Abstract: In this paper, we show that the error propagation coefficient in a slope-type zonal wavefront estimate can be expressed as a function of the eigenvalues of the wavefront reconstruction matrix. With a new indexing order for the grid array of Hudgin model, we show that the new optimized model with an odd-number of dimension size minimizes the error propagation in wavefront reconstruction, which is even better than Noll's theoretical expectation. In fact, Noll's theoretical result is equivalent to the optimized configuration with an even-number of dimension size. This finding illustrates that the new optimized reconstruction scheme with an odd-number dimension grid should be adopted in high-precision wavefront estimation.

OCIS codes: 000.3860, 010.7350, 220.4840

1. Error propagation in wavefront estimation

In optical testing, wavefront reconstruction is a technique that converts wavefront measurements into wavefront phases. Write this conversion in matrix form, which is

$$\mathbf{CW} = \mathbf{S}, \tag{1}$$

where C is the wavefront reconstruction matrix, and S is the wavefront difference vector, and W is the wavefront phase vector. Due to the measurement noise this matrix equation does not have an exact solution but weak solutions, which can be obtained by the least square method. The normal matrix equation is

$$\mathbf{C}^{\mathrm{T}}\mathbf{CW} = \mathbf{C}^{\mathrm{T}}\mathbf{S}. \tag{2}$$

This is a discretization form of Poisson equation

$$\nabla^2 \mathbf{W} = \nabla \mathbf{S} = \mathbf{f(y,z)} \tag{3}$$

The reconstructed wavefront errors are from two sources, one is the algorithm discretization errors, which depends on the basic reconstruction scheme we adopted; the other one is from the wavefront sensor measurements, such as the CCD centroiding errors. If we consider the wavefront measurement noise $\mathbf{N}$ only, we have $\mathbf{S} = \mathbf{S}_0 + \mathbf{N}$, where $\mathbf{S_0} = \nabla \mathbf{W}_0$ is the vector of theoretical wavefront difference, and $\mathbf{N} = \begin{pmatrix} \mathbf{n}_1 & \mathbf{n}_2 & \cdots & \mathbf{n_m} \end{pmatrix}^{\mathbf{T}}$ is the vector of measurement errors in the wavefront difference. Suppose the vector of wavefront error induced from the wavefront slope errors is $\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_{\mathbf{m}} \end{pmatrix}^{\mathbf{T}}$, then we have

$$\mathbf{C^T C \boldsymbol{\varepsilon}} = \mathbf{C^T N} \tag{4}$$

If $\mathbf{C^T C}$ is invertible, then we have

$$\boldsymbol{\varepsilon} = \mathbf{(C^T C)^{-1} C^T N} \tag{5}$$

Therefore,

$$\boldsymbol{\varepsilon\varepsilon}^{\mathbf{T}} = \mathbf{(C^T C)^{-1} C^T N N^T C (C^T C)^{-1}}. \tag{6}$$

Where

$$\boldsymbol{N}^{\mathbf{T}}\boldsymbol{N} = \begin{bmatrix} \boldsymbol{n}_1^2 & \boldsymbol{n}_1\boldsymbol{n}_2 & \cdots & \boldsymbol{n}_1\boldsymbol{n}_m \\ \boldsymbol{n}_2\boldsymbol{n}_1 & \boldsymbol{n}_2^2 & \cdots & \boldsymbol{n}_2\boldsymbol{n}_m \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{n}_m\boldsymbol{n}_1 & \boldsymbol{n}_m\boldsymbol{n}_2 & \cdots & \boldsymbol{n}_m^2 \end{bmatrix} \tag{7}$$

If we assume that the wavefront-difference errors are independent, uncorrelated, of the same variance σ_n^2 with zero mean, the ensemble statistical average of the slope errors provides [1]

$$\left\langle \boldsymbol{n}_i \boldsymbol{n}_j \right\rangle = \sigma_{\boldsymbol{n}}^2 \delta_{\boldsymbol{ij}} = \begin{cases} 0\text{, when i} \neq \text{j} \\ \sigma_{\text{n}}^2\text{, when i} = \text{j} \end{cases} \tag{8}$$

or

$$\left\langle \boldsymbol{N}^{\mathrm{T}} \boldsymbol{N} \right\rangle = \sigma_{\mathrm{n}}^2 \mathbf{I}, \tag{9}$$

Then Eq.(6) becomes

$$\boldsymbol{\varepsilon\varepsilon}^{\mathbf{T}} = \boldsymbol{\sigma}_{\boldsymbol{n}}^{\mathbf{2}} (\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1} \tag{10}$$

Taking Frobenius norms in both sides of the Eq.(10). For the left side, we have

$$\left\| \boldsymbol{\varepsilon\varepsilon}^{\mathrm{T}} \right\|_{\boldsymbol{F}} = \left(\sum_{\boldsymbol{k}=1}^{\boldsymbol{m}} \sum_{\boldsymbol{j}=1}^{\boldsymbol{m}} \left| \varepsilon_{\boldsymbol{k}} \varepsilon_{\boldsymbol{j}} \right|^2 \right)^{1/2} = \left(\sum_{\mathbf{k}=1}^{\mathbf{m}} \left| \varepsilon_{\mathbf{k}} \right|^2 \cdot \sum_{\mathbf{j}=1}^{\mathbf{m}} \left| \varepsilon_{\mathbf{j}} \right|^2 \right)^{1/2} = \sum_{\mathbf{j}=1}^{\mathbf{m}} \left| \varepsilon_{\mathbf{j}} \right|^2 = \left\| \boldsymbol{\varepsilon} \right\|_2^2, \tag{11}$$

where

$$\left\| \boldsymbol{\varepsilon} \right\|_2 = \left(\sum_{\mathbf{j}=1}^{\mathbf{m}} \left| \varepsilon_{\mathbf{j}} \right|^2 \right)^{1/2} \tag{12}$$

is the Euclidian norm of vector $\boldsymbol{\varepsilon}$. For the right side Eq.(10), we have

$$\left\| \sigma_{\mathrm{n}}^2 (\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1} \right\|_{\mathrm{F}} = \sigma_{\mathrm{n}}^2 \left\| (\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1} \right\|_{\mathrm{F}} = \sigma_{\mathrm{n}}^2 \left(\mathbf{tr} \left\{ \left(\mathbf{C}^{\mathrm{T}}\mathbf{C} \right)^{-1} \left[\left(\mathbf{C}^{\mathrm{T}}\mathbf{C} \right)^{-1} \right]^{\mathrm{T}} \right\} \right)^{1/2} \tag{13}$$

Therefore,

$$\left\| \boldsymbol{\varepsilon} \right\|_{\mathbf{2}}^{\mathbf{2}} = \sigma_{\mathrm{n}}^2 \left\{ \mathbf{tr} \left[\left(\mathbf{C}^{\mathrm{T}}\mathbf{C} \right)^{-2} \right] \right\}^{1/2}, \tag{14}$$

where $\mathbf{tr}\left[\left(\mathbf{C}^{\mathrm{T}}\mathbf{C} \right)^{-2} \right]$ is the trace of matrix $\left(\mathbf{C}^{\mathrm{T}}\mathbf{C} \right)^{-2}$. If an eigenvalue of $\mathbf{C}^{\mathrm{T}}\mathbf{C}$ is λ, then

$$\mathbf{C}^{\mathrm{T}}\mathbf{CX} = \boldsymbol{\lambda}\mathbf{X} \tag{15}$$

where X is called an eigenvector for eigenvalue λ. Obviously,

$$\left(\mathbf{C}^{\mathbf{T}}\mathbf{C} \right)^{-2} \mathbf{X} = \lambda^{-2} \mathbf{X}, \tag{16}$$

and the eigenvalue of $\left(\mathbf{C}^{\mathbf{T}}\mathbf{C} \right)^{-2}$ is λ^{-2}. We know that

$$\mathrm{tr}\left(\left(\mathbf{C}^{\mathrm{T}}\mathbf{C}\right)^{-2}\right)=\sum_{\mathrm{i}=1}^{\mathrm{m}}\lambda_{\mathrm{i}}^{-2} \quad , \tag{17}$$

so we have

$$\|\boldsymbol{\varepsilon}\|_2^2=\sigma_{\mathrm{n}}^2\left(\sum_{i=1}^{m}\lambda_i^{-2}\right)^{1/2} . \tag{18}$$

If we define the error propagation coefficient η as the ratio of the mean variance of the wavefront reconstruction error, given by

$$\sigma_w^2=\|\boldsymbol{\varepsilon}\|_2^2/\boldsymbol{m} \tag{19}$$

to the variance of the wavefront difference measurement error σ_{n}^2 [2,3,4], i.e.

$$\eta=\frac{\sigma_w^2}{\sigma_{\mathrm{n}}^2}=\frac{1}{\boldsymbol{m}}\left(\sum_{i=1}^{m}\lambda_i^{-2}\right)^{1/2} . \tag{20}$$

This formula tells that the error propagation in zonal wavefront reconstruction is determined by the eigenvalues and the matrix dimension size of the wavefront reconstruction matrix only. Once the eigenvalues of the reconstruction matrix are determined, the error propagation coefficient is known. Now the problem is how to find the eigenvalues of the wavefront reconstruction matrix. Since the normal equation matrices are symmetrical matrices, we can employ the classical Jacobi method[5] , for example, to evaluate the eigenvalues.

2. Wavefront reconstruction schemes

2.1 Review of the error propagation estimations

Wavefront reconstruction schemes can be categorized into three reconstruction geometries: the Hudgin model,[6] the Southwell model [3], and the Fried model.[1] The discretization errors of this wavefront reconstruction geometry are not our concern in this paper[7] . An analysis of error propagation from the wavefront slope measurement will be detailed in this part.

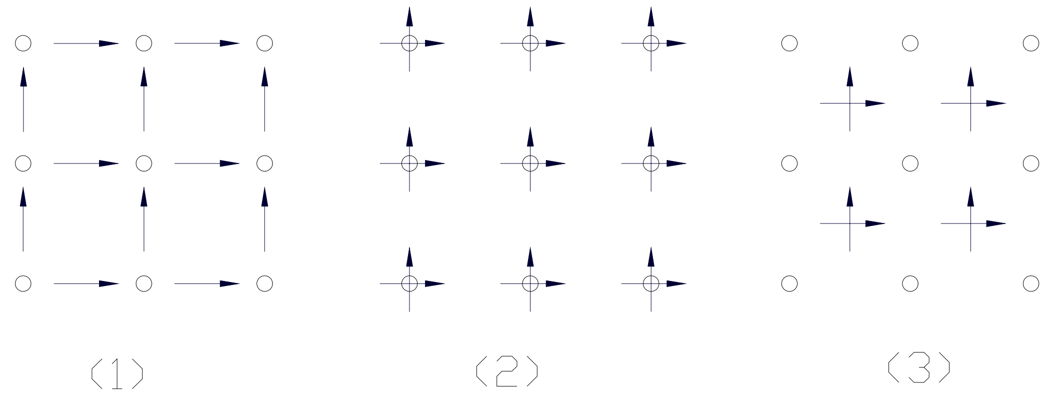

Fig. 1. The Wavefront reconstruction schemes

(1) Hudgin model (2) Southwell model (3) Fried model.

The error propagation coefficient in wavefront reconstruction is defined as the ratio of mean-square phase errors to the variance of the phase difference measurements. The simulation results for the three reconstruction geometries are shown as below:

Fried model[1]

$$\eta_{Fried} = 0.6558 + 0.3206\ln(t)\,, \tag{21}$$

Hudgin model[6]

$$\eta_{Hugin} = 0.561 + 0.103\ln(t) \tag{22}$$

Southwell model[3]

$$\eta_{Southwell} = -0.10447 + 0.2963\ln(t)\,. \tag{23}$$

The Southwell geometry is characterized by taking the wavefront slope measurements and wavefront phase estimations at the same nodes, and it has been demonstrated to have the lowest error propagation in wavefront reconstruction[3], as shown in Figure 2.

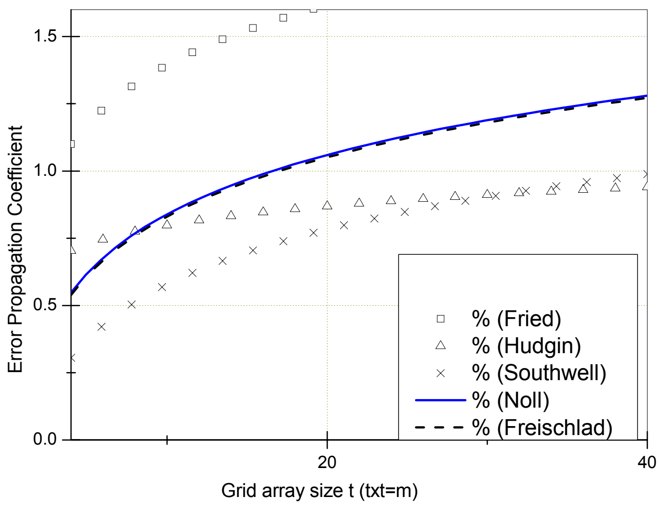

Fig. 2 Previous plots of error propagations

Noll proved analytically that a Ln(t) dependent theoretical formula for error propagation for a square aperture can be expressed as[8]

$$\eta_{Noll} = 0.1072 + 0.318\ln(t) . \tag{24}$$

Based on the Hudgin model and FFT-based algorithm, Freischlad confirmed Noll's result with[9]

$$\eta_{Freischlad} = 0.09753 + \ln(t)/\pi . \tag{25}$$

As shown in Figure 2, regarding the error propagation Hudgin model tends to overcome Southwell model when the grid array size t becomes large, but the Southwell model is superior to all the other models when t is small.

2.2 Error propagation in a new optimized reconstruction scheme

Recently we found that if we take the indexing mode into account in wavefront reconstruction, and index the grid array sequentially from 1 to m row by row as illustrated in Fig. 3, the reconstruction matrix will become extremely regular and sparse.[7]

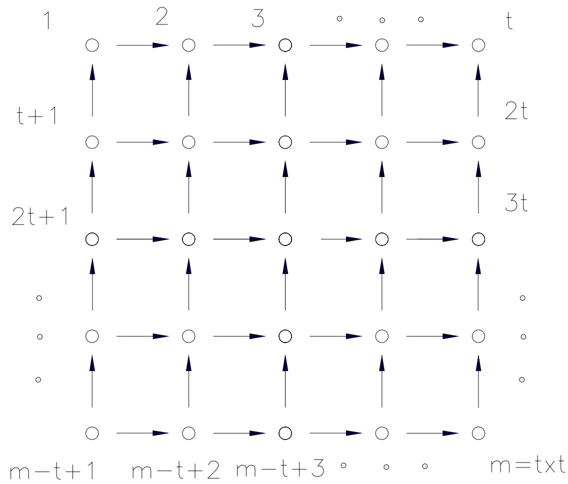

Fig. 3. Reconstruction grid (Hudgin) with the new indexing sequence

Take the Hudgin geometry as example, we have

$$\boldsymbol{w}_{i+1} - \boldsymbol{w}_i = \boldsymbol{s}_{y_{i+1,i}}, \mathrm{i} = 1,2,..., \mathrm{m, but\ i} \neq \mathrm{kt, k\ is\ an\ interger} \tag{26}$$

$$\boldsymbol{w}_i - \boldsymbol{w}_{i+t} = \boldsymbol{s}_{z_{i+1,i}}, \mathrm{i} = 1, 2,..\mathrm{m - t} \tag{27}$$

Where $\boldsymbol{s}_{y_{i+1,i}}$ and $\boldsymbol{s}_{z_{i+1,i}}$ are the wavefront differences in y- and z- directions, which are the slopes at the mid-point between two neighbor grids times the grids interval. Write them in matrix form, we have

$$
\begin{bmatrix}
-1 \quad 1 & & & & & & & & & & \\
 & -1 \quad 1 & & & & & & & & & \\
 & & \ddots \quad \ddots & & & & & & & & \\
 & & & -1 \quad 1 & & & & & & & \\
 & & & & -1 \quad 1 & & & & & & \\
 & & & & & -1 \quad 1 & & & & & \\
 & & & & & & \ddots \quad \ddots & & & & \\
 & & & & & & & -1 \quad 1 & & & \\
 & & & & & & & & \ddots \quad \ddots & & \\
 & & & & & & & & & -1 \quad 1 & \\
 & & & & & & & & & & -1 \quad 1 \\
 & & & & & & & & & & \ddots \quad \ddots \\
 & & & & & & & & & & -1 \quad 1 \\
1 \quad 0 \quad 0 \quad \cdots \quad 0 \quad -1 & & & & & & & & & & \\
 & 1 \quad 0 \quad 0 \quad \cdots \quad 0 \quad -1 & & & & & & & & & \\
 & & \ddots \quad \ddots \quad \ddots \quad \ddots \quad \ddots & & & & & & & & \\
 & & & & & & \ddots \quad \ddots \quad \ddots \quad \ddots \quad \ddots & & & & \\
 & & & & & & & 1 \quad 0 \quad 0 \quad \cdots \quad 0 \quad -1 & & & \\
 & & & & & & & & 1 \quad 0 \quad 0 \quad \cdots \quad 0 \quad -1 & &
\end{bmatrix}
\begin{bmatrix}
\boldsymbol{w}_1 \\ \boldsymbol{w}_2 \\ \vdots \\ \boldsymbol{w}_t \\ \boldsymbol{w}_{t+1} \\ \boldsymbol{w}_{t+2} \\ \vdots \\ \boldsymbol{w}_{2t} \\ \vdots \\ \boldsymbol{w}_{m-2t+1} \\ \boldsymbol{w}_{m-2t+2} \\ \vdots \\ \boldsymbol{w}_{m-t} \\ \boldsymbol{w}_{m-t+1} \\ \vdots \\ \boldsymbol{w}_{m-1} \\ \boldsymbol{w}_m
\end{bmatrix}
=
\begin{bmatrix}
\boldsymbol{s}_{y,2,1} \\ \boldsymbol{s}_{y,3,2} \\ \vdots \\ \boldsymbol{s}_{y,t,t+1} \\ \boldsymbol{s}_{y,t+2,t+1} \\ \vdots \\ \vdots \\ \boldsymbol{s}_{y,m,m-1} \\ \boldsymbol{s}_{z,1,t+1} \\ \boldsymbol{s}_{z,2,t+2} \\ \vdots \\ \boldsymbol{s}_{z,t,2t} \\ \vdots \\ \boldsymbol{s}_{z,i,i+t} \\ \vdots \\ \boldsymbol{s}_{z,m-t-1,m-1} \\ \boldsymbol{s}_{z,m-t,m}
\end{bmatrix}
\tag{28}
$$

or

$$\boldsymbol{A}\mathbf{W} = \mathbf{S} \tag{29}$$

All the coefficients in matrix $\boldsymbol{A}$ is predetermined in this optimized reconstruction scheme, and it is easy to derive its normal matrix equation[7], which is

$$\boldsymbol{A}^T \boldsymbol{A}\mathbf{W} = \boldsymbol{A}^T \mathbf{S} \tag{30}$$

where

$$
\boldsymbol{A}^T \boldsymbol{A} =
\begin{bmatrix}
\boldsymbol{E}_1 & -\boldsymbol{I} & & & \\
-\boldsymbol{I} & \boldsymbol{E}_2 & -\boldsymbol{I} & & \\
 & \ddots & \ddots & \ddots & \\
 & & -\boldsymbol{I} & \boldsymbol{E}_2 & -\boldsymbol{I} \\
 & & & -\boldsymbol{I} & \boldsymbol{E}_1
\end{bmatrix}_{m \times m}
\tag{31}
$$

and

$$
\mathbf{E}_1 = \begin{bmatrix} 2 & -1 & & & \\ -1 & 3 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 3 & -1 \\ & & & -1 & 2 \end{bmatrix}_{t\times t}, \tag{32}
$$

$$
\mathbf{E}_2 = \begin{bmatrix} 3 & -1 & & & \\ -1 & 4 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 4 & -1 \\ & & & -1 & 3 \end{bmatrix}_{t\times t}, \tag{33}
$$

$$
-\boldsymbol{I} = \begin{bmatrix} -1 & & \\ & \ddots & \\ & & -1 \end{bmatrix}_{t\times t} \tag{34}
$$

We can prove that rank(A^TA)=m-1.[7] The eigenvalues of matrix $\boldsymbol{A}^T\boldsymbol{A}$ are sensitive with the position of wavefront zero-point, the matrix dimension size, and even the number parity of the matrix dimension. We demonstrated that the matrix has its smallest condition number when the zero-point is set at the center of the reconstructed wavefront. [7] When the zero-point is determined, matrix $\boldsymbol{A}^T\boldsymbol{A}$ will be positive, and we can employ the classical Jacobi method to compute its eigenvalues. We set the zero-point at the center of the wavefront, and employ the eigenvalue-based formula Eq.(20) we derived to evaluate the curve of the error propagation coefficient versus the grid size. The results are shown in Fig. 4.

We fit the numerical results with least-squares method, and find that the error coefficients of the wavefront reconstruction matrices can be expressed by

$$
\eta_{odd} = 0.14008 + 0.1379\, ln(\boldsymbol{t}), \text{when t is odd} \tag{35}
$$

and

$$\eta_{even} = 0.18123 + 0.27098\,ln(t), \text{when t is even} \tag{36}$$

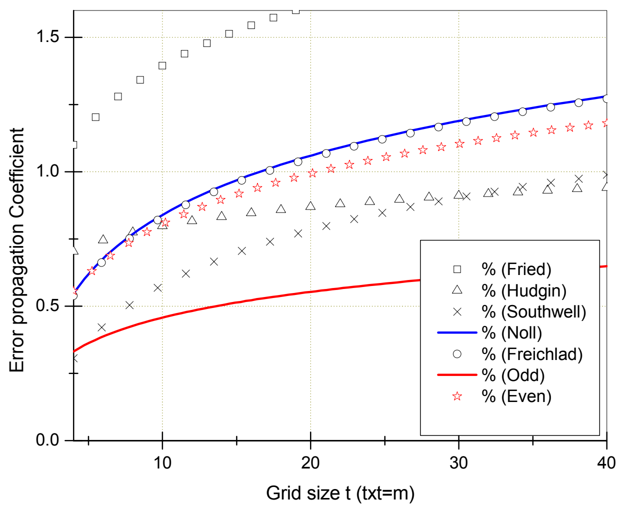

Fig. 4 Comparisons of error propagation with different reconstruction schemes

If we define the condition number of $A^{\mathrm{T}}A$ as

$$cond(A^{\mathrm{T}}A) = \frac{|\lambda_{max}|}{|\lambda_{min}|} \tag{37}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the maximum and minimum of the eigenvalues of matrix $A^{\mathrm{T}}A$, respectively. Then the error propagation of parity dependence is also reflected in the curve of the matrix condition numbers as shown in the Fig.5. By making a least squares fitting of this curve, we obtained the condition number of the wavefront estimation matrix by

$$cond_2(\mathrm{C}^{\mathrm{T}}\mathrm{C}) = \begin{cases} -243.442 + 150.870e^{\frac{t}{7.518}}, & t \text{ is odd} \\ -355.157 + 223.750e^{\frac{t}{6.83}}. & t \text{ is even} \end{cases}. \tag{38}$$

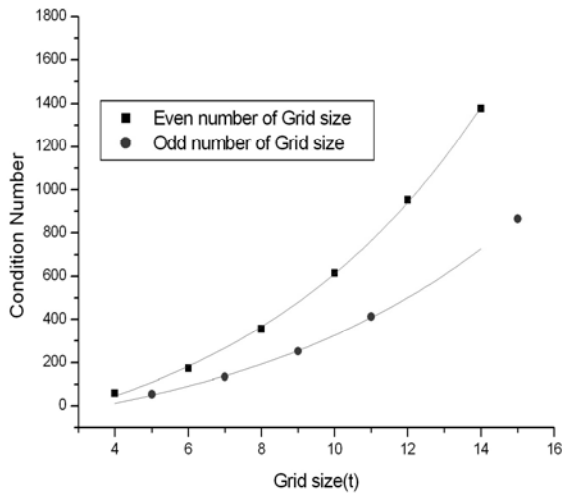

Fig. 5. Normal matrix condition number versus grid dimension size

Comparing these results, we can see that the error coefficients can be even smaller than that of Noll's theoretical result, and even better than Southwell's result. The new optimized reconstruction model with an odd-number of grid dimension performs best regarding the error propagation, and Noll's result described the case when the new optimized model with an even-number of grid dimension.

3. Conclusion

In this paper, we derived the formula for error propagation, which was found to be a function of the eigenvalues of the wavefront reconstruction matrix. With this formula, we evaluated the error propagation of the Hudgin model with the new indexing mode, and found that the new optimized reconstruction model performs the best of all when the dimension number of the grid is odd. This result illustrates that the new optimized reconstruction model should be adopted in wavefront reconstruction, and an odd number of the sampling grid array is preferable to its closest even number.

ACKNOWLEDGEMENTS: This work was funded in part by the National Science Foundation.

REFERENCES

1. Davis L. Fried, "Least-square fitting a wavefront distortion estimate to an array of phase-difference measurements", J. Opt. Soc. AM., Vol. 67, No. 3, pp370-375, 1977.
2. Robert J. Noll, "Phase estimates from slope-type wavefront sensors", J. Opt. Soc. Am., Vol. 68, No. 1, pp139-140, 1978.
3. W. H. Southwell, "Wavefront estimation from wavefront slope measurements", J. Opt. Soc. Am, Vol 70, No.8, pp998-1006, 1980.
4. Klaus 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 Vol. 1755, pp212-218, 1992.
5. Rainer Kress, "Numerical analysis", Graduate texts in mathematics, p127, Springer, 1998.
6. Richard H. Hudgin, "Wavefront reconstruction for compensated imaging", J. Opt. Soc. AM., Vol. 67, No. 3, pp375-378, 1977
7. W. Zou and Z. Zhang, Generalized wavefront reconstruction algorithm applied in a Shack-Hartmann test, Appl. Opt., Vol. 39, No.2, ,pp250-268, 2000.
8. Robert J. Noll, "Phase estimates from slope-type wavefront sensors", J. Opt. Soc. Am., Vol. 68, No. 1, pp139-140, 1978.
9. Klaus 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 Vol. 1755, pp212-218, 1992.