniques," Appl. Opt. **26**, 1668–1673 (1987).
18. F. Roddier and C. Roddier, "Wave-front reconstruction using iterative Fourier transforms," Appl. Opt. **30**, 1325–1327 (1991).
19. W. Zou and Z. Zhang, "Generalized wave-front reconstruction algorithm applied in a Shack–Hartmann test," Appl. Opt. **39**, 250–268 (2000).
20. D. Su, S. Jiang, and L. Shao, "A sort of algorithm of wavefront reconstruction for Shack–Hartmann test (European Southern Observatory, Garching, Germany, 1993), pp. 289–292.
21. J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, 2nd ed., Vol. 12 of Texts in Applied Mathematics (Springer, New York, 1993), pp. 186, 594, and 635.
22. D. S. Watkins, *Fundamentals of Matrix Computations*, 2nd ed. (Wiley, New York, 2002), pp. 3, 58, and 548.
23. A. Quarteroni, R. Sacco, and F. Saleri, *Numerical Mathematics*, Vol. 37 of Texts in Applied Mathematics (Springer, New York, 2000).
24. W. J. Thompson, *Computing for Scientists and Engineers* (Wiley, New York, 1992), p. 333.
25. E. O. Brigham, *The Fast Fourier Transform and Its Applications* (Prentice Hall, Englewood Cliffs, N.J., 1988), pp. 134, 164.
26. R. Kress, *Numerical Analysis*, Graduate Texts in Mathematics (Springer, New York, 1998), p. 127.