

Quantifications of error propagation in slope-based wavefront estimations

Weiyao Zou and Jannick P. Rolland

CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, Florida 32816

Received November 28, 2005; revised March 31, 2006; accepted April 7, 2006; posted May 11, 2006 (Doc. ID 66237)

We discuss error propagation in the slope-based and the difference-based wavefront estimations. The error propagation coefficient can be expressed as a function of the eigenvalues of the wavefront-estimation-related matrices, and we establish such functions for each of the basic geometries with the serial numbering scheme with which a square sampling grid array is sequentially indexed row by row. We first show that for the wavefront estimation with the wavefront piston value determined, the odd-number grid sizes yield better error propagators than the even-number grid sizes for all geometries. We further show that for both slope-based and difference-based wavefront estimations, the Southwell geometry offers the best error propagators with the minimum-norm least-squares solutions. Noll's theoretical result, which was extensively used as a reference in the previous literature for error propagation estimates, corresponds to the Southwell geometry with an odd-number grid size. Typically the Fried geometry is not preferred in slope-based optical testing because it either allows subsize wavefront estimations within the testing domain or yields a two-rank deficient estimations matrix, which usually suffers from high error propagation and the waffle mode problem. The Southwell geometry, with an odd-number grid size if a zero point is assigned for the wavefront, is usually recommended in optical testing because it provides the lowest-error propagation for both slope-based and difference-based wavefront estimations. © 2006 Optical Society of America

OCIS codes: 080.2720, 220.4840, 120.6650, 010.7350.

1. INTRODUCTION

Wavefront estimation is a basic problem in optical testing. On the basis of slope-type data, the method converts wavefront slope or wavefront difference (WFD) measurements into wavefront values (or further into wavefront phase, which is a multiplication of the wavefront value with the constant $2\pi/\lambda$). The wavefront estimation is a numerical solution to a partial differential equation, specifically, Neumann's boundary problem of the Poisson equation. Using a two-dimensional (2-D) finite-difference grid to cover the testing domain (usually the testing is performed over the optical pupil), one can discretize this problem and evaluate the wavefront values at each grid point by solving a difference equation set defined on the grid. The matrix equation set for this problem can be expressed as

$$\mathbf{CW} = \mathbf{S}, \quad (1)$$

where \mathbf{C} is wavefront estimation matrix, \mathbf{W} is the vector of wavefront values, and \mathbf{S} is the vector of the WFD between the neighboring grid points, which can be converted from the wavefront slope vector \mathbf{G} by

$$\mathbf{S} = \mu a \mathbf{M} \mathbf{G}, \quad (2)$$

where μ is a constant related to the given geometry, a is the pitch size of the sampling grid, and \mathbf{M} is the conversion matrix for a given geometry. Owing to the measurement noise, Eq. (1) has its weak solutions only, and an unbiased least-squares approximation can be employed. The normal matrix equation for the least-squares solution is

$$\mathbf{C}^T \mathbf{C} \mathbf{W} = \mathbf{C}^T \mathbf{S}, \quad (3)$$

which is a discretization form of the Poisson equation

$$\nabla^2 \mathbf{W} = \nabla \mathbf{S} = f(y, z), \quad (4)$$

where $\nabla^2 \mathbf{W}$ is the Laplacian of the wavefront, $\nabla = (\partial/\partial x)\mathbf{i} + (\partial/\partial y)\mathbf{j}$, and $f(y, z)$ is a function given. It shall be noted that the Tikhonov regularized least-squares solution, adopted widely in other applications such as adaptive optics, provides a biased estimate that would not be desirable in optical testing.¹

The wavefront estimation errors can be categorized into two types. One type is from the algorithm discretization errors, which depends on the basic estimation scheme we adopted, and the other type is from the WFD error, which can be converted from the slope measurement noise if the measurements are wavefront slopes. The discretization error, the accuracy of a wavefront estimation geometry, some of which were detailed in a separate paper,² are not considered in this paper. The noisy WFD vector can be written as $\mathbf{S} = \mathbf{S}_0 + \mathbf{N}$, where \mathbf{S}_0 is the vector of the true WFD values, and \mathbf{N} is the measurement noise of the WFD. If the induced wavefront error from measurement noise \mathbf{N} is ϵ , then the estimated wavefront can be written as $\mathbf{W} = \mathbf{W}_0 + \epsilon$, where \mathbf{W}_0 is the true wavefront values.

The error propagation coefficient provides a quantitative metric for evaluating wavefront estimation. As defined in the previous literature,³⁻⁵ the error propagation coefficient η is the ratio of the mean variance of the wavefront estimation error σ_w^2 given by

$$\sigma_w^2 = \|\epsilon\|_2^2/m, \quad (5)$$

where $\|\cdot\|_2$ is the Euclidian norm and m is the total number of grid points ($m=t \times t$ for a square array, t is the grid size), to the variance of the WFD measurement error σ_n^2 , i.e.,

$$\eta = \frac{\sigma_w^2}{\sigma_n^2}. \quad (6)$$

In this paper, this definition is referred to as the WFD-based error propagation coefficient, which is a quantitative indicator of the error propagator in the WFD-based wavefront estimation. In Section 2, we briefly review the previous work published on error propagation in wavefront estimation; in Section 3, we generally formulate the WFD-based error propagation coefficient with the matrix eigenvalues; in Section 4, a serial numbering wavefront estimation scheme is introduced; in Section 5, we apply the error propagation formulation for the Hudgin geometry, the Southwell geometry, and the Fried geometry, and we derive a formula of error propagation coefficient for each case. The simulation result is discussed. In Section 6, we introduce the definition of the slope-based error propagation coefficient and formulate it for each of the basic geometries, and in Section 7 we conclude the analysis in this paper.

2. BRIEF REVIEW OF PREVIOUS WORK

Wavefront estimation can be performed on the three introduced geometries: the Hudgin geometry,⁶ the Southwell geometry,⁴ and the Fried geometry.⁷ As shown in Fig. 1, the slope measurements (or alternatively the WFD measurements) are denoted by small arrows, and the grid points are denoted by small circles in each of the geometries to follow the convention adopted previously.⁸ The slope measurements are performed at the midpoint between each neighboring grid points in the Hudgin geometry and at the center point circled by each four neighboring grid points in the Fried geometry. The Southwell geometry is characterized by taking the wavefront slope measurements and the wavefront value estimations at the same nodes. It was given by Southwell that this latter geometry has the lowest error propagation in wavefront estimation across relatively small grid sizes,⁴ as shown in Fig. 2.

We focus the analysis of error propagation on the zonal wavefront estimation induced by the WFD measurements with the estimation geometries. As shown in Fig. 2, the error propagation coefficient for each wavefront estimation

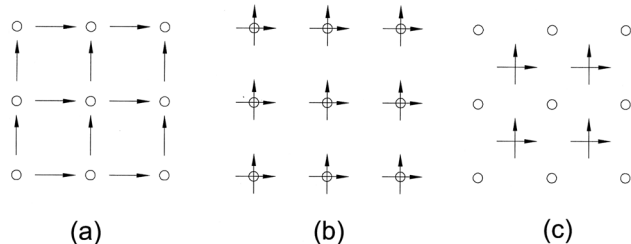


Fig. 1. Wavefront estimation schemes: (a) Hudgin geometry, (b) Southwell geometry, (c) Fried geometry.

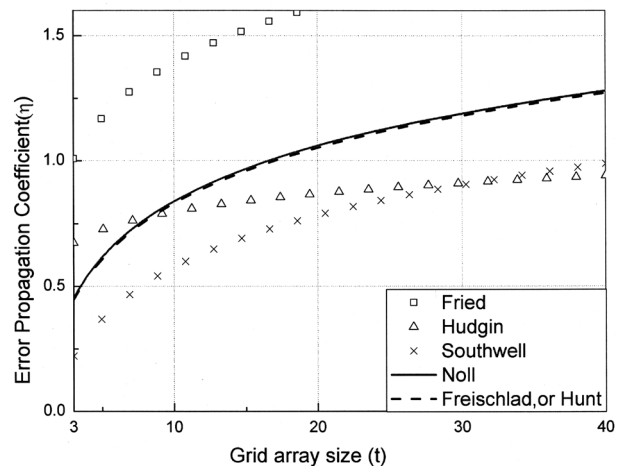


Fig. 2. Previous results on error propagation.

tion geometry was shown to be logarithmic dependent on the grid size t .^{4,6,7} From Eq. (35) in Ref. 7, we obtained the result given by Fried as

$$\eta_{\text{Fried}} = 0.6558 + 0.3206 \ln(t), \quad (7)$$

and from Eq. (32) in Ref. 6, we obtained the result given by Hudgin as

$$\eta_{\text{Hudg}} = 0.561 + 0.103 \ln(t). \quad (8)$$

For the Southwell geometry we least-squares fit our discrete coordinate measurements of curve A in Fig. 2 on page 1003 of Ref. 4 given by Southwell, and we obtain

$$\eta_{\text{South}} = -0.10447 + 0.2963 \ln(t). \quad (9)$$

By using a general Green's function, Noll analytically derived that the error propagation with a square aperture is a logarithmic-dependent function of its grid size t , which is³

$$\eta_{\text{Noll}} = 0.1072 + 0.318 \ln(t). \quad (10)$$

In modal estimation with discrete Fourier transform, Freischlad and Koliopoulos also showed that the error propagation satisfies a logarithmic dependence of $\eta_{\text{H,S;modal}} = a + (1/\pi)\ln(t)$ for the Hudgin (H) geometry and the Southwell (S) geometry, and a logarithmic dependence of $\eta_{\text{F;modal}} = a + (3/\pi)\ln(t-1)$ for the Fried (F) geometry.⁹ On the basis of the Hudgin geometry and a fast-Fourier-transform-based algorithm, which requires an even size, Freischlad confirmed that it is almost identical to Noll's theoretical result with⁵

$$\eta_{\text{Frei}} = 0.09753 + \frac{1}{\pi} \ln(t). \quad (11)$$

Hunt extended Hudgin's method for error coefficient prediction in terms of the discrete Fourier transform and obtained a result that shares the constant term with Noll and the logarithmic term with Fried and Hudgin, which yields little difference from both Noll's and Freischlad's results.¹⁰

Figure 2 shows that the Southwell geometry is superior to all the other geometries regarding the error propagation when t is small, but the Hudgin geometry tends to be

slightly superior to the Southwell geometry when the grid array size t becomes large (when $t > 30$).

3. FORMULATION OF THE ERROR PROPAGATION IN WAVEFRONT ESTIMATION WITH THE MATRIX METHOD

Given Eq. (1) and considering that in the case of zero noise Eq. (1) reduces to $\mathbf{S}_0 = \mathbf{C}\mathbf{W}_0$, the WFD error \mathbf{N} and the induced wavefront error $\boldsymbol{\varepsilon}$ should satisfy the equation

$$\mathbf{C}\boldsymbol{\varepsilon} = \mathbf{N}. \quad (12)$$

In the general case, the least-squares solution for Eq. (12) can be written as

$$\boldsymbol{\varepsilon} = \mathbf{C}^+\mathbf{N}, \quad (13)$$

where \mathbf{C}^+ is the Moore–Penrose pseudo-inverse of matrix \mathbf{C} . Given the relationship between the WFD vector and the wavefront slope vector provided by Eq. (2), similarly, the WFD error \mathbf{N} and the wavefront slope noise \mathbf{N}' should satisfy

$$\mathbf{N} = \mu\alpha\mathbf{M}\mathbf{N}'. \quad (14)$$

Starting with the expression of the mean variance given by Eq. (5), the Euclidian norm for a statistical wavefront error vector is defined as

$$\|\boldsymbol{\varepsilon}\|_2 \equiv \left(\sum_{j=1}^m \langle \boldsymbol{\varepsilon}_j^2 \rangle \right)^{1/2} = (\text{tr}\langle \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T \rangle)^{1/2}, \quad (15)$$

where the bracket $\langle \cdot \rangle$ denotes the averaging operator, $\text{tr}(\cdot)$ is the matrix trace, and

$$\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T = \mathbf{C}^+\mathbf{N}\mathbf{N}^T(\mathbf{C}^+)^T = \mathbf{C}^+\mu\alpha\mathbf{M}\mathbf{N}'\mathbf{N}'^T\mathbf{M}^T\mu\alpha(\mathbf{C}^+)^T, \quad (16)$$

where $\mathbf{N}'\mathbf{N}'^T$ is given by¹

$$\mathbf{N}'\mathbf{N}'^T = \begin{bmatrix} n_1^2 & n_1n_2 & \cdots & n_1n_m \\ n_2n_1 & n_2^2 & \cdots & n_2n_m \\ \vdots & \vdots & \ddots & \vdots \\ n_mn_1 & n_mn_2 & \cdots & n_m^2 \end{bmatrix}. \quad (17)$$

If we simply assume that the wavefront slope noise is random, independent, and has zero mean with the same variance σ_s^2 , the ensemble statistical average of the slope errors yields

$$\langle n_in_j \rangle = \sigma_s^2 \delta_{ij} = \begin{cases} 0, & \text{when } i \neq j \\ \sigma_s^2, & \text{when } i = j \end{cases}. \quad (18)$$

So the slope noise covariance matrix is given by

$$\langle \mathbf{N}'\mathbf{N}'^T \rangle = \sigma_s^2 \mathbf{I}, \quad (19)$$

where \mathbf{I} is the identity matrix. Therefore, Eq. (16) yields

$$\langle \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T \rangle = \mu^2\alpha^2\sigma_s^2\mathbf{C}^+\mathbf{M}\mathbf{M}^T(\mathbf{C}^+)^T. \quad (20)$$

Accordingly the mean variance of the wavefront estimation error σ_w^2 is given by

$$\sigma_w^2 = \frac{1}{m} \|\boldsymbol{\varepsilon}\|_2^2 = \frac{1}{m} \text{tr}\{\langle \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T \rangle\} = \frac{\mu^2\alpha^2\sigma_s^2}{m} \text{tr}[\mathbf{C}^+\mathbf{M}\mathbf{M}^T(\mathbf{C}^+)^T]. \quad (21)$$

Similarly,

$$\sigma_n^2 = \frac{1}{m} \|\mathbf{N}\|_2^2 = \frac{1}{m} \text{tr}\{\langle \mathbf{N}\mathbf{N}^T \rangle\} = \frac{\mu^2\alpha^2\sigma_s^2}{m} \text{tr}\langle \mathbf{M}\mathbf{M}^T \rangle. \quad (22)$$

With the definition of the error propagation coefficient given in Eq. (6), we can write

$$\eta = \frac{\sigma_w^2}{\sigma_n^2} = \frac{\text{tr}[\mathbf{C}^+\mathbf{M}(\mathbf{C}^+\mathbf{M})^T]}{\text{tr}[\mathbf{M}\mathbf{M}^T]}. \quad (23a)$$

This equation is useful for the computation of the error propagation coefficient. It can be noted that Eq. (23a) can also be written in the form of Frobenius norm¹¹ as

$$\eta = \frac{\|\mathbf{C}^+\mathbf{M}\|_F^2}{\|\mathbf{M}\|_F^2}. \quad (23b)$$

Equations (23) provide a generalized expression for the error propagation coefficient in zonal wavefront estimation. This formulation is good for both the slope measurements and the direct WFD measurements, because we can convert either the slope measurements or any other specific WFD measurements into the wavefront piston value difference between each neighboring grid points, which is actually the direct WFD measurement in the Hudgin geometry. Before we present the error propagation coefficient for each of the basic geometries in Section 5, the serial numbering scheme (SNS) adopted for wavefront estimation will be introduced in Section 4.

4. SERIAL NUMBERING SCHEME FOR WAVEFRONT ESTIMATION

The SNS, as adopted in the previous literature,^{1,2} makes wavefront estimation become a standardized process. Instead of numbering the estimation grid array with 2D coordinates (i, j) as previously provided,^{4–7} the grid array is indexed sequentially from 1 to m row by row as illustrated in Fig. 3.

With such a numbering mode, a regular and banded sparse estimation matrix is predefined, which resembles

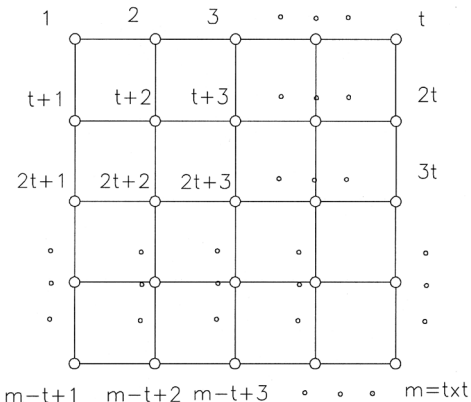


Fig. 3. Grid array with serial numbering scheme (SNS).

Hunt's matrix formulation.^{2,10} One main application of such a numbering scheme was to develop a high-precision iterative wavefront estimation algorithm for any shaped pupils.⁸ Recently we found that adopting this numbering scheme in the wavefront estimation provides new results in the analysis of error propagation.

5. WAVEFRONT DIFFERENCE BASED ERROR PROPAGATORS

In this section we apply the generalized formula of the error propagation coefficient for three basic estimation geometries: the Hudgin geometry, the Southwell geometry, and the Fried geometry. The wavefront estimation matrix for each of the geometries is established, and the eigenvalues for each estimation-related matrix are computed numerically, from which the formulas of the WFD-based error propagation coefficients are established.

A. Hudgin Geometry

In the Hudgin geometry, the wavefront slopes are measured at the midpoint between two neighboring grid points, and the WFD between each pair of neighboring grid points in the x and y directions can be estimated as

$$\begin{aligned} S_{x_{i+1,i}} &= g_{x_{i+1,i}} \cdot a, \\ S_{y_{i,i+t}} &= g_{y_{i,i+t}} \cdot a, \end{aligned} \quad (24)$$

where a is the pitch size of the sampling grid array, $g_{x_{i+1,i}}$ is the wavefront slope in the x direction at the midpoint between point i and point $i+1$, and $g_{y_{i,i+t}}$ is the wavefront slope in the y direction at the midpoint between point $i+t$ and point i . With the SNS the wavefront estimation equations can be written in the x direction as

$$w_{i+1} - w_i = S_{x_{i+1,i}} \quad i = 1, 2, \dots, m, \quad i \neq kt, \quad k \text{ integer}, \quad (25)$$

and in the y direction as

$$w_i - w_{i+t} = S_{y_{i,i+t}}, \quad i = 1, 2, \dots, m-t. \quad (26)$$

When written in matrix form, the estimation equation set for the Hudgin geometry is given by

$$\mathbf{H}\mathbf{W} = \mathbf{S}, \quad (27)$$

where \mathbf{H} is the wavefront estimation matrix, $\mathbf{S} = a\mathbf{G}$ is the WFD vector, and \mathbf{G} is the wavefront slope vector. $\mathbf{S} = a\mathbf{G}$ in comparison with Eq. (2) yields $\mu = 1$ and $\mathbf{M} = \mathbf{I}$, where \mathbf{I} is the identity matrix. From Eqs. (23a) and (23b), where \mathbf{C} equals \mathbf{H} , η takes the form

$$\eta_{\mathbf{H}} = \frac{\text{tr}[\mathbf{H}^+(\mathbf{H}^+)^T]}{\text{tr}[\mathbf{I}]} = \frac{1}{m} \text{tr}[(\mathbf{H}^T\mathbf{H})^+] = \frac{1}{m} \|\mathbf{H}^+\|_F^2. \quad (28)$$

In optical testing, it is the wavefront shape from a reference, not the absolute wavefront piston values, that is desirable. Thus it is common to set a point on the wavefront as the reference point, namely, the "zero point", at which the wavefront value is assigned to zero or another

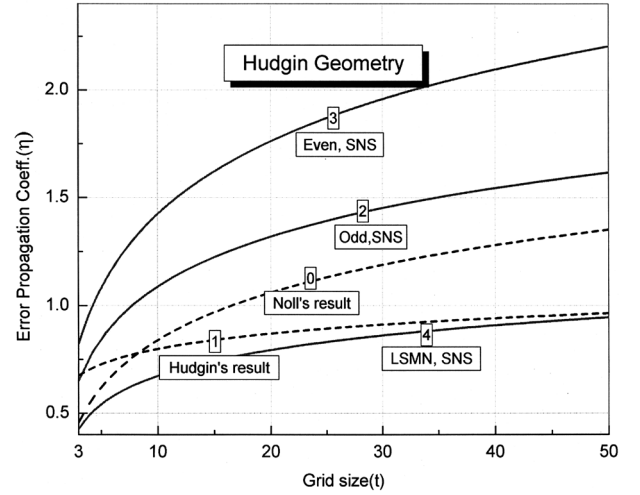


Fig. 4. WFD-based error propagators for the Hudgin geometry.

constant. Setting a "zero point" for the wavefront will not affect the estimated shape, as any piston value in the estimation can be subtracted.

The rank of matrix $\mathbf{H}^T\mathbf{H}$ is $\gamma = m - 1$.² Matrix $\mathbf{H}^T\mathbf{H}$ becomes of full rank ($\gamma = m$) after a "zero point" is set for the wavefront, and its generalized inverse is $\mathbf{H}^+ = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$. Therefore we have $\text{tr}[\mathbf{H}^+(\mathbf{H}^+)^T] = \text{tr}[(\mathbf{H}^T\mathbf{H})^{-1}]$. If the nonzero eigenvalues of matrix $(\mathbf{H}^T\mathbf{H})$ are $\lambda_{\mathbf{H},i} > 0$ ($i = 1, 2, \dots, \gamma$), then the nonzero eigenvalues of matrix $(\mathbf{H}^T\mathbf{H})^+$ are $\lambda_{\mathbf{H},i}^{-1} > 0$ ($i = 1, 2, \dots, \gamma$). The WFD-based error coefficient for the Hudgin geometry can be therefore expressed as

$$\eta_{\mathbf{H}} = \frac{1}{m} \text{tr}[(\mathbf{H}^T\mathbf{H})^+] = \frac{1}{m} \left(\sum_{i=1}^{\gamma} \lambda_{\mathbf{H},i}^{-1} \right). \quad (29)$$

This result is somewhat similar to the result obtained by Southwell for modal wavefront estimation.⁴

In this paper the numerical methods, such as the Jacobi and the singular value decomposition (SVD) methods, are employed to evaluate the matrix eigenvalues. The Jacobi method approximates the eigenvalues of symmetric matrices by reducing the off-diagonal elements to zero, while the SVD method approaches the eigenvalues by decomposing the matrix into a diagonal matrix multiplied by an orthonormal matrix on each side.

The eigenvalues of matrix $\mathbf{H}^T\mathbf{H}$ are sensitive to the position of the wavefront "zero point", the matrix dimension size, and the number parity of the matrix dimension. In this paper, the "zero point" mentioned is located at the center point of the estimation grid array for all geometries, at which the matrix has its smallest condition number.² Given that the coefficients in matrix $\mathbf{H}^T\mathbf{H}$ are predefined,^{2,10} the eigenvalues of matrix $\mathbf{H}^T\mathbf{H}$ for different grid sizes can be computed numerically. The simulation results are plotted in Fig. 4.

Fitting the numerical results (up to $t = 50$) in the least-squares sense, we obtain the error propagation coefficient for the Hudgin geometry as

$$\begin{aligned} \eta_{\mathbf{H},\text{odd}} &= 0.3797 + 0.3171 \ln(t - 0.6672) \\ &\approx 0.3222 + 0.3316 \ln(t) \quad (t \text{ is odd}), \end{aligned} \quad (30)$$

$$\begin{aligned} \eta_{\mathbf{H},\text{even}} &= 0.3294 + 0.479 \ln(t - 0.2136) \\ &\approx 0.3049 + 0.4856 \ln(t) \quad (t \text{ is even}). \end{aligned} \quad (31)$$

When no “zero point” is appointed, the rank of matrix $\mathbf{H}^T\mathbf{H}$ is $\gamma=m-1$. The solution space for the matrix equation has one degree of freedom. With the Moore–Penrose pseudo-inverse of matrix $\mathbf{H}^T\mathbf{H}$, we obtained the least-squares solution with minimum norm (LSMN), which is a least-squares solution that holds the same wavefront shape as that with a “zero point” set but has the minimum Euclidian norm of the wavefront values. In that case, we computed the eigenvalues of the estimation matrix with SVD, and obtained the simulation results with data up to $t=50$ as

$$\begin{aligned} \eta_{\mathbf{H},\text{LSMN}} &= 0.3252 + 0.1593 \ln(t - 1.1208) \\ &\approx 0.2605 + 0.1764 \ln(t). \end{aligned} \quad (32)$$

Figure 4 shows that the behavior of the error propagation is logarithmic dependent on the grid size as previously established in the literature.^{3,6} The error propagation curves for the SNS with the “zero point” given (curves 2 and 3) are higher than the curves of Noll’s result (curve 0) and the Hudgin result (curve 1). However, the error propagation curve for the SNS with the LSMN solution (curve 4) is the lowest of all, nevertheless, at the expense of extensive computations; moreover, it is approaching the Hudgin result when the grid size increases. For wavefront estimation with a “zero point” set, an odd-number grid size is preferable for the Hudgin geometry.

B. Southwell Geometry

In the Southwell geometry the slope measurements and the wavefront estimate are located at each grid point. As Southwell did,⁴ we adopted the average of the measured slopes over two neighboring grid points as an estimate of the slope at the midpoint. The wavefront estimation equation set can be obtained with the SNS in the x and y directions as

$$w_{i+1} - w_i = \frac{a}{2}(g_{x_i} + g_{x_{i+1}}), \quad (33)$$

where $i=1, 2, \dots, m-1$ ($i \neq kt$, k integer), g_{x_i} is the slope data in the x direction, and

$$w_i - w_{i+t} = \frac{a}{2}(g_{y_i} + g_{y_{i+t}}), \quad (34)$$

where $i=1, 2, \dots, m-t$, and g_{y_i} is the slope data in the y direction. In matrix form,

$$\mathbf{H}\mathbf{W} = \mathbf{S}_{\mathbf{H}}, \quad (35)$$

where matrix \mathbf{H} is the same wavefront estimation matrix as for the Hudgin geometry, and $\mathbf{S}_{\mathbf{H}}$ is the WFD vector, which is given by $\mathbf{S}_{\mathbf{H}} = \frac{1}{2}a\mathbf{C}_s\mathbf{G}$. Comparing $\mathbf{S}_{\mathbf{H}}$ with Eq. (2), we obtain $\mu=1/2$ and $\mathbf{M}=\mathbf{C}_s$. It should be noted that matrices \mathbf{C}_s and \mathbf{H} are closely related: Matrix \mathbf{C}_s can be obtained by changing the sign of the coefficient -1 in matrix \mathbf{H} .

According to Eq. (23), we have

$$\eta_{\mathbf{S}} = \frac{\text{tr}[\mathbf{H}^+\mathbf{C}_s(\mathbf{H}^+\mathbf{C}_s)^T]}{\text{tr}[\mathbf{C}_s\mathbf{C}_s^T]} = \frac{\text{tr}[\mathbf{C}_s^T(\mathbf{H}^+)^T\mathbf{H}^+\mathbf{C}_s]}{\text{tr}[\mathbf{H}\mathbf{H}^T]}, \quad (36)$$

where $\text{tr}[\mathbf{H}\mathbf{H}^T]=4t(t-1)=4(m-t)$. Given that $\mathbf{H}^+ = (\mathbf{H}^T\mathbf{H})^+\mathbf{H}^T$,¹² then the WFD-based error propagation coefficient for the Southwell geometry is expressed as

$$\eta_{\mathbf{S}} = \frac{\text{tr}[\mathbf{C}_s^T(\mathbf{H}^+)^T\mathbf{H}^+\mathbf{C}_s]}{4(m-t)} = \frac{\text{tr}[\mathbf{C}_s^T\mathbf{H}[(\mathbf{H}^T\mathbf{H})^+]^2\mathbf{H}^T\mathbf{C}_s]}{4(m-t)}. \quad (37)$$

Finally,

$$\eta_{\mathbf{S}} = \frac{1}{4(m-t)} \left(\sum_{i=1}^{\gamma} \lambda_{s,i} \right), \quad (38)$$

where $\lambda_{s,i}$ ($i=1, 2, \dots, \gamma$) are the nonzero eigenvalues of matrix $\mathbf{C}_s^T\mathbf{H}[(\mathbf{H}^T\mathbf{H})^+]^2\mathbf{H}^T\mathbf{C}_s$, and γ is the rank of this matrix.

When a “zero point” is set, $\mathbf{H}^T\mathbf{H}$ is invertible, and we have $\gamma=m$ and $\mathbf{H}^+ = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$. The simulation results are shown in Fig. 5. We least-squares fit the numerical results (up to $t=50$) and obtain the error propagation coefficients for the Southwell geometry as

$$\begin{aligned} \eta_{\mathbf{S},\text{odd}} &= 0.1489 + 0.2936 \ln(t - 0.06186) \\ &\approx 0.1428 + 0.2952 \ln(t) \quad (t \text{ is odd}), \end{aligned} \quad (39)$$

$$\begin{aligned} \eta_{\mathbf{S},\text{even}} &= 0.04941 + 0.4662 \ln(t + 2.7673) \\ &\approx 0.2861 + 0.41 \ln(t) \quad (t \text{ is even}). \end{aligned} \quad (40)$$

When no “zero point” is appointed, the LSMN solution is sought to make the Euclidian norm of the wavefront minimum. We computed the eigenvalues of the estimation matrix for this case with SVD up to $t=50$ and obtained the error propagation coefficient as

$$\begin{aligned} \eta_{\mathbf{S},\text{LSMN}} &= 0.205 + 0.1487 \ln(t + 0.2562) \approx 0.217 \\ &+ 0.1455 \ln(t). \end{aligned} \quad (41)$$

In Fig. 5, when a wavefront “zero point” is assigned, the error propagation curve for the SNS is above the curve of

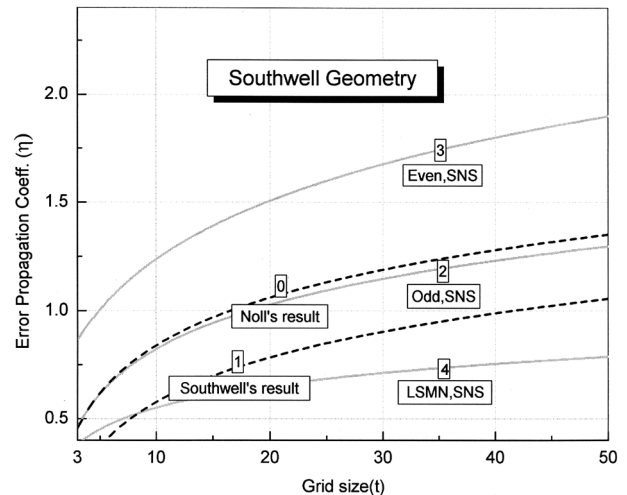


Fig. 5. WFD-based error propagators for the Southwell geometry.

the Noll's result (curve 0) for the even-number sizes of grid (curve 3), and it is equivalent to (or slightly below) it for the odd-number sizes of grid (curve 2). Although both of the above error propagators (curves 2 and 3) are shown to be worse than Southwell's result (curve 1), the error propagator of the LSMN solution (curve 4), when no "zero point" is set for the wavefront, is shown to be clearly better for the large grid size ($t > 8$). It is similar to the Hudgin geometry in that an odd-number size of grid is preferable for the Southwell geometry when a "zero point" is assigned.

C. Fried Geometry

In the Fried geometry, the grid for wavefront slope (or alternatively phase difference) measurements and the grid for wavefront estimation are interleaved by a distance of $a/2$ both in the x and in the y directions as shown in Fig. 1. Fried proposed that the phase difference across the (i, j) th aperture element in the x direction (or in the y direction) can be estimated as the average of the phase differences along its two parallel borders in the x direction (or in the y direction).⁷ To compare the work in this paper with Fried's result, a similar WFD estimation equation set as found in Fried's work was established. For the Fried geometry with the SNS, we have

$$w_{i+1} - w_i + w_{i+t+1} - w_{i+t} = 2s_{x_j}, \quad (42)$$

$$w_i + w_{i+1} - w_{i+t} - w_{i+t+1} = 2s_{y_j}, \quad (43)$$

where $i = 1, 2, \dots, m-t-1$, with $i \neq kt$ (k is an integer); $s_{x_j} = g_{x_j} a$ and $s_{y_j} = g_{y_j} a$ are the WFDs in the x and y directions; and $j = 1, 2, \dots, (t-1)^2$, where $(t-1)^2$ is the total number of the measurement elements in the grid array. Writing them in matrix form, we have

$$\mathbf{F}\mathbf{W} = \mathbf{S}, \quad (44)$$

where \mathbf{F} is a predefined matrix for wavefront estimation for the Fried geometry, $\mathbf{S} = 2a\mathbf{G}$ is the WFD vector, and \mathbf{G} is the wavefront slope vector. As defined in Eq. (2), we obtain $\mu = 2$ and $\mathbf{M} = \mathbf{I}$, where \mathbf{I} is the identity matrix. Given Eqs. (23), the WFD-based error propagation coefficient for the Fried geometry can be written as

$$\eta_{\mathbf{F}} = \frac{\text{tr}[\mathbf{F}^+(\mathbf{F}^+)^T]}{\text{tr}[\mathbf{I}\mathbf{I}^T]} = \frac{1}{m} \text{tr}[(\mathbf{F}^T\mathbf{F})^+]. \quad (45)$$

Herrmann showed that the rank of the normal matrix $\mathbf{F}^T\mathbf{F}$ is $m-2$.¹³ If the nonzero eigenvalues of $\mathbf{F}^T\mathbf{F}$ are $\lambda_{\mathbf{F},i} > 0$ ($i = 1, 2, \dots, \gamma$), where γ is the rank of matrix $\mathbf{F}^T\mathbf{F}$, then

$$\eta_{\mathbf{F}} = \frac{1}{m} \left(\sum_{i=1}^{\gamma} \lambda_{\mathbf{F},i}^{-1} \right). \quad (46)$$

When a grid point is frozen as the "zero point," the rank of matrix $\mathbf{F}^T\mathbf{F}$ becomes $\gamma = m - 1$, and the piston value term of the estimated wavefront is defined. However, there is still one degree of freedom in the least-squares solution space, called the waffle mode, at which the wavefront at the estimation grid points can be pushed up and down with the same value in a checkerboard pattern, while it still satisfies the WFD estimation equation set defined in

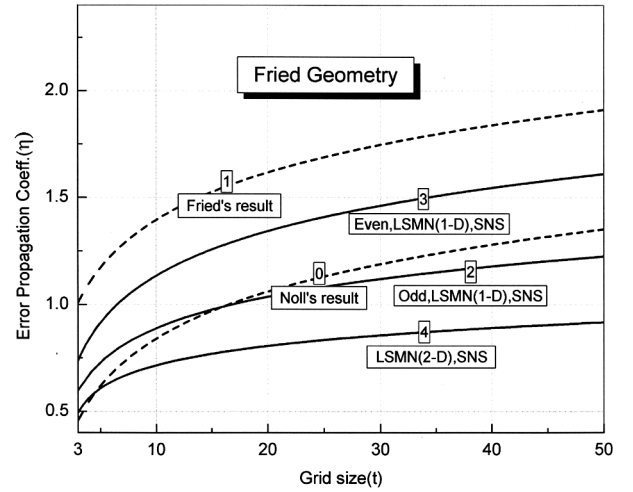


Fig. 6. WFD-based error propagators for the Fried geometry.

Eqs. (42) and (43). In other words, the Fried geometry cannot sense the waffle mode error of the wavefront,¹⁴ which is not desirable in optical testing.

Given a "zero point" for the wavefront, the LSMN solution is to seek a least-squares solution in a one-dimensional (1D) solution space that makes the Euclidian norm of the wavefront shape minimum. While no "zero point" is assigned for wavefront, we have $\gamma = m - 2$, and the LSMN solution is to seek a least-squares solution in a 2D solution space to minimize the Euclidian norm of the wavefront shape, which includes both the overall piston mode and the waffle mode errors. Theoretically, when the wavefront is flat, the LSMN solution can get rid of the waffle mode error. However, since the waffle mode error usually coexists with the wavefront shape, it is impossible to avoid the waffle mode error for the Fried geometry with the LSMN solution. The simulation results are shown in Fig. 6.

For a wavefront with a "zero point" set, the matrix eigenvalues for the LSMN (1-D) solution are computed up to $t = 50$, which yields the error propagation coefficients as

$$\begin{aligned} \eta_{\mathbf{F},\text{odd}} &= 0.4461 + 0.2 \ln(t - 0.8805) \\ &= 0.4146 + 0.207 \ln(t) \quad (t \text{ is odd}), \end{aligned} \quad (47)$$

$$\begin{aligned} \eta_{\mathbf{F},\text{even}} &= 0.4933 + 0.2866 \ln(t - 0.6547) \\ &= 0.4338 + 0.3022 \ln(t) \quad (t \text{ is even}). \end{aligned} \quad (48)$$

For a wavefront with no "zero point" set, the matrix eigenvalues for the LSMN (2-D) solution are computed up to $t = 50$, which yields the error propagation coefficients as

$$\begin{aligned} \eta_{\mathbf{F},\text{LSMN}} &= 0.475 + 0.114 \ln(t - 1.821) \approx 0.4076 \\ &+ 0.1303 \ln(t). \end{aligned} \quad (49)$$

Similar to the Hudgin and Southwell geometries, for a wavefront with a "zero point" set, an odd-number size of grid is preferable for wavefront estimation. The error propagators for the LSMN (1-D) solution (curves 2 and 3) are better than Fried's result (curve 1), worse than Noll's result (curve 0) for the even-number sizes of grid (curve 3), and better than the large odd-number sizes of grid (t

> 16) (curve 2). For a wavefront with no “zero point” set, the LSMN (2-D) solution offers the best error propagator for the Fried geometry.

It should be noticed that for wavefront estimation with the Fried geometry, it can either yield a LSMN solution on the full size of the estimation grid that suffers the waffle mode error, or yield a least-squares solution on a subsized estimation grid by discarding some of the wavefront boundary values (for example, the corner points). However, these solutions may not be acceptable in some optical testing cases.

D. Comparisons of the Error Propagators

To compare the error propagations in different wavefront estimation schemes, the simulation results for the Hudgin, Southwell, and Fried geometries are plotted in Fig. 7. The simulation results quantify how the error propagator depends on the parity of the wavefront estimation grid size, the wavefront estimation scheme, and the solution method used to solve the wavefront estimation.

Importantly, the odd-number grid sizes are shown to yield lower error propagation than the even-number grid sizes for all geometries. This can be explained by the following. If the grid size is odd, the “zero point” can be easily set at the geometrical center of the grid array so that the estimation scheme is symmetric and yields an estimation matrix with its minimum condition number,² which is steady and less sensitive to random noise disturbance. But if the grid size is even, there is no direct way to set a “zero point” at the geometrical center of the sampling grid; therefore, the estimation scheme is typically non-symmetric, which is liable to be affected by random noise disturbance.

For wavefront estimation with a “zero point” set, the Fried geometry offers the best error propagator for the even-number grid sizes, and the Hudgin geometry offers the worst for both even- and odd-number grid sizes. For the odd-number grid sizes, the Southwell ($t < 22$) and Fried geometries ($t > 22$) offer the best error propagators, and the error curve of the Fried geometry almost overlaps that of the Southwell geometry, which well confirms both Noll’s and Freischlad’s results. It is interesting to notice that the error propagator for the Hudgin geometry with the odd-number grid sizes is equivalent to that of the Fried geometry with the even-number grid sizes. When no “zero point” is set for the wavefront, the error curves for Fried and Hudgin geometries superpose together, and the Southwell geometry still offers the best error propagator. The detailed qualitative comparisons are listed in Table 1.

6. SLOPE-BASED ERROR PROPAGATORS

An alternative to Eq. (6) is to define the error propagation coefficient as the ratio of the mean variance of the wavefront estimation error to the variance of the wavefront slope error, i.e.,

$$\eta' = \frac{\sigma_w^2}{\sigma_s^2} = \frac{\mu^2 a^2}{m} \text{tr}[\mathbf{C}^+ \mathbf{M} (\mathbf{C}^+ \mathbf{M})^T], \quad (50)$$

or

$$\eta' = K \eta, \quad (51)$$

where η' is named the slope-based error propagation coefficient, η is the WFD-based error propagation coefficient defined in Eq. (6), and

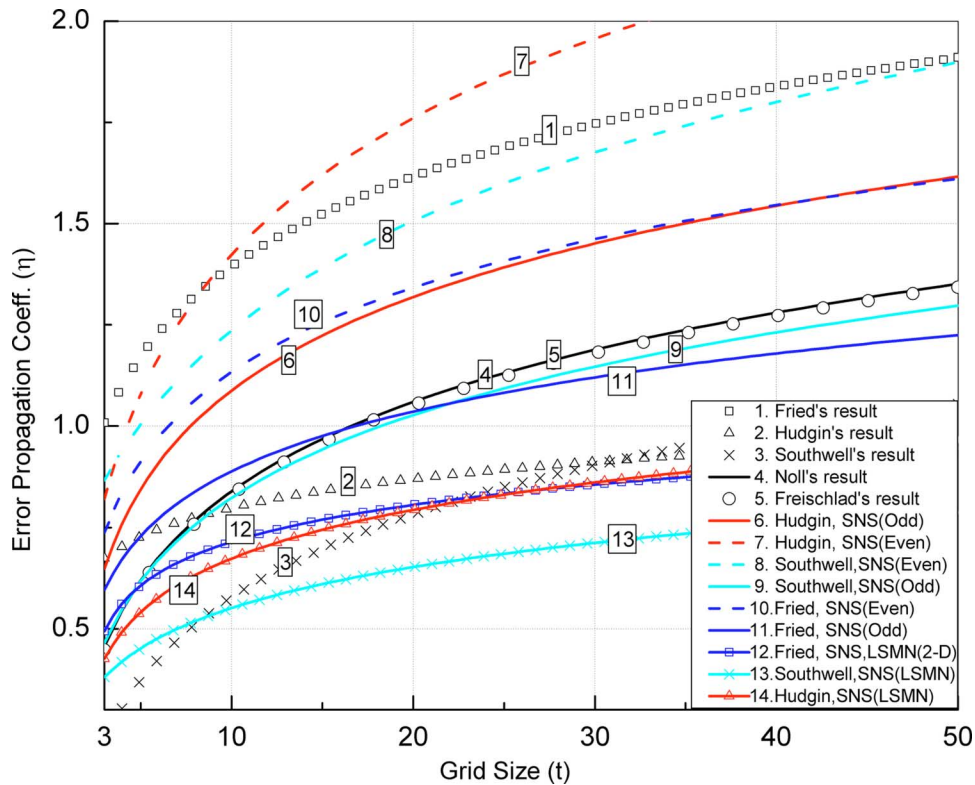


Fig. 7. (Color online) Comparison of the WFD-based error propagators.

Table 1. Qualitative Comparisons of the WFD-Based Error Propagators

With no zero-point set	Southwell (LSMN, 1D) (Curve 13)			Best ↑ ↓ Worst Error Propagator
	Hudgin (LSMN, 1D)(Curve 14) Fried (LSMN, 2D) (Curve 12)	Fried (LSMN, 2D) (Curve 12) Hudgin (LSMN, 1D) (Curve 14)		
	$t < 26$		$t > 26$	
With a zero-point set	Odd-number grid sizes		Even-number grid sizes	
	Southwell (Curve 9)	Fried (Curve 11)	Fried (Curve 10) Southwell (Curve 8) Hudgin (Curve 7)	
	Fried (Curve 11)	Southwell (Curve 9)		
	Hudgin (Curve 6)	Hudgin (Curve 6)		
$t < 22$		$t > 22$		

$$K = \frac{\mu^2 A_0 \text{tr}[\mathbf{M}\mathbf{M}^T]}{m}, \quad (52)$$

where $A_0 = a^2$ is the pitch area of the sampling grid. With this new definition, we can see the direct effect of the pitch area of the sampling grid on the error propagation. The larger the pitch area, the worse the error propagation. To compare the slope-based error propagators for different geometries, the slope-based error propagation coefficient is normalized with $A_0 = 1$. For the Hudgin geometry, it becomes

$$\eta_{\mathbf{H}'} = \frac{A_0}{m} \text{tr}[(\mathbf{H}^T \mathbf{H})^+] = K_{\mathbf{H}} \eta_{\mathbf{H}}, \quad (53)$$

where $K_{\mathbf{H}} = 1$, and $\eta_{\mathbf{H}}$ is the WFD-based error coefficient defined by Eq. (29). For the Southwell geometry, it becomes

$$\eta_{\mathbf{S}'} = \frac{A_0}{4m} \text{tr}[\mathbf{C}_s^T (\mathbf{H}^+)^T \mathbf{H}^+ \mathbf{C}_s] = K_{\mathbf{S}} \eta_{\mathbf{S}}, \quad (54)$$

where $K_{\mathbf{S}} = (m-t)/m$, and $\eta_{\mathbf{S}}$ is the WFD-based error coefficient defined by Eq. (37). For the Fried geometry, it becomes

$$\eta_{\mathbf{F}'} = \frac{4A_0}{m} \text{tr}[(\mathbf{F}^T \mathbf{F})^+] = K_{\mathbf{F}} \eta_{\mathbf{F}}, \quad (55)$$

where $K_{\mathbf{F}} = 4$, and $\eta_{\mathbf{F}}$ is the WFD-based error coefficient defined by Eq. (45).

From the WFD-based definition to the slope-based definition, except for the effect of the pitch size, the error propagator remains the same for the Hudgin geometry,

slightly improves for the Southwell geometry, and becomes four times worse for the Fried geometry. The slope-based coefficient may prove to be useful for the design of a slope-based optical testing system. Given the slope measurement noise and the wavefront error expected, for example, we can determine the pitch size required for the testing system.

The slope-based error propagators for the basic wavefront estimation geometries are plotted in Fig. 8. The Southwell geometry is shown to be better than the Hudgin geometry, and the Fried geometry performs the worst. As previously found, the parity of the sampling grid size also affects the error propagator: A grid with an odd-number size is preferable to a grid with an even-number size. For wavefront estimation with no “zero point” set, the LSMN solution offers lower error propagators in comparison with the wavefront estimation with a “zero point” set for all geometries, among which the Southwell geometry offers the best.

Generally speaking, for a slope-based wavefront estimation in optical testing, the Fried geometry is not preferred because of its interlaced sampling scheme, which yields inadequate slope information for estimating the wavefront over the entire sampling grid, and its high error propagation.

For both the slope-based and the WFD-based wavefront estimation, the Southwell geometry is preferred because it offers the best error propagator for the wavefront estimation over the entire testing domain. The performance of the Southwell geometry can be explained by the noise-averaging effort in the WFD estimates. When slope data (or WFD) are estimated at the midpoint between the neighboring grid points, the estimated slope noise is reduced by the slope data averaging.

7. CONCLUSION

In this paper, the error propagation coefficient for the WFD-based wavefront estimation was derived with the

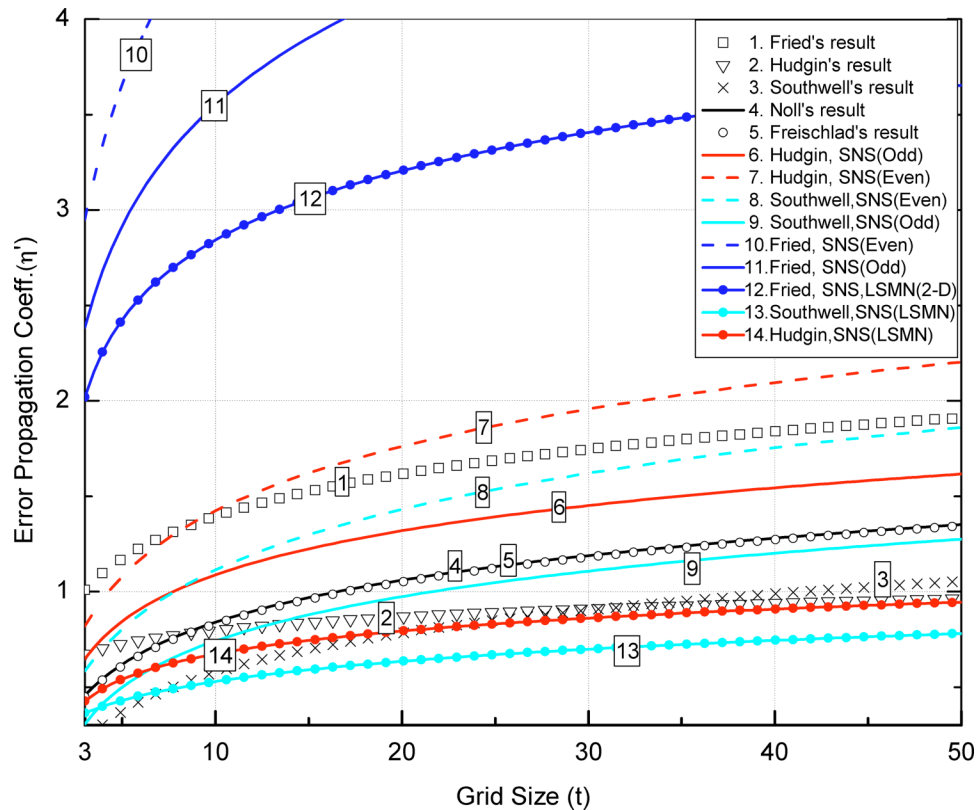


Fig. 8. (Color online) Comparison of the slope-based error propagators.

matrix formulation. We established the functions depicting the error propagation behaviors for three basic wavefront estimation geometries based on the wavefront estimation-related matrix eigenvalues.

The simulation results show that, for wavefront estimation with a “zero point”, the odd-number grid sizes are preferable to the even-number grid sizes. The Southwell geometry ($t < 22$) and the Fried geometry ($t > 22$) with the odd-number grid sizes offer the best error propagators if a “zero point” is set. The Southwell geometry offers the best error propagator for all grid sizes when no “zero point” is set.

Given the popularity of the slope-based wavefront sensors in optical testing, a slope-based error propagation coefficient is defined. The benefit of this metric is that, given the rms noise in slope data, the rms error in wavefront value can be directly estimated. Furthermore, this metric shows the direct dependence of the error propagation on the pitch size. Using this metric, the Southwell geometry offers the best error propagators for wavefront estimation with and without a “zero point” set. Therefore, in optical testing the Southwell geometry is highly desirable. The Fried geometry is not recommended because of its high error propagation and the ambiguity in wavefront shape estimation. Not limited with the three basic geometries listed in this paper, the generalized formulation presented here can be applied in any other estimation geometries, and the corresponding formulas for the error propagation coefficients can be established with numerical simulations.

ACKNOWLEDGMENTS

This work was funded in part by the National Science Foundation grant IIS/HCI-0307189. We thank Eric Clarkson for stimulating discussions about this research, and we also thank the anonymous reviewers for their excellent comments on this paper.

Weiyao Zou can be reached at wzou@creol.ucf.edu, and Jannick Rolland can be reached at rolland@odalab.ucf.edu.

REFERENCES

1. H. H. Barrett and K. J. Myers, *Foundations of Image Science*, Wiley Series in Pure and Applied Optics, B. E. Saleh, ed. (Wiley, 2003), pp. 368, 1036.
2. W. Zou and Z. Zhang, “Generalized wave-front reconstruction algorithm applied in a Shack–Hartmann test,” *Appl. Opt.* **39**, 250–268 (2000).
3. R. J. Noll, “Phase estimates from slope-type wave front sensors,” *J. Opt. Soc. Am.* **68**, 139–140 (1978).
4. W. H. Southwell, “Wave front estimation from wave front slope measurements,” *J. Opt. Soc. Am.* **70**, 998–1006 (1980).
5. K. Freischlad, “Wave front 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).
6. R. H. Hudgin, “Wave front reconstruction for compensated imaging,” *J. Opt. Soc. Am.* **67**, 375–378 (1977).
7. 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).

8. W. Zou and J. P. Rolland, "Iterative zonal wave-front estimation algorithm for optical testing with general-shaped pupils," *J. Opt. Soc. Am. A* **22**, 938–951 (2005).
9. K. Freischlad and C. L. Koliopoulos, "Modal estimation of a wave front from difference measurements using the discrete Fourier transform," *J. Opt. Soc. Am. A* **3**, 1852–1861 (1986).
10. B. R. Hunt, "Matrix formulation of the reconstruction of phase values from phase differences," *J. Opt. Soc. Am.* **69**, 393–399 (1979).
11. R. Kress, *Numerical Analysis*, Vol. 1818 Graduate Texts in Mathematics (Springer, 1998), pp. 22, 127.
12. J. R. Schott, *Matrix Analysis for Statistics* (Wiley, 1997), pp. 138, 171, 177.
13. J. Herrmann, "Least-square wave front errors of minimum norm," *J. Opt. Soc. Am.* **70**, 28–35 (1980).
14. M. D. Olier, "Sensing waffle in the Fried geometry," in *Adaptive Optical System Technologies*, D. Bonaccini and R. K. Tyson, eds., *Proc. SPIE* **3353**, 964–971 (1998).