

Wavefront reconstruction from its gradients

Amos Talmi

Timi Technologies Ltd, Ramat Hashofet 19238, Israel

Erez N. Ribak

Department of Physics, Technion-Israel Institute of Technology, Haifa 32000, Israel

Received April 22, 2005; accepted June 5, 2005

Wavefronts reconstructed from measured gradients are composed of a straightforward integration of the measured data, plus a correction term that disappears when there are no measurement errors. For regions of any shape, this term is a solution of Poisson's equation with Dirichlet conditions ($V=0$ on the boundaries). We show that for rectangular regions, the correct solution is not a periodic one, but one expressed with Fourier cosine series. The correct solution has a lower variance than the periodic Fourier transform solution. Similar formulas exist for a circular region with obscuration. We present a near-optimal solution that is much faster than fast-Fourier-transform methods. By use of diagonal multigrid methods, a single iteration brings the correction term to within a standard deviation of 0.08, two iterations, to within 0.0064, etc. © 2006 Optical Society of America

OCIS codes: 100.0100, 100.5070.

1. INTRODUCTION

We wish to solve efficiently the problem of reconstructing a wavefront from signals obtained with a wavefront sensor, a device that usually measures the wavefront itself, its gradient components, or its curvature. Such sensors are required for many applications to obtain physical quantities such as density variations or temperature gradients, for optical workshop testing,¹ adaptive optics,^{2,3} and more. Specifically, we are interested in the popular class of gradient sensors, such as shearing interferometers and Hartmann or Hartmann-Shack sensors. The reconstruction methods also apply to other fields, such as projecting fringes on an object to find its shape, or finding strains in fabrics that show as gradients in the fabric pattern density. Astronomical telescopes equipped with adaptive optics use deformable mirrors and Hartmann-Shack sensors, where the commands to the mirrors are processed in two stages. In the first stage, the wavefront phase errors are calculated from the measured gradients;^{2,3} in the second, the corrections to produce that phase are found. All of these applications reconstruct the property—wavefront, object shape, fabric density—from its measured slopes or shear maps.^{4,5} In a previous paper⁶ we described an efficient method to obtain the two orthogonal components of the wavefront slope from grids of Hartmann-Shack spots. Now we investigate the problem of reconstructing the phase from these gradients.

Reconstructing the wavefront phase from measurements of its slopes is a simple and well-understood linear algebra problem^{7–20} with a straightforward solution. One has to solve a set of $2N_v$ linear equations/conditions with only N_v variables, where N_v is the number of data points. These N_v variables are the values of the phase at those points. Each measured x slope produces a condition; given N_v such measurements, this makes N_v conditions. The measured y slope produces another set of N_v conditions.

Those $2N_v$ conditions should be satisfied simultaneously with only N_v values of phase. To be precise, boundary points lack some neighbors, so the number of conditions is slightly lower than $2N_v$. One possible solution is a best fit, one that minimizes the sum of deviations between the wavefront derivatives and the measured slopes. The matrix of coefficients is sparse and well behaved. The practical problem is the computation load or the efficiency of the algorithm.

We examine the problem of a large shear map containing, say, 1000×1000 points. Solving a best-fit problem with 10^6 variables tasks the resources of a desk computer and even of a real-time computer. In adaptive optics for very large telescopes, the number of points in each gradient map is large and corrections must be applied within milliseconds. In industrial applications such as quality control of fabrics, defects that are identified as phase abnormalities must be detected and corrected in real time. Such applications have large maps and short times, requiring efficient algorithms. This problem was tackled early on in adaptive optics^{7–12} but with future large telescopes, new solutions are proposed.^{13–15} Sparse matrix and multigrid methods were also proposed.^{16–18} The minimal-variance problem may be solved directly using an inverse matrix, an approach that takes $O(N_v^2)$ operations. It might be solved using sparse-matrix methods, with $O(N_v^{3/2})$ operations. For rectangular geometries, it might be solved using fast Fourier transforms (FFTs),^{11–13} with $O(N_v \log_2 N_v)$ operations. Computationally, the FFT is the fastest approach, requiring roughly thrice the FFTs of N_v items—two forward FFTs for the two slope components, and an inverse FFT for the phase itself.

In order to use digital Fourier transforms and specifically FFT, the problem must be extended into the rigid geometry of a rectangular region and a periodic solution. The method of extending the data outside the measured

$$\tilde{\rho}(q_x, q_y) = \frac{4}{NM} \sum_{n=1}^{N-1} \sum_{m=1}^{M-1} \rho(\hat{n}, \hat{m}) \sin\left(\frac{nq_x\pi}{N}\right) \sin\left(\frac{mq_y\pi}{M}\right), \quad (22)$$

$$\tilde{V}(q_x, q_y) = \frac{4}{NM} \sum_{n=1}^{N-1} \sum_{m=1}^{M-1} V(\hat{n}, \hat{m}) \sin\left(\frac{nq_x\pi}{N}\right) \sin\left(\frac{mq_y\pi}{M}\right). \quad (23)$$

Introducing the shorthand $Q_x = \pi q_x / N$ and $Q_y = \pi q_y / M$ we rewrite Eq. (15) in Fourier terms:

$$\begin{aligned} (\Delta_x^2 + \Delta_y^2) \tilde{V}(q_x, q_y) &= -4 \left(\sin^2 \frac{Q_x}{2} + \sin^2 \frac{Q_y}{2} \right) \tilde{V}(q_x, q_y) \\ &= -\tilde{\rho}(q_x, q_y). \end{aligned} \quad (24)$$

Note that, by definition, neither V nor ρ have $q_x=0$ or $q_y=0$ components. We solve for V :

$$\tilde{V}(q_x, q_y) = \frac{\tilde{\rho}(q_x, q_y)}{T^2(q_x) + T^2(q_y)}, \quad q_x, q_y > 0, \quad (25)$$

where $T(q) = 2 \sin(Q/2)$ for either x or y . The correction terms are

$$C_x(q_x, q_y) = -T(q_y) \tilde{V}(q_x, q_y), \quad \tilde{C}_y(q_x, q_y) = T(q_x) \tilde{V}(q_x, q_y). \quad (26)$$

Thus, the steps of the solution are

1. Calculate $\rho = \Delta_x S_y - \Delta_y S_x$ [Eq. (13)].
2. Transform ρ into the Q space of the sine transform [Eq. (22)].
3. Calculate $V(q)$ [Eq. (25)].
4. Transform $V(q)$ back to real space [Eq. (20)].
5. Calculate $C_x = -\Delta_y V$; $C_y = \Delta_x V$ [Eq. (12)].
6. Form the sums of the phase slopes, $\Phi_x = S_x + C_x$, $\Phi_y = S_y + C_y$ [Eq. (3)].
7. Integrate the difference to get the phase [Eq. (4)].

The steps outlined above use the virtual functions C, V, ρ . Those are intermediate steps in the solution. If we execute the steps analytically, we see that the phase Φ may be expressed directly in terms of S_x and S_y . Expanding V in a sine series over the half-integer set is equivalent to expanding the phase in a cosine series over the integer points.

The phase Φ is expressed in the complete, orthogonal cosine set over grid (n, m) . The functions S_x and $\Delta_x \Phi$ that use half-integer x and integer y are expanded as a mixed sine-cosine series:

$$\begin{aligned} S_x(\hat{n}, m) &= S_x(n + 1/2, m) \\ &= \sum_{q_x=1}^{N-1} \sum_{q_y=0}^{M-1} \tilde{S}_x(q_x, q_y) \sin(nQ_x) \cos[(m - 1/2)Q_y], \end{aligned} \quad (27)$$

and similarly for S_y . The Fourier components are given by

$$\Phi(n, m) = \sum_{q_x=0}^{N-1} \sum_{q_y=0}^{M-1} \tilde{\Phi}(q_x, q_y) \cos[(n - 1/2)Q_x] \cos[(m - 1/2)Q_y], \quad (28)$$

$$\tilde{\Phi}(q_x, q_y) = \frac{T(q_x) \tilde{S}_x(q_x, q_y) + T(q_y) \tilde{S}_y(q_x, q_y)}{T^2(q_x) + T^2(q_y)}, \quad \tilde{\Phi}(0, 0) = 0 \quad (29)$$

$$\begin{aligned} \tilde{S}_x(q_x, q_y) &= \frac{4 - 2\delta(q_y)}{NM} \sum_{n=1}^{N-1} \sum_{m=1}^M S_x(\hat{n}, m) \sin(nQ_x) \cos[(m \\ &- 1/2)Q_y], \end{aligned} \quad (30a)$$

$$\begin{aligned} \tilde{S}_y(q_x, q_y) &= \frac{4 - 2\delta(q_x)}{NM} \sum_{n=1}^N \sum_{m=1}^{M-1} S_y(n, \hat{m}) \cos[(n \\ &- 1/2)Q_x] \sin(mQ_y), \end{aligned} \quad (30b)$$

where we have made use of the Kronecker δ .

Once it is known that the optimal solution may be expressed in a cosine series [Eqs. (30)], it may seem that one could have used a much simpler derivation of the formulas: Expand the phase in cosine series, and find the best value for the coefficients. This approach was taken before, with expansions in periodic Fourier components. Note the difference: We solved from scratch, using the potential V , and found that the best solution could be expressed in a cosine series. Previous works used *a priori* periodic Fourier components, and got suboptimal results.

Both the sine and cosine transforms may be calculated using variations of the FFT algorithm, the fast sine transform and fast cosine transform. Such routines are available over the web in C and FORTRAN forms. In Appendix A we show how to use the more available FFT.

6. EXAMPLES AND COMPARISON

At first glance, Eqs. (28)–(30) look exactly like the result from previous works using the FFT. There are, however, three major differences:

1. The wave frequencies here are $k\pi/N$, whereas usually they are $2k\pi/N$. This means that waves of lower frequencies are being used. Our set is not periodic in the region $1, 2, \dots, N$.
2. The set of functions $\exp(ix2n\pi/N)$ is incomplete if one drops the $n=0$ term. Expanding the phase in the set of $\exp(inx2\pi/N + imy2\pi/M)$ assumes that the average shear in the x (or y) direction is zero. If the shear is constant, or S_x depends only on y , there is an error. In contrast, our set is complete.
3. Our method produces a lower variance.

In Table 1 we take some analytical examples and compare them with the periodic FFT methods. The region R is $n=1, \dots, 64$, $m=1, \dots, 128$. For each example, we calculate the errors in the fit, $e_x = S_x - \Delta_x \Phi$ and $e_y = S_y - \Delta_y \Phi$. We cite the Std, i.e., the normalized standard deviation between the measured and calculated gradients, where

Table 1. Errors Produced in Calculation of Phase from Gradients over a Rectangular Region

Example No.	Phase $\Phi_a(n, m)$	Given Shears S	This Method		Periodic FFT	
			Std	MaxDev	Std	MaxDev
1	n	$S_x(n+1/2, m)=1; S_y=0$	3.38×10^{-15}	1.07×10^{-14}	0.992	1
2	$n^2/2+9.5n$	$S_x(n+1/2, m)=n+10; S_y=0$	1.83×10^{-13}	4.55×10^{-13}	42.17	42.5
3	—	$S_x(n+1/2, m)=m+10; S_y=0$	15.30	29.26	82.51	138
4	—	$S_x(n+1/2, m)=(m-16.5)n(64-n); S_y=0$	3.07	5.06	10.17	23.69
5	$\cos kn \cos lm$	$S_x=\Delta_x\Phi, S_y=\Delta_y\Phi$	1.98×10^{-14}	7.11×10^{-14}	1.41	6.72
6	Random	$S_x=\Delta_x\Phi, S_y=\Delta_y\Phi$	1.42×10^{-14}	4.61×10^{-14}	1.48	3.15
7	Random	Noise added to $S_x=\Delta_x\Phi, S_y=\Delta_y\Phi$	1.04	3.14	1.81	4.33

$\text{Std}^2 \equiv [\text{Var}(e_x) + \text{Var}(e_y)] / (NM)$. We also cite the maximum deviation of the fit, $\text{MaxDev} \equiv \max(|e_x|, |e_y|)$.

Some examples have an exact solution, which is correctly produced in this method (limited by computational accuracy). The third and fourth cases have impossible data, but this method has a better fit. In the last three cases we chose a phase, calculated its slopes, and reconstructed it back. In the last two cases we took a random distribution with a power spectrum of -1 , and in the last case added noise to the calculated slopes. To show the capabilities of the aperiodic solution, we purposely added a small planar tilt to all these last cases (Fig. 1), a tilt which tends to cause the standard reconstruction algorithms²² to fail.

7. THE SOLUTION FOR A CIRCULAR REGION WITH EXCLUSION

A telescope with a circular central obscuration is modeled as a round region R_0 , with a circular, concentric exclusion R_1 . The potential V is zero on the external and the internal circles: $V(R_0) = V(R_1) = 0$. Choosing as base functions a set that fulfills the boundary condition, and that is an eigenfunction of the Laplace operator, makes the solution trivial. Using the variables $z = \log(r)$ and $\varphi = a \tan(x/y)$, the boundary conditions become $V(z_0, \varphi) = V(z_1, \varphi) = 0$.

For expanding V , the set is $V_{km}(z, \varphi) = \sin[\pi k(z - z_0)/(z_1 - z_0)] \exp(im\varphi)$. For expanding the phase, the set is $\Phi_{km}(z, \varphi) = \cos[\pi k(z - z_0)/(z_1 - z_0)] \exp(im\varphi)$. These sets satisfy the criteria that (1) they are orthogonal over the region R with weight $1/r^2$; (2) their gradients, $\nabla\psi_{km} \cdot \nabla\psi_{k'm'}$ are orthogonal with unit weight; and (3) $\nabla^2\psi_{km} = -(k \cdot k + m \cdot m)r^{-2}\psi_{km}$.

The ideal integer grid (measurement points) for expanding in such a set is a rectangular grid in $z-\varphi$ space, $(z, \varphi) = (n, m)$, with $n = 1, 2, \dots, N$ and $m = 1, 2, \dots, M$ as before. In the $r-\varphi$ space this grid is composed of concentric circles (spaced logarithmically) and equispaced rays.

8. SHEARS VERSUS EXACT DERIVATIVES

If the measured quantities are the gradients U_x and U_y on the integer grid (n, m) , then a good estimate for the shears S_x, S_y is the average of the derivatives at the end points, or the trapeze rule:

$$\begin{aligned} \Phi(n+1, m) - \Phi(n, m) &\approx S_x(n+1/2, m) \\ &= 1/2[U_x(n+1, m) + U_x(n, m)], \end{aligned} \quad (31a)$$

$$\begin{aligned} \Phi(n, m+1) - \Phi(n, m) &\approx S_y(n, m+1/2) \\ &= 1/2[U_y(n, m+1) + U_y(n, m)]. \end{aligned} \quad (31b)$$

The calculated S_x, S_y are then inserted into Eqs. (13)–(25) or (29). An improved and more complex estimate is a cubic interpolation:

$$\begin{aligned} S_x(n+1/2, m) &\approx 13/24[U_x(n+1, m) + U_x(n, m)] \\ &\quad - 1/24[U_x(n+2, m) + U_x(n-1, m)], \end{aligned} \quad (32a)$$

$$\begin{aligned} S_y(n, m+1/2) &\approx 13/24[U_y(n, m+1) + U_y(n, m)] \\ &\quad - 1/24[U_y(n, m+2) + U_y(n, m-1)]. \end{aligned} \quad (32b)$$

At the edges this is not defined, so Eqs. (31) must be used.

In every case, the shears S_x and S_y are integrals over the gradient in a small interval. If the measured quantities are accurate derivatives, then Eqs. (31) or (32) introduce some errors of approximating the shear with a second, or fourth, or any finite-order formula. It may be argued that better estimates are obtained by (a) fitting the measured derivatives with analytic functions, (b) carrying the integration analytically, and (c) calculating the shears S_x and S_y from the analytic expression and using them instead.²⁰

A counterargument is that an analytic expansion introduces hidden assumptions about the behavior of the function between the grid points. If the errors are caused by a process with a known power spectrum, such as turbulence, this can be taken care of by using an appropriate filter. In the case of segmented optics this assumption might not hold, and one should sample on a finer grid. The trapeze or cubic interpolations introduce known small errors. Analytical expansion could introduce unknown large ones. Mathematically, the analytical expansion holds where the derivatives are band-limited, while in truth they contain experimental noise, and noise is known not to be band-limited.

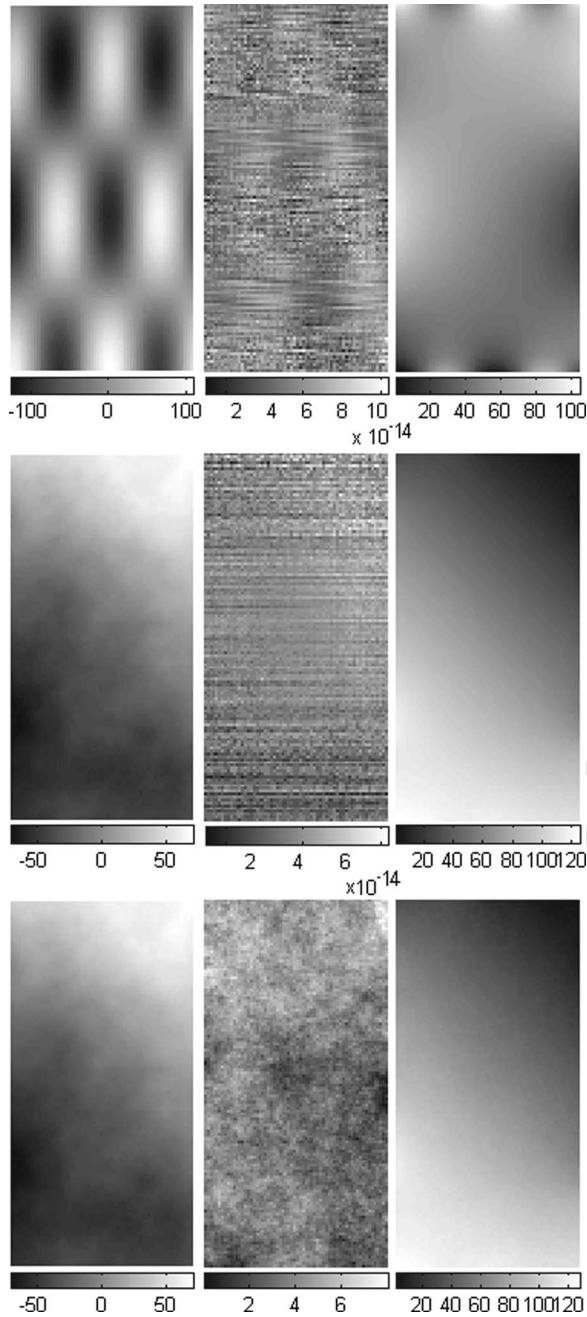


Fig. 1. (Left) input, (center) cosine transform errors, (right) Fourier transform errors for examples 5–7 in Table 1: (top) analytic periodic function, (center) random function, (bottom) with 50% noise on the slopes. In this last case we show the cosine *output* (bottom left) as it changes little compared with the noiseless input above it. The frame sizes were 64×128 .

We proceed, with grave reservations, and solve the analytical expansion for rectangular region R . The analytical solution has $V=0$ on the boundaries ($x=1/2$, $x=N+1/2$, $y=1/2$, $y=M+1/2$) and is not periodic. Thus, it is a sum of sine functions that satisfy the boundary conditions. This is equivalent to expanding Φ in the complete set of continuous cosine functions (functions with continuous x and discrete q), and the derivatives in the mixed sine–cosine series [as in Eqs. (28)–(31)]. The hidden assumption is that all the ignored high-frequency terms are negligible; that is

$$\Phi(x,y) = \sum_{q_x, q_y} \tilde{\Phi}(q_x, q_y) \cos[(x-1/2)Q_x] \cos[(y-1/2)Q_y], \quad (33)$$

which is Eq. (28) with the integer variables n, m replaced by the continuous variables $x \in [1, N], y \in [1, M]$, or

$$U_x(x,y) = \frac{d\Phi}{dx} = \sum_{q_x=1}^N \sum_{q_y=0}^{M-1} \tilde{U}_x(q_x, q_y) \sin[(x-1/2)Q_x] \times \cos[(y-1/2)Q_y], \quad (34a)$$

$$U_y(x,y) = \frac{d\Phi}{dy} = \sum_{q_x=0}^{N-1} \sum_{q_y=1}^M \tilde{U}_y(q_x, q_y) \cos[(x-1/2)Q_x] \times \sin[(y-1/2)Q_y]. \quad (34b)$$

There are NM values of $U_x(x,y)$, which requires at least NM variables $U_x(q_x, q_y)$. The sine term with $q=0$ vanishes identically everywhere, so we must start with $q_x=1$. The expansion of the derivatives uses one extra sine function, instead of the missing $q=0$ term. The sets are orthogonal and complete over the region $x=1, 2, \dots, N; y=1, 2, \dots, M$. The condition for minimum becomes

$$\tilde{\Phi}(q_x, q_y) = - \frac{\tilde{U}_x(q_x, q_y)Q_x + \tilde{U}_y(q_x, q_y)Q_y}{Q_x^2 + Q_y^2} \delta(q_x, q_y), \quad (35)$$

defined for $q_x=0, 1, \dots, N-1$ and $q_y=0, 1, \dots, M-1$, and excluding the term $q_x=q_y=0$. This is similar to F. Roddier and C. Roddier's result with the full FFT.¹⁰ The edge values $U_x(q_x=N, q_y)$ and $U_y(q_x, q_y=M)$ are not used, since their contribution to the phase vanishes on integer grid points.

One could also expand everything in the set $q_x=2, \dots, N+1$. Such an expansion assumes that $\Phi(x,y)$ does not have a $\cos[\pi(x-1/2)/N]F(y)$ term. This is the weakness of the analytical expansion: One assumes *a priori* that the continuous function may be approximated well by some finite series, and lacks those terms that were not included.

Suppose we use this same data with the difference-equation approach. S_x and S_y are the averages of U_x and U_y [see Eqs. (31)], so their transforms are

$$\begin{aligned} \tilde{S}_x(q_x, q_y) &= \tilde{U}_x(q_x, q_y) \cos(Q_x/2), \\ \tilde{S}_y(q_x, q_y) &= \tilde{U}_y(q_x, q_y) \cos(Q_y/2). \end{aligned} \quad (36)$$

Inserting into Eq. (29) we get

$$\tilde{\Phi}(q_x, q_y) = - \frac{1}{4} \frac{\tilde{U}_x(q_x, q_y) \sin(Q_x) + \tilde{U}_y(q_x, q_y) \sin(Q_y)}{\sin^2(Q_x/2) + \sin^2(Q_y/2)} \delta(q_x, q_y). \quad (37)$$

This is very similar to Eq. (35) and even more so to previous solutions,^{12,13} but the highest-frequency components have lower amplitudes.

If one uses the cubic interpolations of Eqs. (32) the results for the x transform are

$$\begin{aligned}\tilde{S}_x(q_x, q_y) &= \tilde{U}_x(q_x, q_y) \left[\frac{13}{12} \cos\left(\frac{Q_x}{2}\right) - \frac{1}{12} \cos\left(\frac{3Q_x}{2}\right) \right] \\ &= \tilde{U}_x(q_x, q_y) \cos\left(\frac{Q_x}{2}\right) P(Q_x),\end{aligned}\quad (38)$$

where $P(Q) = 4/3 - 1/3 \cos^2(Q/2)$, and similarly for the y transform. The phase transform is then

$$\begin{aligned}\tilde{\Phi}(q_x, q_y) &= -\frac{1}{4} \\ &\times \frac{\tilde{U}_x(q_x, q_y) \sin(Q_x) P(Q_x) + \tilde{U}_y(q_x, q_y) \sin(Q_y) P(Q_y)}{\sin^2(Q_x/2) + \sin^2(Q_y/2)} \\ &\times \delta(q_x, q_y).\end{aligned}\quad (39)$$

Note that although the formulas are similar to previous ones,^{10–13} the base set differs—we use the cosine set with $(Q_x, Q_y) = (\pi n/N, \pi m/M)$ while previous results are periodic with twice the values, $(Q_x, Q_y) = (2\pi n/N, 2\pi m/M)$.

9. SUMMARY

We have shown that given the measured shears, the optimal solution for the phase is composed of a main term resulting from direct integration and a small correction term. We have shown that the correction term is a solution of a Poisson equation, where the sources are the errors in the original data, and the boundary conditions are zero potential on the boundaries. In fact, one may argue that the correction term is no more than a sophisticated error-diffusion method that shuffles the errors slightly so as to reduce the variance somewhat. Like a cleaner who sweeps the dirt under the rug, the mathematics attributes the errors to regions without measured data. Errors close to the boundaries contribute little to the variance, while internal ones (that can not be shuffled away) contribute their full share.

We have shown that in the case of rectangular regions, the correct solution is very similar to the established periodic FFT solution, with one big difference: The basic set is the Fourier cosine series, not the periodic one, and the waves have half the frequencies. We have demonstrated with seven simple cases the superiority, namely the smaller variance, compared with the FFT method.

The fast sine/cosine transform is actually quicker than FFT, as it uses only real numbers, not complex ones. Since it produces the best results, it is the preferable choice.

We have shown that for a telescope with an obscuration, the natural set of functions is waves in azimuth and logarithm of the radius. We have shown how to solve (exactly) such a geometry.

We have investigated the solution with analytical extension of measured derivatives, and compared the results with those obtained using approximated shears. The results show that the differences are not significant. Formally, these results are very similar to previous ones. The

similarity is only formal—the actual set of base functions differs: The wave numbers are smaller by half in each direction

We have used an approximating method (multigrid at 45°) to solve the Poisson equation for the correction term. The method is linear, requiring $4N$ first-neighbor averages or $19N$ basic operations, much quicker than $(3)(5) \times (N) \log_2 N$ using the FFT method. The convergence is so quick that one iteration suffices. A variation of it solves any region R with any number (and shape) of excluded subregions, with roughly 50% more operations: Each gradient term S_x and S_y is accompanied by a Boolean invalid/valid flag (or $\varepsilon/1$ weight). Excluded points or regions are simply flagged invalid or multiplied by the very small weight ε .

APPENDIX A

The purpose here is to show the relationship between Fourier, sine, and cosine expansions. We use r and q as the reciprocal indices, in order to represent both x and q_x , y and q_y . We use r_h to denote half-integer values, and r for integer ones.

The set of functions $\sin(q\pi r_h/N)$ with integer q is orthogonal over the set of points $r_h = 0.5, 1.5, \dots, N-0.5$. It forms a complete set of functions if one uses $q = 1, 2, \dots, N$. The function with $q=0$ is zero. We write the set as

$$\begin{aligned}\sum_{r_{1/2}=0.5}^{N-0.5} \sin\left(\frac{r_h \pi k}{N}\right) \sin\left(\frac{r_h \pi j}{N}\right) &= \frac{N}{2} \delta(k-j), \\ \sum_{q=1}^{N-1} \sin\left(\frac{\pi r_h q}{N}\right) \sin\left(\frac{\pi t_h q}{N}\right) &= \frac{N}{2} \delta(t_h - r_h).\end{aligned}\quad (A1)$$

The set of functions $\cos(q\pi r_h/N)$ with integer q is orthogonal over the set of points $r_h = 0.5, 1.5, \dots, N-0.5$. It forms a complete set of functions if one uses $q = 0, 1, \dots, N-1$. The function with $q=N$ is zero at those half-integer r points. We write the set as

$$\begin{aligned}\sum_{r_{1/2}=0.5}^{N-0.5} \cos\left(\frac{r_h \pi k}{N}\right) \cos\left(\frac{r_h \pi j}{N}\right) &= \frac{N}{2} \delta(k-j)[1 + \delta(k)], \\ \sum_{q=0}^{N-1} \frac{1}{1 + \delta(q)} \cos\left(\frac{\pi r_h q}{N}\right) \cos\left(\frac{\pi t_h q}{N}\right) &= \frac{N}{2} \delta(t_h - r_h).\end{aligned}\quad (A2)$$

A different orthogonal series is the sine of integer r values. The set of functions $\sin(q\pi r/N)$ with integer q, r is orthogonal over the set of points $r = 1, 2, \dots, N-1$. The functions vanish at the edges $r=0, r=N$. Note that the function with $q=0$ is identically zero, and the function with $q=N$ vanishes, too, at integer r points. To form a complete set, we use $q = 1, 2, \dots, N-1$:

$$\sum_{r=1}^{N-1} \sin\left(\frac{r\pi k}{N}\right) \sin\left(\frac{r\pi j}{N}\right) = \frac{N}{2} \delta(k-j),$$

$$\sum_{q=1}^{N-1} \sin\left(\frac{\pi r q}{N}\right) \sin\left(\frac{\pi t q}{N}\right) = \frac{N}{2} \delta(t-r). \quad (\text{A3})$$

The functions $\cos(q\pi r/N)$ with integer q, r are orthogonal over the set of points $r=0, 1, \dots, N$ provided that their weights at $r=0$ and at $r=N$ are one-half. The orthogonal functions are $q=0, 1, \dots, N$. To form a complete set, one should use weights of one-half with the functions $q=0$ and $q=N$:

$$\langle k|j \rangle = \sum_{r=1}^{N-1} \cos\left(\frac{r\pi k}{N}\right) \cos\left(\frac{r\pi j}{N}\right) + 1/2 \cos 0 \cos 0$$

$$+ 1/2 \cos(\pi k) \cos(\pi j) = \frac{N}{2} \delta(k-j)[1 + \delta(k) + \delta(k-N)]. \quad (\text{A4})$$

The sine and cosine transforms of N terms may be calculated using a FFT of $2N$ terms. This is an inefficient but handy method, as FFTs are included in computer packages, but fast sine/cosine transforms are not. Thus the integer- r , integer- q cosine and sine expansions are

$$\tilde{R}_{Ci}(q) = \sum_{r=1}^N R(r) \cos\left(\frac{\pi q r}{N}\right) = \sum_{r=1}^N R(r) \cos\left(\frac{2\pi q r}{2N}\right)$$

$$= \frac{R_F(q) + R_F(2N-q)}{2},$$

$$R_F(q) = \sum_{r=1}^N R(r) \exp\left(i \frac{2\pi q r}{2N}\right) + \sum_{r=N+1}^{2N} 0^* \exp\left(i \frac{2\pi q r}{2N}\right), \quad (\text{A5})$$

$$\tilde{R}_{Si}(q) = \sum_{r=1}^N R(r) \sin\left(\frac{\pi q r}{N}\right) = \frac{R_F(q) - R_F(2N-q)}{2i}, \quad (\text{A6})$$

where $R(r)$ is the series to be expanded, and indices Ci , Si , and F indicate cosine (integer r), sine (integer r), and Fourier expansions, respectively. In other words, one fills the first half of a vector of length $2N$ with the series $R(r)$, and pads the second half with zeros. After performing the FFT, the required results are calculated from the q and $-q$ terms.

The transform for half-integer points [such as in Eqs. (30)] is given by mixing sine and cosine transforms of integer-points:

$$\tilde{R}_{Ch}(q) = \sum_{n=1}^N R(n) \cos\left[\frac{\pi q(n-1/2)}{N}\right]$$

$$= \tilde{R}_{Ci}(q) \cos\left(\frac{\pi q}{2N}\right) + \tilde{R}_{Si}(q) \sin\left(\frac{\pi q}{2N}\right), \quad (\text{A7})$$

$$\tilde{R}_{Sh}(q) = \sum_{n=1}^N R(n) \sin\left[\frac{\pi q(n-1/2)}{N}\right]$$

$$= \tilde{R}_{Si}(q) \cos\left(\frac{\pi q}{2N}\right) + \tilde{R}_{Ci}(q) \sin\left(\frac{\pi q}{2N}\right). \quad (\text{A8})$$

The inverse transform for half-integer points, such as Eq. (28), is calculated by

$$R_{Ch}(r) = \frac{1}{N} \sum_{q=0}^{N-1} \tilde{R}_h(q) \cos\left[\frac{\pi q(r-1/2)}{N}\right]$$

$$= \frac{F(r) + F(2N-r+1)}{2}, \quad (\text{A9})$$

$$R_{Sh}(r) = \frac{1}{N} \sum_{q=1}^{N-1} \tilde{R}_h(q) \sin\left[\frac{\pi q(r-1/2)}{N}\right]$$

$$= \frac{F(r) - F(2N-r+1)}{2i}, \quad (\text{A10})$$

where

$$F(r) = \frac{2}{2N} \sum_{q=0}^{N-1} \tilde{R}_h(q) \exp\left(\frac{-i\pi q}{2N}\right) \exp\left(\frac{i2r\pi q}{2N}\right)$$

$$= 2\mathcal{F}^{-1}\left[\tilde{R}_h(q) \exp\left(\frac{-i\pi q}{2N}\right)\right], \quad (\text{A11})$$

and where the index h in (A7)–(A11) signifies expansion with half-integer x values. Here again, the recipe is

“fill only half of a $2N$ -long vector with data, the rest with zeros. Multiply by some phase correction and perform an inverse FFT. Mix the two symmetric FFT terms to get results.”

We assumed the convention of normalization by N following a transform and its inverse.

ACKNOWLEDGMENTS

Trying to solve for ocular wave fronts was a main driver for this research, started with support from the Ministry of Science and Technology in Israel and the European Community Research and Training Network SHARP-EYE.

The authors' e-mail addresses are amos@talmi.net and eribak@physics.technion.ac.il.

REFERENCES

1. D. Malacara, ed., *Optical Shop Testing* (Wiley, 1978).
2. R. K. Tyson, *Principles of Adaptive Optics*, 2nd ed. (Academic, 1998).
3. R. K. Tyson, ed., *Adaptive Optics Engineering Handbook* (Marcel Decker, 2000).
4. Y. Carmon and E. N. Ribak, “Phase retrieval by demodulation of a Hartmann–Shack sensor,” *Opt. Commun.* **215**, 285–288 (2003).
5. Y. Carmon and E. N. Ribak, “Fast Fourier demodulation,” *Appl. Phys. Lett.* **84**, 4656–4657 (2004).
6. A. Talmi and E. N. Ribak, “Direct demodulation of

- Hartmann–Shack patterns,” *J. Opt. Soc. Am. A* **21**, 632–639 (2004).
7. D. L. Fried, “Least-square fitting of a wave-front distortion estimate to an array of phase-difference measurements,” *J. Opt. Soc. Am.* **67**, 370–375 (1977).
 8. R. H. Hudgin, “Wave-front reconstruction for compensated imaging,” *J. Opt. Soc. Am.* **67**, 375–378 (1977).
 9. W. H. Southwell, “Wave front estimation from wave front slope measurement,” *J. Opt. Soc. Am.* **70**, 998–1006 (1980).
 10. F. Roddier and C. Roddier, “Wavefront reconstruction using iterative Fourier transforms,” *Appl. Opt.* **30**, 1325–1327 (1991).
 11. C. Roddier and F. Roddier, “Wavefront reconstruction from defocused images and the testing of ground-based optical telescopes,” *J. Opt. Soc. Am. A* **10**, 2277–2287 (1993).
 12. K. R. 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).
 13. L. A. Poyneer, M. Troy, B. Macintosh, and D. Gavel, “Experimental validation of Fourier transform wave-front reconstruction at the Palomar Observatory,” *Opt. Lett.* **28**, 798–800 (2003).
 14. C. Elster, “Exact two-dimensional wave-front reconstruction from lateral shearing interferograms with large shears,” *Appl. Opt.* **39**, 5353–5359 (2000).
 15. A. Dubra, C. Paterson, and C. Dainty, “Wave-front reconstruction from shear phase maps by use of the discrete Fourier transform,” *Appl. Opt.* **43**, 1108–1113 (2004).
 16. L. Ellerbroek, “Efficient computation of minimum variance wave-front reconstructors using sparse matrix techniques,” *J. Opt. Soc. Am. A* **19**, 1803–1816 (2002).
 17. L. Gilles, C. R. Vogel, and B. L. Ellerbroek, “A multigrid preconditioned conjugate gradient method for large scale wave-front reconstruction,” *J. Opt. Soc. Am. A* **19**, 1817–1822 (2002).
 18. L. Gilles, “Order N sparse minimum-variance open-loop reconstructor for extreme adaptive optics,” *Opt. Lett.* **28**, 1927–1929 (2003).
 19. G. A. Tyler, “Reconstruction and assessment of the least-squares and slope discrepancy components of the phase,” *J. Opt. Soc. Am. A* **17**, 1828–1839 (2000).
 20. S. Rios and E. Acosta, “Orthogonal modal reconstruction of a wave front from phase difference measurements,” *J. Mod. Opt.* **46**, 931–939 (1999).
 21. A. J. Roberts, “Simple and fast multigrid solution of Poisson’s equation using diagonally oriented grids,” *Aust N. Z. Ind. Appl. Math. J.* **43**, E1–36 (2001).
 22. A MATLAB program with these examples can be read at physics.technion.ac.il/~eribak/s2p/.

AUTHOR QUERIES — 005602JOA

- #1 Au: OSA recently adopted the spelling of wavefront as one word.
- #2 Au: "A" seems to fit better here. Do you agree?
- #3 Au: Is "three" correct?
- #4 Au: Is "first" as meant?
- #5 Au: Did you mean Q sub y here?
- #6 Au: Is "functions" correct?

PROOF COPY [61718] 005602JOA