

Principles of Wavefront Sensing and Reconstruction

Gary Chanan

*University of California, Irvine, Department of Physics and Astronomy,
Irvine, California 92697*

Abstract. A variety of approaches to wavefront sensing and reconstruction are surveyed as they are used in adaptive optics and related applications. These include the Gerchberg-Saxton algorithm; shearing interferometry; and Shack-Hartmann, curvature, and pyramid wavefront sensing. Emphasis is placed on the relevant optics and mathematics, which are developed in some detail for Shack-Hartmann and curvature sensing (currently the two most widely-used approaches) and also to a lesser extent for pyramid sensing. Examples are given throughout.

1. Introduction

The basic goal of adaptive optics is easily stated: to measure the aberrations of an incoming wavefront and then cancel these out by applying compensating aberrations, all in real time. Of course, underlying this simple statement is a host of very challenging optical, mathematical, computational, and technological problems. In this work we concentrate on some of the optical and mathematical issues associated with the first task from the above statement, i.e. measuring the aberrations of an incoming wavefront. It is convenient to subdivide this into two separate tasks, which we refer to as wavefront sensing and wavefront reconstruction, where the distinction is articulated below.

We do not measure wavefront aberrations or phases directly; in practice the direct measurements virtually always consist of intensity distributions on a CCD or other area detector. In this work, we use “wavefront sensing” to refer to a technique by which an arbitrary wavefront phase surface is converted into an uniquely defined intensity distribution — which in turn can be (more or less) readily inverted to yield the original phase. This latter inversion is referred to as “wavefront reconstruction”; this may involve inverting a matrix, solving Poisson’s equation subject to certain boundary conditions, or another inversion procedure. The wavefront sensing generally involves manipulation of the original wavefront to facilitate the subsequent inversion. This manipulation may take place in the aperture plane (the insertion of a lenslet array in the Shack-Hartmann procedure) or in the image plane (the insertion of knife edge in the knife edge test or of a pyramid in pyramid wavefront sensing). [For the purposes of this paper we consider the introduction of a deliberate focus error — as in curvature sensing — as a manipulation in the aperture plane because the mathematical condition corresponding to defocus is more easily expressed there

than in the image plane.] Also note that the intensity distribution of interest is sometimes measured in (or near) the image plane (as in Shack-Hartmann sensing or with the Gerchberg-Saxton algorithm) and sometimes in the re-imaged aperture plane (as in pyramid sensing or with the shearing interferometer).

2. Gerchberg-Saxton Algorithm

The Gerchberg-Saxton algorithm (7) is an iterative scheme which reconstructs the wavefront from the out-of-focus intensity distribution in the image plane. Using an out-of-focus, rather than an in-focus, image serves two purposes: (1) the spatial sampling of the image intensity distribution can be greatly improved, and (2) the undesirable degeneracy between the wavefront and its negative, which otherwise both yield the same intensity distribution, is broken; thus the extraneous negative phase solution can be avoided.

We assume that the intensity distributions in both the aperture plane (usually constant over the aperture) and the image plane are known. We write the complex amplitude of the wave in the aperture plane as $A(x, y) \exp[i\phi(x, y)]$ and in the image plane as $B(u, v) \exp[i\theta(u, v)]$, where (x, y) and (u, v) are the coordinates in the two planes and the magnitudes $A(x, y)$ and $B(u, v)$ of the complex amplitudes are obtained directly as the square roots of the assumed or measured intensity distributions.

The task is to obtain the aperture plane phase $\phi(x, y)$. If the image plane phase $\theta(u, v)$ were known, this could be accomplished immediately by performing an inverse Fourier transform. Of course in practice the latter phase is not measurable. However, the desired aperture plane phase can still be obtained iteratively, as summarized by the flow chart shown in Figure 1, and described below.

Suppose that the intensity is known to be constant over the aperture; then so too is the magnitude $A(x, y)$. The algorithm begins by constructing the complex amplitude in the amplitude plane by “grafting” a guess for the unknown phase $\phi(x, y)$ onto the known $A(x, y)$. Absent other information (e.g. the phase at an earlier time), the guess may be identically zero or generated by a random number generator. [Typically, the known defocus error is explicitly included in the guess.] The complex amplitude in the image plane is then constructed by taking the Fourier transform of $A(x, y) \exp[i\phi(x, y)]$. The image plane complex amplitude is revised by substituting the known image plane magnitude $B(u, v)$ for the calculated value, while retaining the calculated phase $\theta(u, v)$. One then goes back to the aperture plane by means of the inverse Fourier transform, revises the magnitude $A(x, y)$ while retaining the newly calculated phase $\phi(x, y)$, and so on. Iterating in this manner, one will eventually converge to the true aperture plane phase.

Figure 2 shows the progression of Gerchberg-Saxton reconstructions for a purely astigmatic wavefront, starting from an initial guess of a perfect wavefront. The standard focus error has been subtracted from all the images in the sequence.

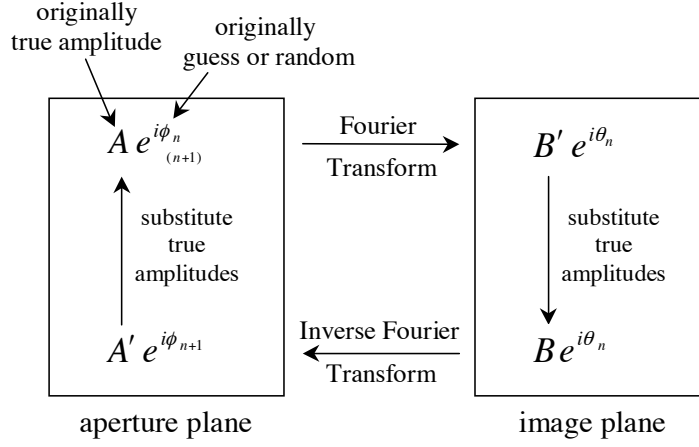


Figure 1. Gerchberg-Saxton flow chart, indicating refinement of aperture plane phase estimate between n -th and $(n+1)$ -st iterations.

A proof that this procedure converges to the correct aperture plane phase is beyond our scope here, but to make the process somewhat less mysterious we briefly review (with some changes in notation) the plausibility argument originally presented by Gerchberg and Saxton (1972). For simplicity we consider a one-dimensional version of the algorithm. At the n -th iteration we represent the complex amplitude at a particular point in the aperture by the complex number h_n , as shown in Figure 3. At the instant shown, h_n has already been revised so that it has the known correct magnitude a . The image of h_n under the action of the Fourier transform is g_n (before correction); g_{n+1} is the corresponding complex number after correction to the known magnitude b , and the value of the correction at this step is

$$c_{n+1} = g_{n+1} - g_n \quad (1)$$

The inverse Fourier transform carries g_{n+1} into h_{n+1} in the aperture. As before, h_{n+1} is revised to h_{n+2} . At this point there are two relevant corrections:

$$d_{n+1} = h_{n+1} - h_n \quad (2)$$

$$d_{n+2} = h_{n+2} - h_{n+1} \quad (3)$$

Now by construction, the following are all Fourier transform/inverse transform pairs: h_n and g_n , h_{n+1} and g_{n+1} , c_{n+1} and d_{n+1} . Furthermore, the root mean square values (averaging over all points in the aperture or image plane) of h_n and g_n are equal by Parseval's theorem, and similarly for the other two pairs. By simple geometry:

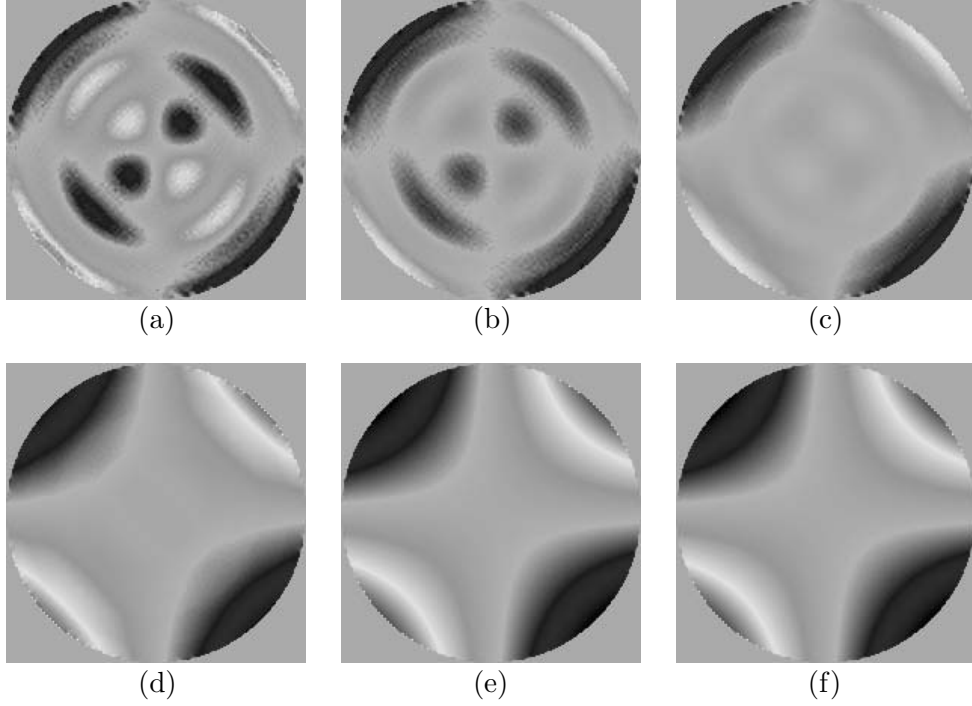


Figure 2. Progressive Gerchberg-Saxton reconstructions for a purely astigmatic wavefront: (a) after 1 iteration, (b) after 3 iterations, (c) after 10 iterations, (d) after 30 iterations, (e) after 100 iterations; (f) the original wavefront.

$$|d_{n+1}|^2 = |h_n|^2 + |h_{n+1}|^2 - 2|h_n||h_{n+1}| \cos \psi \quad (4)$$

and

$$|d_{n+2}|^2 = |h_n|^2 + |h_{n+1}|^2 - 2|h_n||h_{n+1}| \quad (5)$$

Thus we have $d_{n+2}^{rms} \leq d_{n+1}^{rms} = c_{n+1}^{rms}$ and the errors tend to grow smaller in magnitude at every step.

The main advantage of the Gerchberg-Saxton method is ease of implementation since it requires no special purpose hardware and since the algorithm itself is relatively simple. The main disadvantage is that it is computationally intensive — one may have to perform thousands of Fourier transforms in order to obtain satisfactory convergence. Like the shearing interferometer, but unlike all the other methods discussed in this work, the Gerchberg-Saxton algorithm operates in the physical, not geometrical, optics regime. This means that it can be used to phase segmented mirrors (16), but it also implies that there are relatively narrow spectral band limitations and that phase unwrapping algorithms

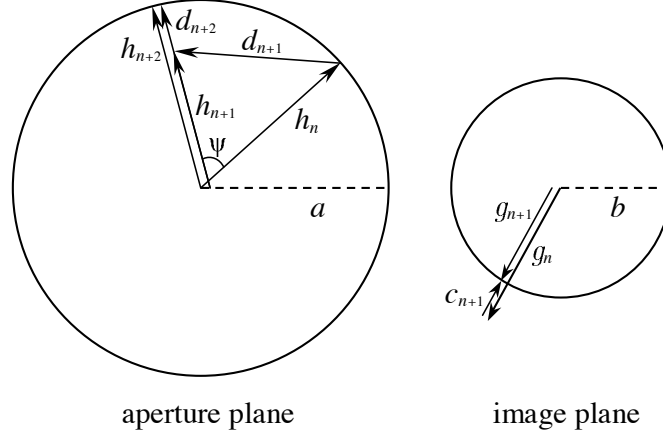


Figure 3. Geometrical relationship of the complex numbers in the Gerchberg-Saxton plausibility argument.

are required for larger aberrations.

A related method of wavefront sensing, beyond our scope here, is the phase diversity approach (10), which utilizes the intensity distributions in both the in-focus image and a simultaneously-measured, slightly defocused image, as well as a numerically intensive computational algorithm, to reconstruct the phase in the aperture plane. (See Rousset (22) and references therein).

3. Lateral Shearing Interferometer

The shearing interferometric approach to wavefront sensing is unusual in that neither the detection nor the manipulation of the wavefront (i.e. the shear) takes place in the image plane; rather both occur in (separately re-imaged) aperture planes. In practice the wavefront is split (e.g. by a 50/50 beamsplitter) in the aperture plane, and the resulting two wavefronts are recombined without a phase shift but with a displacement or lateral shear \vec{s} , as indicated in Figure 4. For reasons of symmetry it is convenient to think of the two wavefronts as shifted by $\pm \frac{1}{2}\vec{s}$ from the original position. The resulting intensity as a function of position \vec{r} in the aperture is then given by (see e.g. 22):

$$I(\vec{r}) = \frac{1}{4} |\exp[i\phi(\vec{r} - \frac{1}{2}\vec{s})] + \exp[i\phi(\vec{r} + \frac{1}{2}\vec{s})]|^2 \quad (6)$$

where the normalization of the intensity has been chosen so that $I(\vec{r}) \rightarrow 1$ as $\vec{s} \rightarrow 0$. Abbreviating

$$\phi_{\pm} = \phi(\vec{r} \pm \frac{1}{2}\vec{s}) \quad (7)$$

we find:

$$I(\vec{r}) = \frac{1}{2} + \frac{1}{2} \cos(\phi_+ - \phi_-) \quad (8)$$

Consider a small shear in the x -direction; in this case:

$$\phi_+ - \phi_- = s \frac{\partial \phi(\vec{r})}{\partial x} + \text{higher order terms} \quad (9)$$

Thus the shearing interferometer is a gradient detector. In this respect it is similar to the Shack-Hartmann wavefront sensor described below; we defer the discussion of wavefront reconstruction from gradient information to the latter section. However, note that unlike the Shack-Hartmann scheme, two separate channels (two separate shearing mechanisms) are necessary to obtain the two components of the gradient in the shearing interferometric case. Examples of shearing interferograms are shown in Figure 5 for an aberration corresponding to pure spherical aberration (the Zernike polynomial $Z_{4,0}$). The four panels show the effect of increasing shear.

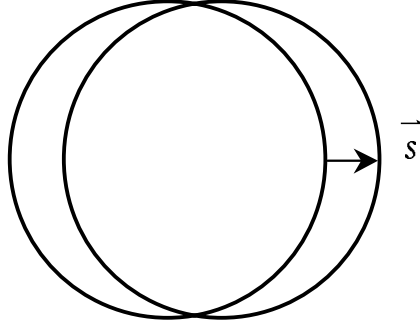


Figure 4. Geometrical relationship between sheared pupils in the lateral shearing interferometer.

From Equation (9) and Figure 5, it is clear that the gain of a shearing interferometer increases with the magnitude s of the shear. This is an advantage in that the signal strength (the variable part of the intensity) can be maintained as the gradient is decreased, for example as the adaptive optics loop is closed. However, it can also be a disadvantage as there is a non-trivial trade-off between the gain and the accuracy of the approximation associated with the neglect of higher order terms in Equation (9). That is, to be well-satisfied, this approximation requires:

$$s\phi_{\text{rms}} \ll D \quad (10)$$

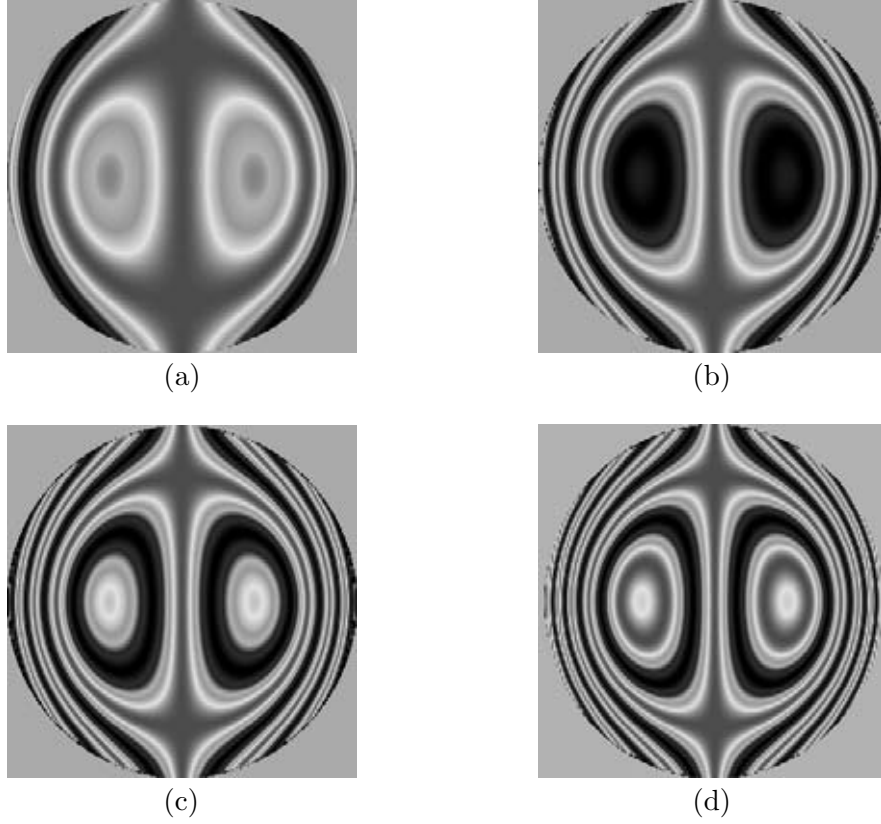


Figure 5. Shearing interferograms for pure spherical aberration, $Z_{4,0}$. The panels show the effect of increasing shear s , expressed as a fraction of the aperture diameter: (a) $s = 0.01$, (b) $s = 0.02$, (c) $s = 0.03$, and (d) $s = 0.04$.

where D is the aperture diameter. When the AO loop is first closed, the rms phase is given by the full atmospheric value (11):

$$\phi_{\text{rms}} \cong \left(\frac{D}{r_o} \right)^{6/5} \quad (11)$$

where r_o is the atmospheric coherence diameter (6). For even the largest current optical telescopes substitution of inequality (10) in Equation (11) yields the clearly restrictive condition $s \ll r_o$ because of the weak dependence on D . In the absence of these considerations one would simply match the shear scale to r_o , and the condition on the gain would not be so restrictive.

Apart from the potential advantages associated with variable gain, (e.g., the

large dynamic range), the shearing interferometer has several disadvantages with respect to Shack-Hartmann sensing: the two-channel requirement, the non-linear response, and bandwidth restrictions. The latter is therefore now in much more widespread use. However, there are still special-purpose applications of non-AO shearing interferometric wavefront sensing, one of which is described below.

Unlike Shack-Hartmann, curvature, or pyramid wavefront sensing, but like the Gerchberg-Saxton algorithm, the shearing interferometer is based on physical, not geometrical, optics, and thus can be used for relative piston sensing in segmented mirrors. It is easy to see that the shearing interferometer has no sensitivity to global piston, and is only sensitive to the component of global tilt that is parallel to the shear. However, in the case of a segmented telescope, if we shear the wavefront by exactly one segment, relative piston errors are manifested by partially destructive interference, and tilt errors (in any direction) by linear fringes in the interferogram. This is illustrated in Figure 6 (pure piston errors) and Figure 7 (pure tip/tilt errors), which show such simulations for random segment alignment errors of the Keck Telescope. The shear corresponds to one segment in the vertical direction. For contrast we also show the effect of a half-segment shear. Such a shearing interferometric system is the basis for segment alignment of the 91-segment Hobby-Eberly Telescope (1); an unrelated physical optics approach is actually used for Keck (4; 3).

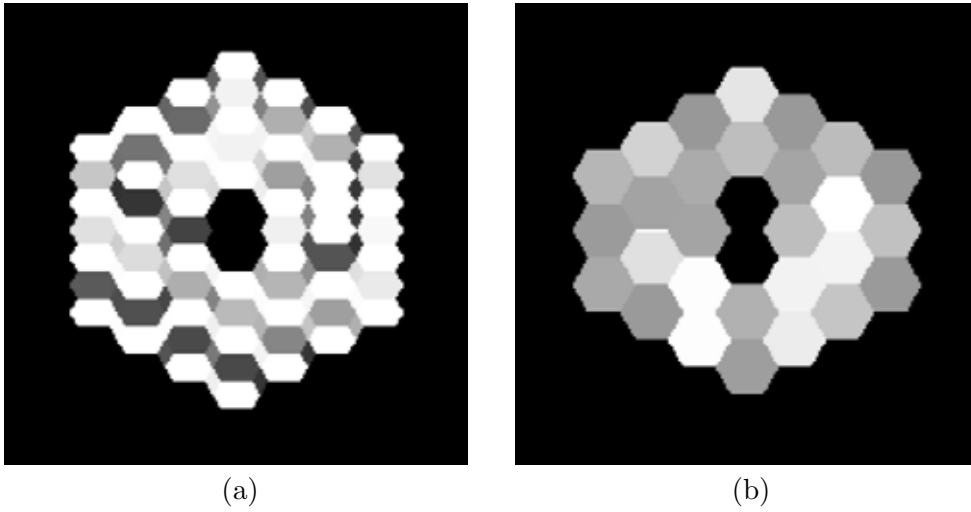


Figure 6. Simulated shearing interferograms corresponding to segment piston errors for the Keck Telescope: (a) half-segment shear, and (b) full-segment shear; both in the vertical direction.

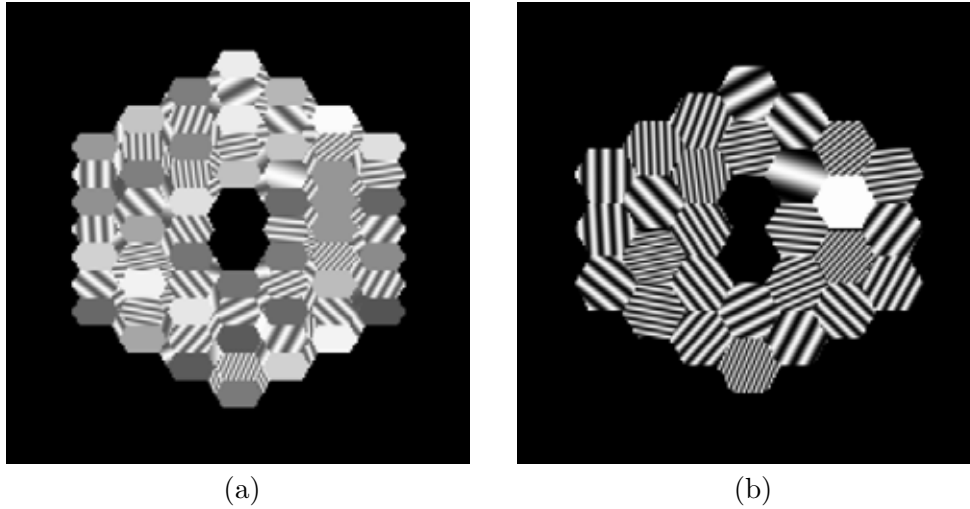


Figure 7. Simulated shearing interferograms corresponding to segment tip/tilt errors for the Keck Telescope: (a) half-segment shear, and (b) full-segment shear; both in the vertical direction.

4. Shack-Hartmann Sensing

Shack-Hartmann wavefront sensing is a variation of the classical Hartmann test (see e.g. 8), which we here review briefly by way of introduction. In the Hartmann test, a mask containing a large number of holes (up to several hundred) is placed over the primary mirror of a telescope, and the telescope is pointed at a star. An area detector such as a CCD is placed far enough in front of (or in back of) the nominal focal plane that the resulting array of stellar subimages fills a substantial portion of the detector, as shown schematically in Figure 8. The displacement of these subimages from their ideal or theoretical positions is then directly proportional to the gradient of the wavefront.

The Hartmann test is relatively simple, but it suffers from several disadvantages. Since the mask has the same diameter as the primary mirror, it is awkward to implement, particularly for large telescopes. Because the subimages are out of focus, only fairly bright stars can be used in the test. And, finally, by its very geometry, dense sampling of the wavefront is precluded.

In the Shack-Hartmann test (18), the primary is re-imaged onto a dense (close-packed) array of micro-lenses, as shown in Figure 9. A CCD is placed at the focus of the lenslet array, and the telescope is again pointed at a star. As in the Hartmann test, the displacement of the stellar subimages from their ideal positions is proportional to the gradient of the wavefront.

The Shack-Hartmann test addresses the major difficulties of the Hartmann test. The test is physically much less awkward to implement because the required

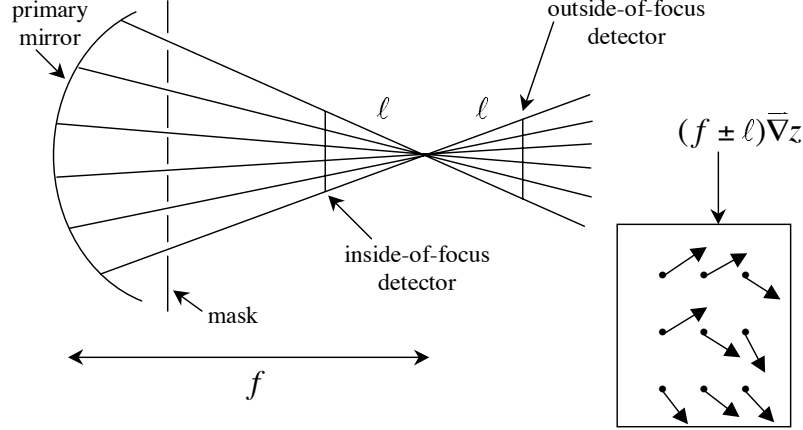


Figure 8. Hartmann concept. Inset indicates typical displacement of out-of-focus subimages.

size of the mask is reduced by as much as two to three orders of magnitude as a result of the demagnification of the primary (although, of course, additional optics are required), fainter stars can be used because the stellar subimages are in focus, and the lenslet array densely samples the wavefront. The ideal positions of the subimages can readily be determined by inserting a reference beam into the focal plane, as for example with a fiber optic. Calibrating the system in this way greatly reduces the optical demands on the lenslet array.

The basic equation of Shack-Hartmann testing is

$$\vec{\delta} = m f_{\ell} \vec{\nabla} z \quad (12)$$

where $\vec{\delta}$ is the vector displacement of the stellar subimages (in units of length), $m = f_{\text{telescope}}/f_{\text{collimator}}$ is the demagnification of the system, f_{ℓ} is the focal length of the lenslets, and $z(x, y)$ is the aberration expressed in units of length, related to the aberration $\phi(x, y)$ expressed as a phase by $\phi = kz = 2\pi z/\lambda$, where λ is the wavelength. To give an indication of the typical parameters of a Shack-Hartmann wavefront sensor on a large telescope adaptive optics system, we consider the AO wavefront sensor on the Keck Telescopes (26; 25). For this system, the demagnification is $m = 2800$, and the square lenslets are 200 microns on a side, with a focal length of 2000 microns. The lenslets thus map to 560 mm on a side at the primary mirror; this is generally equal to the atmospheric coherence diameter r_0 for a wavelength somewhere between 1 and 2 microns. As is commonly the case in such systems, the Keck lenslet array is still considerably larger than the CCD. An additional factor of 3.15 demagnification is provided by a focal reducer which relays the lenslet focal plane onto the detector. The above parameters produce a final image scale of 8.6 microns per arcsecond, which corresponds to 2.4 arcseconds per pixel for the 21 micron pixels of the Keck CCD.

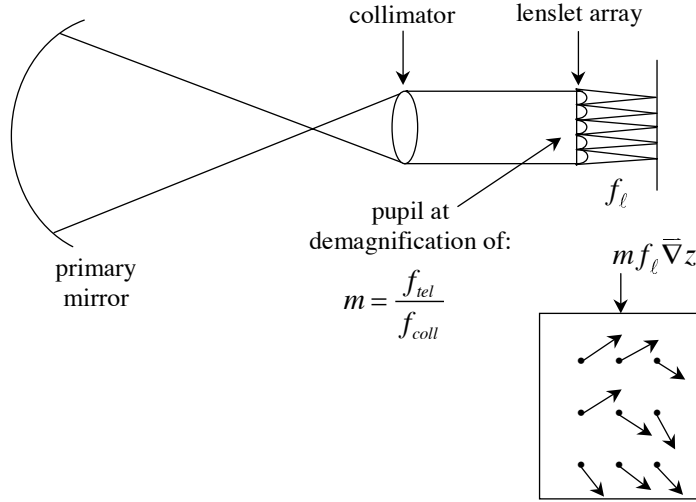


Figure 9. Shack-Hartmann concept. Inset indicates typical displacement of in-focus subimages.

From the above parameters, the Keck 560 mm subapertures map to 63 microns at the detector, i.e. the subimages are 3 pixels apart in each dimension. In practice, the subimages are aligned so that they nominally fall on the vertex of a two pixel by two pixel “quad cell,” with one-pixel-wide guard strips separating the quad cells in each dimension. With the pixels of a single quad cell numbered as in Figure 10, the x and y components of the subimage displacements can be approximated by:

$$\delta_x = \frac{b}{2} \frac{I_1 - I_2 - I_3 + I_4}{I_1 + I_2 + I_3 + I_4} \quad (13)$$

$$\delta_y = \frac{b}{2} \frac{I_1 + I_2 - I_3 - I_4}{I_1 + I_2 + I_3 + I_4} \quad (14)$$

where I_j is the intensity in the j^{th} pixel and b is the subimage diameter (in units of length). These expressions are true only in the limit that the displacements are small compared to b .

Note that the constant of proportionality in the relation (Equations (13) and (14)) between the desired wavefront gradient and the empirically determined intensity ratios includes a factor of the image diameter b , which is not directly measured in the quad cell geometry. This reveals a limitation of the Shack-Hartmann approach — at least in the quad cell implementation — that the overall system gain is imperfectly known.

From the above discussion it is a straightforward matter to obtain the gradient of the wavefront on a grid which covers the aperture plane. Of course what

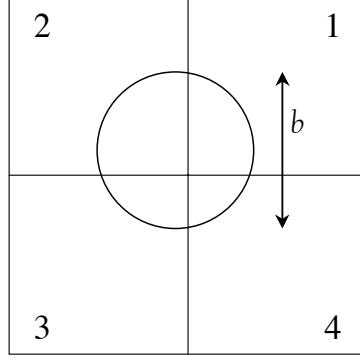


Figure 10. Two pixel by two pixel “quad cell.”

is ultimately desired is not $\vec{\nabla}z$ but the wavefront z itself. We now explore this reconstruction of the wavefront, assuming that the gradient is known.

Figure 11 represents schematically a portion of the aperture plane. Each square represents a lenslet, with the x and y components of the gradient (averaged over each lenslet) assumed to be known. The vertices between lenslets represent the “phase points” at which the unknown phases or aberrations are to be determined. For a square array of $k \times k$ lenslets, there are consequently $2k^2$ known gradient components and $(k+1)^2$ unknown phases, and the system is determined (apart from some exceptions discussed below) provided $k \geq 3$.

We adopt a finite difference approach to the problem of integrating the gradient. That is, with the four typical phase points numbered as in Figure 11, the gradient averaged over the indicated lenslet may be approximated by:

$$\frac{\partial z}{\partial x} = \frac{1}{2\Delta}(z_a - z_b - z_c + z_d) \quad (15)$$

$$\frac{\partial z}{\partial y} = \frac{1}{2\Delta}(z_a + z_b - z_c - z_d) \quad (16)$$

where Δ is the lenslet side length mapped to the primary mirror. The integration is thus reduced to a linear algebraic system of $m = 2k^2$ equations in $n = (k+1)^2$ unknowns. We write:

$$\mathbb{A} \vec{z} = \vec{b} \quad (17)$$

where \mathbb{A} is an $m \times n$ matrix with $m > n$ whose elements are defined by equations of the form of Equations (15) and (16), \vec{z} is an n -dimensional column vector whose elements are the values of the aberrations at the phase points, and \vec{b} is an m -dimensional column vector whose elements are the values of $\partial z/\partial x$ and $\partial z/\partial y$

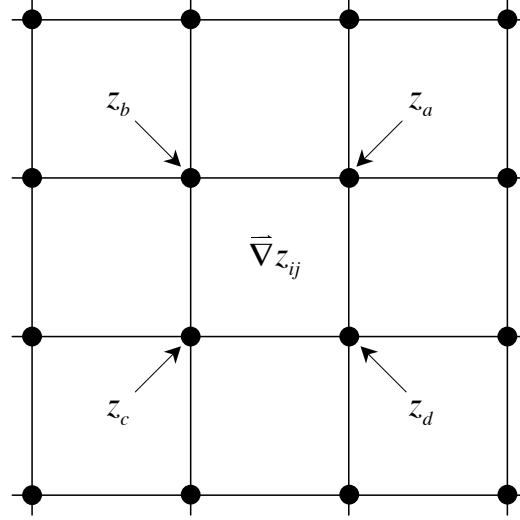


Figure 11. Representation of a portion of the (re-imaged) aperture plane in the square Shack-Hartmann geometry. Squares represent lenslets; filled circles represent phase points.

corresponding to the lenslets, or more precisely, the values of these derivatives averaged over the areas of each of the lenslets. The reconstruction problem can now be concisely stated in the context of Equation (17): given \vec{b} what is \vec{z} ?

Roughly speaking, one would like to multiply both sides of Equation (17) on the left by \mathbb{A}^{-1} . However, since the system is over-determined, the matrix \mathbb{A} is not square and a true inverse matrix (and an exact solution) does not exist. One can produce a reduced set of equations — the so-called normal equations — by multiplying both sides of Equation (17) on the left by the transpose matrix \mathbb{A}^T , resulting in n equations in n unknowns. However, a more robust approach is to construct the pseudo-inverse matrix which effectively solves the problem in a least squares sense by the technique of singular value decomposition or SVD. For a thorough discussion (as well as for the actual SVD code), see Press et al. (1989); here we briefly summarize the procedure, following the notation in that reference.

An arbitrary $m \times n$ matrix \mathbb{A} such as appears in Equation (17) can be decomposed into a product of three matrices:

$$\mathbb{A} = \mathbb{U} \mathbb{W} \mathbb{V}^T \quad (18)$$

where \mathbb{U} is an $m \times n$ column orthonormal matrix, \mathbb{W} is an $n \times n$ diagonal matrix whose diagonal elements are all non-negative, and \mathbb{V}^T is an $n \times n$ row and column orthonormal matrix. The decomposition is almost unique; the exceptions do not concern us here. The appropriate least squares solution can be constructed from the matrices \mathbb{U} , \mathbb{W} , and \mathbb{V} as follows:

$$\vec{z} = \mathbb{V} \mathbb{W}^{-1} \mathbb{U}^T \vec{b} \quad (19)$$

where \mathbb{W}^{-1} is the diagonal matrix whose diagonal elements are the reciprocals of the corresponding elements of \mathbb{W} . Note that the product of the three matrices on the right hand side of Equation (19) is the desired pseudo-inverse of \mathbb{A} .

The power of the SVD method is revealed when there is a subspace z_{null} of \vec{z} of dimension $\ell > 0$ that is mapped to zero by the matrix \mathbb{A} :

$$\mathbb{A} \vec{z}_{null} = 0 \quad (20)$$

(In the specific context of the wavefront reconstruction problem this means that there are non-trivial aberration functions whose [discretized] gradient vanishes; clearly the reconstruction can never recover such components.) Under these circumstances ℓ of the diagonal elements w_{jj} of \mathbb{W} will be zero, and the corresponding elements of \mathbb{W}^{-1} will be singular. The remedy is to replace the corresponding w_{jj}^{-1} by 0 in Equation (19). It can be shown that the resulting solution \vec{z}^* satisfies the least squares condition, i.e. it minimizes the residual R :

$$R = |\mathbb{A} \vec{z} - \vec{b}| \quad (21)$$

and at the same time sets to zero the unmeasurable components of \vec{z} , which lie in \vec{z}_{null} (and which of course cannot contribute to R).

To illustrate the above ideas in the context of the application of SVD to the Shack-Hartmann wavefront reconstruction problem, consider a square lenslet geometry such as that shown in Figure 11, but of arbitrary overall array size. One finds that exactly two elements w_{jj} are always zero, independent of the size of the array. The more obvious of these singular modes corresponds to overall piston; the other mode is the so-called “waffle mode” aberration. In Figure 11 waffle mode corresponds to $z_a = +1$, $z_b = -1$, $z_c = +1$, $z_d = -1$, with the same pattern replicated over all lenslets in the array. Both of these modes are unmeasurable by the square Shack-Hartmann geometry, but there is a fundamental difference between the two. Global piston is not only unmeasurable, but unimportant in the sense that it has no diffraction consequences; waffle mode, on the other hand, does have potentially important diffraction consequences, despite the fact that it is unmeasurable by this particular sensing geometry. In general, steps must be taken to measure waffle mode by some other method — for example by monitoring the actuators of the deformable mirror — otherwise it can increase without limit. [By contrast, one can safely ignore global piston.]

As an aside we note that waffle mode and its associated problems can be avoided by using a hexagonal instead of a square geometry. Thus consider a close packed array of hexagonal lenslets as shown in Figure 12. One might be tempted here to define all six vertices of each lenslet as phase points, analogous to the square geometry case. However, this would result in an excess of phase points, as can

be readily shown: the six vertices of each interior lenslet are shared among three lenslets; thus there are effectively two vertices per interior lenslet, but greater than two vertices per boundary lenslet. Overall, this results in greater than two vertices per lenslet, and since there are exactly two gradient components per lenslet, this assignment would result in more unknowns than equations. Therefore we define a triangular pattern of alternate vertices as phase points, as shown in Figure 12. The corresponding discretized equations for this geometry are:

$$\frac{\partial z}{\partial x} = \frac{\sqrt{3}}{3s}(z_b - z_a) \quad (22)$$

$$\frac{\partial z}{\partial y} = \frac{1}{3s}(z_a + z_b - 2z_c) \quad (23)$$

where s is the side of the hexagonal lenslet, mapped to the original aperture. One can show by construction that this triangular geometry in general has only one singular mode (global piston) and that the singularity corresponding to waffle mode is removed. Although the hexagonal geometry is superior in this respect, the square geometry has the advantage that it is well matched to the quad cell approach, which in turn makes the most efficient use of the limited pixels on a high speed CCD. For this reason, the square Shack-Hartmann geometry is the more common one in adaptive optics, and we shall not consider the hexagonal geometry further here.

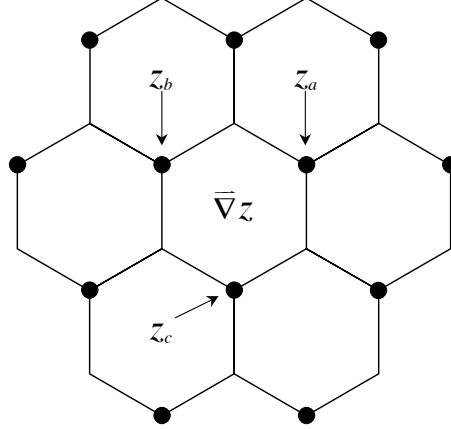


Figure 12. Representation of a portion of the (re-imaged) aperture plane in the hexagonal Shack-Hartmann geometry. Hexagons represent lenslets; filled circles represent phase points.

As an example of Shack-Hartmann/SVD wavefront reconstruction, we consider a square array of 16×16 lenslets (17×17 phase points). The complete set of orthonormal functions on the square are products of Legendre polynomials, identified by an ordered pair of indices, e.g. $L_{1,2} = \frac{1}{2}x(3y^2 - 1)$. Figure 13a shows

the (original) function $L_{2,3}$ which was used to generate the average gradients for purposes of the reconstruction, Figure 13b shows the reconstructed wavefront, and Figure 13c shows the difference of the two (reconstructed minus original function). The fractional rms error of the reconstruction, here equal to 0.0228, is defined to be the rms of the difference function divided by the square root of the variance of the original function.

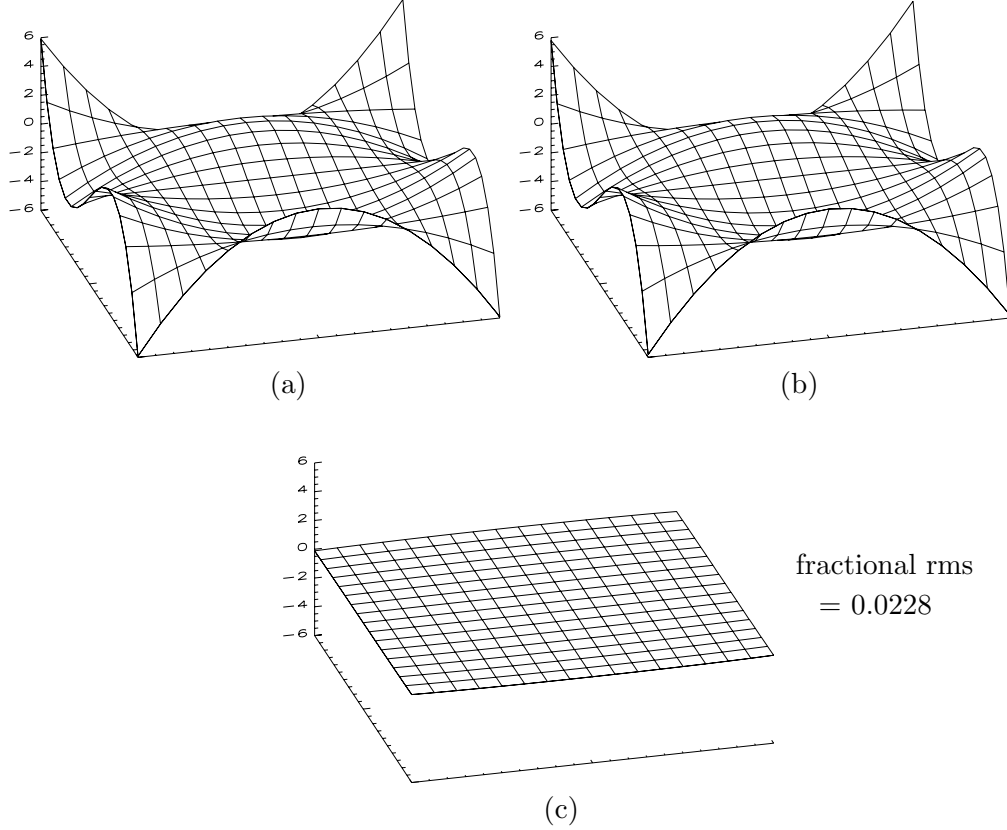


Figure 13. (a) Original function $L_{2,3}$, (b) Shack-Hartmann zonal reconstruction, (c) difference.

As the spatial frequency content of the trial function increases, so too does the fractional error, because the discretization approximation (Equations (15) and (16)) becomes worse with increasing spatial frequency. This is illustrated in Table 1, which gives the fractional error of the reconstruction as a function of the Legendre indices (i,j) (whose order does not matter in this context). For each pair (i,j) we give two fractional errors: the upper entry is the full fractional error calculated over all 17×17 phase points; the lower entry gives the fractional error restricted to the 15×15 interior phase points. Note that the reconstruction is relatively poor for the exterior phase points at high spatial frequencies.

Table 1. Fractional Errors for 16×16 Shack-Hartmann Reconstructions of Products of Legendre Polynomials.*

j	i = 1	i = 2	i = 3	i = 4	i = 5	i = 6	i = 7
0	0.000005 0.000005	0.000431 0.000439	0.000028 0.000029	0.000580 0.000597	0.000150 0.000150	0.000759 0.000786	0.000480 0.000473
1	0.000003 0.000003	0.00775 0.00823	0.0133 0.0140	0.0330 0.0350	0.0382 0.0388	0.0726 0.0754	0.0730 0.0716
2		0.0109 0.0116	0.0228 0.0240	0.0433 0.0450	0.0693 0.0691	0.1020 0.0985	0.138 0.126
3			0.0349 0.0351	0.0585 0.0562	0.0942 0.0862	0.140 0.121	0.194 0.158
4				0.0827 0.0724	0.1240 0.0981	0.179 0.136	0.251 0.175
5					0.167 0.114	0.230 0.142	0.313 0.184
6						0.297 0.168	0.389 0.196
7							0.484 0.229

*Upper entries are for full array; lower entries, interior points only.

The above wavefront reconstruction, which proceeds by integrating the data, is referred to generically as a zonal reconstruction. This is in contradistinction to a modal reconstruction, which proceeds by assuming a theoretical form for the solution (as the sum of a series of well-defined modes), differentiating the theoretical solution (with unknown coefficients) and comparing the result to the data. The reconstruction which we describe under curvature sensing (Section 5.) is also zonal, but any algorithm can be implemented in either the zonal or modal form. In the following we briefly describe the modal reconstruction in the context of the Shack-Hartmann algorithm.

Suppose that we know *a priori* that the wavefront of interest can be well-approximated as a finite sum of known functions or modes; generally this comes from the truncation of an infinite series:

$$z_i = \sum_j c_j P_j(x_i, y_i); \quad i = 1, \dots, m$$

$$j = 1, \dots, n \quad (24)$$

Here m is the number of points, n is the number of modes, the c_j are unknown coefficients, and the P_j are the known modes. Differentiating, we have:

$$\frac{\partial z_i}{\partial x} = \sum_j c_j g_j^x(x_i, y_i) \quad (25)$$

$$\frac{\partial z_i}{\partial y} = \sum_j c_j g_j^y(x_i, y_i) \quad (26)$$

where \vec{g}_j is the gradient of the j -th mode. This can be written as:

$$\mathbb{B} \vec{c} = \vec{g} \quad (27)$$

where \mathbb{B} is a $2m \times n$ matrix with known coefficients, \vec{c} is an n -element column vector whose elements are the desired coefficients of the expansion of the wavefront, and \vec{g} is a $2m$ -element vector whose first m elements are the measured x -components of the wavefront gradient, and whose last m elements are the measured y -components. The j -th column of \mathbb{B} is simply the (appropriately ordered) gradient of the j -th mode. One can solve this linear system for the unknown vector \vec{c} by the same techniques as described in the zonal section above; the wavefront is finally constructed by substituting \vec{c} into Equation (24). Optimally the modes should not only be orthogonal but independent in the sense that the resulting c_j are uncorrelated. Note that the modal approach results in a value for the wavefront everywhere, while the zonal approach only defines the wavefront at discrete points.

There is no simple answer to the question of which type of reconstructor one should use in a given situation. There is an extensive literature on optimal wavefront reconstruction, but that topic is beyond our scope; the reader is referred to reviews by Hardy (9), Rousset (22), and Wild (24). Clearly, modal methods require a greater *a priori* knowledge about the system. Generally speaking these methods are favored as the quality of the approximation leading to Equation (24) improves. By contrast, a zonal approach might be favored, for example, in an AO system that was being used not only to compensate for free atmospheric turbulence, but also for dome seeing and secondary mirror vibration, since no simple modal description could be expected to account for all such aberrations.

5. Curvature Sensing

Figure 14 illustrates (in a partly symbolic way) the basic idea of curvature sensing (19; 21). Out-of-focus images are recorded with an area detector sub-

stantially inside and outside of focus. If the wavefront aberration at a given point has positive curvature, then excess illumination will be recorded at the corresponding point on the inside-of-focus image, with a corresponding lack of illumination at the outside-of-focus image, and conversely for negative curvature.

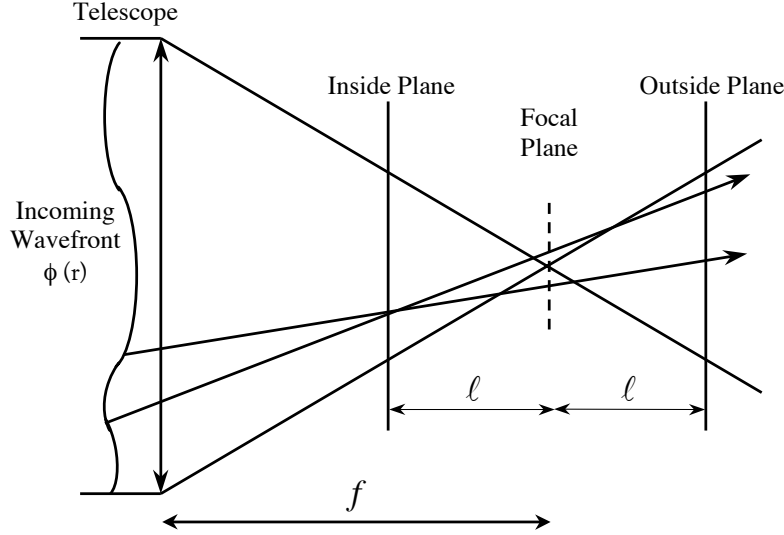


Figure 14. Curvature sensing concept. Positive curvature at a point on the incoming wavefront results in excess illumination at the corresponding point on the inside plane, lack of illumination at the outside plane, and conversely.

Roddier (20) derives the basic equation of curvature sensing from the “equation of radiative transport.” Here we pursue an alternative, somewhat more mathematical approach. [Although it is longer, this derivation is more readily adapted to the physical optics generalization of curvature sensing, which we have found to be useful in the context of phasing segmented mirrors (5) — although we do not pursue this generalization further in the present work.]

In general, in the full diffraction calculation, we go from the pupil or aperture plane to the image plane via a Fourier transform:

$$\hat{A}_k(\vec{\omega}) = \int A(\vec{r}) e^{ik\vec{r} \cdot \vec{\omega}} d^2\vec{r} \quad (28)$$

Here $A(\vec{r})$ is the complex amplitude of the electric field in the aperture plane; \vec{r} , with rectangular coordinates (x, y) and dimensions of length, is the position vector in the aperture plane; $\hat{A}_k(\vec{\omega})$ is the complex amplitude in the image plane; $\vec{\omega}$, with rectangular coordinates (u, v) and dimensions of radians, is the position vector in the image plane; $k = 2\pi/\lambda$; and the integral is over the aperture.

We assume that the illumination is uniform, in which case:

$$A(\vec{r}) = A_0 e^{i\phi(\vec{r})} \quad (29)$$

where A_0 is real; as before, the phase error $\phi(\vec{r})$ in radians is related to the wavefront error $z(\vec{r})$ in units of length according to $\phi = kz$. In the present application we are interested in the geometrical optics limit:

$$\lim_{k \rightarrow \infty} \hat{A}_k(\vec{\omega}) = \lim_{k \rightarrow \infty} A_0 \int e^{ikf(x,y)} dx dy \quad (30)$$

where

$$f(x, y) = ux + vy + z_{tot} \quad (31)$$

Here the subscript on z is to indicate that this includes both the (original) wavefront error of interest and the wavefront error associated with the deliberate defocus. Writing this latter quadratic term explicitly we have:

$$f(x, y) = ux + vy \pm h(x^2 + y^2) + z(x, y) \quad (32)$$

where z is now (just) the wavefront error of interest and the coefficient h is to be determined.

Integrals such as that on the right hand side of Equation (30), with rapidly varying exponential integrands, can be evaluated by the method of stationary phases. A discussion can be found in Born & Wolf (2) and references therein; here we just quote the results (in the context of Equation (30)).

For a given $\vec{\omega} = (u, v)$, only regions in a neighborhood of a “critical point” contribute to the integral. In general there are three kinds of critical points (x_o, y_o) :

- (1) (x_o, y_o) is interior to the aperture and

$$\frac{\partial f}{\partial x} = 0, \quad \frac{\partial f}{\partial y} = 0. \quad (33)$$

- (2) (x_o, y_o) is on the boundary of the curve defining the aperture and $\partial f / \partial s = 0$, where ds is an element of the arc of the bounding curve.
- (3) (x_o, y_o) is at a kink in the boundary curve, e.g. at the corners of a square aperture.

For critical points of the first kind, the leading term in the approximation of the right hand side of Equation (30) is given by:

$$\lim_{k \rightarrow \infty} \hat{A}_k(\vec{\omega}) = \frac{2\pi i}{[f_{xx}f_{yy} - f_{xy}^2]^{\frac{1}{2}}} \frac{e^{ikf(x,y)}}{k} \Big|_{(x_0, y_0)} \quad (34)$$

provided that the argument of the square root is positive (as it can be shown to be here) and where f_{xx} , f_{yy} , and f_{xy} are the partial second derivatives of f . Imposing conditions (33) we obtain:

$$0 = u \pm 2hx + \frac{\partial z}{\partial x} \implies x \cong \mp \frac{u}{2h} \quad (35)$$

$$0 = v \pm 2hy + \frac{\partial z}{\partial y} \implies y \cong \mp \frac{v}{2h} \quad (36)$$

i.e. there is an approximately linear mapping from pupil to image. (The neglect of the gradient terms results from anticipating the approximation below.) Furthermore:

$$f_{xx} = \pm 2h + \frac{\partial^2 z}{\partial x^2} \quad (37)$$

$$f_{yy} = \pm 2h + \frac{\partial^2 z}{\partial y^2} \quad (38)$$

$$f_{xy} = \frac{\partial^2 z}{\partial x \partial y} \quad (39)$$

If we assume that h is sufficiently large that

$$\nabla^2 z \ll 2h \quad (40)$$

then

$$f_{xx}f_{yy} - f_{xy}^2 \cong 4h^2 \left(1 \pm \frac{1}{2h} \nabla^2 z \right) \quad (41)$$

and finally:

$$I_{\pm} = I_0 \left(1 \mp \frac{1}{2h} \nabla^2 z \right) \quad (42)$$

It remains to evaluate h . In practice, we do not physically move the detector; rather we change the focal length, holding the detector fixed. This is represented in Figure 15: f_+ is the focal length corresponding to the positive focus error (with the detector outside-of-focus), and conversely for f_- . The defocus distances are defined by $\ell_+ = f - f_+$ and $\ell_- = f_- - f$, where f is the nominal focal length. Let D be the diameter of the pupil and d be the diameter of the defocused spot. Simple geometry implies

$$\frac{d}{\ell_{\pm}} = \frac{D}{f_{\pm}} \quad (43)$$

which in turn yields:

$$\frac{f + \ell_-}{\ell_-} = \frac{f - \ell_+}{\ell_+} \quad (44)$$

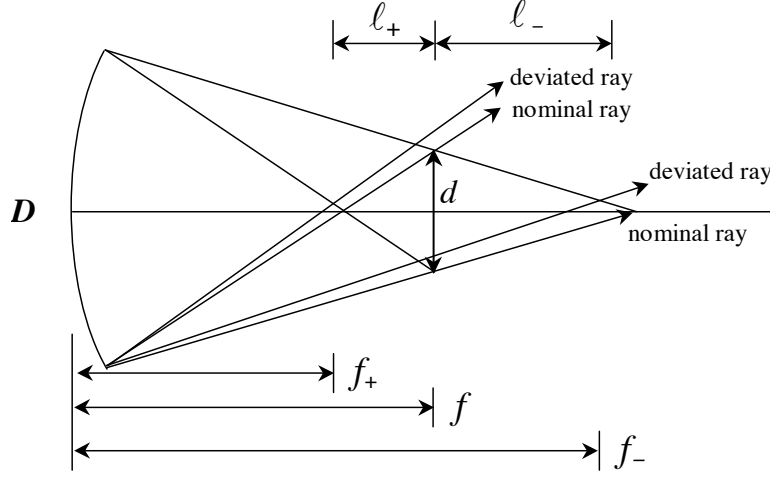


Figure 15. Geometrical relationships of the various curvature sensing foci and of a typical deviated ray from the boundary.

It is easy to show that h is related to f_+ and f_- by

$$\frac{1}{2f} \pm h = \frac{1}{2f_{\pm}} \quad (45)$$

and thus:

$$\frac{1}{2h} = \frac{f(f - \ell_+)}{\ell_+} = \frac{f(f + \ell_-)}{\ell_-} \quad (46)$$

For simplicity we write $\ell = \ell_+$ (but note that ℓ_+ and ℓ_- are not equal; they differ by a term of order ℓ/f). Finally we have

$$I_{\pm} = I_0 \left(1 \mp \frac{f(f - \ell)}{\ell} \nabla^2 z \right) \quad (47)$$

or:

$$\frac{I_- - I_+}{I_- + I_+} = \frac{f(f - \ell)}{\ell} \nabla^2 z \quad (48)$$

The original restriction on h can now be rewritten as a lower bound on the defocus distance:

$$\ell \gg f(f - \ell) \nabla^2 z \quad (49)$$

Note that in principle one can do “one-sided” curvature sensing as defined by either the upper or lower signs in Equation (47). In adaptive optics applications, one shifts rapidly between the plus and minus images and utilizes Equation (48) so that scintillation effects are cancelled out and not mistaken for curvature variations.

We still have to consider critical points of types 2 and 3. Although type 3 points (if they exist) do contain wavefront information, it is often difficult to extract this, and for practical purposes these can be ignored in the wavefront reconstruction (see the discussion of square apertures below). At this point we depart from the above formalism and evaluate the contributions of type 2 points by a simple construction.

We return to Figure 15, which shows both deviated and ideal rays from the boundary of the mirror, taken for the moment to be circular for simplicity. From inspection of the figure we note that too steep a slope at the boundary subtracts width from the image I_- (detector inside-of-focus) and adds width to the image I_+ (detector outside-of-focus). To zeroth order we have:

$$I_- = \begin{cases} I_0 & \rho < \frac{d}{2} - f \frac{\partial z}{\partial r} \\ 0 & \rho > \frac{d}{2} - f \frac{\partial z}{\partial r} \end{cases} \quad (50)$$

$$I_+ = \begin{cases} I_0 & \rho < \frac{d}{2} + f \frac{\partial z}{\partial r} \\ 0 & \rho > \frac{d}{2} + f \frac{\partial z}{\partial r} \end{cases} \quad (51)$$

where ρ is the radial coordinate in the image plane, d is the diameter of the defocused image, and these conditions are expressed at the (defocused) image plane, not the pupil plane. Thus, too steep a slope adds width w to the difference image in the amount $2f \partial z / \partial \rho$ in the image plane, or — referred to the aperture plane:

$$w = \frac{2f(f - \ell)}{\ell} \frac{\partial z}{\partial \rho} \quad (52)$$

A curvature sensing difference image therefore consists of a zeroth order boundary image whose width is proportional to the normal derivative of the wavefront at the boundary, plus a first order interior image whose magnitude is proportional to the Laplacian of the wavefront. [Both the width and the Laplacian or

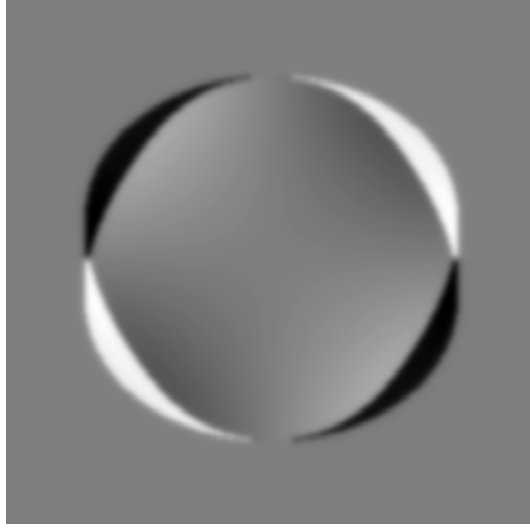


Figure 16. Simulated curvature sensing difference image corresponding to the pure Zernike term $Z_{4,-2}$ for a circular aperture. It is easy to distinguish the zeroth order boundary image from the first order interior image.

curvature are signed quantities.] A simulated difference image corresponding to the pure Zernike term $Z_{4,-2}$ is shown in Figure 16. The reconstruction of the wavefront from the difference image is entirely analogous to solving Poisson's equation with Neumann boundary conditions. Note that ideally the normal derivative at the boundary could be measured directly from the width of the zeroth order boundary term; in practice, seeing (and possibly diffraction) effects blur out the boundary image so that the width is no longer well-defined. Nonetheless, the integrated intensity of the zeroth order terms is preserved so that:

$$\int (I_+ - I_-) dr = 2I_0 f \frac{(f - \ell)}{\ell} \frac{\partial z}{\partial r} \quad (53)$$

where the integral is over a neighborhood of the boundary and the variable of integration is the coordinate that increases perpendicular to the boundary. Ignoring type 3 points, collecting terms, and writing the radial derivative in more general form as the normal derivative, we have:

$$\frac{I_- - I_+}{I_- + I_+} = \frac{f(f - \ell)}{\ell} \left\{ \nabla^2 z - \delta_b \frac{\partial z}{\partial n} \right\} \quad (54)$$

where the term containing the delta function δ_b means that the integral of this term over a small region perpendicular to the boundary yields $-\partial z / \partial n$.

Note that the presence of the boundary term $\partial z/\partial n$ in Equation (54) is not a mere technical detail. For circular apertures, (exactly) two Zernike terms of every order are defined completely by this boundary condition, since the Laplacian vanishes for

$$Z_{n,\pm n}(r, \theta) = r^n \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} n\theta \quad (55)$$

The curvature sensing approach has the advantage of opto-mechanical simplicity over Shack-Hartmann, since no lenslet array, or for that matter re-imaging of the pupil, is required (although a mechanism for rapidly changing the focus is needed). In addition, the sensitivity of a curvature sensor can be varied readily and continuously by varying the out-of-focus distance ℓ . In the Shack-Hartmann test, by contrast, the sensitivity is changed (with difficulty) by changing the focal length of the lenslets; this can only be done discretely, and then only by physically exchanging the entire lenslet array. Finally, a curvature wavefront sensor can be efficiently combined with a bimorph deformable mirror (see Sechaud (23) and references therein). Since the equation which controls the surface of the latter is also a Poisson equation, the wavefront can be reconstructed in analog fashion, without the need for an additional computer calculation. A potential disadvantage of the curvature approach is the fact that the zeroth order intensity effects can overwhelm the first order effects in the neighborhood of the boundary.

The preceding discussion shows that we can extract both $\nabla^2 z$ and $\partial z/\partial n$ from the curvature image or difference image. How then do we reconstruct the wavefront aberration z ? As noted above, this is exactly equivalent to solving Poisson's equation in electricity and magnetism with Neumann boundary conditions. There are a variety of numerical approaches. Here we adopt one that is both simple and closely analogous to the Shack-Hartmann reconstructor discussed in Section 4..

We assume that $\nabla^2 z$ and $\partial z/\partial n$ have already been extracted from the image. For simplicity we consider a square geometry which has been subdivided into small square regions (which may be, for example, pixels on a CCD or groups of pixels). Various subsets of these small square regions are shown for labeling purposes in Figure 17a, 17c, and 17e. We may approximate the second derivatives of z_{ij} by means of finite differences. If the dimensions of these subdivisions mapped to the original aperture plane are Δ on a side, and if z_{ij} lies in an interior region of the aperture, then we have approximately:

$$\frac{\partial^2 z_{ij}}{\partial x^2} = \frac{1}{\Delta} \left\{ \left(\frac{z_{i+1,j} - z_{ij}}{\Delta} \right) - \left(\frac{z_{ij} - z_{i-1,j}}{\Delta} \right) \right\} \quad (56)$$

$$= \frac{1}{\Delta^2} (z_{i+1,j} - 2z_{ij} + z_{i-1,j}) \quad (57)$$

and similarly for $\partial^2 z_{ij}/\partial y^2$. The coefficients of the z_{ij} from the various subdivisions which appear in the expression for $\nabla^2 z_{ij}$ are shown schematically in Figure 17b.

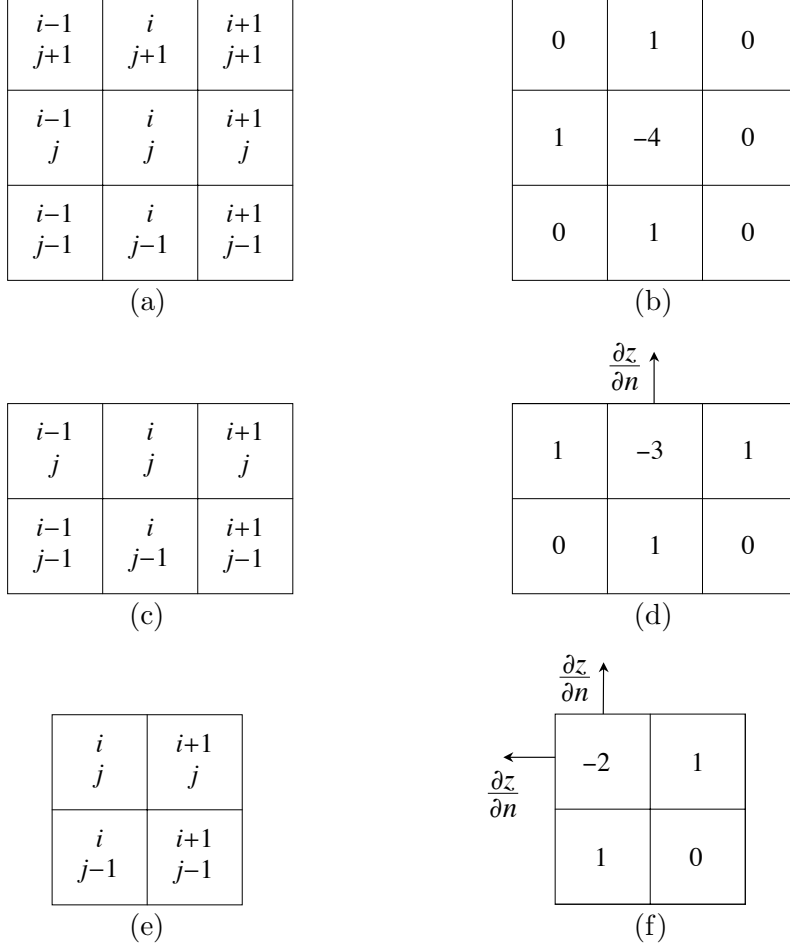


Figure 17. (a-b) Symbolic construction of curvature sensing A-matrix for interior point (i, j) ; (c-d) same as (a) and (b) but for edge point; (e-f) same as (a) and (b) but for corner point.

Now if z_{ij} lies, say, in a boundary region along the top edge of the aperture (see Figure 17c), but not in a corner, then the above construction for $\partial^2 z/\partial y^2$ fails; however, we may utilize the normal derivative in an equivalent approximation:

$$\frac{\partial^2 z_{ij}}{\partial y^2} = \frac{1}{\Delta} \left\{ \frac{\partial z_{ij}}{\partial y} - \frac{1}{\Delta} (z_{ij} - z_{i,j-1}) \right\} \quad (58)$$

Finally, if z_{ij} lies, say, in the upper left corner region (Figure 17e) we have

$$\nabla^2 z_{ij} = \frac{1}{\Delta} \left\{ \frac{\partial z_{ij}}{\partial y} - \frac{\partial z_{ij}}{\partial x} - \frac{(2z_{ij} - z_{i+1,j} - z_{i,j-1})}{\Delta} \right\} \quad (59)$$

The constructions corresponding to the cases treated in Equations (58) and (59) are shown schematically in Figures 17d and 17f, respectively.

In general we can represent the overall finite-difference equations by a linear system of the form:

$$\mathbb{A} \vec{z} = \nabla^2 \vec{z} - \vec{\eta} \quad (60)$$

where \mathbb{A} is a matrix with integer coefficients (divided by Δ^2), $\nabla^2 \vec{z}$ is a data vector whose components are the measured Laplacians of the subregions and $\vec{\eta}$ is another data vector whose components are zero for interior points, the measured normal derivatives $\partial z / \partial n$ (divided by Δ) for edge points, and the sum of two measured normal derivatives (again divided by Δ) for corner points. The differential equation has been reduced to a system of linear algebraic equations whose least squares solution can again be obtained by the technique of singular value decomposition.

As in the above Shack-Hartmann section, we here present a sample curvature sensing wavefront reconstruction using singular value decomposition. We assume that the appropriately averaged Laplacian of the wavefront has already been extracted from a 16×16 array of square subregions and also that the (averaged) normal derivative of the wavefront has been extracted for each boundary region (including two normal derivatives for each corner region as in Figure 17f). Again we consider a function which is the product of two one-dimensional Legendre polynomials. Figure 18a shows the (original) function $L_{3,3}$ which was used to generate the average derivatives for the reconstruction, Figure 18b shows the reconstructed wavefront, and Figure 18c shows the difference of the two. The fractional rms error of the reconstruction, defined as in Section 4., is equal to 0.0175.

As was the case for the Shack-Hartmann reconstruction, the fractional error tends to increase with the spatial frequency content of the trial function, as the discretization approximation (here, Equations (56) – (59)) breaks down. This is illustrated in Table 2, which gives the fractional error of the reconstruction as a function of the Legendre indices (i, j) .

6. Pyramid Wavefront Sensing

Pyramid wavefront sensing is a relatively new technique (13; 14; 17), which is a quantitative two-dimensional version of a classical qualitative one-dimensional optical test — the knife edge test. To place pyramid sensing in its proper context, we first briefly review the knife edge test (see e.g. 12).

In the knife edge test (and also in pyramid wavefront sensing) the situation

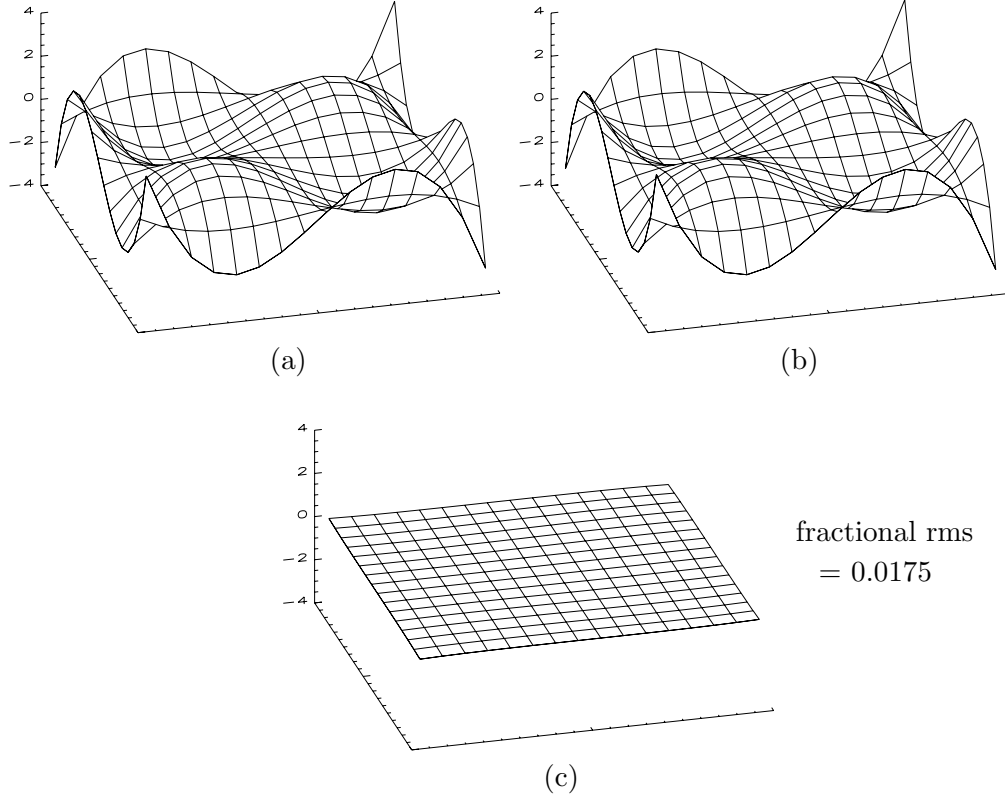


Figure 18. (a) Original function $L_{3,3}$, (b) curvature sensing reconstruction, (c) difference.

is reversed from Shack-Hartmann and curvature sensing in the sense that the wavefront is manipulated in the image plane and the detector is in the pupil plane. Specifically, a knife edge, located in the focal plane and which for the sake of definiteness we take to be parallel to the u -axis, is moved in the $+v$ direction. Rays which cross the focal plane with v -values above the position of the knife edge will be transmitted; otherwise they will be blocked. [See Figure 19.] An exit pupil is formed downstream of the knife-edge assembly.

Consider a point (x, y) in the entrance pupil plane at which the wavefront error is z . When the knife edge is at the position θ_v in the image plane, the corresponding point in the exit pupil will be illuminated if and only if

$$\frac{\partial z}{\partial y} > \theta_v \quad (61)$$

Thus the knife edge data consist of a sequence of images; in each image of the sequence the illumination at each point is either a constant or 0, according to whether or not the inequality (61) is satisfied. Qualitative information about the

Table 2. Fractional Errors for 16×16 Curvature Sensing Reconstructions of Products of Legendre Polynomials

j	i = 1	i = 2	i = 3	i = 4	i = 5	i = 6	i = 7
0	0.000023	0.000014	0.0153	0.0321	0.0761	0.130	0.205
1	0.000036	0.00438	0.00205	0.0291	0.0503	0.124	0.177
2		0.00619	0.0104	0.0223	0.0424	0.091	0.152
3			0.0175	0.0285	0.0462	0.0802	0.132
4				0.0410	0.0590	0.0872	0.128
5					0.0784	0.105	0.139
6						0.130	0.161
7							0.191

wavefront error can be readily inferred from the sequence of images, particularly for simple, low order aberrations. For example, suppose that the wavefront error consists of a positive focus error, i.e. the wavefront is “over-focused” as shown in Figure 19. It is clear from the Figure that the shadow of the knife in the pupil plane will move parallel to the motion of the knife in the image plane for this aberration. Similarly for a negative “under-focused” wavefront, the shadow will move anti-parallel to the motion of the knife, and for “ x - y astigmatism” the shadow will move perpendicular to the motion of the knife. The sequences of images corresponding to these aberrations are shown in Figure 20a through c. On the other hand, for aberrations of even moderate spatial frequency, the knife edge sequences can be very complicated and difficult to interpret: Figure 20d shows the sequence corresponding to the Zernike aberration $Z_{5,3}$.

Note that the knife edge is used to sort points in the pupil plane into two categories according to whether the corresponding slope errors satisfy inequality (61) or not. In pyramid sensing, the points are sorted into four categories defined by two inequalities, one for x and one for y . The actual sorting is accomplished by a transmissive pyramid located in and with its base parallel to the focal plane, with the four categories corresponding to the four facets of the pyramid. As in the knife edge test, an exit pupil — or in this case, a set of four exit pupils — is formed downstream of the pyramid assembly. A side view of a pyramid sensor, showing two facets and the two corresponding exit pupils is presented in Figure 21a. A top view of such a sensor is shown in Figure 21b. For definiteness we take the pyramid to be oriented so that the projections of the edges between

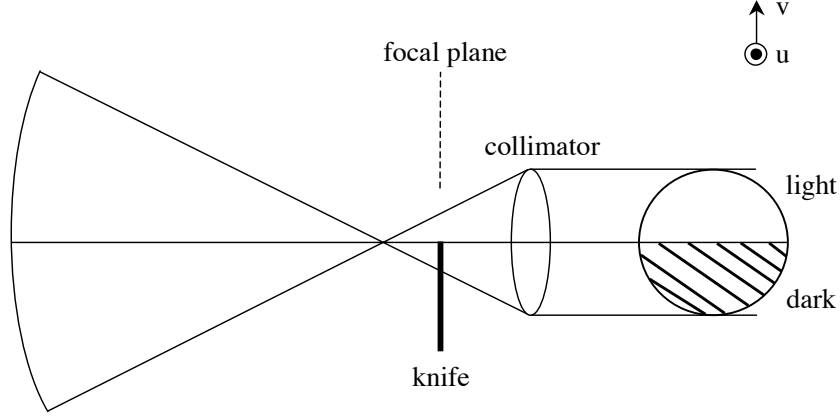


Figure 19. The geometry of the knife edge test. For an over-focused wavefront as shown, the shadow of the knife edge moves parallel to the knife.

the four triangular facets are parallel to the u and v axes. With the facets numbered as shown, rays are sorted into the corresponding exit pupils according to the four inequalities:

$$\begin{aligned}
 1. \quad & \frac{\partial z}{\partial x} > \theta_u, & \frac{\partial z}{\partial y} > \theta_v \\
 2. \quad & \frac{\partial z}{\partial x} < \theta_u, & \frac{\partial z}{\partial y} > \theta_v \\
 3. \quad & \frac{\partial z}{\partial x} < \theta_u, & \frac{\partial z}{\partial y} < \theta_v \\
 4. \quad & \frac{\partial z}{\partial x} > \theta_u, & \frac{\partial z}{\partial y} < \theta_v
 \end{aligned} \tag{62}$$

As with the knife edge test, the pyramid is in motion during the test. For simplicity we suppose that the apex of the pyramid traces out a circle centered on the origin of the focal plane, with the projected edges between triangular facets always parallel to the u and v axes. The radius of the circle is taken to be equal to the largest angular aberration of any ray, and the triangular facets should be large enough (ideally) so that no ray misses the pyramid at any position in its “orbit.”

Note that unlike the knife edge test, in a pyramid sensor the information is two dimensional, not one dimensional, and no photons are wasted — rather every photon shows up in one of the four pupil images; furthermore, as shown below, the information is quantitative, not qualitative, and the information is contained not in a sequence of images, but in the four time-averaged pupil images.

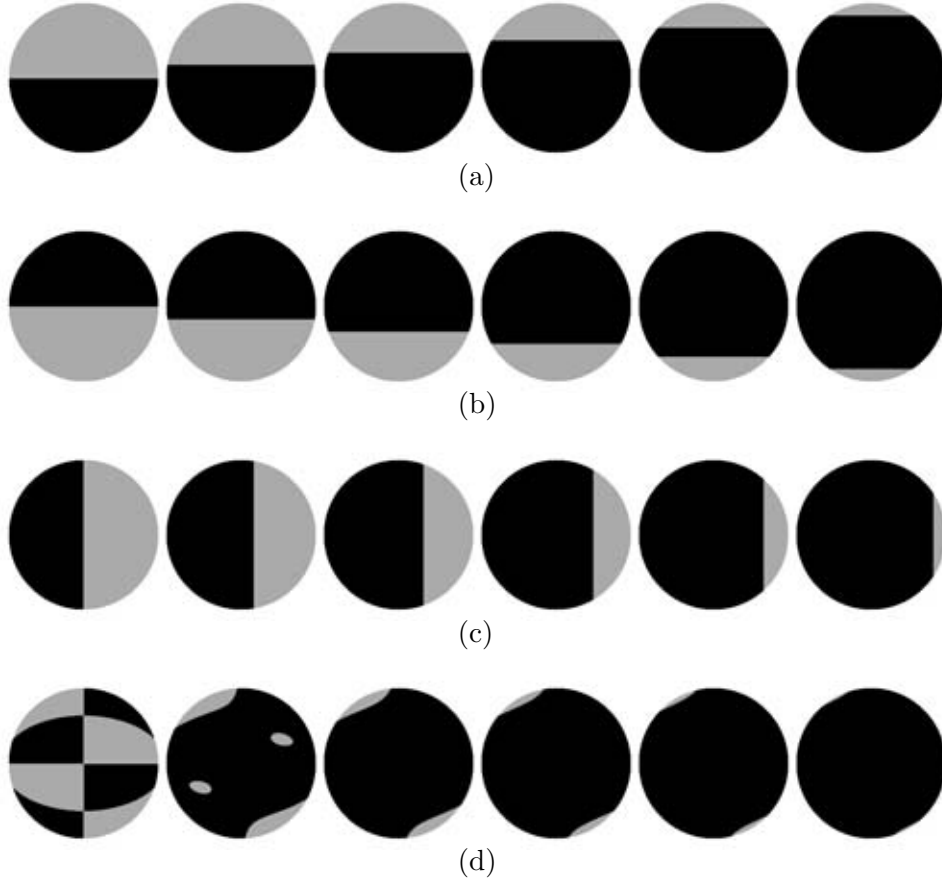


Figure 20. Knife edge sequences corresponding to (a) positive focus error, (b) negative focus error, (c) x - y astigmatism, and (d) Zernike aberration $Z_{5,3}$.

To illustrate the extraction of quantitative slope information from the time-averaged images, we define the fractional illumination of a given pixel (x, y) in images 1 and 4 (combined) as:

$$f_u = \frac{I_1 + I_4}{I_1 + I_2 + I_3 + I_4} \quad (63)$$

where all symbols are implicit functions of (x, y) . Similarly we have

$$f_v = \frac{I_1 + I_2}{I_1 + I_2 + I_3 + I_4} \quad (64)$$

It is also useful to define the related illumination parameters:

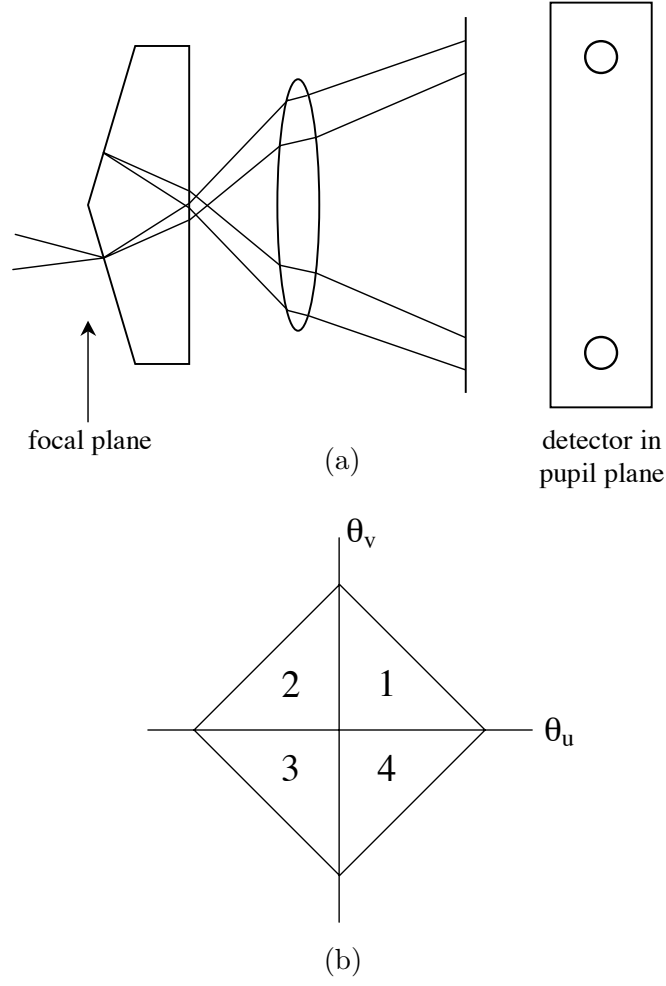


Figure 21. Pyramid sensing geometry: (a) side view, and (b) top view.

$$\alpha_u = \frac{I_1 - I_2 - I_3 + I_4}{I_1 + I_2 + I_3 + I_4} = 2f_u - 1 \quad (65)$$

and

$$\alpha_v = \frac{I_1 + I_2 - I_3 - I_4}{I_1 + I_2 + I_3 + I_4} = 2f_v - 1 \quad (66)$$

Now let $\theta_u(t)$ be the horizontal coordinate of the apex of the pyramid as a function of time t . We have:

$$\theta_u(t) = \theta_m \cos \omega t \quad (67)$$

where θ_m is the radius of the circle traced out by the apex. An arbitrary slope

error $\partial z/\partial x$ associated with the point (x,y) is shown on the plot of $\theta_u(t)$ in Figure 22. When θ_u is above this value the corresponding pixel will be illuminated in pupil image 2 or 3; when it is below this value, the pixel will be illuminated in image 1 or 4. Thus:

$$1 - f_v = \frac{\beta}{\pi} \quad (68)$$

where β is defined by the construction indicated in the Figure:

$$\theta_u = \theta_m \cos \beta \quad (69)$$

Some simple algebra readily yields expressions for the x and y slope errors at a given point in the pupil as functions of the experimentally determined illumination parameters:

$$\frac{\partial z}{\partial x} = \theta_m \cos \frac{\pi}{2} (1 - \alpha_x) \quad (70)$$

$$\frac{\partial z}{\partial y} = \theta_m \cos \frac{\pi}{2} (1 - \alpha_y) \quad (71)$$

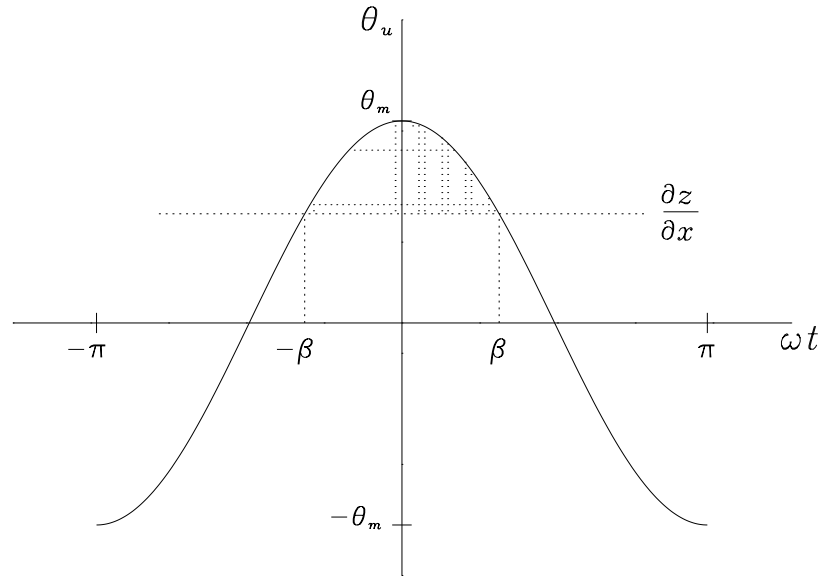
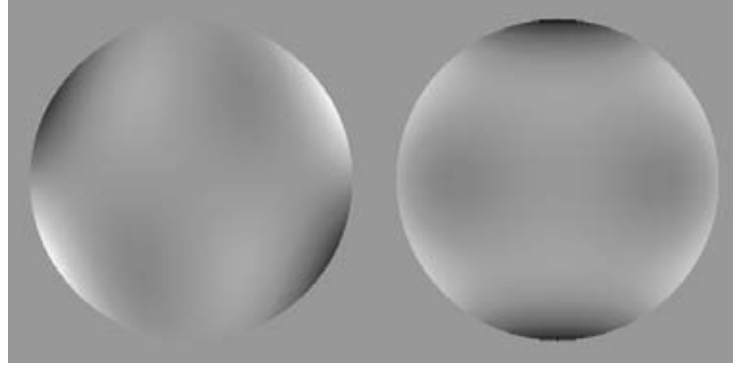


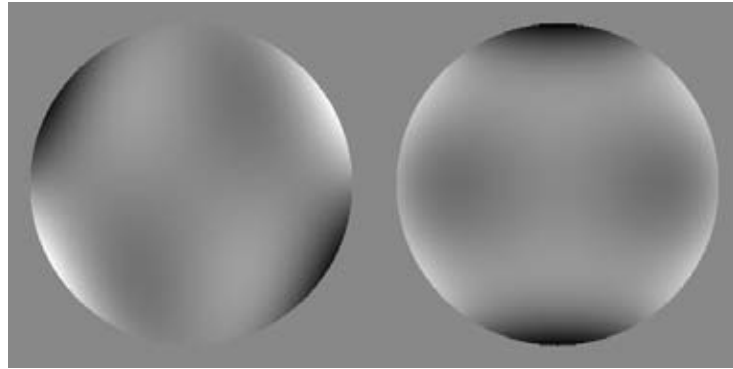
Figure 22. Relationships between the pyramid sensing parameters in Equations (67)–(69).

In Figure 23a we have plotted the slope errors for a pure Zernike aberration $Z_{5,-3}$, obtained from Equations (70) and (71) by means of a symmetrized Monte

Carlo simulation in which the angular position of the pyramid was drawn randomly from a flat distribution. Figure 23b shows the corresponding slope errors obtained by analytically differentiating the same surface error. The close correspondence between the two sets of figures confirms the above analysis.



(a)



(b)

Figure 23. (a) The x (left) and y (right) components of wavefront gradient for pure Zernike term $Z_{5,-3}$ as extracted from a Monte Carlo simulation of pyramid wavefront sensing. (b) Same as (a) but obtained by analytically differentiating the wavefront.

Wavefront reconstruction for a pyramid wavefront sensor is virtually identical to that for a Shack-Hartmann sensor, since both methods sense the wavefront gradient. The pyramid sensor shares the advantage of a curvature sensor that the sensitivity can be varied easily and continuously to match the needs of a given situation; there may be particularly large advantages as one approaches the diffraction limit (15). Furthermore, the pupil sampling may be varied if one employs a zoom lens to form the exit pupils. A disadvantage of the pyramid sensing approach is that it is fundamentally dynamic, and a suitable mechanism must be employed to dither the pyramid with respect to the image with a pe-

riod that is short compared to the characteristic timescale of the aberrations of interest.

7. Acknowledgements

I am grateful to Catherine Ohara, Edwin Sirko, and Patricia Van Buskirk for assistance with the manuscript, figures, and calculations, and to Matthias Schoeck for a critical reading of the manuscript. This research is based upon work supported in part by the Science and Technology Centers Program of the National Science Foundation under Agreement No. AST-9876783.

References

- Barnes, T. G., Adams, M. T., Booth, J. A., Cornell, M. E., Gaffney, N., Fowler, J. R., Hill, G. J., Hill, G. M., Nance, C., Piche, F., Ramsey, L. W., Ricklefs, R. L., Spiesman, W. J., & Worthington, P. T. 2000, in *Telescope Structures, Enclosures, Controls, Assembly/Integration/Validation, and Commissioning*, ed. T. A. Sebring, Vol. 4004, SPIE, 14
- Born, M. & Wolf, E. 1969, *Principles of Optics*, 6th edn. (New York: Pergamon Press)
- Chanan, G. A., Ohara, C., & Troy, M. 2000, *Appl.Optics*, in press
- Chanan, G. A., Troy, M., Dekens, F. G., Michaels, S., Nelson, J., Mast, T., & Kirkman, D. 1998, *Appl.Optics*, 37, 140
- Chanan, G. A., Troy, M., & Sirko, E. 1999, *Appl.Optics*, 38, 704
- Fried, D. L. 1966, *J. Opt. Soc. Am.*, 56, 1372
- Gerchberg, R. W. & Saxton, W. O. 1972, *Optik*, 35, 237
- Ghozeil, I. 1992, in *Optical Shop Testing*, ed. D. Malacara (New York: Wiley), 367
- Hardy, J. W. 1998, *Adaptive Optics for Astronomical Telescopes* (New York: Oxford University Press)
- Kendrick, R. L., Acton, D. S., & Duncan, A. L. 1994, *Appl.Optics*, 33, 6533
- Noll, R. J. 1976, *J. Opt. Soc. Am.*, 66, 207
- Ojeda-Castaneda, J. 1992, in *Optical Shop Testing*, ed. D. Malacara (New York: Wiley), 265
- Pugh, W. N., Lobb, D. R., Walker, D. D., & Williams, T. L. 1995, in *Adaptive Optical Systems and Applications*, ed. R. K. Tyson & R. Q. Fugate, Vol. 2534, SPIE, 312
- Ragazzoni, R. 1996, *J. Mod. Optics*, 43, 289

- Ragazzoni, R. & Farinato, J. 1999, *Astron. Astrophys.*, 350, L23
- Redding, D. C., Basinger, S. A., Lowman, A. E., Kissil, A., Bely, P. Y., Burg, R., Lyon, R. G., Mosier, G. E., Femiano, M., Wilson, M. E., Schunk, R. G., Craig, L., Jacobson, D. N., Rakoczy, J., & Hadaway, J. B. 1998, in *Space Telescopes and Instruments V*, ed. P. Y. Bely & J. B. Breckinridge, Vol. 3356, SPIE, 758
- Riccardi, A., Bindi, N., Ragazzoni, R., Esposito, S., & Stefanini, P. 1998, in *Adaptive Optical System Technologies*, ed. D. Bonaccini & R. K. Tyson, Vol. 3353, SPIE, 941
- Richardson, M. F. 1983, Master's thesis, University of Arizona
- Roddier, F. 1988, *Appl. Optics*, 27, 1223
- . 1990, *Appl. Optics*, 29, 1402
- Roddier, N. 1991, in *Active and Adaptive Optical Systems*, ed. M. A. Ealey, Vol. 1542, SPIE, 120
- Rousset, G. 1999, in *Adaptive Optics in Astronomy*, ed. F. Roddier (Cambridge: Cambridge), 91
- Sechaud, M. 1999, in *Adaptive Optics in Astronomy*, ed. F. Roddier (Cambridge: Cambridge), 57
- Wild, W. J. 2000, in *Adaptive Optics Engineering Handbook*, ed. R. Tyson (New York: Marcel Dekker), 199
- Wizinowich, P., Acton, D. S., Shelton, C., Stomski, P., Gathright, J., Ho, K., Lupton, W., Tsubota, K., Lai, O., Max, C. E., Brase, J., An, J., Avicola, K., Olivier, S., Gavel, D., Macintosh, B., Ghez, A., & Larkin, J. 2000, *PASP*, 112, 315
- Wizinowich, P. L., Acton, D. S., Gregory, T., Stomski, P. J., An, J. R., Avicola, K., Brase, J. M., Friedman, H. W., Gavel, D. T., & Max, C. E. 1998, in *Adaptive Optical System Technologies*, ed. D. Bonaccini & R. K. Tyson, Vol. 3353, SPIE, 568