

A priori information and optimisation in polarimetry

Matthew R. Foreman, Carlos Macías Romero, and Peter Török*

Blackett Laboratory, Department of Physics, Imperial College London, Prince Consort Road, London SW7 2BZ, UK

*Corresponding author: peter.torok@imperial.ac.uk

Abstract: Polarimetric measurements are designed to obtain information pertaining to the system under study, however noise in the system limits the precision and hence information obtainable. Exploitation of *a priori* knowledge of the system allows for an improvement in the precision of experimental data. In this vein we present a framework for system design and optimisation based upon the Fisher information matrix, which allows easy incorporation of such *a priori* information. As such the proposed figure of merit is more complete than the commonly used condition number. Conditions of equivalence are considered, however a number of examples highlight the failings of the condition number under more general scenarios. Bounds on the achievable informational gains via multiple polarimeter arms are also given. Finally we present analytic results concerning error distribution in a Mueller matrix polar decomposition, allowing for a more accurate noise analysis in polarimetric experiments.

© 2008 Optical Society of America

OCIS codes: (110.3055) Information theoretical analysis; (120.5410) Polarimetry; (220.4830) Systems design

References and links

1. *Planets Stars and Nebulae Studied With Photopolarimetry*, T. Gehrels ed. (University of Arizona Press 1974).
2. J. Tinbergen, "Interstellar polarization in the immediate solar neighbourhood," *Astron. Astrophys.* **105**, 53–64 (1982).
3. R. A. Chipman, *Polarimetry, Handbook of Optics Vol 2*, (McGraw-Hill, New York, 1995).
4. R.M.A. Azzam and N.M. Bashara, *Ellipsometry and polarised light*, (Elsevier, North Holland, 1987).
5. J. Zallat, S. Ainouz, and M. Ph. Stoll "Optimal configurations for imaging polarimeters: impact of image noise and systematic errors," *J. Opt. A: Pure Appl. Opt.* **8** 807–814 (2006).
6. J. S. Tyo, "Optimum linear combination strategy for an N -channel polarization sensitive imaging or vision system," *J. Opt. Soc. Am. A* **15** 359–366 (1998).
7. J. S. Tyo, "Design of optimal polarimeters: maximization of signal-to-noise ratio and minimization of systematic error," *Appl. Opt.* **41** 619–630 (2002).
8. A. Ambirajan and D.C. Look, "Optimum angles for a polarimeter," *Opt. Eng.* **34** 1651–1658 (1995).
9. M. Smith, "Optimization of a dual-rotating-retarder Mueller matrix polarimeter," *Appl. Opt.* **41** 2488–2493 (2002).
10. A. De Martino, E. Garcia-Caurel, B. Laude, and B. Drévilion, "General methods for optimized design and calibration of Mueller polarimeters," *Thin Solid Films*, **455–456** 112–119 (2004).
11. D. Mendlovic, and A. W. Lohmann "Spacebandwidth product adaptation and its application to superresolution: fundamentals," *J. Opt. Soc. Am. A* **14** (1997).
12. T. J. Rothenberg, *Efficient estimation with a priori information* (New Haven, Yale University Press, 1973).
13. I. J. Cox, and C. J. R. Sheppard, "Information capacity and resolution in an optical system," *J. Opt. Soc. Am. A* **3** 1152–1158 (1986).
14. S. Y. Lu and R. A. Chipman, "Interpretation of Mueller matrices based on polar decomposition," *J. Opt. Soc. Am. A* **13**, 1106–1113 (1996).

15. R. M. A. Azzam, "Division-of-amplitude Photopolarimeter (DOAP) for the simultaneous measurement of all four Stokes parameters of light," J. Mod. Opt. **29** 685–689 (1982).
16. E. Collett, "Automatic determination of the polarization state of nanosecond laser pulses," U.S. Patent No. 4158506 (1979).
17. H. Mueller, "The foundations of optics" J. Opt. Soc. Am. **38**, 661–661 (1948).
18. V. Delaubert, N. Treps, C. Fabre, H. A. Bachor, and P. Réfrégier, "Quantum limits in image processing," J. Mod. Opt. **29** 685–689 (1982)
19. M. R. Foreman, S. S. Sherif, and P. Török, "Photon statistics in single molecule orientational imaging," Opt. Express **15** 13597–13606 (2007).
20. A. Dubois, K. Grieve, G. Moneron, R. Lecaque, L. Vabre, and C. Boccara, "Ultrahigh-Resolution Full-Field Optical Coherence Tomography," Appl. Opt. **43**, 2874–2883 (2004).
21. B. J. Meers, "Recycling in laser-interferometric gravitational-wave detectors," Phys. Rev. D **38** 2317–2326 (1988).
22. J. W. Goodman, *Statistical Optics* (Wiley, New York, 2000).
23. A. Bénére, F. Goudail, M. Alouini, and D. Dolfi, "Estimation precision of degree of polarization in the presence of signal-dependent and additive Poisson noises," J. Europ. Opt. Soc. Rap. Public. **3** 08002 (2008).
24. H. Cramér, *Mathematical Methods of Statistics*, (Princeton Univ. Press, 1946), ISBN 0-691-08004-6.
25. C. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," Bull. Calcutta Math. Soc. **37** 81-89 (1945)
26. R. Fisher, "On the mathematical foundations of theoretical statistics," Phil. Trans. R. Soc. Lond. **222** 309–368 (1922)
27. R. Fisher, "Theory of statistical estimation," Proc. Cam. Phil. Soc. **22** 700-725 (1925)
28. <http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/calculus.html>
29. H. H. Barrett, J.L. 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)
30. L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis* (Addison-Wesley, 1991)
31. E. W. Barankin, "Locally best unbiased estimates," Ann. Math. Stat. **20** 477-501 (1949).
32. M. F. Kijewski, S. P. Müller, and S. C. Moore, "The Barankin bound: a model of detection with location uncertainty," Proc. SPIE **1768** 153 (1992).
33. S. P. Müller, C. K. Abbey, F. J. Rybicki, S. C. Moore, and M. F. Kijewski, "Measures of performance in nonlinear estimation tasks: prediction of estimation performance at low signal-to-noise ratio," Phys. Med. Biol. **50** 3697–3715 (2005).
34. B. R. Frieden *Physics from Fisher Information: A Unification*, (Cambridge University Press 1998).
35. J. S. Tyo, "Noise equalisation in Stokes parameter images obtained by use of variable-retardance polarimeters," Opt. Lett. **25** 1198-1200 (2000).
36. A. Bénére, F. Goudail, M. Alouini, and D. Dolfi, "Precision of degree of polarization estimation in the presence of additive Gaussian detector noise," Opt. Commun. **278** 264–269 (2007).
37. R. Mehra, "Optimal input signals for parameter estimation in dynamic systems—Survey and new results," IEEE T. Automat. Contr. **19** 753–768 (1974).
38. E. Walter and L. Pronzatom "Qualitative and quantitative experiment design for phenomenological models - a survey," Automatica **26** 195–213 (1990).
39. S. N. Savenkov, "Optimization and structuring of the instrument matrix for polarimetric measurements," Opt. Eng. **41** 965–972 (2002).
40. V. Bhapkar and C. Srinivasan, "On Fisher information inequalities in the presence of nuisance parameters," Ann. Inst. Statist. Math **46** 593–604 (1994).
41. R. M. A. Azzam and F. F. Sudradjat, "Single-layer-coated beam splitters for the division-of-amplitude photopolarimeter," Appl. Opt. **44** 190–196 (2005).
42. A. D. Whalen, *Detection of signals in noise*, (Academic Press Inc., New York, 1971).
43. P. Török, M. Salt, E.E. Kriezis, P.R.T. Munro, H.P. Herzig, and C. Rockstuhl, "Optical disk and reader therefor," Worldwide Patent No. WO 2006/010882 (2006).
44. P. R. T. Munro and P. Török, "Properties of confocal Mueller-matrix polarimeters," (submitted to Opt. Lett.).
45. B. Laude-Boulesteix, A. De Martino, B. Drévillon, and L. Schwartz, "Mueller Polarimetric Imaging System with Liquid Crystals," Appl. Opt. **43**, 2824–2832 (2004)
46. J. M. Bueno, "Depolarization effects in the human eye," Vision Research **41** 2687–2696 (2001).
47. M. Floc'h, G. Le Brun, J. Cariou, and J. Lotrian, "Experimental characterization of immersed targets by polar decomposition of the Mueller matrices," Eur. Phys. J. AP **3** 349–358 (1998).
48. S. M. Nee, "Error analysis for Mueller matrix measurement," J. Opt. Soc. Am. A **20** 1651–1657 (2003).
49. S. N. Savenkov and K. E. Yushin, "Mueller matrix elements error distribution for polarimetric measurements," Proc. SPIE **5158** 251–259 (2003).
50. J. Morio and F. Goudail, "Influence of the order of diattenuator, retarder, and polarizer in polar decomposition of Mueller matrices," Opt. Lett. **29** 2234–2236 (2004).

1. Introduction

Polarimetry is the study and measurement of the polarisation state of light. Although measurement of the state of polarisation of light is often an important objective [1,2] such techniques are also frequently used to obtain information about an optical system, such as its birefringence [3]. One may then subdivide polarimetry into two broad categories; Stokes polarimetry and Mueller polarimetry. The former entails measuring the four Stokes parameters of light [4] whilst the latter is intended to measure the full Mueller matrix of a sample from which parameters of interest can be inferred. Much effort has been invested into determining optimal configurations of polarimeters in terms of their experimental setup [5–10], however invariably little consideration is given to *a priori* information that an experimenter may have about the system they are studying. Information theory however states that if exploited correctly such information can improve the accuracy of any measurements [11, 12]. For example if the position of an object is approximately known the field of view can be reduced, perhaps by using a confocal microscope giving rise to an increase in the bandwidth and hence resolution of the system [13]. To address this omission we consider how such *a priori* information can be represented and incorporated into system optimisation. In doing so we find that common optimisation procedures do not necessarily give the optimal polarimeter configuration. Furthermore our formulation naturally describes the distribution of errors amongst inferred polarisation parameters, such as diattenuation, retardance and depolarisation as may be obtained from a Lu-Chipman Mueller matrix decomposition [14]. These results could potentially be used for a more accurate noise analysis in polarimetry.

In Section 2 we first describe the system model of Stokes and Mueller polarimeters used throughout this text and a discussion of the noise model assumed. Section 3 then proceeds to discuss *a priori* information before we discuss its use in system optimisation in Section 4. A number of examples are given in Section 5 where we also consider and confirm the approach used when no *a priori* information is possessed. Finally in Sections 6 and 7 we discuss how our optimisation strategy can be extended and highlight the inherent description of noise propagation in polarimetry, before concluding in Section 8.

2. System description

2.1. Stokes polarimeter

A division of amplitude polarimeter (DOAP) as originally proposed by Azzam [15] can measure the Stokes vector $\vec{S} = (S_0, S_1, S_2, S_3)$ of incident light by projecting it onto at least four independent polarisation basis states using a polarisation state analyser (PSA) and as such the vector of detected intensities $\vec{D} = (D_0, D_1, \dots, D_{N_D-1})$ can be written

$$\vec{D} = \mathbf{T}\vec{S} \quad (1)$$

where \mathbf{T} is a $N_D \times 4$ matrix with rows corresponding to the Stokes vectors of the N_D polarisation basis states, known as the instrument matrix (see Fig. 1). It is by variation of the instrument matrix that we aim to optimise a Stokes polarimeter. Each row of the instrument matrix can be deduced from the polarisation elements present in each arm of the polarimeter setup used and is thus known. Given a set of measured intensities it is then possible to deduce the state of polarisation of the incident light by application of the inverse operation i.e.

$$\vec{S} = \mathbf{T}^+\vec{D} \quad (2)$$

where \mathbf{T}^+ is the Moore-Penrose pseudoinverse of the instrument matrix.

Although a division of wavefront polarimeter (DOWP) [16] can also be used for polarimetric measurements we neglect this arrangement in this work since it requires beams that are

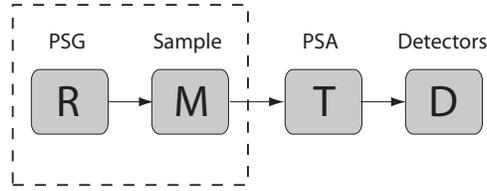


Fig. 1. Schematic diagram of a polarimeter setup.

uniformly polarised and that the beam intensity profile be known *a priori*; conditions which are generally not achieved in practice. DOWPs are hence rarely used.

2.2. Mueller polarimeter

A Mueller polarimeter builds on the principle of a DOAP by addition of a polarisation state generator (PSG) and a sample to the optical setup as shown in Fig. 1. The action of the sample on the incident polarised light can be described by a 4×4 Mueller matrix \mathbf{M} [17] such that the light incident into the DOAP is described by the Stokes vector

$$\vec{S} = \mathbf{M}\vec{R} \quad (3)$$

where \vec{R} is the Stokes vector of the illuminating beam. Since the Mueller matrix has 16 elements, all of which must in general be determined, it is not sufficient to illuminate using a single polarisation state. At minimum four distinct polarisation states must be used to illuminate the sample. Under these circumstances \vec{R} can be written as a $4 \times N_R$ matrix, \mathbf{R} , $N_R \geq 4$ whose columns are the Stokes vectors of the input states. Consequently \vec{D} becomes a $N_D \times N_R$ matrix whose columns correspond to the vector of detected intensities for each input polarisation state. We can hence write

$$\mathbf{D} = \mathbf{TMR} \quad (4)$$

from which the Mueller matrix can be found using the inverse operations i.e.

$$\mathbf{M} = \mathbf{T}^+\mathbf{DR}^+ \quad (5)$$

The incident polarisation matrix \mathbf{R} provides additional degrees of freedom by which a Mueller polarimeter can be optimised.

2.3. Noise model

All practical systems are subject to noise of a stochastic nature. It is this unfortunate fact which limits the precision to which measurements can be made and hence the amount of information we can extract from said measurements. In optical systems noise can be attributed to two sources which Delaubert *et. al* [18] refer to as technical and quantum noise. The former arises from poor experimental setup and can in principle be reduced to an arbitrarily low level, however quantum noise, arising from the discrete nature of light and the stochastic nature of photon arrivals at the detector, can not. Despite being a fundamental limit to experimental precision the assumption of quantum noise is also valid for any shot noise limited setup, a limit that has been reached in many applications, for example single molecule detection, OCT and astronomy [19–21]. Goodman's discussion of the degeneracy parameter [22] also illustrates that at

optical frequencies quantum noise is frequently dominant over other potential fluctuations in the source.

So as not to unnecessarily complicate calculations we adopt a semi-classical model of light whereby before incidence onto the detector light is modelled using traditional electromagnetic theory, whilst the final absorption process is necessarily quantum in nature. Under these assumptions it can be shown [22] that the number of photocounts registered by a photon counting detector is governed by a Poisson probability law

$$f(D|\bar{D}) = \frac{\bar{D}^D}{D!} \exp(-\bar{D}) \quad (6)$$

where $f(D|\bar{D})$ denotes the probability density function (PDF) for registering D photocounts and the bar notation is used to highlight that the PDF is dependent on the ensemble average \bar{D} of possible measurements. Since the number of photocounts is proportional to the integrated intensity received on the detector according to $D = \eta W / h \bar{\nu}$ we refer to D simply as the detected intensity. Here η is the quantum efficiency of the detector, whilst $h \bar{\nu}$ is the average energy of an incident photon.

DOAPs however have multiple detectors and it is necessary to consider the stochastic detection process on each. Assuming that the noise on each detector is independent the joint PDF is

$$f(\mathbf{D}|\bar{\mathbf{D}}) = \prod_{n=1}^{N_D N_R} \frac{(\bar{D}_n + I_n^d)^{D_n}}{D_n!} \exp(-\bar{D}_n - I_n^d) \quad (7)$$

where an additional term I_n^d has been added to account for other potential additive sources of stray photons. The notation \mathbf{D} can be read as either a matrix or vector of intensities, since the formulation in both cases is identical. A good discussion of such possible noise sources is given in [23], however we give the simple examples of detector dark count or a passive background. Although not necessary we make the simplifying assumption that these additional noise sources affect each detector equally such that $I_n^d = I^d$.

3. Information in polarimetry

3.1. Deterministic systems

Optical measurements are intended to extract information about the system under study, however as discussed noise in the system limits the amount obtainable. This relationship can be placed on a more formal basis using the Cramér-Rao lower bound (CRLB) [24, 25] which states that $\mathbf{K} \geq \mathbf{J}_w^{-1}$. Here and henceforth we assume we are estimating a set of parameters \vec{w} . \mathbf{K} is then the covariance matrix of our estimate i.e. the obtainable precision, and \mathbf{J}_w is known as the Fisher information matrix (FIM) [26, 27] defined by

$$\mathbf{J}_w = E_D \left[\frac{\partial \ln f(\mathbf{D}|\vec{w})}{\partial \vec{w}} \frac{\partial \ln f(\mathbf{D}|\vec{w})}{\partial \vec{w}}^T \right] \quad (8)$$

$E_D[\dots]$ denotes the expectation or ensemble average with respect to \mathbf{D} . In essence Fisher information is a statistical measure of the dependence of the observed data on the input parameters. A strong dependence implies that we can measure the parameters more precisely. Application of the chain rule to Eq. (8) quickly yields

$$\mathbf{J}_w = \mathbf{G}^T \mathbf{J}_{\bar{\mathbf{D}}} \mathbf{G} \quad (9)$$

where $\mathbf{G} = \frac{\partial \bar{\mathbf{D}}}{\partial \vec{w}}$ is independent of \mathbf{D} . Various authors use different conventions for matrix calculus however throughout this work we adopt the formalism described in [28]. From Eqs. (7) and (8) we arrive at the well known result that the FIM for estimation of the mean of a Poisson random process is given by

$$(\mathbf{J}_{\bar{\mathbf{D}}})_{ij} = \frac{\delta_{ij}}{\bar{D}_i + I^d} \quad (10)$$

where δ_{ij} is the Kronecker delta. Eqs. (9) and (10) thus define the FIM associated with estimation of a parameter vector \vec{w} from intensity measurements in the presence of Poisson noise. We can consider Stokes (or Mueller matrix) polarimeters by allowing $\vec{w} = \vec{S}$ (or \mathbf{M}) to give the associated FIM, where by equating \vec{w} to \mathbf{M} we imply a vectorisation operation of \mathbf{M} .

3.2. Stochastic systems

Equations given thus far are valid only for a particular value of \vec{w} however the parameter values may differ between different experimental setups or measurements. The experimenter may however know from an existing model or earlier data that the object being studied belongs to a restricted class, that is to say they possess some *a priori* information about the parameters being measured. A fibre-optic communication channel provides a good example whereby it is known that during a measurement window either a pulse will be received or not with equal probability representing logical 1 and 0 respectively. Known restrictions on the possible values of \vec{w} can be conveniently parameterised using a PDF $f(\vec{w})$ which describes the probability of each value of \vec{w} occurring. As shown in the accompanying Appendix when \vec{w} can vary the FIM takes the form

$$\mathbf{J}_w = E_w [\mathbf{J}_w^{nr}] + \mathbf{J}_w^r \quad (11)$$

where \mathbf{J}_w^{nr} is the deterministic FIM given by Eq. (9) and \mathbf{J}_w^r depends only on our *a priori* knowledge via Eq. (46).

Although the CRLB and FIM are useful tools for defining a best obtainable precision it is worth questioning whether such a limit can practically be achieved. In answer we refer to [29] wherein it is shown that each photon can be considered as an independent measurement of the system. Furthermore the maximum-likelihood (ML) estimator [30] asymptotically achieves the CRLB as the number of samples increases. We calculate using the Central Limit Theorem that acceptable convergence requires ~ 8000 photons and hence under these conditions we can achieve the CRLB by using a ML estimator. If such a criteria is not met then the CRLB is not the strongest bound and one must resort to alternative bounds, such as the Barankin bound as discussed in [31–33]. Full exposition of these bounds is beyond the scope of this paper.

3.3. Channel capacity

Frieden [34] defined a metric C_w to quantify the ability of a system to transmit Fisher information about the system being studied, known as the Fisher channel capacity and given mathematically by the trace of the FIM. Parameter estimation can only be accomplished via some degree of data processing on the measured intensities. It is hence sufficient to consider the channel capacity in the context of estimating the mean intensities only i.e. $\text{tr}(\mathbf{J}_{\bar{\mathbf{D}}})$, since it is impossible for data processing steps to introduce additional information into the system. We consider first a Stokes polarimeter. From Eq. (10) we have

$$C_{\bar{\mathbf{D}}} = \sum_{n=1}^{N_D} \frac{1}{\bar{D}_n} \quad (12)$$

which under the constraint $0 < \sum_{n=1}^{N_D} \bar{D}_n \leq S_0$ is a maximum when the intensity is equalised across the detectors, such that $D_n = aS_0/N_D$, where $a \leq 1$ is a positive constant. We note that equalising the intensity on each detector also equalises the noise which is considered a desirable property in optimisation of polarimeters [5,35]. The channel capacity is thus bounded according to

$$C_{\bar{D}} \leq \frac{N_D^2}{S_0 + N_D I^d}. \quad (13)$$

We thus see that although channel capacity initially increases quadratically with the number of detectors, this slows to a linear increase as the additive noise term I^d becomes significant. We note here that if a Gaussian noise model were used, as is frequently done [18, 36], the channel capacity would increase linearly with N_D . Since Fisher information is additive [34] the Fisher channel capacity for a Mueller polarimeter can be calculated by summing the capacities for each input polarisation state. Thus

$$C_{\bar{D}} \leq \frac{N_R N_D^2}{R_0 + N_D I^d} \quad (14)$$

where R_0 is the intensity of the probing polarisation states. We note that this bound is unlikely to be achieved since equality requires the sample to be perfectly transmitting to all incident polarisation states. The channel capacity for a Mueller polarimeter can be seen to scale only linearly with respect to the number of probing polarisation states.

Practically these results embody the intuitive result that by performing more measurements i.e. increase the sampling, we introduce greater redundancy into our experimental data hence allowing a better precision in our parameter estimates to be achieved. Henceforth we assume that $N_D = N_R = 4$ since this is the minimum number of measurements required to determine \vec{S} or \mathbf{M} uniquely. The associated FIMs are thus 4×4 and 16×16 respectively.

4. Optimisation of polarimeters

4.1. Ellipsoids of concentration

Consider a parameter column vector \vec{w} (of length n) and an estimate made by an experimenter of this parameter vector denoted \vec{w}_e . In general there will be a discrepancy between the estimated and true value of \vec{w} due to noise in the system given by

$$\Delta \vec{w} = \vec{w} - \vec{w}_e \quad (15)$$

We note that since \vec{w}_e is derived from random experimental data $\Delta \vec{w}$ is a random variable. Subsequent experiments will give a different value for the vector $\Delta \vec{w}$. In the n -dimensional error space i.e. the space with coordinate axes given by $\Delta w_1, \Delta w_2, \dots, \Delta w_n$ each realisation of $\Delta \vec{w}$ specifies a single point. The scatter of these points is determined by the variations in the experimenters estimate \vec{w}_e as described by the associated covariance matrix \mathbf{K} . Furthermore the extent of scatter of different errors can be visualised using concentration ellipsoids [30] defined by $\Delta \vec{w} \mathbf{K}^{-1} \Delta \vec{w}^T = c$, the volume of which can be used as a metric of aggregate error. Via the CRLB the FIM defines the smallest possible ellipsoid of concentration, which has a volume of

$$V_{min} = V_n \sqrt{c |\mathbf{J}_w^{-1}|} = V_n \sqrt{\frac{c}{|\mathbf{J}_w|}} \quad (16)$$

where V_n is the volume of the n -dimensional unit hypersphere. As such we use the determinant of the FIM as a figure of merit, whereby a larger determinant is preferable. This criterion is

known as D -optimality [37, 38] and has the additional advantage that $\log |\mathbf{J}_w|$ is closely related to the Fisher channel capacity of the system, such that by maximising $|\mathbf{J}_w|$ we are also maximising the channel capacity.

Setting $\vec{w} = \vec{S}$ or \mathbf{M} and using Eqs. (1), (4) and (9) it is possible to calculate the FIM for estimation of the Stokes parameters or the elements of the Mueller matrix respectively as

$$\mathbf{J}_S = \mathbf{T}^T \mathbf{J}_{\bar{D}} \mathbf{T} \quad (17a)$$

$$\mathbf{J}_M = (\mathbf{R} \otimes \mathbf{T}^T) \mathbf{J}_{\bar{D}} (\mathbf{R}^T \otimes \mathbf{T}) \quad (17b)$$

where $\mathbf{J}_{\bar{D}}$ is of the form of Eq. (11) and \otimes denotes the Kronecker product. Hence

$$|\mathbf{J}_S| = |\mathbf{T}|^2 |\mathbf{J}_{\bar{D}}| \quad (18a)$$

$$|\mathbf{J}_M| = |\mathbf{R}|^8 |\mathbf{T}|^8 |\mathbf{J}_{\bar{D}}| \quad (18b)$$

We now restrict to uniform distributions meaning $\mathbf{J}_{\bar{D}}^r = 0$, so as to illustrate the general behaviour of these figures of merit, whereby

$$|\mathbf{J}_S| = |\mathbf{T}|^2 \prod_{i=1}^4 E_S \left[\frac{1}{D_i + I^d} \right] \quad (19a)$$

$$|\mathbf{J}_M| = |\mathbf{R}|^8 |\mathbf{T}|^8 \prod_{i=1}^{16} E_M \left[\frac{1}{D_i + I^d} \right] \quad (19b)$$

Closer inspection of Eqs. (18) reveals there are two factors which influence the amount of information received in an optical experiment. The first of these corresponds to the amount of information acquired during the physical measurement as described by the $|\mathbf{J}_{\bar{D}}|$ term. This component also encompasses any *a priori* information that may be possessed.

The second, and perhaps the more familiar, influence is related to any subsequent data processing used to extract the Stokes vector or Mueller matrix from the measured intensities as per Eqs. (2) and (5). Any noise in the intensity measurements is amplified during data processing, the extent of which is often measured using the condition number of the associated matrices, namely \mathbf{T} (and \mathbf{R}). In this paper we define the condition number of the matrix \mathbf{X} as $\kappa_X = \|\mathbf{X}\|_F \|\mathbf{X}^{-1}\|_F$ where $\|\dots\|_F$ denotes the Frobenius norm. Since the condition number of a matrix is inversely proportional to its determinant (see e.g. [39]) we see that noise amplification is described by the $|\mathbf{T}|$ (and $|\mathbf{R}|$) factors of Eqs. (18).

Frequently the condition number of the instrument matrix (and input polarisation matrix) is used as a figure of merit for polarimeter optimisation [7–10]. Our discussion above however has highlighted the inadequacy of this strategy in general since it gives no regard to potential gains that can be made by improving the precision of the measurement itself or incorporation of *a priori* knowledge. We thus believe that our informational figure of merit is more holistic in terms of measuring the quality of a polarimeter.

4.2. Nuisance parameters

Hitherto it has been assumed that full knowledge of the Stokes vector or Mueller matrix was desired, however this may not always be the case as demonstrated in Section 5.3. Such a scenario thus warrants our attention, although we shall now give a full commentary since it is amply discussed in the literature e.g. [30].

Let us assume that the parameter vector \vec{w} is formed via the concatenation of two vectors \vec{u} and \vec{v} such that $\vec{w} = (\vec{u}, \vec{v})$. \vec{u} is assumed to contain the p parameters that we wish to know,

whilst \vec{v} contains q parameters that are of little interest, known as nuisance parameters. When nuisance parameters are present the relevant FIM \mathbf{J}_u is given by [40]

$$\mathbf{J}_u = \mathbf{J}_{11} - \mathbf{J}_{12}\mathbf{J}_{22}^{-1}\mathbf{J}_{21} \quad (20)$$

where \mathbf{J}_{11} is $p \times p$, \mathbf{J}_{12} is $p \times q$, \mathbf{J}_{21} is $q \times p$, \mathbf{J}_{22} is $q \times q$ and

$$\mathbf{J}_w = \begin{pmatrix} \mathbf{J}_{11} & \mathbf{J}_{12} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{pmatrix} \quad (21)$$

is the FIM were all parameters to be estimated. Accordingly the matrix determinant on which to optimise the polarimeter setup can be shown via rules of matrix algebra to be

$$|\mathbf{J}_u| = \frac{|\mathbf{J}_w|}{|\mathbf{J}_{22}|} \quad (22)$$

5. Examples

Having developed a more suitable framework within which both Stokes and Mueller polarimeters can be optimised we now give a number of examples to highlight some points of interest. We start by highlighting the circumstances under which our informational figure of merit is equivalent to the condition number, however further examples then illustrate that when *a priori* information is introduced this equivalence does not hold.

5.1. Maximal ignorance

Our first example considers the situation in which we are maximally ignorant of the likely incident polarisation states. Initially considering a Stokes polarimeter our *a priori* information (or our lack thereof) can be modelled by assuming each polarisation state is equally likely. The associated PDF is thus uniform over the Poincaré sphere. Assuming that all possible incident states have the same intensity S_0 and degree of polarisation P (defined as the fraction of the total intensity originating from totally polarised light) it is possible to show

$$\begin{aligned} E_S \left[\frac{1}{D_i + I^d} \right] &= \frac{1}{S_0 P} \log \left[\frac{(1+P)S_0 + 2I^d}{(1-P)S_0 + 2I^d} \right] \\ &= \frac{2}{S_0 P} \operatorname{arctanh} \left[\frac{S_0 P}{S_0 + 2I^d} \right] \end{aligned} \quad (23)$$

Hence

$$|\mathbf{J}_S| = \frac{1}{8} |\mathbf{T}|^2 \left(\operatorname{arctanh} \left[\frac{S_0 P}{S_0 + 2I^d} \right] / S_0 P \right)^4 \quad (24)$$

It is thus apparent that to maximise the information obtained we must maximise $|\mathbf{T}|$ or equivalently make the condition number as small as possible. It can be shown geometrically that this corresponds to making the volume of the tetrahedron whose vertices on the Poincaré sphere are defined by \vec{T}_i a maximum, i.e. making the tetrahedron regular [39]. Although the same conclusion has been previously reached via considerations of the structure of the instrument matrix and noise propagation [7, 35, 39, 41] our derivation based on information theory appears to be new. Since a maximum determinant corresponds to minimal noise amplification the SNR, given by

$$\text{SNR} = \left(\frac{S_0}{2P} \operatorname{arctanh} \left[\frac{S_0 P}{S_0 + 2I^d} \right] \right)^{1/2}, \quad (25)$$

is maximum. If $I^d \ll S_0$ this reduces to the familiar $S_0^{1/2}$ scaling associated with Poisson noise.

There are an infinite number of possible instrument matrices corresponding to the rotation of the tetrahedron within the Poincaré sphere about the origin, however given one optimal instrument matrix (as can easily be found numerically) e.g. [39]

$$\mathbf{T}_{opt} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & -0.333 & -0.816 & 0.471 \\ 1 & -0.333 & 0 & -0.943 \\ 1 & -0.333 & 0.816 & 0.471 \end{pmatrix} \quad (26)$$

it is possible to find alternative and perhaps more practical instrument matrices by applying a suitable rotation matrix. For a Mueller matrix polarimeter the same result applies, however we must also maximise the determinant of \mathbf{R} , that is to say make the incident polarisation states as orthogonal as possible. For the completely uniform distribution there is no relationship between \mathbf{T} and \mathbf{R} .

Finally we note with reference to Fig. 2 that the channel capacity is independent of the choice of instrument matrix and increases with the degree of polarisation as would be expected.

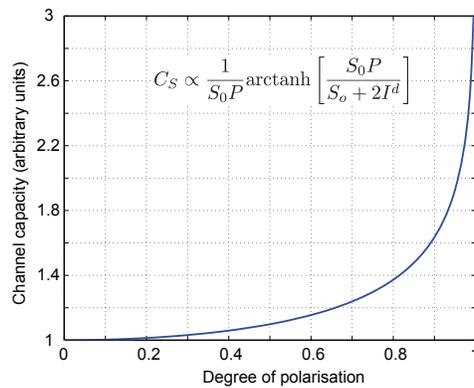


Fig. 2. Dependence of channel capacity on degree of polarisation for $S_0/I^d = 10^4$.

5.2. Matched filter

Adopting the opposite extreme to maximal ignorance we now consider the polarimetric equivalent to the matched filter. Matched filters, which correlate a *known* signal (or template) with a measured signal so as to determine the presence or absence of the known signal against some background, are frequently encountered in signal processing since they maximise the SNR when the desired signal is present [42]. We here restrict our discussion to that of a Stokes polarimeter in which depolarised light constitutes the background signal.

Denoting the known polarisation state by its Stokes vector $\vec{S}_t = (S_{t0}, S_{t1}, S_{t2}, S_{t3})$ our *a priori* knowledge can be represented by the PDF $f(\vec{S}) = \delta(\vec{S} - \vec{S}_t)$ where $\delta(\vec{x})$ is the multi-dimensional Dirac delta function. Since \vec{S} is non-random $\mathbf{J}_{\vec{D}}$ is identically zero. Consequently $(\mathbf{J}_{\vec{D}})_{ij} = \delta_{ij}/(\bar{D}_i + I^d)$ whereby

$$|\mathbf{J}_S| = |\mathbf{T}|^2 \prod_{i=1}^4 \frac{1}{D_i + I^d} \quad (27)$$

If the additive noise term I^d is zero we see that we can obtain infinite information if $D_i = 0$ on a single detector, corresponding to one arm of the polarimeter projecting the incident polarisation

on to the basis state $\vec{T}_i \propto (S_{i0}, -S_{i1}, -S_{i2}, -S_{i3})$. Note the parallel with conventional matched filters whereby the filter corresponds to the template reversed in time. This result can be understood by noting that for a given state of polarisation there are only two PSA configurations capable of uniquely identifying that state, namely $\vec{T}_i \propto (S_{i0}, \pm S_{i1}, \pm S_{i2}, \pm S_{i3})$ corresponding to diametrically opposite points on the Poincaré sphere. For example only a horizontal or vertical polariser can unambiguously identify horizontally polarised light (giving a maximum or null intensity respectively). When taking a single measurement it is not in general possible to know which intensity level corresponds to the maximum, whilst a null intensity is clearly identifiable. Furthermore we have assumed an underlying Poisson process in which noise variations grow as the intensity grows and hence lower intensities give a better precision.

If present, a depolarised background necessitates a second, distinct polarimeter arm and also results in finite information. The situation is similar for non-zero I^d . Additional polarimeter arms improve estimation precision as discussed earlier in Section 3.3. Although unnecessary, we maintain our assertion that $N_D = 4$ to allow easy comparison. Once more we find there are an infinite number of possible instrument matrices that give rise to a maximum in the information, since it is possible to trade off precision in the intensity measurements (corresponding to higher light levels) with a reduction in the noise amplification associated with data processing i.e. smaller condition number. Three possible polarimeter configurations are shown in Fig. 3(a) for a template Stokes vector of $(1, 0, 0, 1)$. The first configuration (shown in green) gives the best condition number possible for a matched polarimeter (and consequently worse measurement precision), whilst the second (red) shows the opposite case, whereby the volume of the inscribed tetrahedron is significantly smaller i.e. larger condition number, yet the precision of measurement is increased since the total detected intensity is smaller. Practically this arrangement is unsuitable since it is highly sensitive to alignment errors in the PSA. The third configuration (blue) illustrates a more general arrangement.

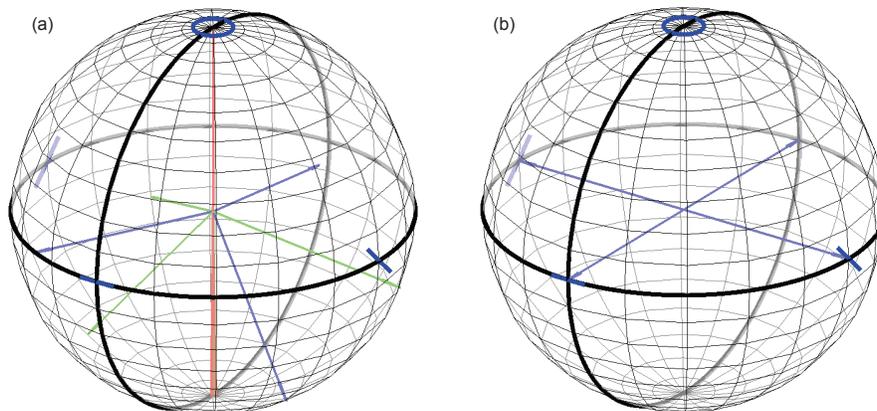


Fig. 3. Poincaré sphere showing possible polarimeter configurations (a) for a Stokes polarimeter matched to the template Stokes vector $(1, 0, 0, 1)$ i.e. left circularly polarised light and (b) a linear polarimeter, assuming the ratio $S_0/I^d = 10^5$. Each arrow denotes the basis Stokes vector of a polarimeter arm.

5.3. Linear polarimeter

Our final example assumes the polarisation incident into a Stokes polarimeter is restricted to lie on the equator of the Poincaré sphere. This could for example correspond to studying the light

from a family of polarisers. Such a model could be useful in polarisation multiplexed optical data storage [43, 44]. Using the PDF $f(\varepsilon, \theta) = \delta(\varepsilon)/\pi$ where ε and θ are angles which define a position on the Poincaré sphere [4] the expectations can be evaluated analytically to give

$$E \left[\frac{1}{D_i + I^d} \right] = \frac{1}{\sqrt{(S_0 + 2I^d)^2 + S_0^2 P^2 \cos^2 2\alpha_i}} \quad (28)$$

where α_i is the equatorial angle on the Poincaré sphere for the i^{th} basis Stokes vector of the instrument matrix. Considering we know *a priori* that the incident polarisation state is linearly polarised we have no need to estimate S_3 since this only describes the ellipticity of the light and can hence treat it as a nuisance parameter as described in Section 4.2. A linear Stokes polarimeter is thus optimised when

$$|\mathbf{J}_{S'}| = |\mathbf{T}|^2 \frac{\prod_{j=1}^4 [(S_0 + 2I^d)^2 + S_0^2 P^2 \cos^2 2\alpha_j]^{-1/2}}{\sum_{j=1}^4 \sin^2(2\alpha_j) [(S_0 + 2I^d)^2 + S_0^2 P^2 \cos^2 2\alpha_j]^{-1/2}} \quad (29)$$

is maximum, where S' denotes the parameter vector (S_0, S_1, S_2) . In agreement with [6] the maximum of this metric occurs when the measurement basis Stokes vectors \vec{T}_i are equally spaced around the equator of the Poincaré sphere as shown in Fig. 3(b). When applied to Mueller polarimeters a similar analysis shows the optimal input polarisation states also be equally spaced about the equator of the Poincaré sphere, although their position need bear no resemblance to those defined by \vec{T}_i .

6. Extension of optimisation results

The above results with regards to the optimisation of polarimeters holds not only for inference of the Stokes parameters or elements of a Mueller matrix but can be further extended for inference of parameters \vec{z} from these quantities. Such a situation may arise when performing a Lu-Chipman polar decomposition [14] on a measured Mueller matrix as is heavily used in the literature [45–47]. Eqs. (17) generalise to

$$\mathbf{J}_z = \frac{\partial \vec{S}^T}{\partial \vec{z}} \mathbf{T}^T \mathbf{J}_D \mathbf{T} \frac{\partial \vec{S}}{\partial \vec{z}} \quad (30a)$$

$$\mathbf{J}_z = \frac{\partial \mathbf{M}^T}{\partial \vec{z}} (\mathbf{R} \otimes \mathbf{T}^T) \mathbf{J}_D (\mathbf{R}^T \otimes \mathbf{T}) \frac{\partial \mathbf{M}}{\partial \vec{z}} \quad (30b)$$

Accordingly the volume of the ellipsoid of concentration as found from $|\mathbf{J}_z|$ is modified by a factor of $|\partial \vec{w} / \partial \vec{z}|^2$ ($\vec{w} = \vec{S}$ or \mathbf{M}), which is independent of both \mathbf{T} and \mathbf{R} . Its significance in terms of optimisation with respect to the experimental setup is thus null hence optimisation of these more complicated inference problems reduces to the optimisation procedure previously discussed.

Although in this work we discuss optimisation when the system is limited by quantum noise, the same technique can be used for alternative noise models. Adopting different noise models only requires a change in the deterministic FIM \mathbf{J}_D^{nr} . For example the FIM when assuming a Gaussian noise model is $\mathbf{J}_D^{nr} = \mathbf{K}^{-1}$.

7. Noise propagation in Mueller matrix polar decomposition

7.1. Single element systems

Noise propagation in inference problems, that is to say how noise in experimental data manifests itself as errors in the parameters of interest, can also be considered by employing Eqs.

(30). Although the mathematics is generally complicated we give here a result pertaining to polar decomposition of Mueller matrices [14]. Before considering the composite systems for which polar decomposition is relevant we must first describe noise propagation for single polarisation element systems, namely pure diattenuators, retarders and depolarisers.

A diattenuator is a non-depolarising polarisation element which preferentially transmits particular states of polarisation and has a Mueller matrix of the form [14]

$$\mathbf{M}_A = T_u \begin{pmatrix} 1 & \vec{A}^T \\ \vec{A} & \mathbf{m}_A \end{pmatrix} \quad (31)$$

where T_u is the transmittance for unpolarised light, $\vec{A} = (A_1, A_2, A_3)$ is the diattenuation vector whose magnitude A is known as the diattenuation and

$$\mathbf{m}_A = \sqrt{1-A^2} \mathbf{I} + (1 - \sqrt{1-A^2}) \frac{\vec{A}\vec{A}^T}{A^2} \quad (32)$$

\mathbf{I} is the 3×3 identity matrix. Any decomposition algorithm will need to estimate all four unknown parameters $\vec{z} = (A_1, A_2, A_3, T_u)$. Lengthy calculations give the derivatives required for evaluation of Eq. (30b) as

$$\frac{\partial \mathbf{M}_A}{\partial A_k} = T_u \begin{pmatrix} 0 & \vec{\delta}_k^T \\ \vec{\delta}_k & \frac{\partial \mathbf{m}_A}{\partial A_k} \end{pmatrix}, \quad \frac{\partial \mathbf{M}_A}{\partial T_u} = \frac{\mathbf{M}_A}{T_u} \quad (33)$$

where $k = 1, 2$ or 3 , $\vec{\delta}_k^T = (\delta_{1k}, \delta_{2k}, \delta_{3k})$ and

$$\frac{\partial m_{Aij}}{\partial A_k} = \left[\frac{A_i \delta_{jk}}{A^2} + \frac{A_j \delta_{ik}}{A^2} \right] \left[1 - \sqrt{1-A^2} \right] + \frac{A_k \delta_{ij}}{\sqrt{1-A^2}} - \frac{A_i A_j A_k}{A^4} \left[2 + \frac{A^2 - 2}{\sqrt{1-A^2}} \right] \quad (34)$$

where m_{Aij} is the $(i, j)^{\text{th}}$ element of \mathbf{m}_A . Using the FIM, \mathbf{J}_A , as calculated from Eqs. (30)-(34) and the CRLB we can calculate the best obtainable precision for estimation of the diattenuation parameters. The error on each parameter will in general be different, a point considered further in [48, 49].

Similarly we can consider a pure retarder which has a Mueller matrix of the general form

$$\mathbf{M}_R = \begin{pmatrix} 1 & \vec{0}^T \\ \vec{0} & \mathbf{m}_R \end{pmatrix} \quad (35)$$

where

$$m_{Rij} = \delta_{ij} \cos R + \frac{R_i R_j}{R^2} (1 - \cos R) + \sum_{q=1}^3 \varepsilon_{ijq} \frac{R_q}{R} \sin R \quad (36)$$

Again (R_1, R_2, R_3) defines a retardance axis and has a norm of R known as the retardance and ε_{ijq} is the Levi-Civita permutation symbol. Calculation of the FIM, \mathbf{J}_R , requires the derivatives

$$\begin{aligned} \frac{\partial m_{Rij}}{\partial R_k} = & \left[\frac{R_i \delta_{jk}}{R^2} + \frac{R_j \delta_{ik}}{R^2} \right] (1 - \cos R) - \frac{R_k \delta_{ij}}{R} \sin R - \frac{R_i R_j R_k}{R^2} \left[\frac{1 - \cos R}{R} + \sin R \right] \\ & + \sum_{q=1}^3 \frac{\varepsilon_{ijq}}{R^2} \left[R_q R_k \cos R + \left(R \delta_{qk} - \frac{R_k}{R} \right) \sin R \right] \end{aligned} \quad (37)$$

The case of a depolariser is however much more difficult to tackle since in general an eigen-analysis of the system is required to find the pertinent depolarisation parameters. This can not be described analytically except in some special cases. For example if it were known *a priori* that the sample were a pure depolariser with Mueller matrix of the form

$$\mathbf{M}_\Delta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ 0 & 0 & b & 0 \\ 0 & 0 & 0 & c \end{pmatrix}, \quad |a|, |b|, |c| \leq 1 \quad (38)$$

where $1 - |a|$, $1 - |b|$ and $1 - |c|$ are the principal depolarisation factors we can easily calculate the derivatives

$$\frac{\partial M_{\Delta ij}}{\partial a} = \delta_{i2}\delta_{j2}, \quad \frac{\partial M_{\Delta ij}}{\partial b} = \delta_{i3}\delta_{j3}, \quad \frac{\partial M_{\Delta ij}}{\partial c} = \delta_{i4}\delta_{j4}, \quad (39)$$

where $M_{\Delta ij}$ is the $(i, j)^{\text{th}}$ element of \mathbf{M}_Δ . The appropriate FIM \mathbf{J}_Δ is then given by substituting Eqs. (39) into Eq. (30b).

7.2. Composite systems

The single element results discussed above can be used for noise analysis when the experimenter has *a priori* knowledge about the structure of the Mueller matrix. If however this is not the case a Lu-Chipman decomposition is frequently performed so as to parameterise the sample. Fundamental to the Lu-Chipman decomposition is the fact that an arbitrary Mueller matrix can be written as the product of three distinct Mueller matrices corresponding to a depolariser, retarder and diattenuator i.e. $\mathbf{M} = \mathbf{M}_\Delta \mathbf{M}_R \mathbf{M}_A$. Morio and Goudail [50] considered the importance of altering the order in which the product is evaluated and found that different decompositions gave either unphysical results or merely comprised of an appropriate rotation compared to the Lu-Chipman decomposition and thus is only a mathematical, not physical, difference. With these results in mind we adhere to the original formulation, since this ensures physicality and furthermore corresponds to common usage.

Calculation of the FIM for a Lu-Chipman decomposition can be achieved by application of the product rule to Eq. (30b) which yields

$$\begin{aligned} \mathbf{J}_z &= \frac{\partial \mathbf{M}_A^T}{\partial \vec{z}} (\mathbf{R} \otimes \mathbf{M}_R^T \mathbf{M}_\Delta^T \mathbf{T}^T) \mathbf{J}_D (\mathbf{R}^T \otimes \mathbf{T} \mathbf{M}_\Delta \mathbf{M}_R) \frac{\partial \mathbf{M}_A}{\partial \vec{z}} \\ &+ \frac{\partial \mathbf{M}_R^T}{\partial \vec{z}} (\mathbf{M}_A \mathbf{R} \otimes \mathbf{M}_\Delta^T \mathbf{T}^T) \mathbf{J}_D (\mathbf{R}^T \mathbf{M}_A^T \otimes \mathbf{T} \mathbf{M}_\Delta) \frac{\partial \mathbf{M}_R}{\partial \vec{z}} \\ &+ \frac{\partial \mathbf{M}_\Delta^T}{\partial \vec{z}} (\mathbf{M}_R \mathbf{M}_A \mathbf{R} \otimes \mathbf{T}^T) \mathbf{J}_D (\mathbf{R}^T \mathbf{M}_A^T \mathbf{M}_R^T \otimes \mathbf{T}) \frac{\partial \mathbf{M}_\Delta}{\partial \vec{z}} \end{aligned} \quad (40)$$

since the structure of the matrices dictates that the cross terms are identically zero. It is important to note that we stack the parameters of interest into a single parameter vector for example $\vec{z} = (R_1, R_2, \dots, b, c)$. \mathbf{J}_z is then block diagonal

$$\mathbf{J}_z = \begin{pmatrix} \mathbf{J}_R & \mathbf{0}^T & \mathbf{0}^T \\ \mathbf{0} & \mathbf{J}_A & \mathbf{0}^T \\ \mathbf{0} & \mathbf{0} & \mathbf{J}_\Delta \end{pmatrix} \quad (41)$$

where the order of the diagonal terms depends only on the ordering of the parameters in \vec{z} . Mathematically the FIMs \mathbf{J}_R , \mathbf{J}_A and \mathbf{J}_Δ are of the same form as the single element FIMs described in the previous section albeit for a slight modification in the effective input polarisation

states and instrument matrix respectively as can be seen by comparing Eqs. (30b) and (40). Fortunately this makes physical sense considering the Mueller matrix polar decomposition models the system as a cascade of three independent polarisation elements. Once more we give the cautionary note that the derivatives required to calculate \mathbf{J}_Δ can not be found analytically in general.

8. Conclusions

In this paper we have addressed the question as to how *a priori* information we may possess about a system being measured can be used to improve the precision of our measurements. In doing so we used a metric based upon the FIM which takes a holistic approach by including the raw information obtained from a measurement and the noise amplification that occurs during data processing. We have shown the conditions under which our figure of merit is equivalent to the frequently used condition number of the instrument (and incident polarisation) matrix, namely maximal ignorance of the likely values of the parameters of interest. We then proceeded to show that when incorporating *a priori* knowledge to achieve a resolution improvement the condition number is unsatisfactory and does not give optimal results. This was illustrated by considering the polarimetric equivalent of a matched filter and linear polarimeters.

Calculation of the informational figure of merit requires calculation of a FIM, which is also beneficial since it provides a simple description of the noise propagation that may arise during data processing and how the resulting errors are distributed among the parameter values. Specifically we have given results pertaining to a Lu-Chipman polar decomposition since this is frequently used in polarimetric analysis.

Although formulated in terms of estimation of Stokes parameters and Mueller matrix elements in the presence of Poisson noise, optimisation with respect to Fisher information is easily extended to inference of alternative sample parameters and noise models and is thus applicable to a wide variety of optical experiments.

Finally we acknowledge the financial support of the EU via NANOPRIM Contract No. NMP3-CT-2007-033310.

A. Fisher information for random parameters

For a deterministic parameter the FIM is defined by Eq. (8) where $f(\mathbf{D}|\vec{w})$ is the PDF of \mathbf{D} conditioned on the value of \vec{w} . If however \vec{w} can vary this definition becomes unsatisfactory since it does not account for our knowledge about the random nature of the parameter, which can be used to improve the precision of any measurement. Instead it is more appropriate to define the FIM in terms of the joint PDF of \mathbf{D} and \vec{w} , namely $f(\mathbf{D}, \vec{w}) = f(\mathbf{D}|\vec{w})f(\vec{w})$. Taking the logarithm gives

$$L(\mathbf{D}, \vec{w}) = \ln f(\mathbf{D}|\vec{w}) + \ln f(\vec{w}) \quad (42)$$

so that the modified FIM is defined by

$$\begin{aligned} \mathbf{J}_w &= \sum_{i=1}^2 \sum_{j=1}^2 E_{D,w} \left[\frac{\partial L_i}{\partial \vec{w}} \frac{\partial L_j}{\partial \vec{w}}^T \right] \\ &= \mathbf{J}_1 + \mathbf{J}_2 + \mathbf{J}_3 + \mathbf{J}_4 \end{aligned} \quad (43)$$

where $L_1(\mathbf{D}, \vec{w}) = \ln f(\mathbf{D}|\vec{w})$ and $L_2(\vec{w}) = \ln f(\vec{w})$, and expectations are now with respect to both \mathbf{D} and \vec{w} . Considering each of these terms in turn we have

$$\begin{aligned} \mathbf{J}_1 &= \iint \frac{\partial L_1}{\partial \vec{w}} \frac{\partial L_1^T}{\partial \vec{w}} f(\mathbf{D}, \vec{w}) d\mathbf{D} d\vec{w} \\ &= \int \left[\int \frac{\partial L_1}{\partial \vec{w}} \frac{\partial L_1^T}{\partial \vec{w}} f(\mathbf{D}|\vec{w}) d\mathbf{D} \right] f(\vec{w}) d\vec{w} \\ &= E_w [\mathbf{J}_w^{nr}] \end{aligned} \quad (44)$$

where $E_w[\dots]$ denotes the expectation with respect to \vec{w} only and \mathbf{J}_w^{nr} is the deterministic (or non-random) Fisher information matrix as defined by Eq. (8).

Adopting a similar treatment of \mathbf{J}_2 we have

$$\begin{aligned} \mathbf{J}_2 &= \iint \frac{\partial L_1}{\partial \vec{w}} \frac{\partial L_2^T}{\partial \vec{w}} f(\mathbf{D}, \vec{w}) d\mathbf{D} d\vec{w} \\ &= \iint \frac{f(\mathbf{D}, \vec{w})}{f(\mathbf{D}|\vec{w})f(\vec{w})} \frac{\partial f(\mathbf{D}|\vec{w})}{\partial \vec{w}} \frac{\partial f(\vec{w})^T}{\partial \vec{w}} d\mathbf{D} d\vec{w} \\ &= \int \left[\frac{\partial}{\partial \vec{w}} \int f(\mathbf{D}|\vec{w}) d\mathbf{D} \right] \frac{\partial f(\vec{w})^T}{\partial \vec{w}} d\vec{w} \\ &= 0 \end{aligned} \quad (45)$$

where the last step follows from $\int f(\mathbf{D}|\vec{w}) d\mathbf{D} = 1$. A similar result follows for \mathbf{J}_3 . Finally consider

$$\begin{aligned} \mathbf{J}_4 &= \iint \frac{\partial L_2}{\partial \vec{w}} \frac{\partial L_2^T}{\partial \vec{w}} f(\mathbf{D}, \vec{w}) d\mathbf{D} d\vec{w} \\ &= \int \frac{\partial L_2}{\partial \vec{w}} \frac{\partial L_2^T}{\partial \vec{w}} f(\vec{w}) d\vec{w} = \mathbf{J}_w^r \end{aligned} \quad (46)$$

Combining these results we find

$$\mathbf{J}_w = E_w [\mathbf{J}_w^{nr}] + \mathbf{J}_w^r \quad (47)$$

We thus see that the Fisher information matrix when trying to estimate a random parameter \vec{w} is given by the average of the deterministic Fisher information with respect to \vec{w} plus an additional term arising from our knowledge of the random behaviour of \vec{w} .

As a final point of interest we note here that if $f(\vec{w})$ is uniform then

$$\frac{\partial \ln f(\vec{w})}{\partial \vec{w}} = 0 \quad (48)$$

such that $\mathbf{J}_w^r = 0$ and $\mathbf{J}_w = E_w [\mathbf{J}_w^{nr}]$. Furthermore if \mathbf{J}_w^{nr} is *not* dependent on \vec{w} then $\mathbf{J}_w = \mathbf{J}_w^{nr}$