

Fundamental estimation bounds for polarimetric imagery

Daniel A. LeMaster

Department of Electrical and Computer Engineering
Air Force Institute of Technology
2950 Hobson Way, Wright-Patterson AFB, OH 45433, USA
daniel.lemaster@afit.edu

Abstract: Precise channel-to-channel registration is a prerequisite for effective exploitation of passive polarimetric imagery. In this paper, the Cramer-Rao bound is employed to determine the limits of registration precision in the presence of scene polarization diversity, channel noise, and random translational registration errors between channels. The effects of mis-registration on Stokes image estimation are also explored in depth. Algorithm bias is discussed in the context of the bound, without being estimator specific. Finally, case studies are presented for polarization insensitive imagery (a special case) and linear polarization imaging systems with three and four channels. An optimum polarization channel arrangement is proposed in the context of the bound.

OCIS codes: (110.5405) Polarimetric imaging; (100.2000) Digital image processing; (110.3925) Metrics

References and links

1. M. Kishimoto, L. E. Kay, R. Antonucci, T. W. Hurt, R. D. Cohen, and J. H. Krolik, "Ultraviolet Imaging Polarimetry of the Seyfert 2 Galaxy Markarian 3," *apj* **565**, 155–162 (2002).
2. W. G. Egan, "Polarization in remote sensing," in *Polarization and remote sensing*, W. G. Egan, ed., Proc. **SPIE** 1747, 2-48 (1992).
3. S. Lin, K. Yemelyanov, E. Pugh, Jr, and N. Engheta, "Separation and contrast enhancement of overlapping cast shadow components using polarization," *Opt. Express* **14**, 7099–7108 (2006).
4. C. M. Persons, D. B. Chenault, M. W. Jones, K. D. Spradley, M. G. Gulley, and C. A. Farlow, "Automated registration of polarimetric imagery using Fourier transform techniques," in *Polarization Measurement, Analysis, and Applications V*, Goldstein, Dennis H.; Chenault, David B., eds., Proc. **SPIE** 4819, 107-117 (2002).
5. X. Wang, S. Yang, J. Ma, and Y. Qiao, "Automated registration of polarimetric image using wavelet transform techniques," vol. 5832, pp. 695–702 (SPIE, 2005).
6. D. A. LeMaster, "A Comparison of Template Matching Registration Methods for Polarimetric Imagery," in *Aerospace Conference, 2008 IEEE*, vol. 1, pp. 1–9 (2008).
7. Guyot, S. and Anastasiadou, M. and Deléchéle, E. and De Martino, A., "Registration scheme suitable to Mueller matrix imaging for biomedical applications," *Opt. Express* **15**, 7393–7400 (2007).
8. S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice Hall, Englewood Cliffs, New Jersey, 1993).
9. D. Robinson and P. Milanfar, "Fundamental performance limits in image registration." *IEEE Trans. Image Process* **13**, 1185–1199 (2004).
10. A. Yetik, I.S.; Nehorai, "Performance bounds on image registration," *IEEE Trans. Signal Process* **54**, 1737–1749 (May 2006).
11. B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing* **21**, 977–1000 (2003).
12. L. L. Scharf, *Statistical Signal Processing: detection, estimation, and time series analysis* (Addison-Wesley, Reading, Massachusetts, 1991).
13. L. L. Scharf, L. T. McWhorter, "Geometry of the Cramer-Rao bound" *Signal Processing* **31**, 301-311 (1993).
14. H. Van Trees, *Detection, estimation, and modulation theory. Part 1: detection, estimation, and linear modulation theory* (Wiley, New York, 2001).

15. J. Tyo, "Design of optimal polarimeters: maximization of signal-to-noise ratio and minimization of systematic error," *Appl. Opt.* **41**, 619–630 (2002).
 16. J. Tyo, "Optimum linear combination strategy for an N-channel polarization-sensitive imaging or vision system," *J. Opt. Soc. Am. A* **15**, 359–366 (1998).
 17. E. Collett, *Polarized Light: Fundamentals and Applications* (Marcel Dekker, Inc., New York, 1992).
 18. M. Healy, *Matrices for Statistics* (Oxford University Press, USA, 1986).
 19. A. Graham, *Kronecker Products and Matrix Calculus With Applications*. (Wiley, New York, 1982).
-

1. Introduction

Passive polarimetric imaging (PI) is a form of remote sensing in which the target scene is characterized in terms of the polarization state of its incoherently reflected or emitted radiation. Polarimetric imaging has been applied to successfully address problems in astrophysics [1], terrestrial remote sensing [2], machine vision [3], and other disciplines. In all such applications, the required polarimetric measurements are achieved by combining multiple imaging channels each with a unique response to polarized light. These channels may be collected in a parallel configuration (i.e. each channel has its own optical path) or in series where one or more polarizing elements are manipulated between channel images. In either case, target polarization state estimates are inherently prone to both channel noise and registration errors. In this paper, we examine the combined effects of misregistration and channel noise in the statistical framework of the Cramer-Rao lower bound (CRLB).

For the present purpose, systems responsive to linear polarization are considered. Persons (et al.) noted that, among common means of collecting polarimetric imagery, translational and rotational registration errors between channels are most prevalent [4]. In this context, Persons and others [5, 6] have made limited attempts to characterize the effects of registration errors on polarimetric imagery using established (but polarization insensitive) registration algorithms. In addition, a recent registration study has also been conducted for the related problem of Mueller matrix imaging [7]. The results of these studies range from successful registration within stated tolerances to complete failure. Each of these efforts is a useful reference point, however, they lack the depth required to make complete comparisons across algorithms or imaging scenarios. Indeed, what is most lacking is a theoretical context in which to place these results.

The first step in providing this context is to identify the problem of estimating the polarimetric signature of a target scene as a joint estimation of the misregistration between channels and of the scene itself. Second, channel noise and misregistration are random variables, therefore, this framework must be statistical in nature. Finally, this framework must account for the influence of deterministic parameters such as channel orientation and spacing as well as polarization diversity in the scene. In this paper, the Cramer-Rao bound is derived for this joint estimation problem and expounded as a theoretical framework that meets all of these criteria.

The Cramer-Rao bound determines the lower limit on the variance of any estimator. The form of the bound is different for biased and unbiased cases. With the exception of section 5, this research focuses mainly on bounds for unbiased estimators. The utility of the CRLB for unbiased estimators is well described by Kay [8]; his arguments may also be tailored to this specific case: if a new estimator is devised, it is more efficient to compare it to a single bound rather than to attempt to simulate results for many different existing estimators. Alternatively, the bound can be used to determine if any existing algorithm meets this best case scenario. If no such algorithm exists, the bound can motivate the search for a better estimator. Similar studies via simulation (e.g. Monte Carlo methods) lack the theoretical foundation which the CRLB readily provides. In the present work, the bound is also shown to provide insight into the relationship between the sensor, target scene, and estimator.

Inspiration for applying this approach to the polarimetric imagery estimation problem comes from two recent works. Robinson and Milanfar [9] applied the Cramer-Rao bound to the prob-

lem of estimating the translational misregistration between pairs of (polarization insensitive) images in cases where the underlying scene is not estimated as a parameter. Yetik and Nehorai expanded this work to describe limits for other common misregistration issues (e.g. rotation, affine transformation) and for both feature based and intensity based registration algorithms [10]. Aside from insensitivity to polarization effects, the underlying scene assumption shared by both sets of authors obviates the need for a joint estimator and is therefore inadequate for the present purpose.

The bound derived in this work is most applicable to so called “feature matching, area-based” or “template matching” registration estimators. Common among these methods is that registration is achieved by comparing image intensities without regard to specific objects in the scene. A survey of these methods can be found in [11].

The newly derived bound determines the minimum achievable variance on estimates of the translational misregistration between channels and of the polarization state at each point in the scene. The inclusion of polarization effects and removal of the known image assumption from Robinson and Milanfar leads to a bound calculation that is computationally impractical using direct methods. As such, matrix theory is applied to the newly derived bound so that extraction of the relevant terms becomes tenable. Additionally, many well known polarization insensitive image registration algorithms exhibit bias and, in fact, some biased estimators have been shown to outperform unbiased estimators in theory and in practice. Though no such study has been conducted with polarimetric imagery, a biased estimator form of the CRLB is provided in anticipation of a similar eventual result.

The bound stands on its own as an algorithm evaluation metric, however, it may also be used in a quantitative evaluation of system design criteria. To this end, the bound is used to determine the conditions under which it is necessary to generate the channel registration parameters externally (as opposed to including them in the estimator). Furthermore, the polarization insensitive case is actually a special case of this new bound. As a consequence, Robinson and Milanfar’s bound is generalized to the case where the underlying image is unknown.

Finally, the bound can also be used to suggest optimal channel configurations. Several researchers have shown that optimum configuration (in terms of maximizing SNR) is achieved by distributing the channels evenly across all possible polarization states. In the most recent of these attempts, Tyo improved on this theory by considering the optimization problem in the presence of random channel calibration errors [15]. In perhaps the most directly related research, Tyo also showed that channels that are evenly spaced across all linear polarization states are optimal in a principal components sense [16]. The analysis presented here complements this body of research in that it demonstrates the same conclusion under different sets of assumptions and using different metrics.

1.1. Stokes vector formulation

The Stokes parameters are a standard tool for describing the polarization state of radiation in passive remote sensing literature. The great utility of the Stokes formalism stems from the fact that these parameters can be determined directly from observables. This section provides an introduction to the Stokes representation of polarized radiation to the extent that such knowledge is required for the present purpose. The interested reader can consult Collett [17] for a more detailed description. In the section that follows, the Stokes parameters are used to provide a concise example of how the inclusion of polarization content can frustrate attempts to explain registration estimator behavior using traditional (i.e. polarization insensitive) techniques.

Consider a theoretical optical system consisting of a solid-state detector and two specialized polarization manipulating elements. The first of these two elements, a retarder, introduces a phase delay, ϕ , between the x and y electric field components of the incident signal. The second

element, a polarizer, transmits the portion of this field along angle θ (measured with respect to the x axis) and completely attenuates the field everywhere else. The intensity measured at the detector is given by:

$$I = \frac{1}{2}(S_0 + S_1 \cos 2\theta + S_2 \cos \phi \sin 2\theta + S_3 \sin \phi \sin 2\theta). \quad (1)$$

Consequently, the measured intensity is a weighted sum of the Stokes parameters. Physically, S_0 is the total incident intensity, S_1 and S_2 together describe the inclination angle of the polarization ellipse, and S_3 and S_0 together describe its ellipticity. The Stokes formalism is also capable of describing partially polarized, broadband radiation whereas a field based representation (e.g. the Jones formalism) is not. As stated previously, system responsiveness to the ellipticity of the incident field is not considered in this research. Consequently, the phase delay parameter ϕ will be assumed to be zero for each imaging channel. In other words, I is not a function of S_3 and all measurable polarization effects will be due to linear polarization.

After setting $\phi = 0$, one must still contend with polarizing optical elements that are less than ideal. That is to say, the preferred signal may be partially attenuated while, at the same time, the remaining signal is not completely suppressed. Consequently, the intensity equation is better represented with weights on the Stokes parameters that are determined via lab calibration.

$$I_x = a_{x0}S_0 + a_{x1}S_1 + a_{x2}S_2 \quad (2)$$

where x refers to the specific channel in question. For an ensemble of channels, it is useful to define vectors for intensity measurements, \mathbf{I} , and Stokes parameters, \mathbf{S} , such that $\mathbf{I} = \mathbf{A}\mathbf{S}$ and matrix \mathbf{A} consists of the channel Stokes weighting parameters:

$$\mathbf{A} = \begin{bmatrix} a_{10} & \cdots & a_{12} \\ \vdots & \ddots & \vdots \\ a_{N0} & \cdots & a_{N2} \end{bmatrix} \quad (3)$$

1.2. A quick example

Building upon the previous section, the results of a simple laboratory experiment are presented in order to illustrate some issues regarding the registration of polarimetric imagery. The underlying assumption in many registration algorithms of the aforementioned “template matching” category (e.g. cross-correlation and variants) is that the images under test contain the same scene content but are translated and corrupted by noise. Other methods in this category, such as registration via mutual information, are perhaps more suited toward the registration of polarimetric imagery in that a statistical dependence between intensity values in each image is assumed rather than a direct correspondence in intensity. To illustrate these differences, the following example was contrived to show how a registration algorithm can utterly fail when strong polarization content violates its underlying assumptions.

Figure 1 contains the result of an attempt to register the three channels of an imaging polarimeter using cross-correlation. The target scene contains two fully polarized bars from a resolution target oriented such that the bars have orthogonal polarization states. The channels are spaced such that each is most sensitive to polarization at 0° , 60° , or -60° . The target polarization is rotated slightly ($\approx 10^\circ$) out of the sensor reference frame so that each bar appears (to a greater or lesser extent) in each channel. False color and an arrow are used to accentuate the weak signal of the top bar in the 0° channel.

Clearly, these three channels do not contain the same scene content which is in violation of the assumptions built into the cross-correlation algorithm. As a consequence, the bottom bar in

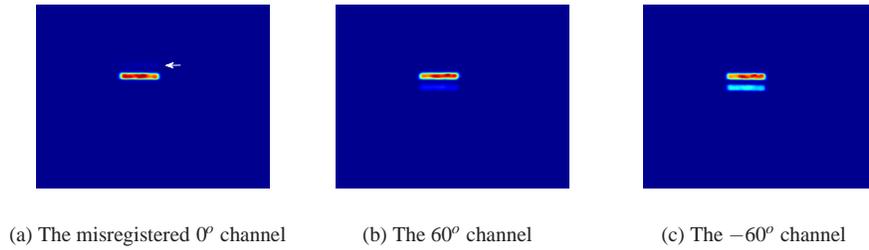


Fig. 1. The misregistered polarimetric images

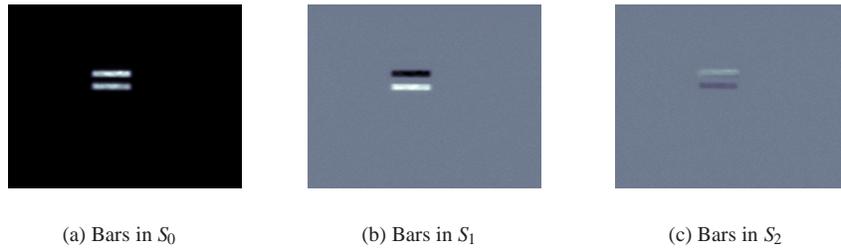


Fig. 2. The Stokes parameter images of the bar target

channel 0° has been misregistered to the top bars in the 60° and -60° channels. The Stokes parameter images of this scene (figure 2) provide an explanation of this behavior.

Note that S_1 and S_2 can take on negative values; the reader should interpret the dark regions in the S_1 and S_2 images in this way. This behavior is unlike S_0 , which is strictly positive. The comparatively higher contrast in the S_1 image occurs because the signal in S_2 is much weaker than in S_1 . If the sensor were not sensitive to polarization then the captured image in each channel would be the S_0 image.

Comparing the Stokes images in figure 2 with the intensity equation (1), the bottom bar in the 0° channel image is bright because its preferred polarization state is nearly parallel to the preferred polarization state of the channel. In this region $I \approx \frac{1}{2}(S_0 + S_1)$ where S_1 is a positive quantity. The small contribution from S_2 in this channel is ignored. The top bar is orthogonally polarized to channel 0° and in this region the form of the intensity equation is the same but S_1 is *negative*. A similar analysis could be conducted for the remaining channels but the point of this section has already been made: the phenomenology of polarization imagery is different than that of traditional intensity imagery and as such, the rules developed for image registration must be reevaluated in this new context.

2. Bound definition and data model

In this section, the Cramer-Rao bound is defined and the components required to calculate the bound are identified. These components are then specified for the specific problem of calculating the bound for a joint registration and Stokes parameter estimator.

2.1. Definition of the Cramer-Rao bound

Let \mathbf{Z} be a vector of random variables that is parameterized by vector θ . Define $\hat{\theta}$ to be any unbiased estimate of these parameters and \mathbf{z} to be one realization of \mathbf{Z} . The Cramer-Rao in-

equality provides the lower bound on this estimator's error covariance matrix in terms of the Fisher Information Matrix (FIM), J :

$$\text{Cov}[\hat{\boldsymbol{\theta}}] \geq J^{-1}, \quad (4)$$

where

$$J = -E \left[\frac{\partial}{\partial \boldsymbol{\theta}} \left(\frac{\partial}{\partial \boldsymbol{\theta}} L(\boldsymbol{\theta}, \mathbf{z}) \right)^T \right] \quad (5)$$

and $L(\boldsymbol{\theta}, \mathbf{z})$ is the data log-likelihood function [12]. Note that $E[\dots]$ represents the expected value operation over \mathbf{Z} . The minimum variance for the estimate of each parameter in $\boldsymbol{\theta}$ is given by the diagonal elements of J^{-1} . When J is not positive definite the Cramer-Rao bound is not defined.

The concept of a log-likelihood function may require some elaboration. A likelihood function describes how the probability density for a given measurement changes as $\boldsymbol{\theta}$ changes. Hence, calculation of the likelihood function requires knowledge of the probability density function for \mathbf{Z} , $p_{\boldsymbol{\theta}}(\mathbf{z})$, at each measured \mathbf{z} . The log-likelihood function is simply the natural log of the likelihood function:

$$L(\boldsymbol{\theta}, \mathbf{z}) = \ln p_{\boldsymbol{\theta}}(\mathbf{z}). \quad (6)$$

2.2. Data model

The bound described in the previous section is defined in terms of a set of random variables \mathbf{Z} , parameter vector, $\boldsymbol{\theta}$, and their corresponding log-likelihood function, $L(\boldsymbol{\theta}, \mathbf{z})$. The purpose of this section is to define these items for the specific problem of generating polarimetric imagery from noisy, misregistered data.

In this scenario, the collected images are realizations of \mathbf{Z} . Consistent with [9] and [10], the images to be registered are modeled as sampled, noisy versions of the continuously varying underlying target scene. Channel-to-channel point spread function variations are assumed to be minimal, sampling is at the Nyquist frequency or better, and noise in the scene is zero-mean, Gaussian and IID. The sampled coordinates in the (arbitrarily selected) first imaging channel, f_1 , are taken to be the reference by which the remaining channels, f_2 to f_N , are specified. Mathematically, the collected image z_i for channel i is given by:

$$z_i(\mathbf{m}_n) = f_i(\mathbf{m}_n) + \varepsilon_i(\mathbf{m}_n), \quad (7)$$

where \mathbf{m}_n is the 2-dimensional coordinates of a given pixel in the image plane and $\varepsilon_i(\mathbf{m}_n)$ is a realization of the channel noise. The boundaries of \mathbf{m}_n are determined by the region over which the collected images overlap. This overlap is assumed to be square ($1 \leq n \leq p^2$ for a $p \times p$ image). In a departure from the cited research, the image content in each channel is determined by a shared Stokes parameter mapping rather than a common intensity mapping:

$$\begin{aligned} f_1(\mathbf{m}_n) &= \sum_{j=0}^2 a_{1j} S_j(\mathbf{m}_n) \\ f_i(\mathbf{m}_n) &= \sum_{j=0}^2 a_{ij} S_j(\mathbf{m}_n - \mathbf{v}_i), \end{aligned} \quad (8)$$

where \mathbf{v}_i is the 2-dimensional translational misregistration between \mathbf{f}_i and \mathbf{f}_1 .

For each channel, there is a one-to-one correspondence between the 2-dimensional image and a 1-dimensional vector:

$$\begin{aligned}\mathbf{z}_i &= \left[z_i(\mathbf{m}_1) \dots z_i(\mathbf{m}_{p^2}) \right]^T \\ \mathbf{f}_i &= \left[f_i(\mathbf{m}_1) \dots f_i(\mathbf{m}_{p^2}) \right]^T,\end{aligned}\quad (9)$$

which, in turn, allows for a very compact expression of the following results.

As previously stated, the per-pixel channel noise is zero mean, IID and Gaussian; consequently, the data log-likelihood function is given by:

$$L(\theta, \mathbf{z}) = \frac{-1}{2\sigma^2} \sum_{i=1}^N (\mathbf{z}_i - \mathbf{f}_i)^T (\mathbf{z}_i - \mathbf{f}_i) + \xi \quad (10)$$

where σ^2 is the noise variance and ξ is a constant term that is not dependent on θ . It is clear from this equation that the log-likelihood is dependent on the region of overlap defined by the intersection of all images \mathbf{f}_i . This region of intersection is, in turn, dependent on the relative misregistration between the images. Following the lead of Yetik, this overlap region is assumed to be constant. The efficacy of this assumption is greatest when the relative misregistration between images is small when compared to the dimensions of the overlap region; it is reasonable to assume that a multi-channel polarimeter will be operating in this regime.

Additionally, calculation of the Fisher information matrix requires an unambiguous ordering in the parameter vector θ . Since the goal is to place a bound on a joint estimator of the translational shifts between images *and* the values of the Stokes parameters at each pixel in the image, this parameter vector will be very large indeed:

$$\theta = \left[\mathbf{v}_2^T \dots \mathbf{v}_N^T \quad \mathbf{S}_0^T \dots \mathbf{S}_2^T \right]^T, \quad (11)$$

where, analogous to the vectorized form of the channels mean images, \mathbf{f}_i , each Stokes parameter vector is given by:

$$\mathbf{S}_i = \left[S_i(\mathbf{m}_1) \dots S_i(\mathbf{m}_{p^2}) \right]^T. \quad (12)$$

Scharf [12] provides an expression for the Fisher information matrix for m observations of an multivariate, normally distributed random vector with covariance matrix, R , and mean, \mathbf{f} :

$$J_{ij} = m \frac{\partial \mathbf{f}^T}{\partial \theta_i} R^{-1} \frac{\partial \mathbf{f}}{\partial \theta_j} \quad (13)$$

As shown in (10), there is one observation (i.e. $m = 1$) for each of N random images. Each image i has a covariance matrix $R = \sigma^2 I$ and mean \mathbf{f}_i . It is straightforward to show that the FIM in this case is simply a sum of N instances of (13):

$$J_{ij} = \frac{1}{\sigma^2} \sum_{n=1}^N \frac{\partial \mathbf{f}_n^T}{\partial \theta_i} \frac{\partial \mathbf{f}_n}{\partial \theta_j} \quad (14)$$

3. Fisher information for the joint estimator

The Fisher information matrix from (14) is developed in this section by partitioning J into sub-matrices. Matrix partitioning is used to exploit the inherent symmetry in the Fisher information matrix and to lend insight into the physical interpretation of J . In [13], Scharf and McWhorter show that a FIM in the form of (14) may be partitioned so that the subset FIMs of the parameter space may be considered individually. In this work, it is useful to partition the parameter space

such the sub-FIM V describes correlations amongst the translational registration parameters and the sub-FIM S describes correlations between Stokes parameters. Define:

$$J = \begin{bmatrix} V & H^T \\ H & S \end{bmatrix} \quad (15)$$

where H , to use Scharf and McWhorter's parlance, relates the intercorrelations between the registration and Stokes partitions in θ . In what follows, each of V , H , and S are described in detail.

The first matrix, V , is actually the FIM for an unbiased estimator of the misregistration between channels when the underlying Stokes images are known a priori. The inverse of V is the Cramer-Rao bound for the registration estimator under this "known prior" condition. In the two channel polarization insensitive case, V is the bound from Robinson and Milanfar. V is composed of $(N-1)^2$ submatrices of the form:

$$V_{ij} = \begin{cases} \frac{1}{\sigma^2} \left(\frac{\partial \mathbf{f}_i^T}{\partial \mathbf{v}_i} \right) \left(\frac{\partial \mathbf{f}_j^T}{\partial \mathbf{v}_j} \right)^T & \text{if } i = j, \\ \mathbf{0}_{2 \times 2} & \text{if } i \neq j \end{cases} \quad (16)$$

where $\mathbf{0}_{2 \times 2}$ is a 2×2 zero matrix. Though it does not appear explicitly in (15), it is also useful to define \tilde{V} with the same form as V but with entries:

$$\tilde{V}_{ij} = \begin{cases} \mathbf{0}_{2 \times 2} & \text{if } i = j, \\ \frac{1}{\sigma^2} \left(\frac{\partial \mathbf{f}_i^T}{\partial \mathbf{v}_i} \right) \left(\frac{\partial \mathbf{f}_j^T}{\partial \mathbf{v}_j} \right)^T & \text{if } i \neq j \end{cases} \quad (17)$$

The role of \tilde{V} will be demonstrated in the following section. For now, consider that the form of V demonstrates that errors in the estimates of the misregistration parameters are uncorrelated between channels when perfect knowledge of the underlying scene exists. When this perfect knowledge does not exist, \tilde{V} will be used to describe the nature of the correlation between channels.

In the lower right quadrant of (15) is the $3p^2 \times 3p^2$ matrix S . Note that S is the FIM that would be used to estimate the bound on a unbiased Stokes estimator if the relative shifts between the collected images were known a priori. S divides into 3×3 submatrices corresponding to combinations of Stokes parameters (images).

$$S_{jk} = \frac{1}{\sigma^2} \left(I_{p^2 \times p^2} \right) \sum_{i=1}^N a_{ij} a_{ik} \quad (18)$$

where $I_{p^2 \times p^2}$ is an identity matrix of rank p^2 . The block symmetry in S allows for substantial further simplification by defining the matrix:

$$C = \begin{bmatrix} c_{00} & \cdots & c_{02} \\ \vdots & \ddots & \vdots \\ c_{20} & \cdots & c_{22} \end{bmatrix}, \quad c_{jk} = \sum_{i=1}^N a_{ij} a_{ik} \quad (19)$$

in which case:

$$S = \frac{1}{\sigma^2} C \otimes I_{p^2 \times p^2} \quad (20)$$

where \otimes is the Kronecker product. This Kronecker product representation significantly simplifies inversion of the S matrix.

$$S^{-1} = \sigma^2 C^{-1} \otimes I_{p^2 \times p^2} \quad (21)$$

In other words, the inversion of the $3p^2 \times 3p^2$ matrix S is solved by simply inverting the 3×3 matrix C .

Connecting the matrices V and S is the $3p^2 \times 2(N-1)$ matrix H . Physically, if H were the zero matrix then the bounds on the registration and Stokes parameters could be determined independently of each other. Proof of this statement is provided in section (4). In the process of defining H , it becomes obvious that this independence condition is never met. H is composed of $3 \times (N-1)$ readily identifiable submatrices:

$$H_{ij} = \frac{a_{ji}}{\sigma^2} \left(\frac{\partial}{\partial \mathbf{v}_j} \mathbf{f}_j^T \right)^T. \quad (22)$$

Unless the collected image is constant everywhere (i.e. the derivative of the image is zero everywhere) then H is non-zero and the covariance bounds must be determined jointly. More on the relationship between H , V and \tilde{V} can be found in appendix A.

4. Bound derivation

The Fisher information matrix has been shown to contain $[2(N-1) + 3p^2] \times [2(N-1) + 3p^2]$ entries. To put the enormity of this matrix into perspective, consider the bound calculation for a four channel polarimeter used to estimate the first three Stokes parameters. Assuming a 512×512 overlap region, the corresponding FIM has approximately 6.18×10^{11} entries. Inversion of such a large matrix is prohibitive. In this section, the partitioning of the Fisher information matrix from the previous section is exploited to make this inversion problem tractable.

4.1. Block matrix inversion

As previously discussed, the variance bound for each parameter in θ is given by the diagonal entries of J^{-1} . Consequently, computational expense can be significantly reduced by avoiding the unnecessary calculation of many of the off diagonal terms in the inverse. A trivial rearrangement of the partitioned inverse of a block matrix in [18] provides the following:

$$B_v = (V - H^T S^{-1} H)^{-1} \quad (23)$$

and

$$B_S = S^{-1} + S^{-1} (H B_v H^T) S^{-1} \quad (24)$$

where B_v and B_S , which are submatrices of J^{-1} , identify the bounding covariance matrices on the shift parameters and Stokes images. As an aside, if $H = \mathbf{0}$ then $B_v = V^{-1}$ and $B_S = S^{-1}$, thus demonstrating the physical interpretation of H put forth in the previous section.

Equations (23) and (24) readily demonstrate how uncertainty in the misregistration between channels impacts uncertainty in the Stokes image estimates and vice versa. Similar to the $H = \mathbf{0}$ case, if the underlying Stokes images are known perfectly then the Cramer-Rao bound on the misregistration estimates would simply be $B_v = V^{-1}$. Likewise, if perfect knowledge of the registration parameters existed then $B_S = S^{-1}$. Consequently, it is clear that B_v and B_S are always larger than V^{-1} and S^{-1} in the absence of perfect knowledge.

4.2. Simplified registration parameter bound

Matrix B_v is addressed first because it is required to calculate B_S . As preliminary work, note that:

$$H^T S^{-1} H = \sigma^2 H^T \left(C^{-1} \otimes I_{p^2 \times p^2} \right) H. \quad (25)$$

Consequently, $H^T S^{-1} H$ can itself be partitioned into a matrix, D , such that:

$$D_{hk} = \sigma^2 \sum_{i=0}^2 \sum_{j=0}^2 C_{ij}^{-1} H_{ih}^T H_{jk}. \quad (26)$$

which provides the opportunity to profitably apply equations (50) and (51) from appendix A:

$$D_{hk} = \begin{cases} \sum_{i=0}^2 \sum_{j=0}^2 C_{ij}^{-1} a_{hi} a_{hj} V_{hh} & \text{if } h = k, \\ \sum_{i=0}^2 \sum_{j=0}^2 C_{ij}^{-1} a_{hi} a_{kj} \tilde{V}_{hk} & \text{if } h \neq k \end{cases} \quad (27)$$

Furthermore, it is mathematically expedient to add and then subtract V into the definition of D such that:

$$D = -W_v \bullet (V + \tilde{V}) + V \quad (28)$$

where \bullet is the Hadamard product and all sensor dependent terms have been bundled into:

$$W_v = (\mathbf{I}_{(N-1) \times (N-1)} - MC^{-T}M^T) \otimes \mathbf{I}_{2 \times 2} \quad (29)$$

The matrix M is formed of the 2 to N rows of A , the matrix of per channel Stokes weighting parameters from (3). The bound on the shift estimates can now be expressed concisely:

$$B_v = (V - D)^{-1} = [W_v \bullet (V + \tilde{V})]^{-1} \quad (30)$$

As anticipated, \tilde{V} has indeed provided the constituents for the cross-covariance terms in B_v .

4.3. Simplified Stokes parameter bound

The impetus behind much of the preceding section was to avoid having to manipulate an unwieldy Fisher information matrix directly while still achieving the Cramer-Rao bound on the registration parameters. The goal in this section is to do the same for the bound on the Stokes image estimators, B_S . We propose that the variance bound on the Stokes parameter estimate for any one pixel in the image is of less interest than the average bound across the image. In turn, this calculation is significantly simplified by applying properties of the trace and of the Kronecker product as shown in appendix B, where the bound is derived in detail. What follows are the highlights of this derivation.

Define B_{S_i} , the covariance matrix for an estimator of Stokes image \mathbf{S}_i , to be a submatrix of B_S . Equivalently, let Γ_i be the submatrix of S corresponding to the Stokes image \mathbf{S}_i :

$$\Gamma_i^{-1} = \sigma^2 C_{ii}^{-1} \otimes \mathbf{I}_{p^2 \times p^2} \quad (31)$$

and define:

$$\Phi_i^{-1} = \sigma^2 C_i^{-1} \otimes \mathbf{I}_{p^2 \times p^2} \quad (32)$$

where

$$C_i^{-1} = [C_{i0}^{-1} \quad \dots \quad C_{i2}^{-1}] \quad (33)$$

then the average bound for Stokes image \mathbf{S}_i is defined to be:

$$\langle B_{S_i} \rangle = \frac{1}{p^2} \text{tr}(B_{S_i}) \quad (34)$$

Expanded out, the trace term is:

$$\text{tr}(B_{S_i}) = \text{tr}(\Gamma_i^{-1}) + \text{tr}[\Phi_i^{-1} (HB_v H^T) \Phi_i^{-T}] \quad (35)$$

and because of the implicit symmetry of V and \tilde{V} :

$$\begin{aligned} \text{tr} [\Phi_i^{-1} (HB_v H^T) \Phi_i^{-T}] &= \sigma^2 \text{vec} [W_{Si} \bullet (V + \tilde{V})]^T \text{vec} (B_v) \\ &= \sigma^2 \text{tr} [W_{Si} \bullet (V + \tilde{V}) B_v] \end{aligned} \quad (36)$$

where

$$W_{Si} = M (C_i^{-T} C_i^{-1}) M^T \otimes \mathbf{1}_{2 \times 2} \quad (37)$$

this result can be folded back into (34) to produce:

$$\langle B_{Si} \rangle = \sigma^2 C_{ii}^{-1} + \frac{\sigma^2}{p^2} \text{tr} [W_{Si} \bullet (V + \tilde{V}) B_v] \quad (38)$$

Hence, substantial simplification of the bound calculation is achieved. The similarities between $\langle B_{Si} \rangle$ and B_v are clear. The terms in B_S that are not contained in B_{Si} can be ignored because they do not influence the trace. Also, it is interesting to note how the sensor itself (realized by W_{Si} and W_v in B_v) plays opposing roles in equation (38) analogous to multiplication and division if this were a purely scalar case.

5. Biased estimators

A study of estimator bias in the registration of polarimetric imagery has not been addressed in the literature. However, there are several well known examples of estimator bias in traditional, polarization insensitive registration algorithms. In anticipation of this future work, what follows is a discussion of incorporating bias into the results from sections 4.3 and 4.2.

According to Scharf [12], the Cramer-Rao bound on the covariance matrix of a biased estimator is given by:

$$\text{Cov} [\hat{\theta}] \geq \Delta^T J^{-1} \Delta \quad (39)$$

where, J , is the Fisher information matrix (identical to the unbiased case), $\hat{\theta}$ is the estimate of θ and

$$\Delta = \frac{\partial}{\partial \theta} E [\hat{\theta}]^T \quad (40)$$

is the partial derivative of the expected value of the estimator. The expected value of $\hat{\theta}$ is estimator specific (i.e. registration algorithm specific) and, furthermore, the CRLB of a biased estimator may be higher or lower than that of an unbiased estimator. Note that in the unbiased case, $\Delta = \mathbf{I}$ and equation (39) reduces to equation (4). In what follows, a specific partition of Δ is defined via subscript. For instance, Δ_v refers to a partition of Δ dealing strictly with estimates of the registration parameters whereas Δ_{S_0} refers to estimates of the Stokes image S_0 .

The biased estimator bound for the registration parameters, \tilde{B}_v , is easily achieved by combining (30) with (39):

$$\tilde{B}_v = \Delta_v^T [W_v \bullet (V + \tilde{V})]^{-1} \Delta_v \quad (41)$$

which is made possible by exchanging J^{-1} in (39) with its submatrix B_v (from the unbiased case). In what appears to be a very small step, this equation shows that the CRLB can be decomposed into an scene specific part, $V + \tilde{V}$, a sensor specific part W_v , and an estimator specific part, Δ_v . This separation may not be complete in that the channel spacing effects $V + \tilde{V}$ and, more than likely, bias in the estimator will be to some extent affected by scene polarimetric content.

Without the context of a specific registration algorithm, interpretation of the biased CRLB for a Stokes estimator is less straightforward. Combining (35) with (39) reveals that the trace on the biased Stokes bound is:

$$\text{tr} (\tilde{B}_{Si}) = \text{tr} (\Delta_{Si}^T \Gamma_i^{-1} \Delta_{Si} + \Delta_{Si}^T \Phi_i^{-1} (HB_v H^T) \Phi_i^{-T} \Delta_{Si}) \quad (42)$$

which may be reduced somewhat by substituting $\Delta_{S_i}^T \Phi_i^{-1}$ in for Φ_i^{-1} in appendix B and noting that:

$$\Phi_i^{-T} \Delta_{S_i} \Delta_{S_i}^T \Phi_i^{-1} = \sigma^4 C_i^{-T} C_i^{-1} \otimes \Delta_{S_i} \Delta_{S_i}^T \quad (43)$$

resulting in:

$$tr(\tilde{B}_{S_i}) = tr(\Delta_{S_i}^T \Gamma_i^{-1} \Delta_{S_i}) + \sigma^4 tr(H^T (C_i^{-T} C_i^{-1} \otimes \Delta_{S_i} \Delta_{S_i}^T) H B_v) \quad (44)$$

which is, if not quite as simple as the registration parameter case, certainly more computationally efficient than directly inverting the FIM. In terms of interpretation, the trace is composed of the sum of two terms, the first of which depends only on the estimator and the sensor. Later, in section 6.2, it is shown that the equivalent to this term in the unbiased case dominates the solution to the problem of an optimum channel spacing. It will be left to future work to show that this is the case for specific biased estimators.

Though there are currently no studies of bias in the registration of polarimetric imagery, the interested reader will be well served by referring to [9] for an in depth discussion of bias in gradient based registration algorithms for polarization insensitive imagery.

6. Example bound calculations

In the examples that follow, an unbiased estimator is assumed throughout.

6.1. Bounds on polarization insensitive imagery

Bounds on polarization insensitive imagery are presented here as a special case of the results derived in section 4. This case serves as both a practical example of how the Cramer-Rao bound can be applied effectively and as a bridge between this work and the work of Robinson and Milanfar.

We seek the bound on a joint estimator of a scene and the registration parameters amongst N noisy samples of this image. Recall that S_0 represents total scene intensity, therefore, S_0 is the only Stokes image of interest. In this scenario, the matrix in equation (19) reduces to a scalar, $C = N$, because there is only one parameter per pixel to be estimated. The channel transmission coefficient, a_{i0} , is assumed to be unity for each channel. In that case, $M = \mathbf{1}_{(N-1) \times 1}$. Therefore, each of W_v and W_{S_0} are the $2(N-1) \times 2(N-1)$ matrices:

$$(W_v)_{ij} = \begin{cases} 1 - \frac{1}{N} & \text{if } i = j, \\ -\frac{1}{N} & \text{if } i \neq j \end{cases} \quad (45)$$

and

$$W_{S_0} = \frac{\sigma^2}{N^2} \otimes \mathbf{1}_{2(N-1) \times 2(N-1)}. \quad (46)$$

To reiterate an important point, both V and \tilde{V} are image dependent. That being said, the overall bound's behavior with increasing N can be predicted by the behavior of W_v and W_{S_0} for any image. Equation (45) shows that $B_v = 2V^{-1}$ in the two channel case and $B_v = V^{-1}$ in the limit of N . Essentially, the \tilde{V} matrix is suppressed by the $1/N$ terms in W_v as N increases. These suppressed terms represent the decreasing influence of each individual image as the true underlying image emerges. In this limit, the bound parameters for each image asymptotically achieve the bound predicted by Robinson and Milanfar. Assuming a few images do not differ substantially from the rest of the ensemble (e.g. due to a large translational error or parallax) then the results follow a $(1 - \frac{1}{N})^{-1}$ progression.

Likewise, the progression of the estimated image toward the true image can be ascertained from the behavior of $\langle B_{S_0} \rangle$. Again, we consider the endpoints. In the two channel case:

$$\langle B_{S_0} \rangle|_{N=2} = \frac{\sigma^2}{p^2} + \frac{\sigma^2}{2} \quad (47)$$

and the large N case:

$$\langle B_{S_0} \rangle|_{N \rightarrow \text{large}} = \frac{\sigma^2}{p^2} \frac{2(N-1)}{N^2} + \frac{\sigma^2}{N} \quad (48)$$

It would appear that the behavior of this bound is dominated by $\frac{\sigma^2}{N}$. As shown in figure (3), the broad applicability of the preceding observations can be verified experimentally.

Twenty realizations of two test images, each with unique spatial characteristics, are generated with white Gaussian noise statistics and an uniformly distributed random shift error between ± 5 pixels in any direction. The scene average signal-to-noise ratio is 3. The bound on the average pixel value estimates, $\langle B_{S_0} \rangle$, and the average misregistration bound (i.e. $tr(B_v)/(N-1)$) are normalized and plotted against the number of frames, N . Normalization is carried out to illustrate the $\frac{1}{N}$ behavior of the intensity variance bound and the $(1 - \frac{1}{N})^{-1}$ behavior of the shift estimator bound. For comparison, each of these predicted curves are shown as a red dashed line underneath the data.

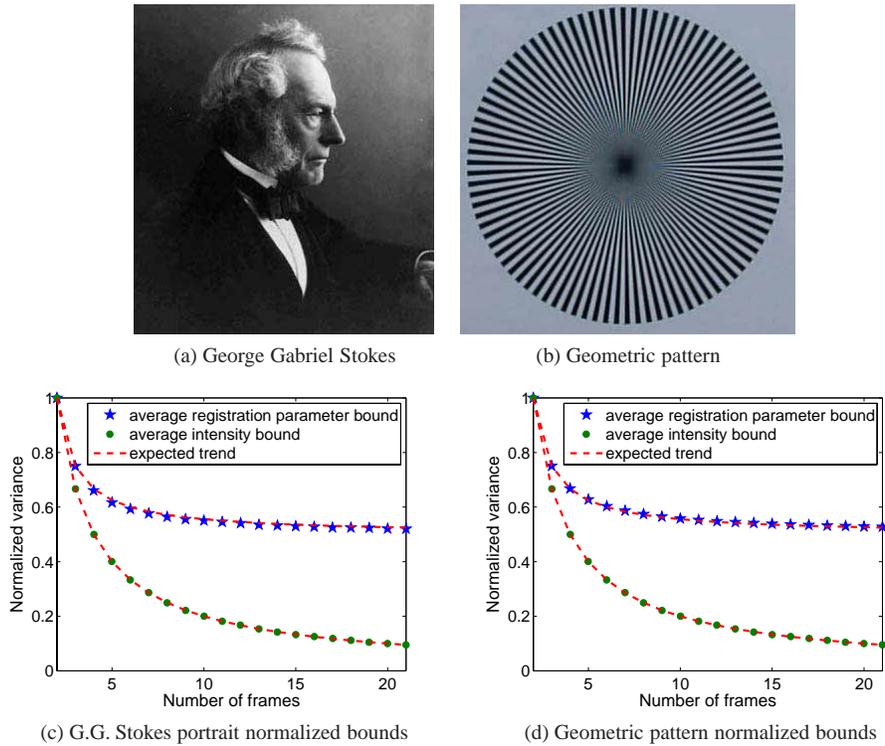


Fig. 3. Polarization insensitive imagery and its estimation bounds

The expected behavior of the bound with increasing N is confirmed for these disparate examples. The deviation of average registration parameter bound in the G.G. Stokes image from the expected trend demonstrates the slight influence of changes in image overlap area on the results.

The bound characteristics are determined by varying the angular spacing between polarization elements in one degree increments. The separation angle between channels is measured with respect to the first channel, which is held fixed. Each channel is evenly spaced and, in the plots that follow, this channel spacing is used to index the bound results. For example, the reader can infer that a channel spacing of 5° corresponds to an orientation of 5° , 10° , and 15° for the second, third, and fourth channels with respect to the first. Figure (6) shows the average bound on the Stokes parameters for each of the three test cases. Each plot is normalized by σ^{-2} since it may be divided out of equation (38).

As should be expected, the largest variances in Stokes image estimates occur when the polarization channels are closely spaced. Somewhat less expected is the observation that scene content appears to have little influence on individual results for the Stokes estimator in terms of the overall trends. Though there are small scale differences, the Stokes parameter bounds for all images closely follow the evolution of C_{ii}^{-1} with increasing channel spacing. This result is significant for two reasons. First, there appears to be global agreement as to which channel spacing is most preferable, at least for the test cases sampled here. Second, it would appear that each B_{S_i} is very well approximated by direct interrogation of S^{-1} , at least in an average sense. In other words, the Stokes parameter bounds are effectively scene independent with respect to channel spacing. Recall that S^{-1} is, by itself, the Cramer-Rao bound on the Stokes estimates when perfect knowledge of the misregistration parameters is available, consequently, these observations apply equally to that scenario.

The bound for the S_1 and S_2 estimates meet at Tyo's predicted optimal channel spacing of 45° , however, S_1 has a minimum bound at a somewhat closer channel spacing. Consequently, there is no global minimum bound for all Stokes parameters. Rather, the 45° spacing is the point where there is no preferred parameter. This point is significant because, as stated previously, the Stokes parameters are defined with respect to some coordinate system and, as this system changes in relation to the target (e.g. through camera motion) then scene content can shift between S_1 and S_2 . With this qualifier, a 45° channel spacing can be said to be optimal for a four channel system.

Figure (6) also contains a plot of the average registration parameter bound for each of the three channels. As before, σ^2 has been normalized out. Unlike the bounds on the Stokes parameters, these bounds depend both on scene content and channel orientation. Consistent with the Robinson and Milanfar analysis of V in the polarization insensitive two channel case, the difference in bound magnitude correlates with the amount of high spatial frequency content in the test images. The PC board image, with its multi-faceted geometric features, generates the lowest bound while the galaxy Markarian 3, which has largely diffuse features, generates the highest bound.

6.3. The three channel, three Stokes case

The three channel case is important because it is the minimum number of channels required to completely describe linear polarization (i.e. S_0 , S_1 , S_2). The interested reader may easily verify this three channel prerequisite for themselves by attempting to calculate B_S with only two channels. Three channel polarimeters are also common in practice; all examples in the previous section were originally collected by different three channel systems. In this section, the three channel polarimeter is examined from a joint estimation perspective.

First, consider the bound matrix B_v for some combination of three polarimeter channels represented by weight matrix A . Immediately, a mathematical difficulty arises:

$$W_v = \mathbf{0}_{4 \times 4} \quad (49)$$

in which case equation (30) is uninvertible or, in other words, the CRLB for the joint estimator

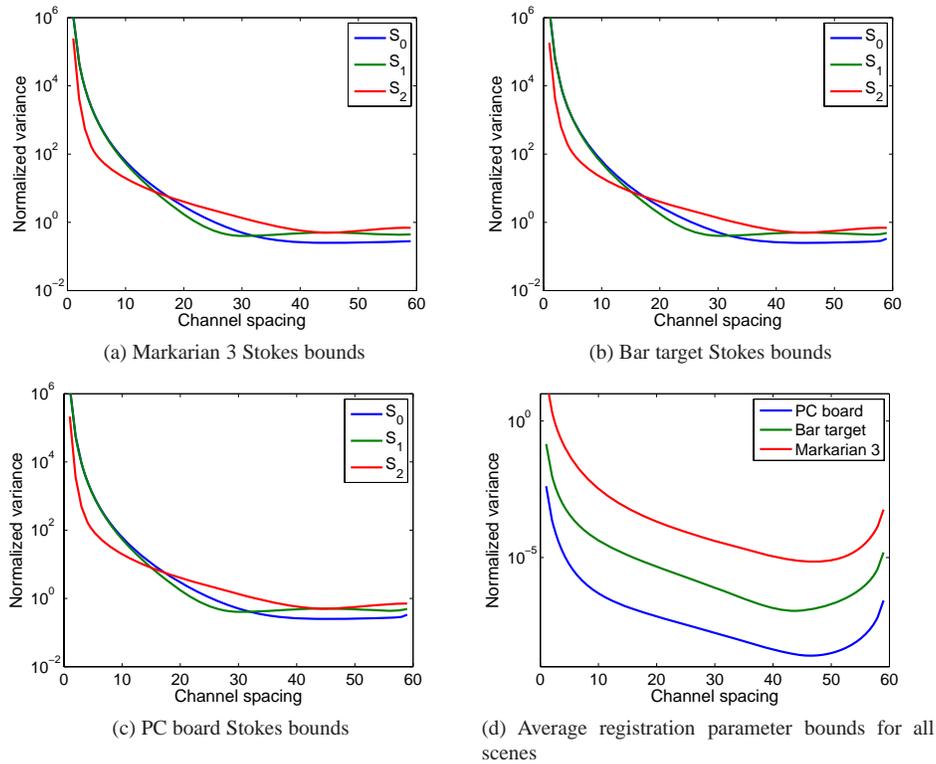


Fig. 6. Bound results for the four channel case

(biased or unbiased) is infinity.

At first glance, the result in (49) appears to be inconsistent with the fact that, three channel polarimeters are routinely employed in practice. The key difference, however, is that the data from these systems are not reduced using *joint* estimation algorithms. Rather, shift estimation and Stokes estimation are treated as separate problems. In other words, external measurement of the registration parameters is always required in the three channel case.

In terrestrial remote sensing applications, it is often the case that much of the scene is dominated by weakly polarized content (an obvious exception is a target scene composed over the water). In this case, the channel-to-channel registration problem for each pair of images is well approximated by the two channel polarization insensitive case described in section (6.1).

Assuming external measurement of the registration parameters by whatever means, there is likely a regime where S^{-1} will dominate equation (24), analogous to the four channel joint estimation examples. Following the prescribed method in the previous section, the diagonal terms in J_S^{-1} will be jointly minimized for S_1 and S_2 when the angular separation between channels is 60° . In this sense, the Cramer-Rao bound for the Stokes estimates can be compared across the three and four channel cases. Note from the table below that S_0 follows the $\frac{1}{N}$ pattern established for the polarization insensitive case while S_1 and S_2 go by $\frac{2}{N}$.

7. Conclusion

In this paper, the Cramer-Rao lower bound is derived for the problem of jointly estimating Stokes images and random misregistration parameters in the presence of channel noise. Direct

Table 1. Average bounds on Stokes parameter estimates in the S^{-1} dominant regime

Parameter	3-channel case	4-channel case
S_0	$\frac{1}{3}\sigma^2$	$\frac{1}{4}\sigma^2$
S_1	$\frac{1}{3}\sigma^2$	$\frac{1}{2}\sigma^2$
S_2	$\frac{1}{3}\sigma^2$	$\frac{1}{2}\sigma^2$

inversion of the prohibitively large Fisher information matrix is bypassed by applying matrix theory to express the resulting bounds in a tractable form. The effects of estimator bias on the bound are discussed up to the point where it becomes necessary to specify a specific estimator. The bound is then used to describe three and four channel polarimetric imaging systems as well as the special case of registering N polarization insensitive images. The bound itself is a useful evaluation metric because it incorporates both sensor and estimation algorithm effects and can be used to describe these effects theoretically in a way that pure simulation can not. In addition, the following general conclusions can be drawn from the results:

- The minimum achievable estimator variance is scene and channel orientation dependent in the unbiased case. The estimator itself also effects the bound in the biased case.
- The bounds on the registration and Stokes parameters are dependent.

Furthermore, specific conclusions can be drawn from the section 6 case studies:

- In the polarization insensitive case, the Cramer-Rao bound derived by Robinson and Milanfar is shown to be the asymptotic limit of the joint estimation bound as the number of frames approach infinity.
- In the three channel case, the Cramer-Rao bound for any joint estimator is infinity. Consequently, the registration parameters and Stokes images must be estimated separately.
- The form of the bound suggests that the optimum channel spacing is 60° and 45° respectively in the three and four channel cases. Optimum, in this context, refers to a joint minimum bound for S_1 and S_2 . These arrangements do not guarantee that the bound on the registration estimator is minimized.

This paper opens the possibility for substantial future work. This research could follow the lead of Robinson and Milanfar [9] and address the problem of registration estimator bias for a specific algorithm as it applies to polarimetric imagery. Additionally, Yetik and Nehorai's work could be expanded to include a polarimetric bound for other types of image deformation (e.g. rotation, shearing, etc.) or bounding estimator error for polarimetric registration algorithms that use feature extraction techniques. Additionally, a search could be conducted for a joint estimator that meets the Cramer-Rao bound. Finally, the idea of representing bounds on image registration as a realization of underlying parameters (in this case, as the weighted sum of the Stokes parameters) may be used to derive bounds for registering multi-modal images if a similar underlying relationship exists.

A. Appendix: More on the V and \tilde{V} matrices

The matrices V and \tilde{V} play a central role in the bound calculation. In addition, the V matrix provides the connection between the previous work by Robinson and Milanfar [9] and the generalization provided here. Consequently, some useful properties of these matrices are developed in this section.

First, there is an important connection between the V , \tilde{V} , and H_{ij} :

$$(H_{ij})^T H_{kj} = \frac{a_{ji}a_{jk}}{\sigma^4} \left(\frac{\partial}{\partial \mathbf{v}_j} \mathbf{f}_j^T \right) \left(\frac{\partial}{\partial \mathbf{v}_j} \mathbf{f}_j^T \right)^T = \frac{a_{ji}a_{jk}}{\sigma^2} V_{jj} \quad (50)$$

and

$$(H_{lj})^T H_{ki} = \frac{a_{jl}a_{ik}}{\sigma^4} \left(\frac{\partial}{\partial \mathbf{v}_j} \mathbf{f}_j^T \right) \left(\frac{\partial}{\partial \mathbf{v}_i} \mathbf{f}_i^T \right)^T = \frac{a_{jl}a_{ik}}{\sigma^2} \tilde{V}_{ji} \quad (51)$$

for all $i \neq j$.

Second, each \mathbf{v}_i is established in the plane of the image. Therefore, a common direction vector, \mathbf{x} , for all relevant derivatives can be defined via the chain rule:

$$\frac{\partial}{\partial \mathbf{v}_i} \mathbf{f}_i^T = \frac{\partial}{\partial \mathbf{x}} \mathbf{f}_i^T. \quad (52)$$

Consequently, each entry in V and \tilde{V} is defined in a common coordinate system.

B. Appendix: Detailed derivation of the Stokes parameter bound

The average bound can be determined by calculating the trace of equation (24) via application of the following properties of the trace and the Kronecker product. For the trace [19]:

$$\text{tr}(A + G) = \text{tr}(A) + \text{tr}(G) \quad (53a)$$

$$\text{tr}(CGD) = \text{tr}(DCG) \quad (53b)$$

$$\text{tr}(EF) = \text{vec}(E^T)^T \text{vec}(F) \quad (53c)$$

where A and G are square matrices and C , D , E , and F are any matrices such that CGD and EF are square matrices. Operator tr represents the trace and, for any matrix A , $\text{vec}(A)$ is defined to be an ordered stack of the columns of A . For the Kronecker product (also from [19]):

$$\text{vec}(AGC) = (C^T \otimes A) \text{vec}(G) \quad (54a)$$

$$(A \otimes C)(G \otimes D) = (AG \otimes CD) \quad (54b)$$

both of which always hold whenever AGC , AG , and CD are defined. Finally, note that, in contrast to regular matrix multiplication:

$$(A \otimes G)^T = A^T \otimes G^T \quad (55a)$$

$$(A \bullet G)^T = A^T \bullet G^T \quad (55b)$$

Armed with the preceding formulae, derivation of the average bound for each Stokes image is straightforward. Starting with the second term in equation (35):

$$\begin{aligned} \text{tr}[\Phi_i^{-1} (HB_v H^T) \Phi_i^{-T}] &= \text{tr}[\Phi_i^{-T} \Phi_i^{-1} (HB_v H^T)] \\ &= \text{vec}(\Phi_i^{-T} \Phi_i^{-1})^T \text{vec}(HB_v H^T) \\ &= \text{vec}(\Phi_i^{-T} \Phi_i^{-1})^T (H \otimes H) \text{vec}(B_v) \\ &= \text{vec}[H^T (\Phi_i^{-T} \Phi_i^{-1}) H]^T \text{vec}(B_v) \end{aligned} \quad (56)$$

Fortunately, expression $H^T (\Phi_i^{-T} \Phi_i^{-1}) H$ is identical in form to equation (25) and can therefore be simplified in the same manner:

$$\begin{aligned} H^T (\Phi_i^{-T} \Phi_i^{-1}) H &= \sigma^4 H^T \left(C_i^{-T} C_i^{-1} \otimes \mathbf{I}_{p^2 \times p^2} \right) H \\ &= \sigma^2 W_{Si} \bullet (V + \tilde{V}) \end{aligned} \quad (57)$$

where

$$W_{Si} = M (C_i^{-T} C_i^{-1}) M^T \otimes \mathbf{1}_{2 \times 2} \quad (58)$$

Invoking the implicit symmetry of W_{Si} , V , and \tilde{V} :

$$\begin{aligned} \text{tr} [\Phi_i^{-1} (H B_v H^T) \Phi_i^{-T}] &= \sigma^2 \text{vec} [W_{Si} \bullet (V + \tilde{V})]^T \text{vec} (B_v) \\ &= \sigma^2 \text{tr} [W_{Si} \bullet (V + \tilde{V}) B_v] \end{aligned} \quad (59)$$

This result can be folded back into (34) to produce:

$$\begin{aligned} \langle B_{Si} \rangle &= \frac{1}{p^2} [\text{tr} (\Gamma_i^{-1})] + \frac{\sigma^2}{p^2} \text{tr} [W_{Si} \bullet (V + \tilde{V}) B_v] \\ &= \sigma^2 C_{ii}^{-1} + \frac{\sigma^2}{p^2} \text{tr} [W_{Si} \bullet (V + \tilde{V}) B_v] \end{aligned} \quad (60)$$

which is identical to equation (38).

Acknowledgments

Thanks to R. Martin for gently suggesting this course of research by including it in my qualifying exam, to M. Oxley for introducing the Kronecker product, and to S. Cain for his guidance and perpetual encouragement. Calibration data for the HST imagery was graciously provided by M. Kishimoto. Some of the data presented in this paper were obtained from the Multimission Archive at the Space Telescope Science Institute (MAST). STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. The opinions and views expressed by the author are not necessarily those of the United States Air Force or the Department of Defense.