Revisiting the comparison between the Shack-Hartmann and the pyramid wavefront sensors via the Fisher information matrix

C. Plantet, S. Meimon, J.M. Conan, T. Fusco

To cite this version:

C. Plantet, S. Meimon, J.M. Conan, T. Fusco. Revisiting the comparison between the Shack-Hartmann and the pyramid wavefront sensors via the Fisher information matrix. Optics Express, Optical Society of America, 2015, 23 (22), pp.28619-28633. 10.1364/OE.23.028619. hal-01390865

HAL Id: hal-01390865
https://hal.archives-ouvertes.fr/hal-01390865
Submitted on 2 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Revisiting the comparison between the Shack-Hartmann and the pyramid wavefront sensors via the Fisher information matrix

C. Plantet,1,* S. Meimon,1 J.-M. Conan,1 T. Fusco1,2

1 Onera, the French Aerospace Lab, 92322 Chatillon, France
2 Aix Marseille Université, CNRS, LAM (Laboratoire d’Astrophysique de Marseille) UMR 7326, 13388 Marseille, France
*cedric.plantet@onera.fr

Abstract: Exoplanet direct imaging with large ground based telescopes requires eXtreme Adaptive Optics that couples high-order adaptive optics and coronagraphy. A key element of such systems is the high-order wavefront sensor. We study here several high-order wavefront sensing approaches, and more precisely compare their sensitivity to noise. Three techniques are considered: the classical Shack-Hartmann sensor, the pyramid sensor and the recently proposed LIFTed Shack-Hartmann sensor. They are compared in a unified framework based on precise diffractive models and on the Fisher information matrix, which conveys the information present in the data whatever the estimation method. The diagonal elements of the inverse of the Fisher information matrix, which we use as a figure of merit, are similar to noise propagation coefficients. With these diagonal elements, so called “Fisher coefficients”, we show that the LIFTed Shack-Hartmann and pyramid sensors outperform the classical Shack-Hartmann sensor. In photon noise regime, the LIFTed Shack-Hartmann and modulated pyramid sensors obtain a similar overall noise propagation. The LIFTed Shack-Hartmann sensor however provides attractive noise properties on high orders.

© 2015 Optical Society of America

OCIS codes: (010.1080) Active or adaptive optics; (010.7350) Wave-front sensing.

References and links
1. T. Fusco, J.-F. Sauvage, C. Petit, A. Costille, K. Dohlen, D. Mouillet, J.-L. Beuzit, M. Kasper, M. Suarez, C. Soenke, E. Fedrigo, M. Downing, P. Baudoz, A. Sevin, D. Perret, P. Barrufol, B. Salasnich, P. Puget, F. Feautrier, S. Rochat, T. Moulin, A. Deboulbè, E. Hugot, A. Vigan, D. Mawet, J. Girard, and N. Hubin, “Final performance and lesson-learned of SAXO, the VLT-SPHERE extreme AO: from early design to on-sky results,” Proc. SPIE 9148, 91481U (2014).
2. B. A. Macintosh, A. Anthony, J. Atwood, N. Barriga, B. Bauman, K. Caputa, J. Chilcote, D. Dillon, R. Doyon, J. Dunn, D. T. Gavel, R. Galvez, S. J. Goodsell, J. R. Graham, M. Hartung, J. Isaacs, D. Kerley, Q. Konopacky, K. Labrie, J. E. Larkin, J. Maire, C. Marois, M. Millar-Blanchaer, A. Nunez, B. R. Oppenheimer, D. W. Palmer, J. Pazder, L. A. Poyneer, C. Quirez, F. Rantakyrö, V. Reshtov, L. Saddlemeyer, N. Sadakuni, D. Savransky, A. Sivaramakrishnan, M. Smith, R. Soummer, S. Thomas, J. K. Wallace, J. Weiss, and S. Wiktorowicz, “The Gemini Planet Imager: integration and status,” Proc. SPIE 8446, 84461U (2012).
3. R. M. Wagner, M. L. Edwards, O. Kuhn, D. Thompson, and C. Veillet, “An overview and the current status of instrumentation at the Large Binocular Telescope Observatory,” Proc. SPIE 9147, 914705 (2014).
4. T. Fusco, G. Rouset, J.-F. Sauvage, C. Petit, J.-L. Beuzit, K. Dohlen, D. Mouillet, J. Charton, M. Nicolle, M. Kasper, P. Baudoz, and P. Puget, “High-order adaptive optics requirements for direct detection of extrasolar planets: application to the SPHERE instrument,” Opt. Express 14, 7515–7534 (2006).
5. F. Rigaut and E. Gendron, “Laser guide star in adaptive optics - the tilt determination problem,” Astronomy Astrophys. 261, 677–684 (1992).

6. G. Rousset, “Wave-front Sensors,” in Adaptive Optics in Astronomy, F. Roddier, ed. (Cambridge University Press, 1999), pp. 91–130.

7. O. Guyon, “Limits of adaptive optics for high-contrast imaging,” Astrophys. J. 629, 592 (2005).

8. C. Véritaud, “On the nature of the measurements provided by a pyramid wave-front sensor,” Opt. Commun. 233, 27–38 (2004).

9. R. Ragazzoni and J. Farinato, “Sensitivity of a pyramidal wave front sensor in closed loop adaptive optics,” Astronomy Astrophys. 350, L23–L26 (1999).

10. B. L. Ellerbroek, B. J. Thelen, D. J. Lee, D. A. Carrara, and R. G. Paxman, “Comparison of Shack–Hartmann wavefront sensing and phase-diverse phase retrieval,” in Optical Science, Engineering and Instrumentation ’97, (ISOP, 1997), pp. 307–320.

11. B. M. Welsh, B. L. Ellerbroek, M. C. Roggemann, and T. L. Pennington, “Fundamental performance comparison of a Hartmann and a shearing interferometer wave-front sensor,” Appl. Opt. 34, 4186–4195 (1995).

12. L. Meynadier, V. Michau, M.-T. Velluet, J.-M. Conan, L. M. Mugnier, and G. Rousset, “Noise propagation in wave-front sensing with phase diversity,” Appl. Opt. 38, 4967–4979 (1999).

13. E. Thiebaut, “Introduction to image reconstruction and inverse problems,” in Optics in Astrophysics, (Springer, 2005), pp. 397–422.

14. J. R. Fienup, B. J. Thelen, R. G. Paxman, and D. A. Carrara, “Comparison of phase diversity and curvature wavefront sensing,” in Astronomical Telescopes and Instrumentation, (ISOP, 1998), pp. 930–940.

15. D. J. Lee, B. M. Welsh, and M. C. Roggemann, “Cramér-Rao analysis of phase diversity imaging,” in Optical Science, Engineering and Instrumentation ’97, (ISOP, 1997), pp. 161–172.

16. T. J. Schulz, W. Sun, and M. C. Roggemann, “Cramér-Rao bounds for estimation of turbulence-induced wavefront aberrations,” in SPIE’s International Symposium on Optical Science, Engineering, and Instrumentation, (International Society for Optics and Photonics, 1999), pp. 23–28.

17. C. Paterson, “Towards practical wavefront sensing at the fundamental information limit,” J. Phys.: Conf. Series 139, 012021.

18. S. Barwick, “Performance comparison between Shack–Hartmann and astigmatic hybrid wavefront sensors,” Appl. Opt. 48, 6967–6972 (2009).

19. R. Ragazzoni, “Pupil plane wavefront sensing with an oscillating prism,” J. Mod. Opt. 43, 289–293 (1996).

20. S. Meimon, T. Fusco, V. Michau, and C. Plantet, “Sensing more modes with fewer sub-apertures: the LIFTed Shack–Hartmann wavefront sensor,” Opt. Lett. 39, 2835–2837 (2014).

21. S. Meimon, T. Fusco, and L. M. Mugnier, “LIFT: a focal-plane wavefront sensor for real-time low-order sensing on faint sources,” Opt. Lett. 35, 3036–3038 (2010).

22. L. M. Mugnier, T. Fusco, and J.-M. Conan, “MISTRAL: a myopic edge-preserving image restoration method, with application to astronomical adaptive-optics-corrected long-exposure images,” J. Opt. Soc. Am. A 21, 1841–1854 (2004).

23. R. Lane and M. Tallon, “Wave-front reconstruction using a Shack–Hartmann sensor,” Appl. Opt. 31, 6902–6908 (1992).

24. R. C. Cannon, “Optimal bases for wave-front simulation and reconstruction on annular apertures,” J. Opt. Soc. Am. A 13, 862–867 (1996).

25. S. Thomas, T. Fusco, A. Tokovinin, M. Nicolle, V. Michau, and G. Rousset, “Comparison of centroid computation algorithms in a Shack–Hartmann wavefront sensor,” Monthly Notices Royal Astron. Soc. 371, 323–336 (2006).

26. M. Nicolle, T. Fusco, G. Rousset, and V. Michau, “Improvement of Shack–Hartmann wave-front sensor measurement for extreme adaptive optics,” Opt. Lett. 29, 2743–2745 (2004).

27. S. Esposito, A. Riccardi, E. Pinna, A. Puglisi, F. Quiro–Pacheco, C. Arcidiacono, M. Xompero, R. Briguglio, G. Agapito, L. Busoni, L. Fini, J. Argomedo, A. Gherardi, G. Brusa, D. Miller, J. C. Guerra, P. Stefanini and P. Salinari, “Large Binocular Telescope Adaptive Optics System: new achievements and perspectives in adaptive optics,” Proc. SPIE 8149, 814902 (2011).

28. C. Plantet, S. Meimon, J.-M. Conan, and T. Fusco, “Experimental validation of LIFT for estimation of low-order modes in low-flux wavefront sensing,” Opt. Express 21, 16337–16352 (2013).

29. V. N. Mahajan and G.-M. Dai, “Orthonormal polynomials in wavefront analysis: analytical solution,” J. Opt. Soc. Am. A 24, 2994–3016 (2007).

30. A. Papoulis, Probability and Statistics, (Prentice-Hall Englewood Cliffs, 1990)

31. H. H. Barrett, J. Denny, R. F. Wagner, and K. J. Myers, “Objective assessment of image quality. II. Fisher information, Fourier crosstalk, and figures of merit for task performance,” J. Opt. Soc. Am. A 12, 834–852 (1995).

32. H. H. Barrett, C. Dainty, and D. Lara, “Maximum-likelihood methods in wavefront sensing: stochastic models and likelihood functions,” J. Opt. Soc. Am. A 24, 391–414 (2007).

33. A. Papoulis and B. Macintosh, “Spatially filtered wave-front sensor for high-order adaptive optics,” J. Opt. Soc. Am. A 21, 810–819 (2004).

34. J.-F. Sauvage, T. Fusco, C. Petit, D. Mouillet, K. Dohlen, A. Costille, J.-L. Beuzit, A. Baruffolo, M. E. Kasper,
1. Introduction

Exoplanet direct imaging is made difficult by the huge intensity contrast between the star and its companion. The contrast can be reduced by a coronagraph, which diffracts the star light (but not the companion’s) away from the nominal ray path. However, coronagraphs quickly lose efficiency in the presence of optical aberrations. High contrast imaging on large ground based telescopes therefore implies adaptive optics (AO) to correct for atmospheric turbulence and aberrations due to the optical system itself. The tight requirements on the amplitude of the residual wavefront lead to high-order AO systems, typically from $30 \times 30$ actuators to $44 \times 44$ actuators on current systems [1–3]. The association of high-order AO with coronagraphy, and more generally high contrast instruments, is generally called eXtreme AO (XAO) [4]. A key element of such systems is the high-order wavefront sensor that has to accurately measure aberrations at a high spatial resolution. This paper focuses on this key element, and more precisely on its sensitivity to noise.

Sensitivity to noise, or noise propagation [5], can be quantified by the covariance matrix of the wavefront estimation error. This metric is commonly used to evaluate the performance of a wavefront sensor (e.g. [5–12]). The error covariance matrix depends on the amount of both information and noise in the data. It also depends on the estimator used, or more precisely, on the way the estimator propagates noise affecting the data into the wavefront estimate. One classical approach is to assume that a maximum likelihood (ML) estimation is performed with a Gaussian noise model on data (e.g. [5,6,8,9]). However, maximum likelihood has proved to be unadapted in some cases (e. g. where not enough modes are sensed with respect to the modal content of the wavefront to be sensed), leading to strong errors in the wavefront estimation, in the same manner as in image reconstruction [13]. In these cases, a prior knowledge of the wavefront statistics can help wavefront sensing reconstruction. Reasonably enough, one may wonder if a comparison method based on the covariance matrix of the wavefront estimation error is still fair apart from the sole maximum likelihood case.

In order to get a fundamental limit of a wavefront sensor’s performance, other studies rely on the Cramér-Rao lower bound [14–18]. The Cramér-Rao lower bound defines a lower limit of the error covariance matrix. This bound depends on the wavefront estimator’s bias – linked to the estimation method and the prior knowledge on the unknown wavefront – and on the inverse of the Fisher information matrix – which conveys the information ultimately extractable from the data (i.e. the ability to estimate the wavefront from the data set), whatever the estimation method.

In this paper, we compare wavefront sensors based on the inverse of the Fisher information matrix. This metric corresponds to the fundamental limit of wavefront sensors sensitivity to noise when using unbiased estimators, but also determines their relative performance when using biased estimators. The proposed method thus allows a fair comparison.
We consider three wavefront sensing approaches: the classical Shack-Hartmann sensor, widely used in AO, and recently implemented in two operational XAO systems SAXO [1] and GPI [2]; the pyramid wavefront sensor, introduced in 1996 by Ragazzoni [19], very promising for high-order AO, and successfully integrated in FLAO [3], the LBT high-order AO system; the LIFTed Shack-Hartmann sensor [20], a recent attractive evolution of the Shack-Hartmann dedicated to high order sensing and that makes use of the LIFT concept [21].

It has been shown that the pyramid sensor has a lower noise propagation than the Shack-Hartmann sensor on low orders, and reaches the same performance at the sensor’s spatial cut-off frequency [7–9]. This so-called full aperture gain is however partially lost when using a modulated pyramid sensor [8], a technique often used in practice to increase the dynamic range of the sensor. Note that these quantitative analyses were performed only in photon noise, and rely on simplifying approximations on the optical and noise model. As regards the LIFTed Shack-Hartmann sensor, a preliminary comparison with the classical Shack-Hartmann sensor is given in [20], but only in terms of interaction matrix eigenvalues.

We present here a detailed comparison, in the context of high-order wavefront sensing, of these three wavefront sensors in a unified framework: modeling each sensor with a precise diffractive model, and then comparing them with the Fisher information matrix, accounting for both photon and read-out noise. This study therefore focuses on noise propagation, but we also briefly discuss the impact of other error sources, such as aliasing, on the wavefront sensing performance.

We first present, in section 2, the diffractive models used for the considered wavefront sensors. We then describe, in section 3, the comparison method based on the Fisher information matrix. Using this method we evaluate on one hand the noise propagation for the classical and LIFTed Shack-Hartmann sensors, and we study on the other hand the pyramid sensor with and without modulation (see section 4). Section 5 finally focuses on the comparison between the three sensors.

2. High-order wavefront sensors

2.1. General model

A wavefront sensor uses optical elements to turn the wavefront deformations into an interpretable intensity distribution on a detector. It thus consists in a hardware part (optics and detector) and a signal processing part. This signal processing step turns the pixels values into the output data of the wavefront sensor, e. g. local slopes in the classical Shack-Hartmann sensor (see section 2.2). We consider here that it does not include the wavefront estimation step.

The pixels values are affected by photon noise and by the detector’s read-out noise. In this study, we assume that the noise on pixels is a zero-mean additive Gaussian noise. We approximate the noise variance on each pixel $p$ by the addition of the mean flux on the pixel $\bar{I}_p$ (photon noise) and the variance of read-out noise $\sigma_e^2$ [22]:

$$\sigma_p^2 = \bar{I}_p + \sigma_e^2$$ (1)

The validity of assumptions on noise is discussed in paragraph 3.2.

The wavefront is usually reconstructed on a polynomial basis. We consider here a reconstruction on the Karhunen-Loève polynomials, which constitute an efficient basis for sensing turbulent aberrations with a high spatial resolution [23], since they are statistically uncorrelated and they maximize the energy in low order modes. There are no general analytical expressions of the Karhunen-Loève polynomials. In this paper, we calculate them with a computationally efficient method proposed by Cannon [24].

The signal, processed from pixels values, then depends on the vector of unknowns $A = [a_2, a_3, \ldots, a_n]^T$, with $a_i$ the coefficient of the $i$-th polynomial. We assume that the relation be-
The aberrations coefficients and the noiseless data is linear around the operating point.

We can thus write the data formation model for any wavefront sensor:

\[ y = DA + n \]  

(2)

with \( y \) the vector of data, \( D \) the interaction matrix and \( n \) the noise. The matrix \( D \) consists of the wavefront sensor response to the Karhunen-Loève polynomials. Each of its columns is the vector of data \( y^i \) corresponding to the \( i \)-th polynomial.

In the following, we describe the models used to compute \( y \) and the noise models for the considered wavefront sensors.

2.2. Shack-Hartmann sensor

**Modeling method** The simulations are made at Shannon sampling, with 16×16 pixels per subaperture. The interferences produced by the lenslet array are neglected. The Shack-Hartmann sensor slopes are computed with either a Center of Gravity (CoG) or an unbiased Weighted Center of Gravity (WCoG), as defined in [25]. CoG is more efficient in photon noise regime, while WCoG is more efficient in read-out noise regime [25].

The vector of data is, for \( N_{sub} \) subapertures:

\[ y^i = [x_1, y_1, x_2, y_2, \ldots, x_{N_{sub}}, y_{N_{sub}}] \]  

(3)

where \( x_i \) and \( y_i \) are the coordinates of the spot centroid in the \( i \)-th subaperture.

**Noise model** The noise variance on slopes are computed with Nicolle’s and Thomas’s theoretical formulas for the center of gravity and the weighted center of gravity (equations 1 and 2 in [26], equations 23 and 24 in [25]), with the following parameters: \( N_t = N_d = N_w = 2 \) pixels and \( N_s = 4 \) pixels.

2.3. Pyramid sensor

**Modeling method** The pyramid phase mask is applied to the complex amplitude in the focal plane, as described by Vérinaud in [8], creating a new complex amplitude from which the intensity in the detector plane, conjugated to the pupil plane, is deduced. With this accurate diffractive model, the pupil images, called hereafter “image”, include the interferences between the four beams leaving the pyramid. In our case, the centers of the pupil images are separated by 2 pupil diameters. The considered radii of modulation are \( 2 \frac{\lambda}{D}, 3 \frac{\lambda}{D} \) and \( 6 \frac{\lambda}{D}, \lambda \) being the sensing wavelength and \( D \) the pupil diameter. They correspond to modulations performed on the LBT [27].

The data of the pyramid are computed as follows for four pixels, denoted \( k \), all corresponding to the same location in each pupil image:

\[
S_x[k] = \frac{(P_1[k] + P_3[k]) - (P_2[k] + P_4[k])}{N} \\
S_y[k] = \frac{(P_1[k] + P_2[k]) - (P_3[k] + P_4[k])}{N} \\
\]  

with \( N = \frac{1}{N_{pix}} \sum_{k=1}^{N_{pix}} P_1[k] + P_2[k] + P_3[k] + P_4[k] \)  

(4)

with \( P_i \) the pupil image from face \( i \) and \( N_{pix} \) the number of pixels in each pupil image. \( S_x \) is the signal linked to local slopes in the \( x \) direction and \( S_y \) is the signal linked to local slopes in the \( y \) direction. \( N \) is the detected flux per pixel averaged over the 4 pupil images.
The vector of data is then:

\[
y' = [S_x[1], S_y[1], S_x[2], S_y[2], \ldots, S_x[N_{pix}], S_y[N_{pix}]]
\]  

(5)

Note that we consider only the pixels inside the geometrical pupil footprints. The flux diffracted outside the pupil footprints, accurately modeled with our diffractive simulations, is thus lost for our data. At diffraction limit, the flux loss is \(\sim 57\%\) of the incoming flux with no modulation. In effect, when the pyramid is not modulated, the focal spot constantly undergoes the diffraction by four edges. For modulations greater or equal to \(\sim \lambda/D\), the spot spends a little time on each edge, and the flux loss becomes negligible.

Noise model  We consider there is enough flux to neglect the noise on \(N\). The noise variance on \(S_x\) and \(S_y\) is thus equal to \(1/N^2\) times the noise variance of their numerators.

In photon noise, the variance on the pixel \(P_i[k]\) is equal to its mean flux (given by the diffractive model). The numerator noise variance is thus the sum of the mean fluxes of \(P_1[k], P_2[k], P_3[k]\) and \(P_4[k]\), which is equal to the mean flux of \(P_1[k] + P_2[k] + P_3[k] + P_4[k]\).

In read-out noise, the variance on the pixel \(P_i[k]\) is equal to \(\sigma_e^2\), with \(\sigma_e^2\) the read-out noise variance. The numerator noise variance is thus \(4\sigma_e^2\) in read-out noise.

Hence, the noise variance on \(S_x[k]\) and \(S_y[k]\) is \((P_1[k] + P_2[k] + P_3[k] + P_4[k])/N^2\) in photon noise and \(4\sigma_e^2/N^2\) in read-out noise.

2.4. LIFTed Shack-Hartmann sensor

The LIFTed Shack-Hartmann sensor consists in using the focal plane wavefront sensor called LIFT on the subapertures of a Shack-Hartmann sensor \([20]\). LIFT performs a maximum likelihood estimation of the phase on a single image, with a small-phase approximation \([21, 28]\). To remove the even modes indetermination, an astigmatism offset is added to the incoming phase. It is therefore possible to implement LIFT in a Shack-Hartmann sensor by using astigmatic lenslets. Since more modes than the two centroids can be estimated per subaperture, it is also possible to have less, hence larger, subapertures.

Modeling method  As for the Shack-Hartmann sensor, the simulations are made at Shannon sampling, with \(16 \times 16\) pixels per subaperture, and the interferences produced by the lenslet array are neglected. The added astigmatism, of amplitude 0.5 rad rms, is taken from a Zernike basis orthonormalized on a square subaperture, computed from equations in \([29]\). The local modes estimated by LIFT are also taken from this basis. The estimation by LIFT returns a vector of local modes coefficients for each subaperture \(i: [a_{i,1}, a_{i,2}, \ldots, a_{i,m}]\), with \(m\) the number of estimated local modes. The vector of data is then the concatenation of all subapertures local modes coefficients:

\[
y' = [a_{1,1}, a_{2,1}, \ldots, a_{m,1}, a_{1,2}, a_{2,2}, \ldots, a_{m,2}, \ldots, a_{1,N_{sub}}, a_{2,N_{sub}}, \ldots, a_{m,N_{sub}}]
\]  

(6)

with \(a_{i,j}\) the \(i\)-th local mode coefficient for the \(j\)-th subaperture and \(N_{sub}\) the number of subapertures.

Noise model  Each element of the data is a local mode coefficient estimated by LIFT. The computation of the noise propagation in a subaperture follows the equations used in \([28]\). One might think that the number of pixels taken into account (\(16 \times 16\)) in each subaperture will affect the sensitivity to read-out noise, but LIFT, similarly to a WCoG, uses weighting functions on the image to make its estimation (examples can be found in \([21]\)). In read-out noise regime, the pixels far from the spot center are weighted at zero, which strongly limits the impact of read-out noise on the estimation.
3. Comparing wavefront sensors via the Fisher information matrix

3.1. Cramér-Rao bound and Fisher Information Matrix

The Cramér-Rao inequality expresses a lower bound on the variance of estimators of a deterministic parameter. Let us consider for now the estimation of a scalar parameter \( a \) from a vector \( y \) of measurement data with an estimator \( \hat{a} \). Then the variance of the estimation error verifies:

\[
\text{var}\{\hat{a} - a\} \geq \frac{[1 + \partial \text{bias} / \partial a]^2}{\text{Fisher}(a)}
\]

where the bias term is defined as \( \text{bias}(a) = E\{\hat{a}\} - a \) and the Fisher Information term \( \text{Fisher}(a) \) corresponds to the amount of information contained in the data \( y \).

In the non-scalar case, the variance \( \text{var}\{\hat{a} - a\} \) is a covariance matrix, the inequality is a matrix inequality (for two matrices \( A \) and \( B \), \( A \geq B \) means that \( A-B \) is positive semidefinite), and the Fisher Information is also a matrix defined by

\[
F_{i,j}(A_p) = E\{[\partial / \partial a_i \ln p(y|A)|_{A=A_p}][\partial / \partial a_j \ln p(y|A)|_{A=A_p}]\}
\]

where \( A = [a_1, a_2, \ldots, a_n]^t \) is the vector of unknowns to estimate, \( A_p \) is the operating point, and \( p(y|A) \) is the likelihood function [30].

In Eq. (7), the numerator term depends on the estimator used and the prior on the data, whereas the denominator depends only on the sensitivity of the data to the parameter to estimate (i.e. data variations with respect to the parameter). Logically, the higher the information, the lower the inferior bound on the estimation error variance, \textit{whatever the estimation method used}. The classical result derived from the Cramér-Rao inequality is that in the case of an unbiased estimator, the lower bound reduces to \( 1 / \text{Fisher}(a) \), and the maximum likelihood estimator (asymptotically) reaches this bound [31]. This means that, in absence of prior knowledge, the maximum likelihood makes an optimal use of the information contained in the data and quantified by the Fisher Information. In this particular case only, the inverse of the Fisher information matrix coincides with the noise covariance matrix (see appendix B). But the Cramér-Rao inequality goes beyond this particular case: it shows that whatever the class of estimators considered, it is always beneficiary to maximize the Fisher Information. To summarize, examining the inverse Fisher Information is equivalent to the classical noise propagation coefficients approach in the maximum likelihood case, but provides a wider framework, as it is still relevant when using another estimation method than the maximum likelihood.

The inverse Fisher Information is therefore a powerful analytical tool to quantify the amount of information in the data \textit{whatever the estimator}. In the following, we use the inverse Fisher Information matrix as the metric to fairly compare wavefront sensors.

3.2. Comparison method

In the assumption of an additive Gaussian noise, the expression of the Fisher information matrix becomes (see appendix A):

\[
F_{i,j}(A_p) = \sum_{k=1}^{N} \frac{1}{\sigma_k^2} \frac{\partial \tilde{y}_k(A)}{\partial a_i} |_{A=A_p} \frac{\partial \tilde{y}_k(A)}{\partial a_j} |_{A=A_p}
\]

with \( \sigma_k^2 \) the noise variance on each data element \( y_k \), \( \tilde{y}_k \) a noiseless data element and \( N \) the total number of data elements. One can recognize here a “signal-to-noise ratio”, as the sensor’s sensitivity to each mode, represented by the derivatives, is weighted by the noise variance. This
expression remains valid in presence of only Poisson noise [32]. The more complicated case of mixed Gaussian and Poisson noise at low flux and low read-out noise is not treated here, as it requires knowledge on the detector’s response [32].

Also, the trace of the inverse of the Fisher information matrix can be expressed (see appendix B):

\[
\text{trace}(F^{-1}) = \frac{\alpha}{n_{\text{ph}}} + \left( \frac{\beta}{n_{\text{ph}}} \right)^2 \alpha \Gamma \left( \sum_{i} \alpha_i \right) + \left( \sum_{i} \beta_i \right) \left( \frac{\sigma_e}{n_{\text{ph}}} \right)^2
\]

(10)

The coefficients \(\alpha_i\) and \(\beta_i\) can be numerically obtained from extreme cases: when computing the inverse Fisher information matrix with \(n_{\text{ph}} = 1\) photo-electron and \(\sigma_e = 0\) electron, we get \(\alpha_i = F^{-1}[i, i]\), the Fisher coefficients for photon noise. Similarly, with \(n_{\text{ph}} = 1\) photo-electron and \(\sigma_e = 1\) electron, without photon noise, we have \(\beta_i = F^{-1}[i, i]\), the Fisher coefficients for read-out noise. The lower these coefficients are, the better performance the wavefront sensor will have. Note that for a maximum likelihood estimation, Eq. (10) can be used to obtain the variance of estimation error due to noise. The Fisher coefficients given in the rest of the paper are such that Eq. (10) provides an estimation error in squared radians.

In the following, we compute the Fisher coefficients of the LIFTed Shack-Hartmann, the classical Shack-Hartmann and the pyramid sensors in the context of high-order wavefront sensing. In section 4, we discuss the consistency of our noise propagation evaluations with former studies. We also quantify the gain brought by the LIFTed Shack-Hartmann sensor on the classical one. In section 5, we compare both Shack-Hartmann sensors to the pyramid sensor, and analyze their respective assets for XAO applications.

4. Fisher coefficients of the considered sensors

Current 8m-class telescope XAO systems use a fine pupil sampling to estimate the incoming wavefront, e.g. \(30 \times 30\) subapertures for FLAO, \(40 \times 40\) subapertures for SAXO and \(44 \times 44\) subapertures for GPI. In order to be representative of current systems, we compute here the Fisher coefficients for a pupil sampling of \(40 \times 40\) subapertures. To do this, we use the diffractive models described in sections 2.2 to 2.4. We consider the estimation of 1000 Karhunen-Loève polynomials at diffraction limit in monochromatic light.

4.1. Classical and LIFTed Shack-Hartmann sensors

The modal Fisher coefficients \(\alpha_i\) and \(\beta_i\) for the classical Shack-Hartmann and the LIFTed Shack-Hartmann sensors are plotted in Fig. 1. In a first study, a LIFTed Shack-Hartmann sensor with \(10 \times 10\) subapertures was compared to a classical Shack-Hartmann sensor with \(20 \times 20\) subapertures [20]. To keep the same ratio, the LIFTed Shack-Hartmann sensor has here a pupil sampling of \(20 \times 20\). In order to have as many data in the LIFTed Shack-Hartmann sensor \(20 \times 20\) as in the classical Shack-Hartmann sensor \(40 \times 40\), we estimate 8 local modes per subaperture. In effect, there are 4 times less valid subapertures in the LIFTed Shack-Hartmann sensor \(20 \times 20\) than in the classical Shack-Hartmann sensor \(40 \times 40\), so we need to compute 2 slopes \(\times 4 = 8\) coefficients per subaperture to reach the same number of data in total.

The Shack-Hartmann sensor has a noise propagation in \(j^{-1}\) with \(j\) the polynomial number. Since \(j \sim (n + 1)^2\), \(n\) being the radial order, this is consistent with the propagation found by Rigaut and Gendron [5]. As expected, the best estimator for local slopes is the center of gravity in photon noise and the weighted center of gravity in read-out noise.

The LIFTed Shack-Hartmann sensor follows approximately the same trend as the classical Shack-Hartmann sensor, but has a lower noise propagation. The gain over the Shack-Hartmann is 2 in photon noise and approximately 1.6 in read-out noise. This gain is brought by the increase...
4. Photon noise

Read-out noise

Fig. 1. Fisher coefficients of the Shack-Hartmann sensor $40 \times 40$ and the LIFTed Shack-Hartmann sensor $20 \times 20$ for the estimation of 1000 Karhunen-Loève modes. The LIFTed Shack-Hartmann sensor estimates 8 modes per subaperture. The dotted line indicates the $j^{-1}$ trend.

of the subaperture diameter, leading to a “large aperture gain” [20] (the diffraction spot is narrower, and the flux is distributed over less pixels).

Note that the amplitude of the added astigmatism could be further optimized for the estimation of 8 local modes, following the strategy used for LIFT tip-tilt-focus sensing in [28], and one could also work on the choice of the local modes basis to gain even more performance with the LIFTed Shack-Hartmann sensor.

4.2. Pyramid sensor

The modal Fisher coefficients $\alpha_i$ and $\beta_i$ for the pyramid wavefront sensor are plotted in Fig. 2.

(a) Photon noise

(b) Read-out noise

Fig. 2. Fisher coefficients of the pyramid sensor for the estimation of 1000 Karhunen-Loève modes. The curves correspond to modulations from $0 \lambda/D$ to $6 \lambda/D$. A modulation at $0.5 \lambda/D$ has been added to show the transition between $0 \lambda/D$ and $2 \lambda/D$. The dotted line indicates the $j^{-1}$ trend.

In photon noise, the non-modulated pyramid sensor has a flat propagation which rises slightly for high frequencies. This slight increase was attributed to the filtering effect of the subaperture size by Vérinaud [8]. Also, the modulation makes the pyramid sensor act as a slope sensor in low orders and its propagation follows the same trend as the propagation of the Shack-Hartmann sensor. In this slope sensor regime, the pyramid sensor’s noise propagation increases propor-
tionally to the square of the modulation radius. In high orders, the non-modulated pyramid sensor has a much lower propagation than the modulated pyramid sensor (factor $\sim 0.36$ with modulations greater or equal to $2 \lambda / D$). This factor is due to a higher sensitivity to wavefront variations (so-called full aperture gain, already discussed in the context of low-order wavefront sensing in [28]), partly counterbalanced by the flux loss related to the diffraction of the pyramid’s edges. We obtain the same ratio by reproducing Vérinaud’s simulations as in [8] (a detailed demonstration can be obtained by contacting C. Plantet). The part of lost flux, as well as the sensitivity to wavefront variations, decreases when increasing the modulation radius. Hence, the change in noise propagation is progressive when varying the modulation between $0 \lambda / D$ and $2 \lambda / D$.

In read-out noise, the pyramid sensor’s noise propagation follows the same trends as in photon noise. The only difference is the factor between the non-modulated and the modulated pyramid sensor in high orders, which is equal to $\sim 0.92$ (not noticeable in Fig. 2, also verified with Vérinaud’s simulations). The flux loss is effectively more penalizing in read-out noise regime than in photon noise regime: indeed, one can see in Eq. (10) that the photon noise term is inversely proportional to the flux, while the read-out noise term is inversely proportional to the squared flux.

5. Comparison of the LIFTed Shack-Hartmann, the classical Shack-Hartmann and the pyramid sensors

We now compare the LIFTed Shack-Hartmann and the classical Shack-Hartmann sensors to the pyramid sensor. We plot in Figs. 3(a) and 3(b) the Fisher coefficients of these sensors. Figures 3(c) and 3(d) show the cumulated coefficients over the estimated modes.

In photon noise, the LIFTed Shack-Hartmann sensor is approximately as efficient as a pyramid sensor with a $6 \lambda / D$ modulation and has a performance close to lower modulations (factor 1.17 with modulation $3 \lambda / D$, 1.26 with modulation $2 \lambda / D$) for the estimation of 1000 modes (Fig. 3(c), abscissa 1000). However, its noise propagation in read-out noise is approximately 5 times as high as the pyramid sensor’s for a $6 \lambda / D$ modulation (Fig. 3(d), abscissa 1000). As regards the classical Shack-Hartmann sensor, it is significantly outperformed by the pyramid sensor, even at the highest considered modulation (factor $\sim 2$ in photon noise and $\sim 10$ in read-out noise).

Also, we can see that, in photon noise, the LIFTed and the classical Shack-Hartmann sensors Fisher coefficients are higher than the pyramid sensor’s in low orders (this point is discussed at the end of this section), but they become lower in high orders. Figure 4 shows the Fisher coefficients for modes 100 to 1000 with a linear abscissa. For modes over 600, the LIFTed Shack-Hartmann sensor has a better performance than the non-modulated pyramid sensor, with a factor going up to $\sim 2$ for the 1000th mode. For modes over 250, it is more efficient than the modulated pyramid sensor, with a factor going up to $\sim 5$ for the 1000th mode (up to $\sim 2$ from mode 450 for the classical Shack-Hartmann sensor). This can be useful for XAO systems, as they need a very precise wavefront correction in order to get rid of residual speckles, which mix with the signal of exoplanets or dust discs. High orders are responsible for speckles far from the image center. On SPHERE, the modes 250 to 1000 would approximately correspond to the second half of the correction zone (i.e. at a distance greater than $10 \lambda / D$ from the spot center for an image corrected up to $20 \lambda / D$). The LIFTed Shack-Hartmann sensor could thus be an attractive alternative to the pyramid and the classical Shack-Hartmann sensors in XAO.

In conclusion, the LIFTed Shack-Hartmann sensor is an important improvement of the classical Shack-Hartmann sensor. We showed that its performance is close to the pyramid sensor’s in photon noise limited applications, with an even better precision than the pyramid sensor in high orders.
These conclusions only concern noise propagation, that is the subject of the present article. An overall performance comparison would of course require to consider other error terms, such as aliasing and temporal error, and possibly also account for the coupling with a coronagraph in the case of high contrast imaging. Although this clearly goes beyond the scope of our study, the subject still deserves a discussion.

As regards aliasing, it is worth noting that the LIFTed Shack-Hartmann sensor could be spatially filtered [33], in the same way it is currently done with the classical Shack-Hartmann on SPHERE [1]. On-sky results of SPHERE show that this technique drastically reduces the aliasing effects [1,34]. Also, it has been shown via end-to-end simulations that the spatially filtered Shack-Hartmann and the pyramid sensors have a similar behavior with respect to aliasing, leading to a similar exoplanet detectability at high flux [35].

One may also be concerned by performance on low orders since this may affect coronagraphic efficiency. Noise propagation on low orders, including tip-tilt, is clearly higher on both Shack-Hartmann sensors than on the pyramid sensor. However, the sensitivity to low order residuals depends on the type of coronagraph: “interferometric” coronographs (e. g. a four-quadrant phase mask [36]), as observed on SPHERE [37], have a high sensitivity to these modes, while “occulting” coronagraph (e. g. Lyot’s coronagraph [38]) are much more permissive [39]. Such a behavior has indeed been observed experimentally on SPHERE [37].
addition, one has to remember that noise propagation is not the sole error term on low orders, temporal error being generally of the same order of magnitude.

An overall performance evaluation would therefore deserve a specific study accounting for: system parameters (number of actuators, sampling frequency...), turbulence conditions (seeing, wind speed), size of the wavefront sensing spatial filter (if any), type of coronagraph... End-to-end simulations would probably be required to obtain a precise performance evaluation.

6. Conclusion

We have used a wavefront sensor comparison method based on the Fisher information matrix, from which we derive Fisher coefficients (similar to noise propagation coefficients). It allows a fair comparison as it evaluates directly the information available in the data, disregarding the estimator used.

We have applied this method to evaluate the noise propagations of three wavefront sensors in a high-order wavefront sensing application: the Shack-Hartmann sensor, the pyramid sensor and the LIFTed Shack-Hartmann sensor, which is able to extract more information from the pixels than the classical one, without a significative loss of computational time. We considered the estimation of 1000 Karhunen-Loève polynomials at diffraction limit on 40×40 subapertures (20×20 for the LIFTed Shack-Hartmann sensor), in both photon noise and read-out noise regimes. Our study is based on an accurate diffractive model of these sensors. This approach could be extended to other wavefront sensors and/or applications.

We have shown that, in terms of Fisher coefficients, the LIFTed Shack-Hartmann sensor outperforms the classical Shack-Hartmann sensor by a factor 2 in photon noise regime and 1.6 in read-out noise regime. Its overall performance is comparable to a pyramid with a 6 \( \lambda/D \) modulation radius in photon noise limited applications. Moreover, in photon noise regime, the LIFTed Shack-Hartmann sensor has a lower noise propagation than the pyramid sensor in high orders, with a gain over a modulated pyramid sensor going from 1 to 5 between the modes 250 and 1000. This could lead to a better attenuation of residual speckles in the second half of the corrected field in an exoplanet imaging system such as SPHERE. The LIFTed Shack-Hartmann
sensor therefore presents a significant asset for XAO. A further study of the LIFTed Shack-Hartmann sensor, comprising other sources of error such as aliasing effects and temporal error, will be the subject of future works.

Acknowledgments

This work was funded by the European Commission under FP7 Grant Agreement No. 312430 Optical Infrared Co-ordination Network for Astronomy, and by the Office National d’Etudes et de Recherches Aérospatiales (ONERA) in the frame of the NAIADE Research Project.

A. Fisher information matrix of data with additive Gaussian noise

Let \( y = \{y_1, y_2, \ldots, y_N\} \) be a set of data, depending on a set of wavefront mode coefficients \( A = \{a_1, a_2, \ldots, a_M\} \) and a noise \( n = \{n_1, n_2, \ldots, n_N\} \). The expression of the Fisher matrix is:

\[
F_{i,j}(A_p) = E\left[\left(\frac{\partial}{\partial a_i} \ln p(y|A)\right) \left(\frac{\partial}{\partial a_j} \ln p(y|A)\right) \right]_{A=A_p} \tag{11}
\]

with \( p(y|A) \) the likelihood function and \( A_p \) the operating point.

For an additive Gaussian noise on data, the likelihood function is:

\[
p(y|A) = p_b(y) = \prod_{k=1}^{N} \frac{1}{\sqrt{2\pi \sigma_k^2}} \exp\left\{-\frac{[y_k - \tilde{y}_k(A)]^2}{2\sigma_k^2}\right\} \tag{12}
\]

with \( \sigma_k^2 \) the noise variance on the data element \( y_k \).

To find \( F(A_p) \), we need to compute \( \frac{\partial}{\partial a_i} \ln p(y|A)\big|_{A=A_p} \). We can write:

\[
\ln p(y|A) = \sum_{k=1}^{N} \frac{1}{2} \ln (2\pi \sigma_k^2) - \frac{[y_k - \tilde{y}_k(A)]^2}{2\sigma_k^2} \tag{13}
\]

\[
\frac{\partial}{\partial a_i} \ln p(y|A)\big|_{A=A_p} = \sum_{k=1}^{N} \frac{1}{\sigma_k^2} \frac{\partial \tilde{y}_k(A)}{\partial a_i} \big|_{A=A_p} \times [y_k - \tilde{y}_k(A)] \tag{14}
\]

Hence:

\[
\frac{\partial}{\partial a_i} \ln p(y|A)\big|_{A=A_p} \times \frac{\partial}{\partial a_j} \ln p(y|A)\big|_{A=A_p} = \sum_{k=1}^{N} \frac{1}{\sigma_k^2} \frac{\partial \tilde{y}_k(A)}{\partial a_i} \big|_{A=A_p} \frac{\partial \tilde{y}_k(A)}{\partial a_j} \big|_{A=A_p} \times [y_k - \tilde{y}_k(A)]^2
\]

\[
+ \sum_{l \neq k} \frac{1}{\sigma_l^2 \sigma_k^2} \frac{\partial \tilde{y}_l(A)}{\partial a_i} \big|_{A=A_p} \frac{\partial \tilde{y}_l(A)}{\partial a_j} \big|_{A=A_p} \times [y_k - \tilde{y}_k(A)] \times [y_l - \tilde{y}_l(A)] \tag{15}
\]

Knowing that \( E\{[y_k - \tilde{y}_k(A)]^2\} = \sigma_k^2 \) and \( E\{[y_k - \tilde{y}_k(A)][y_l - \tilde{y}_l(A)]\} = \sigma_{kl}^2 \), we finally have:

\[
F_{i,j}(A_p) = \sum_{k=1}^{N} \frac{1}{\sigma_k^2} \frac{\partial \tilde{y}_k(A)}{\partial a_i} \big|_{A=A_p} \frac{\partial \tilde{y}_k(A)}{\partial a_j} \big|_{A=A_p} + \sum_{l \neq k} \frac{\sigma_{kl}^2}{\sigma_l^2 \sigma_k^2} \frac{\partial \tilde{y}_l(A)}{\partial a_i} \big|_{A=A_p} \frac{\partial \tilde{y}_l(A)}{\partial a_j} \big|_{A=A_p} \tag{16}
\]

If we consider that the noise is uncorrelated from one data element \( y_i \) to another, the expression is simplified into:

\[
F_{i,j}(A_p) = \sum_{k=1}^{N} \frac{1}{\sigma_k^2} \frac{\partial \tilde{y}_k(A)}{\partial a_i} \big|_{A=A_p} \frac{\partial \tilde{y}_k(A)}{\partial a_j} \big|_{A=A_p} \tag{17}
\]
B. Maximum likelihood estimation and Fisher information matrix

The Fisher information matrix is often seen as a complicated mathematical object which cannot be easily related to physical concepts. The goal of this paragraph is to link the Fisher information matrix with a more familiar figure of merit: the covariance matrix of estimation error for a maximum likelihood estimator with a Gaussian noise model.

B.1. Noise propagation in a maximum likelihood estimation

The solution given by the maximum likelihood estimator for the linear model described by Eq. (2) is:

\[ \hat{A} = (D^t C_n^{-1} D)^{-1} D^t C_n^{-1} y \]  \hspace{1cm} (18)

with \( C_n \triangleq <n n^t> \) the noise covariance matrix. The covariance matrix of the estimation error is then:

\[ <E E^t>_ML = (D^t C_n^{-1} D)^{-1} \]  \hspace{1cm} (19)

The variance of estimation error for each mode is given by the diagonal elements of this matrix.

For any wavefront sensor and an unbiased estimator, we can express the total variance of estimation error by [6]:

\[
\text{trace}(<E E^t>_ML) = C_{\text{ph}} \left( \sum_i C_{\text{ph},i} n_{\text{ph}} \right) + C_{\text{det}} \left( \sum_i C_{\text{det},i} \left( \frac{\sigma_e}{n_{\text{ph}}} \right)^2 \right)
\]  \hspace{1cm} (20)

with \( \hat{A} \) the vector of estimated coefficients, \( E = \hat{A} - A \) the estimation error, \( n_{\text{ph}} \) the incoming flux in photo-electrons and \( \sigma_e \) the standard deviation of the read-out noise. \( C_{\text{ph},i} \) and \( C_{\text{det},i} \) are the noise propagation coefficients on the \( i \)-th mode for photon noise and read-out noise respectively.

For Shack-Hartmann slopes, assuming the noise is homogeneous and uncorrelated from one slope to another, \( C_n \) is diagonal. The noise propagation coefficients on each mode are then proportional to the diagonal elements of \( (D^t D)^{-1} \). Rigaut and Gendron used this result to find an analytical formulation of the noise propagation in the Shack-Hartmann [5]. In the asymptotic case of an infinite number of subapertures, they demonstrated that the noise propagation coefficient for each mode was proportional to \( (n + 1)^{-2} \), with \( n \) the radial order of the considered mode. This result is typical for slope sensors [6].

B.2. Link with the Fisher information matrix

Let us see if we can compare \( <E E^t>_ML = (D^t C_n^{-1} D)^{-1} \) and the Fisher information matrix, as expressed in Eq. (16). We first need to find the expression of the interaction matrix \( D \). From Eq. (2), we can write:

\[ D = \frac{\partial y}{\partial \hat{A}} \]  \hspace{1cm} (21)

For \( m \) modes and \( p \) data elements, the expression of \( D \) is then:

\[
D = \begin{pmatrix}
\frac{\partial y_1}{\partial a_1} & \frac{\partial y_1}{\partial a_2} & \ldots & \frac{\partial y_1}{\partial a_m} \\
\frac{\partial y_2}{\partial a_1} & \frac{\partial y_2}{\partial a_2} & \ldots & \frac{\partial y_2}{\partial a_m} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial y_p}{\partial a_1} & \frac{\partial y_p}{\partial a_2} & \ldots & \frac{\partial y_p}{\partial a_m}
\end{pmatrix}
\]  \hspace{1cm} (22)
We thus have:

$$[D^t C_n^{-1} D]_{i,j} = \sum_k \frac{1}{\sigma_k^2} \frac{\partial y_k}{\partial a_i} \frac{\partial y_k}{\partial a_j} + \sum_{l \neq k} \frac{\sigma_{ij}^2}{\sigma_i^2 \sigma_j^2} \frac{\partial y_l}{\partial a_i} \frac{\partial y_l}{\partial a_j} = F_{i,j}$$

(23)

Hence, $< EE^t >_{ML} = F^{-1}$. The covariance matrix of error estimation of the maximum likelihood estimator is thus equal to the Cramèr-Rao lower bound, as expected for a Gaussian noise model. Moreover, we can write, similarly to Eq. (20):

$$trace(F^{-1}) = \left( \sum_i \alpha_i \right) \frac{1}{n_{ph}} + \left( \sum_i \beta_i \right) \left( \frac{\sigma_e}{n_{ph}} \right)^2$$

(24)

with $\alpha_i = C_{ph,i}^{ML}$ and $\beta_i = C_{det,i}^{ML}$ the Fisher coefficients for photon noise and read-out noise respectively.