THE WISHART-KOTZ CLASSIFIER FOR MULTILOOK POLARIMETRIC SAR DATA

Paul R. Kersten,

No affiliation
Utah, USA

Stian N. Anfinsen, Anthony P. Doulgeris

University of Tromsø
Department of Physics and Technology
N-9037 Tromsø, Norway

ABSTRACT

We presents a classifier for multilook polarimetric synthetic aperture radar (PolSAR) data based upon a new distribution model for the polarimetric sample covariance matrix: the complex Wishart-Kotz distribution. This is a highly flexible model, which exhibits the heavy tails needed to fit the data found in high resolution PolSAR images. In addition, they do not contain the mathematical special functions that limit the usefulness of alternative heavy-tailed distributions by inflicting high computational cost and numerical instability. Classification results on simulated and real data are presented.

Index Terms— Radar imaging, Polarimetric synthetic aperture radar, probability distributions, parameter estimation, statistical analysis

1. INTRODUCTION

Classification of the surface cover is an important task in the analysis of PolSAR images, which has attracted much interest. Although many nonparametric approaches have been tested, classifiers based on parametric models for the data probability density function (pdf) remain the most popular, for several reasons: The statistical models are often founded in a physical description of the backscattering process; They provide good fit with the data; Furthermore, the resulting classifiers are rooted in well known methods from the pattern recognition literature, such as maximum likelihood and maximum a posteriori classification, and also finite mixture models solved with the expectation-maximisation (EM) algorithm.

The Wishart classifer [1] is unarguably the most used classifier for multilook polarimetric data, both due to its ease of implementation, low computational cost and availability in software suites. Improved algorithms have been produced by using more accurate pdfs and also including information about spatial context [2, 3, 4, 5]. The improvements have come at the expense of higher computational cost and complicated implementation, since the heavy-tailed distributions contain special functions such as the Bessel K function and the Kummer U function, whose approximations are required to be both exact and smooth over the entire domain of the function arguments. As an alternative to these models, we have sought an extension of the complex Wishart distribution, analogous to how the generalised gamma distribution (GGD) extends the ordinary gamma distribution.

The desired model was found in the complex Wishart-Kotz (WK) distribution, first described in [6], whereas this paper further extends the parameter range and the flexibility. An important feature of this model is that the pdf contains only exponential functions, determinants, traces and powers, which assures a low computational cost. Moreover, the matrix log-cumulant equations [7] are relatively simple and provide efficient parameter estimators through a method of moment approach. When the model is portrayed in a log-cumulant diagram, it is seen to be more flexible than the U distribution [5], and reduces to special cases that closely approximate the popular K distribution [8] and G^0 distribution [9]. The WK distribution thus encompasses the characteristics of natural surfaces, and provides a very efficient and versatile model.

The paper is structured as follows: Background theory is given in Section 2, including expressions for the the probability density function (pdf) and the log-cumulants of the new distribution. These provide us with data models and parameter estimators that can be directly inserted into known classification algorithms. In Section 3 we discuss the implementation of the model in two particular algorithms: the maximum likelihood classifier, which parallels the well-known Wishart classifer, and a recently developed EM type clustering algorithm [2] which has been further extended with a Markov random field (MRF) stage to incorporate contextual information [3]. In Section 4 we test these classifiers with different data models. Conclusions are given in Section 5.

2. DATA MODELS

Let $\mathbf{s} \in \mathbb{C}^d$ be the target vector holding the complex scattering coefficients measured in d polarimetric channels. Thus, $\mathbf{C} = (1/L)\sum_{\ell=1}^{L} \mathbf{s}_\ell \mathbf{s}_\ell^\dagger$ is a $d \times d$ sample covariance matrix produced from a multilook set of L vectors, $\{\mathbf{s}_\ell\}_{\ell=1}^{L}$, where $(\cdot)^\dagger$ denotes the conjugate transpose. The random matrix $\mathbf{C}$ is defined on the cone of positive definite Hermitian matrices, denoted as Ω_+.

2.1. Density Models

Let $\mathbf{s}$ be zero mean, and assume first that $\mathbf{s}$ follows the complex circular multinormal distribution with pdf

$$p_{\mathbf{s}}(\mathbf{s};\boldsymbol{\Sigma}) = \frac{1}{\pi^d|\boldsymbol{\Sigma}|}\exp(-\mathbf{s}^\dagger\boldsymbol{\Sigma}\mathbf{s})\,, \tag{1}$$

where $\boldsymbol{\Sigma} = \mathrm{E}\{\mathbf{s}\mathbf{s}^H\}$ and $|\cdot|$ is the determinant. Then $\mathbf{C}$ follows the scaled complex Wishart distribution [7], with pdf

$$p_{\mathbf{C}}(\mathbf{C};L,\boldsymbol{\Sigma}) = \frac{L^{Ld}}{\Gamma_d(L)}\frac{|\mathbf{C}|^{L-d}}{|\boldsymbol{\Sigma}|^L}\exp(-L\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{C}))\,, \tag{2}$$

where $\mathrm{tr}(\cdot)$ is the trace operator and $\Gamma_d(\cdot)$ the multivariate gamma function of the complex kind [7].

Next assume that $\mathbf{s}$ follows a complex Kotz-type distribution. This is a generalisation of (1), whose real version is discussed in [10, 11, 12]. Its pdf is

$$p_{\mathbf{s}}(\mathbf{s};\boldsymbol{\Sigma},\beta,\rho) = \frac{1}{\pi^d|\boldsymbol{\Sigma}|}h(\mathbf{s}^\dagger\boldsymbol{\Sigma}\mathbf{s})\,, \tag{3}$$

with the generating function, $h(x)$, defined as

$$h(x) = \frac{\rho\,\Gamma(d)}{\Gamma\left(\frac{d+\beta-1}{\rho}\right)}x^{\beta-1}\exp(-x^\rho)\,. \tag{4}$$

Due to limited space, we state without proof that the resulting density for $\mathbf{C}$ is the WK distribution given in [6] as

$$\begin{aligned} p_{\mathbf{C}}(\mathbf{C};L,\boldsymbol{\Sigma},\gamma,\rho) = c\,\frac{|\mathbf{C}|^{L-d}}{|\boldsymbol{\Sigma}|^L}\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{C})^{\gamma-Ld} \\ \times\exp\big(-\,[L\,\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{C})]^\rho\big)\,, \end{aligned} \tag{5}$$

with the constant normalisation factor

$$c = \frac{|\rho|\,L^\gamma}{\Gamma\left(\frac{\gamma}{\rho}\right)}\frac{\Gamma(Ld)}{\Gamma_d(L)} \tag{6}$$

where $\Gamma(\cdot)$ is the standard gamma function, while $\gamma = Ld + \beta - 1$ and ρ are shape parameters. The scale matrix parameter $\boldsymbol{\Sigma} \in \Omega_+$ is related to $\mathbf{C}$ through

$$\mathrm{E}\{\mathbf{C}\} = \frac{\Gamma\left(\frac{\gamma+1}{\rho}\right)}{\Gamma\left(\frac{\gamma}{\rho}\right)}\frac{\boldsymbol{\Sigma}}{Ld}\,. \tag{7}$$

The domain of γ and ρ were in [6] restricted to $\mathbb{R}^+$, but is now extended to $\mathbb{R}$, with the additional requirement that the shape parameters must have the same sign. The detailed derivation will be presented in a future journal paper.

Remark that when $d = 1$ and $\gamma = L$, (5) reduces to the generalised gamma distribution, which has been promoted as a model for single polarisation multilook SAR data [13].

2.2. Parameter Estimation

The scale matrix is estimated from (7), whereas the shape parameters are estimated from cumulants of $\ln|\mathbf{C}|$ by the method of matrix log-cumulants [7]. Let $\kappa_\nu\{\mathbf{C}\}$ be the νth-order matrix log-cumulant for a given density model. These can be derived as

$$\kappa_\nu\{\mathbf{C}\} = \frac{d^\nu}{ds^\nu}\ln\phi_{\mathbf{C}}(s)\bigg|_{s=d}\,, \tag{8}$$

with the Mellin kind characteristic function (mkcf) of $\mathbf{C}$,

$$\phi_{\mathbf{C}}(s) = \mathcal{M}\{p_{\mathbf{C}}(\mathbf{C})\}(s) = \mathrm{E}\{|\mathbf{C}|^{s-d}\}\,, \tag{9}$$

defined as the matrix-variate Mellin transform of $p_{\mathbf{C}}(\mathbf{C})$, where $s \in \mathbb{C}$ is the transform variable.

Let $\mathbf{W}$ be a Wishart matrix which follows the pdf in (2). Then it has the mkcf [7]

$$\phi_{\mathbf{W}}(s) = \frac{|\boldsymbol{\Sigma}|^{s-d}}{L^{d(s-d)}}\frac{\Gamma_d(s+L-d)}{\Gamma_d(L)}\,. \tag{10}$$

and the matrix log-cumulants

$$\kappa_\nu\{\mathbf{W}\} = \begin{cases} \psi_d^{(0)}(L) + \ln|\boldsymbol{\Sigma}| - d\ln L & ;\ \nu = 1\,, \\ \psi_d^{(\nu-1)}(L) & ;\ \nu > 1\,. \end{cases} \tag{11}$$

The mkcf for a WK distributed $\mathbf{C}$ is

$$\phi_{\mathbf{C}}(s) = \phi_{\mathbf{W}}(s)\,\frac{\Gamma(Ld)}{\Gamma(d(s+L-d))}\frac{\Gamma\left(\frac{d(s-d)+\gamma}{\rho}\right)}{\Gamma\left(\frac{\gamma}{\rho}\right)}\,, \tag{12}$$

with $\phi_{\mathbf{W}}(s)$ identified as a factor. The matrix log-cumulants of $\mathbf{C}$ follow from (8) as [7]

$$\kappa_\nu\{\mathbf{C}\} = \kappa_\nu\{\mathbf{W}\} + d^\nu\kappa_\nu\{\tau\}\,. \tag{13}$$

They decompose as the speckle contribution, $\kappa_\nu\{\mathbf{W}\}$, and the texture log-cumulants (TLC) [6]

$$\kappa_\nu\{\tau\} = \frac{1}{\rho^\nu}\,\psi^{(\nu-1)}\left(\frac{\gamma}{\rho}\right) - \psi^{(\nu-1)}(Ld)\ ;\ \nu \geq 1\,. \tag{14}$$

In the estimation procedure, we extract the parameters that produce the best match between the model TLCs in (14) and the sample TLCs, $\langle\kappa_\nu\{\tau\}\rangle$, evaluated as

$$\langle\kappa_\nu\{\tau\}\rangle = \frac{\langle\kappa_\nu\{\mathbf{C}\}\rangle - \kappa_\nu\{\mathbf{W}\}}{d^\nu}\,. \tag{15}$$

where the $\langle\kappa_\nu\{\mathbf{C}\}\rangle$ are sample log-cumulants of $\mathbf{C}$. These are computed from the sample log-moments:

$$\langle\mu_\nu\{\mathbf{C}\}\rangle = \frac{1}{n}\sum_{i=1}^{n}(\ln|\mathbf{C}_i|)^\nu\,. \tag{16}$$

using a population of n sample covariance matrices, $\{\mathbf{C}_i\}_{i=1}^n$,

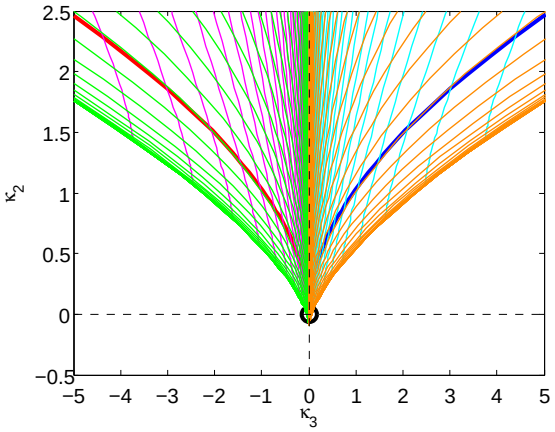

Fig. 1. Texture log-cumulant diagram showing the moments space spanned by the second and third-order TLCs, $\kappa_2\{\tau\}$ and $\kappa_3\{\tau\}$, that can be attained under the WK model.

and the relations found in [7, Eq. (76)-(78)].

We must use a minimum of two log-cumulant equations to invert numerically for the shape parameters γ and ρ. Using more than two equations yields an overdetermined system, and the estimation takes the form of an optimisation problem, in a procedure known as the generalised method of moments [14]. We determine the sign of the shape parameters from

$$\operatorname{sgn}(\gamma) = \operatorname{sgn}(\rho) = -\operatorname{sgn}\left(\langle\kappa_3\{\tau\}\rangle + \psi^{(2)}(Ld)\right), \quad (17)$$

where $\operatorname{sgn}(x) = x/|x|$ is the sign operator, and use this information to limit the search space to $\mathbb{R}^-$ or $\mathbb{R}^+$ in each case.

2.3. Log-cumulant Diagram

Figure 1 shows how the WK model appears in the texture log-cumulant diagram spanned by the TLCs $\kappa_3\{\tau\}$ and $\kappa_2\{\tau\}$. The WK distribution covers the entire gridded area, which demonstrates the flexibility of the model. The K distribution and the G^0 are indicated by the red and the blue curve, respectively. The Wishart model lies at the origo. The green and orange curves represent WK distributions with fixed values of ρ, when $\rho > 0$ and $\rho < 0$, respectively. The magenta and cyan curves represent WK distributions with fixed values of γ, when $\gamma > 0$ and $\gamma < 0$, respectively.

3. CLASSIFICATION ALGORITHMS

The proposed distribution can be applied to many existing model-based image analysis algorithms. Below, we look at two classification algorithms that have been used in the PolSAR literature with alternative density models, and discuss them in the context of the WK distribution.

3.1. Maximum Likelihood Classifier

The WK model can be inserted into a maximum likelihood classification algorithm, by analogy with the Wishart classifier [1] based on (2). This leads to the discriminant function

$$d_{WK}(\mathbf{C}, \boldsymbol{\Sigma}) = \ln\left(\frac{\Gamma\left(\frac{\gamma}{\rho}\right)}{|\rho| L^{\gamma}}\right) + L\ln|\boldsymbol{\Sigma}| + (\gamma - Ld)\ln(\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{C})) + (L\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{C}))^{\rho}, \quad (18)$$

which can be referred to as the WK distance. It is clearly not as simple as the Wishart distance [1],

$$d_W(\mathbf{C}, \boldsymbol{\Sigma}) = \ln|\mathbf{C}| + \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{C}), \quad (19)$$

but comprises a tractable alternative.

3.2. EM-based Contextual Clustering Algorithm

An unsupervised clustering algorithm has recently been developed [2, 3], which combines: (i) flexible non-Gaussian models for the pixelwise distribution of $\mathbf{C}$ with (ii) a Markov random field (MRF) model for contextual smoothing of the class labels, and (iii) goodness-of-fit tests that determine the appropriate number of statistically distinct classes. The classifier is implemented as an EM-type algorithm which iteratively assigns pixels to classes based on posterior class probabilities modified by contextual information and estimates the distribution parameters of the classes using membership weights. A test stage is inserted within the iterations, which assesses the model fit of each class and adapts the number of classes through a split-and-merge procedure. Full details are given in [2]. The extension with an MRF model is described in [3].

The iterations are run until convergence and the algorithm requires a massive number of pdf evaluations. On the other hand, the success of the unique mechanism for selecting the number of classes relies on having accurate density models. Thus, the value of having a flexible model with a low computational cost is obvious.

It is quite straightforward to integrate the WK model into the clustering algorithm by simply substituting the new pdf and log-cumulant expressions. The goodness-of-fit tests and parameter estimation already relies on matrix log-cumulant distances, and so the main framework remains unchanged. The most significant modification needed is that some suitable limits have been set on the search space for the shape parameters to avoid unrealistic parameter domains.

4. RESULTS

We have implemented the EM-based contextual clustering algorithm described in Section 3.2 with the WK distribution as the density model for $\mathbf{C}$. Some results are shown in Fig. 2 for simulated and real data in multilook polarimetric format.

The leftmost image is a Pauli RGB showing a simulated test pattern with six U distributed classes. The class param-

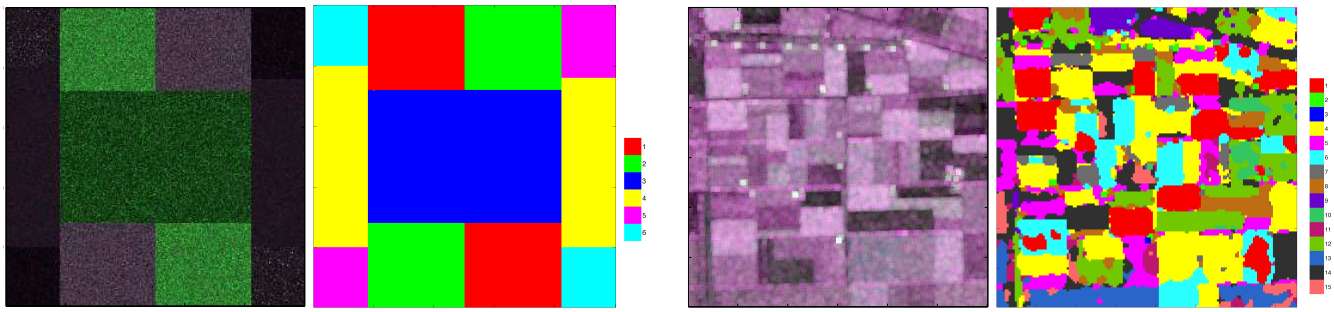

Fig. 2. Classifications obtained with the WK classifier for simulated test pattern and AIRSAR C-band image of Flevoland.

eters are extracted from real data to imitate realistic surface cover and the number of looks is $L = 5$. The WK classifier produces a perfect result, as seen in the classification map second from the left.

The second image from the right is a Pauli RGB of a Radarsat-2 C-band acquisition from 2008 of an agricultural area in Flevoland, The Netherlands (©MDA Ltd., 2008). It has been multilooked with a factor of $L = 25$. The result of the WK classifier is seen in the image at the far right. It has produced 15 classes. The segments are smooth, and they seem to correspond reasonably well to the patches in the Pauli RGB image, as judged by visual inspection.

5. CONCLUSIONS

We have derived a flexible and computationally efficient density model for the polarimetric covariance matrix, which provides a mathematically tractable generalisation of the complex Wishart distribution. The flexibility of the model is demonstrated by projecting it in a texture log-cumulant diagram. The implementation of the model in established classification algorithms have been described. Classification results with a state-of-the-art fully automatic clustering algorithm have been presented.

6. REFERENCES

[1] J.-S. Lee, M. R. Grunes, T. L. Ainsworth, D. L. Schuler, and S. R. Cloude, "Unsupervised classification using polarimetric decomposition and the complex Wishart distribution," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2249–2259, Sep. 1999.

[2] A. P. Doulgeris, S. N. Anfinsen, and T. Eltoft, "Automated non-Gaussian clustering of polarimetric synthetic aperture radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3665–3676, Oct. 2011.

[3] A. P. Doulgeris, V. Akbari, and T. Eltoft, "Automatic PolSAR segmentation with the U-distribution and Markov random fields," in *Proc. EUSAR 2012*, Nuremberg, Germany, 23–26 April 2012, pp. 183–186.

[4] A. C. Frery, A. H. Correia, and C. C. Freitas, "Classifying multifrequency fully polarimetric imagery with multiple sources of statistical evidence and contextual information," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3098–3109, Oct. 2007.

[5] L. Bombrun, "Hierarchical segmentation of polarimetric SAR images using heterogeneous clutter models," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 726–737, Feb. 2011.

[6] P. Kersten and S. N. Anfinsen, "A flexible and computationally efficient density model for the multilook polarimetric covariance matrix," in *Proc. EUSAR 2012*, Nuremberg, Germany, 23-26 April 2012, pp. 760–763.

[7] S. N. Anfinsen and T. Eltoft, "Application of the matrix-variate Mellin transform to analysis of polarimetric radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2281–2295, Jun. 2011.

[8] J.-S. Lee, D. L. Schuler, R. H. Lang, and K. J. Ranson, "K-distribution for multi-look processed polarimetric SAR imagery," in *Proc. IGARSS 1994*, vol. 4, Pasadena, USA, Aug. 1994, pp. 2179–2181.

[9] C. C. Freitas, A. C. Frery, and A. H. Correia, "The polarimetric G distribution for SAR data analysis," *Environmetrics*, vol. 16, no. 1, pp. 13–31, Feb. 2005.

[10] S. Nadarajah, "The Kotz-type distribution with applications," *Statistics*, vol. 37, no. 4, pp. 341–358, Jul. 2003.

[11] A. Sarr and A. K. Gupta, "Estimation of the precision matrix of multivariate Kotz type model," *J. Multivariate Anal.*, vol. 100, no. 4, pp. 742–752, Apr. 2009.

[12] J. A. Diaz-Garcia and R. Gutiérrez-Jáimez, "Compound and scale mixture of matricvariate and matrix variate kotz-type distributions," *J. Korean Statist. Soc.*, vol. 39, no. 1, pp. 75–82, Mar. 2010.

[13] H.-C. Li, W. Hong, Y.-R. Wu, and P.-Z. Fan, "On the empirical-statistical modeling of SAR images with generalized gamma distribution," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 386–397, Jun. 2010.

[14] A. R. Hall, *Generalized Method of Moments*. Oxford, UK: Oxford University Press, 2005.