EM ALGORITHM FOR ESTIMATING THE PARAMETERS OF A MULTIVARIATE COMPLEX RICIAN DENSITY FOR POLARIMETRIC SAR

Thomas L. Marzetta
Nichols Research Corporation
251 Edgewater Drive, Wakefield, MA 01880

ABSTRACT

A polarimetric synthetic aperture radar (SAR) forms a complex vector-valued image where each pixel comprises the polarization-dependent reflectivity of a portion of a target or scene. The most common statistical model for this type of image is the zero-mean, circularly-symmetric, multivariate, complex Gaussian model. A logical generalization of this model is a circularly-symmetric, multivariate, complex *Rician* model which results from having a nonzero-mean complex target reflectivity. Direct maximum-likelihood estimation of the Rician model parameters is infeasible, since setting derivatives equal to zero results in an intractable system of coupled nonlinear equations. The contribution of this paper is a complete iterative solution to the Rician parameter estimation problem by means of the EM (expectation-maximization) algorithm.

1. PROBLEM STATEMENT

A polarimetric SAR separately transmits horizontally and vertically polarized pulses and the receiver separately measures the horizontally and vertically polarized components of the returns, which after processing yield four complex-valued SAR images: HH, HV, VH, and VV [1]. Reciprocity implies that the VH and HV images should be equal, so the resulting polarimetric SAR image at a particular pixel can be denoted by a three-component complex vector, **x**. In turn **x** can be modelled as the three-component complex reflectivity, **y**, phase-shifted by an amount corresponding to the two-way range between the radar and the resolution cell,

$$\mathbf{x} = \mathbf{y}\cdot\exp(i\phi)\,. \tag{1}$$

Given the typically great disparity between the range resolution of the SAR and its wavelength it is usual to model the phase shift, ϕ, as random and uniformly distributed over $[0,2\pi]$. The most common model for polarimetric SAR assumes that the reflectivity, **y**, is zero-mean and Gaussian distributed which results in **x** having the zero-mean, circularly-symmetric, multivariate, complex Gaussian distribution [2]. The underlying assumptions about reflectivity are consistent with having a scattering surface that is rough on a scale comparable to the size of the resolution cell [3].

In cases where the resolution cell is dominated by one large scatterer a more reasonable model would assume that the reflectivity has a nonzero mean. One could still assume the presence of a zero-mean Gaussian component in the reflectivity to account for small scatterers. With this revised assumption the *conditional* density of the image is circularly-symmetric Gaussian with a nonzero mean,

$$p_{\mathbf{x}|\phi}(X|\phi) = \frac{1}{\pi^d \det K}\exp\left\{-[X-Ae^{i\phi}]^H K^{-1}[X-Ae^{i\phi}]\right\}, \tag{2}$$

where A and K are the mean and the covariance matrix respectively of **y**, d is the dimension (typically three), and the superscript "H" denotes "conjugate-transpose". Multiplying (2) by the probability density for ϕ and integrating gives the marginal density for **x** which is *not* Gaussian,

$$p_{\mathbf{x}}(X) = \frac{1}{\pi^d \det K}\exp\left\{-X^H K^{-1} X - A^H K^{-1} A\right\} \cdot I_0(2|A^H K^{-1} X|), \tag{3}$$

and $I_0(z)$ is the modified Bessel function of the first kind of order zero. It is convenient to refer to this density as circularly-symmetric, multivariate, complex Rician, recalling that the conventional Rician density describes the magnitude of **x** for the scalar case (d=1) [4].

 0-7803-2431-5/95 $4.00

To understand the relation between this new probability density and the classical Rician density, consider the density (3) for the case d=1, which is the joint probability density for the real and imaginary parts of the complex scalar, x,

$$p_x(X) = \frac{1}{\pi K}\exp\left\{-(|X|^2+|A|^2)/K\right\} \cdot I_0(2|A||X|/K). \quad (4)$$

Finding the joint probability density for the magnitude and phase of x (e.g. such that $\mathrm{Re}\{x\} = \rho\cos\theta$ and $\mathrm{Im}\{x\}=\rho\sin\theta$) by standard techniques, and integrating over θ gives the marginal probability density for the magnitude of x, which is the classical Rician density,

$$p_\rho(\rho) = \frac{2\rho}{K}\exp\left\{-(\rho^2+|A|^2)/K\right\}\cdot I_0(2|A|\rho/K). \quad (5)$$

The application of the Rician model to polarimetric SAR involves choosing the model parameters so that the model best represents the pixels that comprise some portion of the image. More formally, given N statistically-independent realizations of the random vector $\mathbf{x}$, $X = \{\mathbf{x}_n, 1\le n\le N\}$ (note that both the reflectivities and the translational phase shifts are statistically independent from pixel-to-pixel) the problem is to find the values of A and K that maximize the joint likelihood function according to the probability density (3),

$$\log p_X(X) = \sum_{n=1}^{N}\left\{-d\log\pi - \log\det K - X_n^H K^{-1} X_n - A^H K^{-1} A + \log I_0\left(2\left|A^H K^{-1} X_n\right|\right)\right\}. \quad (6)$$

Setting derivatives of the joint log-likelihood with respect to the parameters equal to zero yields a set of coupled equations with no closed-form solution. It turns out that the EM algorithm [5] is a tractable way to find the ML estimates.

2. EM SOLUTION FOR ML PARAMETER ESTIMATES

The application of the EM algorithm to this problem is based on the fact that possession of the unobserved (hidden) random phases would make this an easy estimation problem. The EM algorithm is explicitly iterative. It begins with an initial guess for the mean and the covariance, A_0 and K_0. At the beginning of the t-th iteration, the current estimates are A_{t-1} and K_{t-1}. The t-th iteration consists of an E-step followed by an M-step, both of which can be performed analytically.

The E-step involves taking the expectation of the joint log-likelihood of the observed and the hidden data, conditioned on the observed data and on the current parameter estimates. For the problem at hand the hidden random variable associated with the n-th pixel is ϕ_n, and its conditional probability density (applying Bayes' rule) is

$$p_{\phi_n|x_n;\, a_{t-1}\, K_{t-1}}(\phi_n|X_n;\, A_{t-1}, K_{t-1}) = \frac{\exp\left\{2\mathrm{Re}\left[e^{-i\phi_n}A_{t-1}^H K_{t-1}^{-1} X_n\right]\right\}}{2\pi I_0\left(2\left|A_{t-1}^H K_{t-1}^{-1} X_n\right|\right)}. \quad (7)$$

Performing the E-step gives

$$E_\phi\left\{\log p_{X\,\phi\,|\,a\,K}(X, \phi|A, K)\middle|X, A_{t-1}, K_{t-1}\right\} = \sum_{n=1}^{N}\Big[-\log(2\pi) - d\log\pi - \log(\det K) - X_n^H K^{-1} X_n - A^H K^{-1} A + 2\cdot\mathrm{Re}\left\{h_{t-1}^*(n)\cdot[A^H K^{-1} X_n]\right\}\Big], \quad (8)$$

where $h_{t-1}(n)$ is a scalar-valued weight,

$$h_{t-1}(n) = \frac{I_1\left(2\left|A_{t-1}^H K_{t-1}^{-1} X_n\right|\right)}{I_0\left(2\left|A_{t-1}^H K_{t-1}^{-1} X_n\right|\right)}\cdot\frac{(A_{t-1}^H K_{t-1}^{-1} X_n)}{\left|A_{t-1}^H K_{t-1}^{-1} X_n\right|}, \quad 1\le n\le N. \quad (9)$$

and $I_1(z)$ is the modified Bessel function of the first kind of order one.

The M-step involves choosing A and K to maximize the conditional expectation (8), resulting in the following updated estimates:

$$A_t = \frac{1}{N}\sum_{n=1}^{N} h^*_{t-1}(n)\cdot X_n, \qquad (10)$$

$$K_t = \frac{1}{N}\sum_{n=1}^{N} X_n X_n^H - A_t A_t^H. \qquad (11)$$

The EM algorithm produces a sequence of estimates whose likelihood is non-decreasing, and convergence to at least a local maximum is guaranteed.

3. COMMENTS AND INTERPRETATION

As usual, the EM iterative estimator combines both intuitively obvious as well as more subtle features. Consider first the A update (10), which is equal to the average of the weighted measurements. The phase associated with the complex-valued scalar weight, $h_{t-1}(n)$, is equal to the conditional maximum a-posteriori estimate for the random phase, ϕ_n, based on the current parameter estimates. (The MAP estimate for the phase maximizes the conditional density (7).) Thus EM is cleverly estimating the hidden random phase and is attempting to compensate for it.

The magnitude of the weight, equal to $I_1(z)/I_0(z)$, is more difficult to explain. The ratio of Bessel functions increases monotonically from zero to one as z goes from zero to infinity, and it serves to compensate for a bias that would otherwise occur. In fact it can be shown that the expectation of A_t, conditioned on A_{t-1} and on K_{t-1}, is equal to A_{t-1}. A direct proof of this fact is difficult; instead we use a result that is employed in the proof of the Cramer-Rao bound [4]: the expectation of the derivative of a log-likelihood with respect to any of its parameters is equal to zero. In particular, taking the derivative of the logarithm of the probability density (3) with respect to the real and imaginary parts of the components of A, and then taking the expectation gives

$$0 = -2K^{-1}A + 2E\left\{\frac{I_1(2|A^H K^{-1} X|)}{I_0(2|A^H K^{-1} X|)}\cdot\frac{(A^H K^{-1} X)^*}{|A^H K^{-1} X|}\cdot K^{-1} X\right\}, \qquad (12)$$

or

$$E\left\{h^* X\right\} = A. \qquad (13)$$

The covariance matrix update (11) is intuitively reasonable given the fact that

$$E\left\{xx^H\right\} = E_\phi\left\{E\left\{xx^H \mid \phi\right\}\right\} = E_\phi\left\{K + AA^H\right\} = K + AA^H. \qquad (14)$$

The covariance matrix update is guaranteed to be nonnegative definite since it can be shown to equal the following expression,

$$K_t = \frac{1}{N}\sum_{n=1}^{N}\left[\left(h^*_{t-1}(n)\cdot X_n - A\right)\left(h^*_{t-1}(n)\cdot X_n - A\right)^H + \left(1-|h_{t-1}(n)|^2\right)X_n X_n^H\right], \qquad (15)$$

which is a sum of outer products (recall that the magnitudes of the weights are less than or equal to one).

The form of the likelihood function (6) implies that only the relative phases of the components of A can be uniquely estimated (for the scalar case only the magnitude of A can be estimated uniquely): if the initial guess, A_0, produces the sequence A_t then the initial guess, $\exp\{i\theta\}A_0$, produces the sequence $\exp\{i\theta\}A_t$ however the likelihoods for the two estimates at each iteration are equal.

The EM algorithm requires an initial guess for the parameter estimates. The problem of choosing good initial guesses requires additional research. One initial guess to avoid is $A_0=0$ which results in $A_t=0$ for all t. Although $A=0$ is a stationary point for the likelihood function it is not a stable point for EM: a perturbation analysis discloses that A_t does not tend to migrate to zero.

4. CONCLUSIONS

The multivariate Rician model represents a logical generalization of the most common probabilistic model for polarimetric SAR images. Without the EM algorithm one might be inhibited from using the Rician model in the absence of any reliable algorithm for estimating the model parameters. The EM algorithm provides an explicit, computationally attractive iterative parameter estimation technique that is "safe" in the sense that convergence to at least a local maximum in likelihood is guaranteed.

Rician distributions are encountered in other problem areas such as communication through fading channels, and the EM algorithm will solve the associated parameter estimation problems.

REFERENCES

[1] H. A. Zebker & J. J. van Zyl, "Imaging Radar Polarimetry: A Review", *Proc. IEEE*, Nov. 1991.

[2] E. Rignot, R. Chellappa, & P. Dubois, "Supervised Segmentation of Polarimetric SAR Data Using the Covariance Matrix", *IEEE Trans. Geosc. Remote Sensing*, July, 1992.

[3] J. W. Goodman, *Statistical Optics*, pp. 347-351, Wiley, 1985.

[4] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, Part I, pp.360-364; p.67 Wiley, 1968.

[5] A. P. Dempster, N. M. Lair, & D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm", *J. Roy. Stat. Soc.*, ser. 39, pp. 1-38, 1977.