On the equivalence of optimization metrics in Stokes polarimetry

Matthew R Foreman, François Goudail

On the equivalence of optimization metrics in Stokes polarimetry

Matthew R. Foreman[a,*], **François Goudail**[b]
[a]Blackett Laboratory, Department of Physics, Imperial College London, Prince Consort Road, London, SW7 2AZ, UK
[b]Laboratoire Charles Fabry, Institut d'Optique Graduate School, CNRS, Université Paris-Saclay, 91127 Palaiseau cedex, France

Abstract. Optimization of polarimeters has historically been achieved using an assortment of performance metrics. Selection of an optimization parameter is frequently made, however, on an ad hoc basis. In this article we rigorously demonstrate that optimization strategies in Stokes polarimetry based on three common metrics; namely the condition number of the instrument matrix, the determinant of the associated Gram matrix, or the equally weighted variance, are formally equivalent. In particular, using each metric we derive the same set of constraints on the measurement states, and show that these can be satisfied using spherical 2 designs. Our conclusions are equally applicable to optimization of the illumination states in Mueller matrix polarimetry.

Keywords: polarimetry, optimization, spherical design, instrument matrix, equally weighted variance.

*Matthew R. Foreman, matthew.foreman@imperial.ac.uk

1 Introduction

Quantitative analysis of the state of polarization of light provides a powerful tool in modern science. Applications vary from microscopy, biomedical diagnosis and astrophysics[1–3] to crystallographic, material and single molecule studies.[4,5] Whilst the polarization state of light itself, can be used to transmit information, hence presenting new opportunities in optical data storage and communications,[6–9] changes in polarization induced by a material can alternatively be used for object detection[10] or to characterize sample properties, such as chirality or molecular orientation.[11–13]

Stokes polarimeters, which allow a complete characterization of the polarization state of input light as described by the associated 4×1 Stokes vector $\mathbf{S}$, comprise of N (≥ 4) distinct measurements that can be multiplexed in, for example, time,[14] frequency[15] or space.[16] Fundamentally, each constituent measurement outputs an intensity I_j ($j \in [1, N]$) which is proportional to the projection of the incident Stokes vector onto an analysis state $\mathbf{W}_j$, i.e., $I_j = \mathbf{W}_j^T \mathbf{S}$. Central to

the description and design of Stokes polarimeters is hence the so-called instrument or measurement matrix $\mathbb{W} = [\mathbf{W}_1, \mathbf{W}_2, \ldots]^T$ formed from stacking the set of analysis vectors. In order to obtain an estimate of the Stokes vector from the set of projections I_j, the measurement matrix must be inverted. So as to limit noise propagation through this inversion process, optimization of the measurement matrix is hence frequently performed. Optimization in this vein has been performed using different metrics including the associated information content,[17–19] matrix determinant,[20–22] signal to noise ratio,[23] equally weighted variance[24,25] and condition number.[21–23,25–29]

Mueller matrix polarimeters, on the other hand, combine a Stokes polarimeter with use of multiple incident polarized states so as to measure the full Mueller matrix of an object. Variation of the probing polarization states (as can be described using an analogous illumination matrix) therefore introduces additional degrees of freedom, hence admitting further optimization.[17,28–32] Application specific optimization of polarimeters has also been reported, for example, in detection and imaging problems the polarization contrast is a more suitable metric.[33,34]

Recently, the equivalence of a number of optimization metrics, namely the equally weighted variance, the condition number of $\mathbb{W}$ and the determinant of the associated Gram matrix, was discussed by Foreman et al.[35] Additionally, Foreman et al. proved that a Stokes polarimeter is optimal (as characterized by these metrics) when the set of analysis states defines a spherical 2 design[36] on the unit Poincaré sphere. A re-examination of the equivalence between these metrics is, however, necessary due to an error in the proof presented in Ref. 35. The goal of this article is therefore to provide a rigorous proof that the conclusions of Ref. 35 hold. Our derivations also elicit greater insight into the optimization of Stokes polarimeters. We additionally note that our results are equally applicable to optimization of the probing states used in Mueller matrix polarimetry due to the similar matrix structure of the problem.[31,37]

2 Optimal polarimetry with spherical 2-designs

The instrument matrix, $\mathbb{W}$, of a polarimeter is an $N \times 4$ matrix, the rows of which are the Stokes vectors of the N polarization states being analyzed, normalized such that the polarimeter is passive. Accordingly the instrument matrix has the parametric form

$$\mathbb{W} = \frac{1}{2} \begin{bmatrix} 1 & \mathbf{w}_1^T \\ 1 & \mathbf{w}_2^T \\ \vdots & \vdots \\ 1 & \mathbf{w}_N^T \end{bmatrix} \triangleq \frac{1}{2} \begin{bmatrix} \mathbf{r} & \mathbb{Q} \end{bmatrix}, \tag{1}$$

where $\mathbf{r}$ is an $N \times 1$ vector of ones and $\mathbb{Q}$ is the matrix formed from the 3×1 vectors $\mathbf{w}_j$ ($j \in [1, N]$) of unit norm. Note that throughout this work bold notation is used to signify column vectors whilst blackboard font denotes matrices. Please also note that we have assumed an "ideal" instrument matrix, in the sense that the transmittance and degree of polarization of all the rows are equal to one. The generalization of the obtained results to arbitrary instrumentation matrices will be discussed in Section 4.

In Stokes polarimetry, one performs N intensity measurements $I_j, j \in [1, N]$ by projecting the input Stokes vector $\mathbf{S}$ onto each of the N analyzers described by the N rows of the matrix $\mathbb{W}$. If these measurements are stacked in an N-dimensional vector $\mathbf{I} = [I_1, I_2, \ldots, I_N]^T$, and if we assume that the measurements are perturbed by white additive noise, we obtain

$$\mathbf{I} = \mathbb{W}\mathbf{S} + \mathbf{B}, \tag{2}$$

where $\mathbf{B}$ is an $N \times 1$ random vector with covariance matrix $\sigma^2 \mathbb{I}_N$ and $\mathbb{I}_n$ denotes the $n \times n$ identity

matrix. The maximum-likelihood estimate of $\mathbf{S}$ is obtained by

$$\widehat{\mathbf{S}} = \mathbb{W}^{+}\mathbf{I} \tag{3}$$

where

$$\mathbb{W}^{+} = (\mathbb{W}^{T}\mathbb{W})^{-1}\mathbb{W}^{T} \tag{4}$$

denotes the pseudo-inverse matrix. The estimate $\widehat{\mathbf{S}}$ is a random vector of mean $\mathbf{S}$ (i.e., the estimator is unbiased), and of covariance matrix[17,23,24]

$$\mathbb{K}_{\mathbf{S}} = \sigma^{2}(\mathbb{W}^{T}\mathbb{W})^{-1}. \tag{5}$$

The estimation variances of each element of the Stokes vector estimate are the diagonal elements of this matrix. A natural goal of polarimeter optimization is to find the measurement matrix $\mathbb{W}$ that minimizes the sum of these variances, that is, the trace of $\mathbb{K}_{\mathbf{S}}$. The corresponding performance metric is called the equally weighted variance (EWV), i.e.,

$$\text{EWV} = \sigma^{2}\text{tr}[\mathbb{G}^{-1}] \tag{6}$$

where

$$\mathbb{G} = \mathbb{W}^{T}\mathbb{W} \tag{7}$$

denotes the Gram matrix associated with $\mathbb{W}$.

To optimize the EWV, we first express the Gram matrix $\mathbb{G}$ in block format, viz.

$$\mathbb{G} = \frac{1}{4}\begin{bmatrix} N & \mathbf{r}^T\mathbb{Q} \\ \mathbb{Q}^T\mathbf{r} & \mathbb{Q}^T\mathbb{Q} \end{bmatrix} \triangleq \begin{bmatrix} A & \mathbf{B}^T \\ \mathbf{C} & \mathbb{D} \end{bmatrix}. \tag{8}$$

The inverse of the Gram matrix can then be expressed in the form[38]

$$\mathbb{G}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}\mathbf{B}^T\mathbb{M}^{-1}\mathbf{C}A^{-1} & -A^{-1}\mathbf{B}^T\mathbb{M}^{-1} \\ -\mathbb{M}^{-1}\mathbf{C}A^{-1} & \mathbb{M}^{-1} \end{bmatrix}, \tag{9}$$

where the matrix

$$\mathbb{M} = (\mathbb{D} - \mathbf{C}A^{-1}\mathbf{B}^T) \tag{10}$$

is the Schur complement of the upper left block of $\mathbb{G}$. This implies that the trace we seek can be written as

$$\text{tr}[\mathbb{G}^{-1}] = A^{-1} + A^{-1}\mathbf{B}^T\mathbb{M}^{-1}\mathbf{C}A^{-1} + \text{tr}[\mathbb{M}^{-1}]. \tag{11}$$

Substituting Eq. (8) into Eq. (10), the Schur complement takes the form:

$$\mathbb{M} = \frac{1}{4}\left(\mathbb{Q}^T\mathbb{Q} - \frac{\mathbf{q}\mathbf{q}^T}{N}\right) \tag{12}$$

where $\mathbf{q} = \mathbb{Q}^T\mathbf{r}$ is an N-dimensional vector. Upon using the identity[38]

$$(\mathbb{Z} + \mathbf{x}\mathbf{y}^T)^{-1} = \mathbb{Z}^{-1} - \frac{\mathbb{Z}^{-1}\mathbf{x}\mathbf{y}^T\mathbb{Z}^{-1}}{1 + \mathbf{y}^T\mathbb{Z}^{-1}\mathbf{x}} \tag{13}$$

with $\mathbf{x} = -\mathbf{y} = \mathbf{q}/\sqrt{N}$ and $\mathbb{Z} = \mathbb{Q}^T\mathbb{Q}$, we find

$$\mathbb{M}^{-1} = 4\,(\mathbb{Q}^T\mathbb{Q})^{-1} + 4\,\frac{(\mathbb{Q}^T\mathbb{Q})^{-1}\mathbf{q}\mathbf{q}^T(\mathbb{Q}^T\mathbb{Q})^{-1}}{N - \mathbf{q}^T(\mathbb{Q}^T\mathbb{Q})^{-1}\mathbf{q}}. \tag{14}$$

Direct substitution from Eqs. (8) and (14) into Eq. (11) yields

$$\begin{aligned}\mathrm{tr}[\mathbb{G}^{-1}] = 4\Bigg\{&\frac{1}{N} + \mathrm{tr}\left[(\mathbb{Q}^T\mathbb{Q})^{-1}\right] \\ &+ \frac{\mathbf{q}^T[N(\mathbb{Q}^T\mathbb{Q})^{-2} + (\mathbb{Q}^T\mathbb{Q})^{-1}]\mathbf{q}}{N[N - \mathbf{q}^T(\mathbb{Q}^T\mathbb{Q})^{-1}\mathbf{q}]}\Bigg\},\end{aligned} \tag{15}$$

where we have also used the cyclic property of the trace operation and the identity $\mathrm{tr}[\mathbb{X}\mathbf{q}^T\mathbf{q}] = \mathbf{q}^T\mathbb{X}\mathbf{q}$ for arbitrary $\mathbb{X}$.[38]

Noting that $N > 0$ and that $\mathbb{Q}^T\mathbb{Q}$ is positive definite, it follows immediately that the first two terms in Eq. (15) are positive. We show in Appendix A that the third term is also positive. Consequently, the trace in Eq. (15) is minimal when its three terms are minimal. The first term is constant, and the third is minimal when it is null, that is, when $\mathbf{q} = \mathbb{Q}^T\mathbf{r} = \mathbf{0}$ or equivalently

$$\sum_{n=1}^{N} \mathbf{w}_n = \mathbf{0}. \tag{16}$$

Importantly, Eq. (16) expresses a polynomial constraint that must be satisfied by an optimal measurement matrix, and is equivalent to that given in Eq. (4) of Ref. 35.

When Eq. (16) holds, minimizing $\mathrm{tr}[\mathbb{G}^{-1}]$ is equivalent to minimizing $\mathrm{tr}\left[(\mathbb{Q}^T\mathbb{Q})^{-1}\right]$. This optimization has to be done under the constraint that the trace of the matrix $\mathbb{Q}^T\mathbb{Q}$ is constant as follows from the normalization of $\mathbf{w}_j$. Indeed, since each row of the matrix $\mathbb{Q}$ is a unit-norm vector, we have

$$\mathrm{tr}[\mathbb{Q}^T\mathbb{Q}] = \mathrm{tr}[\mathbb{Q}\mathbb{Q}^T] = N. \tag{17}$$

We thus have to solve the following constrained optimization problem, set in Lagrange form

$$\Psi(\mathbb{Q}) = \mathrm{tr}[(\mathbb{Q}^T\mathbb{Q})^{-1}] - \mu\left(\mathrm{tr}[\mathbb{Q}^T\mathbb{Q}] - N\right) \tag{18}$$

where μ is a Lagrange multiplier. The Lagrange function can also be expressed as

$$\Psi(\boldsymbol{\beta}) = \sum_{j=1}^{3}\frac{1}{\beta_j} + \mu\left(\sum_{j=1}^{3}\beta_j - N\right) \tag{19}$$

where β_j, $j \in [1,3]$, are the positive eigenvalues of the matrix $\mathbb{Q}^T\mathbb{Q}$. Equating the gradient of Eq. (19) with respect to $\boldsymbol{\beta}$ to zero and enforcing the constraint ($\partial\Psi/\partial\mu = 0$) yields $\beta_j = 1/\sqrt{\mu} = N/3$ for all $j \in [1,3]$, such that

$$\mathbb{Q}^T\mathbb{Q} = \sum_{j=1}^{m}\mathbf{w}_j\mathbf{w}_j^T = \frac{N}{3}\mathbb{I}_3. \tag{20}$$

Eq. (20) is the second set of polynomial constraints derived in Ref. 35. The form of the Gram

matrix $\mathbb{G}$ which hence minimizes the EWV of the instrument matrix is thus

$$\mathbb{G} = \mathbb{W}^T\mathbb{W} = \frac{N}{12}\begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{21}$$

According to Eq. (5), the corresponding covariance of the Stokes vector estimate is hence:

$$\mathbb{K}_{\mathbf{S}} = \frac{4}{N}\sigma^2\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}. \tag{22}$$

This result is important since it specifies, in a very simple closed-from, the fundamental limit of the estimation variance that can be reached by a Stokes polarimeter with a given number of measurement vectors in the presence of additive noise. For example, we note that the minimum achievable variance on an estimate of the intensity (i.e. the first element of the Stokes vector) is three times better than that on the other Stokes parameters. Moreover, the covariance matrix is diagonal, which means that the fluctuations of each element of the Stokes vector estimator are statistically independent. This property is important when performing theoretical computations involving Stokes vector estimators. Incidently, we note that the minimum value of the equally weighted variance is EWV $= 40\sigma^2/N$.

Finally, we can show that the conditions expressed by Eqs. (16) and (20) are satisfied when

the set of measurement states on the Poincaré sphere, defined by $\{\mathbf{w}_j\}$, $j \in [1, N]$, constitute a spherical 2 design (see Appendix B) as reported in Ref. 35. A spherical t-design is defined as a collection of N points $\{\mathbf{w}_j\}$ on the surface of the unit sphere (in our case in $\mathcal{R}^3$) for which the normalized integral of any polynomial function, $f(\mathbf{w})$, of degree t or less is equal to the average taken over the N points. The Platonic solids, i.e., the regular tetrahedron ($N = 4$), the octahedron ($N = 6$), the cube ($N = 8$), the icosahedron ($N = 12$), and the dodecahedron ($N = 20$), are spherical 2 designs, as well as many other measurement frames illustrated in.[35]

3 Equivalence of optimization metrics

We will now demonstrate that the optimization of two other popular metrics, namely the condition number and the determinant of the Gram matrix, lead to exactly the same measurement frames as the EWV, so that these three criteria are strictly equivalent.

3.1 Condition number

The condition number κ of the instrument matrix is defined by $\kappa = \|\mathbb{W}\|\|\mathbb{W}^+\|$ where $\mathbb{W}^+$ is the pseudo-inverse matrix and the Hilbert-Schmidt matrix norm[38] of any matrix $\mathbb{P}$ is defined as:

$$\|\mathbb{P}\| = (\text{tr}[\mathbb{P}^T\mathbb{P}])^{1/2} = (\text{tr}[\mathbb{P}\mathbb{P}^T])^{1/2} \tag{23}$$

Our choice of normalization of the measurement states $\mathbf{W}_j = [1, \mathbf{w}_j]^T/2$ implies that

$$\|\mathbb{W}\|^2 = \text{tr}[\mathbb{W}^T\mathbb{W}] = \frac{N}{2}. \tag{24}$$

Moreover, using the expression of the pseudo-inverse in Eq. (4), it is easily shown that

$$\|\mathbb{W}^+\|^2 = \text{tr}[(\mathbb{W}^+)^T\mathbb{W}^+] = \text{tr}[(\mathbb{W}^T\mathbb{W})^{-1}] = \frac{\text{EWV}}{\sigma^2}. \tag{25}$$

Consequently, one can write

$$\kappa = \frac{\sqrt{N}}{\sqrt{2}\sigma}\sqrt{\text{EWV}}. \tag{26}$$

For any measurement matrix verifying Eq. (1), the condition number is thus simply proportional to the square root of the EWV. It is thus obvious that minimizing the condition number is equivalent to minimizing the EWV.

3.2 Determinant of the Gram matrix

The first works on Stokes polarimeter optimization considered devices with a minimal number ($N = 4$) of measurement vectors.[26] Optimization of such systems used the determinant of the matrix $\mathbb{W}$ (which for this value of N is square and non singular) as a performance metric. In this case the optimal structure found dictated that the measurement vectors defined a regular tetrahedron on the Poincaré sphere, a result that we also found above by optimizing the EWV. We show in this section that this result comes from the strict equivalence of these two optimization metrics. This equivalence can be generalized to any value of N if one considers the optimization of the determinant of the Gram matrix $\mathbb{G}$, since for $N > 4$ the matrix $\mathbb{W}$ itself is rectangular and its determinant is thus not defined. Notice that this equivalence was mentioned in Ref. 35, but there was an erroneous step in the logic presented in that work (see appendix C for more detail).

We intend here to show that maximization of the determinant $|\mathbb{G}|$ yields the same polynomial

constraints embodied in Eqs. (16) and (20). Considering the block form of the Gram matrix in Eq. (8), its determinant can be written as[38]

$$\begin{aligned} |\mathbb{G}| &= |A - \mathbf{B}^T\mathbb{D}^{-1}\mathbf{C}||\mathbb{D}| \\ &= \frac{1}{256}\left[N - \mathbf{r}^T\mathbb{Q}(\mathbb{Q}^T\mathbb{Q})^{-1}\mathbb{Q}^T\mathbf{r}\right]|\mathbb{Q}^T\mathbb{Q}|. \end{aligned} \tag{27}$$

Maximizing this expression means maximizing the two factors appearing in the product. The first factor is maximized if the positive subtractive term is zero, that is to say when the vector $\mathbb{Q}^T\mathbf{r} = \mathbf{0}$, corresponding to the first polynomial constraint expressed in Eq. (16).

For the second factor, we note that $|\mathbb{Q}^T\mathbb{Q}| = \prod_{j=1}^{3}\beta_j$ where β_j, $j \in [1,3]$, are the eigenvalues of the matrix $\mathbb{Q}^T\mathbb{Q}$ which are positive since $\mathbb{Q}^T\mathbb{Q}$ is positive definite. Moreover, according to Eq. (17), the matrix $\mathbb{Q}^T\mathbb{Q}$ has constant trace. Maximization of $|\mathbb{Q}^T\mathbb{Q}|$ is thus once again a constrained optimization problem which can be solved using the method of Lagrange multipliers. We will consider maximization of $\ln|\mathbb{Q}^T\mathbb{Q}|$ since the logarithmic function is monotonically increasing. The Lagrange function then becomes

$$\Psi(\boldsymbol{\beta}) = \sum_{j=1}^{3}\ln\beta_j - \mu\left(\sum_{j=1}^{3}\beta_j - N\right). \tag{28}$$

Following the standard optimization procedure we find, similarly to Section 2, that $\beta_j = 1/\mu = N/3$ for all $j \in [1,2,3]$. As shown in Section 2, the second polynomial constraint expressed in Eq. (20) then follows. Therefore, we have ultimately shown that minimization of the EWV (and thus also of the condition number of the instrument matrix) of a polarimeter yields the same set of optimality constraints as maximizing the determinant of the associated Gram matrix.

4 Discussion

The main conclusion of the previous sections is the following: among all the measurement matrices verifying Eq. (1), the ones that maximize the condition number, the EWV and the determinant are exactly the same. It can be said that this results unifies many previous works on polarimeter optimization: for example, the early work of Azzam[26] (based on determinant), Ambirajan and Look[22] (based on condition number and determinant), Sabatke et al.[24] (based on EWV and determinant), and Tyo[43] (based on condition number), among many others.

Modeling of $\mathbb{W}$ with Eq. (1) implies that the transmittance and the degree of polarization of all rows of $\mathbb{W}$ are equal to one. This hypothesis is frequently done in polarimetry, but it is interesting to consider the case where it is not fulfilled. In the general case, each row of the measurement matrix may have a different transmission $t_i, i \in [1, N]$, and a different degree of polarization $P_i, i \in [1, N]$. In this case, the measurement matrix can be expressed in the following form:

$$\mathbb{W} = \frac{1}{2}\begin{bmatrix} t_1 & t_1 P_1 \mathbf{w}_1^T \\ t_2 & t_2 P_2 \mathbf{w}_2^T \\ \vdots & \vdots \\ t_N & t_N P_N \mathbf{w}_N^T \end{bmatrix} \triangleq \frac{1}{2}\begin{bmatrix} \mathbb{T}\mathbf{r} & \mathbb{T}\mathbb{P}\mathbb{Q} \end{bmatrix}, \tag{29}$$

where $\mathbb{T} = \text{diag}(t_1, \ldots, t_N)$ and $\mathbb{P} = \text{diag}(P_1, \ldots, P_n)$. In this general form, it is easy to demonstrate that

$$\|\mathbb{W}\|^2 = \frac{1}{4}\sum_{i=1}^{N}(1 + P_i^2)t_i^2 \tag{30}$$

Consequently, one can generalize Eq. (26) to:

$$\kappa = \frac{\sqrt{\sum_{i=1}^{N}(1+P_i^2)t_i^2}}{2\sigma}\sqrt{\text{EWV}} \tag{31}$$

This entails a generalization of the result obtained in Section 3: for arbitrary, but fixed values of transmissions and degree of polarization of the analysis vectors, optimization of the vectors' positions on the normalized Poincaré sphere (i.e., of the vectors $\mathbf{w}_n$) yields the same result if the condition number or the EWV are used as criteria. If, in addition, the transmissions and degree of polarizations are identical for all the N measurements, it is easy to show that the resulting optimal structures are spherical 2 designs. If they are not identical, the optimal structures are no longer the spherical designs. Determining the optimal structures in this case is an interesting perspective to the present work.

Another important practical question is which one of the three considered metrics is the most appropriate for evaluating the performance of any polarimeter, be it optimal or not. Indeed, on this point of view, the metrics are not equivalent. This is most easily seen by noticing that the three metrics only depend on the measurement matrix through the eigenvalues of $\mathbb{G}$, denoted $\mu_i, i \in [1,4]$, and have the following form:

$$|\mathbb{G}| = \prod_{i=1}^{4}\mu_i \tag{32}$$

$$\kappa = \left(\sum_{i=1}^{4}\mu_i\right)^{1/2}\left(\sum_{i=1}^{4}\frac{1}{\mu_i}\right)^{1/2} \tag{33}$$

$$\text{EWV} = \sum_{i=1}^{4}\frac{1}{\mu_i} \tag{34}$$

It is seen that although they look similar, they are different. In particular, two different sets of eigenvalues may lead to the same value of κ, but different values of $|\mathbb{G}|$ and EWV, and vice versa. The question of how to choose the best metric could seem arbitrary, but we think there is a strong objective argument in favor of EWV. Indeed, the EWV corresponds to an estimation variance, which has a clear and useful statistical meaning. For example, it makes it handier to compare two different polarimeter structures: saying that polarimeter A has a EWV twice than polarimeter B just means that the variance of the estimated Stokes vector is twice. In sharp contrast, a ratio between two determinants or two condition numbers is much less easy to interpret in terms of estimation error.

Another strong advantage of the EWV is that it can be used for polarimeter optimization in the presence of non-additive noises sources. The EWV has for example been used to derive the optimal measurement frames in the presence of Poisson shot noise.[39,40] In this case, the expression of the covariance matrix of the Stokes estimate was different from Eq.(5). Consequently, the EWV was different from Eq. (6), and was thus no longer proportional of the square of the condition number, which had no statistical meaning. Furthermore, when measurements are simultaneously affected by several types of statistically independent noise sources, the value of the EWV is simply the sum of the EWVs corresponding to each noise source. This property has for example been recently employed to characterize the actual performance of micro-grid based polarimetric cameras in the presence of both additive detection noise and Poisson shot noise.[41]

In conclusion, the message of the present work could be summarized as follows. When optimizing the estimation performance of a polarimeter in the presence of additive Gaussian noise, condition number, determinant, and EWV are three strictly equivalent metrics. When evaluating and comparing the performance of different polarimeters, or when optimizing polarimeters in the

presence of non-additive, non-Gaussian noise sources, the EWV has strong advantages compared with the two other metrics.

5 Conclusions

We have shown that optimization of the EWV, of the condition number, or of the determinant of the Gram matrix of a Stokes polarimeter leads to the same optimal measurement structures, namely, spherical 2 designs. These structures yield a very simple closed-form expression for the covariance matrix of the Stokes vector estimator, and thus of the variances of each element of the Stokes vector. These expressions constitute the fundamental limit of the estimation variance that can be reached by a Stokes polarimeter in the presence of additive noise.

As a conclusion, we would like to stress that although the three considered metrics are equivalent for polarimeter optimization in the presence of additive noise, the EWV has the simplest physical interpretation since it corresponds to an estimation variance, which has a clear and useful statistical meaning. As a consquence, in contrast to the two other metrics, the EWV can be used for polarimeter optimization in the presence of noise sources with non-additive, non-Gaussian, or mixed statistics.

A very interesting perspective is the determination of the optimal measurement matrices in the presence of non-additive noise sources. As said above, this problem has already been addressed by optimizing the EWV obtained after application of the pseudo-inverse estimator.[39,40] Although this procedure give satisfying results in practice,[42] it is not strictly optimal. Indeed, in the presence of non additive and non Gaussian noise, by virtue of the Cramér-Rao lower bound, the appropriate criterion is the trace of the inverse Fisher information matrix.[17,18] The value of this criterion corresponds to the EWV of an efficient estimator, which the pseudo-inverse estimator is not in

general. An exciting perspective is to work at determining the measurement matrices optimizing this Fisher criterion, and at analyzing the differences between the obtained optimal structures and the spherical 2 designs.

Appendix A: Positivity of the third term of Eq. (15)

We demonstrate in this appendix that the third term of the expression of $\mathrm{tr}[\mathbb{G}^{-1}]$ in Eq. (15) is positive definite. Since the matrix $\mathbb{Q}^T\mathbb{Q}$ is by definition a positive matrix, the numerator of this term is also positive. We therefore need only analyse the denominator. Considering then the singular value decomposition $\mathbb{Q} = \mathbb{U}\mathbb{F}\mathbb{V}^T$, where $\mathbb{U}$ and $\mathbb{V}$ are unitary matrices and $\mathbb{F}$ is diagonal, it is easily seen that

$$\mathbb{Q}(\mathbb{Q}^T\mathbb{Q})^{-1}\mathbb{Q}^T = \mathbb{U}\mathbb{F}\mathbb{U}^T \tag{35}$$

where $\mathbb{F} = \mathbb{D}(\mathbb{D}^T\mathbb{D})^{-1}\mathbb{D}^T$ is a diagonal $N \times N$ matrix. The first three diagonal elements of $\mathbb{F}$ are unity, whereas the other elements are zero. We thus have

$$\mathbf{q}^T(\mathbb{Q}^T\mathbb{Q})^{-1}\mathbf{q} = \mathbf{v}^T\mathbb{F}\mathbf{v} = \sum_{i=1}^{3} v_i^2 \tag{36}$$

where $\mathbf{v} = \mathbb{U}^T\mathbf{r}$ is an N-dimensional vector. Moreover

$$\sum_{i=1}^{3} v_i^2 \leq \sum_{i=1}^{N} v_i^2 = \|\mathbf{v}\|^2 = \|\mathbf{r}\|^2 = N \tag{37}$$

since $\mathbb{U}$ is a unitary matrix. Hence

$$\mathbf{q}^T(\mathbb{Q}^T\mathbb{Q})^{-1}\mathbf{q} \leq N, \tag{38}$$

177 which means that the third term of Eq. (15) is positive.

178 Appendix B: Satisfying Eqs. (16) and (20) with spherical t designs

Consider a finite set of points $\{\mathbf{w}_j\}$ ($j \in [1, N]$), which lie on the surface of the three dimensional unit sphere. The set of points $\{\mathbf{w}_j\}$ are said to constitute a spherical t design if for *any* polynomial function $f(\mathbf{w})$ of order t or lower

$$\sum_{j=1}^{N} f(\mathbf{w}_j) = N \int f(\mathbf{w}) d\sigma_{\mathbf{w}}, \tag{39}$$

179 where $d\sigma_{\mathbf{w}}$ is the normalized surface area element of the unit sphere.

Proof that Eqs. (16) and (20) can be satisfied using spherical 2 designs follows by showing that we can generate the constraints through appropriate choice of polynomial functions $f(\mathbf{w})$ of second order degree or less in Eq. (39). Considering first the case $f(\mathbf{w}) = w_s$ ($s \in [1, 3]$), substitution into Eq. (39) yields

$$\sum_{j=1}^{N} w_{sj} = N \int w_s d\sigma_{\mathbf{w}} \tag{40}$$

where w_{sj} is the value of the sth element of $\mathbf{w}_j$. We can express $\mathbf{w}$ in terms of the usual spherical polar coordinates, i.e., $\mathbf{w} = [\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta]^T$ such that $4\pi d\sigma_{\mathbf{w}} = \sin\theta d\theta d\phi$. It is then simple to show that $\int w_s d\sigma_{\mathbf{w}} = 0$ for $s \in [1, 3]$ such that Eq. (40) reduces to Eq. (16). Similarly,

using the polynomial function $f(\mathbf{w}) = w_s w_t$ for $\{s,t\} = \{1,2,3\}$ Eq. (39) becomes

$$\sum_{j=1}^{m} w_{sj} w_{tj} = N \int w_s w_t d\sigma_{\mathbf{w}}. \tag{41}$$

Evaluating the integral on the right hand side yields $\delta_{st}/3$, such that Eq. (41) reduces to Eq. (20), therefore completing our proof. Although we have proven that Eqs. (16) and (20) can be satisfied by a spherical 2 design, it is worthwhile to note that it automatically follows that they can also be satisfied by a spherical design of higher order, $t \geq 2$, because a spherical t design is also a $t-1$ design.

Appendix C: Previous derivation

The constraints derived in Section 2 through direct minimization of the EWV were first derived by Foreman et. al exploiting a claimed equivalence between minimizing the trace of $\mathbb{G}^{-1}$ and maximizing the determinant of $\mathbb{G}$. Specifically, using the definition of the matrix inverse and Jacobi's formula it was first shown that the condition number can be expressed in the form[35]

$$\kappa^2 = \frac{N}{2}\text{tr}[(\mathbb{W}^T\mathbb{W})^{-1}] = \frac{N}{2}\sum_{i=1}^{4}\frac{\partial \ln|\mathbb{G}|}{\partial G_{ii}} \tag{42}$$

where G_{ii} are the diagonal elements of $\mathbb{G}$. Based on Eq. (42) Foreman et al. claim that the equivalence of optimization metrics follows from the differential relation $2d\ln\kappa = -d\ln|\mathbb{G}|$. Regrettably this relation does not follow from Eq. (42), nor in fact does it hold in general, as can be seen by expressing both $\ln[\text{tr}[\mathbb{G}^{-1}]] = 2\ln\kappa + \text{const.}$ and $\ln|\mathbb{G}|$ in terms of the eigenvalues of $\mathbb{G}$.

Disclosures

The authors declare they have no conflicts of interest.

Acknowledgements

The authors would like to thank Dr. A. Favaro for useful discussions. MRF also acknowledges financial support from the Royal Society through a Royal Society University Research Fellowship.

References

1 S. Brasselet, "Polarization-resolved nonlinear microscopy: application to structural molecular and biological imaging," Adv. Opt. Photon. **3**, 205–271 (2011).

2 E. Hadamcik, A. K. Sen, A. C. Levasseur-Regourd, R. Gupta, and J. Lasue, "Polarimetric observations of comet 67P/Churyumov-Gerasimenko during its 2008-2009 apparition," Astron. Astrophys. **517**, A86 (2010).

3 N. Mazumder, J. Qiu, M. R. Foreman, C. M. Romero, C.-W. Hu, H.-R. Tsai, P. Török, and F.-J. Kao, "Polarization-resolved second harmonic generation microscopy with a four-channel Stokes polarimeter," Opt. Express (2012).

4 W. Kaminsky, K. Claborn, and B. Kahr, "Polarimetric imaging of crystals," Chem. Soc. Rev. **33**, 514–525 (2004).

5 M. R. Foreman, C. Macías-Romero, and P. Török, "Determination of the three dimensional orientation of single molecules," Opt. Lett. **33**, 1020–1022 (2008).

6 J. Chon, P. Zijlstra, and M. Gu, "Five-dimensional optical recording mediated by surface plasmons in gold nanorods," Nature **459**, 410–413 (2009).

7 C. Macias-Romero, P. R. T. Munro, and P. Török, "Polarization-multiplexed encoding at nanometer scales," Opt. Express **22**, 26240–26245 (2014).

8 J. Liu, S.-M. Li, L. Zhu, A.-D. Wang, S. Chen, C. Klitis, C. Du, Q. Mo, M. Sorel, S.-Y. Yu, X.-L. Cai, and J. Wang, "Direct fiber vector eigenmode multiplexing transmission seeded by integrated optical vortex emitters," Light: Sci. Appl. **7**, 17148 (2018).

9 J. N. Damask, *Polarization Optics in Telecommunications*, Springer Series in Optical Sciences (Springer-Verlag, New York, New York, 2005).

10 F. Goudail, P. Terrier, Y. Takakura, L. Bigué, F. Galland, and V. Devlaminck, "Target detection with a liquid-crystal-based passive Stokes polarimeter," Appl. Opt. **43**, 274–282 (2004).

11 R. M. A. Azzam and N. M. Bashara, "Ellipsometry and polarized light," (1977).

12 M. Grell and D. D. C. Bradley, "Polarized Luminescence from Oriented Molecular Materials," Adv. Mater. **11**, 895–905 (1999).

13 D. Sofikitis, L. Bougas, G. E. Katsoprinakis, A. K. Spiliotis, B. Loppinet, and T. P. Rakitzis, "Evanescent-wave and ambient chiral sensing by signal-reversing cavity ringdown polarimetry," Nature **514**, 76–79 (2014).

14 N. J. Pust and J. A. Shaw, "Dual-field imaging polarimeter using liquid crystal variable retarders," Appl. Opt. **45**, 5470–5478 (2006).

15 A. S. Alenin and J. S. Tyo, "Generalized channeled polarimetry," J. Opt. Soc. Am. A **31**, 1013–1022 (2014).

16 E. Compain and B. Drevillon, "Broadband division-of-amplitude polarimeter based on uncoated prisms," Appl. Opt. **37**, 5938–5944 (1998).

17 M. R. Foreman, C. Macías-Romero, and P. Török, "A priori information and optimisation in polarimetry," Opt. Express **16**, 15212–15227 (2008).

18 M. R. Foreman and P. Török, "Information and resolution in electromagnetic optical systems," Phys. Rev. A **82**, 043835 (2010).

19 A. S. Alenin, "Matrix Structure for Information Driven Polarimeter Design," Phd, University of Arizona (2015).

20 R. M. A. Azzam, "Optimal beam splitters for the division-of- amplitude photopolarimeter," J. Opt. Soc. Am. **20**, 955–958 (2003).

21 A. Ambirajan and D. C. Look, "Optimum angles for a polarimeter: part II," Opt. Eng. **34**, 1656–1658 (1995).

22 A. Ambirajan and D. C. Look, "Optimum angles for a polarimeter: part I," Opt. Eng. **34**, 1651–1655 (1995).

23 J. S. Tyo, "Design of optimal polarimeters: maximization of signal-to-noise ratio and minimization of systematic error," Appl. Opt. **41**, 619–630 (2002).

24 D. Sabatke, M. R. Descour, E. L. Dereniak, W. C. Sweatt, S. A. Kemme, and G. S. Phipps, "Optimization of retardance for a complete Stokes polarimeter," Opt. Lett. **25**, 802–804 (2000).

25 A. Peinado, A. Lizana, J. Vidal, C. Iemmi, and J. Campos, "Optimization and performance criteria of a Stokes polarimeter based on two variable retarders." Opt. Express **18**, 9815–9830 (2010).

26 R. M. A. Azzam, I. M. Elminyawi, and A. M. El-Saba, "General analysis and optimization of the four-detector photopolarimeter," J. Opt. Soc. Am. A **5**, 681–689 (1988).

27 S. N. Savenkov, "Optimisation and structuring of the instrument matrix for polarimetric measurements," Opt. Eng. **41**, 965–972 (2002).

28 M. H. Smith, "Optimization of a dual-rotating-retarder Mueller matrix polarimeter," Appl. Opt. **41**, 2488–2493 (2002).

29 K. M. Twietmeyer and R. A. Chipman, "Optimization of Mueller matrix polarimeters in the presence of error sources." Opt. Express **16**, 11589–11603 (2008).

30 A. Ambirajan and D. C. Look, "Optimum angles for a Mueller matrix polarimeter," Proc. SPIE **2265**, 314–326 (1995).

31 D. Layden, M. F. G. Wood, and I. A. Vitkin, "Optimum selection of input polarization states in determining the sample Mueller matrix: a dual photoelastic polarimeter approach," Opt. Express **20**, 20466–20481 (2012).

32 F. Goudail, M. Boffety, and S. Roussel, "Optimal configuration of static Mueller imagers for target detection," J. Opt. Soc. Am. A **34**, 1054–1062 (2017).

33 M. Boffety, H. Hu, and F. Goudail, "Contrast optimization in broadband passive polarimetric imaging," Opt. Lett.' **39**, 6759–6762 (2014).

34 F. Goudail and M. Boffety, "Optimal configuration of static polarization imagers for target detection," J. Opt. Soc. Am. A **33**, 9–16 (2016).

35 M. R. Foreman, A. Favaro, and A. Aiello, "Optimal Frames for Polarisation State Reconstruction," Phys. Rev. Lett. **115**, 263901 (2015).

36 P. Delsarte, J. M. Goethals, and J. J. Seidel, "Spherical codes and designs," Geometria Dedicata **6**, 363–388 (1977).

37 F. Goudail, "Optimal Mueller matrix estimation in the presence of additive and Poisson noise for any number of illumination and analysis states," Opt. Lett. **42**, 2153–2156 (2017).

38 M. Brookes, "The Matrix Reference Manual," (2011).

39 F. Goudail, "Noise minimization and equalization for Stokes polarimeters in the presence of signal-dependent Poisson shot noise." Opt. Lett. **34**, 647–649 (2009).

40 F. Goudail, "Equalized estimation of Stokes parameters in the presence of Poisson noise for any number of polarization analysis states," Opt. Lett. **41**, 5772–5775 (2016).

41 S. Roussel, M. Boffety, and F. Goudail, "Polarimetric precision of micropolarizer grid-based camera in the presence of additive and Poisson shot noise," Opt. Express **26**, 29968–29982 (2018).

42 F. Goudail, "Performance comparison of pseudo-inverse and maximum-likelihood estimators of Stokes parameters in the presence of Poisson noise for spherical design-based measurement structures," Opt. Lett. **42**, 1899–1902 (2017).

43 J. S. Tyo, "Noise equalization in Stokes parameter images obtained by use of variable-retardance polarimeters," Opt. Lett. **25**, 1198–2000 (2017).

François Goudail graduated from the École Supérieure d'Optique (Orsay) in 1992 and obtained his PhD in 1997 from the University of Aix-Marseille III. He was an associate professor at Fresnel Institute (Marseille) until 2005. He is now a professor at the Institut d'Optique Graduate School (Palaiseau). His research topics include information extraction in images from different types of passive and active sensors (hyperspectral, SAR, polarimetric), wavefront engineering and joint design of optical systems and image processing algorithms.

Matthew R. Foreman received his MPhys degree from the University of Oxford in 2006 and his PhD from Imperial College London in 2010. He has held research posts at the UK National Physical Laboratory (Teddington) and the Max Planck Institute for the Science of Light (Erlangen), where he held an Alexander von Humboldt Fellowship. Currently, he is a Royal Society University Research Fellow at Imperial College London. His research interests include theoretical aspects of nanophotonics, plasmonics, polarimetry, random scattering and sensing.