Towards practical wavefront sensing at the fundamental information limit

C Paterson
The Blackett Laboratory, Imperial College London, London SW7 2BW, UK
E-mail: carl.paterson@imperial.ac.uk

Abstract. The performance of adaptive optics systems, such as for astronomical telescopes, is often limited by the available light for wavefront sensing. The efficiency of the wavefront sensor is therefore of vital importance. We have recently used a Fisher information approach to derive a fundamental limit for the measurement of wavefront phase for adaptive optics systems at low light levels. The limit, which arises from the quantized nature of light, suggests that it may be possible for adaptive optics systems to operate at light levels close to the one photon per correction mode.

1. Introduction
Adaptive optics systems that compensate dynamically for randomly varying wavefront aberrations are proving invaluable in applications such as ground-based astronomy, retinal imaging and microscopy. These systems involve measuring or estimating the optical wavefront distortion and correcting it with an active optical element such as a deformable mirror. In many situations, the light available for wavefront sensing is extremely limited and it is the efficiency of the wavefront sensor that limits the performance of the system.

The ability to measure the wavefront depends on a number of factors, but is ultimately limited by noise, both from the detector and photon noise. Dyson was the first to investigate the effects of photon noise in adaptive feedback systems, deriving optimal feedback algorithms and calculating theoretical estimates of the performance limits for several wavefront sensing configurations [1]. There have been many of studies on improving wavefront estimation accuracy for the widely used Shack–Hartmann sensor [2, 3, 4, 5, 6, 7, 8, 9]. Cannon used maximum-likelihood on the raw Shack–Hartmann image to show that there was much more information available for estimation than from spot centroiding [10]. Schulz et al used an approach based on the Cramér–Rao (CR) lower bound to show that the information available from a conventional image can be greater than that from the Shack–Hartmann [11]. The CR lower bound provides a theoretical limit to the estimator accuracy and given a physical model for a wavefront sensor, it is relatively simple to calculate [12, 6]. Several studies have also compared the performances of various existing wavefront sensor [13, 14, 15, 16].

In this paper, rather than try to compare particular wavefront sensors, we address the question of what is the fundamental information limit to wavefront sensing. In other words, what is the maximum information that can be obtained from a limited number of photo-detections given complete freedom in the design of the optical measurement.
2. Information Limit

We start with the most general description of a wavefront sensor. Our canonical wavefront sensor consists of any arrangement of passive optical elements (lenses, beamsplitters, mirrors, diffractive elements and so forth) and any number of photo-detectors. The role of the optics is to transform the input optical field so that the spatial distribution of phase results in measurable irradiance modulation at the detectors. We follow a similar approach to ref [17], but whereas in that work we assumed small phase perturbations, here we derive the general case. We write the complex amplitude of the optical field at the input as

\[ u(x) = u_0(x) \exp[i\phi(x)], \]

where \( \phi(x) \) is the spatial phase aberration that we wish to measure, and where \( x \) denotes the generalized coordinates over the input, which may be a simple aperture as in the case of a simple telescope or microscope, or may comprise multiple apertures such as for an interferometer. We represent \( \phi(x) \) in terms of an arbitrary set of real functions \( \phi_j(x) \), orthonormal with weight \( |u_0(x)|^2 \),

\[ \phi(x) = \sum_j a_j \phi_j(x), \]

where \( a_j \) are a set of parameters describing the wavefront that we wish to estimate. Functions \( \Delta u(x) = u_0(x)\phi(x) \) define an orthonormal basis for the optical field perturbations. Without loss of generality, we shall assume that the optics are lossless in that all light entering the system reaches the detectors. Then the optics perform a unitary transformation on \( u(x) \), to give the field at the detectors \( u(y) \) where \( y \) represents the generalized coordinates over the detector surfaces.  

At the detectors, the normalized irradiance is \( I_y(y) = |u(y)|^2 \). The accuracy of the irradiance measurements are limited by noise. Detector noise arises from a variety of mechanisms depending on the detector, e.g., Johnson-Nyquist noise, dark current etc. These sources are technological in nature. For example, electron multiplying CCD devices and PMT arrays that can effectively remove the effects of read noise are becoming widely available. Photon noise on the other hand is fundamental in nature arising from the quantum nature of light. Since we are seeking a fundamental limit, it is the photon noise that shall be our concern here. The normalized irradiance can then be considered as a probability density function for the position \( y \) of a single photon (we will use the word photon to mean a photo-detection event). Photon data are generated by the measurement according to this probability density, from which we then try to estimate the wavefront parameters \( a_j \). We will use the Fisher information to quantify the measurement accuracy. For a vector of unknown parameters \( \mathbf{a} \) the covariance of any unbiased estimator \( \hat{\mathbf{a}} \) must satisfy the CR lower bound [18], thus

\[ \mathbf{C} \geq \mathbf{J}^{-1}, \]

where the estimator covariance matrix elements are \( C_{jk} = \langle (\hat{a}_j - a_j)(\hat{a}_k - a_k) \rangle \), \( \mathbf{J} \) is the Fisher information matrix, and where \( \mathbf{A} \geq \mathbf{B} \) means \( \mathbf{A} - \mathbf{B} \) is positive semi-definite. The Fisher information matrix quantifies the useful information obtained from the measurement process. It is defined as the covariance of the score vector \( \mathbf{V} \),

\[ J_{jk} = \langle V_j(y)V_k(y) \rangle = \iint V_j(y)V_k(y)I_y(y)d^2y, \]

where the score vector elements are given by

\[ V_j(y) = \frac{\partial \log I_y(y)}{\partial a_j} = \frac{\partial \log p_y(y)}{\partial a_j} = \frac{1}{|u(y)|^2} \frac{\partial |u(y)|^2}{\partial a_j}. \]

After some algebraic manipulation we obtain

\[ J_{jk} = \iint 4 \left| \frac{\partial u(y)}{\partial a_j} \right| \left| \frac{\partial u(y)}{\partial a_k} \right| \cos \phi_j(y) \cos \phi_k(y) d^2y, \]

\[ \text{We use } u \text{ to denote the field at both } x \text{ and } y \text{ in recognition that it is the same Hilbert space vector in both coordinate systems.} \]
where $\phi_j(y)$ is the complex angle between $\partial u^*(y)/\partial a_j$ and $u(y)$.

From consideration of the CR bound, we see that in order to minimize the estimator variances we should maximize the diagonal elements of the Fisher matrix and minimize (make zero) the off-diagonal elements. First note that the functions $\partial u(y)/\partial a_j$ are orthonormal. To see this, consider the equivalent terms in the aperture coordinates $x$. From eqs. 1 and 2, we have

$$\frac{\partial u(x)}{\partial a_j} = i\omega_0(x)\phi(x)\exp[i\phi(x)] = \Delta u(x)\exp[i\phi(x)],$$

(7)

which are just the original orthonormal functions $\Delta u(x)$ subject to the unitary transformation $\exp[i\phi(x)]$. From the unitary property of the optics, the same must apply in the detector coordinates $y$. The diagonal elements of $J$ then satisfy,

$$J_j = 4\int\int \left|\frac{\partial u(y)}{\partial a_j}\right|^2 \cos^2 \phi_j(y) d^2y \leq 4,$$

(8)

(9)

the maximum being achieved when $\phi_j(y) = 0$ or $\pi$ over $y$, under which condition the off-diagonal elements are zero. We can view this condition as a generalized maximum phase contrast condition: the information is maximized when the perturbed components of the optical field interfere constructively or destructively with the rest of the field at the detectors. This expression is for a single photon: for multiple photons (with Poisson statistics) we just multiply the Fisher information by the number of photons $n$. Applying the CR bound we find the variance for the estimate for each aberration parameter must satisfy

$$\sigma_j^2 \geq 1/(4n),$$

(10)

and for $N_A$ statistically independent aberration modes, the combined error is

$$\sigma^2 = \sum_{j=1}^{N_A} \sigma_j^2 \geq N_A/(4n).$$

(11)

To illustrate, we apply the limit to an adaptive optics system. For diffraction limited performance, using the Maréchal criterion of Strehl ratio greater than 0.8 (equivalent to $\lambda/14$ rms wavefront error), gives

$$\sigma^2 \leq 0.2 \quad \Rightarrow \quad n \gtrsim (5/4)N_A.$$  

(12)

So, for diffraction limited imaging we require approximately one photon per correction mode.

3. Zernike point-diffraction wavefront sensing

We have derived an information limit for the wavefront sensing problem. We now turn our attention to the practical form of the wavefront sensor required. The generalized phase contrast condition for maximum Fisher information suggests the use of a point diffraction interferometer (PDI) with a $\pi/2$ phase phase mask. The principle is illustrated in figure 1. A phase mask is placed in an image (Fourier) plane with a phase step “pinhole”, approximately the size of the Airy disk, that shifts the phase of the unperturbed optical field by $\pi/2$ with respect to the perturbed components. This requires that the perturbed and unperturbed field components are physically separated in the Fourier plane. The separation increases with the spatial frequency of the aberration so that for high spatial frequencies the wavefront sensor can approach the information limit. The separation is less complete for low spatial frequency components (due to the finite extent of the point-spread function.) Figure 2 plots the estimation accuracy achievable with the Zernike PDI for each of the first 35 Zernike aberrations. We can see in this case that the
Figure 1. Zernike point diffraction wavefront sensor for an aberration corresponding to a single spatial frequency $\phi = a \cos kx$. The left hand image shows the Fourier image of the aberrated wave (the position of the $\pi/2$ phase mask is indicated by the circle). The right image shows the resulting irradiance $I(y)$ at the detector plane with fringes corresponding to the aberration visible.

Figure 2. Estimation variance limits for different Zernike aberration coefficients for a Zernike PDI wavefront sensor with mask pinhole radius $= \lambda f/D$. (The dotted line indicates the fundamental Fisher information limit.)

Estimation accuracy is not as good for the lower-order aberrations (particularly Zernike tip and tilt), but approaches the information limit for higher order aberrations.

Although the CR lower bound gives a limit to the accuracy of an unbiased estimator, it does not guarantee that an estimator exists to achieve the limit. In general such estimators only exist for a limited class of probability distributions (exponential class). However, the maximum-likelihood estimator, if it exists, does achieve the CR bound asymptotically, i.e., for large numbers of data (photons in our case.) For $n$ independent photon events at positions $y = \{y_1, y_2, \ldots, y_n\}$ the combined likelihood is the product of the individual likelihoods, which are just the irradiances at the positions of the photon events.

$$p(y) = \prod_{k=1}^{n} I(y_k).$$

The maximum likelihood estimator is found by solving the maximization problem: $\partial \log p(y)/\partial a_j = 0$ for all $a_j$. The general solution is non-trivial, however, for small perturbations the variations in the irradiance are linear with respect to the aberration parameters, allowing us to write the irradiance in the
The following form

\[ I(y) = I_0(y) + \sum_j \Delta I_j(y) a_j + O[a^2]. \]  

Then, solving the maximum likelihood equation and taking the asymptotic limit for large \( n \) leads to a linear form for the asymptotic maximum-likelihood estimator

\[ \sum_j \left[ \int \int \frac{\Delta I_j(y) \Delta I_j'(y)}{I_0(y)} d^2 y \right] a_j = \frac{1}{n} \sum_{k=1}^{n} \frac{\Delta I_j(y_k)}{I_0(y_k)}. \]  

Figure 3 shows results of a Monte-Carlo type simulation using this estimator with the Zernike PDI.
For a small random aberration, comprising 38 Zernike aberration terms, sample photon-counting data at the detector plane of the wavefront sensor were generated for a range of different exposures. The detector was modelled as a 128x128 pixel electron multiplication CCD camera operating in a thresholded photon-counting mode (in which multiple photon events on a single pixel are seen as a single count.) The mean error in the wavefront estimate is within a factor or two of the fundamental wavefront sensor limit we have derived. Note that diffraction limited performance (σ^2 < 0.2) is achieved with less than 100 photons. The efficiency deteriorates for n ≳ 10^3, when the probability of multiple photons per pixel becomes significant. To avoid this, in a practical sensor we must ensure that the pixel readout rate be much higher than the photon detection rate.

4. Conclusion
We have derived a fundamental information limit for wavefront sensing. This suggests we need approximately one photon per aberration mode to measure a wavefront sufficiently for diffraction-limited correction. The canonical wavefront sensor we used is quite general, and so we can consider the limit a property of the nature of optical propagation and quantized photodetection. Nevertheless, the form of the limit suggests a practical efficient wavefront sensor based on the point diffraction interferometer, using a Zernike phase contrast mask.

The author gratefully acknowledges the support of the Royal Society URF scheme.

References
[1] Dyson F J 1975 J. Opt. Soc. Am. 65 551–558
[2] Irwan R and Lane R G 1999 Appl. Opt. 38 6737–6743
[3] Nicolle M, Fusco T, Rouset G and Michau V 2004 Opt. Lett. 29 2743–2745
[4] Saleh B E A 1974 Proceedings Of The Ieee 62 530–531
[5] Saleh B E A 1974 Appl. Opt. 13 1824–1827
[6] van Dam M A and Lane R G 2000 J. Opt. Soc. Am. A 17 1319–1324
[7] Wallner E P 1983 J. Opt. Soc. Am. 73 1771–1776
[8] Winick K A 1986 J. Opt. Soc. Am. A 3 1809–1815
[9] Barrett H H, Dainty C and Lara D 2007 J. Opt. Soc. Am. A 24 391–414
[10] Cannon R C 1995 J. Opt. Soc. Am. A 12 2031–2039
[11] Schulz T J, Sun W and Roggemann M C 1999 (Proc. SPIE vol 3763) ed Roggemann M C and Bissonnette L R pp 23 – 28 ISSN 0277-786X
[12] Cederquist J, Robinson S R, Kryskowski D, Fienup J R and Wackerman C C 1986 25 586–592
[13] Chew T Y, Clare R M and Lane R G 2006 Opt. Commun. 268 189–195
[14] Dean B H and Bowers C W 2003 J. Opt. Soc. Am. A 20 1490–1504
[15] Welsh B M, Ellerbroek B L, Roggemann M C and Pennington T L 1995 Appl. Opt. 34 4186–4195
[16] Roddier F 1995 Optics Communications 113 357–359
[17] Paterson C 2008 Adaptive Optics for Industry and Medicine: Proceedings of the 6th International Workshop ed Dainty J C (Imperial College Press) pp 157–160
[18] Kay S M 1993 Fundamentals of statistical signal processing: estimation theory vol 1 (Prentice Hall, New Jersey)